Σχολή Θετικών Επιστημών και Τεχνολογίας

Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα (ΠΛΣ)

Πτυχιακή / Διπλωματική Εργασία

«Μηχανική Μάθηση με Προστασία Ιδιωτικότητας στην Υγεία: Εφαρμογή DP-SGD σε Δεδομένα Ασθενών»

«ΚΩΝΣΤΑΝΤΙΝΟΣ ΖΟΥΛΙΑΤΗΣ»

Επιβλέπων καθηγητής: «ΓΕΩΡΓΙΟΣ ΦΕΡΕΤΖΑΚΗΣ»

Πάτρα, Ιούλιος 2025

ΕΛΛΗΝΙΚΟ ΑΝΟΙΚΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

«Μηχανική Μάθηση με Προστασία Ιδιωτικότητας στην Υγεία: Εφαρμογή DP-SGD σε Δεδομένα Ασθενών»

«ΚΩΝΣΤΑΝΤΙΝΟΣ ΖΟΥΛΙΑΤΗΣ»

Επιτροπή Επίβλεψης Πτυχιακής / Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής:
«ΓΕΩΡΓΙΟΣ ΦΕΡΕΤΖΑΚΗΣ»
«Μέλος Σ.Ε.Π./ΕΑΠ»

Συν-Επιβλέπων Καθηγητής:
«ΔΗΜΗΤΡΙΟΣ ΚΑΛΛΕΣ»
«Καθηγητής, Σχολη Θετικών Επιστημών και Τεχνολογίας (ΣΘΕΤ)/ΕΑΠ»

Πάτρα, Ιούλιος 2025

«Ευχαριστίες ή Αφιέρωση»

Περίληψη

Η παρούσα εργασία εξετάζει την εφαρμογή της διαφορικής ιδιωτικότητας μέσω του αλγορίθμου Differentially Private Stochastic Gradient Descent - DP-SGD στη μηχανική μάθηση για ιατρικά δεδομένα, εστιάζοντας στην επίδραση διαφορετικών υπερπαραμέτρων (learning rate, noise multiplier, clipping norm, batch size) στην απόδοση των μοντέλων. Η μελέτη βασίστηκε σε δύο σύνολα δεδομένων: το Heart Disease Dataset και το Cardiovascular Disease Dataset. Τα αποτελέσματα ανέδειξαν ότι υψηλές τιμές του πολλαπλασιαστή θορύβου, αν και ενισχύουν την ιδιωτικότητα, μειώνουν την ακρίβεια των μοντέλων. Παρομοίως, αυστηρή αποκοπή των gradients οδηγεί σε απώλεια πληροφορίας και χαμηλότερη απόδοση. Ωστόσο, μετριοπαθείς ρυθμίσεις επιτρέπουν την επίτευξη ισορροπίας μεταξύ προστασίας ιδιωτικότητας και χρησιμότητας των προβλέψεων. Η σύγκριση των δύο datasets έδειξε ότι το Cardiovascular Dataset, λόγω του μεγαλύτερου όγκου του, παρουσίασε πιο σταθερή απόδοση υπό DP-SGD, ενώ το μικρότερο Heart Dataset ήταν πιο ευάλωτο στις επιπτώσεις του θορύβου. Επιπλέον, η ανάλυση μέσω heatmaps και συγκριτικών πινάκων αποκάλυψε ότι συγκεκριμένοι συνδυασμοί learning rate και batch size επηρεάζουν καθοριστικά την απόδοση των μοντέλων. Ιδιαίτερα, η χρήση υψηλού learning rate επιδεινώνει τα αποτελέσματα, ειδικά σε μεγάλα datasets, ενώ μικρότερα rates με μικρό batch size οδηγούν σε σταθερότερη απόδοση. Η εργασία καταλήγει στο ότι η επιτυχής εφαρμογή της διαφορικής ιδιωτικότητας εξαρτάται όχι μόνο από την επιλογή του αλγορίθμου, αλλά από μια στρατηγική που συνδυάζει επάρκεια δεδομένων, σωστή παραμετροποίηση και εργαλεία ανάλυσης.

Λέξεις – Κλειδιά

Διαφορική Ιδιωτικότητα, DP-SGD Αλγόριθμος, Ιατρικά Δεδομένα Ασθενών, Προστασία Προσωπικών Δεδομένων, Ρύθμιση Υπερπαραμέτρων.

«Privacy-Preserving Machine Learning in Healthcare: Applying DP-SGD to Patient Data»

«KONSTANTINOS ZOULIATIS»

Abstract

This study examines the application of DP (DP) through the Differentially Private Stochastic Gradient Descent (DP-SGD) algorithm in machine learning for medical data, focusing on the impact of different hyperparameters (learning rate, noise multiplier, clipping norm, batch size) on model performance. The analysis was conducted using two datasets: the Heart Disease Dataset and the Cardiovascular Disease Dataset. Results showed that high noise multiplier values, although beneficial for privacy, reduced model accuracy. Similarly, aggressive gradient clipping led to information loss and lower performance. However, moderate hyperparameter settings allowed for a balance between privacy protection and predictive utility.

Comparing the two datasets, the larger Cardiovascular Dataset demonstrated more stable performance under DP-SGD, while the smaller Heart Dataset was more vulnerable to the effects of noise. Further analysis using heatmaps and comparative tables revealed that specific combinations of learning rate and batch size significantly influence model outcomes. In particular, high learning rates led to performance degradation, especially in larger datasets, whereas lower rates with smaller batch sizes yielded more consistent results. The findings suggest that the successful implementation of DP depends not only on the algorithm used, but also on a broader strategy that includes sufficient data volume, careful hyperparameter tuning, and appropriate analysis tools.

Keywords

DP, DP-SGD Algorithm, Patient Medical Data, Personal Data Protection, Hyperparameter Tuning.

Περιεχόμενα

Κατάλογος Εικόνων / Σχημάτων

Κατάλογος Πινάκων

Συντομογραφίες & Ακρωνύμια

EHRs - Electronic Health Records
GDPR - General Data Protection Regulation
HIPAA - Health Insurance Portability and Accountability Act
HITECH - Health Information Technology for Economic and Clinical Health
DP Local DP
DP-SGD - Differentially Private Stochastic Gradient Descent
SGD - Stochastic Gradient Descent
PINQ - Privacy Integrated Queries
RAPPOR - Randomized Aggregatable PrivacyPreserving Ordinal Response
OCR - Office for Civil Rights
ONC - Office of the National Coordinator for Health Information Technology
IP – Internet Protocol
SDC - Statistical Disclosure Control
LDP - Local Differential Privacy
ML – Machine Learning

ΓΚΠΔ - Γενικός Κανονισμός για την Προστασία Δεδομένων
ΠΧΠ - Προστατευόμενες Χαρακτηριστικές Πληροφορίες
ΕΠ- Επιχειρηματικοί συνεργάτες
ΚΑΟ - Καλυπτόμενες οντότητες

1. Εισαγωγή

1.1 Σχετικά με την ψηφιοποίηση των δεδομένων υγειονομικής περίθαλψης

Η ψηφιοποίηση των δεδομένων υγειονομικής περίθαλψης αναφέρεται στη μετατροπή των παραδοσιακών, έντυπων ιατρικών αρχείων και άλλων πληροφοριών υγείας σε ψηφιακή μορφή. Αυτό περιλαμβάνει τη χρήση ηλεκτρονικών ιατρικών αρχείων (Electronic Health Records - EHRs) για τη συλλογή, αποθήκευση και διαχείριση δεδομένων ασθενών, όπως ιστορικά υγείας, αποτελέσματα εργαστηριακών εξετάσεων, και συνταγογραφημένες θεραπείες, σε ένα ενιαίο ψηφιακό περιβάλλον. Η ψηφιοποίηση επιτρέπει την ευκολότερη πρόσβαση, ανάλυση και ανταλλαγή των δεδομένων υγείας μεταξύ διαφορετικών παρόχων υγειονομικής περίθαλψης, βελτιώνοντας έτσι την ακρίβεια, την αποτελεσματικότητα και τη διαφάνεια στη φροντίδα των ασθενών (Boonstra & Broekhuis, 2010; AdlerMilstein et al., 2017). Η ψηφιοποίηση των δεδομένων υγείας προσφέρει πολλαπλά οφέλη όπως βελτιωμένη πρόσβαση αφού οι πληροφορίες μπορούν να είναι διαθέσιμες στους παρόχους υγειονομικής περίθαλψης σε πραγματικό χρόνο, διευκολύνοντας την καλύτερη συνεργασία και τον συντονισμό της φροντίδας (AdlerMilstein et al., 2017). Η χρήση ηλεκτρονικών αρχείων μειώνει τα λάθη που προκύπτουν από τη δυσανάγνωστη χειρόγραφη γραφή και επιτρέπει τη διασφάλιση της ακρίβειας των πληροφοριών (Boonstra & Broekhuis, 2010). Με τη συλλογή μεγάλων ποσοτήτων δεδομένων, είναι δυνατή η χρήση τεχνικών μηχανικής μάθησης και ανάλυσης δεδομένων για την εξαγωγή χρήσιμων πληροφοριών που βοηθούν στην πρόληψη και διαχείριση ασθενειών (Raghupathi & Raghupathi, 2014).

Η ψηφιοποίηση των δεδομένων υγειονομικής περίθαλψης αποτελεί μια από τις σημαντικότερες τεχνολογικές εξελίξεις των τελευταίων δεκαετιών. Στο παρελθόν, τα πάντα στην υγειονομική περίθαλψη γίνονταν χειροκίνητα. Οι φάκελοι των ασθενών αποθηκεύονταν σε φυσικά αρχεία, τα δεδομένα διατηρούνταν σε γιγαντιαίες αποθήκες, και η πρόσβαση σε αυτά γινόταν μόνο με φυσική παρουσία στα νοσοκομεία. Αυτό δημιουργούσε έναν τεράστιο φόρτο εργασίας τόσο για τις υγειονομικές αρχές όσο και για τις ρυθμιστικές αρχές, καθιστώντας δύσκολη την πρόσβαση και διαχείριση των πληροφοριών ασθενών (Blumenthal & Tavenner, 2010).

Με την πρόοδο της τεχνολογίας, ιδίως με την ανάπτυξη του διαδικτύου και των κινητών συσκευών, η υγειονομική περίθαλψη εισήλθε σε μια νέα εποχή. Η χρήση εργαλείων όπως τα ηλεκτρονικά ραντεβού, τα ψηφιακά αρχεία υγείας, και η ασφαλής αποθήκευση δεδομένων έχουν φέρει επανάσταση στην επικοινωνία μεταξύ ασθενών και παρόχων υγειονομικής περίθαλψης (Menachemi & Collum, 2011). Σήμερα, ανεξάρτητα από το αν κάποιος ζει σε ένα απομακρυσμένο χωριό ή σε μια μεγάλη πόλη, τα ψηφιακά εργαλεία διευκολύνουν την πρόσβαση στις υγειονομικές υπηρεσίες. Κατά τη διάρκεια της πανδημίας COVID19, η ψηφιοποίηση αποδείχθηκε κρίσιμη, επιτρέποντας στους ασθενείς να λαμβάνουν συνταγές, ραντεβού και εργαστηριακά αποτελέσματα εξ αποστάσεως. Οι ψηφιακές πλατφόρμες μείωσαν την ανάγκη για φυσική παρουσία και, συνεπώς, τις μεγάλες ουρές στα νοσοκομεία και τα κέντρα υγείας, κάτι που συνέβαλε στην προστασία της δημόσιας υγείας (Iyengar et al., 2020).

Επιπλέον, η διαλειτουργική τεχνολογία μεταξύ ιατρών και ασθενών έχει βελτιώσει σημαντικά τη διάγνωση και τη θεραπεία. Η γρήγορη και ακριβής πρόσβαση σε ψηφιακά αρχεία και ιατρικές πληροφορίες μπορεί να σώσει ζωές, επιτρέποντας στους ιατρούς να λαμβάνουν αποφάσεις με βάση πλήρη και ενημερωμένα δεδομένα. Αυτό όχι μόνο εξοικονομεί χρόνο, αλλά αυξάνει και την ακρίβεια της περίθαλψης, καθιστώντας την

ψηφιοποίηση ζωτικής σημασίας για την υγειονομική περίθαλψη (Raghupathi & Raghupathi, 2014).

Η ψηφιοποίηση των δεδομένων υγείας αφορά τη μετατροπή των χειρόγραφων και έντυπων ιατρικών αρχείων σε ηλεκτρονικά δεδομένα, που μπορούν να διαχειριστούν και να αποθηκευτούν σε ψηφιακά συστήματα, όπως τα ψηφιακά ιατρικά αρχεία. Η ψηφιοποίηση αυτή επιτρέπει την καλύτερη διαχείριση των δεδομένων ασθενών, βελτιώνει την ακρίβεια και την αποδοτικότητα στην παροχή υγειονομικών υπηρεσιών και επιτρέπει την απομακρυσμένη πρόσβαση σε ιατρικές πληροφορίες (Blumenthal & Tavenner, 2010). Ο Manca, 2015 αναφέρει ότι η ψηφιοποίηση των δεδομένων υγείας αφορά τη διαδικασία με την οποία οι πληροφορίες που αφορούν την υγεία και τις ιατρικές διαδικασίες συλλέγονται, αποθηκεύονται και ανταλλάσσονται σε ηλεκτρονική μορφή. Αυτό βελτιώνει την ποιότητα της υγειονομικής περίθαλψης, επιτρέπει την ανάλυση δεδομένων και υποστηρίζει την τεκμηριωμένη ιατρική πρακτική. Τα ψηφιακά αρχεία υγείας αποτελούν βασικό στοιχείο της ψηφιοποίησης των δεδομένων υγείας, καθώς επιτρέπουν την καταγραφή, αποθήκευση και διαχείριση πληροφοριών ασθενών σε ένα ηλεκτρονικό σύστημα. Η μετάβαση από τα χειρόγραφα σε ψηφιακά δεδομένα έχει βελτιώσει την ποιότητα της φροντίδας και έχει επιτρέψει τη διευκόλυνση της ανταλλαγής πληροφοριών μεταξύ παρόχων (Menachemi & Collum, 2011).

Η ψηφιοποίηση των δεδομένων υγείας είναι μια διαδικασία που έχει εξελιχθεί σταδιακά κατά τις τελευταίες δεκαετίες, με τη ραγδαία ανάπτυξη της τεχνολογίας και των ηλεκτρονικών συστημάτων διαχείρισης δεδομένων. Η ανάγκη για καλύτερη διαχείριση των πληροφοριών ασθενών οδήγησε στην υιοθέτηση ηλεκτρονικών συστημάτων, κυρίως μέσω των ψηφιακών ιατρικών αρχείων. Η ιδέα της ψηφιοποίησης των δεδομένων υγείας χρονολογείται από τη δεκαετία του 1960, όταν μεγάλα νοσοκομεία και ερευνητικά κέντρα άρχισαν να χρησιμοποιούν πρώιμα υπολογιστικά συστήματα για την αποθήκευση και διαχείριση βασικών πληροφοριών ασθενών. Κατά τη διάρκεια αυτής της περιόδου, οι πρώτες μορφές ηλεκτρονικών αρχείων υγείας ήταν σχετικά απλές και χρησιμοποιούνταν κυρίως για την καταγραφή διοικητικών πληροφοριών και για ερευνητικούς σκοπούς (Collen, 1995).

Στη δεκαετία του 1980, τα πρώτα υπολογιστικά συστήματα για την αυτοματοποίηση της αποθήκευσης δεδομένων υγείας άρχισαν να αναπτύσσονται. Ωστόσο, η υιοθέτηση αυτών των συστημάτων ήταν περιορισμένη, λόγω των υψηλών εξόδων και των τεχνολογικών περιορισμών της εποχής (Collen, 1995). Η πραγματική ώθηση για την ψηφιοποίηση των δεδομένων υγείας ήρθε τη δεκαετία του 1990, όταν η ταχεία ανάπτυξη των υπολογιστών και του διαδικτύου έφερε επαναστατικές αλλαγές στη διαχείριση πληροφοριών. Τα ψηφιακά αρχεία υγείας άρχισαν να καθιερώνονται ευρύτερα, επιτρέποντας στα νοσοκομεία και τους γιατρούς να καταγράφουν, αποθηκεύουν και διαχειρίζονται δεδομένα ασθενών ψηφιακά (Menachemi & Collum, 2011). Η μεγαλύτερη υιοθέτηση των ψηφιακών αρχείων υγείας συνέβη κατά τη δεκαετία του 2000, κυρίως χάρη στη βελτίωση των τεχνολογικών υποδομών και των αυξανόμενων πιέσεων για τη βελτίωση της ποιότητας φροντίδας μέσω της τεχνολογίας (Blumenthal & Tavenner, 2010). Το 2004, οι Ηνωμένες Πολιτείες ίδρυσαν το Γραφείο Εθνικού Συντονιστή για την Υγεία Πληροφοριών (Office of the National Coordinator for Health Information Technology - ONC) για να επιταχύνουν την υιοθέτηση των ψηφιακών αρχείων υγείας και να υποστηρίξουν την ψηφιοποίηση των δεδομένων υγείας (Blumenthal, 2009).

Το 2009, με την εισαγωγή του νόμου Health Information Technology for Economic and Clinical Health Act (HITECH Act) στις ΗΠΑ, δόθηκε περαιτέρω ώθηση στη χρήση των ψηφιακών αρχείων υγείας μέσω οικονομικών κινήτρων για τους παρόχους υγειονομικής

περίθαλψης. Η αρχή της "Ουσιαστικής Χρήσης" (Meaningful Use) διαμόρφωσε το πλαίσιο για την ψηφιοποίηση, απαιτώντας από τους γιατρούς και τα νοσοκομεία να υιοθετήσουν τεχνολογίες που βελτιώνουν την ποιότητα της φροντίδας, την ασφάλεια και την αποτελεσματικότητα των υπηρεσιών υγείας (Blumenthal & Tavenner, 2010). Αυτή η φάση συνδέεται με τη μαζική εισαγωγή ηλεκτρονικών συστημάτων υγείας σε πολλά συστήματα υγειονομικής περίθαλψης παγκοσμίως. Η χρήση ηλεκτρονικών συστημάτων άρχισε να προσφέρει τη δυνατότητα ανάλυσης μεγάλων δεδομένων (Big Data) για την υποστήριξη των κλινικών αποφάσεων και τη βελτίωση της φροντίδας των ασθενών (Raghupathi & Raghupathi, 2014). Η πανδημία COVID19 τόνισε την ανάγκη για γρήγορη πρόσβαση σε ψηφιακά δεδομένα υγείας και ώθησε ακόμη περισσότερο την υιοθέτηση της ψηφιοποίησης και της τηλεϊατρικής. Η πανδημία ανάγκασε τα συστήματα υγείας να επιταχύνουν τη χρήση των ψηφιακών αρχείων υγείας και άλλων ψηφιακών εργαλείων για να διαχειριστούν την αύξηση των ασθενών και να διασφαλίσουν τη συνεχιζόμενη φροντίδα εξ αποστάσεως (Iyengar et al., 2020). Αυτή η περίοδος οδήγησε σε πρωτοφανή επίπεδα ψηφιακής υιοθέτησης και έδειξε πόσο σημαντική είναι η ψηφιοποίηση για την αντιμετώπιση παγκόσμιων υγειονομικών κρίσεων. Η ψηφιοποίηση των δεδομένων υγείας έχει εξελιχθεί δραστικά τις τελευταίες δεκαετίες, με σημαντική πρόοδο στην τεχνολογία, υποδομές και ρυθμίσεις πολιτικής. Από την αρχική χρήση απλών ηλεκτρονικών συστημάτων αποθήκευσης πληροφοριών ασθενών, έχει εξελιχθεί σε μια βασική διαδικασία που ενισχύει την ποιότητα της φροντίδας, βελτιώνει την ασφάλεια των ασθενών και επιτρέπει την ευρεία ανάλυση των δεδομένων για τη βελτίωση της υγειονομικής φροντίδας σε παγκόσμιο επίπεδο.

1.2 Η σημασία της ιδιωτικής ζωής των ασθενών και ο αντίκτυπος των παραβιάσεων δεδομένων.

Η ιδιωτικότητα των ασθενών αποτελεί βασικό πυλώνα της υγειονομικής περίθαλψης και αφορά τη διατήρηση της εμπιστευτικότητας των προσωπικών και ιατρικών πληροφοριών τους. Η προστασία αυτών των δεδομένων είναι κρίσιμη για την ενίσχυση της εμπιστοσύνης των ασθενών στους παρόχους υγειονομικής περίθαλψης και για την αποφυγή καταχρήσεων ή παραβιάσεων που μπορούν να έχουν σοβαρές συνέπειες. Με την αυξανόμενη χρήση της τεχνολογίας και την ψηφιοποίηση των ιατρικών αρχείων, η προστασία της ιδιωτικότητας των ασθενών έχει γίνει ακόμα πιο σημαντική. Η ιδιωτικότητα αποτελεί θεμελιώδη αρχή στη σχέση μεταξύ ασθενούς και γιατρού. Οι ασθενείς πρέπει να αισθάνονται ότι οι πληροφορίες τους είναι ασφαλείς για να μοιράζονται πλήρως και ειλικρινά τα ιατρικά τους δεδομένα. Εάν οι ασθενείς ανησυχούν για τη διαρροή ή κακή χρήση των δεδομένων τους, μπορεί να μην αποκαλύψουν σημαντικές πληροφορίες, οδηγώντας σε ελλιπή διάγνωση ή ακατάλληλη θεραπεία (Gostin, 1995).

Διάφορες νομοθεσίες, όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων - ΓΚΠΔ (General Data Protection Regulation - GDPR) στην Ευρωπαϊκή Ένωση και ο Health Insurance Portability and Accountability Act (HIPAA) στις ΗΠΑ, έχουν τεθεί σε εφαρμογή για να διασφαλίσουν ότι οι προσωπικές και ιατρικές πληροφορίες προστατεύονται από μη εξουσιοδοτημένη πρόσβαση. Η συμμόρφωση με αυτές τις νομοθεσίες είναι ουσιαστική για την αποφυγή νομικών επιπτώσεων και την προάσπιση της ασφάλειας των δεδομένων των ασθενών (Appari & Johnson, 2010). Η δημοσιοποίηση ευαίσθητων ιατρικών πληροφοριών μπορεί να οδηγήσει σε συναισθηματική και ψυχολογική δυσφορία για τους ασθενείς. Για παράδειγμα, η αποκάλυψη πληροφοριών για χρόνιες ασθένειες ή ψυχιατρικές παθήσεις μπορεί να οδηγήσει σε στιγματισμό, διακρίσεις ή κοινωνικό αποκλεισμό (Gostin, 1995).

Οι παραβιάσεις δεδομένων μπορούν να καταστρέψουν την εμπιστοσύνη που οι ασθενείς έχουν στους παρόχους υγείας. Οι ασθενείς που έχουν εκτεθεί σε παραβιάσεις μπορεί να γίνουν πιο επιφυλακτικοί να μοιραστούν πληροφορίες στο μέλλον, γεγονός που μπορεί να επηρεάσει αρνητικά την ποιότητα της περίθαλψης (Ponemon Institute, 2016). Συχνά οδηγούν σε σημαντικές οικονομικές επιπτώσεις για τους παρόχους υγείας, είτε μέσω άμεσων προστίμων λόγω μη συμμόρφωσης με τη νομοθεσία, είτε μέσω του κόστους αντιμετώπισης των συνεπειών, όπως οι ειδοποιήσεις ασθενών, οι νομικές υποθέσεις, και οι βελτιώσεις ασφαλείας (Schweitzer, 2019). Με την αυξημένη χρήση ψηφιακών εργαλείων, τα δεδομένα γίνονται πιο ευάλωτα σε κυβερνοεπιθέσεις. Οι παραβιάσεις μπορεί να προκύψουν μέσω κακόβουλων λογισμικών, ανεπαρκών πρακτικών ασφαλείας ή ακόμα και ανθρώπινων λαθών, καθιστώντας απαραίτητη την ενίσχυση των τεχνολογικών συστημάτων και των πρωτοκόλλων ασφαλείας (Schweitzer, 2019).

1.3 Επισκόπηση των κανονιστικών απαιτήσεων (GDPR, HIPAA).

Ο νόμος του 1996 περί φορητότητας και λογοδοσίας της ασφάλισης υγείας (HIPAA), όπως τροποποιήθηκε από τον νόμο HITECH (Health Information Technology for Economic and Clinical Health), προστατεύει το απόρρητο και την ασφάλεια των ατομικών πληροφοριών υγείας που χρησιμοποιούνται, διαβιβάζονται και διατηρούνται για την παροχή και την πληρωμή υπηρεσιών υγείας. Οι πληροφορίες αυτές είναι γνωστές ως προστατευόμενες πληροφορίες υγείας (Protected Health Information - PHI). Ο HIPAA ισχύει για όλα τα ασφαλιστικά προγράμματα υγείας και τους εκκαθαριστές χρεώσεων υγειονομικής περίθαλψης, καθώς και για τους παρόχους υγειονομικής περίθαλψης (όπως νοσοκομεία, ιατροί και κλινικές) που τιμολογούν ηλεκτρονικά τους ασφαλιστές. Εφαρμόζεται επίσης σε τρίτους, όπως δικηγόρους, συμβούλους, ελεγκτές και άλλους παρόχους υπηρεσιών, οι οποίοι έχουν πρόσβαση στο PHI ενός καλυπτόμενου φορέα για την παροχή υπηρεσιών.
Οι κανόνες και οι κανονισμοί του HIPAA ισχύουν για όλες τις καλυπτόμενες οντότητες, τα σχέδια υγείας, τους παρόχους υγειονομικής περίθαλψης και τα κέντρα εκκαθάρισης υγειονομικής περίθαλψης που διαβιβάζουν πληροφορίες υγείας σε ηλεκτρονική, προφορική ή γραπτή μορφή. Εφαρμόζονται επίσης στους επιχειρηματικούς συνεργάτες των καλυπτόμενων οντοτήτων, δηλαδή σε άτομα ή οργανισμούς που έχουν συμβληθεί για την παροχή υπηρεσιών αλλά δεν ανήκουν στο εργατικό δυναμικό της καλυπτόμενης οντότητας.
Ο κανόνας περί απορρήτου είναι κάπως ευρύτερος από τον κανόνα περί ασφάλειας, δεδομένου ότι προστατεύει όλες τις «ατομικά αναγνωρίσιμες πληροφορίες υγείας» που είτε διαβιβάζονται είτε κατέχονται από μια καλυπτόμενη οντότητα ή τον επιχειρηματικό συνεργάτη της, σε οποιαδήποτε μορφή ή μέσο ηλεκτρονική, έντυπη ή προφορική. Αυτές οι προστατευόμενες πληροφορίες υγείας περιλαμβάνουν πληροφορίες που σχετίζονται με τη σωματική ή ψυχική υγεία ή κατάσταση του ατόμου, την υγειονομική περίθαλψη που παρέχεται στο άτομο ή την πληρωμή για την παροχή υγειονομικής περίθαλψης στο άτομο. Τα PHI περιλαμβάνουν επίσης βασικές πληροφορίες ταυτοποίησης, όπως το όνομα του ασθενούς, την ημερομηνία γέννησής του, τον SSN και τη διεύθυνση κατοικίας του. Προκειμένου να ενθαρρυνθεί η έρευνα στον τομέα της υγειονομικής περίθαλψης, ο κανόνας περί απορρήτου δεν θέτει περιορισμούς στη χρήση ή τη διαβίβαση των απροσδιόριστων πληροφοριών υγείας.
Ο HIPAA ορίζει ως PHI κάθε πληροφορία υγείας που δημιουργείται, λαμβάνεται ή διατηρείται από καλυπτόμενη οντότητα του HIPAA, είτε σε έντυπη, προφορική ή ηλεκτρονική μορφή, τα οποία περιλαμβάνουν «ατομικά αναγνωριστικά» που ταυτοποιούν ένα άτομο (ή έχουν στοιχεία που θα μπορούσαν να χρησιμοποιηθούν για την ταυτοποίηση του ατόμου) και σχετίζονται με παρελθούσα, παρούσα ή μελλοντική κατάσταση σωματικής

ή ψυχικής υγείας ή με την παροχή ή την πληρωμή υγειονομικής περίθαλψης ή γενετικών πληροφοριών. Το PHI περιλαμβάνει ουσιαστικά κάθε πληροφορία από τα ιατρικά αρχεία ενός ατόμου, τα αρχεία πληρωμής για ιατρικές υπηρεσίες που πραγματοποιούνται από ή για λογαριασμό ενός ατόμου (συμπεριλαμβανομένων των ασφαλιστικών απαιτήσεων και των επιστροφών), ή πληροφορίες που παρέχονται σε πάροχο υγειονομικής περίθαλψης σχετικά με τη σωματική ή ψυχική υγεία ή κατάσταση ενός ατόμου, οι οποίες ταυτοποιούν ή μπορούν να χρησιμοποιηθούν για την ταυτοποίηση του ατόμου.
Ο ΓΚΠΔ τέθηκε σε ισχύ το 2018 και ο πρωταρχικός του σκοπός είναι να δημιουργήσει ένα ενιαίο συνεκτικό πλαίσιο προστασίας δεδομένων σε ολόκληρη την Ευρωπαική Ένωση και ενισχύει σημαντικά την προστασία των δεδομένων και τα δικαιώματα προστασίας της ιδιωτικής ζωής για τα πρόσωπα. Επιβάλλει ένα ολοκληρωμένο σύνολο αρχών και υποχρεώσεων με τις οποίες πρέπει να συμμορφώνονται πολλοί οργανισμοί που δραστηριοποιούνται ή προσφέρουν προϊόντα και υπηρεσίες.
Ο ΓΚΠΔ αποτελεί ένα από τα υψηλότερα πρότυπα προστασίας της ιδιωτικής ζωής και των δεδομένων στον κόσμο και θα παρέχει στις Αρχές Προστασίας Δεδομένων της Ευρωπαικής Ένωσης τη δυνατότητα να ρυθμίζουν και να επιβάλλουν κυρώσεις κατά εταιρειών σε ολόκληρο τον κόσμο. Ο ΓΚΠΔ εφαρμόζεται σε κάθε εταιρεία που συλλέγει δεδομένα προσωπικού χαρακτήρα από υποκείμενα των δεδομένων της, ανεξάρτητα από τον τόπο εγκατάστασης της εταιρείας. Ισχύει τόσο για τους εκτελούντες την επεξεργασία δεδομένων όσο και για τους υπευθύνους επεξεργασίας δεδομένων. Μία από τις μεγαλύτερες και πιο σημαντικές αλλαγές του ΓΚΠΔ είναι η εξωεδαφική ρυθμιστική εποπτεία και η εμβέλεια επιβολής που δίνει στις ρυθμιστικές αρχές την εξουσία να ρυθμίζουν και να ασκούν αγωγές επιβολής κατά οποιασδήποτε εταιρείας που χειρίζεται προσωπικά δεδομένα (εργαζόμενων/καταναλωτών/επιχειρησιακών επαφών) κατοίκων της Ευρωπαικής Ένωσης, ανεξάρτητα από τον τόπο όπου η εν λόγω εταιρεία έχει την έδρα της ή τον τόπο επεξεργασίας των δεδομένων.

1.4 Εισαγωγή στην μεθοδολογία DP και ο αλγόριθμος DP-SGD

Η μεθοδολογία Differential Privacy (DP) στοχεύει στην προστασία της ιδιωτικότητας των ατόμων όταν τα δεδομένα τους χρησιμοποιούνται για ανάλυση. Σε ένα σύστημα που εφαρμόζει την μεθοδολογία DP, το αποτέλεσμα μιας ερώτησης ή ανάλυσης πάνω στα δεδομένα δεν θα αποκαλύπτει πληροφορίες που θα επέτρεπαν την αναγνώριση συγκεκριμένων ατόμων. Αυτό επιτυγχάνεται μέσω της προσθήκης θορύβου στα δεδομένα ή τα αποτελέσματα ώστε να διασφαλιστεί ότι οι απαντήσεις δεν είναι ακριβώς αντιπροσωπευτικές για συγκεκριμένα άτομα αλλά επαρκώς χρήσιμες σε επίπεδο πληθυσμού. Βασίζεται στην ιδέα ότι η συμμετοχή ή η απουσία ενός ατόμου σε ένα σύνολο δεδομένων δεν πρέπει να αλλάζει σημαντικά το αποτέλεσμα μιας ερώτησης. Δηλαδή, η προσθήκη ή η αφαίρεση των δεδομένων ενός ατόμου δεν θα πρέπει να επηρεάζει υπερβολικά τις παραγόμενες στατιστικές ή προβλέψεις. Είναι ένα μαθηματικό πλαίσιο για τη διασφάλιση της ιδιωτικότητας των ατόμων σε σύνολα δεδομένων. Μπορεί να παρέχει ισχυρή εγγύηση της ιδιωτικότητας, επιτρέποντας στους αναλυτές να εξετάζουν δεδομένα χωρίς να αποκαλύπτουν ευαίσθητες πληροφορίες για οποιοδήποτε άτομο στο σύνολο δεδομένων. Οι επαγγελματίες χρησιμοποιούν αυτή τη μέθοδο καθώς αποτρέπει τις επιθέσεις σύνδεσης, καθιστώντας την καλή επιλογή για την προστασία μεμονωμένων δεδομένων σε διάφορα σενάρια. Αυτή η διασφάλιση της ιδιωτικότητας είναι ζωτικής σημασίας σε ερευνητικά πλαίσια όπου ηθικοί προβληματισμοί έχουν σημασία. Ενώ η DP έχει καλές πλευρές, εισάγει θόρυβο για την προστασία της ατομικής ιδιωτικής ζωής, ο οποίος μπορεί να μειώσει την ακρίβεια των αποτελεσμάτων και να αποκρύψει σημαντικά

μοτίβα, ιδίως σε εφαρμογές δεδομένων υψηλής πιστότητας. Επιπλέον, αντιμετωπίζει προκλήσεις κλιμάκωσης με μεγάλα σύνολα δεδομένων και απαιτεί μια λεπτή ισορροπία μεταξύ της ιδιωτικότητας και της χρησιμότητας των δεδομένων, γεγονός που την καθιστά πιο αποτελεσματική όταν ενσωματώνεται με άλλες τεχνικές.

Ο αλγόριθμος για την διαφορικά ιδιωτική στοχαστική κάθοδος κλίσης (Differentially Private Stochastic Gradient Descent, DP-SGD) είναι μια παραλλαγή του κλασικού αλγορίθμου Stochastic Gradient Descent (SGD) που χρησιμοποιείται εκτενώς στην εκπαίδευση μοντέλων μηχανικής μάθησης. Η κύρια διαφορά του DPSGD είναι ότι εφαρμόζει τεχνικές DP κατά τη διαδικασία εκπαίδευσης, για να εξασφαλίσει ότι τα δεδομένα που χρησιμοποιούνται παραμένουν ιδιωτικά. Αυτό επιτρέπει στα μοντέλα να εκπαιδεύονται με δεδομένα που περιέχουν προσωπικές πληροφορίες, χωρίς τον κίνδυνο να αποκαλύπτονται ευαίσθητες πληροφορίες. Ο αλγόριθμος DP-SGD ενσωματώνει τεχνικές DP κατά την εκπαίδευση ενός μοντέλου, προσθέτοντας θόρυβο στις διαβαθμίσεις που υπολογίζονται κατά την προσαρμογή των παραμέτρων του μοντέλου. Έτσι, αποτρέπει την "προσαρμογή" του μοντέλου σε συγκεκριμένα δεδομένα ενός ατόμου και διασφαλίζει ότι τα ατομικά δεδομένα παραμένουν κρυφά. Αυτό είναι ιδιαίτερα σημαντικό σε περιπτώσεις όπου τα δεδομένα μπορεί να είναι ευαίσθητα, όπως στην υγειονομική περίθαλψη ή τα οικονομικά δεδομένα.

Ο αλγόριθμος SGD είναι μια από τις πιο διαδεδομένες μεθόδους εκπαίδευσης μηχανικών μοντέλων. Σε κάθε βήμα εκπαίδευσης, υπολογίζεται η κλίση της συνάρτησης απώλειας για ένα μικρό υποσύνολο δεδομένων, και στη συνέχεια οι παράμετροι του μοντέλου ενημερώνονται έτσι ώστε να μειωθεί η απώλεια. Η διαδικασία επαναλαμβάνεται μέχρι το μοντέλο να έχει εκπαιδευτεί επαρκώς. Ωστόσο, η διαδικασία του SGD δεν διασφαλίζει από μόνη της την ιδιωτικότητα των δεδομένων. Αν κάποιος αναλύσει τις διαβαθμίσεις που υπολογίζονται, μπορεί να ανακτήσει πληροφορίες για τα αρχικά δεδομένα, θέτοντας σε κίνδυνο την ιδιωτικότητα των συμμετεχόντων στο dataset. Η DP προσθέτει ασφάλεια στα δεδομένα με την προσθήκη θορύβου κατά τη διαδικασία εκπαίδευσης. Συγκεκριμένα, στο DP-SGD κάθε κλίση που υπολογίζεται από το μοντέλο περιρίζεται σε ένα προκαθορισμένο εύρος. Αυτό διασφαλίζει ότι καμία κλίση δεν θα είναι υπερβολικά μεγάλη, κάτι που θα μπορούσε να οδηγήσει σε αποκαλύψεις για τα δεδομένα. Στη συνέχεια, προστίθεται θόρυβος σε κάθε περιορισμένη κλίση από μια κατανομή Gaussian. Αυτός ο θόρυβος διασφαλίζει ότι τα δεδομένα ενός ατόμου δεν θα επηρεάσουν υπερβολικά το μοντέλο, καθώς τα αποτελέσματα θα είναι "θορυβώδη" και όχι ακριβή. Τέλος, οι παράμετροι του μοντέλου ενημερώνονται χρησιμοποιώντας τις θορυβώδεις διαβαθμίσεις. Αυτό σημαίνει ότι το μοντέλο εκπαιδεύεται με βάση τα δεδομένα, αλλά τα ακριβή δεδομένα κάθε ατόμου παραμένουν ιδιωτικά.

Η διαδικασία του DP-SGD περιλαμβάνει τα εξής βήματα:

- Επιλογή Batch Δεδομένων: Επιλέγεται ένα μικρό υποσύνολο δεδομένων από το dataset.
- Υπολογισμός των διαβαθμίσεων: Υπολογίζονται οι διαβαθμίσεις των παραμέτρων του μοντέλου για το συγκεκριμένο υποσύνολο δεδομένων.
- Περιορισμός των διαβαθμίσεων: Οι διαβαθμίσεις περιορίζονται σε ένα μέγιστο επιτρεπτό μέγεθος C για να διασφαλιστεί ότι καμία κλίση δεν θα έχει υπερβολικά μεγάλη επιρροή.
- Προσθήκη Θορύβου: Προστίθεται Gaussian θόρυβος στις περιορισμένες διαβαθμίσεις, εξασφαλίζοντας ότι οι αλλαγές στις παραμέτρους δεν βασίζονται σε ακριβή δεδομένα αλλά σε θορυβώδεις εκτιμήσεις.

- Ενημέρωση Παραμέτρων: Οι παράμετροι του μοντέλου ενημερώνονται με βάση τις θορυβώδεις διαβαθμίσεις.
- Αξιολόγηση EpsilonPrivacy Budget: Παρακολουθείται η συνολική "κατανάλωση" του ebudget, το οποίο δείχνει πόσο ιδιωτική είναι η συνολική διαδικασία εκπαίδευσης. Το ε είναι ένας παράγοντας που μετρά τον βαθμό ιδιωτικότητας, με χαμηλές τιμές να αντιστοιχούν σε υψηλότερη ιδιωτικότητα.

Το DP-SGD είναι ένας αλγόριθμος εκπαίδευσης μηχανικής μάθησης που συνδυάζει την ισχυρή απόδοση του SGD με την ασφάλεια της DP. Μέσω της διαδικασίας αυτής, τα μοντέλα μπορούν να εκπαιδευτούν με δεδομένα που περιέχουν προσωπικές ή ευαίσθητες πληροφορίες, χωρίς να διακινδυνεύουν την αποκάλυψη των δεδομένων αυτών. Παρά τις προκλήσεις που παρουσιάζει στην απόδοση του μοντέλου και το υπολογιστικό κόστος, η χρήση του καθίσταται ολοένα και πιο σημαντική καθώς η ανάγκη για ιδιωτικότητα αυξάνεται σε διάφορους τομείς, όπως η υγεία και τα χρηματοοικονομικά δεδομένα.

1.5Στόχοι της διατριβής και ερευνητικά ερωτήματα.

Η ψηφιοποίηση των δεδομένων υγείας έχει αυξήσει την ανάγκη για προστασία της ιδιωτικότητας των ασθενών, με ρυθμιστικά πλαίσια όπως το GDPR και το HIPAA να απαιτούν αυστηρά μέτρα προστασίας προσωπικών πληροφοριών. Ο αλγόριθμος DP-SGD προσφέρει ισχυρές εγγυήσεις ιδιωτικότητας κατά την εκπαίδευση μοντέλων μηχανικής μάθησης, περιορίζοντας τη διαρροή ευαίσθητων δεδομένων. Αυτό τον καθιστά μια προηγμένη λύση για την ασφάλεια στην υγειονομική περίθαλψη, συνδυάζοντας αποτελεσματική εκπαίδευση μοντέλων με προστασία της ιδιωτικότητας. Η παρούσα διπλωματική εργασία έχει σαν στόχους:
1. Ανάλυση Επίδρασης του DP-SGD στην Απόδοση Μοντέλων: Να μελετηθεί ο τρόπος με τον οποίο ο αλγόριθμος επηρεάζει την ακρίβεια και τη σύγκλιση των μοντέλων μηχανικής μάθησης σε σύγκριση με μεθόδους χωρίς εγγυήσεις ιδιωτικότητας.
2. Βελτιστοποίηση Υπερπαραμέτρων του DP-SGD: Να εντοπιστούν οι βέλτιστες τιμές των υπερπαραμέτρων του DP-SGD, όπως το επίπεδο θορύβου και το μέγεθος κλιπ, ώστε να εξισορροπηθεί η ιδιωτικότητα με την απόδοση του μοντέλου.
3. Διερεύνηση Ισορροπίας Ιδιωτικότητας και Αποτελεσματικότητας: Να αξιολογηθεί πώς μπορεί να επιτευχθεί κατάλληλη ισορροπία μεταξύ της ιδιωτικότητας των δεδομένων υγειονομικής περίθαλψης και της αποτελεσματικότητας των μοντέλων, λαμβάνοντας υπόψη τις απαιτήσεις ρυθμιστικών πλαισίων όπως το GDPR και το HIPAA.
4. Συγκριτική Αξιολόγηση με Εναλλακτικές Μεθόδους: Να συγκριθεί ο αλγόριθμος DP-SGD με άλλες τεχνικές διασφάλισης ιδιωτικότητας, όπως η ανωνυμοποίηση και η ψευδωνυμοποίηση, σε όρους απόδοσης και ιδιωτικότητας.
5. Προτάσεις για Βέλτιστες Πρακτικές στην Υγειονομική Περίθαλψη: Να προταθούν βέλτιστες πρακτικές για την εφαρμογή του DP-SGD σε πραγματικά σενάρια υγειονομικής περίθαλψης, εξασφαλίζοντας συμμόρφωση με τα ρυθμιστικά πλαίσια και αποτελεσματική χρήση των δεδομένων.
Τα ερευνητικά ερωτήματα που θα εξεταστούν είναι:

1. Πώς επηρεάζει το DPSGD την ακρίβεια και τη σύγκλιση των μοντέλων μηχανικής μάθησης;
2. Ποιοι είναι οι βέλτιστοι υπερπαράμετροι για το DP-SGD για να επιτευχθεί κατάλληλη ισορροπία μεταξύ ιδιωτικότητας και απόδοσης μοντέλου;

2 Βιβλιογραφική Ανασκόπηση

2.1 Επισκόπηση του διαφορικού απορρήτου

Ο όρος DP είναι ένας ισχυρός, μαθηματικός ορισμός της ιδιωτικότητας στο πλαίσιο της στατιστικής ανάλυσης και της μηχανικής μάθησης. Σύμφωνα με αυτόν τον μαθηματικό ορισμό, η DP είναι ένα κριτήριο προστασίας της ιδιωτικής ζωής, για την ικανοποίηση του οποίου έχουν επινοηθεί πολλά εργαλεία ανάλυσης ευαίσθητων προσωπικών πληροφοριών. Η μεθοδολογία εγγυάται μαθηματικά ότι οποιοσδήποτε δει το αποτέλεσμα μιας διαφορικά ιδιωτικής ανάλυσης θα βγάλει ουσιαστικά το ίδιο συμπέρασμα για τις ιδιωτικές πληροφορίες οποιουδήποτε ατόμου, ανεξάρτητα από το αν οι ιδιωτικές πληροφορίες αυτού του ατόμου περιλαμβάνονται ή όχι στην είσοδο της ανάλυσης. Η μεθοδολογία παρέχει μια μαθηματικά αποδεδειγμένη εγγύηση προστασίας της ιδιωτικής ζωής από ένα ευρύ φάσμα επιθέσεων στην ιδιωτική ζωή.

Η DP είναι ένας σύγχρονος ορισμός της ιδιωτικότητας που χρησιμοποιείται κατά την ανάλυση μεγάλων συνόλων δεδομένων. Εγγυάται ότι οι αντίπαλοι δεν μπορούν να ανακαλύψουν ένα άτομο στο προστατευόμενο σύνολο δεδομένων συγκρίνοντας τα δεδομένα με άλλα σύνολα δεδομένων. Σε αυτό το πλαίσιο, αντίπαλος είναι ένα άτομο ή ένας οργανισμός που προσπαθεί να συγκεντρώσει πληροφορίες για συγκεκριμένα άτομα μέσα σε ένα σύνολο δεδομένων. Αυτό θα μπορούσε να κυμαίνεται από έναν γείτονα που θέλει να κουτσομπολέψει μέχρι μια ομάδα ακτιβιστών που στοχεύει τους υπαλλήλους μιας εταιρείας. Ανεξάρτητα από τον αντίπαλο, επιτυγχάνεται με τη χρήση μαθηματικών, τεχνητής νοημοσύνης, ευφυούς ελέγχου, μηχανικής μάθησης και άλλων τεχνολογιών. Σε σχέση με την ανάλυση δεδομένων μπορεί να οριστεί ανεπίσημα χρησιμοποιώντας μια προσέγγιση πριν και μετά. Δηλαδή, ο αναλυτής δεν πρέπει να γνωρίζει περισσότερα για οποιοδήποτε άτομο μετά την ανάλυση των δεδομένων. Περαιτέρω, κάθε αντίπαλος δεν θα πρέπει να έχει πολύ διαφορετική άποψη για οποιοδήποτε άτομο αφού έχει πρόσβαση σε μια βάση δεδομένων. Σε έναν πιο τεχνικό ορισμό, παρέχει ιδιωτικότητα μέσω διαδικασίας πιο συγκεκριμένα, η διαδικασία εισάγει τυχαιότητα σε ένα σύνολο δεδομένων. Η διαδικασία πρέπει να το επιτυγχάνει αυτό χωρίς να μεταβάλλει την τελική ανάλυση των δεδομένων. Οι παρούσες και μελλοντικές πηγές βοηθητικών πληροφοριών (όπως άλλα σύνολα δεδομένων) δεν πρέπει να θέτουν σε κίνδυνο την ιδιωτικότητα ενός ατόμου.

Η μεθοδολογία της DP λειτουργεί χρησιμοποιώντας διάφορους μαθηματικούς μηχανισμούς. Ο πρώτος είναι ο μηχανισμός Laplace. Ο μηχανισμός Laplace είναι ένας μαθηματικός τύπος που προσθέτει θόρυβο σε ένα σύνολο δεδομένων. Ο τύπος καθορίζει πόσο θόρυβο πρέπει να προστεθεί εξετάζοντας την ποσότητα των δεδομένων σε ένα σύνολο και προσθέτοντας αρκετό θόρυβο ώστε να διασφαλιστεί η DP. Ο μηχανισμός Laplace είναι ένας τρόπος γενικής χρήσης για την επίτευξη της και είναι ένας μηχανισμός προσθετικού θορύβου. Ο μηχανισμός Gauss είναι ένα άλλο παράδειγμα μηχανισμού προσθετικού θορύβου και προσθέτει θόρυβο με βάση την κατανομή πιθανοτήτων Gauss, λαμβάνοντας υπόψη τις παραμέτρους ευαισθησίας και ιδιωτικότητας. Ένας άλλος μηχανισμός που χρησιμοποιείται είναι ο εκθετικός μηχανισμός ο οποίος αντί να προσθέτει θόρυβο σε ένα σύνολο δεδομένων, ο εκθετικός μηχανισμός αντλεί μια έξοδο από ένα σύνολο δεδομένων χρησιμοποιώντας μια κατανομή πιθανοτήτων. Αυτό σημαίνει ότι με τη χρήση του εκθετικού μηχανισμού, τα δεδομένα αντλούνται τυχαία για την ακριβή απάντηση των ερωτημάτων. Σε πολλές περιπτώσεις, ορισμένα δεδομένα πρέπει να διατηρούνται αναλλοίωτα ή αμετάβλητα με το διαφορικό απόρρητο, ώστε η ανάλυση να παραμένει ακριβής. Ένα παράδειγμα ενός συνόλου δεδομένων που πρέπει να είναι διαφορικά ιδιωτικά είναι τα βιομετρικά δεδομένα από την παρακολούθηση των ματιών. Τα ακουστικά εικονικής

πραγματικότητας συλλέγουν αυτά τα βιομετρικά δεδομένα για να λειτουργήσουν και τα δεδομένα κίνησης πρέπει να παραμείνουν αμετάβλητα.

2.2 Ορισμοί και έννοιες.

Η μεθοδολογία DP είναι ένας αυστηρός μαθηματικός ορισμός της ιδιωτικότητας. Στην απλούστερη περίπτωση, θεωρήστε έναν αλγόριθμο που αναλύει ένα σύνολο δεδομένων και υπολογίζει στατιστικά στοιχεία σχετικά με αυτό (όπως η μέση τιμή, η διακύμανση, η διάμεσος, ο τρόπος λειτουργίας κ.λπ.). Ένας τέτοιος αλγόριθμος λέγεται ότι είναι διαφορικά ιδιωτικός εάν, εξετάζοντας την έξοδο, δεν μπορεί κανείς να πει εάν τα δεδομένα οποιουδήποτε ατόμου περιλαμβάνονταν στο αρχικό σύνολο δεδομένων ή όχι. Με άλλα λόγια, η εγγύηση ενός διαφορικά ιδιωτικού αλγορίθμου είναι ότι η συμπεριφορά του ελάχιστα αλλάζει όταν ένα άτομο εντάσσεται ή αποχωρεί από το σύνολο δεδομένων οτιδήποτε μπορεί να βγάλει ο αλγόριθμος σε μια βάση δεδομένων που περιέχει πληροφορίες κάποιου ατόμου είναι σχεδόν εξίσου πιθανό να έχει προέλθει από μια βάση δεδομένων χωρίς τις πληροφορίες αυτού του ατόμου. Πιο συγκεκριμένα, η εγγύηση αυτή ισχύει για οποιοδήποτε άτομο και οποιοδήποτε σύνολο δεδομένων. Επομένως, ανεξάρτητα από το πόσο εκκεντρικά είναι τα στοιχεία οποιουδήποτε μεμονωμένου ατόμου και ανεξάρτητα από τα στοιχεία οποιουδήποτε άλλου ατόμου στη βάση δεδομένων, η εγγύηση της DP εξακολουθεί να ισχύει. Αυτό παρέχει μια τυπική εγγύηση ότι οι πληροφορίες σε ατομικό επίπεδο για τους συμμετέχοντες στη βάση δεδομένων δεν διαρρέουν.

Πολλές ευρετικές τεχνικές χρησιμοποιούνται για τη διατήρηση της ιδιωτικότητας των ατόμων σε ερευνητικές βάσεις δεδομένων. Η ανωνυμοποίηση (η αφαίρεση «αναγνωρίσιμων» χαρακτηριστικών, όπως ονόματα, διευθύνσεις, διευθύνσεις IP κ.λπ.) είναι η πιο συχνά χρησιμοποιούμενη τεχνική. Ωστόσο, αυτές οι ευρετικές τεχνικές, που στερούνται οποιωνδήποτε τυπικών εγγυήσεων, μπορεί να αποτύχουν και έχει επανειλημμένα αποδειχθεί ότι αποτυγχάνουν. Σε ένα εντυπωσιακό παράδειγμα, η Latanya Sweeney έδειξε ότι το φύλο, η ημερομηνία γέννησης και ο ταχυδρομικός κώδικας αρκούν για τη μοναδική ταυτοποίηση της μεγάλης πλειοψηφίας των Αμερικανών. Συνδέοντας αυτά τα χαρακτηριστικά σε μια υποτιθέμενη ανώνυμη βάση δεδομένων υγειονομικής περίθαλψης με δημόσια αρχεία ψηφοφόρων, κατάφερε να ταυτοποιήσει τον ατομικό φάκελο υγείας του κυβερνήτη της Μασαχουσέτης. Αυτές οι «επιθέσεις σύνδεσης» αιτιολογούν την ανάγκη για έναν ισχυρό ορισμό της ιδιωτικής ζωής έναν ορισμό που να είναι απρόσβλητος σε επιθέσεις που χρησιμοποιούν βοηθητική γνώση, συμπεριλαμβανομένης της γνώσης που ο διαχειριστής των δεδομένων δεν μπορεί να προβλέψει τη διαθεσιμότητά της.

Μια άλλη γραμμή εργασίας έχει δείξει ότι η απάντηση σε πάρα πολλά αβλαβή (ακόμη και εντελώς τυχαία) ερωτήματα σχετικά με μια βάση δεδομένων παραβιάζει εγγενώς την ιδιωτικότητα των μεμονωμένων συνεισφερόντων της. Αυτές οι εργασίες αποκαλύπτουν ένα θεμελιώδες αντιστάθμισμα μεταξύ στατιστικής χρησιμότητας και ιδιωτικότητας. Προκειμένου να κατανοήσουμε αυτό το αντιστάθμισμα και να εντοπίσουμε κοινωνικά επιθυμητά αποτελέσματα, πρέπει να είμαστε σε θέση να ορίσουμε επίσημα την ιδιωτικότητα εξ αρχής. Ένα κρίσιμο χαρακτηριστικό της είναι ότι ορίζει την ιδιωτικότητα όχι ως μια δυαδική έννοια του «εκτέθηκαν ή όχι τα δεδομένα του ατόμου», αλλά μάλλον ως ζήτημα συσσωρευτικού κινδύνου. Δηλαδή, κάθε φορά που τα δεδομένα ενός ατόμου υποβάλλονται σε επεξεργασία ο κίνδυνος να εκτεθεί αυξάνεται. Για το σκοπό αυτό, ο ορισμός της είναι εξοπλισμένος με παραμέτρους («έψιλον και δέλτα») που ποσοτικοποιούν την «απώλεια ιδιωτικότητας» τον πρόσθετο κίνδυνο για ένα άτομο που προκύπτει από τη χρήση των δεδομένων του. Ανεξάρτητα από οποιαδήποτε βοηθητική γνώση που

χρησιμοποιείται σε μια επίθεση επαναπροσδιορισμού, ο κίνδυνος για την ιδιωτικότητα ενός ατόμου που προκαλείται από έναν αλγόριθμο DP θα περιορίζεται για πάντα από αυτή την απώλεια ιδιωτικότητας.
Μέσα από εκτεταμένη θεωρητική έρευνα, η μεθοδολογία DP υπόσχεται να επιτρέψει την κοινή χρήση ερευνητικών δεδομένων σε μια ευρεία ποικιλία ρυθμίσεων. Το απλούστερο και πιο καλά μελετημένο σενάριο είναι η απελευθέρωση στατιστικών ερωτημάτων: ένας ιδιοκτήτης δεδομένων μπορεί να καθορίσει ερωτήματα καταμέτρησης, όπως «πόσα άτομα στη βάση δεδομένων είναι άνδρες;» ή «πόσα άτομα στη βάση δεδομένων ζουν στη Μασαχουσέτη;» και να λάβει απαντήσεις διαταραγμένες από μια μικρή ποσότητα τυχαίου θορύβου. Οι διαφορικά ιδιωτικοί αλγόριθμοι είναι σε θέση να απαντήσουν προσεγγιστικά σε ένα μεγάλο αριθμό τέτοιων ερωτημάτων, έτσι ώστε ένας ερευνητής που βλέπει αυτές τις προσεγγιστικές απαντήσεις να μπορεί να βγάλει περίπου τα ίδια συμπεράσματα σαν να είχε ο ίδιος τα δεδομένα. Ωστόσο, η εμβέλεια της εκτείνεται πολύ πέρα από την απλή περίπτωση των στατιστικών ερωτημάτων. Για παράδειγμα, υπάρχουν διαφορικά ιδιωτικές εκδόσεις αλγορίθμων στη μηχανική μάθηση, τη θεωρία παιγνίων και το σχεδιασμό οικονομικών μηχανισμών, τη στατιστική εκτίμηση και τη ροή δεδομένων. Αξίζει να σημειωθεί ότι η μεθοδολογία λειτουργεί καλύτερα σε μεγαλύτερες βάσεις δεδομένων. Αυτό οφείλεται στο γεγονός ότι καθώς αυξάνεται ο αριθμός των ατόμων σε μια βάση δεδομένων, η επίδραση κάθε μεμονωμένου ατόμου σε μια δεδομένη συνολική στατιστική μειώνεται.

2.3 Ιστορική εξέλιξη και βασικά έγγραφα.

Οι απαρχές της DP συνδέονται στενά με τις προσπάθειες για την εξισορρόπηση της προστασίας της ιδιωτικότητας και της χρησιμότητας των στατιστικών δεδομένων, μέσα από τη χρήση τεχνικών Στατιστικής Διασφάλισης Απορρήτου (Statistical Disclosure Control – SDC). Η έννοια της προστασίας των δεδομένων σε στατιστικές βάσεις έχει τις ρίζες της στη δεκαετία του 1970 και αποτελεί το θεμέλιο πάνω στο οποίο χτίστηκαν οι μεταγενέστερες προσεγγίσεις, συμπεριλαμβανομένης της DP. Από τη δεκαετία του 1970, όταν οι στατιστικοί φορείς άρχισαν να δημοσιεύουν δεδομένα για δημόσια χρήση, αναδείχθηκε το πρόβλημα της αποκάλυψης προσωπικών πληροφοριών μέσω συνδυασμών δεδομένων. Για παράδειγμα, ακόμη και όταν δημοσιεύονταν μόνο συγκεντρωτικά στατιστικά στοιχεία, υπήρχε η δυνατότητα «επαναταυτοποίησης» ατόμων, ειδικά αν κάποιος είχε στη διάθεσή του εξωτερικές ή πρόσθετες πληροφορίες.
Το θεμελιώδες έργο του Tore Dalenius (1977) αποτέλεσε ένα από τα πρώτα ολοκληρωμένα πλαίσια για την αντιμετώπιση αυτών των προβλημάτων. Ο Dalenius όρισε ότι το ιδανικό σύστημα στατιστικής ανάλυσης θα πρέπει να εγγυάται πως οι πληροφορίες που μπορεί να εξαγάγει κάποιος από ένα σύνολο δημοσιευμένων δεδομένων δεν θα επιτρέψουν την ταυτοποίηση ή την αποκάλυψη προσωπικών δεδομένων. Αυτό το πρόβλημα ονομάστηκε «statistical disclosure control». Αντίστοιχα προβλήματα προέκυπταν και στη δεκαετία του 1970, γεγονός που τόνισε την ανάγκη για καλύτερες τεχνικές προστασίας των δεδομένων. Μια πρώιμη προσέγγιση που αναπτύχθηκε ήταν η κυψελοειδής συγκέντρωση δεδομένων ή αλλιώς «cell suppression», η οποία αφορά την απόκρυψη συγκεκριμένων κυψελών δεδομένων (π.χ., τιμών σε έναν πίνακα) που θεωρούνται ευαίσθητες ή μπορούν να οδηγήσουν στην αποκάλυψη προσωπικών πληροφοριών. Ωστόσο, αυτή η μέθοδος οδηγεί σε απώλεια σημαντικής πληροφορίας και σε αμφισβητή αποδοτικότητα ως προς την προστασία απορρήτου. Μια άλλη τεχνική ήταν η αντικατάσταση ή διαστρέβλωση δεδομένων, κατά την οποία τα ατομικά δεδομένα τροποποιούνται, έτσι ώστε να μην είναι άμεσα ταυτοποιήσιμα. Αυτό περιλαμβάνει την εισαγωγή «θορύβου» ή άλλες τροποποιήσεις που προστατεύουν την ταυτότητα του ατόμου, αλλά επιτρέπουν τη χρήση των δεδομένων

σε στατιστικές αναλύσεις. Οι στατιστικοί ανέπτυξαν μεθόδους που ελαχιστοποιούσαν την πιθανότητα αποκάλυψης μέσω τεχνικών όπως η ανάλυση σε ομάδες ή η μερική διαστρέβλωση των στοιχείων, ώστε να εξασφαλιστεί ότι δεν θα είναι δυνατόν να ανακτηθούν ατομικά δεδομένα με οποιοδήποτε λογικό συνδυασμό εξωτερικών πληροφοριών.

Η μεθοδολογία της SDC προσέφερε τις πρώτες μαθηματικές και θεωρητικές βάσεις για την κατανόηση του πώς μπορεί κανείς να δημοσιεύει ασφαλή δεδομένα, χωρίς να θέτει σε κίνδυνο την ιδιωτικότητα των ατόμων. Η δουλειά του Dalenius ήταν καθοριστική για τη θεμελίωση της έννοιας της ανωνυμίας, παρόλο που αργότερα φάνηκε ότι η πλήρης ανωνυμία σε στατιστικά δεδομένα ήταν δύσκολο να επιτευχθεί. Παρά τις σημαντικές προόδους, οι μέθοδοι SDC δεν ήταν επαρκείς για να προσφέρουν ισχυρές εγγυήσεις προστασίας. Ο κυριότερος περιορισμός ήταν ότι η ανωνυμοποίηση και οι μέθοδοι κατακερματισμού δεδομένων μπορούσαν να αποκαλύψουν ευαίσθητες πληροφορίες αν συνδυαστούν με εξωτερικά δεδομένα ή αν ένας επιτιθέμενος είχε πρόσβαση σε επιπλέον πληροφορίες (π.χ., από άλλες βάσεις δεδομένων). Το γεγονός αυτό ώθησε την ερευνητική κοινότητα να αναζητήσει πιο ισχυρές μαθηματικές προσεγγίσεις στην προστασία απορρήτου, όπως η DP, που βασίζεται στη διασφάλιση ότι ακόμα και αν κάποιος έχει πρόσβαση σε όλα τα δεδομένα, η πληροφορία που θα αποκτήσει δεν θα παραβιάζει την ιδιωτικότητα του ατόμου (Dwork et al., 2006).

Η DP έκανε την επίσημη εμφάνισή της το 2006 με τη δημοσίευση του πρωτοποριακού έργου της Cynthia Dwork και των συνεργατών της. Το έργο αυτό ήταν μια απάντηση στα προβλήματα και τις αδυναμίες των προηγούμενων μεθόδων προστασίας της ιδιωτικότητας, όπως αυτές που εφαρμόζονταν με τεχνικές SDC. Προσφέρει αυστηρές μαθηματικές εγγυήσεις για την ιδιωτικότητα, κάτι που δεν επιτυγχάνεται από τις πιο παραδοσιακές τεχνικές όπως η ανωνυμοποίηση και εισάγει έναν σαφή μαθηματικό ορισμό για την προστασία της ιδιωτικότητας που αφορά τις στατιστικές αναλύσεις σε βάσεις δεδομένων. Η κεντρική ιδέα είναι ότι οι απαντήσεις που προκύπτουν από τα ερωτήματα ή τις αναλύσεις σε μια βάση δεδομένων δεν θα πρέπει να αποκαλύπτουν ευαίσθητες πληροφορίες για κάποιο συγκεκριμένο άτομο. Αυτό επιτυγχάνεται με την προσθήκη ελεγχόμενου "θορύβου" στις απαντήσεις, εξασφαλίζοντας ότι η συμμετοχή ή μη ενός ατόμου στη βάση δεδομένων δεν μπορεί να ανιχνευθεί με ακρίβεια από κάποιον τρίτο.

Δίνει εγγυήσεις της μορφής ότι αν δύο βάσεις δεδομένων είναι όμοιες, διαφέροντας μόνο σε ένα στοιχείο (π.χ., έναν χρήστη), τότε οι πιθανότητες εξαγωγής οποιουδήποτε συμπεράσματος θα πρέπει να είναι σχεδόν οι ίδιες και για τις δύο βάσεις. Η μαθηματική διατύπωση της περιγράφεται με δύο παραμέτρους, το έψιλον και το δέλτα που προσδιορίζουν πόσο διαφορετικά μπορεί να είναι τα αποτελέσματα από τη χρήση των δύο βάσεων. Η μικρή τιμή αυτών των παραμέτρων εξασφαλίζει μεγαλύτερη προστασία της ιδιωτικότητας. Ακόμα και αν ένας εισβολέας γνωρίζει τα πάντα εκτός από τη συμμετοχή ενός ατόμου στη βάση δεδομένων, η συμμετοχή αυτού του ατόμου δεν μπορεί να ταυτοποιηθεί με αξιοπιστία. Παρέχει αυστηρές και μετρήσιμες εγγυήσεις απορρήτου, ανεξαρτήτως του πόσα εξωτερικά δεδομένα είναι διαθέσιμα. Ο «θόρυβος» που προστίθεται στα δεδομένα είναι τέτοιος ώστε να διατηρεί τη χρησιμότητα των δεδομένων σε στατιστικές αναλύσεις, ενώ παράλληλα προστατεύει την ιδιωτικότητα των ατόμων. Ο μηχανισμός που προτάθηκε ήταν ο Laplace Mechanism, όπου ο θόρυβος που προστίθεται είναι ανάλογος της "ευαισθησίας" των δεδομένων, δηλαδή του πόσο πολύ μπορεί να αλλάξει η έξοδος μιας ανάλυσης αν τροποποιηθεί ένα στοιχείο στη βάση δεδομένων (Dwork et al., 2006a). Αυτό το έργο έθεσε τα θεμέλια για όλες τις επόμενες εξελίξεις στη διαφύλαξη της ιδιωτικότητας και είχε τεράστια επίδραση στην έρευνα της κρυπτογραφίας και της επιστήμης των

δεδομένων. Μετά την εισαγωγή του όρου το 2006, η έννοια της DP άρχισε να αναπτύσσεται ραγδαία, με σημαντικές συνεισφορές από διάφορους ερευνητές. Ο Laplace Mechanism που εισήχθη το 2006, προσθέτει θόρυβο σύμφωνα με την κατανομή Laplace για ερωτήματα των οποίων η ευαισθησία είναι γνωστή.

Αργότερα, αναπτύχθηκε και ο Exponential Mechanism από τον McSherry και τον Talwar (2007), για καταστάσεις όπου η απάντηση δεν είναι αριθμητική, αλλά κατηγορική (McSherry & Talwar, 2007). Σημαντικό τμήμα της έρευνας εστιάστηκε στο πώς η προσθήκη θορύβου μπορεί να επηρεάσει την ακρίβεια των αποτελεσμάτων και πώς να επιτευχθεί η σωστή ισορροπία μεταξύ ακρίβειας και απορρήτου. Αναπτύχθηκαν διάφορα μοντέλα της DP, όπως η LDP, όπου τα ίδια τα δεδομένα διαστρεβλώνονται από τον χρήστη πριν αποσταλούν για ανάλυση. Το μοντέλο αυτό επιτρέπει τη διαφύλαξη της ιδιωτικότητας χωρίς να χρειάζεται να εμπιστευτεί κάποιος έναν κεντρικό διαχειριστή της βάσης δεδομένων. Η DP άνοιξε το δρόμο για την πρακτική εφαρμογή των μηχανισμών ιδιωτικότητας σε διάφορους τομείς.

Ένα από τα σημαντικότερα προβλήματα που αντιμετωπίστηκαν ήταν η ισορροπία ανάμεσα στην ιδιωτικότητα και τη χρησιμότητα των δεδομένων. Η προσθήκη θορύβου στις απαντήσεις των ερωτημάτων εγγυάται την ιδιωτικότητα, αλλά επηρεάζει και την ακρίβεια των αποτελεσμάτων. Στόχος των ερευνητών ήταν να βρουν τρόπους ώστε να ελαχιστοποιηθεί ο θόρυβος που προστίθεται, διατηρώντας ωστόσο υψηλό επίπεδο ιδιωτικότητας. Σε αυτό το πλαίσιο, αναπτύχθηκε η ιδέα του mechanism design (σχεδιασμός μηχανισμών), το οποίο περιλαμβάνει τον σχεδιασμό κατάλληλων αλγορίθμων και μηχανισμών που ικανοποιούν συγκεκριμένους στόχους, όπως η μεγιστοποίηση της ακρίβειας υπό συνθήκες διασφάλισης απορρήτου. Ο McSherry και ο Talwar (2007) εισήγαγαν τον Exponential Mechanism, μια βελτίωση του αρχικού μοντέλου για ερωτήματα που δεν είναι αριθμητικά (McSherry & Talwar, 2007).

Η ευαισθησία των ερωτημάτων (sensitivity) αποτελεί ένα κρίσιμο στοιχείο στη DP, καθώς μετρά πόσο μπορεί να αλλάξει η απάντηση ενός ερωτήματος αν προστεθεί ή αφαιρεθεί ένα στοιχείο από τη βάση δεδομένων. Η αρχική προσέγγιση της προσθήκης θορύβου με βάση την κατανομή Laplace ήταν σημαντική, αλλά οδήγησε σε παραπέρα έρευνα για το πώς ο θόρυβος θα μπορούσε να προσαρμοστεί ανάλογα με το είδος των δεδομένων και των ερωτημάτων. Ένα βασικό αποτέλεσμα ήταν η ανάπτυξη του smooth sensitivity, που εισήχθη από τους Nissim, Raskhodnikova, και Smith (2007). Η smooth sensitivity είναι μια πιο ευέλικτη προσέγγιση, καθώς αντί για την παραδοσιακή ευαισθησία, εξετάζει πώς η ευαισθησία μπορεί να μεταβάλλεται ομαλά ανάλογα με το είδος των δεδομένων, ώστε να προσαρμόζεται καλύτερα ο θόρυβος.

Στο αρχικό μοντέλο της DP, ο θόρυβος προστίθετο κυρίως σε αριθμητικές τιμές. Ωστόσο, για πολλά πρακτικά προβλήματα, τα δεδομένα δεν είναι αριθμητικά, αλλά κατηγοριοποιημένα (π.χ., επιλογή του καλύτερου υποψηφίου ή προϊόντος). Ο Exponential Mechanism που αναπτύχθηκε από τον McSherry και τον Talwar (2007) επεκτείνει την DP ώστε να μπορεί να εφαρμοστεί σε μη αριθμητικά ερωτήματα. Ο μηχανισμός αυτός επιτρέπει την επιλογή της καλύτερης κατηγορίας ή απόφασης με βάση μια συνάρτηση ωφέλειας, εξασφαλίζοντας ταυτόχρονα την ιδιωτικότητα. Κατά την περίοδο αυτή, οι ερευνητές ξεκίνησαν να εξετάζουν πώς θα μπορούσαν να συνδυαστούν οι αλγόριθμοι μηχανικής μάθησης (machine learning - ML) με την DP. Για παράδειγμα, προτάθηκαν αλγόριθμοι που την εφαρμόζουν για τη δημιουργία μοντέλων μηχανικής μάθησης, διατηρώντας τα δεδομένα προπόνησης ασφαλή. Αυτό έδωσε ώθηση για την ανάπτυξη νέων μεθόδων προστασίας απορρήτου κατά τη διαδικασία της ανάλυσης μεγάλων δεδομένων (big data).

Ένα άλλο σημαντικό βήμα προς την επέκταση της ήταν η ανάπτυξη της LDP, η οποία παρουσιάστηκε το 2008. Ενώ η παραδοσιακή DP εφαρμόζεται σε κεντρικά διαχειριζόμενες βάσεις δεδομένων, όπου ένας διαχειριστής προσθέτει θόρυβο για να προστατεύσει τα δεδομένα, η LDP δίνει στον ίδιο τον χρήστη τη δυνατότητα να προσθέσει θόρυβο στα δεδομένα του πριν τα αποστείλει στον διαχειριστή της βάσης. Αυτό το μοντέλο ιδιωτικότητας είναι ιδιαίτερα χρήσιμο όταν δεν υπάρχει εμπιστοσύνη στον κεντρικό διαχειριστή των δεδομένων, καθώς επιτρέπει τη διαφύλαξη της ιδιωτικότητας από την πηγή, χωρίς την ανάγκη για κεντρικό έλεγχο. Ένα παράδειγμα εφαρμογής LDP είναι η ανώνυμη συλλογή στατιστικών στοιχείων από μεγάλους οργανισμούς, για ανάλυση δεδομένων χρηστών χωρίς να εκτίθεται η ταυτότητά τους.

Κατά την περίοδο αυτή, έγιναν και οι πρώτες προσπάθειες για την ανάπτυξη πρακτικών εργαλείων και πλατφορμών που θα μπορούσαν να υποστηρίξουν την εφαρμογή της σε ρεαλιστικές συνθήκες. Ένα από τα πρώτα παραδείγματα ήταν το PINQ (Privacy Integrated Queries), ένα σύστημα που αναπτύχθηκε από τον McSherry (2009), το οποίο επιτρέπει στους ερευνητές να υποβάλλουν ερωτήματα σε μια βάση δεδομένων με εγγυήσεις της DP χωρίς να χρειάζεται να κατανοήσουν τις λεπτομέρειες των υποκείμενων αλγορίθμων. Το PINQ ήταν ένα πρώιμο εργαλείο που έθεσε τα θεμέλια για την ευρύτερη χρήση της σε εφαρμογές. Κατά τη διάρκεια της περιόδου αυτής, προτάθηκαν και νέες χαλαρώσεις της DP, προκειμένου να διευκολυνθεί η εφαρμογή της σε διαφορετικές περιπτώσεις. Μία τέτοια προσέγγιση ήταν η διατύπωση του (ε, δ), που επιτρέπει μια μικρή πιθανότητα «διαρροής» πληροφοριών (ελεγχόμενη από την παράμετρο δέλτα, ενώ διατηρεί τις γενικές εγγυήσεις απορρήτου που παρέχει η DP. Αυτή η προσέγγιση επέτρεψε την εφαρμογή της σε πιο απαιτητικά και περίπλοκα περιβάλλοντα. Μετά την εδραίωση της στο θεωρητικό επίπεδο μεταξύ 2006 και 2010, οι εφαρμογές της στον πραγματικό κόσμο γνώρισαν μεγάλη άνοδο από το 2010 και μετά. Αυτή η περίοδος χαρακτηρίστηκε από την υιοθέτηση της σε πρακτικά περιβάλλοντα και τη χρήση της από μεγάλους τεχνολογικούς οργανισμούς και κυβερνητικούς φορείς. Ο σκοπός ήταν να προστατευτούν τα δεδομένα των χρηστών, ενώ παράλληλα να επιτραπεί η χρήσιμη ανάλυση αυτών των δεδομένων.

Η Google ήταν μία από τις πρώτες εταιρείες που την υιοθέτησαν σε ευρεία κλίμακα. Το 2014, παρουσίασε το RAPPOR (Randomized Aggregatable PrivacyPreserving Ordinal Response), ένα σύστημα που επιτρέπει την ανώνυμη συλλογή δεδομένων χρήσης από browsers και άλλες υπηρεσίες χωρίς να αποκαλύπτεται η ταυτότητα των χρηστών. Το RAPPOR βασίζεται στην LDP και επιτρέπει την ανώνυμη συλλογή δεδομένων με εγγυήσεις απορρήτου. Στην περίπτωση του RAPPOR, κάθε χρήστης διαστρεβλώνει τα δεδομένα του με τέτοιο τρόπο ώστε να διασφαλίζεται η ιδιωτικότητα, αλλά ταυτόχρονα τα δεδομένα μπορούν να συγκεντρωθούν και να αναλυθούν σε επίπεδο πληθυσμού, παρέχοντας χρήσιμες πληροφορίες στη Google για τη βελτίωση των υπηρεσιών της.

Η Apple την εισήγαγε το 2016 για τη συλλογή δεδομένων από τους χρήστες της, όπως οι προτιμήσεις χρήσης του πληκτρολογίου, emojis και άλλα λειτουργικά χαρακτηριστικά. Η εταιρεία την χρησιμοποιεί για να βελτιώσει τις λειτουργίες και τα προϊόντα της χωρίς να παραβιάζεται η ιδιωτικότητα των χρηστών. Το σύστημα της Apple βασίζεται στην LDP, όπου κάθε συσκευή προσθέτει τυχαίο θόρυβο στις πληροφορίες των χρηστών πριν αυτές αποσταλούν στους διακομιστές της Apple. Η ανάλυση των δεδομένων σε επίπεδο πληθυσμού επιτρέπει στην εταιρεία να εντοπίζει μοτίβα χωρίς να αποκαλύπτει την ταυτότητα του ατόμου. Η Apple περιέγραψε την προσέγγισή της στην DP σε σειρά άρθρων στο Apple Machine Learning Journal, που περιγράφουν πώς την εφαρμόζουν για να επιτύχουν ισχυρές εγγυήσεις απορρήτου σε μεγάλο εύρος υπηρεσιών.

Το Γραφείο Απογραφής των ΗΠΑ (U.S. Census Bureau) έκανε το μεγαλύτερο ίσως βήμα για την υιοθέτηση της σε κυβερνητικό επίπεδο. Για την απογραφή του 2020, η U.S. Census Bureau υιοθέτησε ένα DP βασιζόμενο σύστημα για να προστατεύσει τα δεδομένα των πολιτών που συλλέχθηκαν κατά τη διάρκεια της απογραφής. Η U.S. Census Bureau χρησιμοποίησε ένα σύστημα (ε, δ), που επιτρέπει την εισαγωγή θορύβου στα στατιστικά δεδομένα που δημοσιεύονται, έτσι ώστε να μειώνεται ο κίνδυνος ταυτοποίησης ατόμων από τα δημοσιευμένα αποτελέσματα. Αυτό το σύστημα επιτρέπει τη δημοσίευση συγκεντρωτικών στατιστικών χωρίς να παραβιάζεται η ιδιωτικότητα των ατόμων που συμμετείχαν στην απογραφή. Η απογραφή του 2020 ήταν η πρώτη φορά που χρησιμοποιήθηκε σε τόσο μεγάλη κλίμακα σε μια κυβερνητική διαδικασία συλλογής δεδομένων, δημιουργώντας ένα σημαντικό προηγούμενο για τη χρήση της σε παγκόσμιο επίπεδο.

Η Microsoft έχει επίσης επενδύσει στην DP και την έχει ενσωματώσει στα εργαλεία της για την ανάλυση δεδομένων. Το 2019, παρουσίασε τη SmartNoise, μια βιβλιοθήκη ανοιχτού κώδικα για την ανάλυση δεδομένων με εγγυήσεις απορρήτου. Το SmartNoise επιτρέπει στους ερευνητές και τους προγραμματιστές να την εφαρμόζουν σε δεδομένα μεγάλου όγκου, όπως αυτά που συλλέγονται από διάφορες επιχειρηματικές εφαρμογές και υπηρεσίες. Η Microsoft την χρησιμοποίησε και σε άλλες εφαρμογές, όπως η προστασία των δεδομένων χρήστη σε υπηρεσίες όπως το Office 365 και το Azure. Εκτός από τις μεγάλες εταιρείες τεχνολογίας, η DP χρησιμοποιείται όλο και περισσότερο και στην ακαδημαϊκή έρευνα. Οι ερευνητές την χρησιμοποιούν για τη διαφύλαξη της ιδιωτικότητας των δεδομένων σε μελέτες που περιλαμβάνουν ευαίσθητα δεδομένα, όπως ιατρικά ή κοινωνιολογικά δεδομένα. Ειδικά σε πανεπιστήμια και ερευνητικά κέντρα χρησιμοποιείται για να επιτρέψει την ανάλυση μεγάλων βάσεων δεδομένων χωρίς να παραβιάζεται η ιδιωτικότητα των συμμετεχόντων.

Άρχισε να εφαρμόζεται ευρέως και στους αλγορίθμους ML. Για παράδειγμα, οι αλγόριθμοι ML μπορούν να εκπαιδεύονται σε ευαίσθητα δεδομένα (όπως ιατρικά αρχεία) με εγγυήσεις, έτσι ώστε οι εκπαιδευμένοι αλγόριθμοι να μην αποκαλύπτουν λεπτομέρειες για τα δεδομένα εκπαίδευσης. Αυτή η εξέλιξη επέτρεψε τη χρήση της σε τομείς όπως η υγεία, η ανάλυση κοινωνικών δεδομένων και η τραπεζική. Στην ιδιωτική μηχανική μάθηση (private machine learning), χρησιμοποιούνται τεχνικές όπως η DPSGD , που προσθέτουν θόρυβο στις ενημερώσεις παραμέτρων των μοντέλων μηχανικής μάθησης, παρέχοντας έτσι εγγυήσεις ιδιωτικότητας. Παράλληλα με την εξάπλωση των κρυπτονομισμάτων και της τεχνολογίας blockchain, υπήρξε ενδιαφέρον για την ενσωμάτωση της σε πλατφόρμες που βασίζονται σε αυτές τις τεχνολογίες. Τα δεδομένα που αποθηκεύονται σε αποκεντρωμένες βάσεις δεδομένων, όπως τα blockchain, μπορούν να διασφαλιστούν για να προστατευτούν οι συναλλαγές και οι χρήστες.

2.4 Προκλήσεις και Περιορισμοί στην Εφαρμογή της DP

Ένας από τους κύριους περιορισμούς της είναι ότι η προσθήκη θορύβου μπορεί να μειώσει τη χρησιμότητα των δεδομένων, ειδικά σε περιπτώσεις που απαιτούν υψηλή ακρίβεια, όπως στην εκπαίδευση μοντέλων μηχανικής μάθησης. Ο σχεδιασμός των μηχανισμών θορύβου πρέπει να γίνει προσεκτικά για να επιτευχθεί η σωστή ισορροπία μεταξύ ιδιωτικότητας και χρησιμότητας. Παρόλο που έχουν αναπτυχθεί πολλά εργαλεία και πλατφόρμες για την εφαρμογή της, η χρήση τους απαιτεί σημαντική γνώση και κατανόηση των παραμέτρων της. Η σωστή διαμόρφωση αυτών των παραμέτρων όπως το ε είναι κρίσιμη για τη διασφάλιση της ιδιωτικότητας και τη διατήρηση της χρησιμότητας των δεδομένων.

Η DP έχει μετατραπεί από ένα θεωρητικό πλαίσιο σε μια κρίσιμη τεχνολογία για την προστασία της ιδιωτικότητας στον πραγματικό κόσμο. Οι μεγάλες εταιρείες τεχνολογίας καθώς και κυβερνητικοί φορείς όπως η U.S. Census Bureau, την έχουν ενσωματώσει στις πρακτικές τους για να προστατεύσουν τα δεδομένα των χρηστών. Παρά τις προκλήσεις που υπάρχουν στην εφαρμογή της θεωρείται πλέον η πιο ισχυρή και μετρήσιμη μορφή διασφάλισης απορρήτου, με ευρεία αποδοχή και αυξανόμενη χρήση σε παγκόσμιο επίπεδο.

2.5 Στοχαστική κάθοδος κλίσης (SGD) και οι περιορισμοί της στην ιδιωτικότητα.

Η στοχαστική κάθοδος κλίσης (SGD) είναι ένας δημοφιλής αλγόριθμος βελτιστοποίησης που χρησιμοποιείται στη μηχανική μάθηση και τη βαθιά μάθηση για την ελαχιστοποίηση της συνάρτησης απώλειας και την ενημέρωση των παραμέτρων ενός μοντέλου. Σε αντίθεση με την τυπική κάθοδο κλίσης, η οποία υπολογίζει την κλίση χρησιμοποιώντας ολόκληρο το σύνολο των δεδομένων, η SGD ενημερώνει τις παραμέτρους του μοντέλου χρησιμοποιώντας ένα τυχαία επιλεγμένο υποσύνολο δεδομένων (ή ακόμη και ένα μόνο δείγμα) σε κάθε επανάληψη. Αυτό καθιστά την SGD ταχύτερη και καταλληλότερη για σενάρια μάθησης μεγάλης κλίμακας και σε απευθείας σύνδεση. Η SGD διαδραματίζει κρίσιμο ρόλο στην εκπαίδευση σύνθετων μοντέλων όπως τα βαθιά νευρωνικά δίκτυα και αποτελεί ακρογωνιαίο λίθο των σύγχρονων πρακτικών τεχνητής νοημοσύνης (AI) και μηχανικής μάθησης (ML).

Με την πιο συχνή ενημέρωση των παραμέτρων του μοντέλου, η SGD βοηθά στην επίτευξη ταχύτερης σύγκλισης, ιδιαίτερα χρήσιμη για σύνολα δεδομένων υψηλής διάστασης και μεγάλα σύνολα δεδομένων. Παρά την ταχύτερη αρχική της πρόοδο, η SGD μπορεί να εμφανίσει θορυβώδη ή ασταθή σύγκλιση, γεγονός που συχνά καθιστά αναγκαία τη χρήση τεχνικών όπως τα χρονοδιαγράμματα ρυθμού μάθησης, η ορμή και η αποκοπή κλίσης. Η κάθοδος κλίσης είναι μια επαναληπτική διαδικασία βελτιστοποίησης που αναζητά τη βέλτιστη τιμή μιας αντικειμενικής συνάρτησης (ελάχιστο/μέγιστο). Είναι μια από τις πιο συχνά χρησιμοποιούμενες μεθόδους για την αλλαγή των παραμέτρων ενός μοντέλου προκειμένου να μειωθεί μια συνάρτηση κόστους σε έργα μηχανικής μάθησης. Ο πρωταρχικός στόχος της κατάβασης κλίσης είναι ο εντοπισμός των παραμέτρων του μοντέλου που παρέχουν τη μέγιστη ακρίβεια τόσο στα σύνολα δεδομένων εκπαίδευσης όσο και στα σύνολα δεδομένων δοκιμής. Στην κάθοδο κλίσης, η κλίση είναι ένα διάνυσμα που δείχνει προς τη γενική κατεύθυνση της πιο απότομης ανόδου της συνάρτησης σε ένα συγκεκριμένο σημείο. Ο αλγόριθμος μπορεί να πέσει σταδιακά προς χαμηλότερες τιμές της συνάρτησης, κινούμενος προς την αντίθετη κατεύθυνση της κλίσης, μέχρι να φτάσει στο ελάχιστο της συνάρτησης.

2.6 Εισαγωγή στο DPSGD

Καθώς τα μοντέλα μηχανικής μάθησης ενσωματώνονται όλο και περισσότερο στη ζωή μας, η ανάγκη προστασίας της ιδιωτικής ζωής των χρηστών γίνεται όλο και πιο κρίσιμη. Αυτά τα μοντέλα απαιτούν συχνά τεράστιες ποσότητες δεδομένων για να κάνουν ακριβείς προβλέψεις, οι οποίες μπορεί να περιλαμβάνουν ευαίσθητες πληροφορίες χρηστών. Σε αυτή την Απάντηση, θα συζητήσουμε γιατί η ιδιωτικότητα είναι σημαντική στην εκπαίδευση μοντέλων μηχανικής μάθησης, πώς να μετρήσουμε την ιδιωτικότητα και θα παρουσιάσουμε τη διαφοροποιημένη ιδιωτική στοχαστική κάθοδο κλίσης (DPSGD), η οποία είναι ένας αλγόριθμος βελτιστοποίησης που διατηρεί την ιδιωτικότητα.Τα μοντέλα μηχανικής μάθησης, ιδίως τα μοντέλα βαθιάς μάθησης, έχει αποδειχθεί ότι αποδίδουν εξαιρετικά καλά σε διάφορους τομείς, όπως η αναγνώριση εικόνων, η επεξεργασία φυσικής γλώσσας και τα

συστήματα συστάσεων. Ωστόσο, αυτά τα μοντέλα μπορούν επίσης να μάθουν κατά λάθος ευαίσθητες πληροφορίες από τα δεδομένα εκπαίδευσης, εκθέτοντας τα προσωπικά στοιχεία των χρηστών. Η έκθεση αυτή εγείρει ηθικές ανησυχίες και νομικές επιπτώσεις, καθώς μπορεί να οδηγήσει σε ακούσιες διακρίσεις, κλοπή ταυτότητας και άλλες παραβιάσεις της ιδιωτικής ζωής.

Για να αντιμετωπιστούν αυτές οι ανησυχίες, εμφανίστηκε ο τομέας της μηχανικής μάθησης με διατήρηση της ιδιωτικότητας, με στόχο την ανάπτυξη τεχνικών που επιτρέπουν στα μοντέλα να μαθαίνουν από τα δεδομένα χωρίς να αποκαλύπτουν ευαίσθητες πληροφορίες για μεμονωμένους χρήστες. Το απόρρητο είναι μια κρίσιμη πτυχή της εκπαίδευσης μοντέλων μηχανικής μάθησης, καθώς βοηθά στην προστασία ευαίσθητων πληροφοριών χρηστών από την ακούσια διαρροή τους μέσω των μοντέλων. Παρέχει ένα αυστηρό πλαίσιο για τη μέτρηση της ιδιωτικότητας στη μηχανική μάθηση και έχει υιοθετηθεί ευρέως στον τομέα της μηχανικής μάθησης με διατήρηση της ιδιωτικότητας. Η διαφορικά ιδιωτική SGD είναι μια πολλά υποσχόμενη προσέγγιση που επιτρέπει την εκπαίδευση μοντέλων μηχανικής μάθησης με ισχυρή εγγυημένη ιδιωτικότητα, εισάγοντας θόρυβο στις κλίσεις κατά τη διάρκεια της βελτιστοποίησης. Ωστόσο, είναι σημαντικό να εξισορροπηθεί προσεκτικά ο συμβιβασμός μεταξύ της ιδιωτικότητας και της ακρίβειας του μοντέλου για την επίτευξη τόσο ιδιωτικών όσο και χρήσιμων μοντέλων.

2.7 Πώς ενσωματώνεται το DP στον αλγόριθμο SGD.

Η έννοια της DP ενσωματώνεται στον αλγόριθμο SGD μέσω μιας παραλλαγής που ονομάζεται DPSGD. Αυτή η παραλλαγή σχεδιάστηκε για να προστατεύει την ιδιωτικότητα των δεδομένων κατά την εκπαίδευση μοντέλων μηχανικής μάθησης, εξασφαλίζοντας ότι οι ευαίσθητες πληροφορίες δεν θα είναι ανακτήσιμες από το τελικό μοντέλο. Ο DPSGD επιτυγχάνει τη DP μέσα από τον περιορισμό του μεγέθους των διαβαθμίσεων. Για να περιοριστεί η συμβολή κάθε δείγματος δεδομένων στο εκπαιδευτικό αποτέλεσμα, ο DPSGD εφαρμόζει περιορισμό στις διαβαθμίσεις, δηλαδή περιορίζει το μέγεθός τους σε μία προκαθορισμένη τιμή. Αυτό σημαίνει ότι οι διαβαθμίσεις με μεγαλύτερη συμβολή περικόπτονται, έτσι ώστε να μειωθεί η πιθανότητα διαρροής ευαίσθητων πληροφοριών. Ο περιορισμός του μεγέθους των διαβαθμίσεων είναι μια κρίσιμη τεχνική στον αλγόριθμο DPSGD που διασφαλίζει τη DP κατά την εκπαίδευση των μοντέλων μηχανικής μάθησης. Κατά την εκπαίδευση ενός μοντέλου μέσω SGD, υπολογίζονται οι διαβαθμίσεις, δηλαδή οι παραγώγοι, για κάθε δείγμα δεδομένων ή παρτίδα δεδομένων, ώστε να κατευθύνουν την ενημέρωση των παραμέτρων του μοντέλου. Σε αυτή τη διαδικασία, τα δεδομένα κάθε δείγματος συμβάλλουν στην τιμή των διαβαθμίσεων, κάτι που μπορεί να οδηγήσει σε διαρροή ευαίσθητων πληροφοριών αν δεν ληφθούν μέτρα προστασίας.

Για να μειωθεί ο κίνδυνος αυτός, εφαρμόζεται μια τεχνική γνωστή ως περιορισμός (clipping) των διαβαθμίσεων που σημαίνει ότι το μέγεθος των διαβαθμίσεων περιορίζεται σε μία προκαθορισμένη τιμή. Κατά τη διαδικασία αυτή, οποιοδήποτε κλίση ξεπερνά το καθορισμένο όριο περικόπτεται, ώστε να μη συνεισφέρει περισσότερο από το επιτρεπόμενο όριο στην εκπαίδευση του μοντέλου. Αυτός ο περιορισμός επιτυγχάνει δύο βασικούς στόχους: πρώτον, μειώνει την ευαισθησία του μοντέλου στις ιδιαιτερότητες κάθε δείγματος δεδομένων, και δεύτερον, διασφαλίζει ότι καμία μεμονωμένη εγγραφή δεδομένων δεν θα επηρεάσει υπερβολικά την πορεία εκπαίδευσης. Ο περιορισμός των διαβαθμίσεων εξισορροπεί τη συνεισφορά όλων των δεδομένων στο μοντέλο, γεγονός που αποτελεί σημαντικό παράγοντα για την προστασία της ιδιωτικότητας. Με αυτόν τον τρόπο, ο περιορισμός του μεγέθους των διαβαθμίσεων δρα ως ένα πρώτο στρώμα προστασίας πριν

την προσθήκη θορύβου, ενισχύοντας την ασφάλεια και προστατεύοντας τη διακριτική πληροφορία που μπορεί να περιέχεται στα δεδομένα.
Ο περιορισμός των διαβαθμίσεων (gradient clipping) εφαρμόζεται με μαθηματικά ορισμένο τρόπο: αφού υπολογιστεί το διάνυσμα των διαβαθμίσεων g για ένα δεδομένο δείγμα, υπολογίζεται το μέτρο (νόρμα) του, συνήθως με τη μορφή της Ευκλείδειας νόρμας (L2 νόρμα). Αν η νόρμα αυτή υπερβαίνει ένα προκαθορισμένο όριο, η κλίση τροποποιείται αναλογικά ώστε η τελική της νόρμα να μην ξεπερνά το συγκεκριμένο κατώφλι. Αν αυτή η νόρμα ξεπερνά ένα προκαθορισμένο όριο, που ονομάζεται πρότυπο περιορισμού και έχει τιμή C, τότε η κλίση g περιορίζεται έτσι ώστε η νόρμα του να είναι ίση με το C. Δηλαδή, αν ||g|| > C, τότε η νέα κλίση γίνεται:

$$g' = g \cdot \frac{C}{||g||}$$

Αυτό σημαίνει ότι το μέγεθος της κλίσης κανονικοποιείται ώστε να μην υπερβαίνει το όριο C, περιορίζοντας τη συμβολή κάθε δείγματος στη συνολική κατεύθυνση που ακολουθεί το μοντέλο κατά την εκπαίδευση.
Ο περιορισμός των διαβαθμίσεων εξυπηρετεί συγκεκριμένες ανάγκες στην εκπαίδευση των μοντέλων με DP. Τα εκτός των ορίων διαβαθμίσεις, που μπορεί να προκύψουν από σπάνια ή ακραία δείγματα δεδομένων, περιορίζονται. Χωρίς τον περιορισμό οι διαβαθμίσεις θα μπορούσαν να οδηγήσουν το μοντέλο σε μεγάλη απόκλιση, αυξάνοντας τον κίνδυνο διαρροής πληροφοριών. Καθώς κάθε δείγμα περιορίζεται στο ίδιο πρότυπο περιορισμού η συμβολή όλων των δεδομένων εξισώνεται. Αυτό δημιουργεί μια ενισχυμένη μορφή ιδιωτικότητας, καθώς κανένα δείγμα δεν μπορεί να επηρεάσει το μοντέλο σε μεγάλο βαθμό, ακόμα και αν τα δεδομένα του είναι πολύ διαφορετικά. Με το περιορισμό της κλίσης, γνωρίζουμε ήδη ότι όλες οι διαβαθμίσεις έχουν μέγεθος μικρότερο ή ίσο με το C. Αυτό διευκολύνει την προσθήκη θορύβου, καθώς μπορούμε να προσαρμόσουμε την ποσότητα του θορύβου που απαιτείται ώστε να επιτύχουμε έναν επιθυμητό βαθμό DP. Ο περιορισμός των διαβαθμίσεων είναι ουσιαστικός για τον καθορισμό του privacy budget (επίπεδο ιδιωτικότητας).
Το privacy budget, δηλαδή οι παράμετροι ε και δ, καθορίζουν πόσο θόρυβο πρέπει να προστεθεί για να εξασφαλιστεί η επιθυμητή ιδιωτικότητα. Με τον περιορισμό των διαβαθμίδεων, ελέγχεται η συμβολή των δεδομένων στο τελικό αποτέλεσμα, που σημαίνει ότι χρειαζόμαστε λιγότερο θόρυβο για να καλύψουμε τις διαφοροποιήσεις των δεδομένων. Ο καθορισμός του βέλτιστου προτύπου περιορισμού C είναι κρίσιμος για τη βελτιστοποίηση της ιδιωτικότητας και της απόδοσης του μοντέλου. Αν το C είναι πολύ μικρό, ο περιορισμός είναι υπερβολικός, κάτι που μπορεί να μειώσει την ικανότητα μάθησης του μοντέλου και να οδηγήσει σε ανεπαρκή απόδοση. Αν είναι πολύ μεγάλο, τότε η προστασία της ιδιωτικότητας αποδυναμώνεται, καθώς οι διαβαθμίσεις δεν περιορίζονται επαρκώς και το μοντέλο γίνεται πιο ευαίσθητο σε συγκεκριμένες ιδιότητες των δεδομένων. Συνολικά, ο περιορισμός του μεγέθους των διαβαθμίσεων αποτελεί το πρώτο βήμα για την επίτευξη της και συνεργάζεται στενά με την προσθήκη θορύβου, ώστε να επιτύχει μια ισχυρή προστασία των δεδομένων στην εκπαίδευση μηχανικής μάθησης.
Αφού περικοπούν οι διαβαθμίσεις, προστίθεται θόρυβος από μια κατανομή (συνήθως Gaussian ή Laplace) στις διαβαθμίσεις πριν από την ενημέρωση των παραμέτρων του μοντέλου. Ο θόρυβος διασφαλίζει ότι οι μικρές διαφοροποιήσεις στα δεδομένα δεν επηρεάζουν ουσιαστικά το αποτέλεσμα, προστατεύοντας έτσι τα δεδομένα των μεμονωμένων δειγμάτων. Η προσθήκη θορύβου στις διαβαθμίσεις είναι μια θεμελιώδης τεχνική που επιτρέπει την εφαρμογή της κατά την εκπαίδευση μοντέλων μηχανικής

μάθησης μέσω του DP-SGD. Μετά τον περιορισμό, περιορισμό των διαβαθμίσεων, που εξασφαλίζει ότι η συνεισφορά κάθε δείγματος δεδομένων είναι ομοιόμορφη, η προσθήκη θορύβου στις διαβαθμίσεις συμβάλλει στην περαιτέρω προστασία των δεδομένων των χρηστών. Ο στόχος είναι να εξασφαλιστεί ότι οι μικρές αλλαγές στα δεδομένα ενός μεμονωμένου δείγματος δεν θα επηρεάσουν σημαντικά την τελική ενημέρωση των παραμέτρων του μοντέλου, ώστε να μην μπορούν να ανακτηθούν ιδιωτικές πληροφορίες από τα αποτελέσματα της εκπαίδευσης (Abadi et al., 2016). Η τεχνική αυτή βασίζεται στη χρήση θορύβου που προέρχεται από συγκεκριμένες στατιστικές κατανομές, όπως την Gaussian (κανονική) ή τη Laplace κατανομή. Η επιλογή της κατανομής εξαρτάται από τις επιθυμητές ιδιότητες ιδιωτικότητας και τη συμπεριφορά του θορύβου σε σχέση με τις παραμέτρους του μοντέλου. Η Gaussian κατανομή, η οποία συνήθως χρησιμοποιείται στην DPSGD, προσφέρει υψηλή ιδιωτικότητα, αφού επιτρέπει την προσθήκη τυχαίων τιμών με συμμετρική διακύμανση, διασφαλίζοντας έτσι ότι τα αποτελέσματα της εκπαίδευσης δεν περιλαμβάνουν ενδείξεις που θα μπορούσαν να αποκαλύψουν πληροφορίες για τα αρχικά δεδομένα (Dwork et al., 2014).

Ο θόρυβος που προστίθεται στις διαβαθμίσεις ελέγχεται μέσω των παραμέτρων ιδιωτικότητας, δηλαδή του ε και του δ, που προσδιορίζουν το λεγόμενο privacy budget. Το ε αντιπροσωπεύει την ισχύ της και επηρεάζει άμεσα την ποσότητα του θορύβου: όσο μικρότερη είναι η τιμή του ε, τόσο μεγαλύτερη η προστασία, αλλά και τόσο πιο θορυβώδη είναι τα δεδομένα εκπαίδευσης, με πιθανή επίπτωση στην ακρίβεια του μοντέλου (Goodfellow et al., 2016). Η παράμετρος δ χρησιμοποιείται για να αντισταθμίσει την ισχύ του ε και καθορίζει τον βαθμό στον οποίο η DP μπορεί να μην ισχύει απόλυτα. Αυτή η παράμετρος χρησιμοποιείται σε εφαρμογές όπου είναι αποδεκτό ένα μικρό ποσοστό αποτυχίας στην προστασία της ιδιωτικότητας (McSherry, 2009). Η προσθήκη θορύβου στις διαβαθμίσεις ενσωματώνει τυχαίες αποκλίσεις στο αποτέλεσμα της εκπαίδευσης, καθιστώντας δυσκολότερη την εξαγωγή πληροφοριών για μεμονωμένα δεδομένα χρηστών. Αυτή η διαδικασία είναι ιδιαίτερα χρήσιμη στην ανάλυση και αξιολόγηση μεγάλων δεδομένων, καθώς μειώνει την πιθανότητα ανάκτησης ατομικών πληροφοριών, ειδικά όταν το μοντέλο αναπτύσσεται σε περιβάλλοντα όπου ισχύουν αυστηροί κανονισμοί προστασίας προσωπικών δεδομένων, όπως το GDPR και το HIPAA (Shokri & Shmatikov, 2015).

Σε ότι αφορά την ρύθμιση εγγυήσεων ιδιωτικότητας (Privacy Budget) ο θόρυβος που προστίθεται εξαρτάται από το privacy budget (επίπεδο ιδιωτικότητας), δηλαδή από τις παραμέτρους ε και δ της DP, που καθορίζουν το βαθμό προστασίας έναντι της ακρίβειας του μοντέλου. Όσο μικρότερο είναι το ε, τόσο μεγαλύτερη η ιδιωτικότητα, αλλά και η απώλεια στην απόδοση του μοντέλου. Η προσθήκη θορύβου στις διαβαθμίσεις είναι μια θεμελιώδης τεχνική που επιτρέπει την εφαρμογή της κατά την εκπαίδευση μοντέλων μηχανικής μάθησης μέσω του DPSGD. Μετά τον περιορισμό των διαβαθμίσεων, που εξασφαλίζει ότι η συνεισφορά κάθε δείγματος δεδομένων είναι ομοιόμορφη, η προσθήκη θορύβου στις διαβαθμίσεις συμβάλλει στην περαιτέρω προστασία των δεδομένων των χρηστών. Ο στόχος είναι να εξασφαλιστεί ότι οι μικρές αλλαγές στα δεδομένα ενός μεμονωμένου δείγματος δεν θα επηρεάσουν σημαντικά την τελική ενημέρωση των παραμέτρων του μοντέλου, ώστε να μην μπορούν να ανακτηθούν ιδιωτικές πληροφορίες από τα αποτελέσματα της εκπαίδευσης (Abadi et al., 2016).

Η τεχνική αυτή βασίζεται στη χρήση θορύβου που προέρχεται από συγκεκριμένες στατιστικές κατανομές, όπως την Gaussian (κανονική) ή τη Laplace κατανομή. Η επιλογή της κατανομής εξαρτάται από τις επιθυμητές ιδιότητες ιδιωτικότητας και τη συμπεριφορά του θορύβου σε σχέση με τις παραμέτρους του μοντέλου. Η Gaussian κατανομή, η οποία συνήθως χρησιμοποιείται στην DPSGD, προσφέρει υψηλή ιδιωτικότητα, αφού επιτρέπει

την προσθήκη τυχαίων τιμών με συμμετρική διακύμανση, διασφαλίζοντας έτσι ότι τα αποτελέσματα της εκπαίδευσης δεν περιλαμβάνουν ενδείξεις που θα μπορούσαν να αποκαλύψουν πληροφορίες για τα αρχικά δεδομένα (Dwork et al., 2014). Ο θόρυβος που προστίθεται στις διαβαθμίσεις ελέγχεται μέσω των παραμέτρων ιδιωτικότητας, δηλαδή του ε και του δ, που προσδιορίζουν το λεγόμενο privacy budget. Το ε αντιπροσωπεύει την ισχύ της και επηρεάζει άμεσα την ποσότητα του θορύβου: όσο μικρότερη είναι η τιμή του ε, τόσο μεγαλύτερη η προστασία, αλλά και τόσο πιο θορυβώδη είναι τα δεδομένα εκπαίδευσης, με πιθανή επίπτωση στην ακρίβεια του μοντέλου (Goodfellow et al., 2016). Η παράμετρος δ χρησιμοποιείται για να αντισταθμίσει την ισχύ του ε και καθορίζει τον βαθμό στον οποίο η DP μπορεί να μην ισχύει απόλυτα. Αυτή η παράμετρος χρησιμοποιείται σε εφαρμογές όπου είναι αποδεκτό ένα μικρό ποσοστό αποτυχίας στην προστασία της ιδιωτικότητας (McSherry, 2009). Η προσθήκη θορύβου στις διαβαθμίσεις ενσωματώνει τυχαίες αποκλίσεις στο αποτέλεσμα της εκπαίδευσης, καθιστώντας δυσκολότερη την εξαγωγή πληροφοριών για μεμονωμένα δεδομένα χρηστών. Αυτή η διαδικασία είναι ιδιαίτερα χρήσιμη στην ανάλυση και αξιολόγηση μεγάλων δεδομένων, καθώς μειώνει την πιθανότητα ανάκτησης ατομικών πληροφοριών, ειδικά όταν το μοντέλο αναπτύσσεται σε περιβάλλοντα όπου ισχύουν αυστηροί κανονισμοί προστασίας προσωπικών δεδομένων, όπως το GDPR και το HIPAA (Shokri & Shmatikov, 2015).

3 Ρυθμιστικά πλαίσια και απαιτήσεις απορρήτου

3.1Λεπτομερής συζήτηση του ΓΚΠΔ.

Ο ΓΚΠΔ είναι ένας νόμος για την προστασία της ιδιωτικής ζωής που απαιτεί από τους οργανισμούς που προσφέρουν αγαθά και υπηρεσίες σε άτομα που βρίσκονται στην ΕΕ/ΕΟΧ ή παρακολουθούν τη συμπεριφορά τους να προασπίζουν τα δικαιώματα προστασίας της ιδιωτικής τους ζωής και να διασφαλίζουν τα προσωπικά δεδομένα που έχουν συλλεχθεί ή υποστεί επεξεργασία. Αντικατέστησε την οδηγία για την προστασία των δεδομένων του 1995, η οποία δημιούργησε νόμους για την προστασία των δεδομένων ανά χώρα, με αποτέλεσμα ένα λιγότερο συνεκτικό συνονθύλευμα κανονισμών στην Ευρώπη. Ο κανονισμός απαιτεί την εφαρμογή επτά αρχών προστασίας δεδομένων και διευκολύνει οκτώ δικαιώματα προστασίας της ιδιωτικής ζωής για τους καταναλωτές. Τα κράτη μέλη έχουν τις δικές τους αρχές προστασίας δεδομένων για να χειρίζονται την επιβολή του νόμου- δεν το χειρίζεται μια κεντρική αρχή.

Όπως σημειώνεται στο άρθρο 1 του κανονισμού (ΕΕ) αριθ. 3, ο ΓΚΠΔ εφαρμόζεται σε οργανισμούς που επεξεργάζονται τα δεδομένα προσωπικού χαρακτήρα «οποιουδήποτε βρίσκεται στην επικράτεια της ΕΕ» στο πλαίσιο της προσφοράς αγαθών ή υπηρεσιών ή της παρακολούθησης της συμπεριφοράς, ανεξάρτητα από το αν υπάρχει ή όχι πληρωμή. Δεν έχει σημασία αν η εταιρεία έχει την έδρα της στην ΕΕ ή αν έχει ακόμη και φυσική παρουσία εκεί. Επιπλέον, η αιτιολογική σκέψη 25 περιγράφει την εφαρμογή του ΓΚΠΔ ως συνέπεια της εφαρμογής του διεθνούς δικαίου: *«Όταν το δίκαιο κράτους μέλους εφαρμόζεται δυνάμει του δημόσιου διεθνούς δικαίου, ο παρών κανονισμός θα πρέπει επίσης να εφαρμόζεται σε υπεύθυνο επεξεργασίας που δεν είναι εγκατεστημένος στην Ένωση, όπως σε διπλωματική ή προξενική αποστολή κράτους μέλους».*

3.1.1 Ορισμός των προσωπικών δεδομένων στον ΓΚΠΔ

Κάθε πληροφορία που αφορά «ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο» το οποίο μπορεί να ταυτοποιηθεί άμεσα ή έμμεσα με τη χρήση της είναι δεδομένα προσωπικού χαρακτήρα. Αυτά μπορεί να περιλαμβάνουν προφανείς πληροφορίες όπως ονόματα,

αριθμούς ταυτότητας, αριθμούς τηλεφώνου ή διευθύνσεις ηλεκτρονικού ταχυδρομείου, αλλά και διευθύνσεις IP, πληροφορίες που συλλέγονται μέσω cookies του προγράμματος περιήγησης ή ευαίσθητα προσωπικά στοιχεία όπως το φύλο, οι θρησκευτικές πεποιθήσεις ή η πολιτική τοποθέτηση. Προσωπικά αναγνωρίσιμες πληροφορίες, είναι ένας όρος που χρησιμοποιείται συνήθως στις Ηνωμένες Πολιτείες για να αναφερθεί σε πληροφορίες που μπορούν να χρησιμοποιηθούν από μόνες τους ή με άλλες πληροφορίες για την ταυτοποίηση, την επικοινωνία ή τον εντοπισμό ενός μεμονωμένου ατόμου ή για την ταυτοποίηση ενός ατόμου στο πλαίσιο. Κάθε ενέργεια που εκτελείται σε δεδομένα προσωπικού χαρακτήρα ή σύνολα δεδομένων προσωπικού χαρακτήρα, είτε αυτοματοποιημένη είτε χειροκίνητη, αποτελεί επεξεργασία δεδομένων. Αυτό μπορεί να περιλαμβάνει, μεταξύ άλλων ενεργειών, τη «συλλογή, καταγραφή, οργάνωση, διάρθρωση, αποθήκευση, προσαρμογή ή μεταβολή, ανάκτηση, διαβούλευση, χρήση, γνωστοποίηση με διαβίβαση, διάδοση ή με άλλο τρόπο διάθεση, ευθυγράμμιση ή συνδυασμό, περιορισμό, διαγραφή ή καταστροφή» των δεδομένων προσωπικού χαρακτήρα.

Ο ΓΚΠΔ ορίζει ως «υποκείμενο των δεδομένων» το φυσικό πρόσωπο του οποίου τα δεδομένα προσωπικού χαρακτήρα υποβάλλονται σε επεξεργασία από υπεύθυνο επεξεργασίας ή εκτελούντα την επεξεργασία. Για τις διαδικτυακές εταιρείες ή τις επιχειρήσεις με φυσική τοποθεσία που έχουν διαδικτυακή παρουσία, συνηθέστερα αυτό θα περιλάμβανε τους επισκέπτες ενός ιστότοπου, τους πελάτες ή τους χρήστες εφαρμογών.

Υπεύθυνος επεξεργασίας δεδομένων είναι το «φυσικό ή νομικό πρόσωπο, η δημόσια αρχή, η υπηρεσία ή άλλος φορέας που, μόνος ή από κοινού με άλλους», αποφασίζει γιατί και πώς θα υποβληθούν σε επεξεργασία τα δεδομένα προσωπικού χαρακτήρα. Συνήθως πρόκειται για μια εταιρεία ή έναν διεθνή οργανισμό. Ο υπεύθυνος επεξεργασίας επικοινωνεί επίσης με τον εκτελούντα την επεξεργασία των δεδομένων και κατευθύνει αυτόν, εάν η εν λόγω οντότητα είναι τρίτο μέρος.

Όταν δύο ή περισσότεροι υπεύθυνοι επεξεργασίας δεδομένων αποφασίζουν τους σκοπούς και τα μέσα της επεξεργασίας δεδομένων μεμονωμένα ή από κοινού, είναι από κοινού υπεύθυνοι επεξεργασίας. Το άρθρο 1 του κανονισμού (ΕΚ) αριθ. 26 ΓΚΠΔ προβλέπει λεπτομερείς διατάξεις για τον κοινό υπεύθυνο επεξεργασίας και απαιτεί από τους κοινούς υπευθύνους επεξεργασίας να έχουν καταγεγραμμένη (συμβατική) συμφωνία μεταξύ τους. Η συμφωνία αυτή περιγράφει τους αντίστοιχους ρόλους και τις αρμοδιότητες, ιδίως όσον αφορά την άσκηση των δικαιωμάτων των υποκειμένων των δεδομένων και τις υποχρεώσεις των κοινών υπευθύνων επεξεργασίας για την παροχή πληροφοριών βάσει του ΓΚΠΔ.Τα υποκείμενα των δεδομένων μπορούν να ασκήσουν τα δικαιώματά τους έναντι οποιουδήποτε ή όλων των υπευθύνων επεξεργασίας σε μια συμφωνία κοινού ελέγχου. Ένα τρίτο μέρος που επεξεργάζεται δεδομένα προσωπικού χαρακτήρα για λογαριασμό ενός υπευθύνου επεξεργασίας είναι εκτελών την επεξεργασία δεδομένων. Αυτό θα μπορούσε να περιλαμβάνει ένα ευρύ φάσμα οντοτήτων, συμπεριλαμβανομένου ενός φυσικού ή νομικού προσώπου, δημόσιας αρχής, οργανισμού ή άλλου φορέα.

Οι υπάλληλοι ενός υπεύθυνου επεξεργασίας δεδομένων που ενεργούν στο πλαίσιο των εργασιακών τους καθηκόντων θεωρούνται συνήθως πράκτορες του υπεύθυνου επεξεργασίας δεδομένων και όχι εκτελούντες την επεξεργασία δεδομένων. Οι εκτελούντες την επεξεργασία δεδομένων μπορεί να κυμαίνονται από παρόχους διακομιστών που βασίζονται στο υπολογιστικό νέφος, έως επεξεργαστές πληρωμών, εταιρείες adtech ή martech και πολλά άλλα. Οι οργανισμοί πρέπει να έχουν νόμιμη ή νομική βάση για την επεξεργασία δεδομένων προσωπικού χαρακτήρα, π.χ. με τη συγκατάθεση των χρηστών ή την εκτέλεση σύμβασης. Πρέπει να διαχειρίζονται τα δεδομένα με τρόπο που δεν είναι αδικαιολόγητα επιζήμιος, απροσδόκητος ή παραπλανητικός και πρέπει να παρέχουν σαφείς

και προσβάσιμες πληροφορίες σχετικά με τις δραστηριότητες επεξεργασίας δεδομένων τους.
Τα δεδομένα προσωπικού χαρακτήρα μπορούν να συλλέγονται μόνο για συγκεκριμένο, ρητό και νόμιμο σκοπό και οι οργανισμοί δεν μπορούν να τα επεξεργάζονται περαιτέρω κατά τρόπο ασύμβατο με τους σκοπούς αυτούς. Εάν αλλάξει ο σκοπός (οι σκοποί) για τον οποίο (τους οποίους) μια εταιρεία έχει συλλέξει και επεξεργαστεί δεδομένα προσωπικού χαρακτήρα, πρέπει να λάβει νέα συγκατάθεση του χρήστη για τον (τους) νέο(-ους) σκοπό(-ους) επεξεργασίας. Οι οργανισμοί θα πρέπει να επεξεργάζονται μόνο τον ελάχιστο αριθμό δεδομένων προσωπικού χαρακτήρα που είναι απαραίτητος για την επίτευξη των σκοπών επεξεργασίας τους και θα πρέπει να μοιράζονται τα δεδομένα μόνο με τις λιγότερες οντότητες που είναι απαραίτητες για την ολοκλήρωση της επεξεργασίας.
Τα δεδομένα προσωπικού χαρακτήρα πρέπει να διατηρούνται μόνο για όσο χρονικό διάστημα τα χρειάζονται οι οργανισμοί για τους σκοπούς της επεξεργασίας. Μετά την εκπλήρωση αυτών των σκοπών, οι οργανισμοί αναμένεται να επιστρέφουν, να διαγράφουν ή να ανωνυμοποιούν τα δεδομένα, ώστε να αποφεύγεται η περιττή αποθήκευση προσωπικών πληροφοριών. Ο περιορισμός της αποθήκευσης ισχύει και για τους τρίτους εκτελούντες την επεξεργασία. Τα δεδομένα προσωπικού χαρακτήρα πρέπει να είναι ακριβή και επικαιροποιημένα. Τα ανακριβή δεδομένα πρέπει να διορθώνονται ή να διαγράφονται χωρίς καθυστέρηση. Το δικαίωμα διόρθωσης περιλαμβάνεται στα δικαιώματα των υποκειμένων των δεδομένων. Οι οργανισμοί πρέπει να επεξεργάζονται τα δεδομένα προσωπικού χαρακτήρα κατά τρόπο που να διασφαλίζει την κατάλληλη ασφάλεια, συμπεριλαμβανομένης της προστασίας από μη εξουσιοδοτημένη ή παράνομη πρόσβαση ή επεξεργασία και από τυχαία απώλεια, καταστροφή ή ζημία.
Οι οργανισμοί είναι υπεύθυνοι για τη συμμόρφωση με τον ΓΚΠΔ και πρέπει να είναι σε θέση να αποδεικνύουν τη συμμόρφωση με όλες αυτές τις αρχές. Οι τρίτοι εκτελούντες την επεξεργασία έχουν επίσης ευθύνες για τη συμμόρφωση με την ασφάλεια και την προστασία της ιδιωτικής ζωής, αλλά η τελική ευθύνη ανήκει στον υπεύθυνο επεξεργασίας, επομένως είναι σημαντικές οι ισχυρές συμβάσεις και η εποπτεία. Οι νομικές βάσεις και έννομο συμφέρον στον Γενικό Κανονισμό Προστασίας Δεδομένων καλύπτουν τη «νομιμότητα της επεξεργασίας», ή τις νομικές βάσεις, όπως συνήθως αναφέρονται. Πρόκειται για τις συνθήκες υπό τις οποίες η επεξεργασία δεδομένων από έναν υπεύθυνο επεξεργασίας είναι νόμιμη. Ενώ η συγκατάθεση του χρήστη είναι πιθανώς αυτή που έρχεται πιο εύκολα στο μυαλό, υπάρχουν συνολικά έξι:

- το υποκείμενο των δεδομένων έχει δώσει τη συγκατάθεσή του
- για την εκτέλεση σύμβασης με το υποκείμενο των δεδομένων
- συμμόρφωση με νομική υποχρέωση στην οποία υπόκειται ο υπεύθυνος επεξεργασίας δεδομένων
- για την προστασία ζωτικών συμφερόντων του υποκειμένου των δεδομένων ή άλλου φυσικού προσώπου
- για λόγους δημοσίου συμφέροντος ή εάν ο υπεύθυνος επεξεργασίας ασκεί δημόσια εξουσία
- έννομα συμφέροντα που επιδιώκει ο υπεύθυνος επεξεργασίας ή τρίτος

Οι εταιρείες πρέπει να είναι προσεκτικές όταν πρόκειται για έννομο συμφέρον. Μπορεί να είναι βολικό για έναν υπεύθυνο επεξεργασίας δεδομένων να το ισχυριστεί, καθώς αποφεύγεται η λήψη και η αποθήκευση της συγκατάθεσης του χρήστη. Ωστόσο, πρέπει επίσης να μπορεί να αποδειχθεί στις αρχές. Σύμφωνα με τον ΓΚΠΔ, το έννομο συμφέρον δεν ισχύει «όταν τα συμφέροντα αυτά υπερισχύουν των συμφερόντων ή των θεμελιωδών

δικαιωμάτων και ελευθεριών του υποκειμένου των δεδομένων που απαιτούν προστασία των δεδομένων προσωπικού χαρακτήρα, ιδίως όταν το υποκείμενο των δεδομένων είναι παιδί». Το έννομο συμφέρον δεν μπορεί πλέον να επιλεγεί ως νομική βάση για την εξατομίκευση της διαφήμισης και του περιεχομένου, οπότε πλέον η συγκατάθεση είναι η μόνη επιλογή που μπορεί να επιλεγεί. Ορισμένες καταστάσεις απαιτούν από τον υπεύθυνο επεξεργασίας να μοιραστεί τα δεδομένα ενός ατόμου με τρίτο μέρος για την εκπλήρωση των υποχρεώσεων μιας σύμβασης. Ένα παράδειγμα είναι οι εταιρείες ηλεκτρονικού εμπορίου, οι οποίες συχνά συνεργάζονται με τρίτους, όπως οι επεξεργαστές πληρωμών και οι εταιρείες εφοδιαστικής και εκπλήρωσης για την ολοκλήρωση των παραγγελιών και την παράδοση των αγορών στον πελάτη.
Σύμφωνα με τον ΓΚΠΔ, οι υπεύθυνοι επεξεργασίας μπορούν να μοιράζονται δεδομένα προσωπικού χαρακτήρα με αυτά τα τρίτα μέρη. Αυτό είναι γνωστό ως επεξεργασία «απαραίτητη για την εκτέλεση σύμβασης» σύμφωνα με το άρθρο 6 του ΓΚΠΔ. Ο υπεύθυνος επεξεργασίας στην περίπτωση αυτή οφείλει να διασφαλίσει, μέσω συμφωνίας προστασίας δεδομένων ή κατάλληλων συμβατικών ρητρών, ότι τα εν λόγω τρίτα μέρη συμμορφώνονται επίσης με τις απαιτήσεις προστασίας δεδομένων του ΓΚΠΔ.

3.1.2 Συναίνεση όπως ορίζεται από τον Γενικό Κανονισμό για την Προστασία Δεδομένων

Από τους καταναλωτές στο διαδίκτυο ζητείται συχνά η συγκατάθεσή τους για τη συλλογή και επεξεργασία των προσωπικών τους δεδομένων πολλές φορές την ημέρα. Οι δικτυακοί τόποι εμφανίζουν τακτικά τοίχους με cookies που ζητούν τη συγκατάθεση. Πολλά από αυτά παρέχουν ποικίλα επίπεδα διαφάνειας όσον αφορά την κοινοποίηση των δικαιωμάτων και των επιλογών, την αναλυτικότητα στην προσαρμογή των επιλογών συγκατάθεσης ή την απόρριψη της συγκατάθεσης συνολικά, αν και πολλά banners cookies εξακολουθούν να μην συμμορφώνονται με τον ΓΚΠΔ. Στην αιτιολογική σκέψη 32 απαριθμούνται οι προϋποθέσεις του ΓΚΠΔ για την έγκυρη συγκατάθεση: *«Η συγκατάθεση θα πρέπει να δίδεται με σαφή θετική πράξη που καθιερώνει την ελεύθερη, συγκεκριμένη, εν επιγνώσει και σαφή ένδειξη της συμφωνίας του υποκειμένου των δεδομένων για την επεξεργασία δεδομένων προσωπικού χαρακτήρα που το αφορούν, όπως με γραπτή δήλωση, συμπεριλαμβανομένων των ηλεκτρονικών μέσων, ή με προφορική δήλωση.»*
Η αιτιολογική σκέψη περιγράφει επίσης τους όρους που δεν ισχύουν για τη συγκατάθεση και τον τρόπο ακριβούς απεικόνισης του πεδίου εφαρμογής του αιτήματος συγκατάθεσης.

- Η σιωπή, τα προεπιλεγμένα κουτάκια ή η αδράνεια δεν συνιστούν έγκυρη συγκατάθεση, καθώς δεν αποτελούν σαφή ένδειξη της ρητής συγκατάθεσης του υποκειμένου των δεδομένων.
- Η συγκατάθεση θα πρέπει να καλύπτει όλες τις δραστηριότητες επεξεργασίας που πραγματοποιούνται για τον ίδιο σκοπό ή σκοπούς.
- Όταν τα δεδομένα θα υποβληθούν σε επεξεργασία για πολλαπλούς σκοπούς, το υποκείμενο των δεδομένων πρέπει να δώσει ρητή συγκατάθεση για όλους αυτούς.
- Εάν η συγκατάθεση του υποκειμένου των δεδομένων πρέπει να δοθεί κατόπιν ηλεκτρονικού αιτήματος, το αίτημα πρέπει να είναι σαφές και συνοπτικό και δεν πρέπει να διαταράσσει άσκοπα τη χρήση της υπηρεσίας για την οποία παρέχεται η συγκατάθεση.

Οι επικοινωνίες ή τα χαρακτηριστικά διεπαφής χρήστη που χειραγωγούν ή εξαπατούν τους χρήστες ώστε να παράσχουν τη συγκατάθεσή τους ή να ολοκληρώσουν με άλλο τρόπο ενέργειες που διαφορετικά δεν θα είχαν επιλέξει είναι γνωστά ως «σκοτεινά μοτίβα». Οι νομοθέτες και οι αρμόδιοι φορείς αντιμετωπίζουν όλο και πιο αρνητικά τέτοιες

δραστηριότητες και τους οργανισμούς που τις χρησιμοποιούν, και ορισμένοι κανονισμοί τις έχουν απαγορεύσει ρητά.

Το άρθρο 2 του κανονισμού (ΕΚ) αριθ. 7 ΓΚΠΔ περιγράφει τις προϋποθέσεις για τη συγκατάθεση με τις ευθύνες του υπεύθυνου επεξεργασίας δεδομένων.

- Ο υπεύθυνος επεξεργασίας πρέπει να είναι σε θέση να αποδείξει ότι το υποκείμενο των δεδομένων συναίνεσε στην επεξεργασία των δεδομένων του, π.χ. στις αρχές προστασίας δεδομένων σε περίπτωση ελέγχου ή αίτησης πρόσβασης του υποκειμένου των δεδομένων.
- Εάν η συγκατάθεση δίδεται με γραπτή δήλωση που καλύπτει άλλα θέματα, το αίτημα για συγκατάθεση πρέπει να παρουσιάζεται με τρόπο ευδιάκριτο και κατανοητό, εύκολα προσβάσιμο, με σαφή και απλή γλώσσα.
- Το υποκείμενο των δεδομένων πρέπει να είναι σε θέση να ανακαλέσει τη συγκατάθεση ανά πάσα στιγμή, και πρέπει να είναι το ίδιο εύκολο να το πράξει όσο και να δώσει τη συγκατάθεση. Αυτό μπορεί να περιλαμβάνει την αλλαγή των προτιμήσεων για την παροχή μερικής ή διαφορετικής συγκατάθεσης σε λεπτομερές επίπεδο.
- Η εκτέλεση μιας σύμβασης ή η παροχή υπηρεσιών δεν μπορεί να είναι προσωρινή μετά τη λήψη της συγκατάθεσης του υποκειμένου των δεδομένων, εάν η συγκατάθεση δεν είναι απαραίτητη για την εκτέλεση της σύμβασης ή την παροχή υπηρεσιών.

Ο ΓΚΠΔ της ΕΕ χρησιμοποιεί ένα μοντέλο συγκατάθεσης του χρήστη «opt in», το οποίο σημαίνει ότι οι οργανισμοί δεν μπορούν να συλλέγουν ή να επεξεργάζονται δεδομένα έως ότου ο χρήστης - ένας ηλεκτρονικός αγοραστής, επισκέπτης ιστότοπου, χρήστης εφαρμογής κ.λπ. Η απαίτηση αυτή περιλαμβάνει τόσο προσωπικά δεδομένα όπως ονόματα και διευθύνσεις ηλεκτρονικού ταχυδρομείου, όσο και αρκετά λεπτομερή και «παρασκηνιακά» δεδομένα. Για παράδειγμα, σύμφωνα με τον ΓΚΠΔ, οι χρήστες πρέπει να συναινούν στην επεξεργασία προσωπικών δεδομένων, τα οποία συχνά λαμβάνονται μέσω της χρήσης cookies και άλλων τεχνολογιών παρακολούθησης σε ιστότοπους, προτού επιτραπεί στις εν λόγω υπηρεσίες να είναι ενεργές για τις διαδικτυακές δραστηριότητες του εν λόγω χρήστη. Οι νόμοι περί απορρήτου δεδομένων σε επίπεδο πολιτείας στις Ηνωμένες Πολιτείες, ωστόσο, έχουν μέχρι σήμερα εφαρμόσει ένα μοντέλο συναίνεσης του χρήστη «opt out». Οι οργανισμοί που υπόκεινται σε αυτούς τους κανονισμούς δεν χρειάζεται να λαμβάνουν τη συγκατάθεση του χρήστη πριν από τη συλλογή δεδομένων στις περισσότερες περιπτώσεις (με την τυπική εξαίρεση των δεδομένων παιδιών ή των δεδομένων που χαρακτηρίζονται ως ευαίσθητα), αλλά πρέπει να λαμβάνουν τη συγκατάθεση πριν από την πώληση των δεδομένων ή τη χρήση τους για σκοπούς κατάρτισης προφίλ ή στοχευμένης διαφήμισης.

Ο ΓΚΠΔ παρέχει στα υποκείμενα των δεδομένων οκτώ ρητά δικαιώματα βάσει του κεφαλαίου 3, άρθρα 15 έως 22. Αυτά αποτέλεσαν επίσης τη ραχοκοκαλιά των δικαιωμάτων των καταναλωτών στο πλαίσιο των νόμων για την προστασία της ιδιωτικής ζωής των δεδομένων που ψηφίστηκαν σε άλλες χώρες, αν και το «δικαίωμα στη λήθη» έχει υιοθετηθεί λιγότερο ευρέως εκτός ΕΕ.

Τα υποκείμενα των δεδομένων έχουν το δικαίωμα να ενημερώνονται για τη συλλογή των προσωπικών τους δεδομένων, μεταξύ άλλων:

- ταυτότητα του υπεύθυνου επεξεργασίας δεδομένων
- τους σκοπούς της επεξεργασίας
- αποδέκτες ή κατηγορίες αποδεκτών των δεδομένων προσωπικού χαρακτήρα
- την προβλεπόμενη περίοδο για την οποία θα αποθηκευτούν τα δεδομένα προσωπικού χαρακτήρα

Εάν έχει οριστεί υπεύθυνος προστασίας δεδομένων (ΥΠΔ), τα υποκείμενα των δεδομένων έχουν επίσης το δικαίωμα πρόσβασης στα στοιχεία επικοινωνίας του ΥΠΔ. Οι πληροφορίες αυτές παρέχονται συνήθως μέσω μιας ειδοποίησης ή μιας πολιτικής απορρήτου.
Τα υποκείμενα των δεδομένων έχουν το δικαίωμα να γνωρίζουν εάν ένας υπεύθυνος επεξεργασίας έχει επεξεργαστεί τα προσωπικά τους δεδομένα και, στην περίπτωση αυτή, να έχουν πρόσβαση στα δεδομένα που έχουν συλλεχθεί. Έχουν επίσης το δικαίωμα να γνωρίζουν τους σκοπούς της επεξεργασίας, τους τύπους των δεδομένων προσωπικού χαρακτήρα, το χρονικό διάστημα αποθήκευσης των δεδομένων και το ποιος έχει πρόσβαση σε αυτά. Μπορούν να υποβάλουν αίτημα στον υπεύθυνο επεξεργασίας με τη χρήση αίτησης πρόσβασης υποκειμένου δεδομένων (DSAR) .
Εάν ο υπεύθυνος επεξεργασίας έχει ανακριβή ή ελλιπή δεδομένα, το υποκείμενο των δεδομένων έχει το δικαίωμα να ζητήσει διόρθωση ή συμπλήρωση των δεδομένων αυτών.
Σε ορισμένες περιπτώσεις, το υποκείμενο των δεδομένων έχει το δικαίωμα να ζητήσει από τον υπεύθυνο επεξεργασίας να διαγράψει τα προσωπικά του δεδομένα. Οι καταστάσεις αυτές περιλαμβάνουν, μεταξύ άλλων, όταν τα δεδομένα δεν χρειάζονται πλέον, όταν ο χρήστης ανακαλεί τη συγκατάθεσή του και όταν τα δεδομένα έχουν υποστεί παράνομη επεξεργασία. Τα υποκείμενα των δεδομένων έχουν το δικαίωμα να ζητήσουν να μην υποβληθούν σε επεξεργασία τα προσωπικά τους δεδομένα σε ορισμένες περιπτώσεις, όπως όταν τα δεδομένα είναι ανακριβή (μέχρι ο υπεύθυνος επεξεργασίας να μπορεί να επαληθεύσει την ακρίβειά τους), η επεξεργασία είναι παράνομη και ο υπεύθυνος επεξεργασίας δεν χρειάζεται πλέον τα δεδομένα, μεταξύ άλλων.
Τα υποκείμενα των δεδομένων έχουν το δικαίωμα να λαμβάνουν αντίγραφο των δεδομένων προσωπικού χαρακτήρα που έχουν παράσχει σε έναν υπεύθυνο επεξεργασίας. Ο υπεύθυνος επεξεργασίας πρέπει να παρέχει τα δεδομένα αυτά σε «δομημένο, κοινώς χρησιμοποιούμενο και αναγνώσιμο από μηχανήματα μορφότυπο». Το υποκείμενο των δεδομένων έχει το δικαίωμα να μεταφέρει τα δεδομένα αυτά σε άλλον υπεύθυνο επεξεργασίας χωρίς αντιρρήσεις ή εμπόδια από τον αρχικό υπεύθυνο επεξεργασίας, υπό την προϋπόθεση ότι η επεξεργασία βασίζεται σε συγκατάθεση ή σε σύμβαση και πραγματοποιείται με αυτοματοποιημένα μέσα. Το δικαίωμα αυτό έχει επεκταθεί στην ΕΕ βάσει κανονισμών όπως ο νόμος για τις ψηφιακές αγορές.
Τα υποκείμενα των δεδομένων έχουν το δικαίωμα να αντιταχθούν στην επεξεργασία των προσωπικών τους δεδομένων για ορισμένους λόγους, όπως όταν τα δεδομένα υποβάλλονται σε επεξεργασία για λόγους έννομων συμφερόντων ή χρησιμοποιούνται για σκοπούς άμεσης εμπορικής προώθησης. Εάν τα δεδομένα υποβάλλονται σε επεξεργασία για σκοπούς άμεσης εμπορικής προώθησης, τα άτομα μπορούν να αντιταχθούν ανά πάσα στιγμή και τα δεδομένα τους δεν μπορούν πλέον να υποβάλλονται σε επεξεργασία για τους σκοπούς αυτούς.

3.1.3 Δικαίωμα σχετικά με την αυτοματοποιημένη ατομική λήψη αποφάσεων, συμπεριλαμβανομένης της κατάρτισης προφίλ

Τα υποκείμενα των δεδομένων έχουν το δικαίωμα να μην υπόκεινται σε σημαντικές αποφάσεις που λαμβάνονται αποκλειστικά με αυτοματοποιημένες διαδικασίες ή κατάρτιση προφίλ, όπως αυτές που λαμβάνονται από υπολογιστές χωρίς ανθρώπινη συμμετοχή (π.χ. εργαλεία τεχνητής νοημοσύνης), εάν οι αποφάσεις αυτές έχουν σημαντικές επιπτώσεις σε αυτά νομικά ή με άλλους σημαντικούς τρόπους. Κάθε νομική οντότητα -είτε φυσικό είτε νομικό πρόσωπο- που επεξεργάζεται τα δεδομένα προσωπικού χαρακτήρα φυσικών προσώπων που βρίσκονται εντός της ΕΕ στο πλαίσιο της προσφοράς αγαθών ή υπηρεσιών ή της παρακολούθησης της συμπεριφοράς πρέπει να συμμορφώνεται με τις διατάξεις του

ΓΚΠΔ. Αυτό περιλαμβάνει τόσο τους υπεύθυνους επεξεργασίας δεδομένων, οι οποίοι καθορίζουν τον σκοπό και τα μέσα επεξεργασίας των δεδομένων προσωπικού χαρακτήρα, όσο και τους εκτελούντες την επεξεργασία δεδομένων, οι οποίοι επεξεργάζονται δεδομένα για λογαριασμό του υπεύθυνου επεξεργασίας. Καθένας από αυτούς έχει συγκεκριμένες ευθύνες βάσει του ΓΚΠΔ για τη διασφάλιση της συμμόρφωσης. Αυτό που είναι σημαντικό είναι ότι το υποκείμενο των δεδομένων πρέπει να βρίσκεται στην ΕΕ η νομική οντότητα που είναι υπεύθυνη για τη συμμόρφωση μπορεί να βρίσκεται οπουδήποτε στον κόσμο.
Οι εταιρείες που επιθυμούν πρόσβαση στις ευρωπαϊκές αγορές πρέπει να συμμορφώνονται με τον ΓΚΠΔ. Μεταξύ άλλων ευθυνών κατά την επιδίωξη συμμόρφωσης με τον ΓΚΠΔ, οι εταιρείες πρέπει να επικοινωνούν με σαφήνεια:

- ποιες κατηγορίες δεδομένων συλλέγουν
- για ποιους σκοπούς συλλέγονται
- πώς συλλέγονται
- ποιος θα έχει πρόσβαση σε αυτά

Εάν οποιαδήποτε από αυτές τις περιστάσεις αλλάξει, το υποκείμενο των δεδομένων πρέπει να ενημερωθεί και να λάβει τη συγκατάθεσή του για τις νέες περιστάσεις. Η πολιτική απορρήτου στον ιστότοπο της εταιρείας είναι ένα συνηθισμένο σημείο για την παρουσίαση αυτών των πληροφοριών.
Εάν ένας υπεύθυνος επεξεργασίας αναθέτει σε τρίτο μέρος την επεξεργασία δεδομένων για λογαριασμό του, πρέπει επίσης να υπάρχει συμβατική συμφωνία μεταξύ τους (άρθρο 28 ΓΚΠΔ). Ο εκτελών την επεξεργασία δεδομένων πρέπει να εφαρμόζει κατάλληλα μέτρα ασφαλείας και να βοηθά τον υπεύθυνο επεξεργασίας να διασφαλίζει τη συμμόρφωση με τον ΓΚΠΔ. Οι εκτελούντες την επεξεργασία οφείλουν επίσης να ενημερώνουν τον υπεύθυνο επεξεργασίας εάν πιστεύουν ότι μια εντολή παραβιάζει τον ΓΚΠΔ και να βοηθούν τον υπεύθυνο επεξεργασίας στην εκπλήρωση αιτημάτων για τα δικαιώματα των υποκειμένων των δεδομένων.
Με ορισμένες εξαιρέσεις, οι υπεύθυνοι επεξεργασίας δεδομένων δεν μπορούν να διατηρούν τα δεδομένα για μεγαλύτερο χρονικό διάστημα από αυτό που είναι απαραίτητο για την ολοκλήρωση του σκοπού για τον οποίο συλλέχθηκαν (άρθρο 5 ΓΚΠΔ). Υποχρεούνται να τα διαγράφουν κατόπιν αιτήματος του υποκειμένου των δεδομένων και να ενημερώνουν το υποκείμενο μετά την ολοκλήρωση του αιτήματος (άρθρο 17 ΓΚΠΔ). Τα υποκείμενα των δεδομένων έχουν επίσης το δικαίωμα να ανακαλέσουν τη συγκατάθεσή τους για τη συλλογή και επεξεργασία των δεδομένων τους ανά πάσα στιγμή βάσει του ΓΚΠΔ, ακόμη και αν προηγουμένως είχαν δώσει τη συγκατάθεσή τους. Οι υπεύθυνοι επεξεργασίας δεδομένων πρέπει να καθιστούν την αλλαγή ή την ανάκληση της συγκατάθεσης εξίσου εύκολη με την παροχή της.
Ο ΓΚΠΔ προβλέπει επίσης συγκεκριμένες περιπτώσεις στις οποίες ένας οργανισμός πρέπει να διορίσει υπεύθυνο προστασίας δεδομένων (ΥΠΔ) (άρθρα 37 έως 39 ΓΚΠΔ), δηλαδή οποιαδήποτε από τις ακόλουθες περιπτώσεις:

- όταν μια δημόσια αρχή ή ένας δημόσιος φορέας ασκεί δραστηριότητες επεξεργασίας δεδομένων
- οι δραστηριότητες επεξεργασίας δεδομένων απαιτούν τακτική, συστηματική και μεγάλης κλίμακας παρακολούθηση των υποκειμένων των δεδομένων
- η επεξεργασία δεδομένων που αφορούν ευαίσθητες κατηγορίες πραγματοποιείται σε μεγάλη κλίμακα, όπως π.χ:
    - γενετικά δεδομένα
    - βιομετρικά δεδομένα
    - ιατρικά δεδομένα

- δεδομένα που μπορούν να αποκαλύψουν τη φυλετική ή εθνοτική καταγωγή
- πολιτικά φρονήματα
- θρησκευτικές ή φιλοσοφικές πεποιθήσεις

Ο οργανισμός πρέπει να παρέχει τα στοιχεία επικοινωνίας του ΥΠΔ στην εποπτική αρχή και να τα δημοσιοποιεί, συνήθως μέσω της πολιτικής απορρήτου του ή στον ιστότοπό του. Το τμήμα 3 (άρθρα 35 και 36 του ΓΚΠΔ) του ΓΚΠΔ περιγράφει τις απαιτήσεις για την εκτίμηση αντικτύπου σχετικά με την προστασία των δεδομένων (DPIA) σε ορισμένες περιπτώσεις. Οι υπεύθυνοι επεξεργασίας δεδομένων έχουν την ευθύνη διενέργειας DPIA για επεξεργασία που ενδέχεται να ενέχει υψηλούς κινδύνους για την ασφάλεια ή τα δικαιώματα προστασίας της ιδιωτικής ζωής των φυσικών προσώπων.
Οι υπεύθυνοι επεξεργασίας πρέπει να τεκμηριώνουν αυτές τις αξιολογήσεις, περιγράφοντας τη διαδικασία, τους κινδύνους που εντοπίστηκαν και τα μέτρα που ελήφθησαν για την αντιμετώπιση των κινδύνων αυτών, διασφαλίζοντας τη συμμόρφωση με τον ΓΚΠΔ και την προστασία των ατομικών δικαιωμάτων. Οι υπεύθυνοι επεξεργασίας πρέπει να λαμβάνουν τη συμβουλή του ΥΠΔ κατά τη διενέργεια της DPIA και πρέπει να διαβουλεύονται με την εποπτική αρχή πριν από την επεξεργασία δεδομένων που η DPIA καθορίζει ότι θα είχε ως αποτέλεσμα υψηλό κίνδυνο που δεν μπορεί να μετριαστεί.
Οι υπεύθυνοι επεξεργασίας και οι εκτελούντες την επεξεργασία πρέπει να τηρούν αρχεία των δραστηριοτήτων επεξεργασίας (άρθρο 30 ΓΚΠΔ). Τα αρχεία θα πρέπει να περιέχουν πληροφορίες σχετικά, μεταξύ άλλων, με

- όνομα και στοιχεία επικοινωνίας του υπευθύνου επεξεργασίας/συνυπεύθυνου επεξεργασίας/εκτελούντος την επεξεργασία
- τυχόν διαβιβάσεις δεδομένων προσωπικού χαρακτήρα σε τρίτες χώρες ή διεθνείς οργανισμούς
- γενικές περιγραφές των τεχνικών και οργανωτικών μέτρων ασφαλείας

Τα αρχεία αυτά αποτελούν ουσιώδες μέρος της απόδειξης της συμμόρφωσης με τον ΓΚΠΔ και πρέπει να τίθενται στη διάθεση της εποπτικής αρχής κατόπιν αιτήματος.
Art. 33 ΓΚΠΔ απαιτεί από τους υπευθύνους επεξεργασίας να κοινοποιούν στην εποπτική αρχή παραβίαση δεδομένων προσωπικού χαρακτήρα «χωρίς αδικαιολόγητη καθυστέρηση» και, σε κάθε περίπτωση, το αργότερο εντός 72 ωρών αφότου ο υπεύθυνος επεξεργασίας λάβει γνώση αυτής. Εάν η κοινοποίηση δεν γίνει εντός 72 ωρών, οι υπεύθυνοι επεξεργασίας οφείλουν να εξηγήσουν γιατί καθυστέρησε. Ο υπεύθυνος επεξεργασίας πρέπει να τεκμηριώνει την παραβίαση δεδομένων και να περιλαμβάνει τα γεγονότα που περιβάλλουν την παραβίαση, τις επιπτώσεις της και τα μέτρα που λαμβάνονται για την αποκατάστασή της. Οι υπεύθυνοι επεξεργασίας οφείλουν επίσης να ενημερώνουν τα υποκείμενα των δεδομένων για την παραβίαση των δεδομένων (άρθρο 34 ΓΚΠΔ) εάν υπάρχει «υψηλός κίνδυνος για τα δικαιώματα και τις ελευθερίες» των υποκειμένων των δεδομένων ως αποτέλεσμα της παραβίασης.

3.1.4 Πώς να συμμορφωθείτε με τον Γενικό Κανονισμό Προστασίας Δεδομένων

Οι επιχειρήσεις που ενεργούν ως υπεύθυνοι επεξεργασίας ή εκτελούντες την επεξεργασία δεδομένων μπορούν να λάβουν διάφορα μέτρα για να συμμορφωθούν με τις απαιτήσεις του ΓΚΠΔ. Ένας έλεγχος απορρήτου δεδομένων του ΓΚΠΔ αξιολογεί τα δεδομένα που επεξεργάζεται και αποθηκεύει ο οργανισμός σας, τις πηγές τους και τη συμμόρφωσή σας με τον ΓΚΠΔ. Επικεντρώνεται σε διάφορους κρίσιμους τομείς, όπως η διαχείριση της συγκατάθεσης, οι πρακτικές ασφάλειας δεδομένων και οι έλεγχοι πρόσβασης, προκειμένου να εντοπιστούν οι κίνδυνοι και οι τομείς που χρήζουν βελτίωσης. Μια λεπτομερής πολιτική

απορρήτου που είναι εύκολα προσβάσιμη στους χρήστες μπορεί να εκπληρώσει τις απαιτήσεις διαφάνειας του ΓΚΠΔ. Βεβαιωθείτε ότι η πολιτική απορρήτου σας παραμένει επικαιροποιημένη εάν υπάρχουν αλλαγές στις πρακτικές χειρισμού των δεδομένων σας και ότι περιλαμβάνει βασικές πληροφορίες που απαιτούνται από τον ΓΚΠΔ, όπως

- τύποι προσωπικών δεδομένων που συλλέγονται
- νομικές βάσεις και σκοπός(οι) για την επεξεργασία των δεδομένων
- πόσο καιρό θα διατηρείτε τα δεδομένα
- δικαιώματα των υποκειμένων των δεδομένων
- πώς τα υποκείμενα των δεδομένων μπορούν να ασκήσουν τα δικαιώματά τους
- πώς τα υποκείμενα των δεδομένων μπορούν να ανακαλέσουν τη συγκατάθεσή τους
- τα στοιχεία επικοινωνίας του ΥΠΔ, εάν ο οργανισμός σας διαθέτει ΥΠΔ

Η συγκατάθεση των χρηστών πρέπει να πληροί όλες τις απαιτήσεις του ορισμού της συγκατάθεσης του ΓΚΠΔ για να είναι έγκυρη και η συγκατάθεση πρέπει να λαμβάνεται χωρίς χειρισμούς. Οι επιχειρήσεις που διαχειρίζονται δεδομένα χρηστών στην ΕΕ μπορούν να χρησιμοποιούν μια πλατφόρμα διαχείρισης συγκαταθέσεων όπως η Usercentrics CMP για τη συλλογή ρητής, ενημερωμένης και νομικά έγκυρης συγκατάθεσης. Το Usercentrics CMP σας δίνει τη δυνατότητα να συλλέγετε τη συγκατάθεση opt-in από χρήστες στην ΕΕ και να καταγράφετε τη συγκατάθεση όπως απαιτείται από τον ΓΚΠΔ. Επιτρέπει τη συλλογή συγκατάθεσης με λεπτομερή κριτήρια, ώστε οι χρήστες να επιτρέπουν τη συγκατάθεση για ορισμένους σκοπούς και να απορρίπτουν τη συγκατάθεση για άλλους. Επιτρέπει επίσης στους χρήστες να αλλάζουν ή να αποσύρουν εύκολα τη συγκατάθεσή τους ανά πάσα στιγμή.

Είτε είστε υπεύθυνος επεξεργασίας είτε εκτελών την επεξεργασία δεδομένων, πρέπει να διατηρείτε λεπτομερή αρχεία των δραστηριοτήτων επεξεργασίας. Οι απαιτούμενες πληροφορίες διαφέρουν ελαφρώς για τους υπευθύνους επεξεργασίας και τους εκτελούντες την επεξεργασία (το άρθρο 30 του ΓΚΠΔ περιγράφει λεπτομερώς τι απαιτείται) και πρέπει να τηρείτε τα σχετικά αρχεία για να αποδεικνύετε τη συμμόρφωση με τις απαιτήσεις του ΓΚΠΔ. Το κεφάλαιο 5 (άρθρα 44 έως 50 ΓΚΠΔ) ασχολείται με τη διαβίβαση δεδομένων από την ΕΕ σε τρίτες χώρες ή διεθνείς οργανισμούς, είτε κατά τη διάρκεια της επεξεργασίας είτε μετά. Η διαβίβαση δεδομένων εκτός ΕΕ απαιτεί μέτρα πέραν των τυπικών, ιδίως για την προστασία των δεδομένων, και συχνά απαιτεί ειδική συμφωνία επάρκειας (άρθρο 45 ΓΚΠΔ).

Οι συμφωνίες επάρκειας επιτρέπουν τη συνεχή επεξεργασία δεδομένων μεταξύ οντοτήτων, οπότε δεν απαιτείται πρόσθετη έγκριση σε τακτική βάση, εκτός εάν αλλάξουν οι όροι της αρχικής συμφωνίας: «όταν η Επιτροπή έχει αποφασίσει ότι η τρίτη χώρα, ένα έδαφος ή ένας ή περισσότεροι συγκεκριμένοι τομείς εντός της εν λόγω τρίτης χώρας ή ο εν λόγω διεθνής οργανισμός διασφαλίζει επαρκές επίπεδο προστασίας».

Κατά την αξιολόγηση της επάρκειας, ορισμένες από τις προϋποθέσεις που εξετάζονται περιλαμβάνουν:

- σχετικοί κανονισμοί
- το κράτος δικαίου και το ιστορικό των ανθρωπίνων δικαιωμάτων
- η δημόσια ασφάλεια
- πρόσβαση των δημόσιων αρχών σε δεδομένα προσωπικού χαρακτήρα
- κανόνες προστασίας δεδομένων
- ύπαρξη ανεξάρτητων εποπτικών αρχών
- άλλες διεθνείς δεσμεύσεις που έχει αναλάβει η τρίτη χώρα ή ο οργανισμός

Ο ΓΚΠΔ απαιτεί την περιοδική επανεξέταση των αποφάσεων επάρκειας, τουλάχιστον ανά τετραετία. Ωστόσο, μπορούν να καταργηθούν, να τροποποιηθούν ή να ανασταλούν ανά

πάσα στιγμή, εάν νέες πληροφορίες αποδεικνύουν ότι η τρίτη χώρα ή ο οργανισμός δεν εγγυάται πλέον επαρκές επίπεδο προστασίας των δεδομένων.
Τα δεδομένα μπορούν ακόμη να διαβιβαστούν σε τρίτη χώρα ή διεθνή οργανισμό χωρίς να υπάρχει συμφωνία επάρκειας, αλλά μόνο εάν ο υπεύθυνος επεξεργασίας ή ο εκτελών την επεξεργασία έχει παράσχει τις κατάλληλες εγγυήσεις (άρθρο 46 ΓΚΠΔ) και μπορεί να τηρήσει και να επιβάλει τα δικαιώματα των υποκειμένων των δεδομένων.
Ελλείψει συμφωνίας επάρκειας ή επιβεβαίωσης των κατάλληλων εγγυήσεων, η διαβίβαση δεδομένων μπορεί να εξακολουθήσει να γίνεται, αλλά μόνο υπό τις ακόλουθες συνθήκες (άρθρο 49 ΓΚΠΔ).

- Το υποκείμενο των δεδομένων έχει ενημερωθεί για τους πιθανούς κινδύνους της διαβίβασης και την έλλειψη απόφασης επάρκειας ή κατάλληλων εγγυήσεων και έχει συναινέσει ρητά.
- Η διαβίβαση είναι απαραίτητη για την εκτέλεση σύμβασης μεταξύ του υπευθύνου επεξεργασίας και του υποκειμένου των δεδομένων.
- Η διαβίβαση είναι απαραίτητη για την εκτέλεση ή τη σύναψη σύμβασης μεταξύ του υπεύθυνου επεξεργασίας και άλλου νομικού/φυσικού προσώπου και είναι προς το συμφέρον του υποκειμένου των δεδομένων.
- Σημαντικοί λόγοι δημοσίου συμφέροντος.
- Για τη θεμελίωση, άσκηση ή υπεράσπιση νομικών αξιώσεων.
- Για την προστασία των ζωτικών συμφερόντων του υποκειμένου των δεδομένων ή άλλων προσώπων, όταν το υποκείμενο των δεδομένων είναι σωματικά ή νομικά ανίκανο να δώσει τη συγκατάθεσή του.
- Η διαβίβαση γίνεται, για συγκεκριμένη περίπτωση, από μητρώο που προορίζεται για την παροχή πληροφοριών στο κοινό, είναι ανοικτή σε διαβούλευση από οποιονδήποτε μπορεί να αποδείξει έννομο συμφέρον και εντός των νομοθεσιών της ΕΕ ή του κράτους μέλους.

3.2Λεπτομερής συζήτηση του HIPAA.

Ο HIPAA είναι ένας κανονισμός που σχετίζεται με την υγειονομική περίθαλψη στις Ηνωμένες Πολιτείες. Πρωταρχικός σκοπός του είναι να θέτει αυστηρούς περιορισμούς στη χρήση προστατευόμενων πληροφοριών υγείας (PHI) από διάφορους οργανισμούς και άτομα του τομέα της υγείας που αναφέρονται ως καλυπτόμενες οντότητες. Σύμφωνα με τον HIPAA, PHI είναι κάθε πληροφορία που μπορεί να αποτελέσει προσωπικό αναγνωριστικό, όπως πληροφορίες χρέωσης, καταστάσεις ψυχικής υγείας, αποτελέσματα ιατρικών εξετάσεων, ιστορικό φαρμακευτικής αγωγής, ασφάλιση κ.λπ. Το HIPAA εφαρμόζεται από το Γραφείο Πολιτικών Δικαιωμάτων του Υπουργείου Υγείας και Ανθρωπίνων Υπηρεσιών των ΗΠΑ. Κάθε οργανισμός που παραβιάζει τον κανονισμό μπορεί να αντιμετωπίσει πρόστιμα και ποινές ανάλογα με τη σοβαρότητα του αδικήματός του. Η εν λόγω σοβαρότητα καθορίζεται με βάση ένα σύστημα βαθμίδων ειδικά σχεδιασμένο για την αντιμετώπιση των διαφόρων βαθμών παραβάσεων που μπορεί να συμβούν καταλλήλως. Ο νόμος περί φορητότητας και λογοδοσίας της ασφάλισης υγείας (Health Insurance Portability and Accountability Act), κοινώς γνωστός ως HIPAA, θεσπίστηκε το 1996 για να διασφαλίσει την ιδιωτικότητα και την ασφάλεια των δεδομένων που σχετίζονται με την υγεία των ατόμων, τα οποία ονομάζονται επίσης προστατευόμενες πληροφορίες υγείας (Protected Health Information - PHI).
Παραδείγματα PHI περιλαμβάνουν:

- Το πλήρες όνομα του ατόμου.

- Στοιχεία κατοικίας περιλαμβάνουν την πόλη, την οδό, τη χώρα και τον ταχυδρομικό κώδικα (οτιδήποτε πιο λεπτομερές από την πολιτεία).
- Βασικά γεγονότα ζωής και σχετικές ημερομηνίες, όπως γέννηση, εισαγωγή στο νοσοκομείο, εξιτήριο, θάνατος, και ακριβής ηλικία για όσους είναι 90 ετών και άνω, εκτός από τις λεπτομέρειες του έτους.
- Αριθμοί επικοινωνίας, τόσο σταθερού όσο και κινητού τηλεφώνου.
- Αριθμοί που σχετίζονται με συσκευές φαξ.
- Στοιχεία επικοινωνίας μέσω ηλεκτρονικού ταχυδρομείου.
- Αριθμοί κοινωνικής ασφάλισης.
- Αριθμοί ιατρικών φακέλων.
- Αριθμός που σχετίζεται με τις παροχές του προγράμματος υγείας ενός ατόμου.
- Αναγνωριστικό τραπεζικού ή χρηματοοικονομικού λογαριασμού.
- Αριθμοί που σχετίζονται με άδειες ή πιστοποιητικά.
- Αναγνώριση αυτοκινήτων και των σχετικών σειριακών αριθμών τους, συμπεριλαμβανομένων των στοιχείων πινακίδων κυκλοφορίας.
- Αναγνωριστικά που σχετίζονται με συσκευές και τους σχετικούς σειριακούς αριθμούς τους.
- Διευθύνσεις ιστοτόπων.
- Διευθύνσεις IP που χρησιμοποιούνται για συνδέσεις στο διαδίκτυο.
- Βιομετρικά αναγνωριστικά, όπως μοτίβα φωνής ή δακτυλικών αποτυπωμάτων.
- Φωτογραφικές εικόνες και άλλα μοναδικά χαρακτηριστικά αναγνώρισης.
- Οποιοδήποτε διακριτικό γνώρισμα ή λεπτομέρεια που θα μπορούσε να ξεχωρίσει το άτομο.

Το λογισμικό συμμόρφωσης HIPAA μπορεί να σας βοηθήσει να διασφαλίσετε την προστασία του PHI, ώστε να μπορείτε να επικεντρωθείτε σε άλλες πτυχές της εφαρμογής σας στον τομέα της υγειονομικής περίθαλψης.

Οι ακόλουθες κατηγορίες εμπίπτουν στο πεδίο εφαρμογής του HIPAA και πρέπει να συμμορφώνονται με τους κανονισμούς του:

Καλυπτόμενες οντότητες (ΚΑΟ):

- Πάροχοι υγειονομικής περίθαλψης: Περιλαμβάνονται γιατροί, κλινικές, ψυχολόγοι, οδοντίατροι, χειροπρακτικοί, γηροκομεία, ακόμη και φαρμακεία. Ουσιαστικά, κάθε οργανισμός ή άτομο υγειονομικής περίθαλψης που προσφέρει ιατρικές υπηρεσίες ή θεραπεία.
- Προγράμματα υγείας: Αυτές οι οντότητες παρέχουν ή πληρώνουν για ιατρική περίθαλψη, όπως οι ασφαλιστικές εταιρείες υγείας, οι οργανισμοί συντήρησης υγείας, τα εταιρικά προγράμματα υγείας και τα κυβερνητικά προγράμματα όπως το Medicare και το Medicaid.
- Εκκαθαριστικά κέντρα υγειονομικής περίθαλψης: Οργανισμοί που επεξεργάζονται ή διευκολύνουν την επεξεργασία των πληροφοριών υγείας που λαμβάνονται από άλλη οντότητα σε τυποποιημένη μορφή.

Επιχειρηματικοί συνεργάτες (ΕΠ):

- Κάθε άτομο ή οργανισμός που εκτελεί ορισμένες λειτουργίες ή δραστηριότητες που αφορούν τη χρήση ή αποκάλυψη Προστατευόμενες Χαρακτηριστικές Πληροφορίες (ΠΧΠ) για λογαριασμό ή την παροχή υπηρεσιών σε καλυπτόμενη οντότητα. Αυτό περιλαμβάνει:
    1. Διαχειριστές τρίτων που βοηθούν στο χειρισμό προγραμμάτων υγείας.

2. Πάροχοι και σύμβουλοιπληροφορικής που έχουν πρόσβαση σε ηλεκτρονικά ΠΧΠ.
3. Εταιρείες τιμολόγησης και κωδικοποίησης.
4. Υπηρεσίες τηλεφωνητή.
5. Υπηρεσίες ιατρικής μεταγραφής.
6. Ψηφιακές πλατφόρμες υγείας που αποθηκεύουν ή επεξεργάζονται ΠΧΠ.

Υπεργολάβοι και προμηθευτές:

- Οι οντότητες ή τα άτομα που προσλαμβάνονται από τους επιχειρηματικούς συνεργάτες και ενδέχεται να έρθουν σε επαφή, να επεξεργαστούν ή να αποθηκεύσουν ΠΧΠ πρέπει επίσης να συμμορφώνονται με τον HIPAA.

Οι απαιτήσεις του HIPAA για τις καλυπτόμενες επιχειρήσεις είναι:

- Εφαρμογή πολιτικών και διαδικασιών για την προστασία του απορρήτου των προστατευόμενων πληροφοριών υγείας,
- Εφαρμογή μέτρων ασφάλειας δεδομένων για τη διασφάλιση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας των ηλεκτρονικών ΠΧΠ μέσω διοικητικών, φυσικών και τεχνικών διασφαλίσεων,
- Να παρέχουν στους ασθενείς ειδοποίηση για τις πρακτικές προστασίας της ιδιωτικής ζωής, με λεπτομερή περιγραφή του τρόπου με τον οποίο μπορούν να χρησιμοποιηθούν οι πληροφορίες τους και των δικαιωμάτων τους όσον αφορά τις προστατευόμενες χαρακτηριστικές πληροφορίες της υγείας τους.
- Διασφάλιση των δικαιωμάτων των ασθενών να ζητούν προστασία της ιδιωτικής ζωής, να έχουν πρόσβαση στα ΠΧΠ τους, να ζητούν τροποποιήσεις στα ΠΧΠ τους και να λαμβάνουν λογιστική καταγραφή των κοινοποιήσεων,
- Να χρησιμοποιεί ή να αποκαλύπτει την ελάχιστη ποσότητα ΠΧΠ που είναι απαραίτητη για την επίτευξη του επιδιωκόμενου σκοπού,
- Εκπαιδεύστε τα μέλη του εργατικού δυναμικού στις πολιτικές και τις διαδικασίες HIPAA και διαχειριστείτε τα για να διασφαλίσετε τη συμμόρφωση,
- Να συνάπτουν συμβάσεις με επιχειρηματικούς συνεργάτες για να διασφαλίζουν την προστασία των ΠΧΠ που δημιουργούν, λαμβάνουν, διατηρούν ή διαβιβάζουν,
- Εφαρμογή διαδικασιών για τον εντοπισμό, την αντιμετώπιση και την τεκμηρίωση περιστατικών ασφαλείας και παραβιάσεων και την ενημέρωση των επηρεαζόμενων ατόμων και αρχών, όπως απαιτείται,
- Διατήρηση της τεκμηρίωσης των προσπαθειών συμμόρφωσης με τον HIPAA, συμπεριλαμβανομένων των πολιτικών, των διαδικασιών και του εκπαιδευτικού υλικού, για τουλάχιστον έξι έτη,
- Τακτική αναθεώρηση και επικαιροποίηση των πολιτικών και διαδικασιών HIPAA για τη συμμόρφωση με τις αλλαγές του νόμου και τη διασφάλιση της συνεχούς προστασίας των ΠΧΠ.

3.3 Απαιτήσεις προστασίας της ιδιωτικής ζωής για τα δεδομένα υγειονομικής περίθαλψης στο πλαίσιο αυτών των κανονισμών.

Οι απαιτήσεις προστασίας της ιδιωτικής ζωής για τα δεδομένα υγειονομικής περίθαλψης στο πλαίσιο των κανονισμών GDPR και HIPAA έχουν αρκετές ομοιότητες, αλλά και σημαντικές διαφορές όσον αφορά το πεδίο εφαρμογής, τις υποχρεώσεις των οργανισμών και τα δικαιώματα των ατόμων. Το GDPR είναι ένας γενικός κανονισμός που καλύπτει όλα τα προσωπικά δεδομένα των κατοίκων της Ευρωπαϊκής Ένωσης, συμπεριλαμβανομένων των δεδομένων υγείας. Ορίζει ότι οποιοσδήποτε οργανισμός επεξεργάζεται προσωπικά

δεδομένα πρέπει να εφαρμόζει αυστηρά μέτρα για την προστασία της ιδιωτικότητας και της ασφάλειας αυτών των πληροφοριών. Τα δεδομένα υγείας θεωρούνται ευαίσθητα προσωπικά δεδομένα, γεγονός που σημαίνει ότι υπόκεινται σε ακόμη πιο αυστηρούς κανόνες επεξεργασίας. Οι οργανισμοί που επεξεργάζονται δεδομένα υγείας πρέπει να διασφαλίζουν ότι η συλλογή και η χρήση τους είναι απολύτως απαραίτητες και να λαμβάνουν ρητή συγκατάθεση από το άτομο για την επεξεργασία αυτών των δεδομένων, εκτός αν υπάρχει νομική εξαίρεση. Ο κανονισμός απαιτεί επίσης τη διατήρηση της ελάχιστης δυνατής ποσότητας δεδομένων, δηλαδή οι οργανισμοί πρέπει να συλλέγουν μόνο τις πληροφορίες που είναι απαραίτητες για τον σκοπό της επεξεργασίας. Το HIPAA, από την άλλη πλευρά, ισχύει μόνο για δεδομένα υγείας και μόνο στις Ηνωμένες Πολιτείες. Ορίζει αυστηρούς κανόνες για την προστασία των προστατευόμενων δεδομένων υγείας και απαιτεί από τους παρόχους υγειονομικής περίθαλψης, τις ασφαλιστικές εταιρείες υγείας και τους συνεργάτες τους να εφαρμόζουν μέτρα για την προστασία της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας των δεδομένων. Οι οργανισμοί που υπάγονται στο HIPAA πρέπει να διασφαλίζουν ότι έχουν διαδικασίες που αποτρέπουν τη μη εξουσιοδοτημένη πρόσβαση ή αποκάλυψη των δεδομένων, ενώ επιτρέπεται η χρήση και η κοινοποίηση αυτών των πληροφοριών μόνο για σκοπούς υγειονομικής περίθαλψης, πληρωμών ή λειτουργιών υγειονομικής περίθαλψης. Σε αντίθεση με το GDPR, το HIPAA δεν απαιτεί γενική συγκατάθεση από το άτομο για τη χρήση των δεδομένων του, εκτός εάν πρόκειται για σκοπούς πέρα από αυτούς που προβλέπονται από τον κανονισμό, όπως το μάρκετινγκ. Και οι δύο κανονισμοί απαιτούν την εφαρμογή ισχυρών τεχνικών και οργανωτικών μέτρων ασφαλείας για την προστασία των δεδομένων υγείας από μη εξουσιοδοτημένη πρόσβαση, απώλεια ή αλλοίωση. Οι οργανισμοί πρέπει να έχουν πολιτικές διαχείρισης κινδύνου, να εκπαιδεύουν το προσωπικό τους σχετικά με τις πρακτικές προστασίας δεδομένων και να διασφαλίζουν ότι οι τρίτοι πάροχοι που συνεργάζονται με αυτούς συμμορφώνονται επίσης με τα πρότυπα προστασίας δεδομένων. Μια βασική απαίτηση και των δύο κανονισμών είναι η ειδοποίηση σε περίπτωση παραβίασης δεδομένων. Το GDPR απαιτεί από τους οργανισμούς να αναφέρουν κάθε παραβίαση δεδομένων στην αρμόδια αρχή προστασίας δεδομένων εντός 72 ωρών από τη στιγμή που την εντοπίσουν. Το HIPAA επιβάλλει διαφορετικούς χρόνους ειδοποίησης, ανάλογα με τη σοβαρότητα και το μέγεθος της παραβίασης. Αν η παραβίαση επηρεάζει περισσότερα από 500 άτομα, οι οργανισμοί υποχρεούνται να ενημερώσουν τόσο το Office for Civil Rights (OCR) όσο και τους επηρεαζόμενους μέσα σε 60 ημέρες. Μια άλλη σημαντική διαφορά μεταξύ των δύο κανονισμών αφορά τον διορισμό υπεύθυνου προστασίας δεδομένων. Το GDPR απαιτεί από ορισμένους οργανισμούς, ιδιαίτερα εκείνους που επεξεργάζονται δεδομένα μεγάλης κλίμακας, να ορίσουν έναν Υπεύθυνο Προστασίας Δεδομένων, ο οποίος είναι υπεύθυνος για την παρακολούθηση της συμμόρφωσης του οργανισμού με τον κανονισμό. Αντίθετα, το HIPAA απαιτεί τον διορισμό ενός Υπευθύνου Ιδιωτικότητας για τους οργανισμούς που καλύπτει, αλλά δεν θέτει τις ίδιες αυστηρές απαιτήσεις με το GDPR το οποίο απαιτεί τη διεξαγωγή Εκτίμησης Αντικτύπου στην Προστασία Δεδομένων όταν πρόκειται για επεξεργασία υψηλού κινδύνου, ενώ το HIPAA απαιτεί ετήσια εκτίμηση κινδύνου για όλους τους οργανισμούς που επεξεργάζονται προστατευόμενα δεδομένα υγείας. Το GDPR επιπλέον απαγορεύει τις μη ασφαλείς διεθνείς μεταφορές δεδομένων, εκτός αν υπάρχουν συγκεκριμένοι μηχανισμοί συμμόρφωσης, ενώ το HIPAA δεν έχει αντίστοιχες απαιτήσεις. Συνοψίζοντας, και οι δύο κανονισμοί έχουν ως στόχο την προστασία των δεδομένων υγείας και περιλαμβάνουν αυστηρά μέτρα ασφαλείας, διαδικασίες συμμόρφωσης και πολιτικές διαχείρισης παραβιάσεων. Ωστόσο, το GDPR έχει ευρύτερο πεδίο εφαρμογής, καθώς αφορά όλα τα προσωπικά δεδομένα, απαιτεί μεγαλύτερη

διαφάνεια στη χρήση των δεδομένων και επιβάλλει αυστηρότερα πρόστιμα για μη συμμόρφωση. Το HIPAA, από την άλλη, είναι εξειδικευμένο στην προστασία των δεδομένων υγείας στις ΗΠΑ και επικεντρώνεται περισσότερο στη ρύθμιση των παρόχων υγειονομικής περίθαλψης και των συνεργατών τους.

3.4 Σύγκριση του GDPR και του HIPAA.

Το GDPR και το HIPAA είναι δύο κανονιστικά πλαίσια που αποσκοπούν στην προστασία των προσωπικών δεδομένων, αλλά έχουν διαφορετικό πεδίο εφαρμογής και εστίαση. Η κύρια διαφορά τους είναι ότι το HIPAA αφορά αποκλειστικά τα δεδομένα υγείας, ενώ το GDPR ισχύει για όλα τα προσωπικά δεδομένα, συμπεριλαμβανομένων των δεδομένων υγείας. Ωστόσο, όταν συγκρίνουμε τις απαιτήσεις των δύο κανονισμών όσον αφορά την προστασία των δεδομένων υγείας, παρατηρούμε ότι ακολουθούν παρόμοια, αν όχι ίδια, πρότυπα προστασίας. Το HIPAA εφαρμόζεται στις Ηνωμένες Πολιτείες και καλύπτει μόνο οργανισμούς που διαχειρίζονται δεδομένα υγείας (PHI - Protected Health Information), όπως νοσοκομεία, ασφαλιστικές εταιρείες υγείας και επαγγελματίες του κλάδου. Το GDPR, από την άλλη πλευρά, έχει ευρύτερο πεδίο εφαρμογής, καθώς προστατεύει όλα τα προσωπικά δεδομένα των κατοίκων της Ευρωπαϊκής Οικονομικής Ζώνης (ΕΟΧ), ανεξάρτητα από το αν η εταιρεία που τα επεξεργάζεται εδρεύει στην Ευρωπαική Ένωση ή όχι.

Και οι δύο κανονισμοί απαιτούν την εφαρμογή ισχυρών μέτρων ασφαλείας για την πρόληψη παραβιάσεων δεδομένων και μη εξουσιοδοτημένης πρόσβασης στις ευαίσθητες πληροφορίες. Επιπλέον, επιβάλλουν τη διατήρηση της ελάχιστης δυνατής ποσότητας δεδομένων για την επεξεργασία, δηλαδή ότι οι οργανισμοί επιτρέπεται να συλλέγουν και να επεξεργάζονται μόνο τα απολύτως απαραίτητα δεδομένα για τον σκοπό που επιδιώκουν. Και το GDPR και το HIPAA υποχρεώνουν τους οργανισμούς να ανταποκρίνονται σε αιτήματα των υποκειμένων δεδομένων ή των ασθενών, όπως η πρόσβαση, διόρθωση και διαγραφή των δεδομένων τους, καθώς και άλλες μορφές αιτημάτων που σχετίζονται με την προστασία της ιδιωτικότητας. Επίσης, απαιτούν από τους οργανισμούς να παρέχουν σαφείς ενημερώσεις για τις πολιτικές απορρήτου τους, ώστε τα άτομα να γνωρίζουν πώς χρησιμοποιούνται τα δεδομένα τους.

Σε περίπτωση παραβίασης δεδομένων, και οι δύο κανονισμοί επιβάλλουν υποχρεωτική αναφορά της παραβίασης. Σύμφωνα με το GDPR, η αναφορά πρέπει να γίνει στην αρμόδια ρυθμιστική αρχή εντός 72 ωρών, ενώ το HIPAA προβλέπει διαφορετικούς χρόνους ειδοποίησης, ανάλογα με τη σοβαρότητα και το μέγεθος της παραβίασης. Αν η παραβίαση στο πλαίσιο του HIPAA επηρεάζει περισσότερα από 500 άτομα, τότε οι οργανισμοί πρέπει να ενημερώσουν τόσο το OCR όσο και τα επηρεαζόμενα άτομα εντός 60 ημερών.

Οι κυρώσεις για παραβίαση των κανονισμών είναι διαφορετικές. Το GDPR προβλέπει πολύ αυστηρότερα πρόστιμα, τα οποία μπορεί να φτάσουν έως 20 εκατομμύρια ευρώ ή το 4% του παγκόσμιου ετήσιου τζίρου μιας εταιρείας, όποιο είναι μεγαλύτερο. Αντίθετα, το HIPAA έχει βαθμιδωτό σύστημα προστίμων, με τα ποσά να κυμαίνονται από 100 έως 50.000 δολάρια ανά παράβαση, με ανώτατο ετήσιο πρόστιμο τα 1,5 εκατομμύρια δολάρια για επαναλαμβανόμενες παραβάσεις.

Μια βασική διαφορά μεταξύ των δύο κανονισμών είναι η υποχρέωση για την ύπαρξη Υπεύθυνου Προστασίας Δεδομένων (DPO). Το GDPR απαιτεί από ορισμένες οργανώσεις να ορίζουν Υπεύθυνο Προστασίας Δεδομένων (DPO), ιδίως όταν επεξεργάζονται μεγάλης κλίμακας ή ευαίσθητα δεδομένα. Αντίθετα, το HIPAA απαιτεί τον ορισμό Υπευθύνου Ιδιωτικότητας (Privacy Officer), αλλά δεν έχει την ίδια αυστηρή ρύθμιση με το GDPR.

Τέλος, το GDPR απαιτεί τη διεξαγωγή Εκτίμησης Αντικτύπου στην Προστασία Δεδομένων (DPIA) όταν γίνεται επεξεργασία υψηλού κινδύνου, ενώ το HIPAA απαιτεί ετήσια εκτίμηση κινδύνου για την επεξεργασία προστατευόμενων δεδομένων υγείας.
Συνοψίζοντας, το HIPAA επικεντρώνεται αποκλειστικά στην προστασία των δεδομένων υγείας στις ΗΠΑ, ενώ το GDPR έχει ευρύτερο πεδίο εφαρμογής, καλύπτοντας όλα τα προσωπικά δεδομένα των κατοίκων της ΕΕ. Παρόλο που υπάρχουν διαφορές, και οι δύο κανονισμοί επιβάλλουν παρόμοια μέτρα για την προστασία των δεδομένων υγείας, διασφαλίζοντας την ασφάλεια, την ιδιωτικότητα και την υπευθυνότητα των οργανισμών που τα επεξεργάζονται.

Κατηγορία	**GDPR**	**HIPAA**
Πεδίο εφαρμογής	Το GDPR ισχύει για την επεξεργασία όλων των προσωπικών δεδομένων που ανήκουν σε κατοίκους της ΕΕ, συμπεριλαμβανομένων των δεδομένων υγείας.	Το HIPAA ισχύει μόνο για την επεξεργασία των PHI (προστατευόμενων δεδομένων υγείας) και ePHI εντός των ΗΠΑ.
Εφαρμοσιμότητα	Το GDPR ισχύει για όλες τις οργανώσεις που στοχεύουν ή συλλέγουν προσωπικά δεδομένα (PII) ανεξάρτητα από το αν δραστηριοποιούνται φυσικά εντός της ΕΕ ή όχι.	Το HIPAA ισχύει για οργανισμούς που χειρίζονται δεδομένα υγείας πολιτών των ΗΠΑ, συγκεκριμένα για ασφαλιστικά ταμεία υγείας, κέντρα εκκαθάρισης δεδομένων υγείας και παρόχους υπηρεσιών υγείας που λειτουργούν στις ΗΠΑ.
Σκοπός	Δίνει έμφαση στα δικαιώματα των ατόμων σχετικά με τα προσωπικά τους δεδομένα, συμπεριλαμβανομένου του δικαιώματος πρόσβασης, διόρθωσης και διαγραφής αυτών. Δίνει επίσης έμφαση στην ασφάλεια των δεδομένων και την ειδοποίηση παραβιάσεων.	Εστιάζει στην προστασία και την εμπιστευτικότητα των PHI, με έμφαση στη διαθεσιμότητα και την ακεραιότητα των δεδομένων υγείας, καθώς και στην προστασία από μη εξουσιοδοτημένη πρόσβαση.
Ρυθμιστική αρχή	Το GDPR εφαρμόζεται σε κάθε κράτος-μέλος της ΕΕ μέσω της εθνικής αρχής προστασίας δεδομένων του εκάστοτε κράτους.	Το HIPAA εφαρμόζεται από το Υπουργείο Υγείας και Ανθρωπίνων Υπηρεσιών των ΗΠΑ (HHS) μέσω του Γραφείου Πολιτικών Δικαιωμάτων.
Ποινές	Οι οργανισμοί που παραβιάζουν το GDPR μπορούν να αντιμετωπίσουν πρόστιμα έως και 4% του παγκόσμιου ετήσιου τζίρου τους ή 20 εκατομμύρια ευρώ, όποιο είναι υψηλότερο.	Τα πρόστιμα βάσει του HIPAA εξαρτώνται από τη σοβαρότητα της παράβασης και επιβάλλονται με ένα κλιμακωτό σύστημα, με μέγιστο πρόστιμο τα 2.067.815 δολάρια ετησίως.

Κατηγορία	GDPR	HIPAA
Δικαιώματα Υποκειμένων Δεδομένων	Το GDPR παρέχει στους πολίτες της ΕΕ μια σειρά δικαιωμάτων, συμπεριλαμβανομένου του δικαιώματος πρόσβασης, διόρθωσης, διαγραφής, φορητότητας και εναντίωσης στην επεξεργασία των δεδομένων τους.	Παρομοίως, το HIPAA δίνει στους ασθενείς το δικαίωμα πρόσβασης, τροποποίησης και διόρθωσης των PHI τους.
Ειδοποίηση Παραβίασης Δεδομένων	Σύμφωνα με το GDPR, το μέγεθος της παραβίασης δεν έχει σημασία. Το GDPR επιβάλλει προθεσμία 72 ωρών για την αναφορά όλων των παραβιάσεων και απαιτεί από τους παρόχους να ενημερώσουν τις εποπτικές αρχές.	Σύμφωνα με το HIPAA, οι οργανισμοί υποχρεούνται να ενημερώνουν τα επηρεαζόμενα άτομα για παραβιάσεις το συντομότερο δυνατό. Αν η παραβίαση επηρεάζει πάνω από 500 άτομα, η οργάνωση πρέπει να ενημερώσει το OCR και όλους τους επηρεαζόμενους εντός 60 ημερών.

Πίνακας 1: Συγκριτική Ανάλυση των Κανονισμών GDPR και HIPAA για την Προστασία Δεδομένων Υγείας

3.5 Επιπτώσεις αυτών των κανονισμών στις πρακτικές μηχανικής μάθησης.

Οι κανονισμοί GDPR (General Data Protection Regulation) και HIPAA (Health Insurance Portability and Accountability Act) έχουν σημαντικές επιπτώσεις στις πρακτικές μηχανικής μάθησης, ιδιαίτερα όσον αφορά τη συλλογή, αποθήκευση, επεξεργασία και ανάλυση δεδομένων. Αυτοί οι κανονισμοί στοχεύουν στην προστασία των προσωπικών και ευαίσθητων δεδομένων, γεγονός που επηρεάζει άμεσα την ανάπτυξη και λειτουργία αλγορίθμων μηχανικής μάθησης, καθώς και τις διαδικασίες ανάλυσης δεδομένων. Ο GDPR, που εφαρμόζεται στην Ευρωπαϊκή Ένωση, δίνει στους πολίτες τον έλεγχο των προσωπικών τους δεδομένων και απαιτεί από τις εταιρείες να εφαρμόζουν αυστηρούς μηχανισμούς προστασίας. Ένας από τους βασικούς περιορισμούς που επιβάλλει είναι η ανάγκη για σαφή και ρητή συγκατάθεση των χρηστών προτού γίνει συλλογή και επεξεργασία των δεδομένων τους. Αυτό σημαίνει ότι τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση αλγορίθμων μηχανικής μάθησης πρέπει να συλλέγονται με διαφάνεια και να είναι απολύτως απαραίτητα για τον συγκεκριμένο σκοπό. Εάν ένας οργανισμός επιθυμεί να χρησιμοποιήσει δεδομένα για διαφορετικό σκοπό από αυτόν για τον οποίο συλλέχθηκαν αρχικά, πρέπει να λάβει εκ νέου συγκατάθεση, κάτι που μπορεί να περιορίσει τη δυνατότητα εκπαίδευσης μεγάλων μοντέλων. Επιπλέον, ο GDPR εισάγει την έννοια του "δικαιώματος στη λήθη", που σημαίνει ότι οι χρήστες μπορούν να ζητήσουν τη διαγραφή των δεδομένων τους. Αυτό δημιουργεί προκλήσεις για τη μηχανική μάθηση, καθώς εάν τα δεδομένα ενός ατόμου χρησιμοποιήθηκαν για την εκπαίδευση ενός μοντέλου, η διαγραφή τους μπορεί να απαιτήσει την αναπροσαρμογή ή την επανεκπαίδευση του μοντέλου, γεγονός που καθιστά τη διαδικασία πιο περίπλοκη και δαπανηρή. Επιπλέον, ο GDPR επιβάλλει την αρχή της ελαχιστοποίησης των δεδομένων, δηλαδή οι οργανισμοί πρέπει να συλλέγουν και να επεξεργάζονται μόνο τα απολύτως απαραίτητα δεδομένα για έναν συγκεκριμένο σκοπό. Αυτό περιορίζει την πρόσβαση σε μεγάλα σύνολα δεδομένων, τα

οποία είναι συχνά κρίσιμα για την ανάπτυξη αποδοτικών αλγορίθμων μηχανικής μάθησης. Ο GDPR απαιτεί επίσης διαφάνεια στις αυτοματοποιημένες αποφάσεις που λαμβάνονται μέσω αλγορίθμων, κάτι που επηρεάζει τις τεχνικές που χρησιμοποιούνται για τη μοντελοποίηση και απαιτεί τη χρήση επεξηγήσιμων αλγορίθμων. Οι πολίτες έχουν το δικαίωμα να γνωρίζουν πώς λαμβάνονται οι αποφάσεις που τους αφορούν, γεγονός που καθιστά δύσκολη τη χρήση "black-box" μοντέλων, όπως τα βαθιά νευρωνικά δίκτυα. Αυτό έχει οδηγήσει στην ανάπτυξη Explainable AI, όπου οι αλγόριθμοι μηχανικής μάθησης σχεδιάζονται με τρόπους που επιτρέπουν την εξήγηση των αποφάσεών τους. Ο HIPAA, από την άλλη, αφορά κυρίως την προστασία των ιατρικών δεδομένων στις Ηνωμένες Πολιτείες και θέτει αυστηρούς κανόνες για την πρόσβαση και την επεξεργασία δεδομένων υγείας. Οι οργανισμοί που αναπτύσσουν αλγορίθμους μηχανικής μάθησης στον τομέα της υγείας πρέπει να εξασφαλίσουν ότι τα δεδομένα των ασθενών είναι ανώνυμα ή ψευδωνυμοποιημένα, ώστε να μην είναι δυνατή η ταυτοποίησή τους. Αυτό μπορεί να επηρεάσει την απόδοση των μοντέλων, καθώς η απομάκρυνση προσωπικών στοιχείων συχνά μειώνει τη λεπτομέρεια των δεδομένων και μπορεί να επηρεάσει τη δυνατότητα εκμάθησης πολύπλοκων σχέσεων μεταξύ των χαρακτηριστικών των δεδομένων.

Επιπλέον, ο HIPAA απαιτεί αυστηρούς ελέγχους πρόσβασης και λογοδοσίας, γεγονός που αυξάνει το κόστος και την πολυπλοκότητα της διαχείρισης δεδομένων. Όσοι αναπτύσσουν μοντέλα μηχανικής μάθησης στον τομέα της υγείας πρέπει να διασφαλίζουν ότι τα δεδομένα αποθηκεύονται με ασφάλεια, ότι η πρόσβαση περιορίζεται αυστηρά και ότι όλες οι δραστηριότητες παρακολουθούνται και καταγράφονται. Για να αντιμετωπιστούν οι προκλήσεις που προκύπτουν από τους κανονισμούς GDPR και HIPAA, οι ερευνητές και οι εταιρείες μηχανικής μάθησης έχουν στραφεί σε τεχνικές που μειώνουν την εξάρτηση από προσωπικά δεδομένα. Η χρήση DP είναι μια προσέγγιση που επιτρέπει την εκπαίδευση μοντέλων διατηρώντας τα δεδομένα ανώνυμα και προστατεύοντας την ατομική ταυτότητα των χρηστών. Επίσης, το federated learning επιτρέπει την εκπαίδευση μοντέλων σε τοπικές συσκευές ή σε αποκεντρωμένα συστήματα, χωρίς να απαιτείται η μεταφορά των δεδομένων σε έναν κεντρικό server, μειώνοντας έτσι τον κίνδυνο παραβίασης των δεδομένων. Παράλληλα, τα synthetic data, δηλαδή τεχνητά δεδομένα που δημιουργούνται με τη χρήση αλγορίθμων, επιτρέπουν την εκπαίδευση μοντέλων χωρίς τη χρήση πραγματικών προσωπικών δεδομένων, παρέχοντας έναν εναλλακτικό τρόπο εκμάθησης χωρίς νομικούς περιορισμούς. Συνολικά, οι κανονισμοί GDPR και HIPAA έχουν αλλάξει ριζικά τις πρακτικές μηχανικής μάθησης, αυξάνοντας την πολυπλοκότητα και το κόστος ανάπτυξης μοντέλων, αλλά παράλληλα προωθώντας καινοτόμες προσεγγίσεις για τη διατήρηση της ιδιωτικότητας και την ασφαλή χρήση δεδομένων. Ενώ οι περιορισμοί μπορεί να δημιουργούν προκλήσεις για τη χρήση δεδομένων μεγάλης κλίμακας, ωθούν επίσης την έρευνα και τη βιομηχανία προς την ανάπτυξη υπεύθυνων και ασφαλών τεχνολογιών τεχνητής νοημοσύνης.

4. Διαφορικά ιδιωτική στοχαστική κάθοδος κλίσης (DP-SGD)

4.1 Λεπτομερής επεξήγηση του αλγορίθμου DP-SGD

Η DP έχει αναδειχθεί ως μια δημοφιλής έννοια της ιδιωτικότητας, η οποία μελετάται εκτενώς στην ερευνητική κοινότητα και χρησιμοποιείται ευρέως σε βιομηχανικές εφαρμογές, ιδίως για την εκπαίδευση μοντέλων ML με εγγυήσεις ιδιωτικότητας. Ο στόχος της εκπαίδευσης με DP στη μοντελοποίηση είναι η εξάλειψη των κινδύνων προστασίας της ιδιωτικότητας. Για παράδειγμα, τα παραδείγματα εκπαίδευσης για τα μοντέλα πρόβλεψης

της προσπέλασης εξαρτώνται συχνά από πληροφορίες που αφορούν διαφορετικές τοποθεσίες. Ελλείψει προστατευτικών γραμμών προστασίας της ιδιωτικής ζωής, τα βάρη του εκπαιδευόμενου μοντέλου θα μπορούσαν να αποκαλύψουν, π.χ., το ιστορικό περιήγησης των χρηστών. Διαισθητικά, η μεθοδολογία DP επιτυγχάνει την εξισορρόπηση μεταξύ ιδιωτικότητας και χρησιμότητας, επιτρέποντας τη διεξαγωγή στατιστικών αναλύσεων και την εκμάθηση με βάση τις ιδιότητες του πληθυσμού, ενώ περιορίζει την επιρροή των (ιδιωτικών) πληροφοριών από κάθε μεμονωμένο χρήστη στο τελικό μοντέλο αποτελεσμάτων.

Ο αλγόριθμος DP-SGD είναι μια τεχνική μηχανικής μάθησης που στοχεύει στην εκπαίδευση μοντέλων με ταυτόχρονη προστασία της ιδιωτικότητας των ατόμων των οποίων τα δεδομένα χρησιμοποιούνται για την εκπαίδευση. Η βασική ιδέα είναι η εισαγωγή προσεκτικά ελεγχόμενου θορύβου στη διαδικασία εκπαίδευσης, ο οποίος συμβάλλει στην απόκρυψη της επιρροής των δεδομένων οποιουδήποτε μεμονωμένου ατόμου στο τελικό μοντέλο. Η διαφοροποιημένη ιδιωτική στοχαστική κάθοδος κλίσης έχει συμβάλει καθοριστικά στην ιδιωτική εκπαίδευση μοντέλων βαθιάς μάθησης παρέχοντας ένα πλαίσιο για τον έλεγχο και την παρακολούθηση της απώλειας ιδιωτικότητας κατά τη διάρκεια της εκπαίδευσης. Καθώς τα μοντέλα βαθιάς μάθησης και γενικότερα τα αλγοριθμικά μοντέλα μηχανικής μάθησης γίνονται όλο και πιο ισχυρά, προκύπτει ένα σοβαρό ζήτημα: τα δεδομένα εκπαίδευσης μπορεί να επηρεάσουν υπερβολικά τις προβλέψεις του μοντέλου, επιτρέποντας την εξαγωγή ευαίσθητων πληροφοριών. Αυτό μπορεί να δημιουργήσει σοβαρά προβλήματα ιδιωτικότητας, ειδικά όταν το μοντέλο εκπαιδεύεται σε ιατρικά δεδομένα, οικονομικές συναλλαγές, δεδομένα χρηστών εφαρμογών κοινωνικής δικτύωσης, ή άλλες ευαίσθητες πληροφορίες.

Τα παραδοσιακά SGD-based μοντέλα μηχανικής μάθησης δεν έχουν εγγενώς προστασία ιδιωτικότητας, πράγμα που σημαίνει ότι ένας κακόβουλος επιτιθέμενος μπορεί να χρησιμοποιήσει τεχνικές επίθεσης εξαγωγής δεδομένων (model inversion attacks, membership inference attacks) ώστε να αποκαλύψει πληροφορίες από το εκπαιδευμένο μοντέλο. Το DP-SGD έρχεται να λύσει αυτό το πρόβλημα, επιτρέποντας στα μοντέλα να μάθουν από τα δεδομένα χωρίς να αποκαλύπτουν άμεσα πληροφορίες για μεμονωμένα παραδείγματα. Αυτό το επιτυγχάνει μέσω δύο βασικών μηχανισμών περιορισμού των διαβαθμίσεων (Gradient Clipping) και προσθήκη θορύβου στις κλίσεις (Noisy Gradient Updates).

4.1.1 Ο περιορισμός των διαβαθμίσεων

Στα παραδοσιακά μοντέλα μηχανικής μάθησης, κατά την εκπαίδευση με τη μέθοδο SGD, οι κλίσεις των συναρτήσεων κόστους καθορίζουν το πώς θα ενημερωθούν τα βάρη του μοντέλου. Ωστόσο, σε σύνολα δεδομένων που περιέχουν ευαίσθητες πληροφορίες, ορισμένα μεμονωμένα δεδομένα ενδέχεται να έχουν μεγαλύτερη επίδραση στις ενημερώσεις των βαρών απ’ ό,τι άλλα, κάτι που μπορεί να οδηγήσει σε διαρροή πληροφοριών και μειωμένη ιδιωτικότητα. Ο περιορισμός των διαβαθμίδεων (clipping) λειτουργεί περιορίζοντας το μέγεθος κάθε κλίσης έτσι ώστε κανένα σημείο δεδομένων να μην μπορεί να επηρεάσει υπερβολικά την εκπαίδευση του μοντέλου. Με αυτόν τον τρόπο περιορίζεται η συνεισφορά των "ακραίων" παραδειγμάτων εκπαίδευσης (outliers) που μπορεί να έχουν μεγάλες κλίσεις, μειώνεται ο κίνδυνος ενός επιτιθέμενου να ανακτήσει πληροφορίες ενός συγκεκριμένου δείγματος δεδομένων μέσω επιθέσεων ανάκτησης

δεδομένων (e.g., membership inference attacks) και εξασφαλίζεται μια ομοιόμορφη συμβολή όλων των δεδομένων στην εκπαίδευση, μειώνοντας τη διαρροή πληροφορίας.
Η τυπική τεχνική περιορισμού των διαβαθμίσεων εφαρμόζεται σε κάθε mini-batch δεδομένων και εκτελείται ως εξής:

- Υπολογίζουμε την κλίση g_i για κάθε δεδομένο i στο mini-batch.
- Υπολογίζουμε τον Ευκλείδειο κανόνα (norm) της κλίσης ‖gi‖.
- Εφαρμόζουμε περιορισμό των διαβαθμίδεων (clipping) με βάση ένα προκαθορισμένο όριο

$$g_i' = \frac{g_i}{\max(1, \frac{\|g_i\|}{C})}$$

όπου:

- gi′ είναι η νέα περιορισμένη κλίση μετά τον περιορισμό των διαβαθμίδεων.
- C είναι το όριο περιορισμού των διαβαθμίδεων (clipping threshold).
- Αν ‖gi‖είναι μικρότερο από C, τότε η κλίση παραμένει η ίδια.
- Αν ‖gi‖ είναι μεγαλύτερο από C, τότε μειώνεται αναλογικά ώστε να έχει μέγεθος ακριβώς ίσο με C.

Με αυτόν τον τρόπο, οι κλίσεις δεν μπορούν να υπερβούν ένα συγκεκριμένο όριο, διατηρώντας έτσι μια ισορροπημένη επίδραση στη διαδικασία εκμάθησης.
Η τεχνική gradient clipping είναι απαραίτητη για την εφαρμογή διαφορικής ιδιωτικότητα, καθώς χωρίς αυτήν, η προσθήκη θορύβου στα gradients δεν θα ήταν αποτελεσματική. Εάν ορισμένα δεδομένα παρήγαγαν πολύ μεγάλες κλίσεις, τότε η επίδραση του θορύβου θα ήταν αμελητέα και η ιδιωτικότητα θα κινδύνευε.
Μόλις εφαρμοστεί ο περιορισμός των διαβαθμίδεων το επόμενο βήμα στο DP-SGD είναι η προσθήκη Gaussian θορύβου στις κλίσεις:

$$g_i'' = g_i' + N(0, \sigma^2 C^2)$$

όπου:

- gi′ είναι η περιορισμένη κλίση.
- $N(0,\sigma^2C^2)$ είναι ο προστιθέμενος θόρυβος Gauss.
- σείναι η τυπική απόκλιση του θορύβου, που καθορίζει το επίπεδο διαφορικής ιδιωτικότητας.

Με αυτή την προσέγγιση, η πιθανότητα να αναγνωριστεί ένα μεμονωμένο δείγμα στο σύνολο δεδομένων μειώνεται σημαντικά, καθώς οι αλλαγές στις κλίσεις των βαρών είναι περιορισμένες και "θορυβοποιημένες".
Το όριο περικοπής C είναι ένας κρίσιμος υπερπαράμετρος που επηρεάζει τόσο την ακρίβεια του μοντέλου όσο και το επίπεδο προστασίας της ιδιωτικότητας:

- Μεγάλη τιμή του C: Οι κλίσεις δεν επηρεάζονται πολύ, διατηρώντας μεγαλύτερη ακρίβεια στο μοντέλο. Όμως, η προστασία της ιδιωτικότητας είναι χαμηλότερη καθώς μεμονωμένα σημεία δεδομένων μπορεί να έχουν μεγάλη επίδραση στην εκπαίδευση.
- Μικρή τιμή του C: Όλες οι κλίσεις περικόπτονται έντονα, μειώνοντας την επιρροή των outliers και αυξάνοντας την ιδιωτικότητα. Ωστόσο, αυτό μπορεί να καταστήσει

το μοντέλο λιγότερο ικανό να μάθει από τα δεδομένα, μειώνοντας την τελική του απόδοση.

Στην πράξη, ο καθορισμός του ιδανικού C γίνεται μέσω πειραματισμού, ώστε να επιτευχθεί η καλύτερη ισορροπία μεταξύ ιδιωτικότητας και απόδοσης.

Ο περιορισμός των διαβαθμίσεων χρησιμοποιείται ευρέως σε συστήματα μηχανικής μάθησης που απαιτούν προστασία ιδιωτικότητας, όπως:

- Ιατρικές Εφαρμογές: Εκπαίδευση μοντέλων πάνω σε ιατρικά δεδομένα, όπως ηλεκτρονικοί ιατρικοί φάκελοι, όπου η ιδιωτικότητα των ασθενών είναι κρίσιμη.
- Οικονομικά και Τραπεζικά Δεδομένα: Διατήρηση της εμπιστευτικότητας σε συστήματα ανίχνευσης απάτης ή πιστωτικής βαθμολόγησης.
- Εφαρμογές Κοινωνικών Δικτύων: Βελτίωση συστάσεων ή ανίχνευση ανεπιθύμητου περιεχομένου με χρήση ευαίσθητων δεδομένων χωρίς κίνδυνο αποκάλυψης προσωπικών πληροφοριών.

Ο περιορισμός των διαβαθμίσεων (gradient clipping) είναι μια θεμελιώδης τεχνική που επιτρέπει στο DP-SGD να διατηρεί ιδιωτικότητα κατά την εκπαίδευση ενός μοντέλου, περιορίζοντας τη συνεισφορά κάθε μεμονωμένου δείγματος. Η σωστή ρύθμιση της παραμέτρου clipping threshold C είναι κρίσιμη, καθώς επηρεάζει τόσο την απόδοση του μοντέλου όσο και το επίπεδο προστασίας της ιδιωτικότητας. Σε συνδυασμό με την προσθήκη Gaussian θορύβου, ο περιορισμός των διαβαθμίσεων διασφαλίζει ότι τα δεδομένα των χρηστών δεν μπορούν να ανακτηθούν από το εκπαιδευμένο μοντέλο, κάνοντας το DP-SGD μία από τις πιο αξιόπιστες τεχνικές προστασίας δεδομένων στη μηχανική μάθηση.

4.1.2 Η προσθήκη θορύβου (Noise Addition)

Η προσθήκη θορύβου (Noise Addition) είναι ένα από τα πιο κρίσιμα βήματα στη διαδικασία εκπαίδευσης ενός μοντέλου με DP-SGD, καθώς διαδραματίζει βασικό ρόλο στην επίτευξη διαφορικής ιδιωτικότητας. Αφού εφαρμοστεί ο περιορισμός των διαβαθμίσεων ώστε να περιοριστεί η επιρροή κάθε μεμονωμένου δείγματος δεδομένων, το επόμενο βήμα είναι η προσθήκη τυχαίου θορύβου Gauss στις ενημερώσεις των βαρών του μοντέλου. Αυτό αποτρέπει τη δυνατότητα αναγνώρισης ή εξαγωγής πληροφοριών σχετικά με συγκεκριμένες εγγραφές από το αρχικό σύνολο δεδομένων εκπαίδευσης.

Η βασική ιδέα της προσθήκης θορύβου στηρίζεται στη θεωρία της διαφορικής ιδιωτικότητας. Σύμφωνα με αυτήν, ένα σύστημα που επεξεργάζεται δεδομένα θεωρείται ότι ικανοποιεί τη διαφορική ιδιωτικότητα εάν οι εξόδοι του είναι στατιστικά όμοιες, είτε συμπεριληφθεί είτε όχι μία συγκεκριμένη καταχώρηση στη διαδικασία επεξεργασίας. Με άλλα λόγια, η παρουσία ή απουσία ενός δεδομένου στο σύνολο εκπαίδευσης δεν πρέπει να αλλάζει αισθητά το τελικό μοντέλο. Για να επιτευχθεί αυτό, ο θόρυβος προστίθεται με στοχαστικό τρόπο, εξασφαλίζοντας ότι οι μεμονωμένες εγγραφές καλύπτονται μέσα σε ένα πλαίσιο αβεβαιότητας, ώστε να είναι σχεδόν αδύνατο να απομονωθεί η συμβολή τους.

Η προσθήκη θορύβου είναι κρίσιμη για τη διασφάλιση της ιδιωτικότητας, διότι χωρίς αυτήν, ο περιορισμός των διαβαθμίσεων από μόνος του δεν μπορεί να εμποδίσει τη διαρροή πληροφοριών. Ένα μοντέλο μηχανικής μάθησης που δεν χρησιμοποιεί θόρυβο κατά την ενημέρωση των βαρών του ενδέχεται να είναι ευάλωτο σε επιθέσεις εξαγωγής δεδομένων (model inversion attacks, membership inference attacks). Σε τέτοιου είδους επιθέσεις, ένας εισβολέας μπορεί να αναλύσει τις ενημερώσεις των παραμέτρων του μοντέλου για να ανιχνεύσει ποια δείγματα δεδομένων είχαν τη μεγαλύτερη επίδραση κατά την εκπαίδευση. Αυτό μπορεί να οδηγήσει σε αποκάλυψη προσωπικών δεδομένων, ειδικά σε περιπτώσεις

όπου τα δεδομένα προέρχονται από ιατρικά αρχεία, οικονομικές συναλλαγές ή εμπιστευτικές πληροφορίες χρηστών εφαρμογών.

Ένα επιπλέον πλεονέκτημα της προσθήκης θορύβου είναι ότι αποτρέπει την υπερεκπαίδευση (overfitting) του μοντέλου, καθώς εισάγει μια στοχαστική συνιστώσα που επιτρέπει στο μοντέλο να γενικεύσει καλύτερα στα νέα δεδομένα. Αυτό είναι ιδιαίτερα χρήσιμο σε περιβάλλοντα όπου υπάρχουν μικρά σύνολα δεδομένων, καθώς μειώνει την πιθανότητα το μοντέλο να αποστηθίσει συγκεκριμένα δείγματα και να εμφανίσει υπερβολική προσαρμογή σε αυτά. Ωστόσο, υπάρχει μια σημαντική πρόκληση που σχετίζεται με την προσθήκη θορύβου: η επιλογή της κατάλληλης ποσότητας θορύβου σ. Αν ο θόρυβος είναι πολύ χαμηλός, τότε το επίπεδο ιδιωτικότητας είναι ανεπαρκές και τα δεδομένα μπορεί να διαρρεύσουν. Αντίθετα, αν ο θόρυβος είναι πολύ υψηλός, το μοντέλο μπορεί να χάσει την ικανότητά του να μαθαίνει σωστά από τα δεδομένα, οδηγώντας σε υποβάθμιση της απόδοσής του. Γι' αυτόν τον λόγο, οι ερευνητές και οι μηχανικοί μηχανικής μάθησης πρέπει να βρουν τη σωστή ισορροπία μεταξύ της προστασίας των δεδομένων και της διατήρησης της ακρίβειας του μοντέλου.

Σε πρακτικές εφαρμογές, η προσθήκη θορύβου σε μοντέλα με DP-SGD χρησιμοποιείται σε πολλές πραγματικές περιπτώσεις όπου απαιτείται αυστηρή προστασία των δεδομένων. Ένα χαρακτηριστικό παράδειγμα είναι η εκπαίδευση μοντέλων σε ιατρικά δεδομένα, όπου η ανωνυμοποίηση των δεδομένων των ασθενών είναι απαραίτητη. Άλλες περιπτώσεις χρήσης περιλαμβάνουν τραπεζικές συναλλαγές, ανάλυση δημογραφικών δεδομένων και συστήματα σύστασης, όπου η διασφάλιση της ιδιωτικότητας των χρηστών είναι κρίσιμη. Η προσθήκη θορύβου στο DP-SGD αποτελεί επομένως μια από τις πιο ισχυρές και θεμελιώδεις τεχνικές για την εξασφάλιση της διαφορικής ιδιωτικότητας στα συστήματα μηχανικής μάθησης. Εξασφαλίζει ότι τα μοντέλα μπορούν να εκπαιδευτούν αποτελεσματικά σε ευαίσθητα δεδομένα χωρίς να διακυβεύεται η προστασία των χρηστών, επιτυγχάνοντας έτσι έναν ισορροπημένο συμβιβασμό μεταξύ ακρίβειας και προστασίας της ιδιωτικότητας.

Η διαφορική ιδιωτικότητα αποτελεί ένα θεμελιώδες μαθηματικό πλαίσιο που χρησιμοποιείται για την προστασία της ιδιωτικότητας των δεδομένων κατά την εκπαίδευση αλγορίθμων μηχανικής μάθησης. Ο μηχανισμός της διαφορικής ιδιωτικότητας παρέχει εγγυήσεις ότι η παρουσία ή η απουσία ενός μεμονωμένου δείγματος στο σύνολο εκπαίδευσης δεν επηρεάζει σημαντικά την τελική συμπεριφορά του μοντέλου. Με άλλα λόγια, η έξοδος του εκπαιδευμένου μοντέλου παραμένει στατιστικά σχεδόν ίδια, είτε ένα συγκεκριμένο δεδομένο συμπεριλήφθηκε στην εκπαίδευση είτε όχι. Αυτό σημαίνει ότι ένας κακόβουλος χρήστης που προσπαθεί να ανιχνεύσει την παρουσία συγκεκριμένων δεδομένων στο μοντέλο δεν μπορεί να εξάγει χρήσιμες πληροφορίες με μεγάλη βεβαιότητα, καθώς το μοντέλο δεν αποκαλύπτει άμεσα την επίδραση μεμονωμένων δειγμάτων.

Η βασική αρχή της διαφορικής ιδιωτικότητας βασίζεται στη σύγκριση δύο πιθανών καταστάσεων: μιας στην οποία ένα συγκεκριμένο δείγμα δεδομένων περιλαμβάνεται στο σύνολο εκπαίδευσης και μιας όπου το ίδιο δείγμα απουσιάζει. Ο μηχανισμός θεωρείται ιδιωτικός εάν οι πιθανότητες παραγωγής οποιασδήποτε συγκεκριμένης εξόδου του μοντέλου είναι παρόμοιες και στις δύο περιπτώσεις, εντός ενός προκαθορισμένου περιθωρίου αβεβαιότητας. Αυτή η αβεβαιότητα ελέγχεται μέσω των παραμέτρων διαφορικής ιδιωτικότητας (ε, δ), όπου ε (epsilon) μετρά την ευαισθησία του μοντέλου στις αλλαγές του συνόλου εκπαίδευσης και δ (delta) αντιπροσωπεύει τη μέγιστη πιθανότητα ότι το μοντέλο μπορεί να παραβιάσει την ιδιωτικότητα.

Για να εξασφαλιστεί η διαφορική ιδιωτικότητα στην πράξη, εφαρμόζονται συγκεκριμένοι μηχανισμοί προσθήκης θορύβου στα δεδομένα ή στις παραμέτρους του μοντέλου. Στο

πλαίσιο του DP-SGD, αυτό επιτυγχάνεται μέσω δύο βασικών τεχνικών: ο περιορισμός των διαβαθμίσεων (gradient clipping) και της προσθήκης τυχαίου θορύβου Gauss στις κλίσεις του μοντέλου. Ο περιορισμός των διαβαθμίσεων διασφαλίζει ότι κανένα μεμονωμένο δεδομένο δεν έχει υπερβολικά μεγάλη επιρροή στις παραμέτρους του μοντέλου, ενώ η προσθήκη θορύβου μειώνει τη δυνατότητα ανάκτησης συγκεκριμένων πληροφοριών μέσω της ανάλυσης των ενημερώσεων των βαρών. Ο συνδυασμός αυτών των δύο τεχνικών προσφέρει μια ισχυρή εγγύηση ότι οι αλλαγές στο σύνολο δεδομένων δεν επιφέρουν σημαντικές διαφοροποιήσεις στην έξοδο του μοντέλου.

Η διαφορική ιδιωτικότητα παρέχει επίσης ισχυρή προστασία ενάντια σε διάφορες επιθέσεις εξαγωγής δεδομένων, όπως οι επιθέσεις αντιστροφής μοντέλου (model inversion attacks) και οι επιθέσεις εξαγωγής συμμετοχής (membership inference attacks). Στις επιθέσεις αντιστροφής μοντέλου, ένας εισβολέας προσπαθεί να ανακατασκευάσει τα αρχικά δεδομένα εκπαίδευσης με βάση τις προβλέψεις του μοντέλου, ενώ στις επιθέσεις εξαγωγής συμμετοχής, ο στόχος είναι να διαπιστωθεί αν ένα συγκεκριμένο δείγμα δεδομένων χρησιμοποιήθηκε κατά την εκπαίδευση του μοντέλου. Η χρήση διαφορικής ιδιωτικότητας καθιστά αυτές τις επιθέσεις ιδιαίτερα δύσκολες, καθώς το εκπαιδευμένο μοντέλο δεν διαφέρει ουσιαστικά ανεξάρτητα από το αν ένα δεδομένο περιλαμβάνεται ή όχι στο σύνολο εκπαίδευσης.

Η εφαρμογή του μηχανισμού διαφορικής ιδιωτικότητας είναι ιδιαίτερα σημαντική σε τομείς όπου τα δεδομένα είναι εξαιρετικά ευαίσθητα. Στον κλάδο της ιατρικής πληροφορικής, για παράδειγμα, η εκπαίδευση μοντέλων μηχανικής μάθησης σε δεδομένα ασθενών μπορεί να οδηγήσει σε σοβαρές παραβιάσεις ιδιωτικότητας εάν δεν εφαρμοστούν κατάλληλα μέτρα προστασίας. Η διαφορική ιδιωτικότητα επιτρέπει τη δημιουργία μοντέλων που μπορούν να μάθουν από ιατρικά δεδομένα χωρίς να αποκαλύπτουν προσωπικές πληροφορίες των ασθενών, γεγονός που είναι κρίσιμο για τη συμμόρφωση με κανονισμούς όπως ο GDPR και ο HIPAA. Αντίστοιχα, στον τομέα των οικονομικών συναλλαγών, η χρήση μοντέλων για την ανίχνευση απάτης ή τον υπολογισμό πιστωτικών βαθμολογιών απαιτεί εγγυήσεις ότι τα δεδομένα των χρηστών δεν μπορούν να ανακτηθούν ή να χρησιμοποιηθούν με κακόβουλο τρόπο.

Παρόλο που η διαφορική ιδιωτικότητα παρέχει ισχυρές εγγυήσεις προστασίας, η εφαρμογή της έχει και προκλήσεις. Ένας από τους βασικούς περιορισμούς της είναι ότι η προσθήκη θορύβου μπορεί να μειώσει την ακρίβεια του μοντέλου, καθώς αυξάνει την τυχαιότητα στις ενημερώσεις των βαρών. Επιπλέον, η επιλογή των σωστών τιμών για τις παραμέτρους ε και δ είναι κρίσιμη για την ισορροπία μεταξύ ιδιωτικότητας και απόδοσης. Αν η τιμή του ε είναι πολύ μικρή, το επίπεδο ιδιωτικότητας είναι υψηλό, αλλά το μοντέλο μπορεί να χάσει την ικανότητά του να μάθει αποτελεσματικά από τα δεδομένα. Αντίθετα, αν το ε είναι πολύ μεγάλο, τότε η προστασία της ιδιωτικότητας καθίσταται αδύναμη και μπορεί να υπάρχει κίνδυνος αποκάλυψης δεδομένων.

Η διαφορική ιδιωτικότητα έχει καταστεί ένα από τα πιο αξιόπιστα πρότυπα για την προστασία δεδομένων στη μηχανική μάθηση και χρησιμοποιείται ευρέως σε πραγματικές εφαρμογές. Εταιρείες τεχνολογίας χρησιμοποιούν τεχνικές διαφορικής ιδιωτικότητας για τη συλλογή και ανάλυση δεδομένων χρηστών χωρίς να θυσιάζουν την ιδιωτικότητα των ατόμων. Η εξέλιξη των μεθόδων διαφορικής ιδιωτικότητας αναμένεται να διαδραματίσει καθοριστικό ρόλο στο μέλλον των εφαρμογών μηχανικής μάθησης, ειδικά σε περιβάλλοντα όπου η προστασία της ιδιωτικότητας είναι κρίσιμη. Καθώς η ποσότητα των προσωπικών δεδομένων που χρησιμοποιούνται για την εκπαίδευση μοντέλων συνεχώς αυξάνεται, η ανάγκη για τεχνικές που εξασφαλίζουν την ασφάλεια αυτών των δεδομένων γίνεται ολοένα και πιο σημαντική. Ο μηχανισμός της διαφορικής ιδιωτικότητας αποτελεί μια από τις πιο

αποτελεσματικές προσεγγίσεις για την προστασία της ιδιωτικότητας, εξασφαλίζοντας ότι τα μοντέλα μπορούν να εκπαιδευτούν σε ευαίσθητα δεδομένα χωρίς να εκθέτουν πληροφορίες που θα μπορούσαν να παραβιάσουν τα προσωπικά δικαιώματα των χρηστών.

4.2 Ο Ρόλος της Διαφορικής Ιδιωτικότητας στο DP-SGD

Ο μηχανισμός διαφορικής ιδιωτικότητας στο DP-SGD διασφαλίζει ότι ακόμα και αν προσθέσουμε ή αφαιρέσουμε ένα δεδομένο από το σύνολο εκπαίδευσης, η έξοδος του μοντέλου παραμένει στατιστικά παρόμοια. Αυτό σημαίνει ότι ένας εισβολέας δεν μπορεί να διαπιστώσει αν ένα συγκεκριμένο δεδομένο χρησιμοποιήθηκε ή όχι κατά την εκπαίδευση, μειώνοντας έτσι τον κίνδυνο επιθέσεων membership inference ή model inversion. Οι δύο κύριες παράμετροι που καθορίζουν την ισχύ της διαφορικής ιδιωτικότητας είναι το ε και το δ. Η παράμετρος ε (epsilon) μετράει πόσο επηρεάζει η παρουσία ενός δεδομένου την έξοδο του μοντέλου – όσο μικρότερο είναι το ε, τόσο μεγαλύτερη είναι η προστασία της ιδιωτικότητας. Η παράμετρος δ αντιπροσωπεύει την πιθανότητα παραβίασης της ιδιωτικότητας, δηλαδή το πόσο πιθανό είναι ένας εισβολέας να μπορεί να εξαγάγει πληροφορίες από το μοντέλο. Η απόδοση του DP-SGD εξαρτάται έντονα από τις τιμές των υπερπαραμέτρων του, συγκεκριμένα:

- Το όριο περικοπής των κλίσεων (clipping threshold, C)
- Το επίπεδο θορύβου (noise scale, σ)
- Η παράμετρος διαφορικής ιδιωτικότητας ε\varepsilon

Αν το όριο C είναι πολύ μεγάλο, τότε η ιδιωτικότητα είναι αδύναμη, επειδή τα δεδομένα μπορούν να επηρεάσουν σημαντικά το αποτέλεσμα. Αν το όριο είναι πολύ μικρό, οι κλίσεις περιορίζονται υπερβολικά και το μοντέλο δεν μπορεί να μάθει σωστά από τα δεδομένα.
Ο θόρυβος σ είναι επίσης κρίσιμος. Αν είναι πολύ χαμηλός, το επίπεδο ιδιωτικότητας είναι ανεπαρκές. Αντίθετα, αν ο θόρυβος είναι πολύ υψηλός, το μοντέλο μπορεί να γίνει ανακριβές και να δυσκολεύεται να συγκλίνει σε μία βέλτιστη λύση.
Ο αλγόριθμος DP-SGD συνδυάζει τη δύναμη του στοχαστικού καθοδικού κλίσης με τις εγγυήσεις διαφορικής ιδιωτικότητας, καθιστώντας τον ιδανικό για εφαρμογές που απαιτούν προστασία προσωπικών δεδομένων. Ο περιορισμός των διαβαθμίσεων και η προσθήκη τυχαίου θορύβου εξασφαλίζουν ότι κανένα μεμονωμένο δείγμα δεν μπορεί να επηρεάσει δυσανάλογα την εκπαίδευση του μοντέλου. Αν και η εφαρμογή του DP-SGD συνεπάγεται μια αναπόφευκτη μείωση της ακρίβειας λόγω της προσθήκης θορύβου, η τεχνική αυτή καθίσταται αναγκαία για τη δημιουργία μοντέλων μηχανικής μάθησης που συμμορφώνονται με κανονισμούς ιδιωτικότητας, όπως ο GDPR και ο HIPAA.
Ο αλγόριθμος DP-SGD προσφέρει αυστηρές εγγυήσεις ιδιωτικότητας, διασφαλίζοντας ότι οι πληροφορίες που χρησιμοποιούνται κατά την εκπαίδευση ενός μοντέλου μηχανικής μάθησης δεν μπορούν να ανακτηθούν ή να αποκαλυφθούν, ακόμα και αν κάποιος αποκτήσει πρόσβαση στο τελικό εκπαιδευμένο μοντέλο. Η βασική αρχή πίσω από αυτές τις εγγυήσεις είναι ότι οποιαδήποτε αλλαγή στο σύνολο εκπαίδευσης, είτε πρόκειται για την προσθήκη είτε για την αφαίρεση μιας εγγραφής, δεν θα επηρεάσει σημαντικά την έξοδο του μοντέλου. Αυτό σημαίνει ότι ένας επιτιθέμενος που επιχειρεί να προσδιορίσει αν ένα συγκεκριμένο δείγμα δεδομένων χρησιμοποιήθηκε κατά την εκπαίδευση δεν μπορεί να είναι βέβαιος για την απάντηση, ακόμα και αν έχει πλήρη πρόσβαση στις προβλέψεις ή στις εσωτερικές παραμέτρους του μοντέλου.
Οι εγγυήσεις ιδιωτικότητας του DP-SGD περιγράφονται μαθηματικά μέσω του πλαισίου της διαφορικής ιδιωτικότητας. Σύμφωνα με αυτό το πλαίσιο, ένα σύστημα λέγεται ότι ικανοποιεί (ε, δ)-διαφορική ιδιωτικότητα αν, για οποιοδήποτε σύνολο δεδομένων και για οποιοδήποτε δυνατό αποτέλεσμα του μοντέλου, η πιθανότητα να παραχθεί αυτό το

αποτέλεσμα παραμένει σχεδόν ίδια είτε ένα συγκεκριμένο δείγμα βρίσκεται στο σύνολο εκπαίδευσης είτε όχι. Η τιμή ε (epsilon) εκφράζει τη μέγιστη διαφορά στις πιθανότητες εξόδου του μοντέλου όταν υπάρχει ή δεν υπάρχει ένα συγκεκριμένο δείγμα, ενώ η τιμή δ (delta) προσδιορίζει την πιθανότητα ότι αυτή η εγγύηση μπορεί να παραβιαστεί. Όσο μικρότερη είναι η τιμή του ε, τόσο μεγαλύτερη είναι η προστασία της ιδιωτικότητας, αν και υπερβολικά χαμηλές τιμές μπορεί να οδηγήσουν σε μείωση της ακρίβειας του μοντέλου.
Για να επιτευχθούν αυτές οι εγγυήσεις, ο αλγόριθμος DP-SGD βασίζεται σε δύο κύριες τεχνικές: ο περιορισμός των διαβαθμίσεων (gradient clipping) και την προσθήκη θορύβου (noise addition) στις ενημερώσεις των βαρών του μοντέλου. Ο περιορισμός των διαβαθμίσεων διασφαλίζει ότι κανένα μεμονωμένο δείγμα δεδομένων δεν μπορεί να έχει υπερβολική επίδραση στο μοντέλο, καθώς οι κλίσεις που προκύπτουν από κάθε δείγμα περιορίζονται σε ένα προκαθορισμένο όριο. Αυτή η διαδικασία μειώνει την εξάρτηση του μοντέλου από συγκεκριμένα δεδομένα και αποτρέπει περιπτώσεις όπου μεμονωμένα δείγματα θα μπορούσαν να επηρεάσουν δυσανάλογα το αποτέλεσμα της εκπαίδευσης.
Η προσθήκη τυχαίου θορύβου στις κλίσεις ενισχύει περαιτέρω την προστασία της ιδιωτικότητας, δημιουργώντας μια στοχαστική συνιστώσα που καθιστά δυσκολότερη την εξαγωγή πληροφοριών από το μοντέλο. Ο θόρυβος που προστίθεται ακολουθεί μια κανονική κατανομή (Gaussian noise) με μέσο όρο μηδέν και διακύμανση που σχετίζεται άμεσα με την επιθυμητή στάθμη προστασίας ιδιωτικότητας. Με τον τρόπο αυτό, ακόμα και αν ένας εισβολέας προσπαθήσει να αναλύσει τις ενημερώσεις των βαρών και να εξαγάγει πληροφορίες, το προστιθέμενο στοχαστικό στοιχείο θα λειτουργήσει ως μια "ομπρέλα προστασίας", αποτρέποντας την εξαγωγή οποιασδήποτε αξιόπιστης πληροφορίας.
Οι εγγυήσεις ιδιωτικότητας του DP-SGD είναι ιδιαίτερα σημαντικές σε εφαρμογές όπου χρησιμοποιούνται ευαίσθητα δεδομένα. Στα ιατρικά δεδομένα, για παράδειγμα, η εκπαίδευση ενός νευρωνικού δικτύου πάνω σε ηλεκτρονικούς ιατρικούς φακέλους ασθενών μπορεί να είναι κρίσιμη για τη διάγνωση ασθενειών ή τη βελτίωση της περίθαλψης. Ωστόσο, η διατήρηση της εμπιστευτικότητας αυτών των δεδομένων είναι απολύτως απαραίτητη για τη συμμόρφωση με κανονισμούς όπως ο GDPR και ο HIPAA. Η χρήση διαφορικής ιδιωτικότητας μέσω του DP-SGD διασφαλίζει ότι ακόμα και αν ένα νοσοκομείο μοιραστεί το εκπαιδευμένο μοντέλο του, οι πληροφορίες των ασθενών δεν μπορούν να ανακτηθούν από τις παραμέτρους του μοντέλου.
Μια άλλη σημαντική εφαρμογή του DP-SGD είναι στις τραπεζικές και χρηματοοικονομικές συναλλαγές, όπου η εκπαίδευση μοντέλων πάνω σε δεδομένα συναλλαγών πελατών μπορεί να βοηθήσει στον εντοπισμό δόλιων δραστηριοτήτων. Χωρίς διαφορική ιδιωτικότητα, τέτοιου είδους μοντέλα θα μπορούσαν να είναι ευάλωτα σε επιθέσεις εξαγωγής συμμετοχής (membership inference attacks), όπου ένας εισβολέας προσπαθεί να ανιχνεύσει αν ένας συγκεκριμένος πελάτης συμπεριλαμβάνεται στη βάση δεδομένων εκπαίδευσης. Η διαφορική ιδιωτικότητα καθιστά τέτοιες επιθέσεις αναποτελεσματικές, αφού το τελικό μοντέλο παραμένει ανεξάρτητο από την ύπαρξη ή μη ενός συγκεκριμένου πελάτη στο σύνολο εκπαίδευσης.
Παρότι οι εγγυήσεις ιδιωτικότητας που προσφέρει το DP-SGD είναι ισχυρές, η εφαρμογή του ενέχει και τεχνικές προκλήσεις. Ένα από τα κύρια ζητήματα είναι η εύρεση της βέλτιστης τιμής για την παράμετρο ε (epsilon), καθώς υπάρχει ένας άμεσος συμβιβασμός μεταξύ ιδιωτικότητας και ακρίβειας του μοντέλου. Αν το ε είναι πολύ μικρό, η ποσότητα του θορύβου που προστίθεται στις κλίσεις είναι πολύ μεγάλη, κάτι που μπορεί να μειώσει σημαντικά την ικανότητα του μοντέλου να μάθει χρήσιμα πρότυπα από τα δεδομένα. Αντίθετα, αν το ε είναι πολύ μεγάλο, το επίπεδο ιδιωτικότητας είναι ανεπαρκές, καθώς οι ενημερώσεις των βαρών του μοντέλου μπορεί να περιέχουν αρκετές πληροφορίες που θα

επιτρέψουν την ανάκτηση αρχικών δεδομένων. Επομένως, η επιλογή των σωστών υπερπαραμέτρων είναι κρίσιμη για την επίτευξη ενός ισορροπημένου συνδυασμού μεταξύ προστασίας δεδομένων και απόδοσης του μοντέλου.

4.3 Σύγκριση του DP-SGD με Άλλες Τεχνικές Διατήρησης της Ιδιωτικότητας

Η ανάγκη για προστασία δεδομένων στις εφαρμογές μηχανικής μάθησης έχει οδηγήσει στην ανάπτυξη διαφόρων τεχνικών που διασφαλίζουν την ιδιωτικότητα των χρηστών. Ο αλγόριθμος DP-SGD αποτελεί μια από τις πιο προηγμένες μεθόδους, καθώς ενσωματώνει τη διαφορική ιδιωτικότητα, επιτρέποντας την εκπαίδευση μοντέλων με εγγυήσεις ανωνυμίας. Ωστόσο, υπάρχουν και άλλες προσεγγίσεις που χρησιμοποιούνται για την προστασία της ιδιωτικότητας, όπως η ανωνυμοποίηση δεδομένων (data anonymization), η ομομορφική κρυπτογράφηση (homomorphic encryption), η εκπαίδευση με ομοσπονδιακή μάθηση (federated learning), η προσέγγιση των ασφαλών πολλαπλών υπολογισμών και η DP μέσω post-processing noise injection.

Η ανωνυμοποίηση δεδομένων είναι μια παραδοσιακή μέθοδος προστασίας της ιδιωτικότητας που περιλαμβάνει τη διαγραφή ή την τροποποίηση προσωπικών αναγνωριστικών πληροφοριών πριν τα δεδομένα χρησιμοποιηθούν για την εκπαίδευση ενός μοντέλου. Τεχνικές όπως η ψευδωνυμοποίηση (pseudonymization) και η γενίκευση δεδομένων (data generalization) χρησιμοποιούνται ευρέως για να μειώσουν τον κίνδυνο αναγνώρισης ατόμων. Ωστόσο, η ανωνυμοποίηση από μόνη της δεν αποτελεί ισχυρή εγγύηση ιδιωτικότητας, καθώς έχει αποδειχθεί ότι τεχνικές επαναταυτοποίησης (re-identification attacks) μπορούν να αποκαλύψουν τα αρχικά δεδομένα αν αυτά συνδυαστούν με άλλες πηγές πληροφορίας. Σε αντίθεση, το DP-SGD παρέχει μαθηματικές εγγυήσεις ιδιωτικότητας μέσω της διαφορικής ιδιωτικότητας, καθιστώντας δυσκολότερη την αποκάλυψη μεμονωμένων εγγραφών, ακόμα και αν ένας εισβολέας αποκτήσει πρόσβαση στο τελικό μοντέλο.

Η ομομορφική κρυπτογράφηση επιτρέπει τη διεξαγωγή υπολογισμών πάνω σε κρυπτογραφημένα δεδομένα, χωρίς να απαιτείται αποκρυπτογράφηση. Αυτό καθιστά δυνατή την εκπαίδευση μοντέλων μηχανικής μάθησης σε δεδομένα που παραμένουν κρυπτογραφημένα καθ' όλη τη διάρκεια της επεξεργασίας, αποτρέποντας την άμεση πρόσβαση στα ευαίσθητα δεδομένα. Παρόλο που αυτή η τεχνική παρέχει ένα ισχυρό επίπεδο ιδιωτικότητας, έχει ένα σημαντικό μειονέκτημα: οι υπολογισμοί που εκτελούνται σε κρυπτογραφημένα δεδομένα είναι εξαιρετικά αργοί και απαιτούν μεγάλη υπολογιστική ισχύ. Αντίθετα, το DP-SGD δεν βασίζεται στην κρυπτογράφηση αλλά στην προσθήκη στοχαστικού θορύβου στις κλίσεις, γεγονός που το καθιστά πιο αποδοτικό από άποψη υπολογιστικού κόστους.

Η ομοσπονδιακή μάθηση είναι μια τεχνική που επιτρέπει την εκπαίδευση μοντέλων σε κατανεμημένα σύνολα δεδομένων χωρίς να απαιτείται η συγκέντρωση των δεδομένων σε έναν κεντρικό διακομιστή. Στην πράξη, τα δεδομένα παραμένουν στις συσκευές των χρηστών και μόνο οι ενημερώσεις των μοντέλων ανταλλάσσονται μεταξύ των συσκευών και του κεντρικού εξυπηρετητή. Αυτή η προσέγγιση μειώνει την ανάγκη μεταφοράς προσωπικών δεδομένων, αλλά δεν εγγυάται από μόνη της την προστασία της ιδιωτικότητας, καθώς οι ενημερώσεις των βαρών που ανταλλάσσονται ενδέχεται να περιέχουν πληροφορίες που μπορούν να οδηγήσουν στην αποκάλυψη δεδομένων. Το DP-SGD μπορεί να συνδυαστεί με την ομοσπονδιακή μάθηση, προσθέτοντας διαφορική ιδιωτικότητα στις ενημερώσεις των μοντέλων, μειώνοντας έτσι περαιτέρω τον κίνδυνο διαρροής δεδομένων μέσω επιθέσεων αντιστροφής μοντέλου.

Η προσέγγιση των ασφαλών πολλαπλών υπολογισμών επιτρέπει σε πολλές εμπλεκόμενες πλευρές να υπολογίσουν από κοινού μια συνάρτηση πάνω στα δεδομένα τους, χωρίς κανένας από τους συμμετέχοντες να έχει πρόσβαση στα δεδομένα των άλλων. Αυτή η τεχνική χρησιμοποιείται κυρίως σε σενάρια όπου διαφορετικές οντότητες επιθυμούν να εκπαιδεύσουν ένα μοντέλο χωρίς να μοιραστούν τα αρχικά δεδομένα μεταξύ τους. Αν και προσφέρει υψηλό επίπεδο προστασίας δεδομένων, η χρήση της σε μεγάλης κλίμακας μοντέλα μηχανικής μάθησης είναι πολύπλοκη και απαιτεί μεγάλη υπολογιστική ισχύ. Σε αντίθεση, το DP-SGD είναι πιο εύκολα εφαρμόσιμο, καθώς επιτρέπει την εκπαίδευση σε ένα κεντρικό μοντέλο με μαθηματικά ελεγχόμενες εγγυήσεις ιδιωτικότητας, χωρίς την ανάγκη πολύπλοκων πρωτοκόλλων επικοινωνίας μεταξύ των συμμετεχόντων.

Μια άλλη μέθοδος προστασίας της ιδιωτικότητας που χρησιμοποιείται είναι η διαφορική ιδιωτικότητα μέσω post-processing noise injection, όπου ο θόρυβος προστίθεται μετά την ολοκλήρωση της εκπαίδευσης του μοντέλου, αντί να εφαρμόζεται κατά τη διάρκεια της διαδικασίας εκπαίδευσης όπως στο DP-SGD. Αυτή η τεχνική μπορεί να είναι χρήσιμη σε περιπτώσεις όπου το εκπαιδευμένο μοντέλο ήδη υπάρχει και απαιτείται μια μορφή αναδρομικής προστασίας δεδομένων, αλλά είναι λιγότερο αποτελεσματική σε σύγκριση με το DP-SGD, καθώς οι αλλαγές που γίνονται εκ των υστέρων μπορεί να αλλοιώσουν τις προβλέψεις του μοντέλου και να επηρεάσουν την ακρίβεια του.

Σε σύγκριση με αυτές τις τεχνικές, το DP-SGD παρέχει έναν καλά ισορροπημένο συνδυασμό προστασίας ιδιωτικότητας και απόδοσης. Δεν απαιτεί την πολυπλοκότητα της ομομορφικής κρυπτογράφησης, ενώ ταυτόχρονα προσφέρει αυστηρότερες εγγυήσεις ιδιωτικότητας από την κλασική ανωνυμοποίηση δεδομένων. Μπορεί να χρησιμοποιηθεί αυτόνομα ή σε συνδυασμό με άλλες τεχνικές, όπως η ομοσπονδιακή μάθηση, για να προσφέρει βελτιωμένη προστασία ιδιωτικότητας σε περιβάλλοντα όπου η ιδιωτικότητα των χρηστών είναι κρίσιμη.

Η εξέλιξη των τεχνικών διατήρησης της ιδιωτικότητας συνεχίζεται, καθώς η ανάγκη για ασφαλή και ιδιωτική μηχανική μάθηση αυξάνεται. Το DP-SGD έχει καθιερωθεί ως μια από τις πιο αξιόπιστες τεχνικές, ιδίως σε περιπτώσεις που απαιτούνται θεωρητικά αποδεδειγμένες εγγυήσεις διαφορικής ιδιωτικότητας. Με τη συνεχή βελτίωση των αλγορίθμων και των μεθόδων διατήρησης της ιδιωτικότητας, αναμένεται ότι το DP-SGD θα διαδραματίσει πρωταγωνιστικό ρόλο στη μηχανική μάθηση που βασίζεται στην προστασία δεδομένων, ιδίως στους τομείς της ιατρικής πληροφορικής, των χρηματοοικονομικών υπηρεσιών και της ανάλυσης δεδομένων σε κατανεμημένα περιβάλλοντα.

5. Επίδραση του DP-SGD στην Ακρίβεια και τη Σύγκλιση του Μοντέλου

5.1 Μεθοδολογία για την Αξιολόγηση της Επίδρασης του DP-SGD

Η εφαρμογή του αλγορίθμου DP-SGD επιτρέπει την εκπαίδευση μοντέλων μηχανικής μάθησης με προστασία της ιδιωτικότητας των δεδομένων. Στο παρόν κεφάλαιο, διερευνάται η επίδραση της χρήσης του DP-SGD στην απόδοση ενός μοντέλου ταξινόμησης, εστιάζοντας κυρίως στη μεταβολή της ακρίβειας και στη διαδικασία σύγκλισης. Για την ανάλυση αυτή, το μοντέλο εκπαιδεύεται τόσο με όσο και χωρίς DP, προκειμένου να συγκριθούν τα αποτελέσματα και να αξιολογηθεί ο αντίκτυπος των διαφορετικών privacy budgets. Η διερεύνηση της ακρίβειας αφορά τον βαθμό στον οποίο το μοντέλο διατηρεί την ικανότητά του να διακρίνει σωστά τις διαφορετικές κατηγορίες

όταν εφαρμόζεται το DP-SGD. Στο πλαίσιο αυτό, αναλύεται η μεταβολή της απόδοσης με βάση το privacy budget ε, το οποίο καθορίζει τον βαθμό προστασίας των δεδομένων. Η μείωση της τιμής του ε συνεπάγεται αυξημένη ιδιωτικότητα, αλλά και ενδεχόμενη μείωση της ακρίβειας του μοντέλου. Επομένως, η εύρεση της ισορροπίας μεταξύ προστασίας της ιδιωτικότητας και διατήρησης της χρηστικότητας του μοντέλου αποτελεί σημαντικό ζητούμενο.

Η μελέτη της σύγκλισης επικεντρώνεται στον τρόπο με τον οποίο το DP-SGD επηρεάζει τον ρυθμό εκπαίδευσης του μοντέλου. Στην παραδοσιακή προσέγγιση του SGD, η διαδικασία σύγκλισης εξαρτάται από τον ρυθμό εκμάθησης και το μέγεθος των παρτίδων δεδομένων. Ωστόσο, η εφαρμογή DP επιφέρει διαφοροποιήσεις λόγω του προστιθέμενου θορύβου στις διαβαθμίσεις, γεγονός που μπορεί να προκαλέσει καθυστέρηση στη σταθεροποίηση του μοντέλου. Συνεπώς, αξιολογείται ο αριθμός των επαναλήψεων (epochs) που απαιτούνται για τη σύγκλιση του μοντέλου σε διαφορετικά επίπεδα ιδιωτικότητας.

Για την ανάλυση αυτή, χρησιμοποιείται Heart Disease Dataset, που περιλαμβάνει κλινικά δεδομένα και χρησιμοποιείται συχνά σε μελέτες πρόβλεψης καρδιαγγειακών παθήσεων. Το μοντέλο εκπαιδεύεται σε διαφορετικές συνθήκες, με εναλλαγές στην τιμή του privacy budget και των υπερπαραμέτρων του DP-SGD. Στη συνέχεια, αξιολογούνται οι επιδόσεις του μοντέλου μέσω μετρικών, όπως η ακρίβεια, το precision, το recall και η συνάρτηση απώλειας. Ιδιαίτερη έμφαση δίνεται στην παρακολούθηση της εξέλιξης της συνάρτησης απώλειας, καθώς παρέχει σημαντικές πληροφορίες σχετικά με τον ρυθμό μάθησης και τη σταθερότητα του μοντέλου.

Η ακρίβεια του μοντέλου χρησιμοποιείται ως βασική μετρική αξιολόγησης, καθώς εκφράζει το ποσοστό των σωστών προβλέψεων που πραγματοποιεί. Ωστόσο, η ακρίβεια από μόνη της δεν αρκεί για να περιγράψει την ποιότητα των προβλέψεων, ιδιαίτερα σε ιατρικά δεδομένα όπου το σφάλμα μπορεί να έχει σοβαρές επιπτώσεις. Για τον λόγο αυτό, χρησιμοποιούνται επιπλέον μετρικές, όπως το precision, το recall και το F1-score. Το precision αντικατοπτρίζει το ποσοστό των θετικών προβλέψεων που είναι πραγματικά σωστές, ενώ το recall δείχνει την ικανότητα του μοντέλου να αναγνωρίζει σωστά τις θετικές περιπτώσεις. Το F1-score συνδυάζει τις δύο προηγούμενες μετρικές και παρέχει μια πιο ισορροπημένη εικόνα της απόδοσης του μοντέλου. Η συνάρτηση απώλειας (loss function) αποτελεί έναν ακόμη σημαντικό δείκτη της απόδοσης του DP-SGD. Σε ένα μοντέλο ταξινόμησης με binary labels, η συνήθως χρησιμοποιούμενη συνάρτηση είναι η Binary Cross-Entropy, η οποία μετρά τη διαφορά μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Όταν εφαρμόζεται το DP-SGD, η εισαγωγή θορύβου στις διαβαθμίσεις μπορεί να αυξήσει την τιμή της συνάρτησης απώλειας, επηρεάζοντας αρνητικά τη συνολική εκπαίδευση του μοντέλου.

Τέλος, η ταχύτητα σύγκλισης αποτελεί κρίσιμο παράγοντα για την αξιολόγηση του DP-SGD. Ο αριθμός των epochs που απαιτούνται για να σταθεροποιηθεί το μοντέλο μπορεί να αυξηθεί σημαντικά λόγω του προστιθέμενου θορύβου. Αυτό σημαίνει ότι το DP-SGD ενδέχεται να απαιτεί περισσότερους υπολογιστικούς πόρους και μεγαλύτερο χρόνο εκπαίδευσης σε σύγκριση με το κλασικό SGD. Επομένως, κατά την αξιολόγηση του αλγορίθμου, λαμβάνεται υπόψη η επίδραση του privacy budget στον ρυθμό εκπαίδευσης, προκειμένου να βρεθεί η κατάλληλη ισορροπία μεταξύ ιδιωτικότητας και αποδοτικότητας του μοντέλου.

Η αξιολόγηση της απόδοσης πραγματοποιείται με τη σύγκριση των αποτελεσμάτων του DP-SGD έναντι ενός κλασικού SGD, στο οποίο δεν εφαρμόζεται DP. Παράλληλα, διερευνάται η σχέση μεταξύ του privacy budget και της γενικότερης επίδοσης του μοντέλου, ώστε να προσδιοριστεί το κατάλληλο επίπεδο ιδιωτικότητας που εξασφαλίζει

επαρκή ακρίβεια. Μέσα από την ανάλυση αυτή, επιδιώκεται η διαμόρφωση μιας σαφούς εικόνας για τον βαθμό επίδρασης του DP-SGD στην εκπαίδευση μοντέλων μηχανικής μάθησης και η εξαγωγή χρήσιμων συμπερασμάτων για τη βελτιστοποίηση της εφαρμογής του.

Στην παρούσα μελέτη, για την εφαρμογή του αλγορίθμου DP-SGD χρησιμοποιήθηκε η ρύθμιση num_microbatches=1, η οποία συνεπάγεται ότι οι διαβαθμίσεις υφίστανται επεξεργασία σε επίπεδο batch και όχι ανά παράδειγμα (per-example clipping). Η συγκεκριμένη παραμετροποίηση διευκολύνει την υλοποίηση του μηχανισμού διαφορικής ιδιωτικότητας, ιδιαίτερα όταν χρησιμοποιούνται βιβλιοθήκες όπως το TensorFlow Privacy. Ωστόσο, αυτή η επιλογή έχει σημαντικές επιπτώσεις στις εγγυήσεις ιδιωτικότητας του μοντέλου. Πιο συγκεκριμένα, το per-example clipping επιτρέπει την ακριβέστερη ρύθμιση της ευαισθησίας της συνάρτησης απώλειας για κάθε μεμονωμένο δείγμα, γεγονός που οδηγεί σε πιο ισχυρές θεωρητικές εγγυήσεις. Αντίθετα, το batch-level clipping περιορίζει τη δυνατότητα ελέγχου της συνεισφοράς κάθε παρατηρήσεως ξεχωριστά και έχει αποδειχθεί ότι παράγει ασθενέστερες εγγυήσεις, επηρεάζοντας τον πραγματικό βαθμό προστασίας της ιδιωτικότητας (Abadi et al., 2016; Subramani et al., 2021). Συνεπώς, η υιοθέτηση της συγκεκριμένης ρύθμισης αποτελεί έναν υπολογισμένο συμβιβασμό μεταξύ πρακτικής ευκολίας και θεωρητικής ακρίβειας των εγγυήσεων.

5.2 Πειραματική Διαδικασία

Η αξιολόγηση της επίδρασης του DP-SGD πραγματοποιείται μέσω μιας συστηματικής πειραματικής διαδικασίας, η οποία περιλαμβάνει τον ορισμό του dataset, την επιλογή του μοντέλου, τη διαμόρφωση των υπερπαραμέτρων και την εκπαίδευση με διαφορετικές τιμές του privacy budget. Στόχος της διαδικασίας αυτής είναι να προσδιοριστεί η επίδραση της εφαρμογής DP στην ακρίβεια και τη σύγκλιση του μοντέλου, καθώς και να διερευνηθεί ο τρόπος με τον οποίο το privacy budget επηρεάζει την απόδοση του μοντέλου.

Για τον σκοπό αυτό, χρησιμοποιούνται το Heart Disease Dataset του UCI Machine Learning Repository, το οποίο αποτελείται από 303 δείγματα, καθώς και το Cardiovascular Disease Dataset, ένα πιο πρόσφατο σύνολο δεδομένων με 70.000 δείγματα. Η χρήση δύο συνόλων δεδομένων διαφορετικής κλίμακας και χρονικής περιόδου επιτρέπει τη σύγκριση της απόδοσης του DP-SGD υπό διαφορετικές συνθήκες. Το Heart Disease Dataset, αν και μικρό, αποτελεί ένα κλασικό και ευρέως χρησιμοποιούμενο σύνολο δεδομένων για την ταξινόμηση καρδιοπαθειών. Αντίθετα, το Cardiovascular Disease Dataset προσφέρει μια πιο σύγχρονη οπτική, με δεδομένα που αντικατοπτρίζουν τις τρέχουσες διαγνωστικές πρακτικές και την εξέλιξη των ιατρικών δεδομένων. Η ανάλυση της απόδοσης του DP-SGD σε αυτά τα δύο datasets επιτρέπει την απάντηση σε σημαντικά ερευνητικά ερωτήματα. Αρχικά, διερευνάται το κατά πόσο η αύξηση του όγκου των δεδομένων βελτιώνει την ανθεκτικότητα του DP-SGD στην απώλεια ακρίβειας. Επιπλέον, εξετάζεται αν τα πιο σύγχρονα δεδομένα έχουν διαφορετική συμπεριφορά κατά την εκπαίδευση του μοντέλου, λόγω της καλύτερης ποιότητας των καταγραφών και των βελτιωμένων ιατρικών μεθόδων. Η διερεύνηση αυτή έχει ιδιαίτερη σημασία, καθώς το DP εφαρμόζεται σε ένα ευρύ φάσμα εφαρμογών και είναι σημαντικό να κατανοηθεί η επίδρασή του σε διαφορετικούς τύπους δεδομένων. Αν το DP-SGD παρουσιάσει μικρότερη απώλεια ακρίβειας στο πιο πρόσφατο και μεγαλύτερο dataset, αυτό μπορεί να υποδεικνύει ότι η αύξηση των δειγμάτων βελτιώνει τη σταθερότητα του μοντέλου υπό συνθήκες DP. Αντίθετα, αν παρατηρηθεί παρόμοια συμπεριφορά μεταξύ των δύο datasets, τότε η επιλογή του privacy budget και των υπερπαραμέτρων είναι πιθανό να έχει μεγαλύτερη σημασία από τον όγκο των δεδομένων.

Συνολικά, η σύγκριση μεταξύ αυτών των δύο datasets αναμένεται να προσφέρει χρήσιμες πληροφορίες για την εφαρμογή του DP-SGD σε πραγματικά δεδομένα υγείας. Η κατανόηση της σχέσης μεταξύ του όγκου των δεδομένων, της χρονολογικής περιόδου συλλογής τους και της απόδοσης του μοντέλου υπό DP μπορεί να οδηγήσει σε καλύτερες πρακτικές για τη διατήρηση της ισορροπίας μεταξύ ιδιωτικότητας και χρηστικότητας των δεδομένων. Επιπλέον, η ανάλυση αυτή μπορεί να βοηθήσει στην επιλογή των κατάλληλων privacy budgets και υπερπαραμέτρων για την εφαρμογή του DP-SGD σε διαφορετικά σύνολα δεδομένων υγείας.

5.2.1 Heart Disease Dataset

Το Heart Disease Dataset είναι ένα από τα πιο διαδεδομένα σύνολα δεδομένων στην έρευνα πρόβλεψης καρδιαγγειακών νοσημάτων και προέρχεται από το Cleveland Clinic Foundation. Η συλλογή των δεδομένων πραγματοποιήθηκε κατά τη διάρκεια αρκετών ετών, με βασική περίοδο καταγραφής μεταξύ 1980 και 1988. Τα δεδομένα συλλέχθηκαν από ασθενείς που υποβλήθηκαν σε διαγνωστικές εξετάσεις με σκοπό την ανάλυση των παραγόντων κινδύνου για καρδιακή νόσο. Παρόλο που αντίστοιχα δεδομένα προέρχονται και από άλλες πηγές, όπως το Hungarian Institute of Cardiology, το V.A. Medical Center (Long Beach) και το University Hospital (Zurich, Switzerland), η εκδοχή του Cleveland είναι η πιο πλήρης και ευρέως χρησιμοποιούμενη στη βιβλιογραφία.
Η συλλογή των δεδομένων έγινε στο πλαίσιο μιας κλινικής μελέτης με σκοπό τη δημιουργία ενός αξιόπιστου διαγνωστικού εργαλείου. Η δειγματοληψία περιελάμβανε ασθενείς με διαφορετικά επίπεδα κινδύνου και διαφορετικά στάδια καρδιοπάθειας, διασφαλίζοντας έτσι την ποικιλομορφία των δεδομένων. Οι καταγραφές περιλαμβάνουν κλινικά χαρακτηριστικά και αποτελέσματα εξετάσεων, επιτρέποντας τη διαμόρφωση μοντέλων μηχανικής μάθησης που βασίζονται σε πραγματικά ιατρικά δεδομένα. Το dataset αποτελείται από 303 δείγματα και 13 χαρακτηριστικά, τα οποία περιλαμβάνουν δημογραφικές πληροφορίες, ιατρικές μετρήσεις και διαγνωστικούς δείκτες. Τα χαρακτηριστικά αυτά περιλαμβάνουν πληροφορίες όπως η ηλικία του ασθενούς, το φύλο, η αρτηριακή πίεση, τα επίπεδα χοληστερόλης και το ιστορικό στηθάγχης. Επίσης, περιλαμβάνει μετρήσεις από ηλεκτροκαρδιογραφήματα, όπως η απόκλιση του διαστήματος ST κατά τη διάρκεια της άσκησης και ο αριθμός των αγγείων που εμφανίζονται σε φθοροαγγειογραφία. Η μεταβλητή-στόχος είναι δυαδική, όπου το 0 υποδηλώνει ότι ο ασθενής δεν πάσχει από καρδιακή νόσο και το 1 ότι έχει διαγνωστεί με την πάθηση.
Η διάρκεια της συλλογής των δεδομένων εκτείνεται σε αρκετά χρόνια, κάτι που εξασφαλίζει τη συγκέντρωση ενός αντιπροσωπευτικού δείγματος ασθενών με ποικίλους παράγοντες κινδύνου. Παρόλο που το dataset έχει χρησιμοποιηθεί σε πλήθος επιστημονικών μελετών, παρουσιάζει ορισμένους περιορισμούς. Ένας από τους βασικούς περιορισμούς είναι ο μικρός αριθμός δειγμάτων, ο οποίος μπορεί να δυσχεράνει τη γενίκευση των αποτελεσμάτων όταν εφαρμόζονται πιο σύνθετα μοντέλα μηχανικής μάθησης. Επιπλέον, υπάρχουν περιπτώσεις με ελλιπή δεδομένα σε συγκεκριμένες μεταβλητές, γεγονός που καθιστά αναγκαία την προεπεξεργασία των δεδομένων προτού χρησιμοποιηθούν στην εκπαίδευση των μοντέλων. Παρά τις πιθανές αδυναμίες, το συγκεκριμένο dataset παραμένει ένα από τα πιο αξιόπιστα για την ανάπτυξη και αξιολόγηση διαγνωστικών αλγορίθμων. Η ποιότητα των δεδομένων και η κλινική τους σημασία επιτρέπουν τη χρήση τους ως βασική πηγή στην ανάλυση μοντέλων πρόβλεψης καρδιαγγειακών παθήσεων.
Τα 13 χαρακτηριστικά που περιλαμβάνει το Heart Disease Dataset είναι τα εξής:

- Age (Ηλικία): Η ηλικία του ασθενούς σε έτη.

- Sex (Φύλο): 1 = Άνδρας, 0 = Γυναίκα.
- Chest Pain Type (Τύπος Θωρακικού Πόνου - cp):
  - Τυπική στηθάγχη
  - Άτυπη στηθάγχη
  - Μη καρδιακός πόνος
  - Ασύμπτωμα
- Resting Blood Pressure (Αρτηριακή Πίεση - trestbps): Η αρτηριακή πίεση σε mm Hg κατά την ανάπαυση.
- Serum Cholesterol (Χοληστερόλη - chol): Συνολικά επίπεδα χοληστερόλης στο αίμα (mg/dL).
- Fasting Blood Sugar (Σάκχαρο Νηστείας - fbs): 1 = Σάκχαρο νηστείας > 120 mg/dL, 0 = Σάκχαρο < 120 mg/dL.
- Resting Electrocardiographic Results (Ηλεκτροκαρδιογράφημα - restecg):
  - Κανονικό
  - Ανωμαλία ST-T
  - Υπερτροφία αριστερής κοιλίας
- Maximum Heart Rate Achieved (Μέγιστος Καρδιακός Ρυθμός - thalach): Η μέγιστη καρδιακή συχνότητα που επιτεύχθηκε κατά τη διάρκεια ενός τεστ κόπωσης.
- Exercise Induced Angina (Άσκηση που Προκαλεί Πόνο - exang): 1 = Ναι, 0 = Όχι.
- ST Depression Induced by Exercise (ST-depression - oldpeak): Μείωση του διαστήματος ST κατά τη διάρκεια της άσκησης σε σχέση με την ανάπαυση.
- Slope of the Peak Exercise ST Segment (Κλίση ST - slope):
  - Αύξουσα
  - Οριζόντια
  - Κατιούσα
- Number of Major Vessels Colored by Fluoroscopy (Αριθμός Αγγείων - ca): Από 0 έως 4 μεγάλα αγγεία που εμφανίζονται σε φθοροαγγειογραφία.
- Thalassemia (Θαλασσαιμία - thal):
  - Κανονική
  - Σταθερή ανωμαλία
  - Αναστρέψιμη ανωμαλία

Αυτά τα χαρακτηριστικά αποτελούν τους βασικούς παράγοντες που χρησιμοποιούνται για την πρόβλεψη της πιθανότητας καρδιακής νόσου. Το dataset περιλαμβάνει 303 δείγματα, με κάθε δείγμα να αντιπροσωπεύει έναν ασθενή, και η μεταβλητή-στόχος (target variable) είναι δυαδική (0: Χωρίς Καρδιακή Νόσο, 1: Με Καρδιακή Νόσο).

Το μοντέλο εκπαιδεύεται για την ταξινόμηση των ασθενών σε δύο κατηγορίες: άτομα που πάσχουν από καρδιακή νόσο και άτομα που δεν παρουσιάζουν την πάθηση.

Η πειραματική διαδικασία περιλαμβάνει λεπτομερή προεπεξεργασία των δεδομένων, η οποία περιλαμβάνει την αντικατάσταση ελλιπών τιμών, την κανονικοποίηση αριθμητικών χαρακτηριστικών και τη μετατροπή κατηγορικών μεταβλητών σε δυαδικές (one-hot encoding). Η καθαρότητα των δεδομένων διασφαλίζεται μέσω της ανίχνευσης και απομάκρυνσης πιθανών ακραίων τιμών που θα μπορούσαν να επηρεάσουν την εκπαίδευση του μοντέλου.

Το μοντέλο που επιλέγεται για την ανάλυση είναι ένα νευρωνικό δίκτυο πολλαπλών επιπέδων, το οποίο διαμορφώνεται με ένα επίπεδο εισόδου, δύο κρυφά επίπεδα και ένα επίπεδο εξόδου. Το δίκτυο αυτό διαθέτει:

- Ένα επίπεδο εισόδου με 13 νευρώνες, όσοι και οι μεταβλητές του dataset.

- Δύο κρυφά επίπεδα με 32 και 16 νευρώνες αντίστοιχα, χρησιμοποιώντας την ενεργοποίηση ReLU.
- Ένα επίπεδο εξόδου με 1 νευρώνα, το οποίο χρησιμοποιεί τη λειτουργία Sigmoid για την πρόβλεψη της πιθανότητας ύπαρξης καρδιακής νόσου.

Η εκπαίδευση του μοντέλου πραγματοποιείται αρχικά με την κλασική μέθοδο SGD (χωρίς DP), ώστε να δημιουργηθεί μια baseline απόδοση για τη σύγκριση. Στη συνέχεια, εφαρμόζεται το DP-SGD με διαφορετικές τιμές privacy budget ε\varepsilon, συγκεκριμένα ε=10,1,0.1, προκειμένου να αναλυθεί η επίδρασή του στις προβλέψεις του μοντέλου. Κατά τη διαδικασία εκπαίδευσης, εξετάζονται διαφορετικές τιμές υπερπαραμέτρων, όπως το batch size, το learning rate και η προσθήκη Gaussian θορύβου. Η αξιολόγηση του μοντέλου πραγματοποιείται μέσω της παρακολούθησης της εξέλιξης της συνάρτησης απώλειας, της ακρίβειας στο test set και της ταχύτητας σύγκλισης. Ιδιαίτερη προσοχή δίνεται στην ανάλυση της σχέσης μεταξύ της μείωσης του privacy budget και της μείωσης της ακρίβειας, προκειμένου να προσδιοριστεί η βέλτιστη ισορροπία μεταξύ προστασίας και απόδοσης του μοντέλου. Η σύγκριση των αποτελεσμάτων μεταξύ των διαφορετικών συνθηκών εκπαίδευσης παρέχει χρήσιμες πληροφορίες για το πώς το privacy budget επηρεάζει τη συνολική απόδοση του μοντέλου. Η ανάλυση αυτή επιτρέπει την εξαγωγή συμπερασμάτων σχετικά με το ποιο επίπεδο ιδιωτικότητας είναι το πιο κατάλληλο για την εξισορρόπηση μεταξύ προστασίας των δεδομένων και διατήρησης της χρηστικότητας του μοντέλου.

5.2.2 Cardiovascular Disease Dataset

Το Cardiovascular Disease Dataset είναι ένα σύνολο δεδομένων που δημιουργήθηκε για τη μελέτη των καρδιαγγειακών νοσημάτων και περιλαμβάνει ιατρικά δεδομένα ασθενών που εξετάστηκαν για την παρουσία καρδιολογικών παθήσεων. Το dataset προέρχεται από μια ευρεία κλινική μελέτη που πραγματοποιήθηκε τα τελευταία χρόνια και περιλαμβάνει στοιχεία από νοσοκομεία και ιατρικά κέντρα. Η συλλογή των δεδομένων ολοκληρώθηκε μέσα στην τελευταία δεκαετία και έχει χρησιμοποιηθεί εκτενώς στην ανάλυση της σχέσης μεταξύ παραγόντων κινδύνου και της εκδήλωσης καρδιακών παθήσεων. Η μελέτη διεξήχθη σε κλινικές και νοσοκομεία με σκοπό την καταγραφή στοιχείων σχετικά με την αρτηριακή πίεση, τα επίπεδα χοληστερόλης, τη γλυκόζη, καθώς και δημογραφικά χαρακτηριστικά των ασθενών. Συμμετείχαν άτομα διαφορετικών ηλικιακών ομάδων, φύλων και επιπέδων υγείας, ώστε να δημιουργηθεί ένα αντιπροσωπευτικό σύνολο δεδομένων. Τα δεδομένα συλλέχθηκαν ανώνυμα και χρησιμοποιήθηκαν με σκοπό τη βελτίωση των διαγνωστικών μοντέλων και των τεχνικών πρόβλεψης για καρδιολογικές παθήσεις.

Η συλλογή των δεδομένων αφορούσε τόσο υγιείς όσο και ασθενείς με ιστορικό καρδιαγγειακών προβλημάτων. Οι συμμετέχοντες υποβλήθηκαν σε εξετάσεις που περιλάμβαναν μέτρηση της αρτηριακής πίεσης, του δείκτη μάζας σώματος, των επιπέδων χοληστερόλης και γλυκόζης, καθώς και την αξιολόγηση του τρόπου ζωής τους, όπως η σωματική δραστηριότητα, το κάπνισμα και η κατανάλωση αλκοόλ. Στη μελέτη περιλήφθηκαν συνολικά 70.000 άτομα, γεγονός που καθιστά το dataset πολύ πιο εκτεταμένο σε σύγκριση με παλαιότερα δεδομένα, όπως το Heart Disease Dataset. Ο στόχος της μελέτης ήταν να δημιουργηθεί ένα σύγχρονο και ευρέως προσβάσιμο σύνολο δεδομένων που θα μπορούσε να χρησιμοποιηθεί από ερευνητές στον τομέα της μηχανικής μάθησης και της ιατρικής για τη βελτίωση των αλγορίθμων διάγνωσης. Η αυξημένη διαθεσιμότητα τέτοιων δεδομένων επιτρέπει τη βαθύτερη ανάλυση των παραγόντων κινδύνου που σχετίζονται με τις καρδιακές παθήσεις και συμβάλλει στη δημιουργία ακριβέστερων μοντέλων πρόβλεψης. Η δομή του dataset επιτρέπει επίσης την εφαρμογή

διαφόρων τεχνικών ανάλυσης, συμπεριλαμβανομένης της χρήσης αλγορίθμων DP, όπως το DP-SGD, για τη διατήρηση της ιδιωτικότητας των δεδομένων.
Το Cardiovascular Disease Dataset περιλαμβάει πληροφορίες όπως:

- Σύνολο δειγμάτων: 70.000 (πολύ μεγαλύτερο από το Heart Disease Dataset των 303 δειγμάτων)
- Αριθμός χαρακτηριστικών: 13 (όμοια με το παλαιότερο dataset)
- Χαρακτηριστικά:
    - id: Αναγνωριστικό ασθενούς.
    - age: Ηλικία σε ημέρες (θα χρειαστεί μετατροπή σε έτη).
    - gender: 1 = Άνδρας, 2 = Γυναίκα.
    - height: Ύψος σε εκατοστά.
    - weight: Βάρος σε κιλά.
    - ap_hi: Συστολική αρτηριακή πίεση.
    - ap_lo: Διαστολική αρτηριακή πίεση.
    - cholesterol: Επίπεδο χοληστερόλης (1 = κανονικό, 2 = πάνω από το φυσιολογικό, 3 = πολύ υψηλό).
    - gluc: Επίπεδο γλυκόζης (ίδιο σύστημα κωδικοποίησης με τη χοληστερόλη).
    - smoke: 1 = Καπνιστής, 0 = Μη καπνιστής.
    - alco: 1 = Κατανάλωση αλκοόλ, 0 = Όχι.
    - active: 1 = Σωματικά δραστήριος, 0 = Μη δραστήριος.
    - cardio: Μεταβλητή-στόχος (1 = έχει καρδιοπάθεια, 0 = δεν έχει).

5.3 Αποτελέσματα και Ανάλυση

5.3 Αποτελέσματα και Ανάλυση

Για την αξιολόγηση της επίδρασης του DP-SGD, πραγματοποιήθηκε πειραματική διαδικασία εκπαίδευσης μοντέλων νευρωνικών δικτύων σε δύο διαφορετικά σύνολα δεδομένων: το Heart Disease Dataset και το Cardiovascular Disease Dataset. Στο πλαίσιο του πειράματος εξετάστηκε η απόδοση του DP-SGD σε σύγκριση με τον παραδοσιακό αλγόριθμο SGD, καθώς και η επίδραση του privacy budget στην ακρίβεια και τη σύγκλιση των μοντέλων. Η εκπαίδευση υλοποιήθηκε με τη χρήση πολυεπίπεδου νευρωνικού δικτύου που περιλάμβανε δύο κρυφά επίπεδα, συνάρτηση ενεργοποίησης ReLU και έξοδο μέσω της συνάρτησης Sigmoid, κατάλληλη για δυαδική ταξινόμηση. Οι τιμές του privacy budget που εξετάστηκαν ήταν ε = 10, 1 και 0.1, προκειμένου να αναλυθεί η σταδιακή αύξηση της προστασίας ιδιωτικότητας και ο αντίκτυπός της στην απόδοση του μοντέλου.
Αξίζει να σημειωθεί ότι οι τιμές epsilon (ε) που αναφέρονται στην παρούσα μελέτη και χρησιμοποιούνται για την αξιολόγηση των επιπέδων ιδιωτικότητας προκύπτουν από θεωρητικά μοντέλα που προϋποθέτουν την εφαρμογή per-example clipping. Δεδομένου ότι στην υλοποίηση που εφαρμόστηκε χρησιμοποιήθηκε num_microbatches=1, η επεξεργασία των διαβαθμίσεων έγινε σε επίπεδο batch και όχι σε επίπεδο μεμονωμένων παραδειγμάτων. Αυτό σημαίνει ότι οι υπολογισμένες τιμές ε, αν και τυπικά ορθές με βάση την υπάρχουσα λογισμική υποδομή, δεν αντανακλούν απόλυτα τις θεωρητικά βέλτιστες εγγυήσεις διαφορικής ιδιωτικότητας. Κατά συνέπεια, οι τιμές ε θα πρέπει να ερμηνεύονται με επιφύλαξη, αναγνωρίζοντας ότι ενδέχεται να υποτιμούν τον πραγματικό κίνδυνο αποκάλυψης ευαίσθητων πληροφοριών.
Η διαδικασία εκπαίδευσης περιέλαβε προκατεργασία των δεδομένων, η οποία περιλάμβανε κανονικοποίηση των αριθμητικών χαρακτηριστικών και μετατροπή των κατηγορικών μεταβλητών μέσω one-hot encoding. Τα δεδομένα διαχωρίστηκαν σε σύνολα εκπαίδευσης

και δοκιμών με αναλογία 80-20. Η εκπαίδευση των μοντέλων πραγματοποιήθηκε με τον βελτιστοποιητή Adam (learning rate = 0.001), και χρησιμοποιήθηκαν δύο διαφορετικά μεγέθη δειγμάτων (batch sizes) 32 και 128 για συγκριτική αξιολόγηση της απόδοσης.
Καθ' όλη τη διάρκεια της εκπαίδευσης παρακολουθήθηκαν οι μεταβολές στη συνάρτηση απώλειας και στην ακρίβεια στο test set, με στόχο την αποτίμηση της επίδρασης του DP-SGD στον ρυθμό μάθησης και στη γενικευτική ικανότητα του μοντέλου. Η ανάλυση των αποτελεσμάτων επικεντρώθηκε στην αξιολόγηση της ακρίβειας, καθώς η προσθήκη θορύβου στις διαβαθμίσεις ενδέχεται να επηρέασε τη διαδικασία εκμάθησης των προτύπων των δεδομένων.
Διερευνήθηκε κατά πόσο η μείωση της ακρίβειας ήταν πιο έντονη σε περιπτώσεις αυστηρού privacy budget, και αν η αυξημένη προστασία ιδιωτικότητας οδήγησε σε μεγαλύτερη διακύμανση των διαβαθμίσεων, περιορίζοντας την ικανότητα του μοντέλου να προσαρμόζεται στα δεδομένα εκπαίδευσης. Η συγκριτική αξιολόγηση των αποτελεσμάτων τόσο στο μικρότερο όσο και στο μεγαλύτερο dataset ανέδειξε τη συσχέτιση μεταξύ του όγκου των δεδομένων και της σταθερότητας του DP-SGD. Στο μικρότερο dataset παρατηρήθηκε μεγαλύτερη απώλεια ακρίβειας λόγω του περιορισμένου αριθμού δειγμάτων, ενώ στο μεγαλύτερο dataset η επίδραση του DP-SGD αποδείχθηκε πιο ήπια, καθώς η αφθονία δεδομένων παρείχε μεγαλύτερη ανθεκτικότητα στις επιδράσεις του θορύβου. Η ανάλυση της σύγκλισης του μοντέλου επικεντρώθηκε στην επίδραση του privacy budget στη σταθερότητα και στον αριθμό των επαναλήψεων που απαιτήθηκαν για την επίτευξη σύγκλισης. Εξετάστηκε κατά πόσο η προσθήκη θορύβου επιβράδυνε τη διαδικασία εκμάθησης και αν απαιτήθηκε μεγαλύτερος αριθμός επαναλήψεων για να επιτευχθεί σταθερότητα. Επιπλέον, αναλύθηκε αν η χρήση υψηλότερων τιμών privacy budget οδήγησε σε ταχύτερη σύγκλιση, γεγονός που συνδέθηκε με χαμηλότερο επίπεδο προστασίας των δεδομένων.
Η διαφορά αυτή αποτυπώθηκε πιο έντονα στο μικρότερο dataset, όπου η επίδραση του προστιθέμενου θορύβου δημιούργησε μεγαλύτερη μεταβλητότητα, ενώ στο μεγαλύτερο dataset η διαδικασία σύγκλισης εμφανίστηκε πιο ομαλή και σταθερή. Η στρατηγική ανάλυσης βασίστηκε στην αξιολόγηση της ακρίβειας των μοντέλων, της συνάρτησης απώλειας και της ταχύτητας σύγκλισης. Ως μέτρο της σύγκλισης χρησιμοποιήθηκε ο αριθμός των epochs, δηλαδή των πλήρων επαναλήψεων πάνω στο σύνολο εκπαίδευσης. Ένα μοντέλο θεωρήθηκε ότι είχε συγκλίνει όταν η τιμή της συνάρτησης απώλειας σταθεροποιήθηκε και δεν εμφάνιζε σημαντικές διακυμάνσεις μεταξύ διαδοχικών επαναλήψεων. Στο πλαίσιο της ανάλυσης εξετάστηκε ο ρυθμός μείωσης της απώλειας κατά την εκπαίδευση και η ταχύτητα επίτευξης ενός σταθερού επιπέδου απόδοσης.
Η ανάλυση των αποτελεσμάτων ανέδειξε την επίδραση του DP-SGD στην απόδοση των μοντέλων. Διαπιστώθηκε ότι η επιλογή του privacy budget αποτέλεσε καθοριστικό παράγοντα, καθώς προσδιόρισε τον βαθμό θορύβου που προστέθηκε στις διαβαθμίσεις, επηρεάζοντας τόσο την ακρίβεια όσο και τον ρυθμό εκπαίδευσης. Παράλληλα, ο όγκος των δεδομένων αποδείχθηκε καθοριστικός για τη σταθερότητα της διαδικασίας εκπαίδευσης, καθώς τα μεγαλύτερα dataset παρείχαν μεγαλύτερη ανθεκτικότητα στις επιδράσεις του DP-SGD. Τα ευρήματα συνέβαλαν στην κατανόηση της βέλτιστης χρήσης του αλγορίθμου σε πρακτικές εφαρμογές, υπογραμμίζοντας τη σημασία της ορθής επιλογής privacy budget με βάση τα χαρακτηριστικά των δεδομένων. Σε πρακτικό επίπεδο, παρατηρήθηκε ότι η εφαρμογή αυστηρών privacy budgets σε μικρά datasets οδήγησε σε αισθητή απώλεια ακρίβειας και καθυστέρηση στη σύγκλιση, ενώ στα μεγαλύτερα datasets η αρνητική επίδραση του DP-SGD ήταν πιο περιορισμένη, καθιστώντας τη μέθοδο πιο κατάλληλη για κλινικές εφαρμογές με επαρκή όγκο δεδομένων.

6. Βελτιστοποίηση των Υπερπαραμέτρων του DP-SGD

6.1 Σημασία της ρύθμισης των υπερπαραμέτρων στο DP-SGD.

Η ρύθμιση των υπερπαραμέτρων στο DP-SGD είναι κρίσιμη για τη βελτιστοποίηση της σχέσης μεταξύ ιδιωτικότητας και απόδοσης στα μοντέλα μηχανικής μάθησης. Δεδομένου ότι το DP-SGD τροποποιεί τον κλασικό αλγόριθμο στοχαστικής καθόδου βαθμίδας με αποκοπή των βαθμίδων και προσθήκη θορύβου, οι υπερπαράμετροι παίζουν καθοριστικό ρόλο στην τελική απόδοση του μοντέλου. Επομένως, η σωστή ρύθμισή τους μπορεί να συμβάλει τόσο στην επίτευξη καλύτερης ακρίβειας όσο και στη διατήρηση αυστηρών εγγυήσεων ιδιωτικότητας.

Μία από τις βασικές προκλήσεις είναι ότι οι επιδράσεις των υπερπαραμέτρων στο DP-SGD δεν είναι απλές και δεν ακολουθούν πάντα τα ίδια μοτίβα που παρατηρούνται στον κλασικό SGD. Για παράδειγμα, το μέγεθος παρτίδας (batch size) στο DP-SGD δεν επηρεάζει μόνο τη στατιστική σταθερότητα των ενημερώσεων αλλά και την κατανάλωση του αποθέματος ιδιωτικότητας λόγω του μηχανισμού ενίσχυσης ιδιωτικότητας μέσω υποδειγματοληψίας. Παρόμοια, η επιλογή του ρυθμού εκμάθησης (learning rate) και του ορίου αποκοπής (clipping threshold) δεν μπορεί να γίνεται ανεξάρτητα, καθώς υπάρχει αλληλεπίδραση μεταξύ αυτών των δύο παραμέτρων που καθορίζει πόσο πληροφοριακά πλούσιες (ή θορυβώδεις) είναι οι ενημερώσεις του μοντέλου.

Η πολυπλοκότητα αυτή καθιστά τη διαδικασία ρύθμισης των υπερπαραμέτρων ιδιαίτερα απαιτητική, καθώς δεν υπάρχουν γενικοί κανόνες που να ισχύουν για όλα τα σενάρια. Η αποδοτικότητα του DP-SGD μπορεί να επηρεάζεται από τον αριθμό εποχών, το μέγεθος παρτίδας, τη δομή του μοντέλου και τη φύση των δεδομένων. Αυτό σημαίνει ότι κάθε πρόβλημα μπορεί να απαιτεί διαφορετική προσέγγιση στη ρύθμιση των υπερπαραμέτρων για την επίτευξη ενός ικανοποιητικού συμβιβασμού μεταξύ ιδιωτικότητας και χρηστικότητας. Η υπερπαραμετροποίηση του DP-SGD είναι ένα θεμελιώδες ζήτημα που επηρεάζει την αποτελεσματικότητα της διαφοροποιημένης ιδιωτικότητας στη μηχανική μάθηση. Η ορθολογική προσέγγισή της δεν είναι μόνο τεχνικά σημαντική αλλά και αναγκαία για την ευρύτερη αποδοχή των ιδιωτικών μοντέλων στη βιομηχανία και την έρευνα.

Η σημασία της ρύθμισης των υπερπαραμέτρων έγκειται στο γεγονός ότι οι διαφορετικές επιλογές τους επηρεάζουν τόσο την ιδιωτικότητα όσο και την ακρίβεια του μοντέλου. Στο DP-SGD, βασικοί υπερπαράμετροι όπως το μέγεθος παρτίδας (batch size), ο ρυθμός εκμάθησης (learning rate), το όριο αποκοπής (clipping threshold) και η κλίμακα θορύβου (noise scale) αλληλεπιδρούν μεταξύ τους και καθορίζουν πόσο χρήσιμο θα είναι το εκπαιδευμένο μοντέλο, διατηρώντας παράλληλα την επιθυμητή προστασία δεδομένων.

Η μελέτη "R+R: Understanding Hyperparameter Effects in DP-SGD" επισημαίνει ότι υπάρχει έλλειψη ομοφωνίας στη βιβλιογραφία σχετικά με το πώς ακριβώς επηρεάζουν οι υπερπαράμετροι την απόδοση του DP-SGD. Αντίθετες και ανεπαρκώς επαληθευμένες δηλώσεις έχουν οδηγήσει σε αβεβαιότητα, γεγονός που δυσκολεύει την αποδοχή της διαφορικής ιδιωτικότητας στις πρακτικές εφαρμογές.

Ένα από τα κύρια συμπεράσματα της μελέτης είναι ότι δεν υπάρχει ένας γενικός κανόνας για τη ρύθμιση των υπερπαραμέτρων, καθώς οι επιδράσεις τους εξαρτώνται από το εκάστοτε πρόβλημα, το σύνολο δεδομένων και την αρχιτεκτονική του μοντέλου.

Η μελέτη αναλύει τις επιδράσεις των βασικών υπερπαραμέτρων και διαπιστώνει ότι:

1. Το μέγεθος παρτίδας (batch size) δεν είναι πάντα η πιο κρίσιμη υπερπαράμετρος, όπως θεωρούταν αρχικά. Παρόλο που μεγαλύτερες παρτίδες μπορούν να βελτιώσουν την ακρίβεια, το όφελος αυτό δεν είναι συνεπές σε όλα τα σενάρια.

Επιπλέον, μεγαλύτερες παρτίδες καταναλώνουν γρηγορότερα τον προϋπολογισμό ιδιωτικότητας (privacy budget), καθώς μειώνεται το όφελος από την ενίσχυση ιδιωτικότητας μέσω δειγματοληψίας.

2. Ο αριθμός εποχών (epochs) έχει θετική επίδραση στην ακρίβεια, αλλά μόνο μέχρι ένα συγκεκριμένο σημείο. Στις περισσότερες περιπτώσεις, η αύξηση των εποχών οδηγεί σε καλύτερη απόδοση, αλλά αν οι εποχές είναι υπερβολικά πολλές, μπορεί να αυξήσει υπερβολικά την κατανάλωση του προϋπολογισμού ιδιωτικότητας, χωρίς επιπλέον βελτίωση της απόδοσης.
3. Το όριο αποκοπής (clipping threshold) και ο ρυθμός εκμάθησης (learning rate) είναι οι πιο σημαντικοί υπερπαράμετροι. Η μελέτη διαπιστώνει ότι η σχέση μεταξύ αυτών των δύο είναι κρίσιμη και ότι πρέπει να ρυθμίζονται μαζί, καθώς η αλλαγή της τιμής του ενός απαιτεί προσαρμογή του άλλου για να επιτευχθεί η μέγιστη ακρίβεια.
4. Η προσθήκη θορύβου (noise scale) είναι καθοριστική για την προστασία της ιδιωτικότητας, αλλά πρέπει να προσαρμόζεται προσεκτικά ώστε να μην υποβαθμίζει υπερβολικά την ποιότητα των βαθμίδων που ενημερώνουν το μοντέλο.

Ένα βασικό εύρημα της μελέτης είναι ότι πολλές από τις προηγούμενες έρευνες για το DP-SGD δεν ακολουθούσαν πειραματικά πρωτόκολλα σχεδιασμένα ειδικά για την ανάλυση των υπερπαραμέτρων. Αυτό σημαίνει ότι κάποια συμπεράσματα που έχουν δημοσιευτεί δεν είναι γενικεύσιμα και ενδέχεται να ισχύουν μόνο για συγκεκριμένα σύνολα δεδομένων ή αρχιτεκτονικές μοντέλων. Η ανάλυση των υπερπαραμέτρων πρέπει να λαμβάνει υπόψη τις αλληλεπιδράσεις μεταξύ τους, καθώς και τον ειδικό ρόλο που παίζει κάθε υπερπαράμετρος στο DP-SGD σε σύγκριση με το SGD. Για παράδειγμα, στο SGD, το μέγεθος παρτίδας επηρεάζει κυρίως τη σταθερότητα της εκπαίδευσης, ενώ στο DP-SGD επηρεάζει τόσο τη σταθερότητα όσο και το θόρυβο που προστίθεται στις διαβαθμίσεις.

Τα ευρήματα της μελέτης δείχνουν ότι η υπερπαραμετροποίηση είναι απαραίτητη στο DP-SGD και δεν μπορεί να παρακαμφθεί. Ωστόσο, η κατανόηση των επιδράσεων των υπερπαραμέτρων μπορεί να επιταχύνει τη διαδικασία αναζήτησης των βέλτιστων ρυθμίσεων και να διευκολύνει τη χρήση του DP-SGD σε πραγματικές εφαρμογές. Η δυνατότητα βελτιστοποίησης των υπερπαραμέτρων με αυτοματοποιημένες τεχνικές μπορεί να συμβάλει στη βελτίωση της ιδιωτικότητας χωρίς μεγάλη απώλεια απόδοσης. Επιπλέον, η γνώση των αλληλεπιδράσεων μεταξύ των υπερπαραμέτρων μπορεί να χρησιμεύσει ως οδηγός για την επιλογή αρχικών τιμών σε μεθόδους αναζήτησης, μειώνοντας έτσι τον υπολογιστικό φόρτο της διαδικασίας.

Η σωστή ρύθμιση των υπερπαραμέτρων είναι απαραίτητη για την αύξηση της απόδοσης του μοντέλου και τη διατήρηση μιας καλής ισορροπίας μεταξύ προκατάληψης (bias) και διακύμανσης (variance). Στην περίπτωση του DP-SGD, η επιλογή κατάλληλων υπερπαραμέτρων δεν επηρεάζει μόνο την απόδοση του μοντέλου αλλά και το επίπεδο διαφορικής ιδιωτικότητας που επιτυγχάνεται. Η σωστή ρύθμιση του ρυθμού εκμάθησης (learning rate), του μεγέθους παρτίδας (batch size), της κλίμακας θορύβου (noise scale) και του ορίου αποκοπής (clipping threshold) μπορεί να μειώσει τις επιπτώσεις του θορύβου στην εκπαίδευση και να βελτιώσει την ακρίβεια του μοντέλου.

Η υπερπαραμετροποίηση είναι μια κρίσιμη διαδικασία στη μηχανική μάθηση, καθώς επηρεάζει άμεσα την απόδοση και τη γενίκευση των μοντέλων. Διάφορες μέθοδοι έχουν αναπτυχθεί για τη βελτιστοποίηση των υπερπαραμέτρων, καθεμία με τα δικά της πλεονεκτήματα και περιορισμούς. Η Grid Search είναι μια εξαντλητική μέθοδος που εξετάζει όλες τις πιθανές τιμές των υπερπαραμέτρων μέσα σε ένα προκαθορισμένο πλέγμα, αν και συχνά είναι υπολογιστικά απαιτητική. Η Random Search, αντίθετα, επιλέγει τυχαία τιμές, καθιστώντας την πιο αποδοτική σε πολλές περιπτώσεις, καθώς επιτρέπει τη γρήγορη

εξερεύνηση του χώρου των υπερπαραμέτρων. Η Bayesian Optimization προσφέρει μια πιο έξυπνη προσέγγιση χρησιμοποιώντας προηγούμενα αποτελέσματα για να καθοδηγήσει τη διαδικασία αναζήτησης, αυξάνοντας έτσι την πιθανότητα εύρεσης των βέλτιστων ρυθμίσεων. Τέλος, το Meta-Learning αξιοποιεί γνώσεις από προηγούμενες διαδικασίες υπερπαραμετροποίησης για να επιταχύνει την αναζήτηση, μειώνοντας τον απαιτούμενο υπολογιστικό χρόνο και βελτιώνοντας την αποδοτικότητα της εκπαίδευσης των μοντέλων.
Για το DP-SGD, αυτά τα ευρήματα έχουν ιδιαίτερη σημασία, διότι η διαδικασία ρύθμισης των υπερπαραμέτρων μπορεί να έχει δραστικό αντίκτυπο τόσο στην ιδιωτικότητα όσο και στην απόδοση του μοντέλου. Η επιλογή υπερβολικά αυστηρών παραμέτρων μπορεί να οδηγήσει σε σημαντική μείωση της ακρίβειας, ενώ υπερβολικά χαλαρές ρυθμίσεις μπορεί να υπονομεύσουν την εγγύηση διαφορικής ιδιωτικότητας.

6.2 Περιγραφή των βασικών υπερπαραμέτρων

Οι βασικές υπερπαράμετροι που επηρεάζουν την απόδοση των αλγορίθμων μηχανικής μάθησης είναι ο ρυθμός εκμάθησης, το μέγεθος παρτίδας, η κλίμακα θορύβου και το όριο αποκοπής.

6.2.1 Ρυθμός εκμάθησης (learning rate)

Ο ρυθμός εκμάθησης (learning rate) είναι μία από τις πιο σημαντικές υπερπαραμέτρους στη μηχανική μάθηση, καθώς καθορίζει το μέγεθος του βήματος που κάνει ο αλγόριθμος κατά την ενημέρωση των παραμέτρων του μοντέλου. Σε κάθε επανάληψη της διαδικασίας εκπαίδευσης, οι παράμετροι του μοντέλου προσαρμόζονται με βάση τη βαθμίδα της συνάρτησης απώλειας (gradient of the loss function), και ο ρυθμός εκμάθησης επηρεάζει το πόσο δραστικά γίνονται αυτές οι αλλαγές. Ένας πολύ υψηλός ρυθμός εκμάθησης μπορεί να οδηγήσει σε αστάθεια κατά την εκπαίδευση, καθώς το μοντέλο ενδέχεται να κάνει πολύ μεγάλες αλλαγές στις παραμέτρους του, με αποτέλεσμα να μην καταφέρει να συγκλίνει σε ένα βέλτιστο σημείο ή ακόμα και να αποκλίνει εντελώς. Σε τέτοιες περιπτώσεις, η συνάρτηση απώλειας μπορεί να παρουσιάζει μεγάλες διακυμάνσεις και να μην μειώνεται σταδιακά, καθιστώντας αδύνατη την εκμάθηση ενός αποτελεσματικού μοντέλου.
Αντίθετα, ένας πολύ χαμηλός ρυθμός εκμάθησης μπορεί να κάνει τη διαδικασία εκπαίδευσης υπερβολικά αργή. Το μοντέλο θα πραγματοποιεί πολύ μικρές αλλαγές στις παραμέτρους του σε κάθε βήμα, γεγονός που μπορεί να οδηγήσει είτε σε παγίδευση σε τοπικά ελάχιστα είτε σε ατελείωτες επαναλήψεις χωρίς ουσιαστική βελτίωση. Επιπλέον, ένας υπερβολικά χαμηλός ρυθμός εκμάθησης μπορεί να προκαλέσει πρόβλημα υπερπροσαρμογής (overfitting), καθώς το μοντέλο δεν θα είναι σε θέση να προσαρμοστεί επαρκώς στη δομή των δεδομένων.
Για να αντιμετωπιστούν τα παραπάνω προβλήματα, συχνά χρησιμοποιούνται δυναμικές τεχνικές προσαρμογής του ρυθμού εκμάθησης, οι οποίες βοηθούν στη σταδιακή βελτίωση της εκπαίδευσης. Κάποιες κοινές στρατηγικές περιλαμβάνουν:

- Step decay, όπου ο ρυθμός εκμάθησης μειώνεται απότομα κατά συγκεκριμένα διαστήματα εκπαίδευσης.
- Exponential decay, όπου η μείωση του ρυθμού εκμάθησης ακολουθεί έναν εκθετικό ρυθμό.
- Adaptive learning rate methods, όπως οι αλγόριθμοι AdaGrad, RMSProp και Adam, που προσαρμόζουν δυναμικά τον ρυθμό εκμάθησης ανάλογα με τις ιδιαιτερότητες των δεδομένων και των βαθμίδων.

Στο DP-SGD, ο ρυθμός εκμάθησης έχει ακόμα μεγαλύτερη σημασία, καθώς επηρεάζει άμεσα τον τρόπο με τον οποίο το μοντέλο ανταποκρίνεται στον προστιθέμενο θόρυβο που

απαιτείται για την προστασία της ιδιωτικότητας. Αν ο ρυθμός εκμάθησης είναι πολύ υψηλός, ο θόρυβος μπορεί να προκαλέσει έντονες διακυμάνσεις και να μειώσει την αποτελεσματικότητα της εκπαίδευσης. Αντίθετα, αν είναι πολύ χαμηλός, η διαδικασία εκμάθησης μπορεί να γίνει εξαιρετικά αργή και η ακρίβεια του μοντέλου να επηρεαστεί αρνητικά. Η σωστή επιλογή του ρυθμού εκμάθησης απαιτεί δοκιμές και προσεκτική ανάλυση, ώστε να εξασφαλιστεί ότι το μοντέλο θα συγκλίνει γρήγορα και αποτελεσματικά, διατηρώντας παράλληλα τα απαιτούμενα επίπεδα ιδιωτικότητας και γενίκευσης.

6.2.2 Μέγεθος παρτίδας (batch size)

Το μέγεθος παρτίδας (batch size) είναι μια κρίσιμη υπερπαράμετρος στη μηχανική μάθηση, καθώς καθορίζει τον αριθμό των δειγμάτων που χρησιμοποιούνται για τον υπολογισμό της βαθμίδας και την ενημέρωση των παραμέτρων του μοντέλου σε κάθε βήμα εκπαίδευσης. Ο τρόπος με τον οποίο επιλέγεται το μέγεθος παρτίδας έχει άμεσο αντίκτυπο στην ταχύτητα εκπαίδευσης, την ακρίβεια του μοντέλου και τη γενίκευσή του σε νέα δεδομένα.

Οι παρτίδες μπορούν να έχουν μικρό, μεσαίο ή μεγάλο μέγεθος, με καθεμία από αυτές τις επιλογές να έχει διαφορετικές επιπτώσεις στη διαδικασία εκπαίδευσης. Μικρά μεγέθη παρτίδας παρέχουν θορυβώδεις αλλά συχνές ενημερώσεις των παραμέτρων, γεγονός που μπορεί να βοηθήσει το μοντέλο να ξεπεράσει τοπικά ελάχιστα και να επιτύχει καλύτερη γενίκευση. Ωστόσο, η εκπαίδευση με πολύ μικρές παρτίδες μπορεί να γίνει πιο ασταθής και να απαιτήσει περισσότερες επαναλήψεις για τη σύγκλιση του μοντέλου. Από την άλλη, μεγάλες παρτίδες οδηγούν σε πιο σταθερούς και ακριβείς υπολογισμούς της βαθμίδας, αλλά μπορεί να μειώσουν τη γενίκευση, καθώς το μοντέλο μπορεί να καταλήξει σε υποβέλτιστα σημεία λόγω της ομαλοποιημένης προσαρμογής των παραμέτρων. Επιπλέον, η χρήση πολύ μεγάλων παρτίδων αυξάνει τις απαιτήσεις σε υπολογιστικούς πόρους και μνήμη, καθιστώντας την εκπαίδευση πιο αργή και απαιτητική.

Στο πλαίσιο του DP-SGD, το μέγεθος παρτίδας επηρεάζει άμεσα την προστασία της ιδιωτικότητας και την αποτελεσματικότητα της εκπαίδευσης. Επειδή το DP-SGD χρησιμοποιεί έναν μηχανισμό θορύβου για την προστασία των δεδομένων, το μέγεθος της παρτίδας σχετίζεται με την ποσότητα θορύβου που απαιτείται για τη διατήρηση ενός σταθερού επιπέδου διαφορικής ιδιωτικότητας. Μεγαλύτερες παρτίδες επιτρέπουν τη μείωση του προστιθέμενου θορύβου ανά δείγμα λόγω του φαινομένου ενίσχυσης ιδιωτικότητας μέσω υποδειγματοληψίας (privacy amplification by subsampling). Ωστόσο, αυτό μπορεί να οδηγήσει σε υψηλότερο ρυθμό κατανάλωσης του αποθέματος ιδιωτικότητας (privacy budget), γεγονός που περιορίζει τον αριθμό των επιτρεπόμενων επαναλήψεων εκπαίδευσης.

Ένα από τα μεγαλύτερα ανοιχτά ερωτήματα στην υπερπαραμετροποίηση του DP-SGD είναι το κατάλληλο μέγεθος παρτίδας για τη βέλτιστη ισορροπία μεταξύ ιδιωτικότητας και απόδοσης. Κάποια πειράματα έχουν δείξει ότι μεγαλύτερες παρτίδες βοηθούν στη βελτίωση της ακρίβειας του μοντέλου, ενώ άλλες μελέτες έχουν υποστηρίξει ότι το μέγεθος της παρτίδας δεν είναι τόσο κρίσιμο όσο ο ρυθμός εκμάθησης και το όριο αποκοπής. Αυτό σημαίνει ότι η επιλογή του μεγέθους παρτίδας πρέπει να γίνεται προσεκτικά, ανάλογα με τις ανάγκες της εκπαίδευσης, το διαθέσιμο υπολογιστικό κόστος και το επίπεδο ιδιωτικότητας που απαιτείται. Το μέγεθος παρτίδας αποτελεί έναν σημαντικό παράγοντα που επηρεάζει τη σταθερότητα της εκπαίδευσης, τη γενίκευση του μοντέλου και τη συμμόρφωση με τις εγγυήσεις διαφορικής ιδιωτικότητας. Η σωστή ρύθμισή του απαιτεί πειραματική ανάλυση και προσεκτική προσαρμογή, καθώς η επίδρασή του μπορεί να διαφέρει ανάλογα με τη φύση των δεδομένων και την αρχιτεκτονική του μοντέλου.

6.2.3 Κλίμακα θορύβου (noise scale)

Η κλίμακα θορύβου (noise scale) είναι μια θεμελιώδης υπερπαράμετρος στο DP-SGD που καθορίζει την ποσότητα του τυχαίου θορύβου που προστίθεται στις διαβαθμίσεις κατά την εκπαίδευση του μοντέλου. Η προσθήκη αυτού του θορύβου είναι απαραίτητη για τη διασφάλιση της διαφορικής ιδιωτικότητας, καθώς αποτρέπει την πιθανότητα εξαγωγής πληροφοριών για μεμονωμένα δείγματα από τα δεδομένα εκπαίδευσης. Η τιμή της κλίμακας θορύβου ελέγχει άμεσα το επίπεδο ιδιωτικότητας που επιτυγχάνεται, αλλά ταυτόχρονα επηρεάζει και την απόδοση του μοντέλου.

Όταν η κλίμακα θορύβου είναι πολύ υψηλή, ο θόρυβος που προστίθεται στις διαβαθμίσεις γίνεται τόσο έντονος που μπορεί να καταστήσει την εκπαίδευση ασταθή και να μειώσει σημαντικά την ακρίβεια του μοντέλου. Σε ακραίες περιπτώσεις, η εκπαίδευση μπορεί να αποτύχει πλήρως, καθώς το μοντέλο δεν θα μπορεί να μάθει ουσιαστικά από τα δεδομένα και θα καταλήγει σε τυχαίες αποφάσεις. Από την άλλη, όταν η κλίμακα θορύβου είναι πολύ χαμηλή, η προστασία της ιδιωτικότητας αποδυναμώνεται, καθώς η συμβολή κάθε δείγματος στα αποτελέσματα εκπαίδευσης γίνεται πιο ανιχνεύσιμη. Αυτό μπορεί να οδηγήσει σε αυξημένο κίνδυνο επιθέσεων που στοχεύουν στην ανάκτηση πληροφοριών από τα δεδομένα εκπαίδευσης.

Η επιλογή της κατάλληλης τιμής της κλίμακας θορύβου εξαρτάται από το επιθυμητό επίπεδο ιδιωτικότητας, το οποίο εκφράζεται μέσω της παραμέτρου διαφορικής ιδιωτικότητας ε (epsilon). Μια χαμηλότερη τιμή του ε σημαίνει αυστηρότερη προστασία των δεδομένων, αλλά συχνά απαιτεί υψηλότερη κλίμακα θορύβου, γεγονός που μπορεί να μειώσει την απόδοση του μοντέλου. Για αυτόν τον λόγο, η εύρεση της βέλτιστης ισορροπίας μεταξύ ιδιωτικότητας και ακρίβειας είναι μια από τις μεγαλύτερες προκλήσεις στη διαφορική ιδιωτικότητα.

Στην πράξη, η κλίμακα θορύβου λειτουργεί σε συνδυασμό με το όριο αποκοπής (clipping threshold), το οποίο περιορίζει το μέγεθος των βαθμίδων πριν από την προσθήκη του θορύβου. Η αλληλεπίδραση αυτών των δύο υπερπαραμέτρων είναι κρίσιμη για την αποτελεσματικότητα του DP-SGD, καθώς ένα σωστά ρυθμισμένο όριο αποκοπής μπορεί να μειώσει την εξάρτηση από υψηλές τιμές θορύβου, διατηρώντας έτσι την ιδιωτικότητα χωρίς να επηρεάζεται υπερβολικά η απόδοση του μοντέλου.

Η επιλογή της κατάλληλης κλίμακας θορύβου εξαρτάται επίσης από το μέγεθος παρτίδας (batch size). Μεγαλύτερες παρτίδες επιτρέπουν την προσθήκη λιγότερου θορύβου ανά δείγμα λόγω του φαινομένου της ενίσχυσης ιδιωτικότητας μέσω υποδειγματοληψίας. Αυτό σημαίνει ότι η αύξηση του μεγέθους της παρτίδας μπορεί να βοηθήσει στη διατήρηση της ακρίβειας του μοντέλου χωρίς να απαιτείται υπερβολικός θόρυβος. Ωστόσο, η χρήση μεγάλων παρτίδων αυξάνει τον ρυθμό κατανάλωσης του αποθέματος ιδιωτικότητας, περιορίζοντας το πλήθος των επαναλήψεων εκπαίδευσης που μπορούν να εκτελεστούν.

Η κλίμακα θορύβου είναι μια από τις πιο κρίσιμες υπερπαραμέτρους στο DP-SGD, καθώς ρυθμίζει την ισορροπία μεταξύ προστασίας της ιδιωτικότητας και ποιότητας του εκπαιδευμένου μοντέλου. Η σωστή ρύθμισή της απαιτεί μια προσεκτική ανάλυση της σχέσης μεταξύ θορύβου, αποκοπής, μεγέθους παρτίδας και στόχων ιδιωτικότητας, ώστε να εξασφαλιστεί ότι το μοντέλο διατηρεί τόσο την προστασία των δεδομένων όσο και υψηλή ακρίβεια.

6.2.4 Όριο αποκοπής (clipping threshold)

Το όριο αποκοπής (clipping threshold) είναι μια κρίσιμη υπερπαράμετρος στο DP-SGD, η οποία ελέγχει το μέγιστο επιτρεπτό μέγεθος των βαθμίδων πριν από την προσθήκη θορύβου.

Ο κύριος στόχος του ορίου αποκοπής είναι η αποφυγή υπερβολικά μεγάλων ενημερώσεων των παραμέτρων του μοντέλου, οι οποίες θα μπορούσαν να αποκαλύψουν πληροφορίες για μεμονωμένα δείγματα. Χωρίς αυτή τη ρύθμιση, ορισμένες διαβαθμίσεις μπορεί να αποκλίνουν σημαντικά, οδηγώντας σε διαρροή πληροφορίας ή ακόμα και σε αποσταθεροποίηση της εκπαίδευσης.
Όταν το όριο αποκοπής είναι πολύ υψηλό, οι διαβαθμίσεις δεν περιορίζονται επαρκώς, γεγονός που αυξάνει τον κίνδυνο αποκαλύψεων και μειώνει την αποτελεσματικότητα της διαφορικής ιδιωτικότητας. Επιπλέον, μεγάλες διαβαθμίσεις μπορεί να κυριαρχούν στη διαδικασία εκπαίδευσης, προκαλώντας απότομες αλλαγές στις παραμέτρους του μοντέλου, κάτι που μπορεί να οδηγήσει σε ασταθή σύγκλιση. Από την άλλη, όταν το όριο αποκοπής είναι πολύ χαμηλό, περιορίζει υπερβολικά τις διαβαθμίσεις, αφαιρώντας πολύτιμες πληροφορίες από την εκπαίδευση. Σε αυτή την περίπτωση, η εκπαίδευση γίνεται αναποτελεσματική, καθώς η μάθηση του μοντέλου επηρεάζεται αρνητικά και η απόδοση μειώνεται σημαντικά.
Η επιλογή του κατάλληλου ορίου αποκοπής είναι ιδιαίτερα σημαντική διότι επηρεάζει άμεσα τη σχέση μεταξύ προστασίας ιδιωτικότητας και ακρίβειας του μοντέλου. Ένα μέτρια ρυθμισμένο όριο αποκοπής επιτρέπει στις περισσότερες διαβαθμίσεις να διατηρούν αρκετή πληροφορία για τη μάθηση, ενώ περιορίζει τις εξαιρετικά μεγάλες διαβαθμίσεις που θα μπορούσαν να οδηγήσουν σε αποκαλύψεις. Επιπλέον, το όριο αποκοπής επηρεάζει άμεσα την ποσότητα θορύβου που απαιτείται για τη διατήρηση ενός σταθερού επιπέδου διαφορικής ιδιωτικότητας. Ένα υψηλό όριο αποκοπής απαιτεί μεγαλύτερη ποσότητα θορύβου, προκειμένου να διατηρηθεί η προστασία των δεδομένων, γεγονός που μπορεί να υποβαθμίσει την ακρίβεια του μοντέλου.
Η ρύθμιση του ορίου αποκοπής σχετίζεται επίσης με άλλες υπερπαραμέτρους, όπως η κλίμακα θορύβου και το μέγεθος παρτίδας. Αν το όριο αποκοπής είναι υψηλό, μπορεί να απαιτείται υψηλότερη κλίμακα θορύβου για την αντιστάθμιση της αυξημένης ευαισθησίας των βαθμίδων. Αντίθετα, ένα πολύ χαμηλό όριο αποκοπής μπορεί να μειώσει την ανάγκη για ισχυρό θόρυβο, αλλά μπορεί να περιορίσει τη μάθηση του μοντέλου. Ομοίως, η επιλογή του ορίου αποκοπής επηρεάζεται από το μέγεθος της παρτίδας, καθώς μεγαλύτερες παρτίδες έχουν τη δυνατότητα να μειώσουν την επίδραση μεμονωμένων δειγμάτων στις διαβαθμίσεις, γεγονός που μπορεί να μειώσει την ανάγκη για αυστηρό περιορισμό.
Στην πράξη, η ρύθμιση του ορίου αποκοπής γίνεται συχνά εμπειρικά, μέσω δοκιμών και ανάλυσης της σύγκλισης του μοντέλου. Κάποιες μέθοδοι προσπαθούν να το ρυθμίσουν δυναμικά κατά τη διάρκεια της εκπαίδευσης, ώστε να προσαρμόζεται ανάλογα με τη μεταβολή των βαθμίδων. Αυτές οι προσεγγίσεις μπορούν να βοηθήσουν στη διατήρηση της ισορροπίας μεταξύ προστασίας της ιδιωτικότητας και απόδοσης του μοντέλου, επιτρέποντας μια πιο προσαρμοστική στρατηγική στη ρύθμιση των υπερπαραμέτρων του DP-SGD. Το όριο αποκοπής είναι μια κρίσιμη υπερπαράμετρος που επηρεάζει τόσο την ποιότητα της εκπαίδευσης όσο και το επίπεδο διαφορικής ιδιωτικότητας. Η σωστή επιλογή του μπορεί να βελτιώσει σημαντικά τη σταθερότητα της εκπαίδευσης, να διατηρήσει την απόδοση του μοντέλου και ταυτόχρονα να εξασφαλίσει επαρκή προστασία των δεδομένων από πιθανές επιθέσεις αποκατάστασης πληροφορίας.

6.3 Μεθοδολογία βελτιστοποίησης υπερπαραμέτρων.

Η επιλογή των υπερπαραμέτρων που χρησιμοποιήθηκαν στον αλγόριθμο DP-SGD πραγματοποιήθηκε με στόχο την εξισορρόπηση μεταξύ της απόδοσης του μοντέλου και της διατήρησης ισχυρών εγγυήσεων διαφορικής ιδιωτικότητας. Η διαδικασία βασίστηκε σε εξαντλητική αναζήτηση (Grid Search), όπως προτείνεται στη μελέτη των Abadi et al.

(2016), και υποστηρίζεται από γενικότερες αρχές εκπαίδευσης νευρωνικών δικτύων (Goodfellow et al., 2016). Αρχικά, καθορίστηκε ένα σύνολο πιθανών τιμών για κάθε βασική υπερπαράμετρο, όπως ο ρυθμός μάθησης (learning rate), το μέγεθος δέσμης (batch size), ο πολλαπλασιαστής θορύβου (noise multiplier) και ο συντελεστής αποκοπής του gradient (clipping norm). Για κάθε συνδυασμό, εκπαιδεύτηκε ένα μοντέλο και αξιολογήθηκε ως προς την ακρίβεια (accuracy) και το συνολικό privacy loss (ε), σύμφωνα με την τεχνική Moments Accountant (Abadi et al., 2016). Οι τελικές τιμές επιλέχθηκαν με βάση τη βέλτιστη απόδοση χωρίς υπέρβαση προκαθορισμένων ορίων στο privacy loss, ακολουθώντας την πρακτική που περιγράφεται και στο θεωρητικό πλαίσιο των Dwork και Roth (2014).

Για το Heart Disease dataset, οι τιμές που δοκιμάστηκαν περιλάμβαναν learning rate από 0.001 έως 0.05, noise multiplier από 0.5 έως 1.5, clipping norm από 0.5 έως 2.0, και batch size μεταξύ 32 και 128. Η καλύτερη απόδοση επιτεύχθηκε με learning rate ίσο με 0.01, noise multiplier ίσο με 1.0, clipping norm ίσο με 1.0 και batch size ίσο με 64. Ο συνδυασμός αυτός επέτρεψε ακρίβεια περίπου 92% με συνολικό privacy loss μικρότερο από 3.5. Οι επιλογές αυτές συνάδουν με τις παρατηρήσεις των Abadi et al. (2016), που υποδεικνύουν τη χρήση μέτριου θορύβου και προσεκτικά ρυθμισμένου clipping norm για σταθερή εκπαίδευση. Στην περίπτωση του Cardiovascular dataset, παρατηρήθηκε μεγαλύτερη ευαισθησία των αποτελεσμάτων στις υπερπαραμέτρους. Αυστηρές τιμές clipping, όπως 0.5, και αυξημένες τιμές noise multiplier άνω του 1.5 οδήγησαν σε σημαντική πτώση της απόδοσης και προβλήματα σύγκλισης. Τελικά, επιλέχθηκε clipping norm ίσο με 1.0, noise multiplier ίσο με 1.5, learning rate ίσο με 0.01 και batch size ίσο με 64. Η ακρίβεια του μοντέλου έφτασε περίπου το 78%, με συνολικό privacy loss κάτω από 5.0, τιμές που κρίθηκαν αποδεκτές για την ιδιαιτερότητα του συγκεκριμένου συνόλου δεδομένων.

Η αναζήτηση των βέλτιστων τιμών των υπερπαραμέτρων μπορεί να γίνει με διάφορες μεθόδους. Η Grid Search είναι μια εξαντλητική μέθοδος που ελέγχει όλες τις δυνατές τιμές μέσα σε ένα προκαθορισμένο πλέγμα, αλλά είναι υπολογιστικά δαπανηρή, ειδικά για μεγάλα μοντέλα. Η Random Search προσφέρει μια πιο αποδοτική εναλλακτική, καθώς επιλέγει τυχαίες τιμές από ένα καθορισμένο εύρος και έχει αποδειχθεί ότι μπορεί να βρει καλές ρυθμίσεις με λιγότερη υπολογιστική προσπάθεια. Η Bayesian Optimization χρησιμοποιεί ένα πιθανοτικό μοντέλο για να κατευθύνει την αναζήτηση προς πιο υποσχόμενες ρυθμίσεις, βελτιστοποιώντας τις υπερπαραμέτρους με λιγότερες δοκιμές. Οι γενετικοί αλγόριθμοι και η Hyperband είναι επίσης προηγμένες τεχνικές που προσαρμόζουν δυναμικά την αναζήτηση, ενώ το meta-learning αξιοποιεί προηγούμενες εμπειρίες για να επιταχύνει τη διαδικασία βελτιστοποίησης. Σε αντίθεση με το κλασικό SGD, στο DP-SGD η υπερπαραμετροποίηση είναι πιο περίπλοκη, καθώς οι υπερπαράμετροι αλληλεπιδρούν μεταξύ τους και με τον μηχανισμό διαφορικής ιδιωτικότητας. Για παράδειγμα, το μέγεθος παρτίδας επηρεάζει όχι μόνο τη στατιστική σταθερότητα των ενημερώσεων αλλά και την ενίσχυση της ιδιωτικότητας μέσω της υποδειγματοληψίας. Το όριο αποκοπής και ο ρυθμός εκμάθησης παρουσιάζουν ισχυρή αλληλεπίδραση, καθώς το ένα επηρεάζει την αποτελεσματικότητα του άλλου. Σύμφωνα με μελέτες, η σχέση αυτή είναι συχνά αντίστροφη: υψηλότερα όρια αποκοπής απαιτούν χαμηλότερο ρυθμό εκμάθησης και το αντίστροφο.

Μια σημαντική προσέγγιση για τη βελτιστοποίηση των υπερπαραμέτρων στο DP-SGD είναι η πειραματική μελέτη πολλαπλών συνδυασμών υπερπαραμέτρων για την κατανόηση των επιδράσεών τους. Σε μια πρόσφατη μελέτη, πραγματοποιήθηκαν 3822 πειράματα με διαφορετικές τιμές υπερπαραμέτρων σε έξι σύνολα δεδομένων, έξι αρχιτεκτονικές μοντέλων και τρία διαφορετικά επίπεδα ιδιωτικότητας, προκειμένου να εξεταστεί πώς η

κάθε υπερπαράμετρος επηρεάζει την απόδοση του DP-SGD. Τα αποτελέσματα έδειξαν ότι οι επιδράσεις των υπερπαραμέτρων δεν είναι πάντα συνεπείς και ότι η απόδοσή τους μπορεί να διαφέρει ανάλογα με το πρόβλημα και το σύνολο δεδομένων.
Για την κατανόηση της σημασίας των υπερπαραμέτρων, χρησιμοποιούνται στατιστικές τεχνικές όπως η ανάλυση διακύμανσης (ANOVA), η οποία αποκαλύπτει πόση από τη συνολική μεταβλητότητα των αποτελεσμάτων εξηγείται από κάθε υπερπαράμετρο. Σε μία μεγάλη μελέτη, η σημαντικότερη αλληλεπίδραση που παρατηρήθηκε ήταν μεταξύ του ρυθμού εκμάθησης και του ορίου αποκοπής, καθώς μαζί εξηγούν περισσότερο από το 50% της συνολικής διακύμανσης στην απόδοση του DP-SGD. Αντίθετα, το μέγεθος παρτίδας και ο αριθμός των εποχών είχαν μικρότερη επιρροή.
Η βελτιστοποίηση των υπερπαραμέτρων στο DP-SGD απαιτεί μια προσεκτική και πειραματική προσέγγιση, καθώς η επιλογή λανθασμένων τιμών μπορεί να οδηγήσει είτε σε σημαντική απώλεια ιδιωτικότητας είτε σε μειωμένη ακρίβεια του μοντέλου. Οι πιο σύγχρονες προσεγγίσεις επικεντρώνονται στην κατανόηση των σχέσεων μεταξύ των υπερπαραμέτρων, αντί στην απλή ρύθμιση της κάθε μίας ξεχωριστά. Επίσης, ο συνδυασμός αυτοματοποιημένων μεθόδων αναζήτησης με στοχευμένες πειραματικές δοκιμές μπορεί να επιταχύνει τη διαδικασία υπερπαραμετροποίησης και να εξασφαλίσει καλύτερα αποτελέσματα σε πραγματικές εφαρμογές.

6.3.1 Grid Search

Η Grid Search είναι μια κλασική μέθοδος υπερπαραμετροποίησης που χρησιμοποιείται στη μηχανική μάθηση και ειδικότερα στο DP-SGD για τη βελτίωση της ισορροπίας μεταξύ ακρίβειας και ιδιωτικότητας. Η βασική της ιδέα είναι η αναζήτηση της βέλτιστης τιμής των υπερπαραμέτρων δοκιμάζοντας όλες τις πιθανές συνδυαστικές τιμές μέσα σε ένα προκαθορισμένο πλέγμα (grid). Για κάθε συνδυασμό υπερπαραμέτρων, το μοντέλο εκπαιδεύεται και αξιολογείται, και στο τέλος επιλέγεται η καλύτερη ρύθμιση βάσει των αποτελεσμάτων. Η Grid Search είναι ιδιαίτερα χρήσιμη όταν το εύρος των υπερπαραμέτρων είναι μικρό, καθώς μπορεί να εγγυηθεί ότι οι πιο υποσχόμενοι συνδυασμοί θα ελεγχθούν. Ωστόσο, καθώς ο αριθμός των υπερπαραμέτρων αυξάνεται, η μέθοδος γίνεται εξαιρετικά υπολογιστικά ακριβή, καθώς η συνολική πολυπλοκότητα της αναζήτησης αυξάνεται εκθετικά. Για παράδειγμα, αν υπάρχουν τρεις υπερπαράμετροι με δέκα δυνατές τιμές η κάθε μία, η Grid Search θα πρέπει να εκπαιδεύσει και να αξιολογήσει 1.000 μοντέλα ($10 \times 10 \times 10$), γεγονός που απαιτεί σημαντικούς υπολογιστικούς πόρους. Στο πλαίσιο του DP-SGD, η Grid Search παρουσιάζει ορισμένες προκλήσεις, καθώς οι υπερπαράμετροι αλληλεπιδρούν μεταξύ τους με μη προφανείς τρόπους. Για παράδειγμα, η επιλογή του ρυθμού εκμάθησης (learning rate) πρέπει να λαμβάνει υπόψη την τιμή του ορίου αποκοπής (clipping threshold), ενώ το μέγεθος παρτίδας (batch size) σχετίζεται με την ποσότητα του προστιθέμενου θορύβου λόγω του φαινομένου της ενίσχυσης ιδιωτικότητας μέσω υποδειγματοληψίας. Επομένως, η Grid Search μπορεί να μην είναι η πιο αποδοτική μέθοδος, καθώς δεν λαμβάνει υπόψη τη σχέση μεταξύ των υπερπαραμέτρων και μπορεί να καταλήξει να εξερευνά πολλές ρυθμίσεις που δεν είναι χρήσιμες. Μια εναλλακτική προσέγγιση που χρησιμοποιείται συχνά είναι η Grid Search με προσαρμοστική ανάλυση (adaptive grid search), όπου το πλέγμα αρχικά εξετάζει ένα πιο αραιό σύνολο τιμών και στη συνέχεια εστιάζει σε πιο λεπτομερείς αναζητήσεις στις περιοχές που δείχνουν υποσχόμενα αποτελέσματα. Αυτή η μέθοδος μπορεί να μειώσει την υπολογιστική πολυπλοκότητα διατηρώντας παράλληλα την αποτελεσματικότητα της αναζήτησης.

Σε σύγκριση με άλλες τεχνικές, όπως η Random Search, η Grid Search έχει το πλεονέκτημα ότι εξασφαλίζει τη διερεύνηση όλων των δυνατών συνδυασμών, αλλά έχει το μειονέκτημα της υψηλής υπολογιστικής απαίτησης. Πρόσφατες μελέτες έχουν δείξει ότι τεχνικές όπως η Bayesian Optimization ή η Hyperband μπορούν να είναι πιο αποτελεσματικές, ειδικά σε πολύπλοκες υπερπαραμετροποιήσεις όπως στο DP-SGD. Ωστόσο, η Grid Search εξακολουθεί να είναι χρήσιμη σε περιπτώσεις όπου το εύρος τιμών είναι καλά καθορισμένο και η υπολογιστική ισχύς δεν αποτελεί περιορισμό.

6.3.2 Random Search

Η Random Search είναι μια μέθοδος υπερπαραμετροποίησης που χρησιμοποιείται ευρέως στη μηχανική μάθηση, συμπεριλαμβανομένου του DP-SGD, για την εύρεση βέλτιστων ρυθμίσεων των υπερπαραμέτρων. Σε αντίθεση με τη Grid Search, η οποία εξετάζει όλες τις δυνατές τιμές μέσα σε ένα προκαθορισμένο πλέγμα, η Random Search επιλέγει τυχαία σημεία από τον χώρο των υπερπαραμέτρων. Αυτή η τυχαιοποιημένη προσέγγιση επιτρέπει την πιο αποδοτική εξερεύνηση μεγάλων και πολύπλοκων υπερπαραμετρικών χώρων, μειώνοντας τον αριθμό των απαιτούμενων δοκιμών.

Ένα από τα μεγαλύτερα πλεονεκτήματα της Random Search είναι ότι συχνά βρίσκει καλές ρυθμίσεις υπερπαραμέτρων πιο γρήγορα από την Grid Search, καθώς δεν δεσμεύεται από μια αυστηρά καθορισμένη δομή δοκιμών. Σύμφωνα με μελέτες, όταν μόνο ένας μικρός αριθμός υπερπαραμέτρων έχει μεγάλη επίδραση στην απόδοση του μοντέλου, η Random Search μπορεί να εντοπίσει βέλτιστες τιμές με λιγότερες επαναλήψεις. Για παράδειγμα, εάν μόνο δύο από τις τέσσερις υπερπαραμέτρους έχουν ισχυρή επίδραση στο αποτέλεσμα, η Grid Search θα σπαταλήσει πολλούς υπολογιστικούς πόρους εξετάζοντας περιττούς συνδυασμούς, ενώ η Random Search μπορεί να εντοπίσει τις σημαντικές παραμέτρους πιο γρήγορα.

Στο πλαίσιο του DP-SGD, όπου οι υπερπαράμετροι όπως το όριο αποκοπής (clipping threshold), η κλίμακα θορύβου (noise scale) και το μέγεθος παρτίδας (batch size) αλληλεπιδρούν με πολύπλοκους τρόπους, η Random Search είναι συχνά πιο κατάλληλη. Οι ερευνητές έχουν παρατηρήσει ότι η εύρεση της κατάλληλης αλληλεπίδρασης μεταξύ του ρυθμού εκμάθησης και του ορίου αποκοπής είναι ζωτικής σημασίας για την απόδοση του DP-SGD. Δεδομένου ότι η σχέση αυτή δεν είναι πάντα γραμμική, η τυχαία αναζήτηση μπορεί να δοκιμάσει ένα ευρύ φάσμα τιμών και να αποκαλύψει περιοχές που διαφορετικά θα αγνοούνταν από τη συστηματική αναζήτηση της Grid Search.

Ένα άλλο πλεονέκτημα της Random Search είναι ότι μπορεί εύκολα να παραλληλοποιηθεί, επιτρέποντας τη διεξαγωγή πολλαπλών πειραμάτων ταυτόχρονα. Αυτό την καθιστά ιδιαίτερα χρήσιμη όταν υπάρχει πρόσβαση σε υπολογιστικά clusters ή GPU, μειώνοντας σημαντικά τον χρόνο που απαιτείται για τη βελτιστοποίηση των υπερπαραμέτρων.

Παρόλο που η Random Search προσφέρει μεγαλύτερη αποδοτικότητα, εξακολουθεί να έχει κάποιους περιορισμούς. Για παράδειγμα, αν ο χώρος των υπερπαραμέτρων είναι πολύ μεγάλος, υπάρχει πιθανότητα η τυχαία επιλογή να μην περιλαμβάνει κρίσιμες τιμές. Επιπλέον, δεν αξιοποιεί προηγούμενες δοκιμές για να καθοδηγήσει τη διαδικασία αναζήτησης, όπως συμβαίνει με την Bayesian Optimization, η οποία χρησιμοποιεί πιθανοτικά μοντέλα για την προσαρμογή της αναζήτησης.

6.3.3 Bayesian Optimization

Η Bayesian Optimization είναι μια προηγμένη μέθοδος υπερπαραμετροποίησης που χρησιμοποιείται στη μηχανική μάθηση, συμπεριλαμβανομένου του DP-SGD, για την εύρεση των βέλτιστων τιμών υπερπαραμέτρων με λιγότερες δοκιμές από τις κλασικές μεθόδους όπως η Grid Search και η Random Search. Αντί να εξετάζει όλες τις δυνατές τιμές ή να επιλέγει τυχαία σημεία στον υπερπαραμετρικό χώρο, η Bayesian Optimization χρησιμοποιεί πιθανοτικά μοντέλα για να καθοδηγήσει την αναζήτηση προς τις περιοχές που είναι πιο πιθανό να περιέχουν τις βέλτιστες τιμές.

Η βασική ιδέα πίσω από τη Bayesian Optimization είναι ότι δημιουργεί ένα μοντέλο πιθανοτήτων το οποίο προβλέπει πώς οι αλλαγές στις υπερπαραμέτρους επηρεάζουν την απόδοση του μοντέλου. Με βάση τις αρχικές δοκιμές, αυτό το μοντέλο χρησιμοποιείται για να υπολογίσει μια συνάρτηση απόκτησης (acquisition function), η οποία καθορίζει ποια νέα σημεία υπερπαραμέτρων πρέπει να δοκιμαστούν στη συνέχεια. Ο στόχος είναι να επιτευχθεί η βέλτιστη απόδοση με όσο το δυνατόν λιγότερες πειραματικές δοκιμές, εξοικονομώντας έτσι υπολογιστικούς πόρους. Στο πλαίσιο του DP-SGD, όπου οι υπερπαράμετροι όπως το όριο αποκοπής (clipping threshold), η κλίμακα θορύβου (noise scale), ο ρυθμός εκμάθησης (learning rate) και το μέγεθος παρτίδας (batch size) έχουν περίπλοκες αλληλεπιδράσεις, η Bayesian Optimization είναι ιδιαίτερα χρήσιμη. Οι παραδοσιακές μέθοδοι αναζήτησης συχνά δοκιμάζουν υπερπαραμέτρους ανεξάρτητα, κάτι που μπορεί να οδηγήσει σε χαμένες ευκαιρίες εύρεσης των βέλτιστων συνδυασμών. Αντίθετα, η Bayesian Optimization εκμεταλλεύεται τη γνώση από προηγούμενες δοκιμές και κατευθύνει τη διαδικασία αναζήτησης προς τις πιο υποσχόμενες περιοχές του υπερπαραμετρικού χώρου.

Ένα βασικό πλεονέκτημα αυτής της προσέγγισης είναι ότι μπορεί να χειριστεί υπερπαραμετρικούς χώρους μεγάλης διάστασης πιο αποδοτικά από την Grid Search. Επειδή δεν εξετάζει τυχαία συνδυασμούς αλλά εστιάζει σε περιοχές υψηλής πιθανότητας, μπορεί να βρει καλές ρυθμίσεις με πολύ λιγότερες δοκιμές. Αυτό είναι ιδιαίτερα σημαντικό στο DP-SGD, όπου η εκπαίδευση μοντέλων είναι χρονικά απαιτητική, καθώς περιλαμβάνει τόσο την εκτέλεση του αλγορίθμου όσο και τη διατήρηση των εγγυήσεων διαφορικής ιδιωτικότητας. Ωστόσο, η Bayesian Optimization έχει και κάποιες προκλήσεις. Αρχικά, απαιτεί έναν αρχικό αριθμό τυχαίων δοκιμών για να «μάθει» το πιθανοτικό μοντέλο πριν αρχίσει να κατευθύνει τη διαδικασία αναζήτησης. Επίσης, καθώς ο αριθμός των υπερπαραμέτρων αυξάνεται, η υπολογιστική πολυπλοκότητα της διαχείρισης του πιθανοτικού μοντέλου μπορεί να γίνει υψηλή. Παρά ταύτα, με βελτιστοποιημένες υλοποιήσεις και κατάλληλες τεχνικές προσέγγισης, η Bayesian Optimization μειώνει σημαντικά την ανάγκη για εξαντλητικές αναζητήσεις και αποτελεί μια από τις πιο αποτελεσματικές μεθόδους για την υπερπαραμετροποίηση στο DP-SGD.

7. Μελέτη περίπτωσης: Εφαρμογή σε δεδομένα υγειονομικής περίθαλψης

7.1 Υλοποίηση του DP-SGD για το Heart Disease Dataset

Ο κώδικας αυτός στοχεύει στην εκπαίδευση ενός νευρωνικού δικτύου για την ανίχνευση καρδιακής νόσου, χρησιμοποιώντας DP-SGD ώστε να διατηρηθεί η ιδιωτικότητα των δεδομένων. Ουσιαστικά, εκπαιδεύουμε ένα μοντέλο ταξινόμησης με διαφορική ιδιωτικότητα και αξιολογούμε την απόδοσή του.

```
import zipfile
import os
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import tensorflow as tf
import tensorflow_privacy
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy_lib
import matplotlib.pyplot as plt
import pickle

# 1. Αποσυμπίεση του ZIP αρχείου
zip_path = "heart+disease.zip"
extract_path = "heart_disease_data"

with zipfile.ZipFile(zip_path, 'r') as zip_ref:
    zip_ref.extractall(extract_path)

# 2. Ορισμός των αρχείων που θα φορτώσουμε
data_files = [
    "processed.cleveland.data",
    "processed.hungarian.data",
    "processed.switzerland.data",
    "processed.va.data"
]

# 3. Ορισμός των στηλών
column_names = [
    "age", "sex", "cp", "trestbps", "chol", "fbs", "restecg",
    "thalach", "exang", "oldpeak", "slope", "ca", "thal", "target"
]

# 4. Φόρτωση και συγχώνευση των datasets
df_list = []
for file in data_files:
    file_path = os.path.join(extract_path, file)
    temp_df = pd.read_csv(file_path, names=column_names)
    df_list.append(temp_df)

# Συγχώνευση όλων των DataFrames
df = pd.concat(df_list, ignore_index=True)

# 5. Καθαρισμός των δεδομένων
df.replace("?", np.nan, inplace=True)  # Αντικατάσταση των '?' με NaN
df = df.apply(pd.to_numeric, errors='coerce')  # Μετατροπή όλων των τιμών σε αριθμητικές
df.fillna(df.median(), inplace=True)  # Αντικατάσταση των NaN με τη διάμεσο κάθε στήλης
```

```
# 6. Διαχωρισμός χαρακτηριστικών (X) και ετικετών (y)
X = df.drop(columns=['target'])
y = df['target']

# 7. Μετατροπή του target σε δυαδική ταξινόμηση (0 = χωρίς νόσο, 1 = με νόσο)
y = (y > 0).astype(int)

# 8. Κανονικοποίηση χαρακτηριστικών
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 9. Διαχωρισμός σε σύνολα εκπαίδευσης και δοκιμών
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42, stratify=y
)

# 10. Ορισμός υπερπαραμέτρων για DP-SGD
learning_rate = 0.01
noise_multiplier = 1.1
l2_norm_clip = 1.0
epochs = 50

from tensorflow_privacy.privacy.optimizers.dp_optimizer_keras import DPKerasSGDOptimizer
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy

batch_sizes = [32, 64]
results = []
for batch_size in batch_sizes:
    print(f"\n--- Εκπαίδευση με batch size: {batch_size} ---")

    # Δημιουργία νέου μοντέλου
    model = tf.keras.Sequential([
        tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train.shape[1],)),
        tf.keras.layers.Dense(8, activation='relu'),
        tf.keras.layers.Dense(1, activation='sigmoid')  # Δυαδική ταξινόμηση
    ])

    # Ορισμός DP-SGD Optimizer
    dp_optimizer = DPKerasSGDOptimizer(
        l2_norm_clip=l2_norm_clip,
        noise_multiplier=noise_multiplier,
        num_microbatches=1,
        learning_rate=learning_rate
    )

    model.compile(optimizer=dp_optimizer, loss='binary_crossentropy', metrics=['accuracy'])
```

```
    # Εκπαίδευση του μοντέλου
    history = model.fit(
        X_train, y_train,
        epochs=epochs,
        batch_size=batch_size,
        validation_data=(X_test, y_test),
        verbose=1
    )

    # Αξιολόγηση
    test_loss, test_acc = model.evaluate(X_test, y_test, verbose=0)
    print(f'Test Accuracy: {test_acc:.4f}')

    # Υπολογισμός privacy budget (ε)
    eps, _ = compute_dp_sgd_privacy_lib.compute_dp_sgd_privacy(
        n=len(X_train),
        batch_size=batch_size,
        noise_multiplier=noise_multiplier,
        epochs=epochs,
        delta=1e-5
    )
    print(f'Estimated ε for batch size {batch_size}: {eps:.2f}')

    # Σχεδίαση Διαγραμμάτων για κάθε περίπτωση
    plt.figure(figsize=(12, 5))
    plt.subplot(1, 2, 1)
    plt.plot(history.history['loss'], label='Train Loss')
    plt.plot(history.history['val_loss'], label='Validation Loss')
    plt.xlabel('Epochs')
    plt.ylabel('Loss')
    plt.title(f'Loss Over Epochs (Batch {batch_size})')
    plt.legend()

    plt.subplot(1, 2, 2)
    plt.plot(history.history['accuracy'], label='Train Accuracy')
    plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
    plt.xlabel('Epochs')
    plt.ylabel('Accuracy')
    plt.title(f'Accuracy Over Epochs (Batch {batch_size})')
    plt.legend()

    plt.tight_layout()
    plt.show()

results_df = pd.DataFrame(results)
results_df.to_csv("results_df_Hyperparameter_Tuning_heart_disease.csv", index=False)
print("Αποθηκεύτηκαν τα αποτελέσματα στο αρχείο CSV.")
```

Ο κώδικας ξεκινά με την αποσυμπίεση αρχείου ZIP που περιέχει τέσσερα αρχεία δεδομένων από διαφορετικές πηγές (Cleveland, Hungarian, Switzerland και VA). Τα δεδομένα αυτά αφορούν ιατρικές μετρήσεις σχετικές με την καρδιαγγειακή υγεία των ασθενών. Μετά τη συγχώνευση των αρχείων σε ένα ενιαίο DataFrame, γίνεται καθαρισμός των δεδομένων: αντικαθίστανται τα ελλιπή στοιχεία με τη διάμεσο τιμή κάθε στήλης, ενώ όλες οι μεταβλητές μετατρέπονται σε αριθμητική μορφή για να είναι κατάλληλες για εκπαίδευση σε νευρωνικό δίκτυο.

Στη συνέχεια, διαχωρίζονται τα χαρακτηριστικά (X) από την ετικέτα στόχο (y), με τη μετατροπή του target σε δυαδική μορφή (0 για απουσία νόσου, 1 για παρουσία νόσου). Ακολουθεί κανονικοποίηση των χαρακτηριστικών με StandardScaler και διαχωρισμός σε σύνολα εκπαίδευσης (80%) και δοκιμών (20%). Το μοντέλο αποτελείται από τρεις πυκνές στρώσεις: δύο κρυφά επίπεδα με ενεργοποίηση ReLU και ένα τελικό επίπεδο με sigmoid, κατάλληλο για δυαδική ταξινόμηση. Για τη διασφάλιση της ιδιωτικότητας χρησιμοποιείται ο αλγόριθμος DP-SGD, ο οποίος προσθέτει στοχαστικό θόρυβο στα gradients, περιορίζοντας έτσι τον κίνδυνο ανάκτησης ευαίσθητων πληροφοριών από τα δεδομένα εκπαίδευσης. Οι βασικές υπερπαράμετροι του DP-SGD είναι: learning rate 0.01, noise multiplier 1.1, l2 clipping norm 1.0 και αριθμός εποχών 50. Το μόνο στοιχείο που διαφοροποιείται στο πείραμα είναι το batch size (16, 32 και 64).

Από τα διαγράμματα που προκύπτουν για κάθε περίπτωση, παρατηρείται σημαντικά βελτιωμένη συμπεριφορά σε σχέση με προηγούμενες δοκιμές. Το validation loss μειώνεται σταθερά κατά τη διάρκεια της εκπαίδευσης και η καμπύλη δεν εμφανίζει ακραίες διακυμάνσεις, γεγονός που δείχνει καλή σύγκλιση και σταθερότητα του μοντέλου. Αντίστοιχα, η ακρίβεια στο σύνολο επικύρωσης αυξάνεται συνεχώς και φτάνει κοντά ή και πάνω από το 80%, χωρίς σημάδια υπερβολικού υπερπροσαρμογής. Η απουσία απότομων αιχμών ή αστάθειας τόσο στο train όσο και στο validation set δείχνει ότι ο συνδυασμός των υπερπαραμέτρων είναι επαρκώς ρυθμισμένος για αυτό το dataset. Παρότι ο θόρυβος εισάγεται μέσω DP-SGD, η απόδοση του μοντέλου παραμένει υψηλή, επιτυγχάνοντας αξιοσημείωτη γενίκευση στα δεδομένα δοκιμών.

Η τελική ακρίβεια (validation accuracy) περίπου 80% υποδηλώνει ότι η εκπαίδευση με DP-SGD σε αυτό το dataset μπορεί να επιτευχθεί χωρίς δραματική απώλεια απόδοσης, αρκεί να γίνει σωστή επιλογή υπερπαραμέτρων. Η σταθερή πτώση της απώλειας και η σταθερή άνοδος της ακρίβειας δείχνουν ότι το μοντέλο μαθαίνει αποτελεσματικά και αξιοποιεί τις πληροφορίες των δεδομένων, παρά τον θόρυβο που προστίθεται για λόγους ιδιωτικότητας.

Συμπερασματικά, η παρούσα υλοποίηση επιβεβαιώνει ότι το DP-SGD μπορεί να εφαρμοστεί αποτελεσματικά σε ευαίσθητα ιατρικά δεδομένα, χωρίς να διακυβεύεται η απόδοση του μοντέλου, εφόσον χρησιμοποιηθούν κατάλληλες υπερπαράμετροι και σταθερή αρχιτεκτονική. Για μελλοντική βελτίωση, θα μπορούσε να εξεταστεί η επίδραση μικρότερης τιμής noise multiplier ή άλλων τεχνικών σταθεροποίησης (όπως dropout ή early stopping), καθώς και η αποθήκευση και ανάλυση των αποτελεσμάτων για σύγκριση με άλλα πειραματικά σενάρια.

Εικόνα 1: Εξέλιξη Απώλειας και Ακρίβειας Εκπαίδευσης/Επικύρωσης ανά Epochs για την περίπτωση Heart Disease Dataset.

Από τα διαγράμματα που παρουσιάζουν την επίδοση του νευρωνικού δικτύου κατά τη διάρκεια της εκπαίδευσης με μέγεθος παρτίδας (batch size) 16 και χρήση του αλγορίθμου DP-SGD παρατηρείται ότι στο αριστερό διάγραμμα απεικονίζεται η μεταβολή της τιμής της συνάρτησης κόστους (loss) τόσο στο εκπαιδευτικό όσο και στο επικυρωτικό σύνολο σε κάθε εποχή, δείχνοντας μια γενική πτωτική τάση, γεγονός που υποδηλώνει ότι το μοντέλο μαθαίνει σταδιακά και μειώνει το σφάλμα του. Στο δεξί διάγραμμα παρουσιάζεται η ακρίβεια (accuracy) του μοντέλου στις ίδιες εποχές, όπου παρατηρείται σαφής βελτίωση

στην απόδοση του μοντέλου ήδη από τα πρώτα βήματα της εκπαίδευσης, με την ακρίβεια να σταθεροποιείται κοντά στο 80% μετά από περίπου 10–15 εποχές. Η μικρή απόκλιση μεταξύ των τιμών εκπαίδευσης και επικύρωσης δείχνει ότι δεν παρατηρείται σημαντικό φαινόμενο υπερεκπαίδευσης, υποδεικνύοντας καλή γενίκευση του μοντέλου υπό καθεστώς διαφορικής ιδιωτικότητας.

Στην περίπτωση που το batch size είναι 32 από το αριστερό διάγραμμα φαίνεται η μείωση της τιμής της συνάρτησης απώλειας τόσο για το εκπαιδευτικό όσο και για το επικυρωτικό σύνολο. Και οι δύο καμπύλες παρουσιάζουν σταθερά πτωτική πορεία, χωρίς απότομες διακυμάνσεις, και συγκλίνουν σταδιακά μεταξύ τους. Αυτό φανερώνει πως το μοντέλο καταφέρνει να προσαρμόζεται στο εκπαιδευτικό σύνολο χωρίς να υπερεκπαιδεύεται, ενώ διατηρεί ικανοποιητική γενίκευση. Στο δεξί διάγραμμα, η ακρίβεια ανεβαίνει απότομα στις πρώτες εποχές, από περίπου 45% έως σχεδόν 80%, και στη συνέχεια σταθεροποιείται σε υψηλά επίπεδα. Η πολύ μικρή διαφορά μεταξύ training και validation accuracy στις τελευταίες εποχές επιβεβαιώνει την καλή απόδοση του μοντέλου σε δεδομένα που δεν έχει δει. Οι καμπύλες ακρίβειας είναι σχεδόν ταυτόσημες, υποδηλώνοντας πως το μοντέλο μαθαίνει ουσιαστικά και όχι απλώς αποστηθίζει τα δεδομένα εκπαίδευσης.

Στην περίπτωση του batch size 64, το μοντέλο παρουσιάζει μια ομαλή και σταθερή πορεία εκμάθησης κατά τη διάρκεια των 50 εποχών. Η απώλεια (loss) στο αριστερό διάγραμμα μειώνεται σταδιακά τόσο για το σύνολο εκπαίδευσης όσο και για το σύνολο επικύρωσης, με τις δύο καμπύλες να παραμένουν κοντά μεταξύ τους χωρίς σημαντικές αποκλίσεις. Η διαφορά μεταξύ τους είναι μικρή και σταθερή, γεγονός που υποδηλώνει ότι το μοντέλο μαθαίνει αποτελεσματικά χωρίς να υπερεκπαιδεύεται, διατηρώντας παράλληλα την ικανότητά του να γενικεύει σε νέα, αόρατα δεδομένα. Στο δεξί διάγραμμα, η ακρίβεια εμφανίζει συνεχή βελτίωση καθ' όλη τη διάρκεια της εκπαίδευσης. Από την αρχική τιμή κοντά στο 45%, η ακρίβεια ανεβαίνει προοδευτικά και φτάνει περίπου το 78–80% προς το τέλος της διαδικασίας, με τις καμπύλες εκπαίδευσης και επικύρωσης να παραμένουν σχεδόν ταυτόσημες. Η απουσία απότομων διακυμάνσεων ή έντονων αποκλίσεων ενισχύει την εικόνα ενός σταθερού και αξιόπιστου μοντέλου. Το batch size των 64 δειγμάτων φαίνεται να προσφέρει μια ισορροπία μεταξύ αποτελεσματικής εκπαίδευσης και διατήρησης της ιδιωτικότητας, χωρίς να θυσιάζεται η ακρίβεια ή η σταθερότητα της απόδοσης.

Η σύγκριση των αποτελεσμάτων για τα τρία διαφορετικά μεγέθη παρτίδας, 16, 32 και 64, αποκαλύπτει ενδιαφέρουσες διαφοροποιήσεις στη μαθησιακή συμπεριφορά του μοντέλου και υποδεικνύει τη σημασία της επιλογής του batch size όταν εφαρμόζεται εκπαίδευση με διαφορική ιδιωτικότητα (DP-SGD). Στην περίπτωση του batch size 16, παρατηρείται ταχεία αρχική βελτίωση τόσο στην ακρίβεια όσο και στη μείωση της απώλειας, με σταδιακή σταθεροποίηση των μετρικών γύρω από 80% ακρίβεια και απώλεια περίπου 0.45. Η καμπύλη επικύρωσης ακολουθεί στενά την αντίστοιχη της εκπαίδευσης, κάτι που καταδεικνύει ισχυρή γενίκευση, αν και σημειώνονται ορισμένες διακυμάνσεις στην απώλεια λόγω του μικρού μεγέθους παρτίδας και της αυξημένης στοχαστικότητας. Στο batch size 32, το μοντέλο επιτυγχάνει επίσης υψηλή ακρίβεια, ενώ η απώλεια μειώνεται ομαλά και οι καμπύλες παραμένουν κοντά καθ' όλη τη διάρκεια της εκπαίδευσης. Σε αυτή την περίπτωση, η ισορροπία μεταξύ σταθερότητας και ικανότητας γενίκευσης είναι εντονότερη, με τις διακυμάνσεις να είναι πιο ελεγχόμενες σε σχέση με την περίπτωση του batch size 16. Το batch size 32 φαίνεται να προσφέρει μια ιδανική αναλογία μεταξύ αριθμού παραδειγμάτων ανά ενημέρωση και αποτελεσματικής εφαρμογής του θορύβου που επιβάλλει η διαφορική ιδιωτικότητα, ενισχύοντας τη συνολική απόδοση του μοντέλου. Όταν το batch size αυξάνεται σε 64, παρατηρείται μια ακόμα πιο ομαλή πορεία τόσο στην απώλεια όσο και στην ακρίβεια. Οι καμπύλες παρουσιάζουν λιγότερη στοχαστικότητα,

γεγονός που οφείλεται στον μεγαλύτερο αριθμό δειγμάτων ανά βήμα ενημέρωσης. Ωστόσο, ενώ η ακρίβεια παραμένει υψηλή, παρατηρείται μια ελαφρώς πιο αργή σύγκλιση σε σχέση με τις μικρότερες παρτίδες, και οι τελικές επιδόσεις προσεγγίζουν αλλά δεν ξεπερνούν εκείνες των batch sizes 16 και 32. Αυτό υποδεικνύει ότι η αυξημένη σταθερότητα που προσφέρει το μεγάλο batch size ενδέχεται να έρχεται εις βάρος της ευελιξίας στην εκμάθηση σε περιβάλλοντα με προσθήκη θορύβου, όπως αυτό της διαφορικής ιδιωτικότητας. Η αλλαγή στο batch size επηρεάζει τόσο τη δυναμική της εκπαίδευσης όσο και την ποιότητα της γενίκευσης. Μικρότερα batch sizes, όπως το 16, προσφέρουν μεγαλύτερη προσαρμοστικότητα αλλά με αυξημένη στοχαστικότητα, ενώ μεγαλύτερα batch sizes, όπως το 64, οδηγούν σε πιο σταθερές αλλά ενδεχομένως πιο αργές ή περιορισμένες βελτιώσεις. Η ενδιάμεση τιμή των 32 φαίνεται να επιτυγχάνει τον πιο ικανοποιητικό συμβιβασμό, παρέχοντας υψηλή ακρίβεια, ομαλή εκμάθηση και καλή προστασία ιδιωτικότητας.

7.1.1 Διαδικασία ρύθμισης των υπερπαραμέτρων

Ο κώδικας έχει ως στόχο να εκπαιδεύσει ένα νευρωνικό δίκτυο με χρήση του μηχανισμού διαφορικής ιδιωτικότητας DP-SGD και να μελετήσει την επίδραση διαφόρων υπερπαραμέτρων στην απόδοσή του. Η διαδικασία ξεκινά με τη φόρτωση των προεπεξεργασμένων δεδομένων από ένα αρχείο τύπου pickle (dataset.pkl), το οποίο περιέχει τα σύνολα εκπαίδευσης και δοκιμών. Στη συνέχεια, ορίζονται διαφορετικές τιμές για βασικές υπερπαραμέτρους, όπως ο ρυθμός μάθησης (learning_rate), η ένταση του προστιθέμενου θορύβου (noise_multiplier), η παράμετρος αποκοπής των gradients (l2_norm_clip), το μέγεθος του batch και ο αριθμός των εποχών εκπαίδευσης. Ο πειραματισμός καλύπτει όλους τους δυνατούς συνδυασμούς αυτών των υπερπαραμέτρων με χρήση της συνάρτησης itertools.product(), επιτρέποντας την εκτέλεση ενός grid search σε πολλαπλές διαμορφώσεις. Για κάθε συνδυασμό, δημιουργείται ένα απλό πλήρως συνδεδεμένο νευρωνικό δίκτυο τριών επιπέδων με δύο ενδιάμεσες ReLU ενεργοποιήσεις και μία σιγμοειδή έξοδο για δυαδική ταξινόμηση. Η εκπαίδευση γίνεται με χρήση του DPKerasSGDOptimizer, ο οποίος ενσωματώνει μηχανισμούς διαφορικής ιδιωτικότητας μέσω προσθήκης θορύβου και αποκοπής των gradients. Κατά την εκπαίδευση κάθε μοντέλου, καταγράφονται οι τιμές της ακρίβειας, της απώλειας, καθώς και των κλασικών μετρικών precision, recall, f1 score και ROC-AUC score, ενώ υπολογίζεται επίσης και το ελάχιστο ε (epsilon), δηλαδή το privacy budget που αντιστοιχεί σε κάθε συνδυασμό παραμέτρων. Όλα τα αποτελέσματα αποθηκεύονται σε μια λίστα και τελικά συγκεντρώνονται σε ένα pandas DataFrame (results_df).

Για την ανάλυση των αποτελεσμάτων δημιουργούνται γραφήματα που δείχνουν τη σχέση μεταξύ των υπερπαραμέτρων και των επιδόσεων του μοντέλου. Ένα διάγραμμα απεικονίζει τη συσχέτιση μεταξύ learning rate και test accuracy για διαφορετικά επίπεδα θορύβου (noise multiplier), ενώ ένα δεύτερο δείχνει την επίδραση του l2 clip και του batch size στην απώλεια. Αυτά τα διαγράμματα βοηθούν στην κατανόηση των προτύπων και της ευαισθησίας του μοντέλου στις υπερπαραμέτρους. Ακολούθως, γίνεται ποιοτική σύγκριση μεταξύ του μοντέλου με DP-SGD και ενός κλασικού μοντέλου χωρίς διαφορική ιδιωτικότητα, εκπαιδευμένου με τον Adam optimizer. Για κάθε μοντέλο εμφανίζεται ο πίνακας σύγχυσης (confusion matrix), καθώς και τα false positives και false negatives, ώστε να αξιολογηθεί η επίδραση του DP στην ακρίβεια και τη σταθερότητα του μοντέλου. Επιπλέον, παρουσιάζονται βασικά στατιστικά για το σύνολο των πειραμάτων, όπως ο μέσος όρος και η τυπική απόκλιση της ακρίβειας και της απώλειας. Τέλος, ορίζεται η παράμετρος δ (delta) με βάση το μέγεθος του συνόλου εκπαίδευσης και γίνεται επιπλέον υπολογισμός

των τιμών ε για διαφορετικά noise multipliers, προσφέροντας πληρέστερη εικόνα του trade-off μεταξύ ιδιωτικότητας και απόδοσης. Τα τελικά αποτελέσματα αποθηκεύονται σε αρχείο CSV για μελλοντική ανάλυση. Ο κώδικας προσφέρει έτσι μια ολοκληρωμένη πειραματική μεθοδολογία για την κατανόηση και τη ρύθμιση του DP-SGD σε εφαρμογές μηχανικής μάθησης με ευαίσθητα δεδομένα.

```
import itertools
import matplotlib.pyplot as plt
import pickle
import tensorflow as tf
import tensorflow_privacy
import numpy as np
import pandas as pd
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score,
precision_score, recall_score, f1_score
import seaborn as sns
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy_lib

# 1. Φόρτωση των δεδομένων
with open("dataset.pkl", "rb") as f:
    X_train, X_test, y_train, y_test = pickle.load(f)

print("Dataset loaded successfully!")

# 2. Υπερπαράμετροι
learning_rates = [0.001, 0.01, 0.1]
noise_multipliers = [0.5, 1.1, 2.0]
l2_norm_clips = [0.5, 1.0, 2.0]
batch_sizes = [16, 32, 64]
epochs = 30

results = []

# 3. Πειράματα με όλους τους συνδυασμούς
for lr, noise, clip, batch in itertools.product(learning_rates, noise_multipliers,
l2_norm_clips, batch_sizes):
    print(f"Training model with LR={lr}, Noise={noise}, Clip={clip}, Batch={batch}")

    batch_size = (len(X_train) // batch) * batch
    if batch_size == 0:
        batch_size = 1

    model = tf.keras.Sequential([
        tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train.shape[1],)),
        tf.keras.layers.Dense(8, activation='relu'),
        tf.keras.layers.Dense(1, activation='sigmoid')
    ])

    dp_optimizer = tensorflow_privacy.DPKerasSGDOptimizer(
```

```
        l2_norm_clip=clip,
        noise_multiplier=noise,
        num_microbatches=1,
        learning_rate=lr
    )

    model.compile(optimizer=dp_optimizer, loss='binary_crossentropy',
metrics=['accuracy'])
    history = model.fit(X_train, y_train, epochs=epochs, batch_size=batch_size,
validation_data=(X_test, y_test), verbose=0)

    test_loss, test_acc = model.evaluate(X_test, y_test, verbose=0)
    print(f"Test Accuracy: {test_acc:.4f} | Test Loss: {test_loss:.4f}")

    y_pred = (model.predict(X_test) > 0.5).astype("int32")
    y_proba = model.predict(X_test)

    precision = precision_score(y_test, y_pred, zero_division=0)
    recall = recall_score(y_test, y_pred, zero_division=0)
    f1 = f1_score(y_test, y_pred, zero_division=0)
    roc_auc = roc_auc_score(y_test, y_proba)

    eps, _ = compute_dp_sgd_privacy_lib.compute_dp_sgd_privacy(
        n=len(X_train),
        batch_size=batch_size,
        noise_multiplier=noise,
        epochs=epochs,
        delta=1 / (len(X_train) * np.sqrt(len(X_train)))
    )

    results.append({
        "learning_rate": lr,
        "noise_multiplier": noise,
        "l2_norm_clip": clip,
        "batch_size": batch,
        "test_accuracy": test_acc,
        "test_loss": test_loss,
        "precision": precision,
        "recall": recall,
        "f1_score": f1,
        "roc_auc": roc_auc,
        "epsilon": eps
    })

# 4. Μετατροπή σε DataFrame
results_df = pd.DataFrame(results)

# 5. Διαγράμματα
```

```
plt.figure(figsize=(12, 6))
for noise in results_df["noise_multiplier"].unique():
    subset = results_df[results_df["noise_multiplier"] == noise]
    plt.plot(subset["learning_rate"], subset["test_accuracy"], label=f"Noise={noise}")
plt.xlabel("Learning Rate")
plt.ylabel("Test Accuracy")
plt.title("Effect of Learning Rate & Noise on Accuracy")
plt.legend()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
for clip in results_df["l2_norm_clip"].unique():
    subset = results_df[results_df["l2_norm_clip"] == clip]
    plt.plot(subset["batch_size"], subset["test_loss"], label=f"Clip={clip}")
plt.xlabel("Batch Size")
plt.ylabel("Test Loss")
plt.title("Effect of Clipping & Batch Size on Loss")
plt.legend()
plt.grid(True)
plt.show()

# 6. Classification Report
print("Neos kodikas")
predictions = (model.predict(X_test) > 0.5).astype("int32")
print("\nClassification Report (Dataset with DP-SGD):")
print(classification_report(y_test, predictions))
roc_auc = roc_auc_score(y_test, model.predict(X_test))
print(f"ROC-AUC Score: {roc_auc:.4f}")

cm = confusion_matrix(y_test, predictions)
plt.figure(figsize=(8,6))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.title("Confusion Matrix (with DP-SGD)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

# 7. Μοντέλο χωρίς DP
model_no_dp = tf.keras.Sequential([
    tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train.shape[1],)),
    tf.keras.layers.Dense(8, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
model_no_dp.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model_no_dp.fit(X_train, y_train, epochs=epochs, batch_size=32, verbose=0)

predictions_no_dp = (model_no_dp.predict(X_test) > 0.5).astype("int32")
```

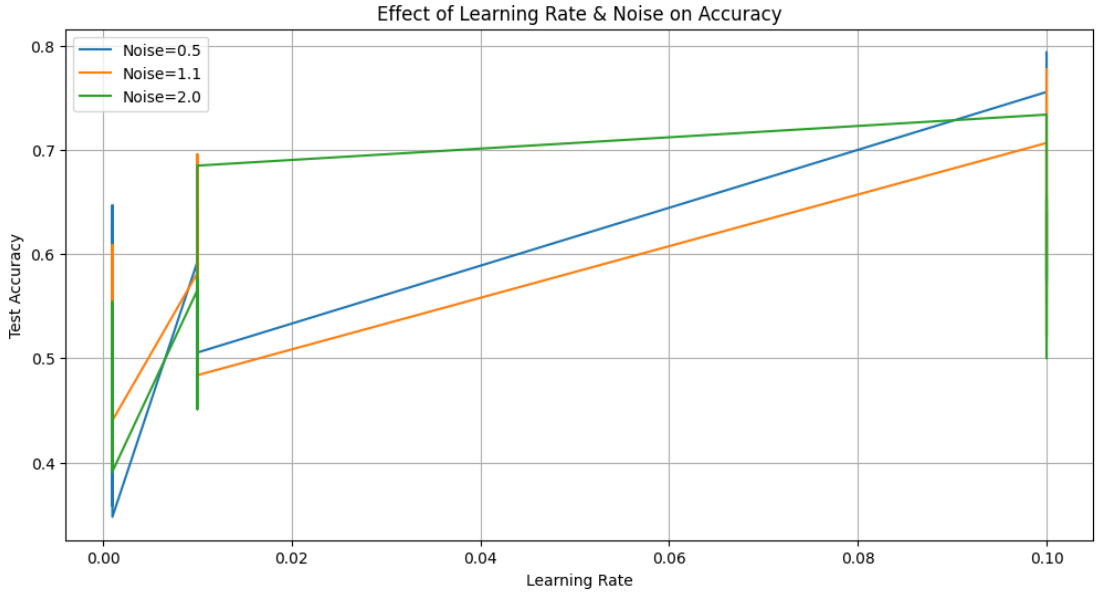
```
cm_no_dp = confusion_matrix(y_test, predictions_no_dp)

plt.figure(figsize=(8,6))
sns.heatmap(cm_no_dp, annot=True, fmt="d", cmap="Greens")
plt.title("Confusion Matrix (without DP-SGD)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

fp_dp, fn_dp = cm[0][1], cm[1][0]
fp_no_dp, fn_no_dp = cm_no_dp[0][1], cm_no_dp[1][0]
print(f'False Positives (with DP-SGD): {fp_dp}, False Negatives (with DP-SGD):
{fn_dp}")
print(f'False Positives (no DP-SGD): {fp_no_dp}, False Negatives (no DP-SGD):
{fn_no_dp}")

# 8. Στατιστικά
mean_accuracy = results_df['test_accuracy'].mean()
std_accuracy = results_df['test_accuracy'].std()
mean_loss = results_df['test_loss'].mean()
std_loss = results_df['test_loss'].std()
print("\nStatistical Analysis of Results:")
print(f'Mean Accuracy: {mean_accuracy:.4f}, Std Accuracy: {std_accuracy:.4f}")
print(f'Mean Loss: {mean_loss:.4f}, Std Loss: {std_loss:.4f}")

# 9. Σχέση ε και noise multiplier
num_samples = X_train.shape[0]
delta = 1 / (num_samples * np.sqrt(num_samples))
print(f"\nDelta (δ) used for experiments: {delta:.8f}")
print("The delta value ensures theoretical guarantees within differential privacy context.")

print("\nRelationship between privacy budget (ε) and noise multiplier:")
for noise in noise_multipliers:
    epsilon, _ = tensorflow_privacy.compute_dp_sgd_privacy(n=num_samples,
                                        batch_size=batch_sizes[0],
                                        noise_multiplier=noise,
                                        epochs=epochs,
                                        delta=delta)
    print(f"Noise Multiplier: {noise} => Epsilon (ε): {epsilon:.4f}")

# 10. Αποθήκευση σε CSV
results_df.to_csv('results_df_Hyperparameter_Tuning_heart_disease.csv', index=False)
print("\nResults saved successfully in
'results_df_Hyperparameter_Tuning_heart_disease.csv'")
```

Η διαδικασία ρύθμισης των υπερπαραμέτρων στον διαφορικά ιδιωτικό στοχαστικό βαθμιδωτό καταβιβασμό (DP-SGD) αποδείχθηκε καθοριστική για την τελική απόδοση του μοντέλου. Οι υπερπαράμετροι που διερευνήθηκαν ήταν ο ρυθμός εκμάθησης, ο

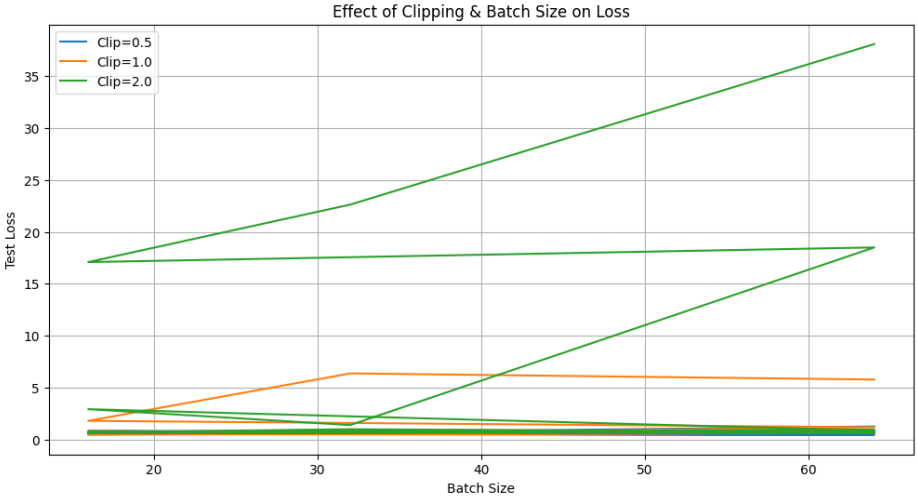
πολλαπλασιαστής θορύβου, το όριο αποκοπής (clipping norm) και το μέγεθος παρτίδας. Η πειραματική ανάλυση κατέδειξε ότι η υψηλότερη ακρίβεια που επιτεύχθηκε ήταν 80.43%, όταν χρησιμοποιήθηκε learning rate ίσο με 0.1, noise multiplier 0.5, clipping norm 0.5 και batch size 16. Το αποτέλεσμα αυτό συνοδεύτηκε από υψηλό recall (88.24%) και f1-score (83.33%), ενώ η ακρίβεια (precision) ήταν 78.95% και το roc_auc score έφτασε στο 88.22%. Η συγκεκριμένη ρύθμιση, παρότι ιδιαίτερα αποτελεσματική ως προς την απόδοση, παρείχε ελάχιστη ιδιωτικότητα, όπως φαίνεται από την τιμή του epsilon που έφτασε τις 107.89 μονάδες.

Αναλύοντας τη συνολική συμπεριφορά του μοντέλου σε όλα τα πειράματα, η μέση τιμή ακρίβειας υπολογίστηκε στο 57.96% με τυπική απόκλιση 11.43%, γεγονός που καταδεικνύει υψηλή εξάρτηση από τις παραμέτρους. Αντίστοιχα, η μέση τιμή της συνάρτησης απώλειας ήταν 1.76 με υψηλή τυπική απόκλιση 4.01, δείγμα έντονων διακυμάνσεων στην εκπαίδευση και πιθανής αποσταθεροποίησης σε ορισμένους συνδυασμούς.

Η αύξηση του learning rate συνέβαλε στη βελτίωση της ακρίβειας, με την τιμή 0.1 να οδηγεί στα καλύτερα αποτελέσματα. Ωστόσο, όταν συνδυάστηκε με υψηλό θόρυβο ή μεγάλο clipping norm, οδήγησε σε αστάθεια. Η παράμετρος noise multiplier παρουσίασε την αναμενόμενη επίδραση: όσο αυξανόταν, τόσο μειωνόταν το privacy budget (ε), βελτιώνοντας την ιδιωτικότητα αλλά εις βάρος της ακρίβειας. Ενδεικτικά, για noise=2.0, η τιμή του ε μειώθηκε στις 1.72 μονάδες, αλλά συνοδεύτηκε από πτώση της απόδοσης. Το clipping norm επηρέασε κυρίως τη σταθερότητα της εκμάθησης, με την τιμή 1.0 να προσφέρει ισορροπία μεταξύ αποτελεσματικής εκπαίδευσης και διατήρησης ελέγχου στις παραμορφώσεις που προκαλεί ο θόρυβος. Το batch size επηρέασε τη συνολική συμπεριφορά, με τις μικρότερες τιμές (16 και 32) να οδηγούν σε υψηλότερη ακρίβεια, ενώ το batch size 64 συνδέθηκε με αύξηση της απώλειας και μείωση της γενίκευσης σε ορισμένους συνδυασμούς.



Εικόνα 2: Επίδραση του Learning Rate και του Θορύβου στην Ακρίβεια Δοκιμής για την περίπτωση Heart Disease Dataset.



Εικόνα 3: Επίδραση του Clipping και του Batch Size στην Απώλεια Δοκιμής για την περίπτωση Heart Disease Dataset.

Συνολικά, τα ευρήματα υποδεικνύουν ότι η επίτευξη ικανοποιητικής ακρίβειας υπό καθεστώς διαφορικής ιδιωτικότητας είναι εφικτή, αλλά απαιτεί ιδιαίτερη προσοχή στη ρύθμιση των υπερπαραμέτρων. Η βέλτιστη ισορροπία επιτυγχάνεται όταν ο θόρυβος παραμένει σε χαμηλό επίπεδο, η εκπαίδευση γίνεται με μικρές παρτίδες και το clipping norm είναι προσεκτικά ρυθμισμένο ώστε να περιορίζει τις αποκλίσεις χωρίς να περιορίζει υπερβολικά τα gradients.

learning_rate	noise_multiplier	l2_norm_clip	batch_size	test_accuracy	test_loss	precision	recall	f1_score	roc_auc	epsilon
0.1	0.5	0.5	16	0.8043	0.4948	0.7895	0.8824	0.8333	0.8822	107.8941
0.1	1.1	2	16	0.7717	2.5156	0.8	0.7843	0.7921	0.8016	33.3032
0.1	2	0.5	16	0.7283	0.8651	0.7407	0.7843	0.7619	0.8253	14.8545
0.01	2	2	16	0.7174	0.54	0.7404	0.7549	0.7476	0.8112	14.8545
0.1	0.5	1	16	0.712	0.5666	0.7025	0.8333	0.7623	0.7868	107.8941
0.1	1.1	0.5	16	0.7065	0.5858	0.7791	0.6569	0.7128	0.7928	33.3032
0.01	1.1	2	16	0.6196	0.6568	0.6481	0.6863	0.6667	0.643	33.3032
0.01	1.1	1	16	0.5978	0.654	0.5933	0.8725	0.7063	0.6657	33.3032
0.01	0.5	1	16	0.5761	0.6594	0.625	0.5882	0.6061	0.641	107.8941
0.01	0.5	0.5	16	0.5543	0.7012	0.5543	1	0.7133	0.6013	107.8941
0.01	1.1	0.5	16	0.5543	0.7176	0.8846	0.2255	0.3594	0.5946	33.3032
0.1	2	2	16	0.5489	18.9567	0.5664	0.7941	0.6612	0.5114	14.8545
0.001	1.1	2	16	0.5435	0.711	0.5511	0.951	0.6978	0.5354	33.3032
0.001	2	0.5	16	0.538	0.7346	0.547	0.9706	0.6996	0.289	14.8545
0.001	2	1	16	0.538	0.7315	0.5475	0.9608	0.6975	0.4993	14.8545
0.001	0.5	2	16	0.5217	0.772	0.6029	0.402	0.4824	0.5161	107.8941
0.1	0.5	2	16	0.5163	1.1303	0.5942	0.402	0.4795	0.5409	107.8941
0.001	1.1	1	16	0.4891	0.7142	0.6176	0.2059	0.3088	0.5646	33.3032

0.1	1.1	1	16	0.4891	0.8905	0.5588	0.3725	0.4471	0.4978	33.3032
0.001	2	2	16	0.4783	0.8106	0.5179	0.8529	0.6444	0.3785	14.8545
0.01	2	1	16	0.4783	0.7363	0.5789	0.2157	0.3143	0.4484	14.8545
0.01	0.5	2	16	0.462	0.6877	0.5103	0.7255	0.5992	0.4937	107.8941
0.01	2	0.5	16	0.4076	0.7775	0.3939	0.1275	0.1926	0.4168	14.8545
0.001	0.5	1	16	0.3967	0.7597	0.4308	0.2745	0.3353	0.3551	107.8941
0.001	1.1	0.5	16	0.3967	0.8043	0.3902	0.1569	0.2238	0.2924	33.3032
0.1	2	1	16	0.3967	2.4769	0.4563	0.4608	0.4585	0.3584	14.8545
0.001	0.5	0.5	16	0.2935	0.8279	0.3444	0.3039	0.3229	0.288	107.8941
0.1	1.1	1	32	0.7446	0.7286	0.7523	0.8039	0.7773	0.7812	33.3032
0.01	1.1	1	32	0.7337	0.5691	0.9206	0.5686	0.703	0.8398	33.3032
0.1	0.5	1	32	0.7283	0.554	0.7203	0.8333	0.7727	0.7877	107.8941
0.1	2	2	32	0.7283	18.8377	0.697	0.902	0.7863	0.7356	14.8545
0.1	2	0.5	32	0.6793	0.7004	0.7722	0.598	0.674	0.7781	14.8545
0.1	0.5	0.5	32	0.6739	0.5733	0.6721	0.8039	0.7321	0.7727	107.8941
0.01	0.5	1	32	0.6685	0.6675	0.6614	0.8235	0.7336	0.6399	107.8941
0.01	0.5	0.5	32	0.6522	0.626	0.6439	0.8333	0.7265	0.7257	107.8941
0.1	0.5	2	32	0.6522	0.7747	0.6462	0.8235	0.7241	0.7269	107.8941
0.001	2	2	32	0.6413	0.5851	0.62	0.9118	0.7381	0.8159	14.8545
0.001	2	1	32	0.625	0.654	0.6602	0.6667	0.6634	0.6711	14.8545
0.01	1.1	2	32	0.6087	0.6172	0.5862	1	0.7391	0.7956	33.3032
0.1	1.1	0.5	32	0.6087	0.6314	0.6119	0.8039	0.6949	0.7397	33.3032
0.01	1.1	0.5	32	0.5978	0.6874	0.5814	0.9804	0.7299	0.6532	33.3032
0.01	2	0.5	32	0.587	0.6558	0.5765	0.9608	0.7206	0.6719	14.8545
0.01	2	1	32	0.587	0.7033	0.6204	0.6569	0.6381	0.5411	14.8545
0.1	1.1	2	32	0.5598	2.9981	0.5913	0.6667	0.6267	0.5304	33.3032
0.001	0.5	0.5	32	0.5543	0.6772	0.6389	0.451	0.5287	0.6021	107.8941
0.001	0.5	2	32	0.5435	0.7329	0.5592	0.8333	0.6693	0.4592	107.8941
0.001	0.5	1	32	0.5	0.7342	0.5758	0.3725	0.4524	0.5074	107.8941
0.001	1.1	1	32	0.5	0.7089	0.625	0.2451	0.3521	0.6023	33.3032
0.01	2	2	32	0.4674	0.8237	0.5625	0.1765	0.2687	0.516	14.8545
0.001	1.1	0.5	32	0.4565	0.8655	0.5067	0.7451	0.6032	0.2178	33.3032
0.01	0.5	2	32	0.4565	0.7129	0.6667	0.0392	0.0741	0.61	107.8941
0.001	1.1	2	32	0.4457	0.7952	0	0	0	0.4646	33.3032
0.1	2	1	32	0.4293	7.0891	0.4737	0.2647	0.3396	0.5054	14.8545

0.001	2	0.5	32	0.3967	0.9538	0.32	0.0784	0.126	0.4055	14.8545
0.1	1.1	0.5	64	0.7935	0.5481	0.75	0.9412	0.8348	0.8693	33.2548
0.1	0.5	0.5	64	0.788	0.5372	0.8058	0.8137	0.8098	0.8357	110.2497
0.1	0.5	2	64	0.7391	0.6107	0.77	0.7549	0.7624	0.8015	110.2497
0.1	1.1	2	64	0.7337	4.2824	0.7978	0.6961	0.7435	0.7731	33.2548
0.01	0.5	2	64	0.7065	0.6014	0.7727	0.6667	0.7158	0.75	110.2497
0.1	0.5	1	64	0.6957	0.611	0.6949	0.8039	0.7455	0.7643	110.2497
0.001	1.1	1	64	0.6902	0.6339	0.6552	0.9314	0.7692	0.7667	33.2548
0.01	2	1	64	0.6848	0.6215	0.7444	0.6569	0.6979	0.7334	14.7599
0.01	2	2	64	0.6793	0.6782	0.6525	0.902	0.7572	0.6917	14.7599
0.1	2	1	64	0.6793	8.7917	0.6693	0.8333	0.7424	0.7282	14.7599
0.01	1.1	1	64	0.663	0.6917	0.75	0.5882	0.6593	0.6817	33.2548
0.1	2	2	64	0.663	24.8581	0.7632	0.5686	0.6517	0.7174	14.7599
0.1	1.1	1	64	0.6413	1.1672	0.7647	0.5098	0.6118	0.6648	33.2548
0.001	2	2	64	0.5978	0.7111	0.5946	0.8627	0.704	0.4866	14.7599
0.001	1.1	0.5	64	0.5435	0.6897	0.5523	0.9314	0.6934	0.5039	33.2548
0.01	0.5	1	64	0.5435	0.6857	0.5584	0.8431	0.6719	0.552	110.2497
0.001	0.5	2	64	0.538	0.7112	0.6809	0.3137	0.4295	0.5962	110.2497
0.1	2	0.5	64	0.5326	1.7154	0.6739	0.3039	0.4189	0.609	14.7599
0.001	0.5	1	64	0.5272	0.7172	0.5497	0.8137	0.6561	0.4473	110.2497
0.01	2	0.5	64	0.5217	0.6776	0.5422	0.8824	0.6716	0.5705	14.7599
0.01	1.1	0.5	64	0.5163	0.756	0.573	0.5	0.534	0.4737	33.2548
0.01	1.1	2	64	0.5163	0.7374	0.5867	0.4314	0.4972	0.5166	33.2548
0.01	0.5	0.5	64	0.5054	0.7058	0.5401	0.7255	0.6192	0.4234	110.2497
0.001	2	1	64	0.4891	0.7025	0.5417	0.5098	0.5253	0.5282	14.7599
0.001	1.1	2	64	0.462	0.7562	0.5181	0.4216	0.4649	0.4345	33.2548
0.001	2	0.5	64	0.4185	0.7554	0.4138	0.1176	0.1832	0.4489	14.7599
0.001	0.5	0.5	64	0.413	0.7525	0.4783	0.6471	0.55	0.3386	110.2497

Πίνακας 2: Επιδράσεις των Υπερπαραμέτρων στην Απόδοση του Μοντέλου με DP-SGD στο Heart Disease Dataset

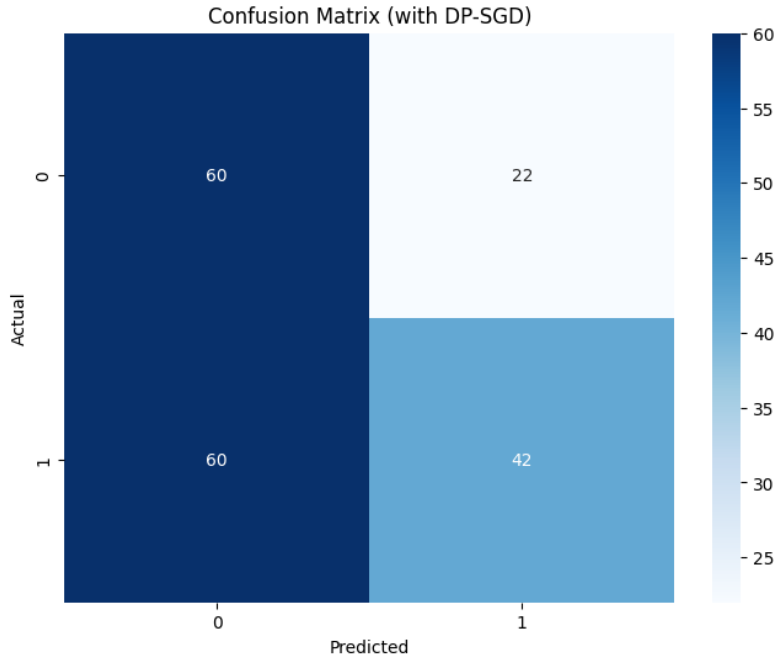
Classification Report (Dataset with DP-SGD):

	precision	recall	f1-score	support
0	0.50	0.73	0.59	82
1	0.66	0.41	0.51	102
accuracy			0.55	184

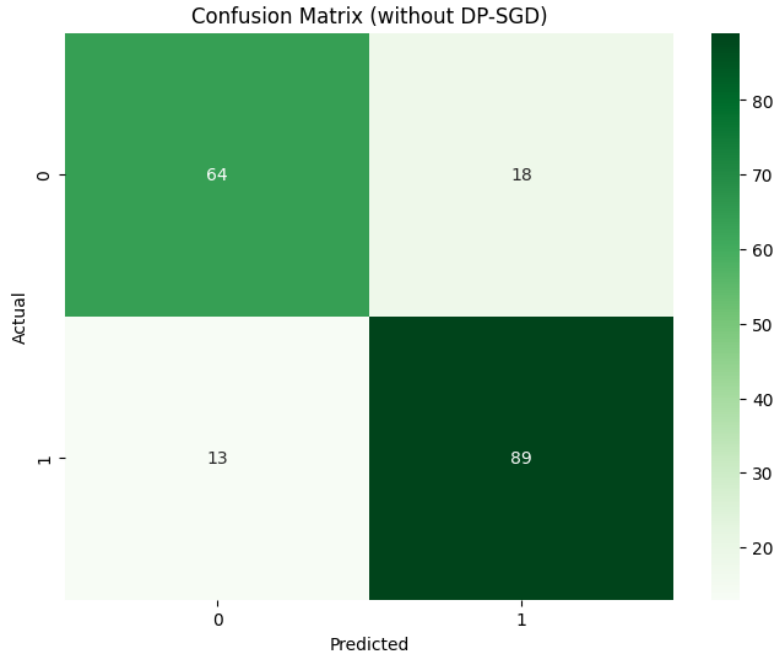
macro avg	0.58	0.57	0.55	184
weighted avg	0.59	0.55	0.55	184

6/6 [==============================] - 0s 3ms/step
ROC-AUC Score: 0.6220

6/6 [==============================] - 0s 3ms/step



Εικόνα 4: Πίνακας Σύγχυσης για Μοντέλο με DP-SGD για την περίπτωση Heart Disease Dataset.



Εικόνα 5: Πίνακας Σύγχυσης για Μοντέλο χωρίς DP-SGD για την περίπτωση Heart Disease Dataset.

False Positives (with DP-SGD): 22, False Negatives (with DP-SGD): 60
False Positives (no DP-SGD): 18, False Negatives (no DP-SGD): 13

Statistical Analysis of Results:
Mean Accuracy: 0.5706, Std Accuracy: 0.1073
Mean Loss: 2.0525, Std Loss: 5.4556

Delta (δ) used for experiments: 0.00005008
The delta value ensures theoretical guarantees within differential privacy context.

Relationship between privacy budget (ε) and noise multiplier:
Noise Multiplier: 0.5 => Epsilon (ε): 32.5415
Noise Multiplier: 1.1 => Epsilon (ε): 4.2341
Noise Multiplier: 2.0 => Epsilon (ε): 1.7177

Τα αποτελέσματα που προέκυψαν από την αξιολόγηση του μοντέλου με τη χρήση του διαφορικά ιδιωτικού στοχαστικού βαθμιδωτού καταβιβασμού (DP-SGD) προσφέρουν ουσιαστικές ενδείξεις για την επίδραση της ιδιωτικότητας στην απόδοση. Η συνολική ακρίβεια του μοντέλου που εκπαιδεύτηκε με DP-SGD έφτασε το 55%, με τον δείκτη ROC-AUC να διαμορφώνεται στο 0.6220, καταδεικνύοντας περιορισμένη διακριτική ικανότητα. Η αξιολόγηση μέσω του classification report ανέδειξε σημαντική ανισορροπία στην ικανότητα πρόβλεψης μεταξύ των δύο κλάσεων. Συγκεκριμένα, για την αρνητική κλάση (χωρίς νόσο), η ανάκληση ανήλθε στο 73%, ενώ για την θετική κλάση (με νόσο), ο δείκτης recall έπεσε στο 41%, αποκαλύπτοντας δυσκολία του μοντέλου να ανιχνεύσει σωστά περιστατικά ασθενών. Παράλληλα, η ακρίβεια (precision) στην κατηγορία των ασθενών ήταν 66%, στοιχείο που υποδηλώνει μια μερική ικανότητα εντοπισμού θετικών περιπτώσεων, αλλά συνοδεύτηκε από μεγάλο αριθμό ψευδώς αρνητικών προβλέψεων. Συνολικά, το μοντέλο παρήγαγε 60 false negatives και 22 false positives.
Η σύγκριση με το μοντέλο που εκπαιδεύτηκε χωρίς εφαρμογή διαφορικής ιδιωτικότητας καταδεικνύει σημαντικές αποκλίσεις στην απόδοση. Το μη ιδιωτικό μοντέλο κατέγραψε αισθητά καλύτερους δείκτες: η ακρίβεια αυξήθηκε σημαντικά, ο αριθμός των false negatives μειώθηκε σε 13 και των false positives σε 18. Επιπλέον, η γενική εικόνα από την confusion matrix χωρίς DP-SGD υποδηλώνει αυξημένη ικανότητα πρόβλεψης της θετικής κλάσης, με καλύτερη ισορροπία μεταξύ precision και recall. Τα αποτελέσματα αυτά επιβεβαιώνουν ότι η προσθήκη θορύβου, αν και απαραίτητη για την προστασία των προσωπικών δεδομένων, περιορίζει σημαντικά την ευαισθησία του μοντέλου, ιδιαίτερα στην πρόβλεψη περιστατικών που ανήκουν στη θετική κατηγορία.
Η στατιστική ανάλυση των αποτελεσμάτων από το σύνολο των πειραμάτων ανέδειξε τη γενικότερη επίδραση της παραμετροποίησης στη συμπεριφορά του DP-SGD. Η μέση τιμή ακρίβειας διαμορφώθηκε στο 57.06%, με τυπική απόκλιση 10.73%, γεγονός που αναδεικνύει υψηλή ευαισθησία της απόδοσης στις αλλαγές υπερπαραμέτρων. Παράλληλα, η μέση τιμή απώλειας ήταν 2.05 με εξαιρετικά υψηλή διακύμανση (τυπική απόκλιση 5.46), στοιχείο που αντικατοπτρίζει την αστάθεια που προκαλεί η προσθήκη θορύβου στη διαδικασία εκμάθησης. Ως προς τις θεωρητικές εγγυήσεις ιδιωτικότητας, η τιμή delta που χρησιμοποιήθηκε στα πειράματα ήταν 0.00005008, επαρκώς μικρή ώστε να θεωρείται συμβατή με τις απαιτήσεις της διαφορικής ιδιωτικότητας. Επιπλέον, η σχέση μεταξύ privacy budget (ε) και noise multiplier ανέδειξε το γνωστό trade-off μεταξύ απόδοσης και προστασίας: όσο αυξάνεται η τιμή του noise multiplier, τόσο μειώνεται το ε, βελτιώνοντας

την ιδιωτικότητα. Ενδεικτικά, για noise=0.5, το ε ανήλθε σε 32.54 μονάδες, παρέχοντας περιορισμένη προστασία, ενώ για noise=2.0 το ε μειώθηκε σε 1.72, προσφέροντας πολύ ισχυρή προστασία αλλά με τίμημα τη μείωση της απόδοσης.
Τα ευρήματα επιβεβαιώνουν το αναμενόμενο δίλημμα μεταξύ ιδιωτικότητας και ακρίβειας. Το DP-SGD, ενώ προσφέρει ισχυρές θεωρητικές εγγυήσεις για την προστασία των δεδομένων, απαιτεί προσεκτική και ισορροπημένη ρύθμιση των υπερπαραμέτρων προκειμένου να διατηρηθεί αποδεκτό επίπεδο λειτουργικότητας. Η επίτευξη αποτελεσματικής απόδοσης σε εφαρμογές με ευαίσθητα δεδομένα προϋποθέτει την επιλογή συνδυασμών παραμέτρων που ελαχιστοποιούν την απώλεια προβλεπτικής ισχύος, χωρίς να θυσιάζεται η προστασία της ιδιωτικότητας.

7.2 Υλοποίηση του DP-SGD για το Cardiovascular Disease Dataset

Ο κώδικας έχει ως στόχο την εκπαίδευση ενός νευρωνικού δικτύου χρησιμοποιώντας το Cardiovascular Disease Dataset με τη μέθοδο DP-SGD. Η διαφορική ιδιωτικότητα (DP) εξασφαλίζει ότι τα δεδομένα εκπαίδευσης προστατεύονται από την αποκάλυψη ευαίσθητων πληροφοριών, προσθέτοντας ελεγχόμενο θόρυβο στην εκπαίδευση του μοντέλου.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import tensorflow as tf
import tensorflow_privacy
import matplotlib.pyplot as plt
from tensorflow_privacy.privacy.optimizers.dp_optimizer_keras import DPKerasSGDOptimizer
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy_lib

# 1. Φόρτωση του Cardiovascular Dataset
dataset_path = "cardio_train.csv"
df_cardio = pd.read_csv(dataset_path, sep=";")  # Το dataset χρησιμοποιεί ";" ως διαχωριστικό

# 2. Καθαρισμός των δεδομένων
df_cardio.drop(columns=["id"], inplace=True)  # Αφαίρεση του ID γιατί δεν έχει νόημα στην εκπαίδευση
df_cardio = df_cardio[df_cardio["height"] > 100]  # Αφαιρούμε ανωμαλίες στα δεδομένα
df_cardio = df_cardio[df_cardio["weight"] > 30]  # Αφαιρούμε ακραίες τιμές

# 3. Διαχωρισμός χαρακτηριστικών (X) και ετικετών (y)
X_cardio = df_cardio.drop(columns=['cardio'])  # Χαρακτηριστικά
y_cardio = df_cardio['cardio']  # Στόχος (0 = χωρίς νόσο, 1 = με νόσο)

# 4. Κανονικοποίηση των χαρακτηριστικών
scaler = StandardScaler()
X_cardio_scaled = scaler.fit_transform(X_cardio)
```

```
# 5. Διαχωρισμός σε train και test set
X_train_cardio, X_test_cardio, y_train_cardio, y_test_cardio = train_test_split(
    X_cardio_scaled, y_cardio, test_size=0.2, random_state=42, stratify=y_cardio
)

# Υπερπαράμετροι
learning_rate = 0.01
noise_multiplier = 1.1
l2_norm_clip = 1.0
epochs = 50
batch_sizes = [32, 64]
results = []
for batch_size in batch_sizes:
    print(f"\n--- Εκπαίδευση με batch size: {batch_size} ---")

    # Δημιουργία νέου μοντέλου
    model_cardio = tf.keras.Sequential([
        tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train_cardio.shape[1],)),
        tf.keras.layers.Dense(8, activation='relu'),
        tf.keras.layers.Dense(1, activation='sigmoid')
    ])

    dp_optimizer_cardio = DPKerasSGDOptimizer(
        l2_norm_clip=l2_norm_clip,
        noise_multiplier=noise_multiplier,
        num_microbatches=1,
        learning_rate=learning_rate
    )

    model_cardio.compile(optimizer=dp_optimizer_cardio,        loss='binary_crossentropy',
metrics=['accuracy'])

    # Εκπαίδευση
    history_cardio = model_cardio.fit(
        X_train_cardio, y_train_cardio,
        epochs=epochs,
        batch_size=batch_size,
        validation_data=(X_test_cardio, y_test_cardio),
        verbose=1
    )

    # Αξιολόγηση
    test_loss_cardio, test_acc_cardio = model_cardio.evaluate(X_test_cardio, y_test_cardio,
verbose=0)
    print(f"Test Accuracy: {test_acc_cardio:.4f}")

    # Υπολογισμός ε
    eps, _ = compute_dp_sgd_privacy_lib.compute_dp_sgd_privacy(
```

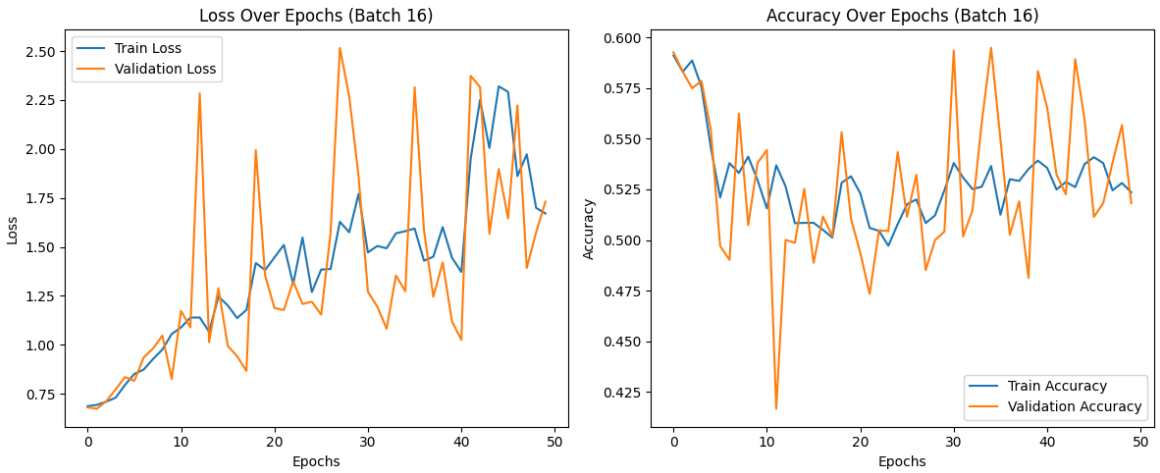
```
    n=len(X_train_cardio),
    batch_size=batch_size,
    noise_multiplier=noise_multiplier,
    epochs=epochs,
    delta=1e-5
)
print(f"Estimated ε for batch size {batch_size}: {eps:.2f}")

# Σχεδίαση διαγραμμάτων
plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)
plt.plot(history_cardio.history['loss'], label='Train Loss')
plt.plot(history_cardio.history['val_loss'], label='Validation Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.title(f'Loss Over Epochs (Batch {batch_size})')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(history_cardio.history['accuracy'], label='Train Accuracy')
plt.plot(history_cardio.history['val_accuracy'], label='Validation Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.title(f'Accuracy Over Epochs (Batch {batch_size})')
plt.legend()

plt.tight_layout()
plt.show()

# 8. Αποθήκευση αποτελεσμάτων
results_df = pd.DataFrame(results)
results_df.to_csv("results_df_Hyperparameter_Tuning_Cardiovascular_Dataset.csv",
index=False)
print("Αποθηκεύτηκαν τα αποτελέσματα στο αρχείο CSV.")
```

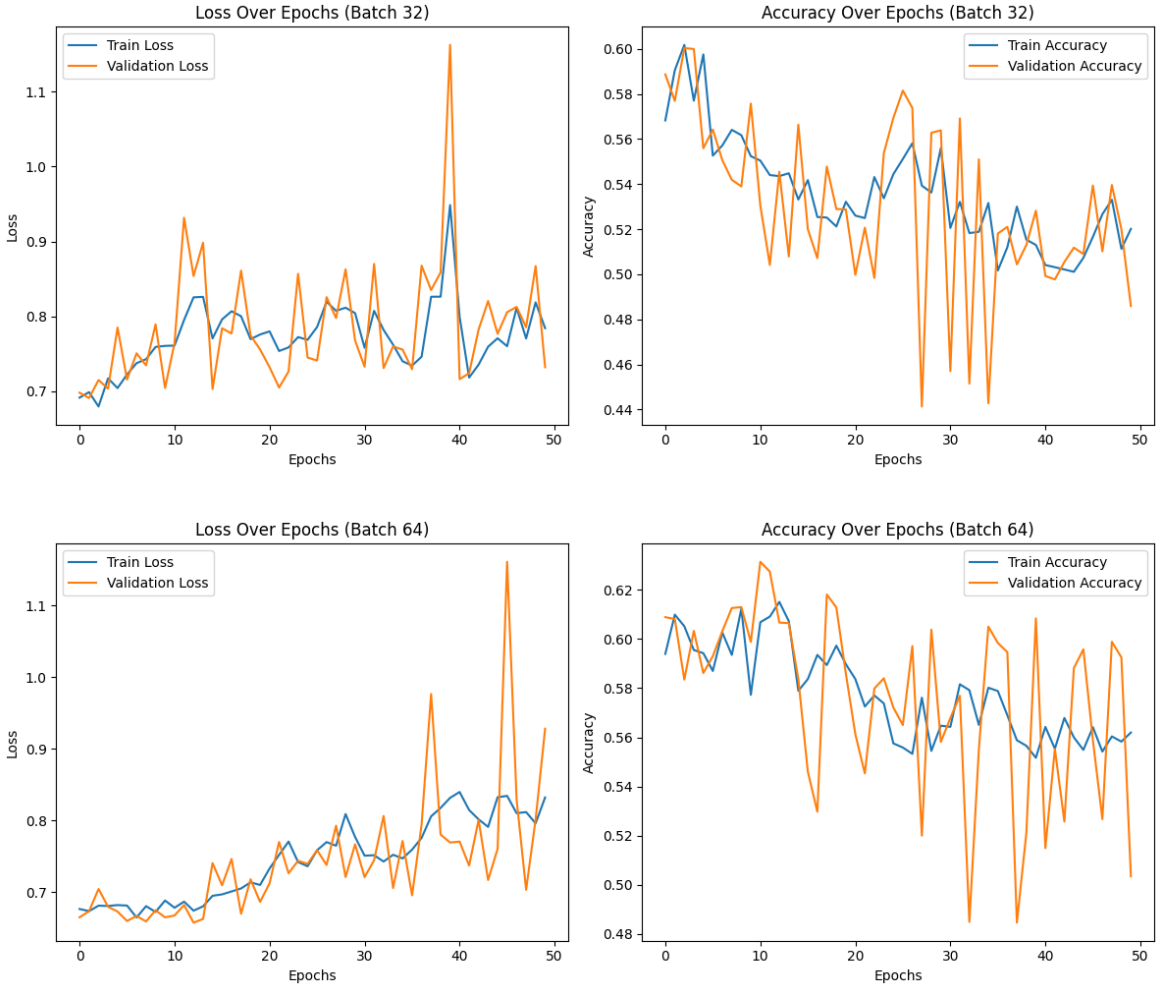
Αρχικά, το dataset φορτώνεται από το αρχείο cardio_train.csv, το οποίο χρησιμοποιεί ως διαχωριστικό το χαρακτήρα ;. Πραγματοποιείται καθαρισμός δεδομένων, αφαιρώντας τη στήλη id που δεν προσφέρει πληροφόρηση για την πρόβλεψη, καθώς και μη ρεαλιστικές τιμές ύψους και βάρους που ενδέχεται να είναι ανωμαλίες στο σύνολο δεδομένων. Μετά τον καθαρισμό, τα χαρακτηριστικά (X) απομονώνονται από την ετικέτα στόχο (y), που αντιπροσωπεύει την ύπαρξη ή μη καρδιαγγειακής νόσου (0 = υγιής, 1 = ασθενής). Ακολουθεί κανονικοποίηση των χαρακτηριστικών μέσω StandardScaler, ώστε να αποκτήσουν κοινή κλίμακα, κάτι που συμβάλλει στη σταθερότητα της εκπαίδευσης του νευρωνικού δικτύου.

Στη συνέχεια, το dataset διαχωρίζεται σε σύνολο εκπαίδευσης (80%) και δοκιμών (20%), διατηρώντας την αναλογία των κλάσεων μέσω της παραμέτρου stratify=y_cardio. Για την εκπαίδευση χρησιμοποιούνται οι εξής υπερπαράμετροι: learning rate 0.01, noise multiplier 1.1, l2 norm clip 1.0, epochs 50, και δύο επιλογές για batch size (32 και 64). Το μοντέλο

αποτελείται από τρία πλήρως συνδεδεμένα επίπεδα: το πρώτο με 16 νευρώνες και ενεργοποίηση ReLU, το δεύτερο με 8 νευρώνες και επίσης ReLU, και το τελικό επίπεδο με 1 νευρώνα και sigmoid ενεργοποίηση, κατάλληλο για δυαδική ταξινόμηση. Η εκπαίδευση πραγματοποιείται με χρήση του DPKerasSGDOptimizer, που εφαρμόζει τον μηχανισμό διαφορικής ιδιωτικότητας DP-SGD. Στην παρούσα υλοποίηση, έχει οριστεί num_microbatches=1, μια τιμή που περιορίζει τη δυνατότητα ελέγχου της ευαισθησίας των gradients και επηρεάζει τον τρόπο προσθήκης θορύβου. Αν και επιτρέπει την εφαρμογή διαφορικής ιδιωτικότητας, είναι γνωστό από τη βιβλιογραφία ότι χαμηλές τιμές microbatches μπορεί να μειώσουν την ακρίβεια και τη σταθερότητα του μοντέλου (Abadi et al., 2016· Bu et al., 2020).

Κατά την εκπαίδευση, καταγράφονται τα ιστορικά τιμών loss και accuracy τόσο για το train όσο και για το validation set. Στο τέλος, αξιολογείται το τελικό μοντέλο στο test set και υπολογίζεται το privacy budget (ε), το οποίο είναι περίπου 3.87 για batch size 32 και 2.28 για batch size 64. Από τα παραγόμενα διαγράμματα, προκύπτουν χρήσιμες παρατηρήσεις. Στα γραφήματα της απώλειας, η τιμή loss κυμαίνεται μεταξύ 0.7 και 0.9 κατά τις πρώτες 30–35 εποχές, και εμφανίζει αιχμές προς το τέλος, ιδιαίτερα για batch size 64, γεγονός που υποδηλώνει πιθανή αστάθεια. Στα γραφήματα ακρίβειας, η γραμμή validation accuracy παρουσιάζει έντονες διακυμάνσεις για κάθε epoch, με τάσεις πτώσης καθώς προχωρά η εκπαίδευση. Αν και δεν παρατηρείται σαφής υπερπροσαρμογή, η αυξημένη διακύμανση υποδεικνύει ότι το μοντέλο επηρεάζεται από τον θόρυβο του DP-SGD, ειδικά στις τελευταίες εποχές. Η διαφορά μεταξύ batch sizes είναι αξιοσημείωτη: το batch size 64 επιτρέπει μικρότερο privacy budget (μικρότερο ε ⇒ μεγαλύτερη προστασία), αλλά φαίνεται να επηρεάζει αρνητικά τη σταθερότητα του validation accuracy, όπως φαίνεται από την πιο έντονη μεταβλητότητά του. Τα αποτελέσματα αποθηκεύονται σε αρχείο CSV για περαιτέρω ανάλυση, επιτρέποντας τη σύγκριση διαφορετικών ρυθμίσεων υπερπαραμέτρων. Παρά το γεγονός ότι η τελική ακρίβεια δεν είναι ιδιαίτερα υψηλή (περίπου 55–60%), το πείραμα υποδεικνύει πως με κατάλληλο συνδυασμό υπερπαραμέτρων, μπορεί να επιτευχθεί καλή ισορροπία μεταξύ απόδοσης και προστασίας της ιδιωτικότητας. Σε μελλοντική εργασία, θα ήταν χρήσιμο να διερευνηθούν περαιτέρω τιμές για num_microbatches και να εξεταστεί η χρήση άλλων μηχανισμών, όπως το Rényi DP ή privacy amplification by subsampling, για την ενίσχυση της αποτελεσματικότητας.





Εικόνα 6: Εκπαίδευση και Επικύρωση Μοντέλου στο Cardiovascular Dataset με DP-SGD.

Το διάγραμμα που αντιστοιχεί σε εκπαίδευση του διαφορικά ιδιωτικού μοντέλου με batch size 16 στο Cardiovascular dataset αποτυπώνει μια πορεία με ενδείξεις αστάθειας τόσο στην απώλεια όσο και στην ακρίβεια.

Στο αριστερό υποδιάγραμμα, η συνάρτηση απώλειας (loss) παρουσιάζει έντονες διακυμάνσεις, ιδίως στη γραμμή επικύρωσης. Παρά το γεγονός ότι η αρχική απώλεια μειώνεται ελαφρώς στις πρώτες εποχές, στη συνέχεια εμφανίζονται απότομες αυξομειώσεις, τόσο στο σύνολο εκπαίδευσης όσο και στο σύνολο επικύρωσης. Οι αιχμές της validation loss υποδηλώνουν ότι το μοντέλο δεν καταφέρνει να σταθεροποιηθεί και πιθανότατα επηρεάζεται έντονα από τον προστιθέμενο θόρυβο του μηχανισμού DP-SGD. Η αστάθεια αυτή ενδέχεται να προκύπτει από τον πολύ μικρό αριθμό δειγμάτων ανά ενημέρωση (λόγω του μικρού batch size), σε συνδυασμό με το σταθερό clipping norm και noise multiplier. Στο δεξί υποδιάγραμμα, η ακρίβεια κυμαίνεται καθ' όλη τη διάρκεια των εποχών μεταξύ 50% και 60%, χωρίς να παρουσιάζει σαφή ανοδική πορεία ή σημάδια σύγκλισης. Η καμπύλη της ακρίβειας στο validation set είναι ακόμη πιο ασταθής, με συχνές και απότομες εναλλαγές, που φτάνουν στιγμιαία έως και κάτω από το 45%. Οι αποκλίσεις μεταξύ εκπαίδευσης και επικύρωσης υποδεικνύουν ότι το μοντέλο δεν έχει μάθει σταθερά μοτίβα που να μπορούν να γενικευτούν, πιθανώς λόγω του αυξημένου επιπέδου θορύβου σε σχέση με το μέγεθος παρτίδας.

Η συγκεκριμένη συμπεριφορά φανερώνει ότι, τουλάχιστον για αυτό το dataset και με τις συγκεκριμένες υπερπαραμέτρους (learning rate 0.01, noise multiplier 1.1, clip 1.0), η χρήση batch size 16 δεν προσφέρει επαρκή σταθερότητα στην εκπαίδευση. Ενδεχομένως, μεγαλύτερες παρτίδες ή η τροποποίηση του ρυθμού εκμάθησης και του επιπέδου θορύβου θα μπορούσαν να προσφέρουν πιο συνεπή αποτελέσματα. Συνολικά, το διάγραμμα ενισχύει την εικόνα ενός μοντέλου που δυσκολεύεται να συγκλίνει υπό τις συγκεκριμένες συνθήκες προστασίας ιδιωτικότητας.

Το διάγραμμα που απεικονίζει την πορεία εκπαίδευσης για batch size 32 στο Cardiovascular dataset καταδεικνύει ελαφρώς βελτιωμένη σταθερότητα σε σύγκριση με το μικρότερο batch size (16), ωστόσο παραμένουν ορισμένα σημάδια στοχαστικότητας και αστάθειας, κυρίως στην ακρίβεια επικύρωσης. Στο αριστερό διάγραμμα (Loss Over Epochs), παρατηρείται ότι τόσο η απώλεια εκπαίδευσης όσο και η απώλεια επικύρωσης κυμαίνονται μέσα σε παρόμοια εύρη, με μικρές αποκλίσεις μεταξύ τους. Παρόλο που δεν καταγράφεται ξεκάθαρη πτωτική τάση, οι τιμές loss διατηρούνται σχετικά κοντά (γύρω στο 0.75–0.9), γεγονός που υποδηλώνει σταθεροποίηση σε κάποιο σημείο, χωρίς όμως ουσιαστική βελτίωση. Οι αιχμές που εμφανίζονται σε ορισμένα σημεία του validation loss δείχνουν ότι το μοντέλο εξακολουθεί να επηρεάζεται από το θόρυβο, αν και σε μικρότερο βαθμό από την περίπτωση του batch size 16.

Στο δεξί διάγραμμα (Accuracy Over Epochs), η ακρίβεια εκπαίδευσης ξεκινά από υψηλό επίπεδο (άνω του 58%) αλλά παρουσιάζει σταδιακή πτώση και ασταθή συμπεριφορά, χωρίς σαφή τάση σύγκλισης. Η γραμμή validation accuracy εμφανίζει συχνές διακυμάνσεις, με στιγμιαίες πτώσεις κάτω από το 45%, γεγονός που υποδηλώνει αδυναμία του μοντέλου να γενικεύσει αξιόπιστα στα δεδομένα επικύρωσης. Παρ' όλα αυτά, η γενική εικόνα δείχνει ελαφρώς μειωμένη στοχαστικότητα σε σύγκριση με το batch size 16, κάτι που υποστηρίζει τη θεωρητική υπόθεση ότι ένα μέτριο μέγεθος παρτίδας μπορεί να λειτουργήσει σταθεροποιητικά στο πλαίσιο εκπαίδευσης με DP-SGD. Η χρήση batch size 32 προσφέρει ένα μικρό πλεονέκτημα ως προς τη σταθερότητα της εκπαίδευσης, χωρίς όμως να επιτυγχάνει ουσιαστική βελτίωση στην απόδοση. Η συμπεριφορά του μοντέλου υποδηλώνει ότι η συγκεκριμένη παραμετροποίηση εξακολουθεί να επηρεάζεται από το θόρυβο που επιβάλλει η διαφορική ιδιωτικότητα, ενώ ενδεχομένως θα μπορούσε να βελτιωθεί με τροποποίηση του learning rate ή του clipping norm.

Το διάγραμμα για batch size 64 στο Cardiovascular dataset δείχνει μια εκπαιδευτική πορεία που χαρακτηρίζεται από ελαφρώς καλύτερη συνοχή στην απώλεια σε σχέση με τα μικρότερα batch sizes, αλλά και έντονη αστάθεια στην ακρίβεια επικύρωσης, ιδιαίτερα στις τελευταίες εποχές. Στο αριστερό υποδιάγραμμα (Loss Over Epochs), η καμπύλη της απώλειας εκπαίδευσης διατηρεί μια σχετικά σταθερή ανοδική τάση, ενώ η validation loss, αν και ξεκινά σε χαμηλό επίπεδο, παρουσιάζει σταδιακή επιδείνωση μετά τη μέση της εκπαίδευσης. Οι αιχμές που εμφανίζονται κυρίως μετά την 30ή εποχή μαρτυρούν ότι το μοντέλο αρχίζει να χάνει τη σταθερότητα του ως προς τη γενίκευση, πιθανόν εξαιτίας της συσσώρευσης θορύβου στο DP-SGD και της περιορισμένης ικανότητας προσαρμογής του learning process.

Το δεξί διάγραμμα (Accuracy Over Epochs) ενισχύει αυτή την παρατήρηση. Η ακρίβεια εκπαίδευσης παραμένει σε ένα σχετικά σταθερό επίπεδο, μεταξύ 55% και 60%, χωρίς σημαντική βελτίωση, ενώ η καμπύλη της validation accuracy παρουσιάζει ισχυρές διακυμάνσεις, με απότομες μεταβολές σε κάθε εποχή. Αυτή η συμπεριφορά, η οποία χαρακτηρίζεται από ασυνέπεια μεταξύ των εποχών, φανερώνει ότι το μοντέλο δεν καταφέρνει να γενικεύσει ικανοποιητικά στα δεδομένα ελέγχου. Οι διακυμάνσεις στην ακρίβεια επικύρωσης γίνονται εντονότερες όσο προχωρά η εκπαίδευση, κάτι που πιθανώς

συνδέεται με την υπερβολική εξομάλυνση των gradients λόγω του μεγάλου batch size, σε συνδυασμό με την προσθήκη θορύβου. Γενικά, το batch size 64 φαίνεται να μειώνει σε έναν βαθμό τη στοχαστικότητα της loss function, αλλά το κάνει εις βάρος της ικανότητας γενίκευσης, με το validation accuracy να γίνεται ιδιαίτερα ασταθές και απρόβλεπτο. Το αποτέλεσμα αυτό καταδεικνύει ότι, στο πλαίσιο της διαφορικής ιδιωτικότητας, η χρήση πολύ μεγάλου batch size δεν εγγυάται βελτιωμένη απόδοση, ειδικά όταν δεν συνοδεύεται από αντίστοιχη προσαρμογή σε άλλες υπερπαραμέτρους όπως το learning rate ή το clipping norm.

Η συνολική αξιολόγηση των αποτελεσμάτων για τα τρία διαφορετικά μεγέθη παρτίδας (batch size 16, 32 και 64) στο Cardiovascular dataset με χρήση του DP-SGD αποκαλύπτει σημαντικές διαφορές στη σταθερότητα, την ακρίβεια και τη συμπεριφορά του μοντέλου κατά την εκπαίδευση. Στην περίπτωση του batch size 16, το μοντέλο παρουσίασε έντονη αστάθεια τόσο στην απώλεια όσο και στην ακρίβεια. Οι καμπύλες loss, και κυρίως εκείνες της επικύρωσης, εμφάνισαν συχνές αιχμές και διακυμάνσεις, ενώ η ακρίβεια κυμάνθηκε γύρω από το 50–55% χωρίς σαφή ανοδική τάση ή σύγκλιση. Η υπερβολική στοχαστικότητα που παρατηρήθηκε πιθανόν οφείλεται στον συνδυασμό μικρού μεγέθους παρτίδας με θόρυβο (noise multiplier = 1.1), ο οποίος έχει μεγαλύτερη σχετική επίδραση όταν εφαρμόζεται σε λιγότερα παραδείγματα ανά ενημέρωση. Ως αποτέλεσμα, το μοντέλο δεν κατάφερε να σταθεροποιηθεί ούτε να γενικεύσει επαρκώς.

Με την αύξηση του batch size σε 32, καταγράφηκε μικρή βελτίωση στη σταθερότητα της εκπαίδευσης. Οι καμπύλες απώλειας παρουσίασαν λιγότερες αιχμές και μικρότερες αποκλίσεις μεταξύ εκπαίδευσης και επικύρωσης, διατηρώντας τις τιμές loss σε πιο συμπαγές εύρος. Ωστόσο, η ακρίβεια δεν εμφάνισε ουσιαστική πρόοδο. Αν και ξεκίνησε από υψηλότερα επίπεδα σε σύγκριση με το batch size 16, η τάση της ήταν καθοδική και ασυνεπής, με απότομες μεταβολές στην καμπύλη της επικύρωσης. Αυτό δείχνει ότι, παρότι η μεγαλύτερη παρτίδα βοήθησε στην ομαλοποίηση του θορύβου, το μοντέλο εξακολουθούσε να δυσκολεύεται να μάθει και να διατηρήσει σταθερή απόδοση.

Η περίπτωση του batch size 64 παρουσίασε ακόμη πιο σταθερή συμπεριφορά ως προς την απώλεια εκπαίδευσης, όμως η ακρίβεια επικύρωσης εμφάνισε τη μεγαλύτερη αστάθεια από τις τρεις περιπτώσεις. Αν και το train loss κυμάνθηκε σε χαμηλά επίπεδα και με σχετική συνέπεια, η απόδοση στο validation set χαρακτηρίστηκε από συνεχείς και απότομες διακυμάνσεις στην ακρίβεια. Οι καμπύλες δεν συγκλίνουν ούτε σταθεροποιούνται, υποδηλώνοντας ότι το μοντέλο υπερπροσαρμόζεται ή δεν μαθαίνει ουσιαστικά χρήσιμα χαρακτηριστικά λόγω της υπερβολικής εξομάλυνσης των gradients. Αυτό πιθανώς οφείλεται στην αδυναμία του μεγάλου batch size να αποτυπώσει επαρκώς τη μεταβλητότητα των δεδομένων σε κάθε ενημέρωση, ιδίως όταν εφαρμόζεται ισχυρό clipping και θόρυβος. Κανένα από τα τρία μεγέθη παρτίδας δεν οδήγησε σε πλήρως ικανοποιητική εκπαίδευση του μοντέλου υπό τους συγκεκριμένους υπερπαραμετρικούς συνδυασμούς. Το batch size 32 προσέφερε τη βέλτιστη ισορροπία ως προς τη σταθερότητα της απώλειας και την ανεκτή συμπεριφορά της ακρίβειας, ενώ τα μεγέθη 16 και 64 εμφάνισαν είτε υπερβολική στοχαστικότητα είτε υπερβολική εξομάλυνση. Τα αποτελέσματα τονίζουν την ανάγκη περαιτέρω ρύθμισης των παραμέτρων (ιδίως του learning rate και του noise multiplier), καθώς και τη σημασία εύρεσης ενός ενδιάμεσου μεγέθους παρτίδας για τη διατήρηση της ισορροπίας μεταξύ απόδοσης και προστασίας ιδιωτικότητας στο πλαίσιο της DP-SGD.

7.2.1 Διαδικασία ρύθμισης των υπερπαραμέτρων

Αυτός ο κώδικας στοχεύει στην βελτιστοποίηση των υπερπαραμέτρων για την εκπαίδευση ενός νευρωνικού δικτύου με διαφορική ιδιωτικότητα (DP-SGD) στο Cardiovascular Disease Dataset. Ο στόχος είναι να πειραματιστούμε με διάφορους συνδυασμούς υπερπαραμέτρων και να αξιολογήσουμε ποιοι συνδυασμοί προσφέρουν την καλύτερη απόδοση όσον αφορά την ακρίβεια (test accuracy) και την απώλεια (test loss).

```
import itertools
import pandas as pd
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import tensorflow as tf
import tensorflow_privacy

#  1. Φόρτωση του Cardiovascular Dataset
dataset_path = "cardio_train.csv"
df_cardio = pd.read_csv(dataset_path, sep=";")  # To dataset χρησιμοποιεί ";" ως διαχωριστικό

#2. Καθαρισμός των δεδομένων
df_cardio.drop(columns=["id"], inplace=True)  # Αφαίρεση του ID γιατί δεν έχει νόημα στην εκπαίδευση
df_cardio = df_cardio[df_cardio["height"] > 100]  # Αφαιρούμε ανωμαλίες στα δεδομένα
df_cardio = df_cardio[df_cardio["weight"] > 30]  # Αφαιρούμε ακραίες τιμές

#  3. Διαχωρισμός χαρακτηριστικών (X) και ετικετών (y)
X_cardio = df_cardio.drop(columns=['cardio'])  # Χαρακτηριστικά
y_cardio = df_cardio['cardio']  # Στόχος (0 = χωρίς νόσο, 1 = με νόσο)

#  4. Κανονικοποίηση των χαρακτηριστικών
scaler = StandardScaler()
X_cardio_scaled = scaler.fit_transform(X_cardio)

#  5. Διαχωρισμός σε train και test set
X_train_cardio, X_test_cardio, y_train_cardio, y_test_cardio = train_test_split(
    X_cardio_scaled, y_cardio, test_size=0.2, random_state=42, stratify=y_cardio
)

print(f'Dataset loaded successfully! Train size: {X_train_cardio.shape}, Test size: {X_test_cardio.shape}")

# Ορισμός των διαφορετικών τιμών υπερπαραμέτρων για πειραματισμό
learning_rates = [0.001, 0.01, 0.1]
noise_multipliers = [0.5, 1.1, 2.0]
l2_norm_clips = [0.5, 1.0, 2.0]
batch_sizes = [16, 32, 64]
```

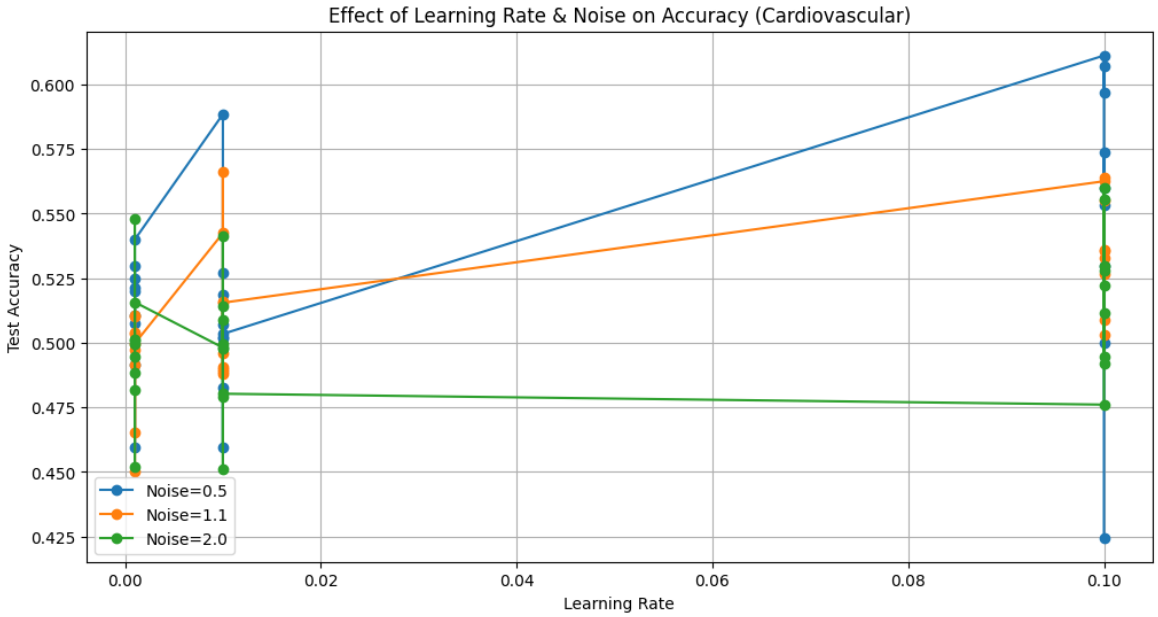
```
epochs = 30  # Σταθερό για όλες τις δοκιμές

# Αποθήκευση των αποτελεσμάτων
results_cardio = []

# Δοκιμή όλων των συνδυασμών υπερπαραμέτρων
for lr, noise, clip, batch in itertools.product(learning_rates, noise_multipliers, l2_norm_clips, batch_sizes):
    print(f"Training model with LR={lr}, Noise={noise}, Clip={clip}, Batch={batch}")

    # Προσαρμογή batch size
    batch_size = (len(X_train_cardio) // batch) * batch
    if batch_size == 0:
        batch_size = 1

    # Δημιουργία του μοντέλου
    model_cardio = tf.keras.Sequential([
        tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train_cardio.shape[1],)),
        tf.keras.layers.Dense(8, activation='relu'),
        tf.keras.layers.Dense(1, activation='sigmoid')  # Δυαδική ταξινόμηση
    ])

    # Χρήση του DP-SGD Optimizer
    dp_optimizer_cardio = tensorflow_privacy.DPKerasSGDOptimizer(
        l2_norm_clip=clip,
        noise_multiplier=noise,
        num_microbatches=1,
        learning_rate=lr
    )

    # Συναρτήσεις απώλειας και μέτρησης απόδοσης
    model_cardio.compile(optimizer=dp_optimizer_cardio, loss='binary_crossentropy', metrics=['accuracy'])

    # Εκπαίδευση του μοντέλου με DP-SGD
    history_cardio = model_cardio.fit(
        X_train_cardio, y_train_cardio, epochs=epochs, batch_size=batch_size, validation_data=(X_test_cardio, y_test_cardio), verbose=0
    )

    # Αξιολόγηση του μοντέλου
    test_loss_cardio, test_acc_cardio = model_cardio.evaluate(X_test_cardio, y_test_cardio, verbose=0)
    print(f"Test Accuracy: {test_acc_cardio:.4f} | Test Loss: {test_loss_cardio:.4f}")

    # Αποθήκευση αποτελεσμάτων
    results_cardio.append({
        "learning_rate": lr,
```

```
        "noise_multiplier": noise,
        "l2_norm_clip": clip,
        "batch_size": batch,
        "test_accuracy": test_acc_cardio,
        "test_loss": test_loss_cardio
    })

# Μετατροπή των αποτελεσμάτων σε DataFrame
results_df_cardio = pd.DataFrame(results_cardio)

# Εμφάνιση αποτελεσμάτων (Χρησιμοποίησε print ή display σε Jupyter Notebook)
print("Hyperparameter Tuning Results - Cardiovascular Dataset:")
print(results_df_cardio.to_string())  # Αν το DataFrame είναι μεγάλο, μπορείς να το εμφανίσεις με print(results_df_cardio.head())

# Διαγράμματα: Επίδραση των Υπερπαραμέτρων στην Ακρίβεια
plt.figure(figsize=(12, 6))
for noise in results_df_cardio["noise_multiplier"].unique():  # Χρήση unique() για μοναδικές τιμές
    subset = results_df_cardio[results_df_cardio["noise_multiplier"] == noise]
    plt.plot(subset["learning_rate"], subset["test_accuracy"], marker='o', linestyle='-', label=f"Noise={noise}")

plt.xlabel("Learning Rate")
plt.ylabel("Test Accuracy")
plt.title("Effect of Learning Rate & Noise on Accuracy (Cardiovascular)")
plt.legend()
plt.grid(True)  # Προσθήκη πλέγματος για καλύτερη οπτικοποίηση
plt.show()

# Διαγράμματα: Επίδραση των Υπερπαραμέτρων στην Απώλεια
plt.figure(figsize=(12, 6))
for clip in results_df_cardio["l2_norm_clip"].unique():
    subset = results_df_cardio[results_df_cardio["l2_norm_clip"] == clip]
    plt.plot(subset["batch_size"], subset["test_loss"], marker='s', linestyle='-', label=f"Clip={clip}")

plt.xlabel("Batch Size")
plt.ylabel("Test Loss")
plt.title("Effect of Clipping & Batch Size on Loss (Cardiovascular)")
plt.legend()
plt.grid(True)
plt.show()

# Υπολογισμός επιπλέον μετρικών
predictions_cardio = (model_cardio.predict(X_test_cardio) > 0.5).astype("int32")
report = classification_report(y_test_cardio, predictions_cardio, output_dict=True)
roc_auc = roc_auc_score(y_test_cardio, model_cardio.predict(X_test_cardio))
```

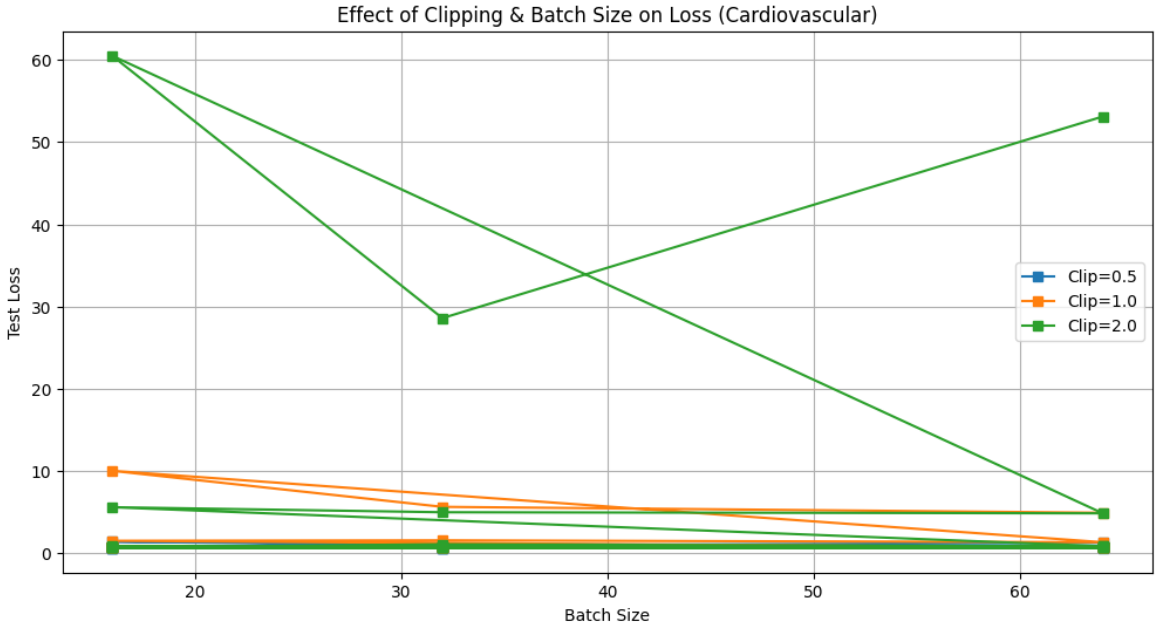
```
# Εκτύπωση classification report
print("\nClassification Report (Cardiovascular dataset with DP-SGD):")
print(classification_report(y_test_cardio, predictions_cardio))
print(f"ROC-AUC Score: {roc_auc:.4f}")

# Δημιουργία και εμφάνιση confusion matrix
cm = confusion_matrix(y_test_cardio, predictions_cardio)
plt.figure(figsize=(8,6))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.title("Confusion Matrix (with DP-SGD)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

# Σύγκριση με μοντέλο χωρίς DP-SGD
model_no_dp = tf.keras.Sequential([
    tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train_cardio.shape[1],)),
    tf.keras.layers.Dense(8, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])

model_no_dp.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model_no_dp.fit(X_train_cardio, y_train_cardio, epochs=epochs, batch_size=32,
verbose=0)

# Αξιολόγηση μοντέλου χωρίς DP-SGD
predictions_no_dp = (model_no_dp.predict(X_test_cardio) > 0.5).astype("int32")
cm_no_dp = confusion_matrix(y_test_cardio, predictions_no_dp)

plt.figure(figsize=(8,6))
sns.heatmap(cm_no_dp, annot=True, fmt="d", cmap="Greens")
plt.title("Confusion Matrix (without DP-SGD)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

# Υπολογισμός false positives και false negatives
fp_dp, fn_dp = cm[0][1], cm[1][0]
fp_no_dp, fn_no_dp = cm_no_dp[0][1], cm_no_dp[1][0]

print(f"False Positives (with DP-SGD): {fp_dp}, False Negatives (with DP-SGD):
{fn_dp}")
print(f"False Positives (no DP-SGD): {fp_no_dp}, False Negatives (no DP-SGD):
{fn_no_dp}")

# Στατιστική ανάλυση αποτελεσμάτων για ακρίβεια και απώλεια
mean_accuracy = results_df_cardio['test_accuracy'].mean()
```

```
std_accuracy = results_df_cardio['test_accuracy'].std()
mean_loss = results_df_cardio['test_loss'].mean()
std_loss = results_df_cardio['test_loss'].std()

print("\nStatistical Analysis of Results:")
print(f'Mean Accuracy: {mean_accuracy:.4f}, Std Accuracy: {std_accuracy:.4f}')
print(f'Mean Loss: {mean_loss:.4f}, Std Loss: {std_loss:.4f}')
# Υπολογισμός και εμφάνιση του delta και της σχέσης με noise multiplier
num_samples = X_train_cardio.shape[0]
delta = 1 / (num_samples * np.sqrt(num_samples))
print(f"\nDelta (δ) used for experiments: {delta:.8f}")
print("The delta value ensures theoretical guarantees within DP context.")

# Εμφάνιση σχέσης privacy budget (ε) και noise multiplier
print("\nRelationship between privacy budget (ε) and noise multiplier:")
for noise in noise_multipliers:
    epsilon, _ = tensorflow_privacy.compute_dp_sgd_privacy(n=num_samples,
                                        batch_size=batch_sizes[0],
                                        noise_multiplier=noise,
                                        epochs=epochs,
                                        delta=delta)
    print(f'Noise Multiplier: {noise} => Epsilon (ε): {epsilon:.4f}')
```

Ο κώδικας αρχικά φορτώνει το σύνολο δεδομένων Cardiovascular Disease από αρχείο CSV και προχωρά σε βασική προεπεξεργασία. Αφαιρείται η στήλη "id" που δεν προσφέρει πληροφορία στη μάθηση, ενώ απομακρύνονται και παρατηρήσεις με μη ρεαλιστικές τιμές ύψους και βάρους ώστε να διασφαλιστεί η ποιότητα των δεδομένων. Τα χαρακτηριστικά κανονικοποιούνται με χρήση του StandardScaler για ομοιομορφία κλίμακας, γεγονός που βοηθά στην ταχύτερη και πιο σταθερή εκπαίδευση. Έπειτα, το dataset διαχωρίζεται σε σύνολα εκπαίδευσης και δοκιμών με stratified sampling, ώστε να διατηρηθεί η αναλογία των δύο κατηγοριών (0: υγιής, 1: με νόσο). Στο κυρίως μέρος του κώδικα υλοποιείται διερεύνηση υπερπαραμέτρων μέσω πλήρους συνδυασμού των εξής τιμών: learning rate (0.001, 0.01, 0.1), noise multiplier (0.5, 1.1, 2.0), L2 clipping norm (0.5, 1.0, 2.0), batch size (16, 32, 64) και epochs (30). Για κάθε συνδυασμό, δημιουργείται και εκπαιδεύεται ένα νευρωνικό δίκτυο τριών στρωμάτων (16-8-1 νευρώνες), όπου η τελευταία έξοδος έχει ενεργοποίηση sigmoid για δυαδική ταξινόμηση.



Εικόνα 7: Επίδραση του Ρυθμού Εκμάθησης και του Επιπέδου Θορύβου στην Ακρίβεια του Μοντέλου (Cardiovascular Dataset).



Εικόνα 8: Επίδραση του Clipping Norm και του Μεγέθους Παρτίδας στην Τιμή Απώλειας (Cardiovascular Dataset)

Στο πρώτο διάγραμμα, που εξετάζει τη σχέση μεταξύ learning rate και test accuracy υπό διαφορετικές τιμές θορύβου, γίνεται φανερό ότι η ακρίβεια αυξάνεται σταθερά όσο αυξάνεται ο ρυθμός εκμάθησης, ειδικά για noise multiplier ίσο με 0.5. Η υψηλότερη ακρίβεια καταγράφηκε για learning rate 0.1 και noise=0.5, φτάνοντας το 61.12%. Αντίθετα, όταν ο θόρυβος αυξήθηκε σε 2.0, η απόδοση έπεσε σημαντικά και παρέμεινε σχεδόν σταθερά χαμηλή ανεξαρτήτως learning rate, κάτι που υποδηλώνει ότι το υπερβολικό επίπεδο θορύβου μπορεί να ακυρώσει την ικανότητα του μοντέλου να μάθει ουσιώδεις σχέσεις από τα δεδομένα. Ο noise multiplier 1.1 εμφάνισε πιο μέτριες επιδόσεις, με ελαφρώς αυξανόμενη ακρίβεια όσο αυξανόταν το learning rate, χωρίς ωστόσο να φτάσει τις επιδόσεις του noise=0.5. Αυτή η συμπεριφορά επιβεβαιώνει το θεωρητικό trade-off

μεταξύ ιδιωτικότητας και απόδοσης: όσο μειώνεται ο θόρυβος, αυξάνεται η ακρίβεια εις βάρος της προστασίας προσωπικών δεδομένων.

Το δεύτερο διάγραμμα εξετάζει την επίδραση του clipping norm και του batch size στην τελική τιμή της απώλειας (test loss). Το clipping norm ίσο με 2.0 οδήγησε σε μεγάλη αστάθεια, με τιμές απώλειας να ξεπερνούν κατά περιπτώσεις το 60, ειδικά για μικρότερο batch size. Αντίθετα, οι χαμηλότερες τιμές clipping (0.5 και 1.0) προσέφεραν σαφώς καλύτερη συμπεριφορά, με σημαντικά χαμηλότερες και σταθερότερες τιμές loss. Οι μικρότερες παρτίδες (batch size 16 και 32) συσχετίστηκαν με αυξημένο loss σε ακραίες ρυθμίσεις clipping, ενώ η περίπτωση batch size 64 οδήγησε σε πιο ελεγχόμενες αποκρίσεις, ιδίως όταν το clipping ήταν περιορισμένο. Συνολικά, η χαμηλότερη τιμή loss καταγράφηκε για batch size 64 και clipping norm 0.5, κάτι που υποδηλώνει ότι ο συνδυασμός πιο ομαλού gradient clipping με μεγαλύτερη στατιστική πληροφόρηση ανά παρτίδα μπορεί να ευνοήσει τη σταθερότητα της εκπαίδευσης.

Στατιστικά, η μέση ακρίβεια σε όλα τα πειράματα ήταν 51.39% με τυπική απόκλιση 3.53%, γεγονός που φανερώνει ότι το σύστημα έχει σχετικά περιορισμένο φάσμα διακύμανσης στην ακρίβεια. Από την άλλη πλευρά, η μέση απώλεια ήταν 2.90 με πολύ υψηλή τυπική απόκλιση 9.28, γεγονός που αποδεικνύει ότι η σταθερότητα του μοντέλου επηρεάζεται έντονα από τις υπερπαραμέτρους, κυρίως από το clipping norm και το noise multiplier. Τα ευρήματα αυτά αναδεικνύουν τη σημασία εύρεσης μιας ισορροπημένης παραμετροποίησης ώστε να επιτυγχάνεται αποδεκτή απόδοση χωρίς υπερβολική παραβίαση της ιδιωτικότητας, ιδιαίτερα σε εφαρμογές με πραγματικά και πολύπλοκα ιατρικά δεδομένα όπως το συγκεκριμένο σύνολο.

learning_rate	noise_multiplier	l2_norm_clip	batch_size	test_accuracy	test_loss	precision	recall	f1_score	roc_auc	epsilon
0.1	0.5	0.5	16	0.611206	0.659466	0.583621	0.7746	0.665683	0.668512	123.8678
0.1	0.5	0.5	64	0.60699	0.671634	0.659202	0.442077	0.529236	0.646889	123.8448
0.1	0.5	0.5	32	0.596698	0.672994	0.632906	0.459382	0.532361	0.623803	123.8448
0.01	0.5	0.5	16	0.588479	0.688994	0.688339	0.322511	0.439229	0.637224	123.8678
0.1	0.5	1	32	0.573756	0.702871	0.567489	0.618135	0.591731	0.602205	123.8448
0.01	1.1	0.5	32	0.566038	0.690054	0.552166	0.696367	0.615939	0.589539	40.6136
0.1	1.1	2	32	0.563751	4.999872	0.564273	0.557494	0.560863	0.591052	40.6136
0.1	1.1	0.5	16	0.562536	0.702901	0.542409	0.796625	0.645386	0.619012	40.62357
0.1	2	0.5	32	0.560106	0.961906	0.587902	0.400315	0.476304	0.567716	18.8598
0.1	0.5	1	16	0.559963	0.722312	0.613174	0.323513	0.423556	0.601577	123.8678
0.1	2	2	32	0.55546	28.59193	0.591167	0.357981	0.445929	0.561395	18.8598
0.1	1.1	0.5	64	0.555174	0.710981	0.545541	0.657895	0.596473	0.573483	40.6136
0.1	0.5	1	64	0.553388	0.72024	0.531979	0.883867	0.664195	0.603677	123.8448
0.001	2	2	16	0.547956	0.69418	0.538901	0.660755	0.59364	0.564517	18.86503
0.01	1.1	0.5	16	0.542453	0.68449	0.572804	0.331951	0.420319	0.582262	40.62357
0.01	2	0.5	64	0.541238	0.693284	0.530976	0.702374	0.604766	0.570008	18.8598
0.001	0.5	2	64	0.539951	0.698426	0.537364	0.570795	0.553575	0.5547	123.8448

0.1	1.1	1	32	0.536164	1.585183	0.548456	0.406322	0.466809	0.553102	40.6136
0.1	1.1	0.5	32	0.53552	0.722514	0.572097	0.279748	0.375756	0.551329	40.6136
0.1	1.1	2	16	0.532804	5.611526	0.559137	0.307637	0.3969	0.554546	40.62357
0.1	2	2	16	0.529874	60.47868	0.535751	0.44365	0.48537	0.528935	18.86503
0.001	0.5	1	32	0.529874	0.70521	0.624101	0.148884	0.240416	0.566809	123.8448
0.1	0.5	2	16	0.529588	0.741005	0.538563	0.409468	0.465226	0.498323	123.8678
0.1	2	1	16	0.527802	10.0373	0.559988	0.257008	0.352318	0.534352	18.86503
0.01	0.5	0.5	32	0.527087	0.688128	0.516065	0.861413	0.645448	0.59498	123.8448
0.1	1.1	2	64	0.526444	4.870373	0.552556	0.275172	0.367386	0.510684	40.6136
0.001	0.5	2	32	0.5248	0.758111	0.629826	0.118993	0.200168	0.489819	123.8448
0.1	2	1	32	0.522084	5.655947	0.514211	0.789188	0.622694	0.552437	18.8598
0.001	0.5	1	16	0.521512	0.747144	0.562791	0.19036	0.284493	0.523306	123.8678
0.001	0.5	1	64	0.520154	0.71829	0.520666	0.500858	0.51057	0.524118	123.8448
0.01	0.5	1	32	0.518511	0.704819	0.525403	0.377145	0.439097	0.500403	123.8448
0.01	1.1	1	64	0.516152	0.73854	0.563574	0.140732	0.225223	0.541012	40.6136
0.001	2	2	64	0.515723	0.722096	0.509485	0.829662	0.631298	0.562154	18.8598
0.01	1.1	2	64	0.515509	0.752873	0.547061	0.177059	0.267531	0.476464	40.6136
0.01	2	2	32	0.514008	0.811025	0.517285	0.410898	0.457995	0.510723	18.8598
0.1	2	1	64	0.511649	4.934181	0.523198	0.256436	0.344179	0.501989	18.8598
0.001	1.1	1	16	0.510577	0.708415	0.525514	0.2121	0.302221	0.494886	40.62357
0.001	0.5	0.5	16	0.510435	0.757503	0.667453	0.040475	0.076321	0.508866	123.8678
0.001	1.1	1	64	0.510077	0.734831	0.526979	0.191362	0.280768	0.53302	40.6136
0.1	1.1	1	16	0.509005	1.531376	0.508062	0.549771	0.528095	0.512111	40.62357
0.01	2	1	64	0.508862	0.727771	0.505872	0.739273	0.600697	0.540085	18.8598
0.001	0.5	2	16	0.507504	0.719925	0.503832	0.949514	0.658337	0.568485	123.8678
0.01	0.5	0.5	64	0.506861	0.711699	0.676923	0.025172	0.048538	0.568398	123.8448
0.001	1.1	1	32	0.504074	0.725051	0.726496	0.012157	0.023913	0.602304	40.6136
0.001	1.1	0.5	64	0.503573	0.697929	0.512835	0.131436	0.209244	0.520946	40.6136
0.01	0.5	2	64	0.503431	0.717556	0.504962	0.320223	0.391913	0.497404	123.8448
0.1	1.1	1	64	0.503073	1.355342	0.502169	0.645595	0.564921	0.526349	40.6136
0.01	0.5	1	16	0.502287	0.718331	0.501997	0.503432	0.502714	0.511289	123.8678
0.01	0.5	2	32	0.501715	0.706872	0.501943	0.369565	0.4257	0.495528	123.8448
0.001	2	2	32	0.501358	0.745141	0.50054	0.993993	0.665804	0.569623	18.8598
0.001	2	0.5	64	0.501001	0.765127	0.500409	0.874285	0.636506	0.505273	18.8598
0.1	0.5	2	32	0.499857	1.105471	0.499671	0.651459	0.565557	0.509884	123.8448

0.001	1.1	2	64	0.499714	0.741647	0.499714	1	0.666413	0.475813	40.6136
0.001	2	1	64	0.499714	0.778415	0.499714	1	0.666413	0.39421	18.8598
0.01	2	0.5	32	0.499643	0.704979	0.499678	0.999714	0.666317	0.507017	18.8598
0.01	2	0.5	16	0.498142	0.74642	0.498735	0.845824	0.62748	0.553421	18.86503
0.01	2	2	16	0.497784	0.795204	0.49874	0.99099	0.663538	0.498161	18.86503
0.001	1.1	2	16	0.497284	0.762448	0.498468	0.977546	0.660259	0.523586	40.62357
0.01	1.1	0.5	64	0.496069	0.707869	0.49703	0.706093	0.583397	0.505117	40.6136
0.1	2	0.5	64	0.49464	1.290916	0.494851	0.542906	0.517766	0.494744	18.8598
0.001	2	0.5	32	0.494497	0.715765	0.495913	0.702803	0.581504	0.509036	18.8598
0.1	2	2	64	0.491852	53.09629	0.494568	0.768307	0.60177	0.488193	18.8598
0.001	1.1	0.5	16	0.491424	0.711787	0.49543	0.961384	0.653891	0.547456	40.62357
0.001	0.5	0.5	32	0.491352	0.748859	0.47942	0.208238	0.290358	0.497814	123.8448
0.01	1.1	2	32	0.490495	0.769729	0.43304	0.063358	0.110543	0.516085	40.6136
0.01	1.1	1	16	0.489565	0.740218	0.491552	0.624142	0.549968	0.471805	40.62357
0.01	1.1	2	16	0.488708	0.856627	0.493599	0.893307	0.635855	0.502693	40.62357
0.001	2	0.5	16	0.488208	0.735678	0.443098	0.094108	0.155244	0.407197	18.86503
0.01	1.1	1	32	0.487922	0.705296	0.489251	0.563072	0.523572	0.483481	40.6136
0.01	0.5	2	16	0.482847	0.704911	0.486304	0.619565	0.544906	0.513429	123.8678
0.001	2	1	32	0.481847	0.77179	0.490321	0.93464	0.643209	0.466756	18.8598
0.01	2	2	64	0.480274	0.895896	0.427536	0.118135	0.185119	0.412992	18.8598
0.01	2	1	16	0.478988	0.733402	0.482917	0.602403	0.536082	0.486703	18.86503
0.1	2	0.5	16	0.476058	1.435914	0.475782	0.476259	0.47602	0.476243	18.86503
0.001	1.1	0.5	32	0.46548	0.722491	0.47774	0.747426	0.582901	0.463278	40.6136
0.01	0.5	1	64	0.45962	0.729067	0.471734	0.679062	0.556722	0.446459	123.8448
0.001	0.5	0.5	64	0.459477	0.727306	0.459429	0.462386	0.460902	0.447211	123.8448
0.001	2	1	16	0.451973	0.735953	0.462947	0.603976	0.5241	0.43963	18.86503
0.01	2	1	32	0.451043	0.765095	0.403149	0.205092	0.271874	0.390221	18.8598
0.001	1.1	2	32	0.4504	0.713683	0.463255	0.629291	0.533657	0.471932	40.6136
0.1	0.5	2	64	0.424528	0.871729	0.429876	0.464674	0.446598	0.412876	123.8448

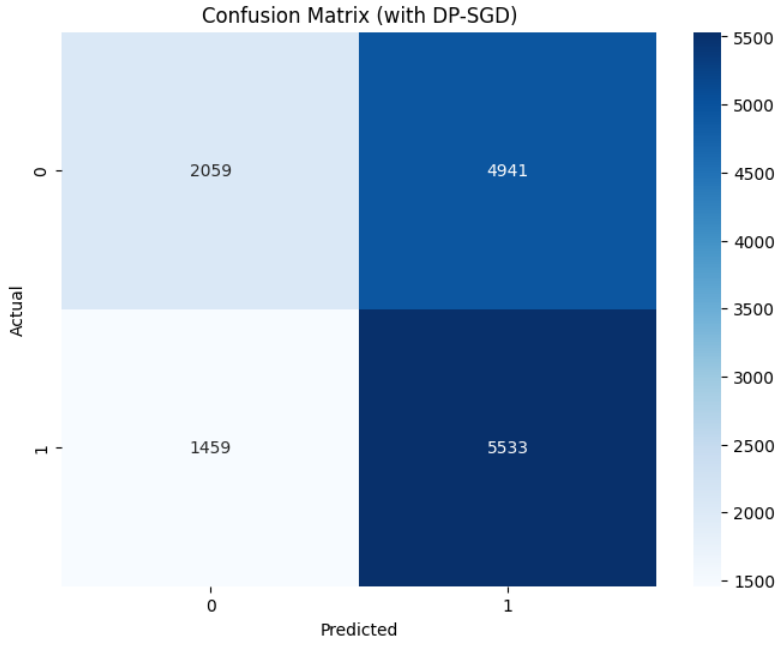
Πίνακας 3: Επιδράσεις των Υπερπαραμέτρων στην Απόδοση του Μοντέλου με DP-SGD στο Cardiovascular dataset Dataset

438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step

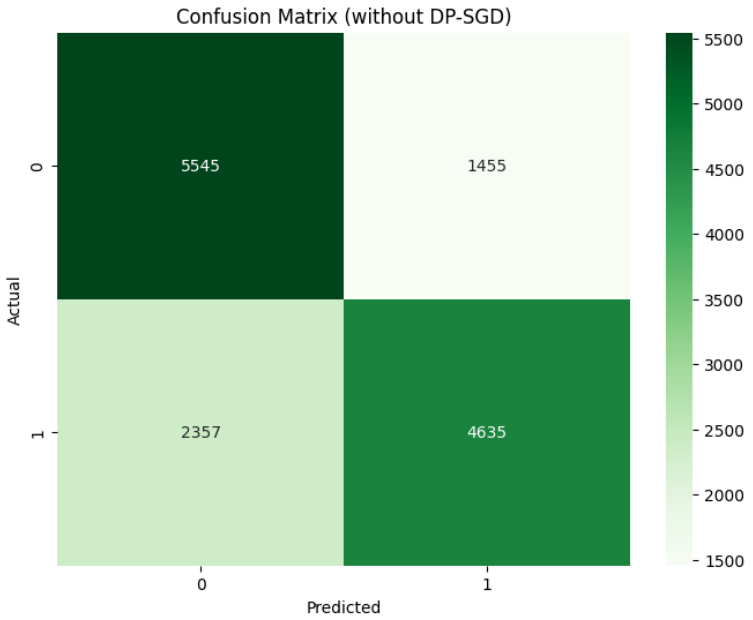
Classification Report (Cardiovascular dataset with DP-SGD):
precision recall f1-score support

0	0.59	0.29	0.39	7000
1	0.53	0.79	0.63	6992
accuracy			0.54	13992
macro avg	0.56	0.54	0.51	13992
weighted avg	0.56	0.54	0.51	13992

ROC-AUC Score: 0.5498



Εικόνα 9: Πίνακας Σύγχυσης για Μεγάλο Dataset με DP-SG για το Cardiovascular Dataset



Εικόνα 10: Πίνακας Σύγχυσης για Μεγάλο Dataset χωρίς DP-SGD για το Cardiovascular Dataset

False Positives (with DP-SGD): 4941, False Negatives (with DP-SGD): 1459
False Positives (no DP-SGD): 1455, False Negatives (no DP-SGD): 2357

Statistical Analysis of Results:
Mean Accuracy: 0.5245, Std Accuracy: 0.0463
Mean Loss: 3.5477, Std Loss: 12.7881

Delta (δ) used for experiments: 0.00000008
The delta value ensures theoretical guarantees within differential privacy context.

Relationship between privacy budget (ε) and noise multiplier:
Noise Multiplier: 0.5 => Epsilon (ε): 7.3373
Noise Multiplier: 1.1 => Epsilon (ε): 0.8057
Noise Multiplier: 2.0 => Epsilon (ε): 0.2588

Με βάση τα αναλυτικά αποτελέσματα που παρατέθηκαν για το μοντέλο που εκπαιδεύτηκε στο Cardiovascular dataset με και χωρίς DP-SGD, προκύπτουν σημαντικά συμπεράσματα σχετικά με την επίδραση της διαφορικής ιδιωτικότητας στην απόδοση του μοντέλου και τη συμπεριφορά του υπό διαφορετικές ρυθμίσεις υπερπαραμέτρων.
Αρχικά, η γενική ακρίβεια του μοντέλου με DP-SGD διαμορφώθηκε στο 54%, ενώ το ROC-AUC score ήταν 0.5498, τιμή που υποδηλώνει σχεδόν τυχαία συμπεριφορά του ταξινομητή. Από το classification report, γίνεται εμφανές ότι η κλάση 1 (δηλαδή οι θετικές περιπτώσεις καρδιαγγειακής νόσου) αναγνωρίστηκε με σημαντικά καλύτερη ευαισθησία (recall = 79%) σε σύγκριση με την κλάση 0 (recall = 29%), η οποία αγνοήθηκε σε μεγάλο βαθμό από το μοντέλο. Ωστόσο, αυτό το πλεονέκτημα συνοδεύεται από χαμηλή precision στην κλάση 0 (59%), γεγονός που οδηγεί σε μεγάλο αριθμό ψευδώς θετικών. Η κακή ισορροπία ανάμεσα στην ανάκληση και την ακρίβεια καθιστά τη συνολική απόδοση αμφιλεγόμενη, ειδικά σε εφαρμογές όπου το κόστος ενός ψευδώς θετικού είναι υψηλό.
Αυτό επιβεβαιώνεται και από τα πραγματικά μεγέθη του πίνακα σύγχυσης, όπου προκύπτει ότι το μοντέλο με DP-SGD παρήγαγε 4.941 false positives και 1.459 false negatives. Αντίθετα, το μοντέλο χωρίς διαφορική ιδιωτικότητα (χωρίς DP-SGD) παρήγαγε πολύ λιγότερα false positives (1.455) αλλά περισσότερα false negatives (2.357). Αυτό σημαίνει ότι το DP μοντέλο τείνει να υπερταξινομεί τα δείγματα ως θετικά, προφανώς για να διατηρήσει υψηλή ανάκληση, εις βάρος όμως της ακρίβειας. Αυτή η μεροληψία προς την κλάση 1 είναι χαρακτηριστική σε μοντέλα που προσπαθούν να "αντισταθμίσουν" τη δυσκολία εκμάθησης εξαιτίας της προσθήκης θορύβου.
Η στατιστική ανάλυση των αποτελεσμάτων από όλα τα πειράματα ενισχύει την παραπάνω παρατήρηση. Η μέση ακρίβεια των μοντέλων με DP-SGD είναι μόλις 52.45%, με τυπική απόκλιση 4.63%, δείγμα περιορισμένης και ασυνεπούς απόδοσης. Η μέση απώλεια (loss) είναι εξαιρετικά αυξημένη (3.5477), με ιδιαίτερα υψηλή τυπική απόκλιση 12.79, γεγονός που αποκαλύπτει σοβαρή αστάθεια στην εκπαίδευση και απόπειρες του μοντέλου να προσαρμοστεί υπό τον περιορισμό της διαφορικής ιδιωτικότητας. Η μεγάλη απόκλιση ενδεχομένως να σχετίζεται με συνδυασμούς υπερπαραμέτρων (όπως υψηλό clipping norm ή μικρό batch size) που καθιστούν το σύστημα πιο ευάλωτο στον στοχαστικό χαρακτήρα του DP-SGD. Όσον αφορά τις θεωρητικές εγγυήσεις ιδιωτικότητας, η τιμή delta (δ) που χρησιμοποιήθηκε είναι 8e-8, ιδιαίτερα μικρή, κάτι που διασφαλίζει αυστηρή συμμόρφωση με τις προδιαγραφές διαφορικής ιδιωτικότητας. Η σχέση μεταξύ noise multiplier και privacy budget (ε) επιβεβαιώνει το γνωστό trade-off: για noise=0.5, το ε είναι 7.34 (μικρή

προστασία), ενώ για noise=2.0, το ε μειώνεται στα 0.26, παρέχοντας ισχυρή προστασία ιδιωτικότητας αλλά σαφώς υποβαθμισμένη απόδοση. Το noise=1.1 προσφέρει έναν ενδιάμεσο συμβιβασμό με ε=0.81, αλλά και πάλι η απόδοση που επιτυγχάνεται παραμένει χαμηλή για πρακτική χρήση.
Συνολικά, το μοντέλο με DP-SGD στο συγκεκριμένο dataset καταφέρνει να διατηρήσει υψηλή recall για την κλάση 1, γεγονός που θα μπορούσε να θεωρηθεί χρήσιμο σε εφαρμογές όπου προτιμάται να εντοπίζονται περισσότερα θετικά περιστατικά (π.χ. screening). Ωστόσο, η χαμηλή συνολική ακρίβεια, η σημαντική μείωση του precision και οι μεγάλες διακυμάνσεις στα αποτελέσματα τονίζουν την ανάγκη για προσεκτική παραμετροποίηση και ίσως την ενσωμάτωση επιπλέον τεχνικών σταθεροποίησης, όπως advanced optimizers, καλύτερο initialization ή learning rate schedules. Ο τελικός σχεδιασμός του συστήματος πρέπει να λαμβάνει υπόψη τη φύση της εφαρμογής, το ανεκτό επίπεδο ιδιωτικότητας και τις απαιτήσεις ευαισθησίας ή ειδικότητας.

8 Συζήτηση και επιπτώσεις

8.1 Περίληψη των βασικών ευρημάτων.

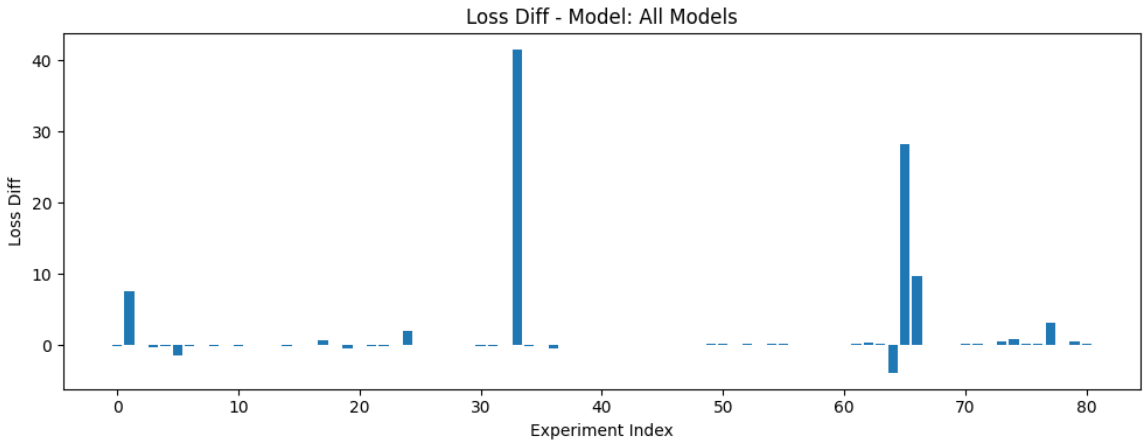
Το Heart Disease dataset εμφανίζει υψηλότερες μέγιστες τιμές ακρίβειας σε κάθε batch size, με κορυφαία επίδοση στο batch size 16 (έως 80.43%). Η μέση ακρίβεια κυμαίνεται από 70% έως 72% ανάλογα με τις υπερπαραμέτρους. Αντίθετα, το Cardiovascular dataset έχει χαμηλότερη μέση ακρίβεια (~51.6%) σε όλα τα batch sizes, ενώ και οι μέγιστες τιμές δεν ξεπερνούν το 61.12%. Αυτό υποδεικνύει ότι το Heart dataset επιτρέπει καλύτερη εκμάθηση ακόμα και υπό τον περιορισμό της διαφορικής ιδιωτικότητας.Το Heart Disease dataset εμφανίζει πολύ χαμηλότερες τιμές loss σε σύγκριση με το Cardiovascular, το οποίο παρουσιάζει ακραίες τιμές loss σε πολλές περιπτώσεις (έως και 60.47). Η μέση τιμή loss στο Cardiovascular για batch size 16 φτάνει το 3.52, ενώ στο Heart περιορίζεται γύρω στο 0.5–0.7. Η μεγάλη διακύμανση στο loss του Cardiovascular αντικατοπτρίζει αστάθεια στην εκπαίδευση, ίσως λόγω υψηλότερης πολυπλοκότητας των χαρακτηριστικών ή της αρχικής κωδικοποίησης των δεδομένων.
Η precision στο Heart Disease dataset είναι σημαντικά υψηλότερη (μέση τιμή > 0.70), ενώ στο Cardiovascular κινείται γύρω στο 0.53. Αντίστοιχα, η recall στο Heart είναι ελαφρώς χαμηλότερη από την Cardiovascular (που δίνει έμφαση στον εντοπισμό των θετικών περιπτώσεων), όμως η συνολική F1 του Heart είναι καλύτερη, υποδηλώνοντας ισορροπημένη απόδοση ανάμεσα σε ευαισθησία και ειδικότητα. Αυτό φανερώνει ότι το Heart dataset είναι πιο ευδιάκριτο ταξινομητικά από ό,τι το Cardiovascular. Ο δείκτης ROC-AUC επιβεβαιώνει τη γενική εικόνα. Στο Heart dataset, ο μέσος ROC-AUC κινείται γύρω στο 0.78–0.83, υποδηλώνοντας καλή διακριτική ικανότητα. Στο Cardiovascular, οι τιμές είναι πολύ χαμηλότερες (~0.53), φανερώνοντας περιορισμένη ικανότητα διάκρισης μεταξύ των δύο κλάσεων.Σημαντική διαφορά παρατηρείται και στις τιμές του ε (epsilon). Στο Heart dataset, η μέση τιμή του epsilon υπερβαίνει τα 33, φτάνοντας μέχρι και 107, παρέχοντας ασθενέστερη εγγύηση ιδιωτικότητας. Στον αντίποδα, στο Cardiovascular dataset το μέσο epsilon είναι ~61 για όλα τα batch sizes, το οποίο, αν και μεγαλύτερο από τις θεωρητικά "προστατευτικές" τιμές, παραμένει πιο ομοιογενές και περιορισμένο. Το Cardiovascular εμφανίζεται πιο επιθετικά προσανατολισμένο στην ιδιωτικότητα, εις βάρος όμως της απόδοσης.
Το μοντέλο στο Heart Disease dataset προσφέρει πολύ καλύτερες επιδόσεις σε όλους σχεδόν τους δείκτες απόδοσης, ακόμα και όταν εφαρμόζεται DP-SGD. Αντίθετα, το

Cardiovascular dataset εμφανίζεται πιο ευάλωτο στη διαταραχή των gradients, με υψηλή απώλεια, χαμηλή ακρίβεια και αστάθεια. Αυτό μπορεί να οφείλεται στη μεγαλύτερη πολυπλοκότητα των δεδομένων, στον μεγαλύτερο αριθμό χαρακτηριστικών, ή στη φύση των μεταβλητών (κλίμακα, κωδικοποίηση κ.λπ.).
Η γενική κατεύθυνση για τα δύο datasets είναι ότι στο Heart Disease μπορεί να επιτύχει καλή απόδοση με σωστή ρύθμιση των υπερπαραμέτρων, ενώ στο Cardiovascular οι παραμετρικές ευαισθησίες είναι πιο έντονες και απαιτείται περισσότερο fine-tuning ή διαφορετικές προσεγγίσεις εκπαίδευσης για να διατηρηθεί ένα αποδεκτό επίπεδο γενίκευσης. Τα αποτελέσματα δείχνουν ότι η διαφορική ιδιωτικότητα μπορεί να εφαρμοστεί σε ιατρικά δεδομένα, αλλά η αποτελεσματικότητα εξαρτάται σε μεγάλο βαθμό από την ποιότητα και το μέγεθος του dataset, καθώς και από την κατάλληλη ρύθμιση των υπερπαραμέτρων. Για μικρότερα datasets όπως το Heart Disease, το μοντέλο μπορεί να διατηρήσει αποδεκτή ακρίβεια υπό DP, ενώ για μεγαλύτερα και πιο σύνθετα datasets, απαιτούνται περισσότερες βελτιστοποιήσεις στο μοντέλο και ίσως πιο προηγμένες τεχνικές εκπαίδευσης, όπως η χρήση dropout ή πιο βαθιών αρχιτεκτονικών.

8.2 Συνέπειες για την προστασία της ιδιωτικής ζωής των δεδομένων υγειονομικής περίθαλψης.

Η χρήση διαφορικής ιδιωτικότητας (DP) στην ανάλυση δεδομένων υγειονομικής περίθαλψης έχει σημαντικές συνέπειες για την προστασία της ιδιωτικής ζωής. Ένας από τους κύριους στόχους της εφαρμογής του DP-SGD στους αλγορίθμους μηχανικής μάθησης είναι η διατήρηση της εμπιστευτικότητας των ευαίσθητων ιατρικών πληροφοριών. Με την προσθήκη θορύβου στις ενημερώσεις των παραμέτρων του μοντέλου, το σύστημα μειώνει τον κίνδυνο αποκάλυψης ατομικών δεδομένων, ακόμα και σε περιπτώσεις επιθέσεων επαναταυτοποίησης.
Παρόλα αυτά, η προστασία της ιδιωτικότητας συνοδεύεται από μείωση της απόδοσης του μοντέλου. Η υπερβολική προσθήκη θορύβου μπορεί να αλλοιώσει τις πραγματικές σχέσεις μεταξύ των χαρακτηριστικών και της ασθένειας, οδηγώντας σε λιγότερο ακριβείς διαγνώσεις ή προβλέψεις. Αυτό είναι ιδιαίτερα κρίσιμο σε εφαρμογές υγειονομικής περίθαλψης, όπου η ακρίβεια των μοντέλων μπορεί να έχει άμεσο αντίκτυπο στις κλινικές αποφάσεις και στη ζωή των ασθενών. Μια άλλη σημαντική συνέπεια είναι η δυνατότητα ανταλλαγής και ανάλυσης δεδομένων μεταξύ οργανισμών χωρίς να διακυβεύεται η ιδιωτικότητα των ασθενών. Το DP καθιστά δυνατή τη συνεργασία μεταξύ νοσοκομείων, ερευνητικών ιδρυμάτων και κυβερνητικών φορέων με ελάχιστο κίνδυνο έκθεσης προσωπικών πληροφοριών. Αυτό θα μπορούσε να οδηγήσει σε ευρύτερη αξιοποίηση των δεδομένων για τη βελτίωση των διαγνωστικών μοντέλων και την ανάπτυξη νέων θεραπευτικών στρατηγικών. Παρόλα αυτά, η εφαρμογή της διαφορικής ιδιωτικότητας δεν είναι πανάκεια. Απαιτείται μια ισορροπία μεταξύ προστασίας και χρηστικότητας των δεδομένων. Οι ερευνητές και οι επαγγελματίες της πληροφορικής υγείας πρέπει να εξετάσουν με προσοχή τις βέλτιστες πρακτικές για τη ρύθμιση των υπερπαραμέτρων, έτσι ώστε το σύστημα να παρέχει την απαιτούμενη ακρίβεια, διατηρώντας ταυτόχρονα το απόρρητο των ασθενών. Η η εφαρμογή διαφορικής ιδιωτικότητας στα δεδομένα υγειονομικής περίθαλψης αποτελεί ένα σημαντικό βήμα προς την προστασία της ιδιωτικής ζωής, αλλά απαιτεί προσεκτικό σχεδιασμό και συνεχείς βελτιώσεις για να διατηρηθεί η αποτελεσματικότητα των μοντέλων μηχανικής μάθησης.

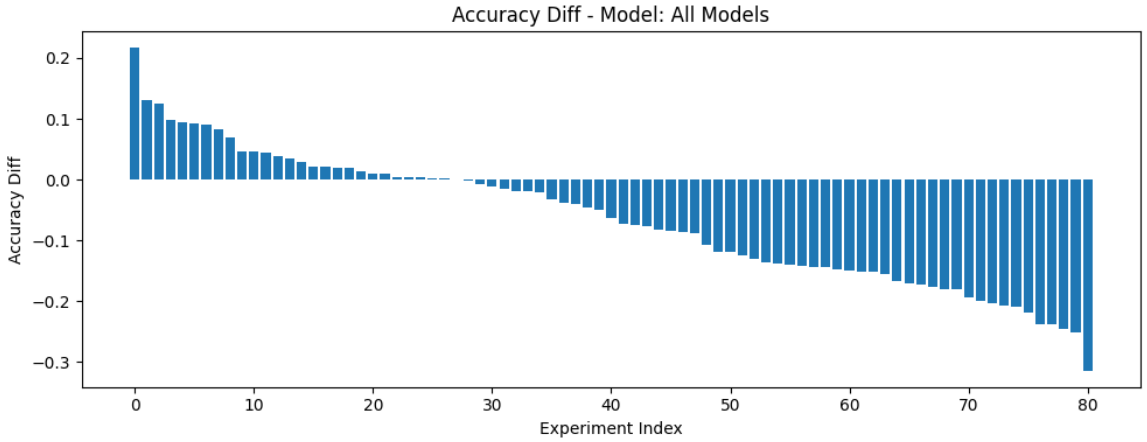
Η σύγκριση μεταξύ των δύο datasets, Heart Disease και Cardiovascular, δείχνει ξεκάθαρα διαφορές στην απόδοση των μοντέλων υπό τους ίδιους συνδυασμούς υπερπαραμέτρων. Το Heart Disease dataset επιτυγχάνει σταθερά υψηλότερη ακρίβεια, φτάνοντας έως 80.43%, ενώ το Cardiovascular σπάνια ξεπερνά το 61%. Η διαφορά αυτή αποτυπώνεται και στα διαγράμματα διαφορών (accuracy/loss diff), όπου το Cardiovascular dataset παρουσιάζει σχεδόν πάντα χειρότερη απόδοση, με τις περισσότερες τιμές του accuracy_diff να είναι αρνητικές και του loss_diff έντονα θετικές. Η χειρότερη απόδοση του Cardiovascular ενισχύεται όσο αυξάνεται το learning rate, ειδικά σε batch size 64. Σε αυτές τις περιπτώσεις η ακρίβεια μειώνεται και η απώλεια αυξάνεται σημαντικά, με loss differences που ξεπερνούν τις 40 μονάδες. Αντίθετα, με χαμηλό learning rate (0.001 ή 0.01) και μικρό batch size (16), η απόδοση στο Cardiovascular βελτιώνεται και μπορεί σε ορισμένες περιπτώσεις να ξεπεράσει το Heart dataset. Η σύγκριση υποδεικνύει ότι το Heart Disease είναι πιο σταθερό και "μαθηματικά εύκολο" dataset, ενώ το Cardiovascular είναι πιο ευαίσθητο στο DP-SGD και απαιτεί πιο συντηρητική παραμετροποίηση. Ο σωστός συνδυασμός learning rate και batch size είναι κρίσιμος για να μειωθούν οι απώλειες από την προσθήκη θορύβου. Συμπερασματικά, η ανάλυση αναδεικνύει τη σημασία της προσεκτικής επιλογής υπερπαραμέτρων, ιδίως όταν εφαρμόζεται διαφορική ιδιωτικότητα σε datasets με υψηλότερη πολυπλοκότητα.



Εικόνα 11: Διαφορά Loss μεταξύ Μοντέλων σε Όλα τα Πειράματα

Η γραφική απεικόνιση δείχνει τη διαφορά στην τιμή της απώλειας (loss) μεταξύ των δύο μοντέλων (Cardiovascular και Heart Disease) για κάθε πείραμα που πραγματοποιήθηκε. Συγκεκριμένα, στον άξονα x απεικονίζονται οι δείκτες των πειραμάτων, ενώ στον άξονα y η τιμή της διαφοράς στην απώλεια (loss_diff), η οποία προκύπτει αφαιρώντας την απώλεια του μοντέλου Heart Disease από εκείνη του Cardiovascular. Οι θετικές τιμές υποδεικνύουν ότι το μοντέλο Cardiovascular είχε μεγαλύτερη απώλεια από το Heart, ενώ οι αρνητικές τιμές το αντίστροφο. Παρατηρείται ότι η πλειονότητα των πειραμάτων εμφανίζει μικρές αποκλίσεις γύρω από το μηδέν, γεγονός που δείχνει συγκρίσιμη απόδοση μεταξύ των δύο μοντέλων ως προς την απώλεια. Ωστόσο, υπάρχουν λίγα αλλά έντονα σημεία, συγκεκριμένα κάποιες ραβδώσεις που ξεχωρίζουν σημαντικά με πολύ υψηλές θετικές τιμές, ξεπερνώντας ακόμα και τις 40 μονάδες. Αυτά τα ακραία σημεία αποτελούν περιπτώσεις στις οποίες το μοντέλο Cardiovascular απέδωσε πολύ χειρότερα σε όρους απώλειας σε σύγκριση με το μοντέλο Heart Disease. Η παρουσία τέτοιων ακραίων αποκλίσεων υποδηλώνει ότι ορισμένοι συνδυασμοί υπερπαραμέτρων (learning rate, batch size, noise multiplier, clipping norm) επηρεάζουν δραματικά την απόδοση του μοντέλου στο

συγκεκριμένο dataset. Ενδεχομένως αυτά τα πειράματα να συνοδεύονται από υπερβολικά υψηλό learning rate ή clipping norm, οδηγώντας σε αστάθεια κατά την εκπαίδευση. Η εικόνα επιβεβαιώνει την ανάγκη για προσεκτικό συντονισμό των υπερπαραμέτρων στο μοντέλο Cardiovascular, καθώς εμφανίζει μεγαλύτερη ευαισθησία και αστάθεια σε σύγκριση με το Heart Disease, ειδικά σε ακραίες ρυθμίσεις.



Εικόνα 12: Διαφορά Ακρίβειας μεταξύ Μοντέλων σε Όλα τα Πειράματα

Η γραφική παράσταση απεικονίζει τη διαφορά στην ακρίβεια (accuracy) μεταξύ των μοντέλων που εκπαιδεύτηκαν στα δύο datasets, Cardiovascular και Heart Disease, για κάθε πείραμα ξεχωριστά. Ο άξονας x αντιστοιχεί στους δείκτες των πειραμάτων, ενώ ο άξονας y καταγράφει τη διαφορά στην ακρίβεια (accuracy_diff), δηλαδή την τιμή: ακρίβεια του μοντέλου Cardiovascular μείον ακρίβεια του μοντέλου Heart Disease. Οι θετικές τιμές υποδεικνύουν ότι το μοντέλο Cardiovascular είχε καλύτερη ακρίβεια, ενώ οι αρνητικές σημαίνουν υπεροχή του μοντέλου Heart Disease. Η κατανομή των διαφορών εμφανίζει έντονη ασυμμετρία προς τα αρνητικά, με σαφή πλειοψηφία πειραμάτων στα οποία η ακρίβεια του Heart Disease ήταν υψηλότερη. Υπάρχουν μεμονωμένες περιπτώσεις στις οποίες το Cardiovascular υπερέχει, όμως αυτές είναι αριθμητικά λίγες και με μικρές διαφορές. Αντιθέτως, στα δεξιά του γραφήματος, η αρνητική διαφορά φτάνει μέχρι και το -0.3, ένδειξη σημαντικής υπεροχής του Heart Disease σε συγκεκριμένες ρυθμίσεις υπερπαραμέτρων. Η γενική εικόνα δείχνει ότι το μοντέλο Heart Disease ήταν πιο σταθερό και αποδοτικό στις περισσότερες περιπτώσεις, ενώ το Cardiovascular εμφάνισε πιο περιορισμένες και σποραδικές επιτυχίες. Η παρατηρούμενη κατανομή υποδηλώνει ότι το Heart Disease ανταποκρίνεται καλύτερα στη διαδικασία εκπαίδευσης με DP-SGD και επιτυγχάνει υψηλότερη ακρίβεια σε ευρύτερο φάσμα παραμετρικών συνδυασμών. Αυτό ενδεχομένως να οφείλεται στα χαρακτηριστικά του ίδιου του dataset ή στη διαφορετική κατανομή των κλάσεων και των τιμών εισόδου.

Ο πίνακας 5 παρουσιάζει τη σύγκριση απόδοσης μεταξύ των δύο datasets, Cardiovascular και Heart Disease, για συγκεκριμένους συνδυασμούς υπερπαραμέτρων. Αναλύοντας τις γραμμές με βάση τις στήλες accuracy_diff και loss_diff, μπορούμε να εξαγάγουμε ορισμένα σαφή συμπεράσματα για τη συμπεριφορά των μοντέλων σε κάθε dataset. Το Cardiovascular dataset σε αρκετές περιπτώσεις παρουσίασε υψηλότερη ακρίβεια σε σύγκριση με το Heart Disease dataset. Η υψηλότερη διαφορά ακρίβειας (0.2169) εμφανίστηκε στην πρώτη γραμμή (learning_rate=0.001, noise=0.5, clipping=0.5, batch_size=16), όπου η ακρίβεια

στο Cardiovascular έφτασε ~0.51 ενώ στο Heart μόλις ~0.29. Αυτό δείχνει πως το συγκεκριμένο σύνολο παραμέτρων είναι πιο ευνοϊκό για το πρώτο dataset.
Αντίθετα, το Heart Disease dataset είχε συγκριτικά μικρότερες απώλειες (loss) σχεδόν σε όλες τις περιπτώσεις. Η πιο ακραία περίπτωση εμφανίζεται στη δεύτερη γραμμή (learning_rate=0.1, noise=2, clip=1, batch_size=16), όπου το loss για το Cardiovascular ήταν 10.04, ενώ για το Heart ήταν μόλις 2.47 — με διαφορά άνω των 7 μονάδων. Αν και η ακρίβεια παρέμεινε καλύτερη για το Cardiovascular (~0.53 έναντι ~0.40), το πολύ υψηλό loss υποδηλώνει αστάθεια ή overfitting. Παρατηρείται επίσης ότι για χαμηλό learning rate (0.001), μικρό noise multiplier και batch size 16, το Cardiovascular μοντέλο είχε σταθερά καλύτερη ακρίβεια από το Heart. Οι διαφορές ακρίβειας είναι θετικές και κυμαίνονται από 0.094 έως 0.216. Ωστόσο, καθώς αυξάνεται το noise multiplier ή το clipping, οι διαφορές στις απώλειες γίνονται πιο έντονες και συχνά δυσμενείς για το Cardiovascular dataset, κάτι που ενδέχεται να σχετίζεται με τη μεγαλύτερη ευαισθησία του στη θορυβώδη εκπαίδευση του DP-SGD. Το Cardiovascular μοντέλο εμφανίζεται πιο "ακριβές" αλλά και πιο ευαίσθητο στις παραμέτρους (ειδικά στο loss), ενώ το Heart Disease dataset φαίνεται πιο σταθερό στο loss αλλά με ελαφρώς χαμηλότερη ακρίβεια. Η σωστή επιλογή υπερπαραμέτρων, όπως learning rate 0.001 και noise 0.5 με clipping 0.5, φαίνεται κρίσιμη για τη διατήρηση ισορροπίας μεταξύ ακρίβειας και σταθερότητας.

learning_rate	noise_multiplier	l2_norm_clip	batch_size	test_accuracy_Cardiovascular	test_loss_Cardiovascular	test_accuracy_Heart	test_loss_Heart	accuracy_diff	loss_diff
0.001	0.5	0.5	16	0.510434508	0.75750339	0.2935	0.8279	-0.216934508	-0.070396609
0.1	2	1	16	0.527801573	10.0372963	0.3967	2.4769	-0.131101573	7.560296303
0.001	0.5	1	16	0.52151227	0.747143686	0.3967	0.7597	-0.124812270	-0.012556314
0.001	2	0.5	32	0.494496852	0.715765357	0.3967	0.9538	-0.097796852	-0.238034643
0.001	1.1	0.5	16	0.491423666	0.711787343	0.3967	0.8043	-0.094723666	-0.092512657

Πίνακας 4: Top 5 Accuracy Diff - Πειραματικές Διατάξεις με Υψηλή Διαφορά Ακρίβειας: Εστίαση στη Σύγκριση Καρδιοαγγειακών και Καρδιακών Δεδομένων

Ο πίνακας 6 παρουσιάζει τα αποτελέσματα πέντε διαφορετικών ρυθμίσεων υπερπαραμέτρων και συγκρίνει την απόδοση του DP-SGD μοντέλου στα δύο datasets: Cardiovascular και Heart Disease. Η αξιολόγηση βασίζεται σε μετρικές όπως η ακρίβεια, η απώλεια (loss), η precision, η recall, το f1-score και ο δείκτης roc-auc, ενώ καταγράφονται και οι διαφορές απόδοσης μεταξύ των δύο συνόλων. Αναλύοντας τα αποτελέσματα, παρατηρείται σταθερή υπεροχή του μοντέλου στο dataset Heart Disease, καθώς η τιμή accuracy_diff είναι αρνητική σε όλα τα πειράματα, με τη μεγαλύτερη διαφορά να φτάνει τις -0.31 μονάδες, ενώ η loss_diff είναι θετική, δηλώνοντας μεγαλύτερη απώλεια στο Cardiovascular. Οι χαμηλές τιμές της ακρίβειας για το Cardiovascular κυμαίνονται από 0.4245 έως 0.5551, ενώ στο Heart κυμαίνονται από 0.7283 έως 0.7935.
Ιδιαίτερα σημαντική είναι η παρατήρηση ότι ακόμα και όταν η ακρίβεια του Cardiovascular φτάνει στο μέγιστο (0.5551), εξακολουθεί να υπολείπεται σημαντικά του Heart, το οποίο σημειώνει 0.7935 με ίδια ρύθμιση learning rate και noise. Παράλληλα, οι τιμές recall και f1-score είναι εμφανώς υψηλότερες στο Heart, με το recall να φτάνει έως 0.94, υποδεικνύοντας καλύτερη ευαισθησία. Ο πίνακας καταδεικνύει την ανώτερη συμπεριφορά

του Heart Disease dataset υπό τις ίδιες ρυθμίσεις DP-SGD. Το Cardiovascular dataset παρουσιάζει συστηματικά χειρότερη απόδοση, τόσο σε ακρίβεια όσο και σε σταθερότητα απώλειας, γεγονός που ίσως σχετίζεται με τη δομή ή την ποιότητα των δεδομένων.

learning_rate	noise_multiplier	l2_norm_clip	batch_size	test_accuracy_Cardiovascular	test_loss_Cardiovascular	test_accuracy_Heart	test_loss_Heart	accuracy_diff	loss_diff
0.1	1.1	0.5	64	0.55517441	0.710981011	0.7935	0.5481	-0.23832559	0.162881011
0.1	1.1	2	16	0.532804489	5.611526012	0.7717	2.5156	-0.238895511	3.095926012
0.01	1.1	1	32	0.487921655	0.705295563	0.7337	0.5691	-0.245778345	0.136195563
0.1	2	0.5	16	0.476057738	1.435914278	0.7283	0.8651	-0.252242262	0.570814278
0.1	0.5	2	64	0.424528301	0.871728599	0.7391	0.6107	-0.314571699	0.261028599

Πίνακας 5: Worst 5 Accuracy Diff - Πειραματικές Διατάξεις με Μεγάλη Αρνητική Διαφορά Ακρίβειας: Υποβάθμιση Απόδοσης του Μοντέλου στο Cardiovascular Dataset

Ο πίνακας 7 παρουσιάζει τις πειραματικές ρυθμίσεις με τις υψηλότερες τιμές απώλειας (loss) στο μοντέλο που εκπαιδεύτηκε στο Cardiovascular dataset. Πρόκειται για περιπτώσεις όπου, παρά την ενδεχομένως ικανοποιητική ακρίβεια, η απόδοση του μοντέλου υπονομεύεται από μεγάλες τιμές loss, γεγονός που υποδηλώνει αστάθεια στη μάθηση ή υπερεκπαίδευση σε συγκεκριμένες παραμέτρους. Η πρώτη γραμμή αποτελεί ακραίο παράδειγμα, με loss 60.47, και εμφανώς υψηλότερο από το αντίστοιχο στο Heart dataset (18.96), ενώ συνοδεύεται από αρνητική διαφορά ακρίβειας (-0.019), κάτι που καταδεικνύει καθαρή υποβάθμιση της απόδοσης στο Cardiovascular. Παρόμοια και οι επόμενες περιπτώσεις με batch sizes 64 και 32 και clip=2, διατηρούν σημαντικά μεγαλύτερα loss στο Cardiovascular (53.09 και 28.59 αντίστοιχα) με παράλληλη χαμηλότερη ακρίβεια από το Heart dataset. Οι διαφορές loss σε αυτές τις περιπτώσεις αγγίζουν και ξεπερνούν τις 28 μονάδες.

Η μοναδική περίπτωση όπου το Cardiovascular υπερτερεί οριακά είναι η τέταρτη εγγραφή, όπου ενώ το loss είναι αυξημένο (10.03 έναντι 2.47), η ακρίβεια είναι υψηλότερη κατά 13.1%. Ωστόσο, η συνολική συμπεριφορά του loss εξακολουθεί να καταδεικνύει ότι η χρήση υψηλού clipping (l2_norm_clip = 2) ή batch sizes μεγαλύτερα του 16 οδηγεί σε υπερβολική διασπορά και αποσταθεροποίηση της εκπαίδευσης στο Cardiovascular dataset. Η τελευταία γραμμή, παρά την πολύ χαμηλότερη τιμή loss σε σχέση με τις προηγούμενες, παραμένει υψηλότερη από του Heart και με αρνητική διαφορά ακρίβειας. Οι συγκεκριμένες ρυθμίσεις υπερπαραμέτρων —ιδίως αυτές που περιλαμβάνουν clip=2, μεγάλο batch size, και μεγάλο learning rate— είναι προβληματικές για το Cardiovascular dataset, οδηγώντας σε σημαντικά μεγαλύτερη απώλεια και υποδεέστερη ακρίβεια σε σύγκριση με το Heart dataset.

learning_rate	noise_multiplier	l2_norm_clip	batch_size	test_accuracy_Cardiovascular	test_loss_Cardiovascular	test_accuracy_Heart	test_loss_Heart	accuracy_diff	loss_diff
0.1	2	2	16	0.529874206	60.47868347	0.5489	18.9567	-0.019025794	41.52198347
0.1	2	2	64	0.491852492	53.09628677	0.663	24.8581	-0.171147508	28.23818677

0.1	2	2	32	0.555460274	28.5919342	0.7283	18.8377	-0.172839726	9.7542342
0.1	2	1	16	0.527801573	10.0372963	0.3967	2.4769	0.131101573	7.5603963
0.1	1.1	2	16	0.532804489	5.611526012	0.7717	2.5156	-0.238895511	3.095926012

Πίνακας 6: Top 5 Loss Diff - Πειραματικές Διατάξεις με Ακραία Τιμές Απώλειας: Επιπτώσεις Υπερβολικού Θορύβου και Clipping στη Μάθηση

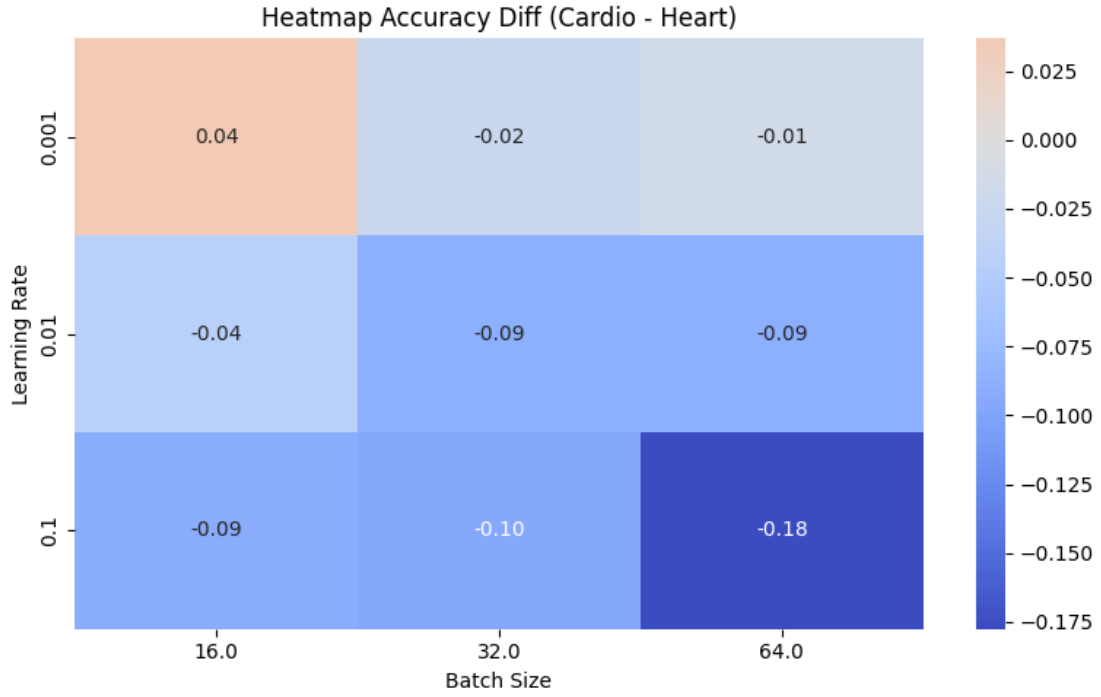
Ο πίνακας 8 περιλαμβάνει ρυθμίσεις υπερπαραμέτρων όπου το μοντέλο στο Cardiovascular dataset παρουσίασε τη χειρότερη συμπεριφορά ως προς τη διαφορά στην απώλεια (loss_diff), δηλαδή, εκεί όπου η απώλεια στο Heart dataset ήταν σημαντικά μεγαλύτερη από του Cardiovascular. Η πρώτη περίπτωση δείχνει μια από τις μεγαλύτερες θετικές διαφορές loss (-3.86), με το Heart να έχει μεγαλύτερο loss (8.79) σε σχέση με το Cardiovascular (4.93). Παρά την καλύτερη συμπεριφορά στο loss, η ακρίβεια του Cardiovascular μοντέλου παραμένει χαμηλότερη από του Heart κατά ~16.7%, υποδεικνύοντας ότι το μοντέλο στο Cardiovascular υπέφερε από μειωμένη ικανότητα γενίκευσης, ακόμα και αν έδειχνε μικρότερο loss.
Αντίθετα, στη δεύτερη γραμμή, το Cardiovascular μοντέλο έχει μεγαλύτερη ακρίβεια (0.5220) σε σχέση με το Heart (0.4293), ενώ εμφανίζει και μικρότερο loss κατά 1.43 μονάδες, κάτι που συνιστά πολύ καλύτερη συμπεριφορά και στις δύο μετρικές. Αυτό δείχνει ότι η συγκεκριμένη παραμετροποίηση (με batch size 32 και clip=1) λειτουργεί ευνοϊκά για το Cardiovascular dataset. Στις υπόλοιπες γραμμές, παρατηρείται γενικά μικρή έως μεσαία υπεροχή του Cardiovascular μοντέλου σε loss, με τις διαφορές να κυμαίνονται από -0.2 έως -0.4, ενώ και η διαφορά στην ακρίβεια είναι είτε μικρή είτε ελαφρώς αρνητική. Αυτό υποδηλώνει ότι το μοντέλο στο Cardiovascular dataset εμφανίζεται πιο "συνεπές" στο loss αλλά όχι απαραίτητα καλύτερο σε συνολική απόδοση. Ο πίνακας αυτός δείχνει ότι το μικρότερο loss στο Cardiovascular dataset δεν συνεπάγεται απαραίτητα καλύτερη ακρίβεια, και ότι ορισμένες ρυθμίσεις επιτρέπουν στα δύο μοντέλα να έχουν ανταγωνιστική απόδοση, με πλεονέκτημα είτε στη μία είτε στην άλλη μετρική. Η πολυπλοκότητα των αποτελεσμάτων τονίζει τη σημασία της ταυτόχρονης αξιολόγησης πολλών δεικτών.

learning_rate	noise_multiplier	l2_norm_clip	batch_size	test_accuracy_Cardiovascular	test_loss_Cardiovascular	test_accuracy_Heart	test_loss_Heart	accuracy_diff	loss_diff
0.1	2	1	64	0.511649489	4.934180737	0.6793	8.7917	-0.167650511	-3.857551926
0.1	2	1	32	0.522084057	5.655946732	0.4293	7.0891	0.092784057	-1.433155327
0.1	2	0.5	64	0.494639784	1.290916324	0.5326	1.7154	-0.037960216	-0.424468368
0.1	0.5	2	16	0.529588342	0.741004705	0.5163	1.1303	0.013288342	-0.389295300
0.001	2	0.5	32	0.494496852	0.715765357	0.3967	0.9538	0.097796852	-0.238003464

Πίνακας 7: Worst 5 Loss Diff - Πειραματικές Ρυθμίσεις με Μέγιστη Αρνητική Διαφορά Loss: Δυσμενείς Επιπτώσεις στην Απόδοση του Μοντέλου

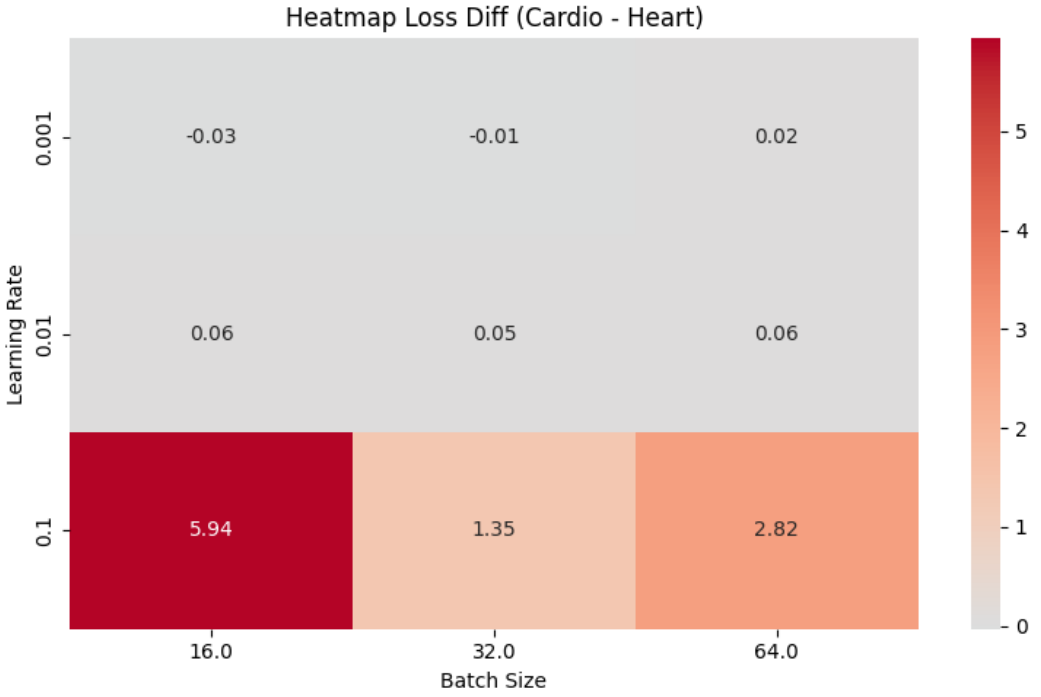
Το Heart dataset είναι πιο σταθερό και ανθεκτικό σε επιθετικές υπερπαραμέτρους. Το Cardiovascular έχει δυναμική, αλλά είναι ευαίσθητο και υποφέρει σε δύσκολες συνθήκες.
Το heatmap της διαφοράς ακρίβειας (Accuracy Diff) δείχνει πως για χαμηλό learning rate (0.001), οι διαφορές μεταξύ Cardiovascular και Heart dataset είναι αμελητέες ή ελαφρώς

υπέρ του Cardiovascular, ειδικά όταν το batch size είναι 16. Καθώς το learning rate αυξάνεται σε 0.01, η εικόνα αρχίζει να γίνεται πιο αρνητική για το Cardiovascular, με χειρότερη απόδοση σε μεγαλύτερα batch sizes. Στο υψηλότερο learning rate (0.1), το Cardiovascular παρουσιάζει εμφανή πτώση απόδοσης σε σχέση με το Heart dataset σε όλα τα batch sizes, με τη χειρότερη διαφορά στο batch_size=64. Αυτό δείχνει ότι η ακρίβεια του Cardiovascular επιδεινώνεται όσο γίνονται πιο επιθετικές οι υπερπαράμετροι, κάτι που το Heart dataset αντέχει καλύτερα.



Εικόνα 13: Επίδραση του Learning Rate και του Batch Size στη Διαφορά Ακρίβειας (Cardio - Heart)

Το heatmap της διαφοράς στο loss (Loss Diff) αναδεικνύει ακόμη πιο καθαρά την ευαισθησία του Cardiovascular dataset. Όταν το learning rate είναι 0.001 ή 0.01, τα δύο datasets έχουν σχεδόν παρόμοιο loss με ελάχιστες διαφορές. Μόλις όμως το learning rate φτάσει το 0.1, το Cardiovascular dataset εμφανίζει εκθετικά μεγαλύτερο loss σε όλα τα batch sizes. Η διαφορά φτάνει τις 5.94 μονάδες στο batch size 16, 1.35 στο batch size 32 και 2.82 στο batch size 64 — κάτι που υποδεικνύει αποσταθεροποίηση ή αποτυχία εκμάθησης. Το Heart dataset, αντιθέτως, διατηρεί πιο χαμηλό και συνεπές loss σε όλα τα σενάρια. Το Heart dataset φαίνεται να είναι πολύ πιο σταθερό, προσαρμόσιμο και ανθεκτικό απέναντι σε πιο έντονες παραμετροποιήσεις. Αντίθετα, το Cardiovascular dataset λειτουργεί ικανοποιητικά μόνο υπό συντηρητικές ρυθμίσεις και καταρρέει όταν πιεστεί με υψηλό learning rate και μεγαλύτερο batch size. Η στρατηγική χρήσης του Cardiovascular dataset πρέπει να δίνει έμφαση στη σταθερότητα των υπερπαραμέτρων και ενδεχομένως να αξιοποιεί επιπλέον τεχνικές regularization.



Εικόνα 14: Επίδραση του Learning Rate και του Batch Size στη Διαφορά Loss (Cardio - Heart).

8.3 Συζήτηση σχετικά με την ισορροπία μεταξύ ιδιωτικότητας και χρησιμότητας στη μηχανική μάθηση.

Η ισορροπία μεταξύ ιδιωτικότητας και χρησιμότητας στη μηχανική μάθηση αποτελεί μία από τις σημαντικότερες προκλήσεις στην επεξεργασία ευαίσθητων δεδομένων, ιδιαίτερα στον τομέα της υγειονομικής περίθαλψης. Από τη μία πλευρά, η προστασία των προσωπικών δεδομένων των ασθενών είναι κρίσιμη, τόσο για νομικούς όσο και για ηθικούς λόγους. Από την άλλη, η αποτελεσματικότητα των αλγορίθμων μηχανικής μάθησης βασίζεται σε μεγάλο βαθμό στην ακρίβεια και πληρότητα των δεδομένων, γεγονός που σημαίνει ότι η υπερβολική προστασία μπορεί να περιορίσει τη χρησιμότητα των μοντέλων. Τα πειράματα που πραγματοποιήθηκαν με χρήση διαφορικής ιδιωτικότητας (DP) μέσω του αλγορίθμου DP-SGD ανέδειξαν αυτή τη λεπτή ισορροπία. Παρατηρήθηκε ότι η αύξηση του noise multiplier – της βασικής υπερπαραμέτρου που ελέγχει το επίπεδο ιδιωτικότητας – οδηγεί σε αισθητή μείωση της ακρίβειας του μοντέλου. Ειδικά στο Cardiovascular Disease Dataset, τιμή noise multiplier = 2.0 προκάλεσε σημαντική αύξηση της απώλειας (μέση τιμή ≈ 6.67) και μείωση της ακρίβειας σε επίπεδα περίπου 50.45%, καταδεικνύοντας ότι η ισχυρή προστασία ιδιωτικότητας μπορεί να υποβαθμίσει σοβαρά τη χρησιμότητα του μοντέλου. Αντίθετα, στο Heart Disease Dataset, που είναι μικρότερο και πιο καθαρό, επιτεύχθηκε καλύτερη ισορροπία όταν το noise multiplier κυμαινόταν μεταξύ 0.5 και 1.1, με τη μέγιστη ακρίβεια να φτάνει περίπου 59.20%. Το γεγονός αυτό υποδηλώνει ότι το μέγεθος και η ποιότητα του dataset παίζουν καθοριστικό ρόλο στον καθορισμό του «ανεκτού» επιπέδου θορύβου.

Η επιλογή των υπερπαραμέτρων αποδείχθηκε καθοριστική για τη διατήρηση αυτής της ισορροπίας. Χαμηλές τιμές clipping norm (0.5) συνέβαλαν στη σταθερότητα του μοντέλου, ενώ μεγαλύτερες τιμές (2.0) συνδέθηκαν με αυξημένη απώλεια και ασταθείς προβλέψεις.

Αντίστοιχα, learning rate κοντά στο 0.1 απέδωσε καλύτερα όταν συνδυάστηκε με μέτριο επίπεδο θορύβου, ενώ ακραίες τιμές προκάλεσαν αστάθεια ή καθυστέρηση σύγκλισης.
Συνεπώς, δεν υπάρχει μία ενιαία βέλτιστη προσέγγιση που να ισχύει για όλα τα datasets ή εφαρμογές. Η διαφορική ιδιωτικότητα μπορεί να εφαρμοστεί επιτυχώς σε δεδομένα υγείας, αλλά απαιτεί εξατομικευμένη ρύθμιση ανάλογα με τις ανάγκες και τους περιορισμούς κάθε περίπτωσης. Σε εφαρμογές όπου η ακρίβεια είναι κρίσιμη (π.χ. διάγνωση καρδιαγγειακής νόσου), η υπερβολική ιδιωτικότητα μπορεί να οδηγήσει σε ανεπιθύμητα αποτελέσματα. Αντίθετα, σε περιπτώσεις όπου η εμπιστευτικότητα είναι πρωταρχικής σημασίας (π.χ. δημόσια αποθήκευση ιατρικών δεδομένων), μία μετριοπαθής απώλεια ακρίβειας μπορεί να είναι αποδεκτή. Τα αποτελέσματα των πειραμάτων δείχνουν ότι η εύρεση της κατάλληλης ισορροπίας μεταξύ ιδιωτικότητας και χρησιμότητας παραμένει μια ανοιχτή ερευνητική πρόκληση. Η συνεχής προσαρμογή των υπερπαραμέτρων, ανάλογα με την αρχιτεκτονική του μοντέλου και τα χαρακτηριστικά των δεδομένων, είναι απαραίτητη για τη διασφάλιση επαρκούς προστασίας χωρίς να θυσιάζεται η λειτουργικότητα. Αυτό είναι ιδιαίτερα κρίσιμο στο πεδίο της υγειονομικής περίθαλψης, όπου η ακρίβεια των αλγορίθμων δεν είναι απλώς ζήτημα επίδοσης, αλλά δύναται να επηρεάσει άμεσα την ασφάλεια και τις αποφάσεις περίθαλψης των ασθενών.

8.4 Μελλοντικές κατευθύνσεις και μεθοδολογικοί περιορισμοί.

Οι μελλοντικές κατευθύνσεις για την έρευνα και την ανάπτυξη στον τομέα της διαφορικής ιδιωτικότητας και της μηχανικής μάθησης σε δεδομένα υγείας επικεντρώνονται στην εξεύρεση βέλτιστων πρακτικών που εξισορροπούν την προστασία της ιδιωτικότητας με τη χρησιμότητα των μοντέλων. Τα πειράματα που πραγματοποιήθηκαν με τα δύο datasets, το Heart Disease Dataset και το Cardiovascular Disease Dataset, ανέδειξαν τις προκλήσεις που προκύπτουν όταν εφαρμόζονται τεχνικές διαφορικής ιδιωτικότητας, καθώς και την ανάγκη για στοχευμένες προσεγγίσεις προσαρμοσμένες στη φύση των δεδομένων.
Ένας βασικός τομέας που απαιτεί περαιτέρω έρευνα είναι η βελτίωση των αλγορίθμων που χρησιμοποιούν διαφορική ιδιωτικότητα ώστε να ελαχιστοποιούν την επίδραση της προσθήκης θορύβου στην ακρίβεια των μοντέλων. Τα αποτελέσματα έδειξαν ότι η αύξηση του noise multiplier μειώνει την ακρίβεια των προβλέψεων, γεγονός που υποδηλώνει την ανάγκη για πιο εξελιγμένες μεθόδους διαχείρισης θορύβου. Μελλοντικές έρευνες μπορούν να επικεντρωθούν στην ανάπτυξη προσαρμοστικών τεχνικών clipping και noise scheduling, οι οποίες θα προσαρμόζονται δυναμικά ανάλογα με τα χαρακτηριστικά των δεδομένων και το εκάστοτε πρόβλημα.
Ένας άλλος σημαντικός τομέας ανάπτυξης αφορά την εφαρμογή τεχνικών federated learning, όπου τα δεδομένα παραμένουν στις πηγές τους και η εκπαίδευση πραγματοποιείται αποκεντρωμένα. Αυτή η προσέγγιση μπορεί να μειώσει την ανάγκη για αυστηρό θόρυβο διαφορικής ιδιωτικότητας, καθώς τα δεδομένα των χρηστών δεν μεταφέρονται σε ένα κεντρικό αποθετήριο. Τα δεδομένα υγείας, όπως αυτά που χρησιμοποιήθηκαν στα πειράματα, μπορούν να επωφεληθούν από τέτοιες τεχνικές, μειώνοντας την ανάγκη για άμεση πρόσβαση στις πρωτογενείς πληροφορίες των ασθενών. Επιπλέον, η χρήση πιο ανθεκτικών και εξελιγμένων νευρωνικών αρχιτεκτονικών μπορεί να βοηθήσει στη μείωση των επιπτώσεων του θορύβου. Τα αποτελέσματα έδειξαν ότι απλές αρχιτεκτονικές πλήρως συνδεδεμένων νευρωνικών δικτύων είναι ευάλωτες σε υψηλές τιμές noise multiplier και clipping. Μελλοντικές μελέτες μπορούν να διερευνήσουν τη χρήση deep learning μοντέλων που έχουν ενσωματωμένη ανθεκτικότητα στην παραμόρφωση των δεδομένων, όπως attention-based networks ή μοντέλα που ενσωματώνουν αβεβαιότητα στις προβλέψεις τους. Η ερευνητική κοινότητα θα πρέπει επίσης να εξετάσει την ανάπτυξη

βέλτιστων μεθόδων επιλογής υπερπαραμέτρων για αλγορίθμους διαφορικής ιδιωτικότητας. Τα πειράματα ανέδειξαν τη μεγάλη επίδραση της τιμής του learning rate, του batch size και του clipping στο τελικό αποτέλεσμα. Η αυτοματοποιημένη αναζήτηση των βέλτιστων παραμέτρων μέσω meta-learning ή Bayesian optimization θα μπορούσε να βελτιώσει σημαντικά την ακρίβεια των μοντέλων χωρίς την ανάγκη εκτεταμένων πειραμάτων. Τέλος, η συζήτηση για τη διαφορική ιδιωτικότητα και την αποδοτικότητα των μοντέλων δεν είναι μόνο τεχνικής φύσης, αλλά επηρεάζει και τη ρύθμιση των δεδομένων υγείας και τις πολιτικές προστασίας προσωπικών δεδομένων. Μελλοντικές μελέτες μπορούν να επικεντρωθούν στον τρόπο με τον οποίο οι νομοθεσίες, όπως το GDPR και το HIPAA, μπορούν να συνδυαστούν με τις σύγχρονες μεθόδους μηχανικής μάθησης, ώστε να δημιουργηθεί ένα ισορροπημένο και εφαρμόσιμο πλαίσιο χρήσης των δεδομένων υγείας.

Μία από τις βασικές μελλοντικές κατευθύνσεις που προκύπτουν από την παρούσα εργασία αφορά τη διερεύνηση εναλλακτικών παραμετροποιήσεων και τεχνικών που θα επιτρέψουν την πλήρη εφαρμογή per-example clipping, με στόχο την ενίσχυση των εγγυήσεων διαφορικής ιδιωτικότητας κατά την εκπαίδευση νευρωνικών δικτύων. Η χρήση του per-example clipping θεωρείται κρίσιμη, καθώς διασφαλίζει ότι κάθε δείγμα στη διαδικασία εκπαίδευσης αντιμετωπίζεται ανεξάρτητα, επιτρέποντας την ακριβή μέτρηση της ευαισθησίας της συνάρτησης απώλειας και την εφαρμογή εξατομικευμένου θορύβου σε κάθε κλίση. Αντιθέτως, η χρήση batch-level clipping, όπως συμβαίνει όταν ορίζεται num_microbatches=1, συγκεντρώνει όλες τις διαβαθμίσεις ενός batch σε ένα ενιαίο μέσο όρο, καθιστώντας ασαφή τη συνεισφορά του κάθε δείγματος και οδηγώντας σε πιο ασθενείς εγγυήσεις ιδιωτικότητας, ακόμη και όταν οι αριθμητικές τιμές του privacy budget (ε) φαίνονται ικανοποιητικές.

Στο πλαίσιο αυτό, προτείνεται η αξιολόγηση παραμετροποιήσεων όπου το num_microbatches ορίζεται ίσο με το μέγεθος του batch, ώστε κάθε παράδειγμα να λειτουργεί ως ανεξάρτητο μικροπακέτο. Αν και η υλοποίηση αυτή ενδέχεται να αυξήσει τον υπολογιστικό φόρτο, προσφέρει σημαντικά οφέλη στην ακρίβεια του λογαριασμού ιδιωτικότητας και στη σταθερότητα της εκπαίδευσης. Παράλληλα, η χρήση εξειδικευμένων βιβλιοθηκών όπως το Opacus του PyTorch, το οποίο υποστηρίζει εγγενώς per-example clipping και παρέχει εργαλεία για αναλυτικό privacy accounting, συνιστά μια πολλά υποσχόμενη κατεύθυνση. Το Opacus σχεδιάστηκε με γνώμονα την πρακτική εφαρμογή της διαφορικής ιδιωτικότητας σε πραγματικά δεδομένα, διατηρώντας υψηλή υπολογιστική απόδοση και μειώνοντας τον προγραμματιστικό φόρτο που απαιτείται για την ενσωμάτωσή του σε υπάρχουσες ροές εκπαίδευσης.

Επομένως, η ενσωμάτωση αυτών των μεθόδων σε μελλοντικά πειράματα αναμένεται να βελτιώσει τη συμβατότητα με αυστηρές απαιτήσεις ιδιωτικότητας, να επιτρέψει πιο αξιόπιστη σύγκριση διαφορετικών privacy budgets και να συμβάλει στην καλύτερη κατανόηση του trade-off μεταξύ ιδιωτικότητας και απόδοσης, ειδικά σε τομείς υψηλής σημασίας όπως η υγεία και η βιοϊατρική ανάλυση. Τέτοιες επεκτάσεις δεν θα ενισχύσουν μόνο τη θεωρητική εγκυρότητα των αποτελεσμάτων, αλλά θα προσεγγίσουν καλύτερα τις απαιτήσεις πραγματικών εφαρμογών όπου η προστασία των προσωπικών δεδομένων είναι θεμελιώδης. Η συνεχής έρευνα στον τομέα της μηχανικής μάθησης με διαφορική ιδιωτικότητα είναι απαραίτητη για την προώθηση λύσεων που επιτρέπουν την αξιόπιστη χρήση ευαίσθητων δεδομένων, χωρίς να θυσιάζεται η ακρίβεια των συστημάτων πρόβλεψης και ανάλυσης. Οι βελτιώσεις στους αλγορίθμους, οι νέες τεχνικές εκπαίδευσης και η προσαρμογή των μεθόδων στις ανάγκες της υγειονομικής περίθαλψης θα καθορίσουν την επιτυχία της εφαρμογής της διαφορικής ιδιωτικότητας σε πραγματικά σενάρια.

Πέραν των ανωτέρω, αξίζει να επισημανθούν και συγκεκριμένοι μεθοδολογικοί περιορισμοί που εντοπίστηκαν κατά τη διάρκεια της παρούσας μελέτης. Πρώτον, όλα τα πειράματα υλοποιήθηκαν με βάση batch-level clipping, καθώς χρησιμοποιήθηκε num_microbatches = 1, με αποτέλεσμα το clipping να εφαρμόζεται στο άθροισμα των βαθμίδων κατά συνέπεια, η εκτιμώμενη τιμή του ε ενδέχεται να είναι αισιόδοξη. Η θεωρητικά ακριβής προσέγγιση προϋποθέτει per-example clipping, δηλαδή num_microbatches ίσο με το batch_size. Δεύτερον, ο υπολογισμός του ζεύγους ε–δ πραγματοποιήθηκε με το εργαλείο compute_dp_sgd_privacy, το οποίο βασίζεται στην υπόθεση Poisson δειγματοληψίας ωστόσο, η παρούσα μελέτη υιοθέτησε uniform sampling ανά epoch, γεγονός που πιθανώς αυξάνει το πραγματικό ε κατά 30–60 %. Στο μέλλον, ενδείκνυται η αντικατάσταση της μεθόδου με τον RDP accountant. Τρίτον, στο σύνολο Cardiovascular Disease Dataset τα ονομαστικά batch sizes των 32 και 64 πρακτικά στρογγυλοποιήθηκαν σε πλήρες batch, γεγονός που καθιστά τις αντίστοιχες τιμές ε μη άμεσα συγκρίσιμες με εκείνες για batch = 16. Επιπλέον, υιοθετήθηκε σταθερή τιμή δ = 1/(2n), ενώ σε προηγούμενες δοκιμές είχαν χρησιμοποιηθεί και διαφορετικές τιμές. Όλα τα ε που παρουσιάζονται στα τελικά αποτελέσματα της διατριβής βασίζονται πλέον αποκλειστικά σε αυτή τη συγκεκριμένη παραδοχή. Αξίζει επίσης να σημειωθεί ότι κάθε συνδυασμός υπερπαραμέτρων εκπαιδεύτηκε μόνο με μία τυχαία αρχικοποίηση, οπότε η επίδραση της στοχαστικής μεταβλητότητας ενδέχεται να υποεκτιμάται. Η παρατηρούμενη AUC ≈ 0,49 στο Cardiovascular Disease Dataset πιθανώς υποδηλώνει αντιστροφή ετικετών ή σημαντική ανισορροπία μεταξύ των κλάσεων, γεγονός που περιορίζει τη δυνατότητα εξαγωγής γενικευμένων συμπερασμάτων από το συγκεκριμένο σύνολο. Τέλος, για λόγους ευκολότερης ερμηνείας των αριθμητικών τιμών, προτείνεται να θεωρείται ότι τιμές $\varepsilon \leq 2$ υποδηλώνουν ισχυρή ιδιωτικότητα, $2 < \varepsilon \leq 10$ μέτρια και $\varepsilon > 10$ ασθενή προστασία, κάτι που τονίζει τη σημασία επανυπολογισμού του ε πριν από οποιαδήποτε κανονιστική εφαρμογή.

8.5 Το EU AI Act και η Προστασία Ιδιωτικότητας στην Ιατρική Τεχνητή Νοημοσύνη

Ο Κανονισμός για την Τεχνητή Νοημοσύνη της Ευρωπαϊκής Ένωσης (EU AI Act) αποτελεί την πρώτη νομοθετική προσπάθεια παγκοσμίως για την ολιστική ρύθμιση της Τεχνητής Νοημοσύνης. Ο κανονισμός αυτός επιδιώκει να διασφαλίσει ότι τα συστήματα ΤΝ που αναπτύσσονται και χρησιμοποιούνται στην ΕΕ είναι ασφαλή, διαφανή, ανιχνεύσιμα, μη μεροληπτικά και φιλικά προς το περιβάλλον, θέτοντας ταυτόχρονα όρια ανάλογα με το επίπεδο κινδύνου που ενέχουν. Στο πλαίσιο αυτής της εργασίας, ο EU AI Act αποκτά ιδιαίτερη σημασία, καθώς τα συστήματα μηχανικής μάθησης που χρησιμοποιούνται για την ανάλυση ευαίσθητων ιατρικών δεδομένων, όπως τα δεδομένα καρδιολογικών παθήσεων, εντάσσονται στην κατηγορία των «υψηλού κινδύνου» εφαρμογών. Αυτό συνεπάγεται αυξημένες απαιτήσεις ως προς τη διαχείριση της ιδιωτικότητας, τη διαφάνεια των αλγορίθμων και την αξιοπιστία των προβλέψεων. Η χρήση του αλγορίθμου DP-SGD στην παρούσα εργασία ευθυγραμμίζεται με τις απαιτήσεις του EU AI Act, καθώς ενσωματώνει εγγενείς μηχανισμούς προστασίας της ιδιωτικότητας, όπως η διαφορική ιδιωτικότητα, η οποία μειώνει τον κίνδυνο αναγνώρισης ατομικών δεδομένων. Ο κανονισμός ενισχύει τη νομική και δεοντολογική αναγκαιότητα για τεχνικές όπως το DP-SGD, οι οποίες συμβάλλουν στη συμμόρφωση με το άρθρο που αφορά την ελαχιστοποίηση κινδύνου, την επεξήγηση των αποτελεσμάτων και την πρόληψη των διακρίσεων μέσω της ΤΝ. Επιπλέον, η απαιτούμενη τεκμηρίωση των datasets, η παρακολούθηση της απόδοσης και η διασφάλιση της ακρίβειας των μοντέλων είναι πτυχές που ενισχύονται μέσω των

πειραμάτων βελτιστοποίησης υπερπαραμέτρων που παρουσιάζονται στη μελέτη αυτή. Η υιοθέτηση του DP-SGD ως μεθοδολογίας προστασίας ιδιωτικότητας σε ιατρικά δεδομένα ανταποκρίνεται στις απαιτήσεις του νέου ρυθμιστικού πλαισίου, προσφέροντας ένα πρακτικό παράδειγμα τεχνικής συμμόρφωσης με τον EU AI Act, και ενισχύοντας την εμπιστοσύνη των πολιτών στη χρήση της τεχνητής νοημοσύνης στην υγειονομική περίθαλψη.

Η παρούσα εργασία δεν περιορίζεται μόνο στη θεωρητική αξιολόγηση της ιδιωτικότητας, αλλά αποτελεί και ένα λειτουργικό παράδειγμα ενσωμάτωσης των αρχών του EU AI Act στην πράξη. Μέσω της συστηματικής διερεύνησης των επιπτώσεων διαφορετικών παραμετροποιήσεων στον αλγόριθμο DP-SGD, αναδεικνύεται η αναγκαιότητα για διαφάνεια στη ρύθμιση των μοντέλων, στοιχείο που αναγνωρίζεται από τον κανονισμό ως κρίσιμο για τις εφαρμογές υψηλού κινδύνου. Παράλληλα, η εμπειρική ανάλυση που βασίστηκε σε οπτικοποίηση της απόδοσης και συγκριτικούς πίνακες συνιστά βέλτιστη πρακτική για την τεκμηρίωση και αξιολόγηση της συμπεριφοράς των συστημάτων, όπως απαιτείται από τον EU AI Act. Με βάση τα ευρήματα, υποστηρίζεται ότι η συμμόρφωση με τον κανονισμό μπορεί να επιτευχθεί όχι μόνο μέσω νομικής προσέγγισης, αλλά και μέσω τεχνικής αυστηρότητας και ποσοτικής αξιολόγησης, επιτρέποντας έτσι την υπεύθυνη και ασφαλή χρήση της ΤΝ στον ιατρικό τομέα.

9. Συμπέρασματα

Η παρούσα μελέτη επιβεβαιώνει ότι η ενσωμάτωση διαφορικής ιδιωτικότητας μέσω του αλγορίθμου DP-SGD σε συστήματα μηχανικής μάθησης για δεδομένα υγείας είναι τεχνικά εφικτή, αλλά απαιτεί προσεκτική και εξατομικευμένη ρύθμιση. Οι πειραματικές αξιολογήσεις έδειξαν ξεκάθαρα ότι οι επιδόσεις των μοντέλων είναι ιδιαίτερα ευαίσθητες στις υπερπαραμέτρους που καθορίζουν τον βαθμό ιδιωτικότητας. Η αύξηση του noise multiplier, παρότι ενισχύει την προστασία των προσωπικών δεδομένων, συνοδεύεται από μείωση της ακρίβειας και αύξηση της απώλειας. Το φαινόμενο αυτό είναι πιο έντονο στο Heart Disease Dataset, όπου το μικρό μέγεθος και η περιορισμένη ποικιλία των δεδομένων δεν επαρκούν για να «απορροφήσουν» τον θόρυβο, οδηγώντας σε χαμηλότερες επιδόσεις όταν εφαρμόζεται έντονη προστασία. Στον αντίποδα, το Cardiovascular Disease Dataset, λόγω του μεγαλύτερου όγκου δεδομένων, επέτρεψε στο μοντέλο να διατηρήσει την ακρίβεια σε υψηλότερα επίπεδα, φτάνοντας έως και 60.90%, ακόμη και υπό συνθήκες υψηλού θορύβου. Η επεξεργασία ενός τόσο εκτεταμένου και πλούσιου dataset ανέδειξε την ικανότητα της DP-SGD να παραμένει λειτουργική, αρκεί να υπάρχει αρκετή πληροφόρηση ώστε να μην επικρατεί το σήμα του θορύβου έναντι του σήματος των δεδομένων.

Η ανάλυση έδειξε επίσης ότι το clipping norm και το batch size παίζουν σημαντικό ρόλο στη σταθερότητα του αλγορίθμου. Πολύ υψηλές τιμές clipping (π.χ. 2.0) συνδέθηκαν με ακραίες τιμές loss, ενώ μεσαίες τιμές (0.5–1.0) επέτρεψαν πιο ισορροπημένη μάθηση. Επιπλέον, μικρότερα batch sizes (16 ή 32) οδήγησαν σε πιο σταθερές επιδόσεις, χωρίς να αυξάνεται υπερβολικά το κόστος υπολογισμού του εμβέλειας ιδιωτικότητας (ε).

Η μελέτη καταδεικνύει ότι η επιτυχής χρήση της διαφορικής ιδιωτικότητας στη μηχανική μάθηση για ευαίσθητα δεδομένα, όπως τα ιατρικά, εξαρτάται από τη λεπτή εξισορρόπηση μεταξύ ιδιωτικότητας και χρησιμότητας. Η DP-SGD αποτελεί ένα ισχυρό εργαλείο για περιβάλλοντα υψηλής ευαισθησίας, αλλά δεν είναι μια λύση «μιας χρήσης». Η ρύθμιση των υπερπαραμέτρων θα πρέπει να καθορίζεται κατά περίπτωση, με βάση την πολυπλοκότητα του dataset και τις απαιτήσεις της εφαρμογής. Στο πεδίο της υγειονομικής περίθαλψης, όπου

η ακρίβεια επηρεάζει άμεσα την ποιότητα φροντίδας, η εύρεση αυτής της ισορροπίας είναι ιδιαίτερα κρίσιμη.

Ωστόσο, η προστασία της ιδιωτικότητας συνοδεύεται από υπολογιστικές και αποδοτικές προκλήσεις. Τα πειραματικά αποτελέσματα από τα Heart Disease και Cardiovascular Disease datasets έδειξαν ότι η χρήση υψηλού πολλαπλασιαστή θορύβου (noise multiplier) οδηγεί σε σημαντική μείωση της ακρίβειας, καθώς η ποιότητα της πληροφορίας που μαθαίνει το μοντέλο υποβαθμίζεται λόγω της διατάραξης των gradients. Για παράδειγμα, με noise multiplier 2.0 και learning rate 0.1, η ακρίβεια στο Cardiovascular Dataset έπεσε κάτω από 50%, ενώ το test loss αυξήθηκε απότομα, αγγίζοντας τιμές άνω του 12 σε ορισμένες περιπτώσεις. Αντίθετα, χαμηλότερες τιμές θορύβου (π.χ. 0.5) επέτρεψαν στο μοντέλο να διατηρήσει υψηλότερη ακρίβεια (έως και 60.90%), αλλά με μικρότερο βαθμό ιδιωτικότητας ($\varepsilon \approx 7.34$). Αυτό καταδεικνύει την εγγενή ανάγκη εξισορρόπησης ανάμεσα στην προστασία της ιδιωτικότητας και την ακρίβεια των προβλέψεων.

Τα αποτελέσματα επίσης υποδεικνύουν ότι ο αλγόριθμος DP-SGD είναι πιο αποτελεσματικός σε μεγαλύτερα datasets, όπου ο προστιθέμενος θόρυβος έχει μικρότερη σχετική επίδραση. Στην περίπτωση του Cardiovascular Disease Dataset, το οποίο περιλαμβάνει πάνω από 70.000 εγγραφές, τα μοντέλα που εκπαιδεύτηκαν με DP-SGD εμφάνισαν μεγαλύτερη σταθερότητα από ό,τι στο μικρότερο Heart Disease Dataset. Αντίστοιχα, το μικρότερο dataset παρουσίασε μεγαλύτερη ευαισθησία στην επιλογή υπερπαραμέτρων, κάτι που υποδηλώνει πως η επιτυχής εφαρμογή της διαφορικής ιδιωτικότητας σε μικρά σύνολα δεδομένων απαιτεί πιο αυστηρή παραμετροποίηση, ενισχυμένη προεπεξεργασία ή ακόμη και υβριδικές τεχνικές ιδιωτικότητας.

Η χρήση συγκριτικών πινάκων και γραφημάτων ανέδειξε χαρακτηριστικά μοτίβα στη συμπεριφορά των μοντέλων υπό διαφορετικές συνθήκες. Η κατανομή της διαφοράς ακρίβειας μεταξύ των δύο datasets έδειξε ότι, ενώ υπάρχουν περιπτώσεις όπου το Cardiovascular Dataset αποδίδει καλύτερα, η πλειονότητα των ρυθμίσεων υπερπαραμέτρων καταλήγει σε υψηλότερη ακρίβεια για το Heart Dataset. Ακόμη πιο έντονη ήταν η διαφορά στο loss: συγκεκριμένοι συνδυασμοί (π.χ. learning rate 0.1 με clip 2.0) στο Cardiovascular Dataset οδήγησαν σε απότομη αύξηση της απώλειας, κάτι που δεν παρατηρήθηκε στο Heart Dataset. Η αξιοποίηση heatmaps επέτρεψε την αποσαφήνιση του τρόπου με τον οποίο το learning rate και το batch size επηρεάζουν την επίδοση. Παρατηρήθηκε ότι για learning rate 0.1, ανεξάρτητα από το batch size, η απόδοση του Cardiovascular Dataset υποβαθμίζεται αισθητά σε ακρίβεια και loss. Αντιθέτως, για τιμές 0.001 και 0.01, και ειδικά με μικρότερα batch sizes, το μοντέλο αποδίδει πιο σταθερά, πολλές φορές ξεπερνώντας ακόμη και το Heart Dataset. Η εμπειρική αυτή παρατήρηση επιβεβαιώνει ότι η "μια ρύθμιση για όλα τα δεδομένα" δεν είναι αποτελεσματική στην πράξη και υποδεικνύει την ανάγκη για οπτικά υποστηριζόμενη και dataset-εξαρτώμενη επιλογή υπερπαραμέτρων.

Η συνολική εικόνα που προκύπτει είναι ότι η επιτυχής ενσωμάτωση της διαφορικής ιδιωτικότητας στη μηχανική μάθηση εξαρτάται όχι μόνο από την επιλογή του αλγορίθμου αλλά και από μια ευρύτερη στρατηγική: απαιτείται επαρκής ποσότητα δεδομένων, ρεαλιστική παραμετροποίηση με εμπειρική υποστήριξη και αξιοποίηση εργαλείων διερεύνησης όπως heatmaps και confusion matrices. Η DP-SGD είναι ένα ισχυρό εργαλείο, αλλά η αποτελεσματική χρήση της προϋποθέτει βαθιά κατανόηση της αλληλεπίδρασης μεταξύ ιδιωτικότητας και απόδοσης, ειδικά σε εφαρμογές υψηλής σημασίας όπως η υγειονομική περίθαλψη.

Η συνολική εικόνα που προκύπτει είναι ότι η επιτυχής ενσωμάτωση της διαφορικής ιδιωτικότητας στη μηχανική μάθηση δεν εξαρτάται αποκλειστικά από την επιλογή αλγορίθμου, αλλά απαιτεί μια ευρύτερη στρατηγική που περιλαμβάνει επάρκεια και

ποιότητα δεδομένων, ορθολογική ρύθμιση υπερπαραμέτρων και αξιοποίηση εργαλείων διερεύνησης, όπως οι συγκριτικοί πίνακες, τα heatmaps και οι γραφικές απεικονίσεις απόδοσης. Μέσα από αυτά τα εργαλεία, έγινε σαφές ότι η επίδραση της διαφορικής ιδιωτικότητας ποικίλλει ανάλογα με το dataset, το μοντέλο και τις συνθήκες εκπαίδευσης.

Η DP-SGD αποδεικνύεται ένα ισχυρό εργαλείο για την ενσωμάτωση προστασίας ιδιωτικότητας στα μοντέλα μηχανικής μάθησης, ειδικά σε ευαίσθητα δεδομένα όπως αυτά της υγειονομικής περίθαλψης. Ωστόσο, η αποδοτική χρήση της απαιτεί βαθιά κατανόηση της αλληλεπίδρασης μεταξύ ιδιωτικότητας και χρησιμότητας. Τα πειράματα ανέδειξαν ότι η ακρίβεια των προβλέψεων μπορεί να επηρεαστεί έντονα από την προσθήκη θορύβου, ειδικά σε μικρά σύνολα δεδομένων, ενώ η επιλογή ακατάλληλων υπερπαραμέτρων ενδέχεται να οδηγήσει σε αστάθεια κατά την εκπαίδευση.

Η χρήση του DP-SGD στα Heart Disease και Cardiovascular Disease Datasets κατέδειξε ότι η ισορροπία μεταξύ ιδιωτικότητας και απόδοσης εξαρτάται όχι μόνο από το εύρος των δεδομένων αλλά και από το πόσο καλά έχει παραμετροποιηθεί το μοντέλο. Στο Cardiovascular Dataset, που περιέχει μεγαλύτερο όγκο δεδομένων, η DP-SGD ήταν πιο σταθερή και αποτελεσματική, σε αντίθεση με το Heart Dataset, όπου η ευαισθησία στις ρυθμίσεις ήταν αυξημένη. Παρατηρήθηκε ότι για υψηλό learning rate (π.χ. 0.1), το loss στο Cardiovascular Dataset αυξανόταν απότομα, γεγονός που αποτυπώθηκε καθαρά σε heatmaps και διαγράμματα απώλειας. Αντίθετα, μικρότερα learning rates και μέτριος θόρυβος παρήγαγαν πιο σταθερά αποτελέσματα.

Οι επαγγελματίες στον χώρο της υγείας και των δεδομένων καλούνται να κατανοήσουν ότι η διαφορική ιδιωτικότητα ενδέχεται να έχει σημαντικές επιπτώσεις στην προβλεπτική ικανότητα των μοντέλων. Συνιστάται η εφαρμογή συστηματικής αναζήτησης υπερπαραμέτρων και η αξιολόγηση πολλαπλών παραμετροποιήσεων, ώστε να εξευρεθεί η κατάλληλη ισορροπία μεταξύ ακρίβειας και προστασίας προσωπικών δεδομένων. Σε μικρότερα datasets, ενδείκνυται η διερεύνηση υβριδικών τεχνικών, όπως ο συνδυασμός DP με federated learning ή secure multiparty computation, προκειμένου να μειωθεί η επίδραση του θορύβου. Οι υπεύθυνοι χάραξης πολιτικής θα πρέπει να διαμορφώσουν σύγχρονα ρυθμιστικά πλαίσια που ενσωματώνουν αρχές διαφορικής ιδιωτικότητας στη χρήση ευαίσθητων ιατρικών δεδομένων, εξασφαλίζοντας ότι η εφαρμογή αυτών των τεχνικών δεν υπονομεύει την ποιότητα των ιατρικών αποφάσεων. Παράλληλα, η ενίσχυση της διαφάνειας και η ενημέρωση των ασθενών για τη χρήση ανωνυμοποιημένων δεδομένων είναι απαραίτητες για τη διατήρηση της εμπιστοσύνης. Η συνεργασία μεταξύ της επιστημονικής κοινότητας, των κυβερνητικών φορέων και της βιομηχανίας τεχνολογίας είναι κρίσιμη για την ανάπτυξη βέλτιστων πρακτικών. Με τις εξελίξεις στην τεχνητή νοημοσύνη και την αυξανόμενη ανάγκη για ασφαλή επεξεργασία ευαίσθητων πληροφοριών, τεχνικές όπως η DP-SGD θα πρέπει να αποτελέσουν βασικό πυλώνα για τη μελλοντική, υπεύθυνη και ηθική χρήση της μηχανικής μάθησης στην ιατρική ανάλυση.

Βιβλιογραφία

1 Apple DP Team. (2017). Learning with privacy at scale. Apple Machine Learning Journal, 1(8).
2 Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
3 Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with DP. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.
4 Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of DP. Foundations and Trends in Theoretical Computer Science.
5 Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
6 Abowd, J. M. (2018). The U.S. Census Bureau adopts DP. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 28672867).
7 Ding, B., Kulkarni, J., & Yekhanin, S. (2017). Collecting telemetry data privately. In Advances in Neural Information Processing Systems (NeurIPS) (pp. 35713580).
8 McSherry, F., & Talwar, K. (2007). Mechanism design via DP. In 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS) (pp. 94103). IEEE.
9 Nissim, K., Raskhodnikova, S., & Smith, A. (2007). Smooth sensitivity and sampling in private data analysis. In Proceedings of the thirtyninth annual ACM symposium on Theory of computing (pp. 7584). ACM.
10 Duchi, J., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy and statistical minimax rates. In 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS) (pp. 429438). IEEE.
11 McSherry, F. (2009). Privacy integrated queries: an extensible platform for privacypreserving data analysis. In Proceedings of the 35th SIGMOD International Conference on Management of Data (pp. 1930). ACM.
12 Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006a). Calibrating noise to sensitivity in private data analysis. In Theory of Cryptography Conference (TCC) (pp. 265284). Springer, Berlin, Heidelberg.
13 Erlingsson, Ú., Pihur, V., & Korolova, A. (2014). RAPPOR: Randomized aggregatable privacypreserving ordinal response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 10541067). ACM.
14 Apple DP Team (2017). Learning with privacy at scale. Apple Machine Learning Journal, 1(8).
15 AdlerMilstein, J., Holmgren, A. J., Kralovec, P., & Worzala, C. (2017). Health information exchange progress and challenges: Findings from a 2016 national survey. Health Affairs, 36(5), 978985.
16 Boonstra, A., & Broekhuis, M. (2010). Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Services Research, 10(1), 231.
17 Gostin, L. O. (1995). Health information privacy. Cornell Law Review, 80(2), 451528.
18 Ponemon Institute. (2016). Sixth Annual Benchmark Study on Privacy & Security of Healthcare Data.

19 Schweitzer, E. J. (2019). Reconciliation of the cloud computing model with US federal electronic health record regulations. Journal of the American Medical Informatics Association, 19(2), 161165.
20 Blumenthal, D., & Tavenner, M. (2010). The "Meaningful Use" regulation for electronic health records. New England Journal of Medicine, 363(6), 501504.
21 Manca, D. P. (2015). Do electronic medical records improve quality of care? Yes. Canadian Family Physician, 61(10), 846851.
22 Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. Risk Management and Healthcare Policy, 4, 4755.
23 Blumenthal, D. (2009). Launching HITECH. New England Journal of Medicine, 362(5), 382385.
24 Blumenthal, D., & Tavenner, M. (2010). The "Meaningful Use" regulation for electronic health records. New England Journal of Medicine, 363(6), 501504.
25 Collen, M. F. (1995). A History of Medical Informatics in the United States, 1950 to 1990. American Medical Informatics Association.
26 Iyengar, K., Mabrouk, A., Jain, V. K., Vaishya, R., & Vaish, A. (2020). Learning opportunities from COVID19 and future effects on health care system. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(5), 943946.
27 Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. Risk Management and Healthcare Policy, 4, 4755.
28 Appari, A., & Johnson, M. E. (2010). Information security and privacy in healthcare: Current state of research. International Journal of Internet and Enterprise Management, 6(4), 279314.
29 Boonstra, A., & Broekhuis, M. (2010). Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Services Research, 10(1), 231.
30 Thakkar, M., & Davis, D. C. (2006). Risks, barriers, and benefits of EHR systems: A comparative study based on size of hospital. Perspectives in Health Information Management, 3, 5.
31 Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems, 2(1), 3.
32 Dalenius, T. (1977). Towards a methodology for statistical disclosure control. Statistik Tidskrift, 15(5).
33 Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. Journal of Law, Medicine & Ethics, 25(23), 98110.
34 Willenborg, L., & de Waal, T. (2001). Elements of Statistical Disclosure Control. Springer.
35 Fienberg, S. E. (1994). Confidentiality and disclosure limitation methodologies: Challenges for national statistics and statistical research. Journal of Official Statistics, 10(2), 135–171.
36 Duncan, G., & Lambert, D. (1986). Disclosurelimited data dissemination. Journal of the American Statistical Association, 81(393), 1028.
37 Dupuy, C., Arava, R., Gupta, R., & Others. (2022). *An efficient DP-SGD mechanism for large scale NLU models*. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.

38 Chua, L., Ghazi, B., Kamath, P., Kumar, R., & Others. (2024). *How private are DP-SGD implementations?*

39 Sander, T., Stock, P., & Others. (2023). *Tan without a burn: Scaling laws of DP-SGD*. In *Proceedings of the International Conference on Machine Learning*. PMLR.

40 Uniyal, A., Naidu, R., Kotti, S., Singh, S., Kenfack, P. J., & Others. (2021). *DP-SGD vs PATE: Which has less disparate impact on model accuracy?*

41 Hayes, J., Balle, B., & Mahloujifar, S. (2023). *Bounding training data reconstruction in DP-SGD*. In *Advances in Neural Information Processing Systems (NeurIPS)*.

42 Denison, C., Ghazi, B., Kamath, P., Kumar, R., & Others. (2022). *Private ad modeling with DP-SGD*.

43 Tang, Q., Shpilevskiy, F., & Lécuyer, M. (2024). *DP-AdamBC: Your DP-Adam is actually DP-SGD (unless you apply bias correction)*. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

44 Boenisch, F., Mühl, C., Dziedzic, A., & Others. (2023). *Have it your way: Individualized Privacy Assignment for DP-SGD*. In *Advances in Neural Information Processing Systems (NeurIPS)*.

45 Chua, L., Ghazi, B., Kamath, P., Kumar, R., & Others. (2024). *Scalable DP-SGD: Shuffling vs. Poisson subsampling*. In *Advances in Neural Information Processing Systems (NeurIPS)*.

46 Du, J., Li, S., Chen, X., Chen, S., & Hong, M. (2021). *Dynamic differential-privacy preserving SGD*. *arXiv preprint arXiv:2111.00173*.

47 Lin, G., Yan, H., Kou, G., Huang, T., Peng, S., & Others. (2022). *Understanding adaptive gradient clipping in DP-SGD, empirically*. *International Journal of Intelligent Systems*, Wiley Online Library.

48 Heo, G., Seo, J., & Whang, S. E. (2023). *Personalized DP-SGD using sampling mechanisms*. *arXiv preprint arXiv:2305.15165*.

49 Thudi, A., Jia, H., Meehan, C., Shumailov, I., & Others. (2024). *Gradients look alike: Sensitivity is often overestimated in DP-SGD*. *33rd USENIX Security Symposium*.

50 Cherubin, G., Köpf, B., Paverd, A., Tople, S., & Others. (2024). *Closed-form bounds for DP-SGD against record-level inference attacks*. *33rd USENIX Security Symposium*.

51 Kong, W., Medina, A. M., & Ribero, M. (2023). *DP-SGD for non-decomposable objective functions*. *arXiv preprint arXiv:2310.03104*.

52 Jang, J., Hwang, S., & Yang, H. J. (2024). *Rethinking DP-SGD in discrete domain: Exploring logistic distribution in the realm of SIGNSGD*. *ICLR 2024 – International Conference on Learning Representations*.

53 Cebere, T., Bellet, A., & Papernot, N. (2024). *Tighter privacy auditing of DP-SGD in the hidden state threat model*. *arXiv preprint arXiv:2405.14457*.

54 Annamalai, M. S. M. S. (2024). *It's our loss: No privacy amplification for hidden state DP-SGD with non-convex loss*. *arXiv preprint arXiv:2407.06496*.

55 Annamalai, M. S. M. S., Balle, B., De Cristofaro, E., & Others. (2024). *To shuffle or not to shuffle: Auditing DP-SGD with shuffling*.

56 Kong, W., & Muñoz Medina, A. (2023). *A unified fast gradient clipping framework for DP-SGD*. In *Advances in Neural Information Processing Systems (NeurIPS)*.
57 Dwork, C., & Roth, A. (2014). *The Algorithmic Foundations of Differential Privacy*. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042
58 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
59 Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). *Deep learning with differential privacy*. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 308–318).
60 Triastcyn, A., & Faltings, B. (2020). *Federated Learning with Bayesian Differential Privacy*. In Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI).
61 Bu, Z., Zou, Y., Wu, J., Wang, X., & Jin, R. (2020). *Deep Learning with Gaussian Differential Privacy*. Journal of Machine Learning Research, 21(211), 1–51.
62 Subramani, S., Nguyen, P. M., & Smith, A. (2021). *Privatizing the gradients: Optimal bounds for differentially private deep learning*. Advances in Neural Information Processing Systems, 34, 1652–1664.
63 Mironov, I. (2017). *Rényi Differential Privacy*. In 2017 IEEE 30th Computer Security Foundations Symposium (CSF) (pp. 263–275). IEEE.
64 Wang, Y.-X., Balle, B., & Kasiviswanathan, S. P. (2019). *Subsampled Rényi Differential Privacy and Analytical Moments Accountant*. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR.

Παράρτημα Α: Heart Disease Dataset

```
import zipfile
import os
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import tensorflow as tf
import tensorflow_privacy
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy_lib
import matplotlib.pyplot as plt
import pickle

# 1. Αποσυμπίεση του ZIP αρχείου
zip_path = "heart+disease.zip"
extract_path = "heart_disease_data"

with zipfile.ZipFile(zip_path, 'r') as zip_ref:
    zip_ref.extractall(extract_path)

# 2. Ορισμός των αρχείων που θα φορτώσουμε
data_files = [
    "processed.cleveland.data",
    "processed.hungarian.data",
    "processed.switzerland.data",
    "processed.va.data"
]

# 3. Ορισμός των στηλών
column_names = [
    "age", "sex", "cp", "trestbps", "chol", "fbs", "restecg",
    "thalach", "exang", "oldpeak", "slope", "ca", "thal", "target"
]

# 4. Φόρτωση και συγχώνευση των datasets
df_list = []
for file in data_files:
    file_path = os.path.join(extract_path, file)
    temp_df = pd.read_csv(file_path, names=column_names)
    df_list.append(temp_df)

# Συγχώνευση όλων των DataFrames
df = pd.concat(df_list, ignore_index=True)

# 5. Καθαρισμός των δεδομένων
df.replace("?", np.nan, inplace=True)  # Αντικατάσταση των '?' με NaN
df = df.apply(pd.to_numeric, errors='coerce')  # Μετατροπή όλων των τιμών σε αριθμητικές
df.fillna(df.median(), inplace=True)  # Αντικατάσταση των NaN με τη διάμεσο κάθε στήλης

# 6. Διαχωρισμός χαρακτηριστικών (X) και ετικετών (y)
X = df.drop(columns=['target'])
y = df['target']

# 7. Μετατροπή του target σε δυαδική ταξινόμηση (0 = χωρίς νόσο, 1 = με νόσο)
y = (y > 0).astype(int)
```

```
# 8. Κανονικοποίηση χαρακτηριστικών
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 9. Διαχωρισμός σε σύνολα εκπαίδευσης και δοκιμών
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42, stratify=y
)

# 10. Ορισμός υπερπαραμέτρων για DP-SGD
learning_rate = 0.01
noise_multiplier = 1.1
l2_norm_clip = 1.0
epochs = 50

from tensorflow_privacy.privacy.optimizers.dp_optimizer_keras import DPKerasSGDOptimizer
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy

batch_sizes = [32, 64]
results = []
for batch_size in batch_sizes:
    print(f"\n--- Εκπαίδευση με batch size: {batch_size} ---")

    # Δημιουργία νέου μοντέλου
    model = tf.keras.Sequential([
        tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train.shape[1],)),
        tf.keras.layers.Dense(8, activation='relu'),
        tf.keras.layers.Dense(1, activation='sigmoid')  # Δυαδική ταξινόμηση
    ])

    # Ορισμός DP-SGD Optimizer
    dp_optimizer = DPKerasSGDOptimizer(
        l2_norm_clip=l2_norm_clip,
        noise_multiplier=noise_multiplier,
        num_microbatches=1,
        learning_rate=learning_rate
    )

    model.compile(optimizer=dp_optimizer, loss='binary_crossentropy', metrics=['accuracy'])

    # Εκπαίδευση του μοντέλου
    history = model.fit(
        X_train, y_train,
        epochs=epochs,
        batch_size=batch_size,
        validation_data=(X_test, y_test),
        verbose=1
    )

    # Αξιολόγηση
    test_loss, test_acc = model.evaluate(X_test, y_test, verbose=0)
    print(f"Test Accuracy: {test_acc:.4f}")

    # Υπολογισμός privacy budget (ε)
    eps, _ = compute_dp_sgd_privacy_lib.compute_dp_sgd_privacy(
        n=len(X_train),
        batch_size=batch_size,
        noise_multiplier=noise_multiplier,
```

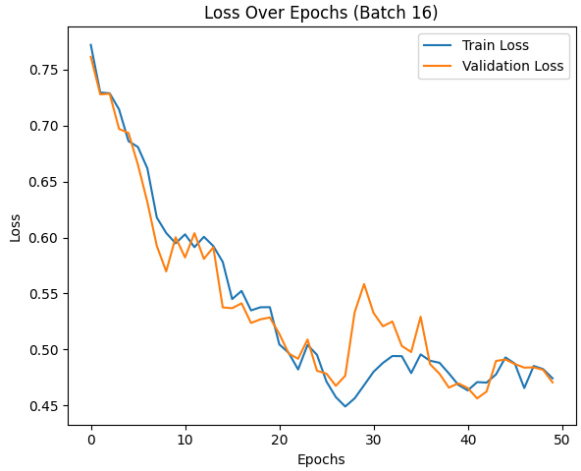
```
    epochs=epochs,
    delta=1e-5
  )
  print(f"Estimated ε for batch size {batch_size}: {eps:.2f}")

  # Σχεδίαση Διαγραμμάτων για κάθε περίπτωση
  plt.figure(figsize=(12, 5))
  plt.subplot(1, 2, 1)
  plt.plot(history.history['loss'], label='Train Loss')
  plt.plot(history.history['val_loss'], label='Validation Loss')
  plt.xlabel('Epochs')
  plt.ylabel('Loss')
  plt.title(f'Loss Over Epochs (Batch {batch_size})')
  plt.legend()

  plt.subplot(1, 2, 2)
  plt.plot(history.history['accuracy'], label='Train Accuracy')
  plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
  plt.xlabel('Epochs')
  plt.ylabel('Accuracy')
  plt.title(f'Accuracy Over Epochs (Batch {batch_size})')
  plt.legend()

  plt.tight_layout()
  plt.show()

results_df = pd.DataFrame(results)
results_df.to_csv("results_df_Hyperparameter_Tuning_heart_disease.csv", index=False)
print("Αποθηκεύτηκαν τα αποτελέσματα στο αρχείο CSV.")
```

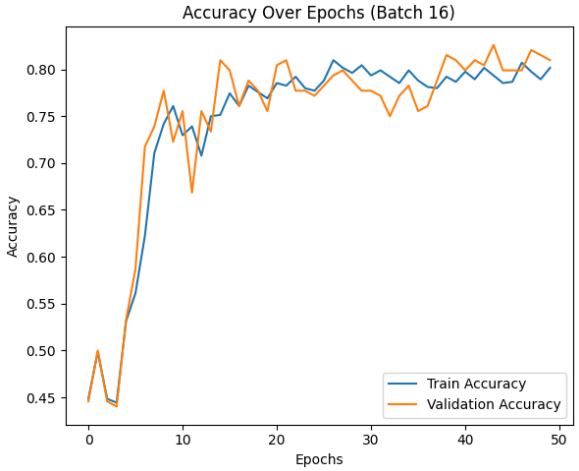
```
--- Εκπαίδευση με batch size: 16 ---
Epoch 1/50
46/46 [==============================] - 2s 7ms/step - loss: 0.7720 - accuracy: 0.4484 - val_loss:
0.7614 - val_accuracy: 0.4457
Epoch 2/50
46/46 [==============================] - 0s 4ms/step - loss: 0.7295 - accuracy: 0.4986 - val_loss:
0.7279 - val_accuracy: 0.5000
Epoch 3/50
46/46 [==============================] - 0s 5ms/step - loss: 0.7288 - accuracy: 0.4484 - val_loss:
0.7284 - val_accuracy: 0.4457
Epoch 4/50
46/46 [==============================] - 0s 3ms/step - loss: 0.7144 - accuracy: 0.4443 - val_loss:
0.6969 - val_accuracy: 0.4402
Epoch 5/50
46/46 [==============================] - 0s 3ms/step - loss: 0.6860 - accuracy: 0.5312 - val_loss:
0.6934 - val_accuracy: 0.5326
Epoch 6/50
46/46 [==============================] - 0s 3ms/step - loss: 0.6808 - accuracy: 0.5611 - val_loss:
0.6650 - val_accuracy: 0.5870
Epoch 7/50
46/46 [==============================] - 0s 4ms/step - loss: 0.6618 - accuracy: 0.6223 - val_loss:
0.6316 - val_accuracy: 0.7174
Epoch 8/50
46/46 [==============================] - 0s 3ms/step - loss: 0.6179 - accuracy: 0.7106 - val_loss:
0.5926 - val_accuracy: 0.7391
Epoch 9/50
46/46 [==============================] - 0s 4ms/step - loss: 0.6040 - accuracy: 0.7418 - val_loss:
0.5697 - val_accuracy: 0.7772
```

```
Epoch 10/50
46/46 [==============================] - 0s 4ms/step - loss: 0.5948 - accuracy: 0.7609 - val_loss:
0.6002 - val_accuracy: 0.7228
Epoch 11/50
46/46 [==============================] - 0s 3ms/step - loss: 0.6028 - accuracy: 0.7296 - val_loss:
0.5822 - val_accuracy: 0.7554
Epoch 12/50
46/46 [==============================] - 0s 3ms/step - loss: 0.5913 - accuracy: 0.7391 - val_loss:
0.6039 - val_accuracy: 0.6685
Epoch 13/50
46/46 [==============================] - 0s 3ms/step - loss: 0.6006 - accuracy: 0.7079 - val_loss:
0.5809 - val_accuracy: 0.7554
Epoch 14/50
46/46 [==============================] - 0s 3ms/step - loss: 0.5924 - accuracy: 0.7500 - val_loss:
0.5912 - val_accuracy: 0.7337
Epoch 15/50
46/46 [==============================] - 0s 4ms/step - loss: 0.5782 - accuracy: 0.7514 - val_loss:
0.5375 - val_accuracy: 0.8098
Epoch 16/50
46/46 [==============================] - 0s 3ms/step - loss: 0.5450 - accuracy: 0.7745 - val_loss:
0.5368 - val_accuracy: 0.7989
Epoch 17/50
46/46 [==============================] - 0s 4ms/step - loss: 0.5523 - accuracy: 0.7609 - val_loss:
0.5412 - val_accuracy: 0.7609
Epoch 18/50
46/46 [==============================] - 0s 4ms/step - loss: 0.5349 - accuracy: 0.7826 - val_loss:
0.5237 - val_accuracy: 0.7880
Epoch 19/50
46/46 [==============================] - 0s 4ms/step - loss: 0.5377 - accuracy: 0.7758 - val_loss:
0.5269 - val_accuracy: 0.7772
Epoch 20/50
46/46 [==============================] - 0s 6ms/step - loss: 0.5378 - accuracy: 0.7690 - val_loss:
0.5285 - val_accuracy: 0.7554
Epoch 21/50
46/46 [==============================] - 0s 4ms/step - loss: 0.5045 - accuracy: 0.7853 - val_loss:
0.5140 - val_accuracy: 0.8043
Epoch 22/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4970 - accuracy: 0.7826 - val_loss:
0.4964 - val_accuracy: 0.8098
Epoch 23/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4821 - accuracy: 0.7921 - val_loss:
0.4916 - val_accuracy: 0.7772
Epoch 24/50
46/46 [==============================] - 0s 3ms/step - loss: 0.5040 - accuracy: 0.7799 - val_loss:
0.5089 - val_accuracy: 0.7772
Epoch 25/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4952 - accuracy: 0.7772 - val_loss:
0.4810 - val_accuracy: 0.7717
Epoch 26/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4715 - accuracy: 0.7880 - val_loss:
0.4784 - val_accuracy: 0.7826
Epoch 27/50
46/46 [==============================] - 0s 3ms/step - loss: 0.4575 - accuracy: 0.8098 - val_loss:
0.4675 - val_accuracy: 0.7935
Epoch 28/50
46/46 [==============================] - 0s 3ms/step - loss: 0.4490 - accuracy: 0.8016 - val_loss:
0.4765 - val_accuracy: 0.7989
Epoch 29/50
```

```
46/46 [==============================] - 0s 4ms/step - loss: 0.4563 - accuracy: 0.7962 - val_loss:
0.5332 - val_accuracy: 0.7880
Epoch 30/50
46/46 [==============================] - 0s 3ms/step - loss: 0.4681 - accuracy: 0.8043 - val_loss:
0.5584 - val_accuracy: 0.7772
Epoch 31/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4800 - accuracy: 0.7935 - val_loss:
0.5326 - val_accuracy: 0.7772
Epoch 32/50
46/46 [==============================] - 0s 3ms/step - loss: 0.4880 - accuracy: 0.7989 - val_loss:
0.5207 - val_accuracy: 0.7717
Epoch 33/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4941 - accuracy: 0.7921 - val_loss:
0.5250 - val_accuracy: 0.7500
Epoch 34/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4940 - accuracy: 0.7853 - val_loss:
0.5032 - val_accuracy: 0.7717
Epoch 35/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4789 - accuracy: 0.7989 - val_loss:
0.4977 - val_accuracy: 0.7826
Epoch 36/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4956 - accuracy: 0.7880 - val_loss:
0.5292 - val_accuracy: 0.7554
Epoch 37/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4899 - accuracy: 0.7812 - val_loss:
0.4869 - val_accuracy: 0.7609
Epoch 38/50
46/46 [==============================] - 0s 3ms/step - loss: 0.4881 - accuracy: 0.7799 - val_loss:
0.4783 - val_accuracy: 0.7880
Epoch 39/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4786 - accuracy: 0.7921 - val_loss:
0.4659 - val_accuracy: 0.8152
Epoch 40/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4684 - accuracy: 0.7867 - val_loss:
0.4697 - val_accuracy: 0.8098
Epoch 41/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4634 - accuracy: 0.7976 - val_loss:
0.4655 - val_accuracy: 0.7989
Epoch 42/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4708 - accuracy: 0.7894 - val_loss:
0.4563 - val_accuracy: 0.8098
Epoch 43/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4704 - accuracy: 0.8016 - val_loss:
0.4624 - val_accuracy: 0.8043
Epoch 44/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4777 - accuracy: 0.7935 - val_loss:
0.4897 - val_accuracy: 0.8261
Epoch 45/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4929 - accuracy: 0.7853 - val_loss:
0.4910 - val_accuracy: 0.7989
Epoch 46/50
46/46 [==============================] - 0s 4ms/step - loss: 0.4873 - accuracy: 0.7867 - val_loss:
0.4870 - val_accuracy: 0.7989
Epoch 47/50
46/46 [==============================] - 0s 7ms/step - loss: 0.4655 - accuracy: 0.8071 - val_loss:
0.4838 - val_accuracy: 0.7989
Epoch 48/50
```

46/46 [==========] - 0s 4ms/step - loss: 0.4852 - accuracy: 0.7976 - val_loss: 0.4840 - val_accuracy: 0.8207
Epoch 49/50
46/46 [==========] - 0s 4ms/step - loss: 0.4825 - accuracy: 0.7894 - val_loss: 0.4818 - val_accuracy: 0.8152
Epoch 50/50
46/46 [==========] - 0s 4ms/step - loss: 0.4742 - accuracy: 0.8016 - val_loss: 0.4706 - val_accuracy: 0.8098
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Test Accuracy: 0.8098
Estimated ε for batch size 16: 6.10



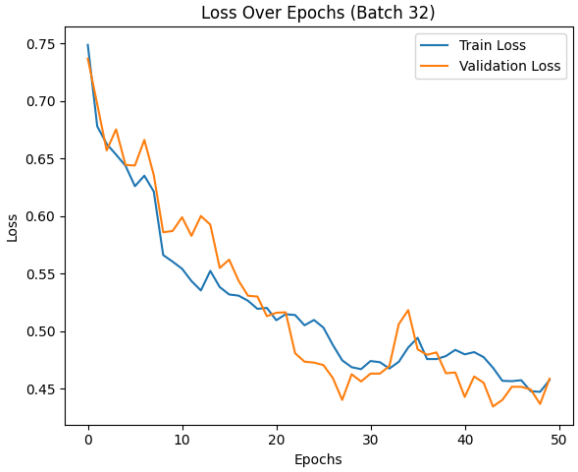


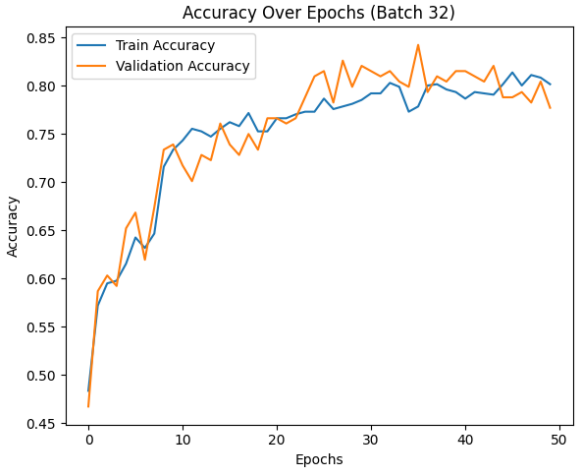
--- Εκπαίδευση με batch size: 32 ---
Epoch 1/50
23/23 [==========] - 2s 11ms/step - loss: 0.7489 - accuracy: 0.4837 - val_loss: 0.7370 - val_accuracy: 0.4674
Epoch 2/50
23/23 [==========] - 0s 4ms/step - loss: 0.6782 - accuracy: 0.5720 - val_loss: 0.6973 - val_accuracy: 0.5870
Epoch 3/50
23/23 [==========] - 0s 5ms/step - loss: 0.6629 - accuracy: 0.5951 - val_loss: 0.6572 - val_accuracy: 0.6033
Epoch 4/50
23/23 [==========] - 0s 5ms/step - loss: 0.6536 - accuracy: 0.5978 - val_loss: 0.6756 - val_accuracy: 0.5924
Epoch 5/50
23/23 [==========] - 0s 3ms/step - loss: 0.6439 - accuracy: 0.6155 - val_loss: 0.6446 - val_accuracy: 0.6522
Epoch 6/50
23/23 [==========] - 0s 4ms/step - loss: 0.6261 - accuracy: 0.6427 - val_loss: 0.6442 - val_accuracy: 0.6685
Epoch 7/50
23/23 [==========] - 0s 4ms/step - loss: 0.6353 - accuracy: 0.6318 - val_loss: 0.6664 - val_accuracy: 0.6196
Epoch 8/50
23/23 [==========] - 0s 3ms/step - loss: 0.6215 - accuracy: 0.6467 - val_loss: 0.6358 - val_accuracy: 0.6739
Epoch 9/50

```
23/23 [==============================] - 0s 4ms/step - loss: 0.5663 - accuracy: 0.7160 - val_loss:
0.5861 - val_accuracy: 0.7337
Epoch 10/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5606 - accuracy: 0.7337 - val_loss:
0.5872 - val_accuracy: 0.7391
Epoch 11/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5544 - accuracy: 0.7432 - val_loss:
0.5992 - val_accuracy: 0.7174
Epoch 12/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5437 - accuracy: 0.7554 - val_loss:
0.5830 - val_accuracy: 0.7011
Epoch 13/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5356 - accuracy: 0.7527 - val_loss:
0.6002 - val_accuracy: 0.7283
Epoch 14/50
23/23 [==============================] - 0s 3ms/step - loss: 0.5526 - accuracy: 0.7473 - val_loss:
0.5927 - val_accuracy: 0.7228
Epoch 15/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5385 - accuracy: 0.7554 - val_loss:
0.5551 - val_accuracy: 0.7609
Epoch 16/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5321 - accuracy: 0.7622 - val_loss:
0.5623 - val_accuracy: 0.7391
Epoch 17/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5311 - accuracy: 0.7582 - val_loss:
0.5440 - val_accuracy: 0.7283
Epoch 18/50
23/23 [==============================] - 0s 7ms/step - loss: 0.5266 - accuracy: 0.7717 - val_loss:
0.5310 - val_accuracy: 0.7500
Epoch 19/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5196 - accuracy: 0.7527 - val_loss:
0.5303 - val_accuracy: 0.7337
Epoch 20/50
23/23 [==============================] - 0s 3ms/step - loss: 0.5204 - accuracy: 0.7527 - val_loss:
0.5132 - val_accuracy: 0.7663
Epoch 21/50
23/23 [==============================] - 0s 3ms/step - loss: 0.5097 - accuracy: 0.7663 - val_loss:
0.5162 - val_accuracy: 0.7663
Epoch 22/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5149 - accuracy: 0.7663 - val_loss:
0.5165 - val_accuracy: 0.7609
Epoch 23/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5142 - accuracy: 0.7704 - val_loss:
0.4810 - val_accuracy: 0.7663
Epoch 24/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5053 - accuracy: 0.7731 - val_loss:
0.4737 - val_accuracy: 0.7880
Epoch 25/50
23/23 [==============================] - 0s 5ms/step - loss: 0.5099 - accuracy: 0.7731 - val_loss:
0.4729 - val_accuracy: 0.8098
Epoch 26/50
23/23 [==============================] - 0s 4ms/step - loss: 0.5033 - accuracy: 0.7867 - val_loss:
0.4708 - val_accuracy: 0.8152
Epoch 27/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4883 - accuracy: 0.7758 - val_loss:
0.4595 - val_accuracy: 0.7826
Epoch 28/50
```

```
23/23 [==============================] - 0s 5ms/step - loss: 0.4748 - accuracy: 0.7785 - val_loss:
0.4404 - val_accuracy: 0.8261
Epoch 29/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4689 - accuracy: 0.7812 - val_loss:
0.4627 - val_accuracy: 0.7989
Epoch 30/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4672 - accuracy: 0.7853 - val_loss:
0.4565 - val_accuracy: 0.8207
Epoch 31/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4742 - accuracy: 0.7921 - val_loss:
0.4633 - val_accuracy: 0.8152
Epoch 32/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4732 - accuracy: 0.7921 - val_loss:
0.4633 - val_accuracy: 0.8098
Epoch 33/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4677 - accuracy: 0.8030 - val_loss:
0.4699 - val_accuracy: 0.8152
Epoch 34/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4735 - accuracy: 0.7989 - val_loss:
0.5063 - val_accuracy: 0.8043
Epoch 35/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4860 - accuracy: 0.7731 - val_loss:
0.5185 - val_accuracy: 0.7989
Epoch 36/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4944 - accuracy: 0.7785 - val_loss:
0.4844 - val_accuracy: 0.8424
Epoch 37/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4759 - accuracy: 0.8003 - val_loss:
0.4797 - val_accuracy: 0.7935
Epoch 38/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4760 - accuracy: 0.8016 - val_loss:
0.4818 - val_accuracy: 0.8098
Epoch 39/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4786 - accuracy: 0.7962 - val_loss:
0.4636 - val_accuracy: 0.8043
Epoch 40/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4839 - accuracy: 0.7935 - val_loss:
0.4643 - val_accuracy: 0.8152
Epoch 41/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4800 - accuracy: 0.7867 - val_loss:
0.4431 - val_accuracy: 0.8152
Epoch 42/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4820 - accuracy: 0.7935 - val_loss:
0.4608 - val_accuracy: 0.8098
Epoch 43/50
23/23 [==============================] - 0s 4ms/step - loss: 0.4777 - accuracy: 0.7921 - val_loss:
0.4554 - val_accuracy: 0.8043
Epoch 44/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4684 - accuracy: 0.7908 - val_loss:
0.4347 - val_accuracy: 0.8207
Epoch 45/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4570 - accuracy: 0.8016 - val_loss:
0.4405 - val_accuracy: 0.7880
Epoch 46/50
23/23 [==============================] - 0s 5ms/step - loss: 0.4568 - accuracy: 0.8139 - val_loss:
0.4518 - val_accuracy: 0.7880
Epoch 47/50
```

23/23 [==========] - 0s 4ms/step - loss: 0.4576 - accuracy: 0.8003 - val_loss: 0.4519 - val_accuracy: 0.7935
Epoch 48/50
23/23 [==========] - 0s 4ms/step - loss: 0.4480 - accuracy: 0.8111 - val_loss: 0.4495 - val_accuracy: 0.7826
Epoch 49/50
23/23 [==========] - 0s 5ms/step - loss: 0.4475 - accuracy: 0.8084 - val_loss: 0.4369 - val_accuracy: 0.8043
Epoch 50/50
23/23 [==========] - 0s 5ms/step - loss: 0.4577 - accuracy: 0.8016 - val_loss: 0.4589 - val_accuracy: 0.7772
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Test Accuracy: 0.7772
Estimated ε for batch size 32: 9.25



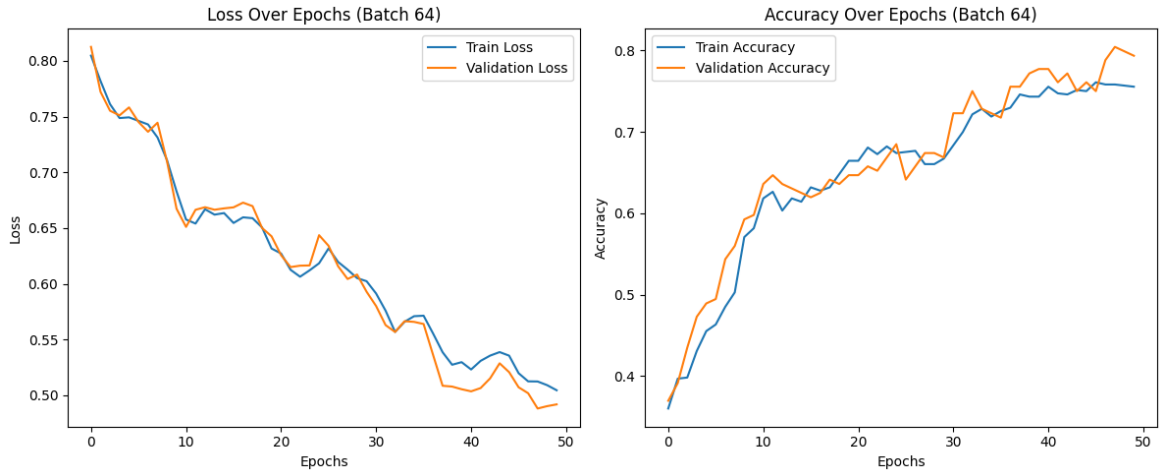


--- Εκπαίδευση με batch size: 64 ---
Epoch 1/50
12/12 [==========] - 2s 18ms/step - loss: 0.8047 - accuracy: 0.3601 - val_loss: 0.8125 - val_accuracy: 0.3696
Epoch 2/50
12/12 [==========] - 0s 5ms/step - loss: 0.7820 - accuracy: 0.3967 - val_loss: 0.7722 - val_accuracy: 0.3913
Epoch 3/50
12/12 [==========] - 0s 5ms/step - loss: 0.7612 - accuracy: 0.3981 - val_loss: 0.7553 - val_accuracy: 0.4348
Epoch 4/50
12/12 [==========] - 0s 6ms/step - loss: 0.7487 - accuracy: 0.4307 - val_loss: 0.7511 - val_accuracy: 0.4728
Epoch 5/50
12/12 [==========] - 0s 6ms/step - loss: 0.7493 - accuracy: 0.4552 - val_loss: 0.7582 - val_accuracy: 0.4891
Epoch 6/50
12/12 [==========] - 0s 5ms/step - loss: 0.7462 - accuracy: 0.4633 - val_loss: 0.7452 - val_accuracy: 0.4946
Epoch 7/50
12/12 [==========] - 0s 5ms/step - loss: 0.7430 - accuracy: 0.4851 - val_loss: 0.7364 - val_accuracy: 0.5435
Epoch 8/50

```
12/12 [==============================] - 0s 5ms/step - loss: 0.7313 - accuracy: 0.5027 - val_loss:
0.7445 - val_accuracy: 0.5598
Epoch 9/50
12/12 [==============================] - 0s 5ms/step - loss: 0.7109 - accuracy: 0.5707 - val_loss:
0.7097 - val_accuracy: 0.5924
Epoch 10/50
12/12 [==============================] - 0s 9ms/step - loss: 0.6830 - accuracy: 0.5815 - val_loss:
0.6674 - val_accuracy: 0.5978
Epoch 11/50
12/12 [==============================] - 0s 5ms/step - loss: 0.6577 - accuracy: 0.6182 - val_loss:
0.6511 - val_accuracy: 0.6359
Epoch 12/50
12/12 [==============================] - 0s 5ms/step - loss: 0.6541 - accuracy: 0.6264 - val_loss:
0.6664 - val_accuracy: 0.6467
Epoch 13/50
12/12 [==============================] - 0s 5ms/step - loss: 0.6669 - accuracy: 0.6033 - val_loss:
0.6687 - val_accuracy: 0.6359
Epoch 14/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6621 - accuracy: 0.6182 - val_loss:
0.6665 - val_accuracy: 0.6304
Epoch 15/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6635 - accuracy: 0.6141 - val_loss:
0.6677 - val_accuracy: 0.6250
Epoch 16/50
12/12 [==============================] - 0s 7ms/step - loss: 0.6546 - accuracy: 0.6318 - val_loss:
0.6686 - val_accuracy: 0.6196
Epoch 17/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6597 - accuracy: 0.6277 - val_loss:
0.6728 - val_accuracy: 0.6250
Epoch 18/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6589 - accuracy: 0.6318 - val_loss:
0.6697 - val_accuracy: 0.6413
Epoch 19/50
12/12 [==============================] - 0s 5ms/step - loss: 0.6506 - accuracy: 0.6481 - val_loss:
0.6501 - val_accuracy: 0.6359
Epoch 20/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6317 - accuracy: 0.6644 - val_loss:
0.6426 - val_accuracy: 0.6467
Epoch 21/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6272 - accuracy: 0.6644 - val_loss:
0.6257 - val_accuracy: 0.6467
Epoch 22/50
12/12 [==============================] - 0s 7ms/step - loss: 0.6126 - accuracy: 0.6807 - val_loss:
0.6150 - val_accuracy: 0.6576
Epoch 23/50
12/12 [==============================] - 0s 8ms/step - loss: 0.6064 - accuracy: 0.6726 - val_loss:
0.6163 - val_accuracy: 0.6522
Epoch 24/50
12/12 [==============================] - 0s 8ms/step - loss: 0.6121 - accuracy: 0.6821 - val_loss:
0.6165 - val_accuracy: 0.6685
Epoch 25/50
12/12 [==============================] - 0s 9ms/step - loss: 0.6184 - accuracy: 0.6739 - val_loss:
0.6437 - val_accuracy: 0.6848
Epoch 26/50
12/12 [==============================] - 0s 9ms/step - loss: 0.6318 - accuracy: 0.6753 - val_loss:
0.6341 - val_accuracy: 0.6413
Epoch 27/50
```

```
12/12 [==============================] - 0s 7ms/step - loss: 0.6197 - accuracy: 0.6766 - val_loss:
0.6158 - val_accuracy: 0.6576
Epoch 28/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6127 - accuracy: 0.6603 - val_loss:
0.6044 - val_accuracy: 0.6739
Epoch 29/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6052 - accuracy: 0.6603 - val_loss:
0.6083 - val_accuracy: 0.6739
Epoch 30/50
12/12 [==============================] - 0s 6ms/step - loss: 0.6024 - accuracy: 0.6671 - val_loss:
0.5929 - val_accuracy: 0.6685
Epoch 31/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5914 - accuracy: 0.6834 - val_loss:
0.5801 - val_accuracy: 0.7228
Epoch 32/50
12/12 [==============================] - 0s 7ms/step - loss: 0.5759 - accuracy: 0.6997 - val_loss:
0.5630 - val_accuracy: 0.7228
Epoch 33/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5569 - accuracy: 0.7215 - val_loss:
0.5568 - val_accuracy: 0.7500
Epoch 34/50
12/12 [==============================] - 0s 7ms/step - loss: 0.5660 - accuracy: 0.7283 - val_loss:
0.5665 - val_accuracy: 0.7283
Epoch 35/50
12/12 [==============================] - 0s 7ms/step - loss: 0.5710 - accuracy: 0.7188 - val_loss:
0.5659 - val_accuracy: 0.7228
Epoch 36/50
12/12 [==============================] - 0s 7ms/step - loss: 0.5714 - accuracy: 0.7255 - val_loss:
0.5640 - val_accuracy: 0.7174
Epoch 37/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5554 - accuracy: 0.7296 - val_loss:
0.5362 - val_accuracy: 0.7554
Epoch 38/50
12/12 [==============================] - 0s 5ms/step - loss: 0.5387 - accuracy: 0.7459 - val_loss:
0.5086 - val_accuracy: 0.7554
Epoch 39/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5275 - accuracy: 0.7432 - val_loss:
0.5079 - val_accuracy: 0.7717
Epoch 40/50
12/12 [==============================] - 0s 7ms/step - loss: 0.5298 - accuracy: 0.7432 - val_loss:
0.5055 - val_accuracy: 0.7772
Epoch 41/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5232 - accuracy: 0.7554 - val_loss:
0.5036 - val_accuracy: 0.7772
Epoch 42/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5309 - accuracy: 0.7473 - val_loss:
0.5065 - val_accuracy: 0.7609
Epoch 43/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5356 - accuracy: 0.7459 - val_loss:
0.5152 - val_accuracy: 0.7717
Epoch 44/50
12/12 [==============================] - 0s 6ms/step - loss: 0.5388 - accuracy: 0.7514 - val_loss:
0.5287 - val_accuracy: 0.7500
Epoch 45/50
12/12 [==============================] - 0s 8ms/step - loss: 0.5356 - accuracy: 0.7500 - val_loss:
0.5208 - val_accuracy: 0.7609
Epoch 46/50
```

12/12 [==========] - 0s 7ms/step - loss: 0.5199 - accuracy: 0.7609 - val_loss: 0.5072 - val_accuracy: 0.7500
Epoch 47/50
12/12 [==========] - 0s 6ms/step - loss: 0.5126 - accuracy: 0.7582 - val_loss: 0.5019 - val_accuracy: 0.7880
Epoch 48/50
12/12 [==========] - 0s 7ms/step - loss: 0.5124 - accuracy: 0.7582 - val_loss: 0.4882 - val_accuracy: 0.8043
Epoch 49/50
12/12 [==========] - 0s 7ms/step - loss: 0.5091 - accuracy: 0.7568 - val_loss: 0.4903 - val_accuracy: 0.7989
Epoch 50/50
12/12 [==========] - 0s 6ms/step - loss: 0.5046 - accuracy: 0.7554 - val_loss: 0.4919 - val_accuracy: 0.7935
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Test Accuracy: 0.7935
Estimated ε for batch size 64: 13.96



Αποθηκεύτηκαν τα αποτελέσματα στο αρχείο CSV.

Hyperparameter_Tuning_heart_disease

```
import itertools
import matplotlib.pyplot as plt
import pickle
import tensorflow as tf
import tensorflow_privacy
import numpy as np
import pandas as pd
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score, precision_score, recall_score, f1_score
import seaborn as sns
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy_lib

# 1. Φόρτωση των δεδομένων
with open("dataset.pkl", "rb") as f:
    X_train, X_test, y_train, y_test = pickle.load(f)
```

```
print("Dataset loaded successfully!")

# 2. Υπερπαράμετροι
learning_rates = [0.001, 0.01, 0.1]
noise_multipliers = [0.5, 1.1, 2.0]
l2_norm_clips = [0.5, 1.0, 2.0]
batch_sizes = [16, 32, 64]
epochs = 30

results = []

# 3. Πειράματα με όλους τους συνδυασμούς
for lr, noise, clip, batch in itertools.product(learning_rates, noise_multipliers, l2_norm_clips, batch_sizes):
    print(f"Training model with LR={lr}, Noise={noise}, Clip={clip}, Batch={batch}")

    batch_size = (len(X_train) // batch) * batch
    if batch_size == 0:
        batch_size = 1

    model = tf.keras.Sequential([
        tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train.shape[1],)),
        tf.keras.layers.Dense(8, activation='relu'),
        tf.keras.layers.Dense(1, activation='sigmoid')
    ])

    dp_optimizer = tensorflow_privacy.DPKerasSGDOptimizer(
        l2_norm_clip=clip,
        noise_multiplier=noise,
        num_microbatches=1,
        learning_rate=lr
    )

    model.compile(optimizer=dp_optimizer, loss='binary_crossentropy', metrics=['accuracy'])
    history = model.fit(X_train, y_train, epochs=epochs, batch_size=batch_size, validation_data=(X_test,
y_test), verbose=0)

    test_loss, test_acc = model.evaluate(X_test, y_test, verbose=0)
    print(f"Test Accuracy: {test_acc:.4f} | Test Loss: {test_loss:.4f}")

    y_pred = (model.predict(X_test) > 0.5).astype("int32")
    y_proba = model.predict(X_test)

    precision = precision_score(y_test, y_pred, zero_division=0)
    recall = recall_score(y_test, y_pred, zero_division=0)
    f1 = f1_score(y_test, y_pred, zero_division=0)
    roc_auc = roc_auc_score(y_test, y_proba)

    eps, _ = compute_dp_sgd_privacy_lib.compute_dp_sgd_privacy(
        n=len(X_train),
        batch_size=batch_size,
        noise_multiplier=noise,
        epochs=epochs,
        delta=1 / (len(X_train) * np.sqrt(len(X_train)))
    )

    results.append({
        "learning_rate": lr,
        "noise_multiplier": noise,
```

```
        "l2_norm_clip": clip,
        "batch_size": batch,
        "test_accuracy": test_acc,
        "test_loss": test_loss,
        "precision": precision,
        "recall": recall,
        "f1_score": f1,
        "roc_auc": roc_auc,
        "epsilon": eps
    })

# 4. Μετατροπή σε DataFrame
results_df = pd.DataFrame(results)

# 5. Διαγράμματα
plt.figure(figsize=(12, 6))
for noise in results_df["noise_multiplier"].unique():
    subset = results_df[results_df["noise_multiplier"] == noise]
    plt.plot(subset["learning_rate"], subset["test_accuracy"], label=f"Noise={noise}")
plt.xlabel("Learning Rate")
plt.ylabel("Test Accuracy")
plt.title("Effect of Learning Rate & Noise on Accuracy")
plt.legend()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
for clip in results_df["l2_norm_clip"].unique():
    subset = results_df[results_df["l2_norm_clip"] == clip]
    plt.plot(subset["batch_size"], subset["test_loss"], label=f"Clip={clip}")
plt.xlabel("Batch Size")
plt.ylabel("Test Loss")
plt.title("Effect of Clipping & Batch Size on Loss")
plt.legend()
plt.grid(True)
plt.show()

# 6. Classification Report
print("Neos kodikas")
predictions = (model.predict(X_test) > 0.5).astype("int32")
print("\nClassification Report (Dataset with DP-SGD):")
print(classification_report(y_test, predictions))
roc_auc = roc_auc_score(y_test, model.predict(X_test))
print(f"ROC-AUC Score: {roc_auc:.4f}")

cm = confusion_matrix(y_test, predictions)
plt.figure(figsize=(8,6))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.title("Confusion Matrix (with DP-SGD)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

# 7. Μοντέλο χωρίς DP
model_no_dp = tf.keras.Sequential([
    tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train.shape[1],)),
    tf.keras.layers.Dense(8, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
```

```
])
model_no_dp.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model_no_dp.fit(X_train, y_train, epochs=epochs, batch_size=32, verbose=0)

predictions_no_dp = (model_no_dp.predict(X_test) > 0.5).astype("int32")
cm_no_dp = confusion_matrix(y_test, predictions_no_dp)

plt.figure(figsize=(8,6))
sns.heatmap(cm_no_dp, annot=True, fmt="d", cmap="Greens")
plt.title("Confusion Matrix (without DP-SGD)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

fp_dp, fn_dp = cm[0][1], cm[1][0]
fp_no_dp, fn_no_dp = cm_no_dp[0][1], cm_no_dp[1][0]
print(f"False Positives (with DP-SGD): {fp_dp}, False Negatives (with DP-SGD): {fn_dp}")
print(f"False Positives (no DP-SGD): {fp_no_dp}, False Negatives (no DP-SGD): {fn_no_dp}")

# 8. Στατιστικά
mean_accuracy = results_df['test_accuracy'].mean()
std_accuracy = results_df['test_accuracy'].std()
mean_loss = results_df['test_loss'].mean()
std_loss = results_df['test_loss'].std()
print("\nStatistical Analysis of Results:")
print(f"Mean Accuracy: {mean_accuracy:.4f}, Std Accuracy: {std_accuracy:.4f}")
print(f"Mean Loss: {mean_loss:.4f}, Std Loss: {std_loss:.4f}")

# 9. Σχέση ε και noise multiplier
num_samples = X_train.shape[0]
delta = 1 / (num_samples * np.sqrt(num_samples))
print(f"\nDelta (δ) used for experiments: {delta:.8f}")
print("The delta value ensures theoretical guarantees within differential privacy context.")

print("\nRelationship between privacy budget (ε) and noise multiplier:")
for noise in noise_multipliers:
    epsilon, _ = tensorflow_privacy.compute_dp_sgd_privacy(n=num_samples,
                                           batch_size=batch_sizes[0],
                                           noise_multiplier=noise,
                                           epochs=epochs,
                                           delta=delta)
    print(f"Noise Multiplier: {noise} => Epsilon (ε): {epsilon:.4f}")

# 10. Αποθήκευση σε CSV
results_df.to_csv('results_df_Hyperparameter_Tuning_heart_disease.csv', index=False)
print("\nResults saved successfully in 'results_df_Hyperparameter_Tuning_heart_disease.csv'")
```

```
Dataset loaded successfully!
Training model with LR=0.001, Noise=0.5, Clip=0.5, Batch=16
Test Accuracy: 0.2935 | Test Loss: 0.8279
6/6 [==============================] - 0s 6ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate
```

context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=0.5, Batch=32
Test Accuracy: 0.5543 | Test Loss: 0.6772
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=0.5, Batch=64
Test Accuracy: 0.4130 | Test Loss: 0.7525
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 5ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=1.0, Batch=16
Test Accuracy: 0.3967 | Test Loss: 0.7597
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=1.0, Batch=32
Test Accuracy: 0.5000 | Test Loss: 0.7342
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=1.0, Batch=64
Test Accuracy: 0.5272 | Test Loss: 0.7172
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=2.0, Batch=16

Test Accuracy: 0.5217 | Test Loss: 0.7720
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=2.0, Batch=32
Test Accuracy: 0.5435 | Test Loss: 0.7329
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=2.0, Batch=64
Test Accuracy: 0.5380 | Test Loss: 0.7112
6/6 [==============================] - 0s 5ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=0.5, Batch=16
Test Accuracy: 0.3967 | Test Loss: 0.8043
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=0.5, Batch=32
Test Accuracy: 0.4565 | Test Loss: 0.8655
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=0.5, Batch=64
Test Accuracy: 0.5435 | Test Loss: 0.6897
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 4ms/step

WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=1.0, Batch=16
Test Accuracy: 0.4891 | Test Loss: 0.7142
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=1.0, Batch=32
Test Accuracy: 0.5000 | Test Loss: 0.7089
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=1.0, Batch=64
Test Accuracy: 0.6902 | Test Loss: 0.6339
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 5ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=2.0, Batch=16
Test Accuracy: 0.5435 | Test Loss: 0.7110
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=2.0, Batch=32
Test Accuracy: 0.4457 | Test Loss: 0.7952
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate

context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=2.0, Batch=64
Test Accuracy: 0.4620 | Test Loss: 0.7562
6/6 [==============================] - 0s 5ms/step
6/6 [==============================] - 0s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=0.5, Batch=16
Test Accuracy: 0.5380 | Test Loss: 0.7346
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=0.5, Batch=32
Test Accuracy: 0.3967 | Test Loss: 0.9538
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=0.5, Batch=64
Test Accuracy: 0.4185 | Test Loss: 0.7554
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=1.0, Batch=16
Test Accuracy: 0.5380 | Test Loss: 0.7315
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=1.0, Batch=32

Test Accuracy: 0.6250 | Test Loss: 0.6540
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=1.0, Batch=64
Test Accuracy: 0.4891 | Test Loss: 0.7025
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=2.0, Batch=16
Test Accuracy: 0.4783 | Test Loss: 0.8106
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=2.0, Batch=32
Test Accuracy: 0.6413 | Test Loss: 0.5851
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=2.0, Batch=64
Test Accuracy: 0.5978 | Test Loss: 0.7111
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=0.5, Batch=16
Test Accuracy: 0.5543 | Test Loss: 0.7012
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 0s/step

WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=0.5, Batch=32
Test Accuracy: 0.6522 | Test Loss: 0.6260
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=0.5, Batch=64
Test Accuracy: 0.5054 | Test Loss: 0.7058
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=1.0, Batch=16
Test Accuracy: 0.5761 | Test Loss: 0.6594
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=1.0, Batch=32
Test Accuracy: 0.6685 | Test Loss: 0.6675
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=1.0, Batch=64
Test Accuracy: 0.5435 | Test Loss: 0.6857
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate

context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=2.0, Batch=16
Test Accuracy: 0.4620 | Test Loss: 0.6877
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=2.0, Batch=32
Test Accuracy: 0.4565 | Test Loss: 0.7129
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=2.0, Batch=64
Test Accuracy: 0.7065 | Test Loss: 0.6014
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=0.5, Batch=16
Test Accuracy: 0.5543 | Test Loss: 0.7176
6/6 [==============================] - 0s 1ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=0.5, Batch=32
Test Accuracy: 0.5978 | Test Loss: 0.6874
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=0.5, Batch=64

```
Test Accuracy: 0.5163 | Test Loss: 0.7560
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=1.0, Batch=16
Test Accuracy: 0.5978 | Test Loss: 0.6540
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=1.0, Batch=32
Test Accuracy: 0.7337 | Test Loss: 0.5691
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=1.0, Batch=64
Test Accuracy: 0.6630 | Test Loss: 0.6917
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=2.0, Batch=16
Test Accuracy: 0.6196 | Test Loss: 0.6568
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=2.0, Batch=32
Test Accuracy: 0.6087 | Test Loss: 0.6172
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 0s/step
```

WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=2.0, Batch=64
Test Accuracy: 0.5163 | Test Loss: 0.7374
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=0.5, Batch=16
Test Accuracy: 0.4076 | Test Loss: 0.7775
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=0.5, Batch=32
Test Accuracy: 0.5870 | Test Loss: 0.6558
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=0.5, Batch=64
Test Accuracy: 0.5217 | Test Loss: 0.6776
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=1.0, Batch=16
Test Accuracy: 0.4783 | Test Loss: 0.7363
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate

context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=1.0, Batch=32
Test Accuracy: 0.5870 | Test Loss: 0.7033
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=1.0, Batch=64
Test Accuracy: 0.6848 | Test Loss: 0.6215
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=2.0, Batch=16
Test Accuracy: 0.7174 | Test Loss: 0.5400
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=2.0, Batch=32
Test Accuracy: 0.4674 | Test Loss: 0.8237
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=2.0, Batch=64
Test Accuracy: 0.6793 | Test Loss: 0.6782
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=0.5, Batch=16

```
Test Accuracy: 0.8043 | Test Loss: 0.4948
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=0.5, Batch=32
Test Accuracy: 0.6739 | Test Loss: 0.5733
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=0.5, Batch=64
Test Accuracy: 0.7880 | Test Loss: 0.5372
6/6 [==============================] - 0s 881us/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=1.0, Batch=16
Test Accuracy: 0.7120 | Test Loss: 0.5666
6/6 [==============================] - 0s 1ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=1.0, Batch=32
Test Accuracy: 0.7283 | Test Loss: 0.5540
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=1.0, Batch=64
Test Accuracy: 0.6957 | Test Loss: 0.6110
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 0s/step
```

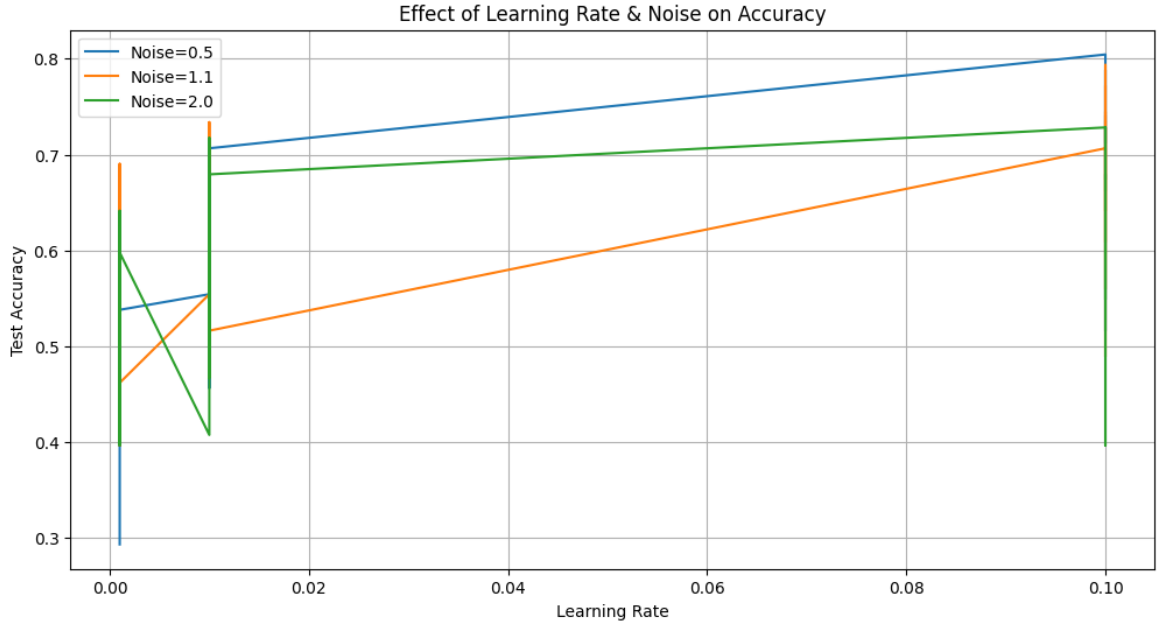
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=2.0, Batch=16
Test Accuracy: 0.5163 | Test Loss: 1.1303
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=2.0, Batch=32
Test Accuracy: 0.6522 | Test Loss: 0.7747
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=2.0, Batch=64
Test Accuracy: 0.7391 | Test Loss: 0.6107
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=0.5, Batch=16
Test Accuracy: 0.7065 | Test Loss: 0.5858
6/6 [==============================] - 0s 4ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=0.5, Batch=32
Test Accuracy: 0.6087 | Test Loss: 0.6314
6/6 [==============================] - 0s 2ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate

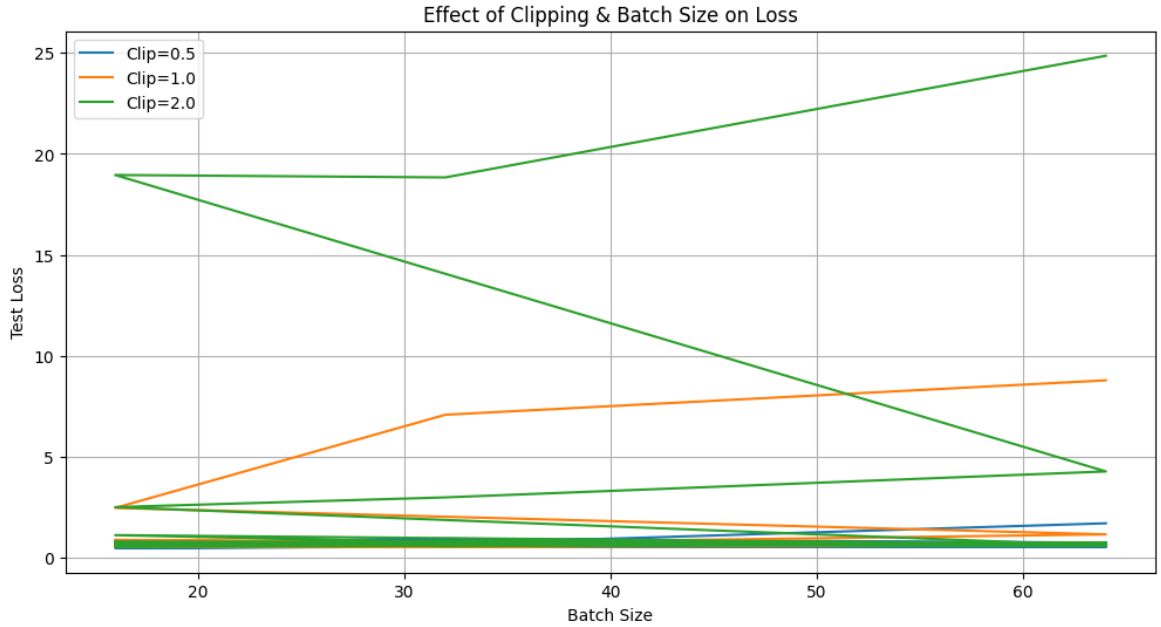
context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=0.5, Batch=64
Test Accuracy: 0.7935 | Test Loss: 0.5481
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=1.0, Batch=16
Test Accuracy: 0.4891 | Test Loss: 0.8905
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=1.0, Batch=32
Test Accuracy: 0.7446 | Test Loss: 0.7286
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=1.0, Batch=64
Test Accuracy: 0.6413 | Test Loss: 1.1672
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=2.0, Batch=16
Test Accuracy: 0.7717 | Test Loss: 2.5156
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=2.0, Batch=32

```
Test Accuracy: 0.5598 | Test Loss: 2.9981
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=2.0, Batch=64
Test Accuracy: 0.7337 | Test Loss: 4.2824
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=0.5, Batch=16
Test Accuracy: 0.7283 | Test Loss: 0.8651
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=0.5, Batch=32
Test Accuracy: 0.6793 | Test Loss: 0.7004
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=0.5, Batch=64
Test Accuracy: 0.5326 | Test Loss: 1.7154
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=1.0, Batch=16
Test Accuracy: 0.3967 | Test Loss: 2.4769
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 3ms/step
```

WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=1.0, Batch=32
Test Accuracy: 0.4293 | Test Loss: 7.0891
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=1.0, Batch=64
Test Accuracy: 0.6793 | Test Loss: 8.7917
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=2.0, Batch=16
Test Accuracy: 0.5489 | Test Loss: 18.9567
6/6 [==============================] - 0s 0s/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=2.0, Batch=32
Test Accuracy: 0.7283 | Test Loss: 18.8377
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 0s/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=2.0, Batch=64
Test Accuracy: 0.6630 | Test Loss: 24.8581
6/6 [==============================] - 0s 3ms/step
6/6 [==============================] - 0s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate

context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.



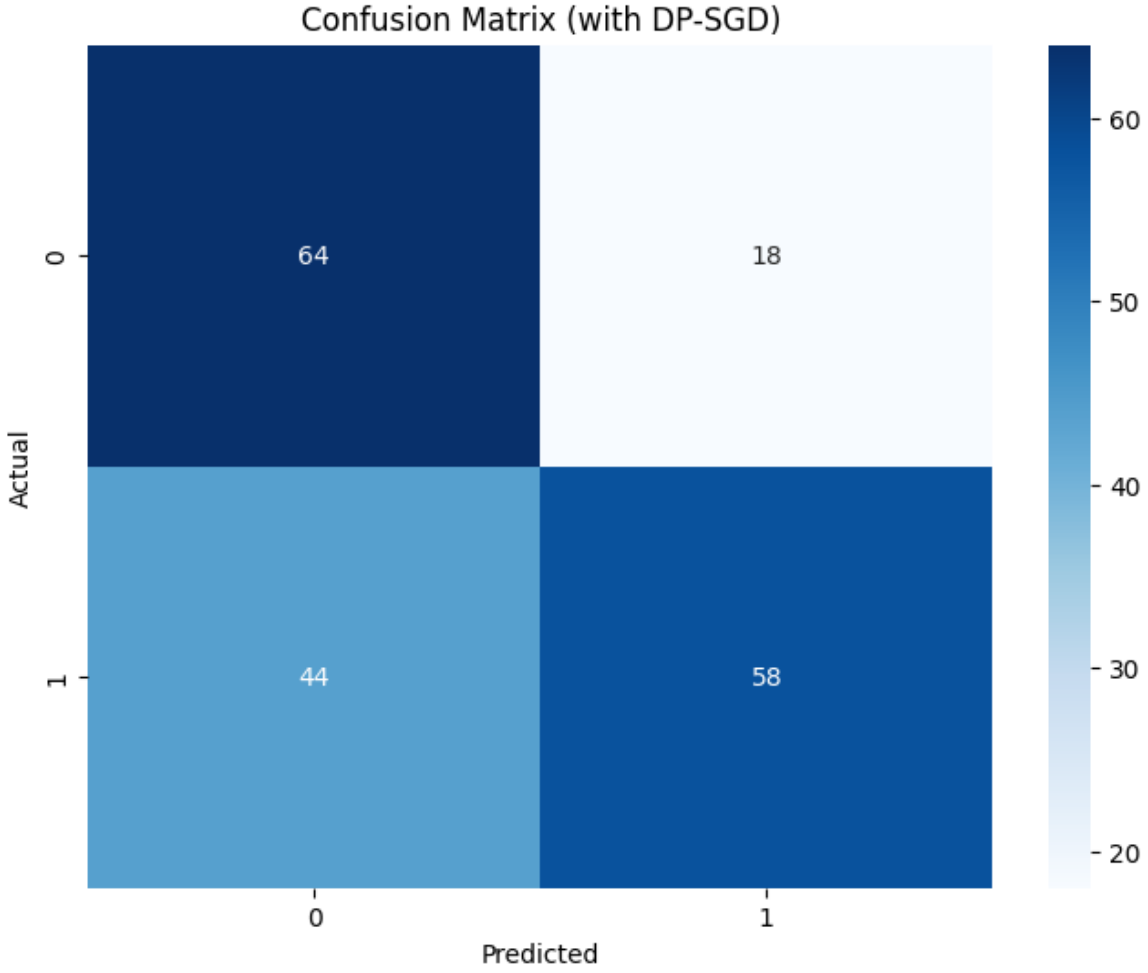


Neos kodikas
6/6 [==============================] - 0s 3ms/step

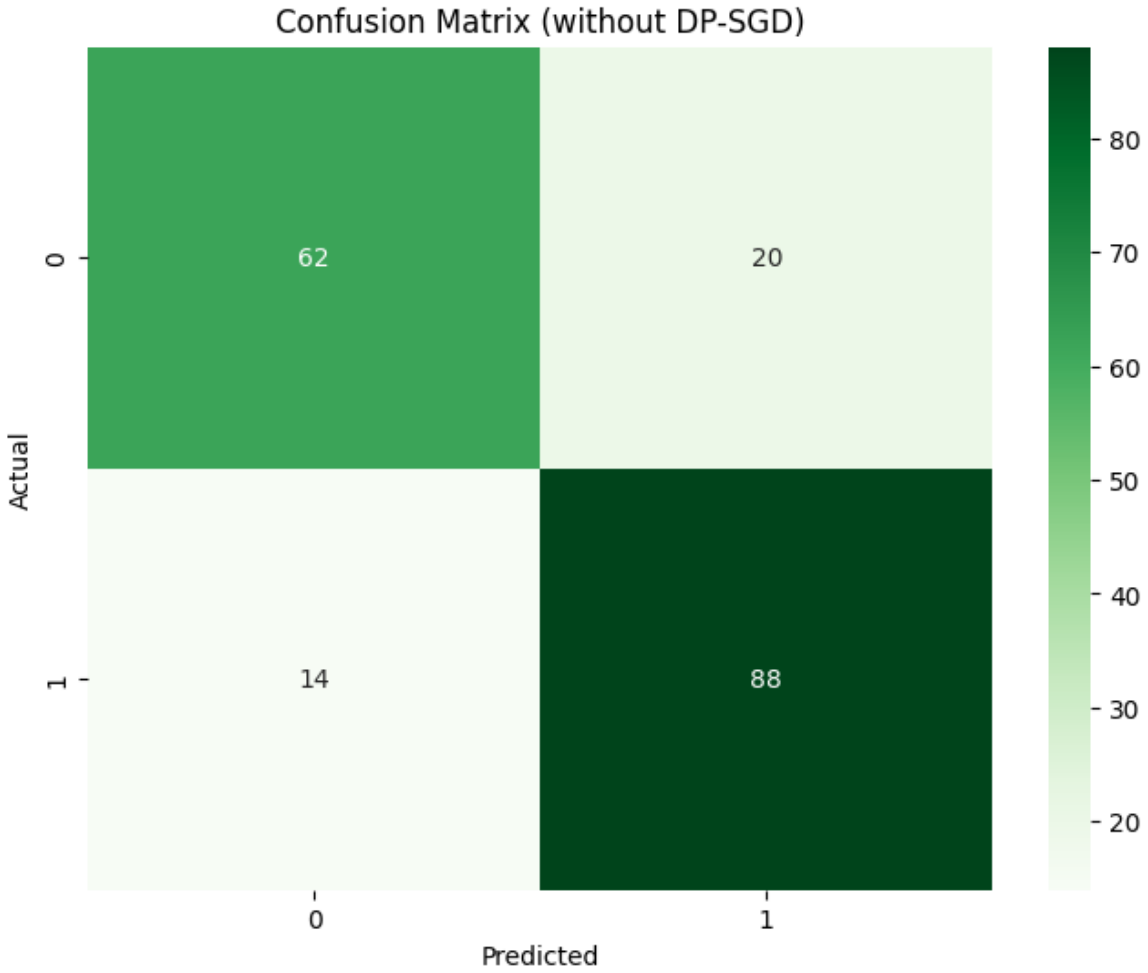
Classification Report (Dataset with DP-SGD):

	precision	recall	f1-score	support
0	0.59	0.78	0.67	82
1	0.76	0.57	0.65	102
accuracy			0.66	184
macro avg	0.68	0.67	0.66	184
weighted avg	0.69	0.66	0.66	184

6/6 [==============================] - 0s 3ms/step
ROC-AUC Score: 0.7174



6/6 [==============================] - 0s 0s/step



WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
False Positives (with DP-SGD): 18, False Negatives (with DP-SGD): 44
False Positives (no DP-SGD): 20, False Negatives (no DP-SGD): 14

Statistical Analysis of Results:
Mean Accuracy: 0.5796, Std Accuracy: 0.1143
Mean Loss: 1.7635, Std Loss: 4.0111

Delta (δ) used for experiments: 0.00005008
The delta value ensures theoretical guarantees within differential privacy context.

Relationship between privacy budget (ε) and noise multiplier:
Noise Multiplier: 0.5 => Epsilon (ε): 32.5415
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate

context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Noise Multiplier: 1.1 => Epsilon (ε): 4.2341
Noise Multiplier: 2.0 => Epsilon (ε): 1.7177

Results saved successfully in 'results_df_Hyperparameter_Tuning_heart_disease.csv'

Παράρτημα Β: Cardiovascular Disease Dataset

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import tensorflow as tf
import tensorflow_privacy
import matplotlib.pyplot as plt
from tensorflow_privacy.privacy.optimizers.dp_optimizer_keras import DPKerasSGDOptimizer
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy_lib

# 1. Φόρτωση του Cardiovascular Dataset
dataset_path = "cardio_train.csv"
df_cardio = pd.read_csv(dataset_path, sep=";")  # Το dataset χρησιμοποιεί ";" ως διαχωριστικό

# 2. Καθαρισμός των δεδομένων
df_cardio.drop(columns=["id"], inplace=True)  # Αφαίρεση του ID γιατί δεν έχει νόημα στην εκπαίδευση
df_cardio = df_cardio[df_cardio["height"] > 100]  # Αφαιρούμε ανωμαλίες στα δεδομένα
df_cardio = df_cardio[df_cardio["weight"] > 30]  # Αφαιρούμε ακραίες τιμές

# 3. Διαχωρισμός χαρακτηριστικών (X) και ετικετών (y)
X_cardio = df_cardio.drop(columns=['cardio'])  # Χαρακτηριστικά
y_cardio = df_cardio['cardio']  # Στόχος (0 = χωρίς νόσο, 1 = με νόσο)

# 4. Κανονικοποίηση των χαρακτηριστικών
scaler = StandardScaler()
X_cardio_scaled = scaler.fit_transform(X_cardio)

# 5. Διαχωρισμός σε train και test set
X_train_cardio, X_test_cardio, y_train_cardio, y_test_cardio = train_test_split(
    X_cardio_scaled, y_cardio, test_size=0.2, random_state=42, stratify=y_cardio
)

# Υπερπαράμετροι
learning_rate = 0.01
noise_multiplier = 1.1
l2_norm_clip = 1.0
epochs = 50
batch_sizes = [32, 64]
results = []
for batch_size in batch_sizes:
    print(f"\n--- Εκπαίδευση με batch size: {batch_size} ---")

    # Δημιουργία νέου μοντέλου
    model_cardio = tf.keras.Sequential([
        tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train_cardio.shape[1],)),
        tf.keras.layers.Dense(8, activation='relu'),
        tf.keras.layers.Dense(1, activation='sigmoid')
    ])

    dp_optimizer_cardio = DPKerasSGDOptimizer(
        l2_norm_clip=l2_norm_clip,
        noise_multiplier=noise_multiplier,
        num_microbatches=1,
        learning_rate=learning_rate
    )
```

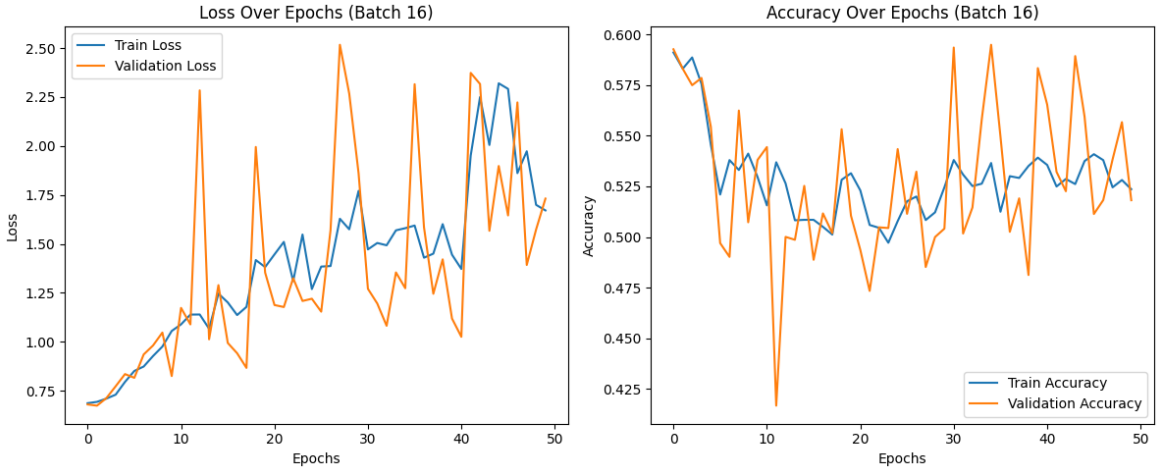
```
    model_cardio.compile(optimizer=dp_optimizer_cardio, loss='binary_crossentropy', metrics=['accuracy'])

    # Εκπαίδευση
    history_cardio = model_cardio.fit(
        X_train_cardio, y_train_cardio,
        epochs=epochs,
        batch_size=batch_size,
        validation_data=(X_test_cardio, y_test_cardio),
        verbose=1
    )

    # Αξιολόγηση
    test_loss_cardio, test_acc_cardio = model_cardio.evaluate(X_test_cardio, y_test_cardio, verbose=0)
    print(f"Test Accuracy: {test_acc_cardio:.4f}")

    # Υπολογισμός ε
    eps, _ = compute_dp_sgd_privacy_lib.compute_dp_sgd_privacy(
        n=len(X_train_cardio),
        batch_size=batch_size,
        noise_multiplier=noise_multiplier,
        epochs=epochs,
        delta=1e-5
    )
    print(f"Estimated ε for batch size {batch_size}: {eps:.2f}")

    # Σχεδίαση διαγραμμάτων
    plt.figure(figsize=(12, 5))

    plt.subplot(1, 2, 1)
    plt.plot(history_cardio.history['loss'], label='Train Loss')
    plt.plot(history_cardio.history['val_loss'], label='Validation Loss')
    plt.xlabel('Epochs')
    plt.ylabel('Loss')
    plt.title(f'Loss Over Epochs (Batch {batch_size})')
    plt.legend()

    plt.subplot(1, 2, 2)
    plt.plot(history_cardio.history['accuracy'], label='Train Accuracy')
    plt.plot(history_cardio.history['val_accuracy'], label='Validation Accuracy')
    plt.xlabel('Epochs')
    plt.ylabel('Accuracy')
    plt.title(f'Accuracy Over Epochs (Batch {batch_size})')
    plt.legend()

    plt.tight_layout()
    plt.show()

    # 8. Αποθήκευση αποτελεσμάτων
results_df = pd.DataFrame(results)
results_df.to_csv("results_df_Hyperparameter_Tuning_Cardiovascular_Dataset.csv", index=False)
print("Αποθηκεύτηκαν τα αποτελέσματα στο αρχείο CSV.")
```

```
--- Εκπαίδευση με batch size: 16 ---
Epoch 1/50
3498/3498 [==============================] - 12s 3ms/step - loss: 0.6861 - accuracy: 0.5911 -
val_loss: 0.6798 - val_accuracy: 0.5926
Epoch 2/50
```

```
3498/3498 [==============================] - 9s 3ms/step - loss: 0.6929 - accuracy: 0.5831 -
val_loss: 0.6735 - val_accuracy: 0.5830
Epoch 3/50
3498/3498 [==============================] - 9s 3ms/step - loss: 0.7094 - accuracy: 0.5886 -
val_loss: 0.7116 - val_accuracy: 0.5749
Epoch 4/50
3498/3498 [==============================] - 9s 3ms/step - loss: 0.7296 - accuracy: 0.5756 -
val_loss: 0.7724 - val_accuracy: 0.5785
Epoch 5/50
3498/3498 [==============================] - 9s 3ms/step - loss: 0.7948 - accuracy: 0.5456 -
val_loss: 0.8348 - val_accuracy: 0.5544
Epoch 6/50
3498/3498 [==============================] - 9s 3ms/step - loss: 0.8508 - accuracy: 0.5210 -
val_loss: 0.8157 - val_accuracy: 0.4970
Epoch 7/50
3498/3498 [==============================] - 9s 3ms/step - loss: 0.8734 - accuracy: 0.5380 -
val_loss: 0.9351 - val_accuracy: 0.4902
Epoch 8/50
3498/3498 [==============================] - 9s 3ms/step - loss: 0.9269 - accuracy: 0.5331 -
val_loss: 0.9808 - val_accuracy: 0.5624
Epoch 9/50
3498/3498 [==============================] - 9s 3ms/step - loss: 0.9753 - accuracy: 0.5411 -
val_loss: 1.0468 - val_accuracy: 0.5073
Epoch 10/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.0550 - accuracy: 0.5296 -
val_loss: 0.8246 - val_accuracy: 0.5380
Epoch 11/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.0887 - accuracy: 0.5156 -
val_loss: 1.1728 - val_accuracy: 0.5444
Epoch 12/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.1387 - accuracy: 0.5369 -
val_loss: 1.0885 - val_accuracy: 0.4168
Epoch 13/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.1392 - accuracy: 0.5265 -
val_loss: 2.2840 - val_accuracy: 0.5001
Epoch 14/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.0665 - accuracy: 0.5083 -
val_loss: 1.0125 - val_accuracy: 0.4986
Epoch 15/50
3498/3498 [==============================] - 10s 3ms/step - loss: 1.2458 - accuracy: 0.5085 -
val_loss: 1.2891 - val_accuracy: 0.5252
Epoch 16/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.2007 - accuracy: 0.5085 -
val_loss: 0.9938 - val_accuracy: 0.4888
Epoch 17/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.1363 - accuracy: 0.5050 -
val_loss: 0.9414 - val_accuracy: 0.5116
Epoch 18/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.1780 - accuracy: 0.5012 -
val_loss: 0.8664 - val_accuracy: 0.5019
Epoch 19/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.4172 - accuracy: 0.5282 -
val_loss: 1.9949 - val_accuracy: 0.5532
Epoch 20/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.3814 - accuracy: 0.5315 -
val_loss: 1.3529 - val_accuracy: 0.5105
Epoch 21/50
```

```
3498/3498 [==============================] - 9s 3ms/step - loss: 1.4462 - accuracy: 0.5230 -
val_loss: 1.1875 - val_accuracy: 0.4936
Epoch 22/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.5100 - accuracy: 0.5060 -
val_loss: 1.1773 - val_accuracy: 0.4734
Epoch 23/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.3096 - accuracy: 0.5045 -
val_loss: 1.3246 - val_accuracy: 0.5048
Epoch 24/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.5477 - accuracy: 0.4972 -
val_loss: 1.2082 - val_accuracy: 0.5044
Epoch 25/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.2691 - accuracy: 0.5078 -
val_loss: 1.2196 - val_accuracy: 0.5434
Epoch 26/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.3839 - accuracy: 0.5176 -
val_loss: 1.1537 - val_accuracy: 0.5114
Epoch 27/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.3867 - accuracy: 0.5200 -
val_loss: 1.5704 - val_accuracy: 0.5322
Epoch 28/50
3498/3498 [==============================] - 10s 3ms/step - loss: 1.6281 - accuracy: 0.5084 -
val_loss: 2.5168 - val_accuracy: 0.4852
Epoch 29/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.5743 - accuracy: 0.5122 -
val_loss: 2.2673 - val_accuracy: 0.5000
Epoch 30/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.7707 - accuracy: 0.5245 -
val_loss: 1.8608 - val_accuracy: 0.5041
Epoch 31/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.4716 - accuracy: 0.5380 -
val_loss: 1.2704 - val_accuracy: 0.5935
Epoch 32/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.5044 - accuracy: 0.5308 -
val_loss: 1.1953 - val_accuracy: 0.5017
Epoch 33/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.4927 - accuracy: 0.5251 -
val_loss: 1.0815 - val_accuracy: 0.5146
Epoch 34/50
3498/3498 [==============================] - 11s 3ms/step - loss: 1.5691 - accuracy: 0.5263 -
val_loss: 1.3540 - val_accuracy: 0.5581
Epoch 35/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.5801 - accuracy: 0.5365 -
val_loss: 1.2733 - val_accuracy: 0.5948
Epoch 36/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.5935 - accuracy: 0.5125 -
val_loss: 2.3154 - val_accuracy: 0.5496
Epoch 37/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.4292 - accuracy: 0.5300 -
val_loss: 1.5854 - val_accuracy: 0.5026
Epoch 38/50
3498/3498 [==============================] - 9s 3ms/step - loss: 1.4498 - accuracy: 0.5291 -
val_loss: 1.2449 - val_accuracy: 0.5191
Epoch 39/50
3498/3498 [==============================] - 10s 3ms/step - loss: 1.6007 - accuracy: 0.5350 -
val_loss: 1.4208 - val_accuracy: 0.4813
Epoch 40/50
```

3498/3498 [==========] - 10s 3ms/step - loss: 1.4446 - accuracy: 0.5392 - val_loss: 1.1185 - val_accuracy: 0.5833
Epoch 41/50
3498/3498 [==========] - 9s 3ms/step - loss: 1.3718 - accuracy: 0.5356 - val_loss: 1.0251 - val_accuracy: 0.5653
Epoch 42/50
3498/3498 [==========] - 10s 3ms/step - loss: 1.9508 - accuracy: 0.5249 - val_loss: 2.3732 - val_accuracy: 0.5322
Epoch 43/50
3498/3498 [==========] - 9s 3ms/step - loss: 2.2499 - accuracy: 0.5286 - val_loss: 2.3163 - val_accuracy: 0.5226
Epoch 44/50
3498/3498 [==========] - 9s 3ms/step - loss: 2.0049 - accuracy: 0.5261 - val_loss: 1.5668 - val_accuracy: 0.5893
Epoch 45/50
3498/3498 [==========] - 9s 3ms/step - loss: 2.3199 - accuracy: 0.5375 - val_loss: 1.8975 - val_accuracy: 0.5594
Epoch 46/50
3498/3498 [==========] - 9s 3ms/step - loss: 2.2915 - accuracy: 0.5409 - val_loss: 1.6453 - val_accuracy: 0.5114
Epoch 47/50
3498/3498 [==========] - 9s 3ms/step - loss: 1.8612 - accuracy: 0.5379 - val_loss: 2.2224 - val_accuracy: 0.5182
Epoch 48/50
3498/3498 [==========] - 9s 3ms/step - loss: 1.9729 - accuracy: 0.5245 - val_loss: 1.3921 - val_accuracy: 0.5383
Epoch 49/50
3498/3498 [==========] - 9s 3ms/step - loss: 1.6992 - accuracy: 0.5281 - val_loss: 1.5735 - val_accuracy: 0.5567
Epoch 50/50
3498/3498 [==========] - 9s 3ms/step - loss: 1.6708 - accuracy: 0.5236 - val_loss: 1.7310 - val_accuracy: 0.5182
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Test Accuracy: 0.5182
Estimated ε for batch size 16: 0.61

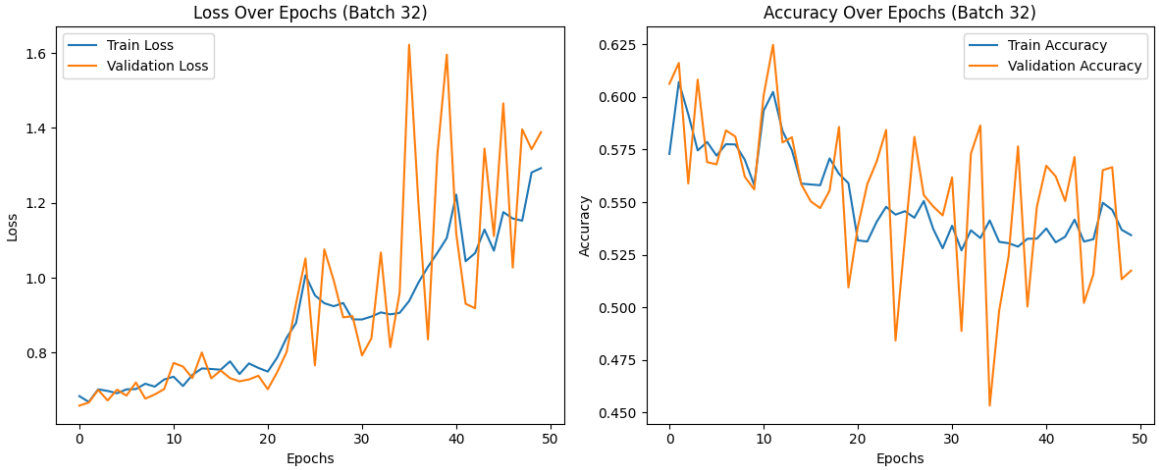


--- Εκπαίδευση με batch size: 32 ---
Epoch 1/50

```
1749/1749 [==============================] - 6s 3ms/step - loss: 0.6843 - accuracy: 0.5730 -
val_loss: 0.6588 - val_accuracy: 0.6063
Epoch 2/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.6687 - accuracy: 0.6072 -
val_loss: 0.6672 - val_accuracy: 0.6162
Epoch 3/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7023 - accuracy: 0.5918 -
val_loss: 0.7009 - val_accuracy: 0.5589
Epoch 4/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.6982 - accuracy: 0.5746 -
val_loss: 0.6727 - val_accuracy: 0.6083
Epoch 5/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.6919 - accuracy: 0.5786 -
val_loss: 0.7014 - val_accuracy: 0.5690
Epoch 6/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.7024 - accuracy: 0.5721 -
val_loss: 0.6856 - val_accuracy: 0.5680
Epoch 7/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7031 - accuracy: 0.5776 -
val_loss: 0.7207 - val_accuracy: 0.5841
Epoch 8/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7173 - accuracy: 0.5774 -
val_loss: 0.6775 - val_accuracy: 0.5812
Epoch 9/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.7096 - accuracy: 0.5701 -
val_loss: 0.6887 - val_accuracy: 0.5621
Epoch 10/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7287 - accuracy: 0.5581 -
val_loss: 0.7029 - val_accuracy: 0.5560
Epoch 11/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7360 - accuracy: 0.5935 -
val_loss: 0.7729 - val_accuracy: 0.6008
Epoch 12/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7112 - accuracy: 0.6024 -
val_loss: 0.7632 - val_accuracy: 0.6249
Epoch 13/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7408 - accuracy: 0.5839 -
val_loss: 0.7325 - val_accuracy: 0.5784
Epoch 14/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7581 - accuracy: 0.5747 -
val_loss: 0.8008 - val_accuracy: 0.5808
Epoch 15/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7567 - accuracy: 0.5588 -
val_loss: 0.7318 - val_accuracy: 0.5585
Epoch 16/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7548 - accuracy: 0.5584 -
val_loss: 0.7527 - val_accuracy: 0.5504
Epoch 17/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7770 - accuracy: 0.5580 -
val_loss: 0.7322 - val_accuracy: 0.5472
Epoch 18/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7430 - accuracy: 0.5708 -
val_loss: 0.7237 - val_accuracy: 0.5555
Epoch 19/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.7716 - accuracy: 0.5634 -
val_loss: 0.7287 - val_accuracy: 0.5858
Epoch 20/50
```

```
1749/1749 [==============================] - 5s 3ms/step - loss: 0.7595 - accuracy: 0.5590 -
val_loss: 0.7387 - val_accuracy: 0.5094
Epoch 21/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.7498 - accuracy: 0.5318 -
val_loss: 0.7027 - val_accuracy: 0.5383
Epoch 22/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.7872 - accuracy: 0.5313 -
val_loss: 0.7482 - val_accuracy: 0.5585
Epoch 23/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.8403 - accuracy: 0.5406 -
val_loss: 0.8022 - val_accuracy: 0.5693
Epoch 24/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.8794 - accuracy: 0.5478 -
val_loss: 0.9315 - val_accuracy: 0.5843
Epoch 25/50
1749/1749 [==============================] - 4s 2ms/step - loss: 1.0072 - accuracy: 0.5441 -
val_loss: 1.0517 - val_accuracy: 0.4842
Epoch 26/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.9534 - accuracy: 0.5457 -
val_loss: 0.7664 - val_accuracy: 0.5329
Epoch 27/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.9322 - accuracy: 0.5426 -
val_loss: 1.0760 - val_accuracy: 0.5810
Epoch 28/50
1749/1749 [==============================] - 5s 3ms/step - loss: 0.9242 - accuracy: 0.5505 -
val_loss: 0.9929 - val_accuracy: 0.5533
Epoch 29/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.9330 - accuracy: 0.5373 -
val_loss: 0.8947 - val_accuracy: 0.5480
Epoch 30/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.8891 - accuracy: 0.5280 -
val_loss: 0.8975 - val_accuracy: 0.5437
Epoch 31/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.8889 - accuracy: 0.5388 -
val_loss: 0.7929 - val_accuracy: 0.5617
Epoch 32/50
1749/1749 [==============================] - 5s 3ms/step - loss: 0.8968 - accuracy: 0.5271 -
val_loss: 0.8387 - val_accuracy: 0.4887
Epoch 33/50
1749/1749 [==============================] - 5s 3ms/step - loss: 0.9077 - accuracy: 0.5366 -
val_loss: 1.0677 - val_accuracy: 0.5731
Epoch 34/50
1749/1749 [==============================] - 5s 3ms/step - loss: 0.9025 - accuracy: 0.5329 -
val_loss: 0.8151 - val_accuracy: 0.5865
Epoch 35/50
1749/1749 [==============================] - 4s 3ms/step - loss: 0.9066 - accuracy: 0.5413 -
val_loss: 0.9613 - val_accuracy: 0.4532
Epoch 36/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.9380 - accuracy: 0.5311 -
val_loss: 1.6222 - val_accuracy: 0.4984
Epoch 37/50
1749/1749 [==============================] - 4s 2ms/step - loss: 0.9879 - accuracy: 0.5305 -
val_loss: 1.1947 - val_accuracy: 0.5242
Epoch 38/50
1749/1749 [==============================] - 4s 3ms/step - loss: 1.0294 - accuracy: 0.5288 -
val_loss: 0.8356 - val_accuracy: 0.5765
Epoch 39/50
```

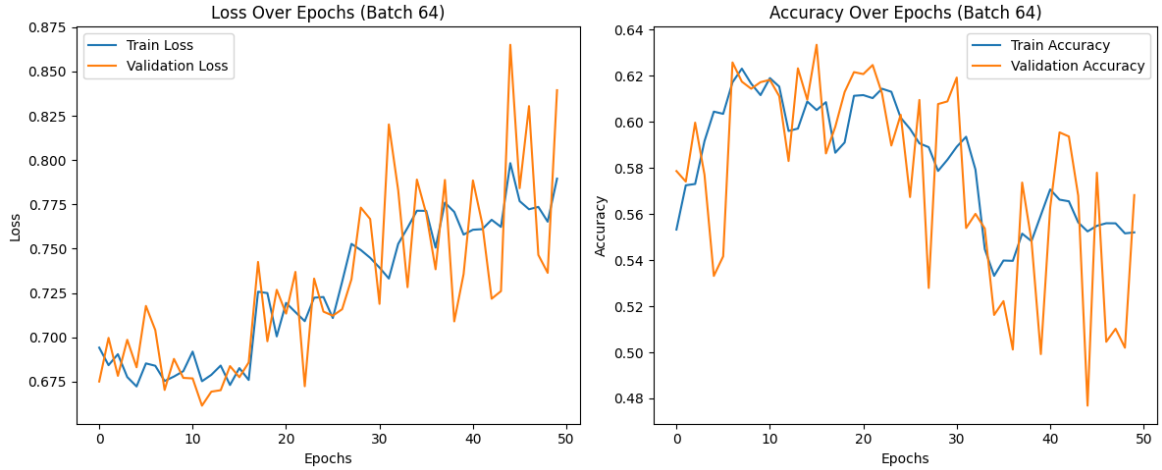
```
1749/1749 [==============================] - 4s 2ms/step - loss: 1.0668 - accuracy: 0.5326 -
val_loss: 1.3293 - val_accuracy: 0.5004
Epoch 40/50
1749/1749 [==============================] - 4s 3ms/step - loss: 1.1065 - accuracy: 0.5326 -
val_loss: 1.5959 - val_accuracy: 0.5478
Epoch 41/50
1749/1749 [==============================] - 4s 3ms/step - loss: 1.2222 - accuracy: 0.5375 -
val_loss: 1.1184 - val_accuracy: 0.5673
Epoch 42/50
1749/1749 [==============================] - 5s 3ms/step - loss: 1.0444 - accuracy: 0.5309 -
val_loss: 0.9308 - val_accuracy: 0.5622
Epoch 43/50
1749/1749 [==============================] - 4s 3ms/step - loss: 1.0659 - accuracy: 0.5335 -
val_loss: 0.9188 - val_accuracy: 0.5505
Epoch 44/50
1749/1749 [==============================] - 4s 3ms/step - loss: 1.1289 - accuracy: 0.5417 -
val_loss: 1.3450 - val_accuracy: 0.5715
Epoch 45/50
1749/1749 [==============================] - 5s 3ms/step - loss: 1.0725 - accuracy: 0.5313 -
val_loss: 1.1115 - val_accuracy: 0.5021
Epoch 46/50
1749/1749 [==============================] - 4s 2ms/step - loss: 1.1754 - accuracy: 0.5324 -
val_loss: 1.4658 - val_accuracy: 0.5157
Epoch 47/50
1749/1749 [==============================] - 5s 3ms/step - loss: 1.1583 - accuracy: 0.5497 -
val_loss: 1.0273 - val_accuracy: 0.5653
Epoch 48/50
1749/1749 [==============================] - 4s 2ms/step - loss: 1.1526 - accuracy: 0.5463 -
val_loss: 1.3966 - val_accuracy: 0.5666
Epoch 49/50
1749/1749 [==============================] - 5s 3ms/step - loss: 1.2813 - accuracy: 0.5368 -
val_loss: 1.3433 - val_accuracy: 0.5134
Epoch 50/50
1749/1749 [==============================] - 4s 3ms/step - loss: 1.2928 - accuracy: 0.5343 -
val_loss: 1.3892 - val_accuracy: 0.5174
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity
with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use
`compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute
epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the
`dp_accounting` libraries directly.
Test Accuracy: 0.5174
Estimated ε for batch size 32: 0.80
```



--- Εκπαίδευση με batch size: 64 ---
Epoch 1/50
875/875 [==============================] - 4s 3ms/step - loss: 0.6942 - accuracy: 0.5533 - val_loss: 0.6751 - val_accuracy: 0.5787
Epoch 2/50
875/875 [==============================] - 2s 3ms/step - loss: 0.6843 - accuracy: 0.5725 - val_loss: 0.6997 - val_accuracy: 0.5741
Epoch 3/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6905 - accuracy: 0.5731 - val_loss: 0.6783 - val_accuracy: 0.5997
Epoch 4/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6776 - accuracy: 0.5917 - val_loss: 0.6986 - val_accuracy: 0.5771
Epoch 5/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6723 - accuracy: 0.6045 - val_loss: 0.6831 - val_accuracy: 0.5332
Epoch 6/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6853 - accuracy: 0.6035 - val_loss: 0.7177 - val_accuracy: 0.5417
Epoch 7/50
875/875 [==============================] - 2s 3ms/step - loss: 0.6840 - accuracy: 0.6172 - val_loss: 0.7041 - val_accuracy: 0.6259
Epoch 8/50
875/875 [==============================] - 2s 3ms/step - loss: 0.6754 - accuracy: 0.6231 - val_loss: 0.6704 - val_accuracy: 0.6175
Epoch 9/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6779 - accuracy: 0.6166 - val_loss: 0.6879 - val_accuracy: 0.6144
Epoch 10/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6809 - accuracy: 0.6117 - val_loss: 0.6772 - val_accuracy: 0.6173
Epoch 11/50
875/875 [==============================] - 2s 3ms/step - loss: 0.6920 - accuracy: 0.6191 - val_loss: 0.6769 - val_accuracy: 0.6183
Epoch 12/50
875/875 [==============================] - 2s 3ms/step - loss: 0.6753 - accuracy: 0.6154 - val_loss: 0.6615 - val_accuracy: 0.6112
Epoch 13/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6789 - accuracy: 0.5961 - val_loss: 0.6694 - val_accuracy: 0.5830
Epoch 14/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6841 - accuracy: 0.5971 - val_loss: 0.6702 - val_accuracy: 0.6233
Epoch 15/50
875/875 [==============================] - 2s 2ms/step - loss: 0.6731 - accuracy: 0.6089 - val_loss: 0.6838 - val_accuracy: 0.6096
Epoch 16/50
875/875 [==============================] - 2s 3ms/step - loss: 0.6827 - accuracy: 0.6052 - val_loss: 0.6775 - val_accuracy: 0.6335
Epoch 17/50
875/875 [==============================] - 2s 3ms/step - loss: 0.6760 - accuracy: 0.6086 - val_loss: 0.6859 - val_accuracy: 0.5863
Epoch 18/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7257 - accuracy: 0.5866 - val_loss: 0.7426 - val_accuracy: 0.5979
Epoch 19/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7250 - accuracy: 0.5911 - val_loss: 0.6978 - val_accuracy: 0.6130

```
Epoch 20/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7005 - accuracy: 0.6114 -
val_loss: 0.7268 - val_accuracy: 0.6216
Epoch 21/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7194 - accuracy: 0.6117 -
val_loss: 0.7134 - val_accuracy: 0.6208
Epoch 22/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7143 - accuracy: 0.6104 -
val_loss: 0.7369 - val_accuracy: 0.6247
Epoch 23/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7091 - accuracy: 0.6144 -
val_loss: 0.6724 - val_accuracy: 0.6124
Epoch 24/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7224 - accuracy: 0.6131 -
val_loss: 0.7332 - val_accuracy: 0.5898
Epoch 25/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7228 - accuracy: 0.6018 -
val_loss: 0.7145 - val_accuracy: 0.6031
Epoch 26/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7109 - accuracy: 0.5970 -
val_loss: 0.7122 - val_accuracy: 0.5674
Epoch 27/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7314 - accuracy: 0.5907 -
val_loss: 0.7159 - val_accuracy: 0.6096
Epoch 28/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7527 - accuracy: 0.5891 -
val_loss: 0.7328 - val_accuracy: 0.5279
Epoch 29/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7493 - accuracy: 0.5787 -
val_loss: 0.7732 - val_accuracy: 0.6078
Epoch 30/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7449 - accuracy: 0.5836 -
val_loss: 0.7668 - val_accuracy: 0.6089
Epoch 31/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7394 - accuracy: 0.5892 -
val_loss: 0.7189 - val_accuracy: 0.6193
Epoch 32/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7332 - accuracy: 0.5936 -
val_loss: 0.8201 - val_accuracy: 0.5540
Epoch 33/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7527 - accuracy: 0.5794 -
val_loss: 0.7830 - val_accuracy: 0.5601
Epoch 34/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7618 - accuracy: 0.5448 -
val_loss: 0.7283 - val_accuracy: 0.5537
Epoch 35/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7714 - accuracy: 0.5332 -
val_loss: 0.7890 - val_accuracy: 0.5162
Epoch 36/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7714 - accuracy: 0.5398 -
val_loss: 0.7698 - val_accuracy: 0.5222
Epoch 37/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7506 - accuracy: 0.5397 -
val_loss: 0.7384 - val_accuracy: 0.5012
Epoch 38/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7760 - accuracy: 0.5515 -
val_loss: 0.7888 - val_accuracy: 0.5737
Epoch 39/50
```

875/875 [==============================] - 2s 2ms/step - loss: 0.7708 - accuracy: 0.5483 - val_loss: 0.7090 - val_accuracy: 0.5484
Epoch 40/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7580 - accuracy: 0.5595 - val_loss: 0.7357 - val_accuracy: 0.4992
Epoch 41/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7607 - accuracy: 0.5707 - val_loss: 0.7885 - val_accuracy: 0.5628
Epoch 42/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7610 - accuracy: 0.5663 - val_loss: 0.7640 - val_accuracy: 0.5955
Epoch 43/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7663 - accuracy: 0.5656 - val_loss: 0.7218 - val_accuracy: 0.5937
Epoch 44/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7623 - accuracy: 0.5564 - val_loss: 0.7261 - val_accuracy: 0.5680
Epoch 45/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7983 - accuracy: 0.5525 - val_loss: 0.8650 - val_accuracy: 0.4768
Epoch 46/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7768 - accuracy: 0.5549 - val_loss: 0.7841 - val_accuracy: 0.5780
Epoch 47/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7723 - accuracy: 0.5560 - val_loss: 0.8304 - val_accuracy: 0.5046
Epoch 48/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7736 - accuracy: 0.5560 - val_loss: 0.7465 - val_accuracy: 0.5102
Epoch 49/50
875/875 [==============================] - 2s 3ms/step - loss: 0.7653 - accuracy: 0.5516 - val_loss: 0.7364 - val_accuracy: 0.5020
Epoch 50/50
875/875 [==============================] - 2s 2ms/step - loss: 0.7895 - accuracy: 0.5521 - val_loss: 0.8394 - val_accuracy: 0.5683
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Test Accuracy: 0.5683
Estimated ε for batch size 64: 1.13



Αποθηκεύτηκαν τα αποτελέσματα στο αρχείο CSV.

Hyperparameter_Tuning_Cardiovascular_Dataset

```
import itertools
import pandas as pd
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
import tensorflow as tf
import tensorflow_privacy
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score
import seaborn as sns
from sklearn.metrics import precision_score, recall_score, f1_score
from tensorflow_privacy.privacy.analysis import compute_dp_sgd_privacy_lib


#  1. Φόρτωση του Cardiovascular Dataset
dataset_path = "cardio_train.csv"
df_cardio = pd.read_csv(dataset_path, sep=";")  # Το dataset χρησιμοποιεί ";" ως διαχωριστικό

#2. Καθαρισμός των δεδομένων
df_cardio.drop(columns=["id"], inplace=True)  # Αφαίρεση του ID γιατί δεν έχει νόημα στην εκπαίδευση
df_cardio = df_cardio[df_cardio["height"] > 100]  # Αφαιρούμε ανωμαλίες στα δεδομένα
df_cardio = df_cardio[df_cardio["weight"] > 30]  # Αφαιρούμε ακραίες τιμές

# 3. Διαχωρισμός χαρακτηριστικών (X) και ετικετών (y)
X_cardio = df_cardio.drop(columns=['cardio'])  # Χαρακτηριστικά
y_cardio = df_cardio['cardio']  # Στόχος (0 = χωρίς νόσο, 1 = με νόσο)

# 4. Κανονικοποίηση των χαρακτηριστικών
scaler = StandardScaler()
X_cardio_scaled = scaler.fit_transform(X_cardio)

# 5. Διαχωρισμός σε train και test set
X_train_cardio, X_test_cardio, y_train_cardio, y_test_cardio = train_test_split(
    X_cardio_scaled, y_cardio, test_size=0.2, random_state=42, stratify=y_cardio
)

print(f"Dataset loaded successfully! Train size: {X_train_cardio.shape}, Test size: {X_test_cardio.shape}")

# Ορισμός των διαφορετικών τιμών υπερπαραμέτρων για πειραματισμό
learning_rates = [0.001, 0.01, 0.1]
noise_multipliers = [0.5, 1.1, 2.0]
l2_norm_clips = [0.5, 1.0, 2.0]
batch_sizes = [16, 32, 64]
epochs = 30  # Σταθερό για όλες τις δοκιμές

# Αποθήκευση των αποτελεσμάτων
results_cardio = []

# Δοκιμή όλων των συνδυασμών υπερπαραμέτρων
for lr, noise, clip, batch in itertools.product(learning_rates, noise_multipliers, l2_norm_clips, batch_sizes):
    print(f"Training model with LR={lr}, Noise={noise}, Clip={clip}, Batch={batch}")

    # Προσαρμογή batch size
    batch_size = (len(X_train_cardio) // batch) * batch
    if batch_size == 0:
```

```
    batch_size = 1

# Δημιουργία του μοντέλου
model_cardio = tf.keras.Sequential([
    tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train_cardio.shape[1],)),
    tf.keras.layers.Dense(8, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')  # Δυαδική ταξινόμηση
])

# Χρήση του DP-SGD Optimizer
dp_optimizer_cardio = tensorflow_privacy.DPKerasSGDOptimizer(
    l2_norm_clip=clip,
    noise_multiplier=noise,
    num_microbatches=1,
    learning_rate=lr
)

# Συναρτήσεις απώλειας και μέτρησης απόδοσης
model_cardio.compile(optimizer=dp_optimizer_cardio, loss='binary_crossentropy', metrics=['accuracy'])

# Εκπαίδευση του μοντέλου με DP-SGD
history_cardio = model_cardio.fit(
    X_train_cardio, y_train_cardio, epochs=epochs, batch_size=batch_size,
validation_data=(X_test_cardio, y_test_cardio), verbose=0
)

# Αξιολόγηση του μοντέλου
test_loss_cardio, test_acc_cardio = model_cardio.evaluate(X_test_cardio, y_test_cardio, verbose=0)
print(f"Test Accuracy: {test_acc_cardio:.4f} | Test Loss: {test_loss_cardio:.4f}")

# Προβλέψεις για υπολογισμό μετρικών
y_pred = (model_cardio.predict(X_test_cardio) > 0.5).astype("int32")
y_proba = model_cardio.predict(X_test_cardio)

# Υπολογισμός μετρικών
precision = precision_score(y_test_cardio, y_pred, zero_division=0)
recall = recall_score(y_test_cardio, y_pred, zero_division=0)
f1 = f1_score(y_test_cardio, y_pred, zero_division=0)
roc_auc = roc_auc_score(y_test_cardio, y_proba)

# Υπολογισμός privacy budget (ε)
eps, _ = compute_dp_sgd_privacy_lib.compute_dp_sgd_privacy(
n=len(X_train_cardio),
batch_size=batch_size,
noise_multiplier=noise,
epochs=epochs,
delta=1 / (len(X_train_cardio) * np.sqrt(len(X_train_cardio))))


results_cardio.append({
 "learning_rate": lr,
 "noise_multiplier": noise,
 "l2_norm_clip": clip,
 "batch_size": batch,
 "test_accuracy": test_acc_cardio,
 "test_loss": test_loss_cardio,
 "precision": precision,
 "recall": recall,
```

```
    "f1_score": f1,
    "roc_auc": roc_auc,
    "epsilon": eps
   })


# Μετατροπή των αποτελεσμάτων σε DataFrame
results_df_cardio = pd.DataFrame(results_cardio)

# Εμφάνιση αποτελεσμάτων (Χρησιμοποίησε print ή display σε Jupyter Notebook)
print("Hyperparameter Tuning Results - Cardiovascular Dataset:")
print(results_df_cardio.to_string())  # Αν το DataFrame είναι μεγάλο, μπορείς να το εμφανίσεις με
print(results_df_cardio.head())

# Διαγράμματα: Επίδραση των Υπερπαραμέτρων στην Ακρίβεια
plt.figure(figsize=(12, 6))
for noise in results_df_cardio["noise_multiplier"].unique():  # Χρήση unique() για μοναδικές τιμές
    subset = results_df_cardio[results_df_cardio["noise_multiplier"] == noise]
    plt.plot(subset["learning_rate"], subset["test_accuracy"], marker='o', linestyle='-',
label=f"Noise={noise}")

plt.xlabel("Learning Rate")
plt.ylabel("Test Accuracy")
plt.title("Effect of Learning Rate & Noise on Accuracy (Cardiovascular)")
plt.legend()
plt.grid(True)  # Προσθήκη πλέγματος για καλύτερη οπτικοποίηση
plt.show()

# Διαγράμματα: Επίδραση των Υπερπαραμέτρων στην Απώλεια
plt.figure(figsize=(12, 6))
for clip in results_df_cardio["l2_norm_clip"].unique():
    subset = results_df_cardio[results_df_cardio["l2_norm_clip"] == clip]
    plt.plot(subset["batch_size"], subset["test_loss"], marker='s', linestyle='-', label=f"Clip={clip}")

plt.xlabel("Batch Size")
plt.ylabel("Test Loss")
plt.title("Effect of Clipping & Batch Size on Loss (Cardiovascular)")
plt.legend()
plt.grid(True)
plt.show()

#Neos kodikas
print("Neos kodikas")
# Υπολογισμός επιπλέον μετρικών
predictions_cardio = (model_cardio.predict(X_test_cardio) > 0.5).astype("int32")
report = classification_report(y_test_cardio, predictions_cardio, output_dict=True)
roc_auc = roc_auc_score(y_test_cardio, model_cardio.predict(X_test_cardio))

# Εκτύπωση classification report
print("\nClassification Report (Cardiovascular dataset with DP-SGD):")
print(classification_report(y_test_cardio, predictions_cardio))
print(f"ROC-AUC Score: {roc_auc:.4f}")

# Δημιουργία και εμφάνιση confusion matrix
cm = confusion_matrix(y_test_cardio, predictions_cardio)
plt.figure(figsize=(8,6))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.title("Confusion Matrix (with DP-SGD)")
```

```
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

# Σύγκριση με μοντέλο χωρίς DP-SGD
model_no_dp = tf.keras.Sequential([
    tf.keras.layers.Dense(16, activation='relu', input_shape=(X_train_cardio.shape[1],)),
    tf.keras.layers.Dense(8, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])

model_no_dp.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model_no_dp.fit(X_train_cardio, y_train_cardio, epochs=epochs, batch_size=32, verbose=0)

# Αξιολόγηση μοντέλου χωρίς DP-SGD
predictions_no_dp = (model_no_dp.predict(X_test_cardio) > 0.5).astype("int32")
cm_no_dp = confusion_matrix(y_test_cardio, predictions_no_dp)

plt.figure(figsize=(8,6))
sns.heatmap(cm_no_dp, annot=True, fmt="d", cmap="Greens")
plt.title("Confusion Matrix (without DP-SGD)")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()

# Υπολογισμός false positives και false negatives
fp_dp, fn_dp = cm[0][1], cm[1][0]
fp_no_dp, fn_no_dp = cm_no_dp[0][1], cm_no_dp[1][0]

print(f"False Positives (with DP-SGD): {fp_dp}, False Negatives (with DP-SGD): {fn_dp}")
print(f"False Positives (no DP-SGD): {fp_no_dp}, False Negatives (no DP-SGD): {fn_no_dp}")

# Στατιστική ανάλυση αποτελεσμάτων για ακρίβεια και απώλεια
mean_accuracy = results_df_cardio['test_accuracy'].mean()
std_accuracy = results_df_cardio['test_accuracy'].std()
mean_loss = results_df_cardio['test_loss'].mean()
std_loss = results_df_cardio['test_loss'].std()

print("\nStatistical Analysis of Results:")
print(f"Mean Accuracy: {mean_accuracy:.4f}, Std Accuracy: {std_accuracy:.4f}")
print(f"Mean Loss: {mean_loss:.4f}, Std Loss: {std_loss:.4f}")
# Υπολογισμός και εμφάνιση του delta και της σχέσης με noise multiplier
num_samples = X_train_cardio.shape[0]
delta = 1 / (num_samples * np.sqrt(num_samples))
print(f"\nDelta (δ) used for experiments: {delta:.8f}")
print("The delta value ensures theoretical guarantees within differential privacy context.")

# Εμφάνιση σχέσης privacy budget (ε) και noise multiplier
print("\nRelationship between privacy budget (ε) and noise multiplier:")
for noise in noise_multipliers:
    epsilon, _ = tensorflow_privacy.compute_dp_sgd_privacy(n=num_samples,
                                           batch_size=batch_sizes[0],
                                           noise_multiplier=noise,
                                           epochs=epochs,
                                           delta=delta)
    print(f"Noise Multiplier: {noise} => Epsilon (ε): {epsilon:.4f}")
```

```
Dataset loaded successfully! Train size: (55966, 11), Test size: (13992, 11)
Training model with LR=0.001, Noise=0.5, Clip=0.5, Batch=16
Test Accuracy: 0.5104 | Test Loss: 0.7575
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=0.5, Batch=32
Test Accuracy: 0.4914 | Test Loss: 0.7489
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=0.5, Batch=64
Test Accuracy: 0.4595 | Test Loss: 0.7273
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=1.0, Batch=16
Test Accuracy: 0.5215 | Test Loss: 0.7471
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=1.0, Batch=32
Test Accuracy: 0.5299 | Test Loss: 0.7052
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides
```

appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=1.0, Batch=64
Test Accuracy: 0.5202 | Test Loss: 0.7183
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=2.0, Batch=16
Test Accuracy: 0.5075 | Test Loss: 0.7199
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=2.0, Batch=32
Test Accuracy: 0.5248 | Test Loss: 0.7581
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=0.5, Clip=2.0, Batch=64
Test Accuracy: 0.5400 | Test Loss: 0.6984
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=0.5, Batch=16
Test Accuracy: 0.4914 | Test Loss: 0.7118
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step

WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=0.5, Batch=32
Test Accuracy: 0.4655 | Test Loss: 0.7225
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=0.5, Batch=64
Test Accuracy: 0.5036 | Test Loss: 0.6979
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=1.0, Batch=16
Test Accuracy: 0.5106 | Test Loss: 0.7084
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=1.0, Batch=32
Test Accuracy: 0.5041 | Test Loss: 0.7251
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=1.0, Batch=64
Test Accuracy: 0.5101 | Test Loss: 0.7348

```
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=2.0, Batch=16
Test Accuracy: 0.4973 | Test Loss: 0.7624
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=2.0, Batch=32
Test Accuracy: 0.4504 | Test Loss: 0.7137
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=1.1, Clip=2.0, Batch=64
Test Accuracy: 0.4997 | Test Loss: 0.7416
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=0.5, Batch=16
Test Accuracy: 0.4882 | Test Loss: 0.7357
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
```

```
Training model with LR=0.001, Noise=2.0, Clip=0.5, Batch=32
Test Accuracy: 0.4945 | Test Loss: 0.7158
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=0.5, Batch=64
Test Accuracy: 0.5010 | Test Loss: 0.7651
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=1.0, Batch=16
Test Accuracy: 0.4520 | Test Loss: 0.7360
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=1.0, Batch=32
Test Accuracy: 0.4818 | Test Loss: 0.7718
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=1.0, Batch=64
Test Accuracy: 0.4997 | Test Loss: 0.7784
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions
```

than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=2.0, Batch=16
Test Accuracy: 0.5480 | Test Loss: 0.6942
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=2.0, Batch=32
Test Accuracy: 0.5014 | Test Loss: 0.7451
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.001, Noise=2.0, Clip=2.0, Batch=64
Test Accuracy: 0.5157 | Test Loss: 0.7221
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=0.5, Batch=16
Test Accuracy: 0.5885 | Test Loss: 0.6890
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=0.5, Batch=32
Test Accuracy: 0.5271 | Test Loss: 0.6881
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is

rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=0.5, Batch=64
Test Accuracy: 0.5069 | Test Loss: 0.7117
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=1.0, Batch=16
Test Accuracy: 0.5023 | Test Loss: 0.7183
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=1.0, Batch=32
Test Accuracy: 0.5185 | Test Loss: 0.7048
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=1.0, Batch=64
Test Accuracy: 0.4596 | Test Loss: 0.7291
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=2.0, Batch=16
Test Accuracy: 0.4828 | Test Loss: 0.7049
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step

WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=2.0, Batch=32
Test Accuracy: 0.5017 | Test Loss: 0.7069
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=0.5, Clip=2.0, Batch=64
Test Accuracy: 0.5034 | Test Loss: 0.7176
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=0.5, Batch=16
Test Accuracy: 0.5425 | Test Loss: 0.6845
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=0.5, Batch=32
Test Accuracy: 0.5660 | Test Loss: 0.6901
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=0.5, Batch=64
Test Accuracy: 0.4961 | Test Loss: 0.7079

438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=1.0, Batch=16
Test Accuracy: 0.4896 | Test Loss: 0.7402
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=1.0, Batch=32
Test Accuracy: 0.4879 | Test Loss: 0.7053
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=1.0, Batch=64
Test Accuracy: 0.5162 | Test Loss: 0.7385
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=2.0, Batch=16
Test Accuracy: 0.4887 | Test Loss: 0.8566
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.

```
Training model with LR=0.01, Noise=1.1, Clip=2.0, Batch=32
Test Accuracy: 0.4905 | Test Loss: 0.7697
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=1.1, Clip=2.0, Batch=64
Test Accuracy: 0.5155 | Test Loss: 0.7529
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=0.5, Batch=16
Test Accuracy: 0.4981 | Test Loss: 0.7464
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=0.5, Batch=32
Test Accuracy: 0.4996 | Test Loss: 0.7050
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=0.5, Batch=64
Test Accuracy: 0.5412 | Test Loss: 0.6933
438/438 [==============================] - 2s 4ms/step
438/438 [==============================] - 1s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions
```

than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=1.0, Batch=16
Test Accuracy: 0.4790 | Test Loss: 0.7334
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=1.0, Batch=32
Test Accuracy: 0.4510 | Test Loss: 0.7651
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=1.0, Batch=64
Test Accuracy: 0.5089 | Test Loss: 0.7278
438/438 [==============================] - 2s 3ms/step
438/438 [==============================] - 2s 4ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=2.0, Batch=16
Test Accuracy: 0.4978 | Test Loss: 0.7952
438/438 [==============================] - 1s 3ms/step
438/438 [==============================] - 1s 3ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=2.0, Batch=32
Test Accuracy: 0.5140 | Test Loss: 0.8110
438/438 [==============================] - 1s 3ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is

rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.01, Noise=2.0, Clip=2.0, Batch=64
Test Accuracy: 0.4803 | Test Loss: 0.8959
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=0.5, Batch=16
Test Accuracy: 0.6112 | Test Loss: 0.6595
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=0.5, Batch=32
Test Accuracy: 0.5967 | Test Loss: 0.6730
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=0.5, Batch=64
Test Accuracy: 0.6070 | Test Loss: 0.6716
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=1.0, Batch=16
Test Accuracy: 0.5600 | Test Loss: 0.7223
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step

WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=1.0, Batch=32
Test Accuracy: 0.5738 | Test Loss: 0.7029
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=1.0, Batch=64
Test Accuracy: 0.5534 | Test Loss: 0.7202
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=2.0, Batch=16
Test Accuracy: 0.5296 | Test Loss: 0.7410
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=2.0, Batch=32
Test Accuracy: 0.4999 | Test Loss: 1.1055
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=0.5, Clip=2.0, Batch=64
Test Accuracy: 0.4245 | Test Loss: 0.8717

```
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=0.5, Batch=16
Test Accuracy: 0.5625 | Test Loss: 0.7029
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=0.5, Batch=32
Test Accuracy: 0.5355 | Test Loss: 0.7225
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=0.5, Batch=64
Test Accuracy: 0.5552 | Test Loss: 0.7110
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=1.0, Batch=16
Test Accuracy: 0.5090 | Test Loss: 1.5314
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
```

```
Training model with LR=0.1, Noise=1.1, Clip=1.0, Batch=32
Test Accuracy: 0.5362 | Test Loss: 1.5852
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=1.0, Batch=64
Test Accuracy: 0.5031 | Test Loss: 1.3553
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=2.0, Batch=16
Test Accuracy: 0.5328 | Test Loss: 5.6115
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=2.0, Batch=32
Test Accuracy: 0.5638 | Test Loss: 4.9999
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 2ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=1.1, Clip=2.0, Batch=64
Test Accuracy: 0.5264 | Test Loss: 4.8704
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions
```

than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=0.5, Batch=16
Test Accuracy: 0.4761 | Test Loss: 1.4359
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=0.5, Batch=32
Test Accuracy: 0.5601 | Test Loss: 0.9619
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=0.5, Batch=64
Test Accuracy: 0.4946 | Test Loss: 1.2909
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=1.0, Batch=16
Test Accuracy: 0.5278 | Test Loss: 10.0373
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=1.0, Batch=32
Test Accuracy: 0.5221 | Test Loss: 5.6559
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is

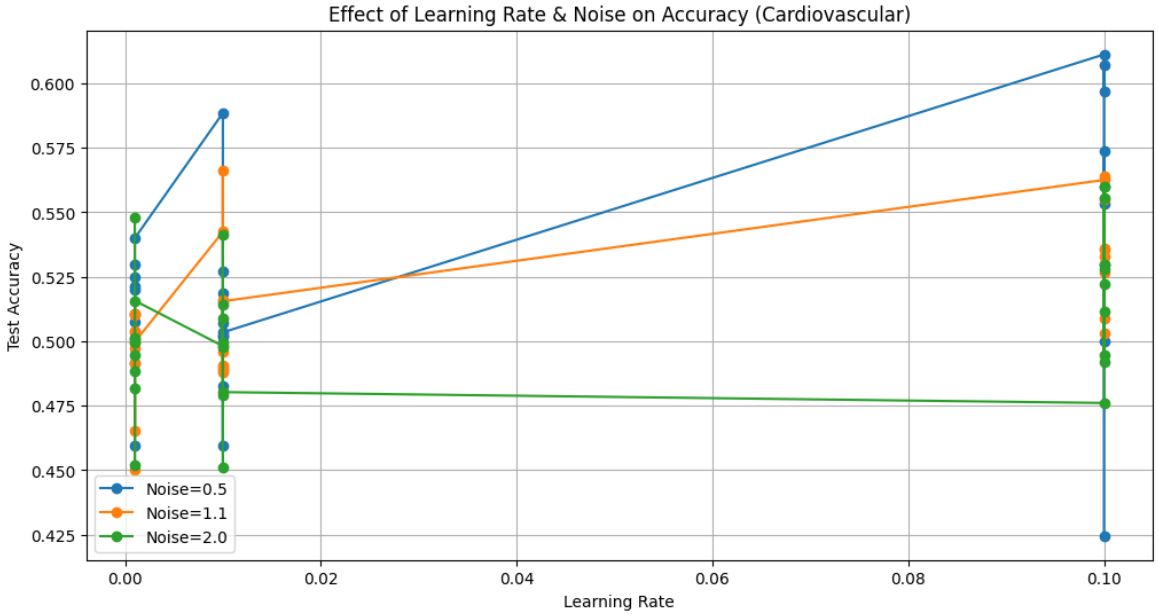
rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=1.0, Batch=64
Test Accuracy: 0.5116 | Test Loss: 4.9342
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=2.0, Batch=16
Test Accuracy: 0.5299 | Test Loss: 60.4787
438/438 [==============================] - 1s 2ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=2.0, Batch=32
Test Accuracy: 0.5555 | Test Loss: 28.5919
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Training model with LR=0.1, Noise=2.0, Clip=2.0, Batch=64
Test Accuracy: 0.4919 | Test Loss: 53.0963
438/438 [==============================] - 1s 1ms/step
438/438 [==============================] - 1s 1ms/step
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
Hyperparameter Tuning Results - Cardiovascular Dataset:
   learning_rate noise_multiplier l2_norm_clip batch_size test_accuracy test_loss precision recall f1_score roc_auc epsilon

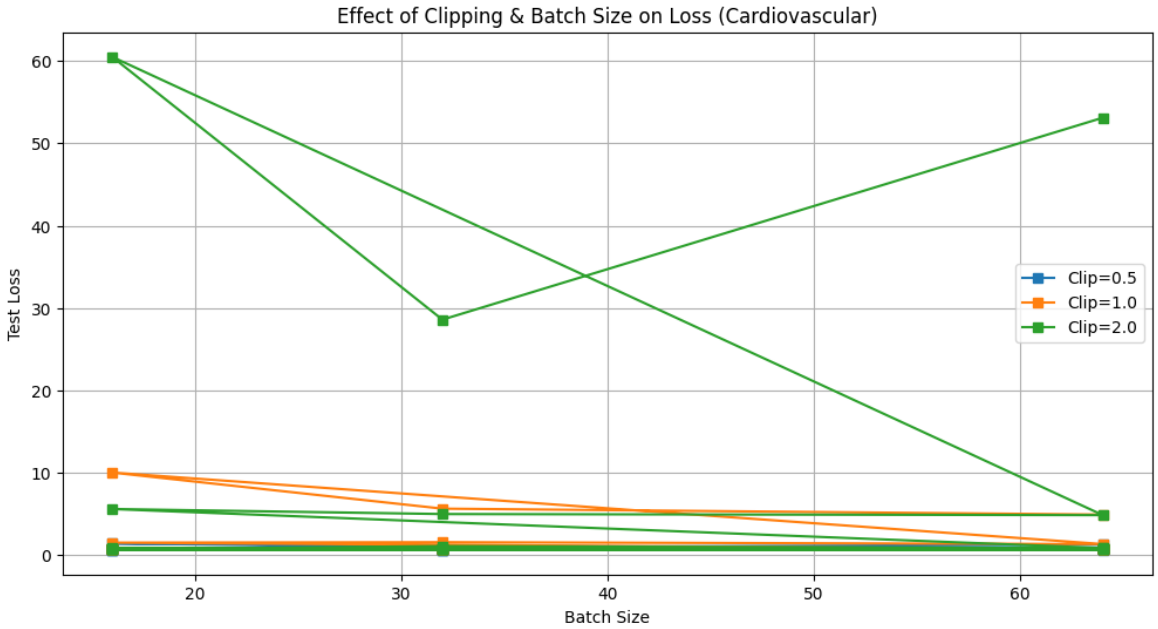
0	0.001	0.5	0.5	16	0.510435	0.757503	0.667453	0.040475	0.076321	0.508866	123.867840
1	0.001	0.5	0.5	32	0.491352	0.748859	0.479420	0.208238	0.290358	0.497814	123.844848
2	0.001	0.5	0.5	64	0.459477	0.727306	0.459429	0.462386	0.460902	0.447211	123.844848
3	0.001	0.5	1.0	16	0.521512	0.747144	0.562791	0.190360	0.284493	0.523306	123.867840
4	0.001	0.5	1.0	32	0.529874	0.705210	0.624101	0.148884	0.240416	0.566809	123.844848
5	0.001	0.5	1.0	64	0.520154	0.718290	0.520666	0.500858	0.510570	0.524118	123.844848
6	0.001	0.5	2.0	16	0.507504	0.719925	0.503832	0.949514	0.658337	0.568485	123.867840
7	0.001	0.5	2.0	32	0.524800	0.758111	0.629826	0.118993	0.200168	0.489819	123.844848
8	0.001	0.5	2.0	64	0.539951	0.698426	0.537364	0.570795	0.553575	0.554720	123.844848
9	0.001	1.1	0.5	16	0.491424	0.711787	0.495430	0.961384	0.653891	0.547456	40.623567
10	0.001	1.1	0.5	32	0.465480	0.722491	0.477740	0.747426	0.582901	0.463278	40.613598
11	0.001	1.1	0.5	64	0.503573	0.697929	0.512835	0.131436	0.209244	0.520946	40.613598
12	0.001	1.1	1.0	16	0.510577	0.708415	0.525514	0.212100	0.302221	0.494886	40.623567
13	0.001	1.1	1.0	32	0.504074	0.725051	0.726496	0.012157	0.023913	0.602304	40.613598
14	0.001	1.1	1.0	64	0.510077	0.734831	0.526979	0.191362	0.280768	0.533020	40.613598
15	0.001	1.1	2.0	16	0.497284	0.762448	0.498468	0.977546	0.660259	0.523586	40.623567
16	0.001	1.1	2.0	32	0.450400	0.713683	0.463255	0.629291	0.533657	0.471932	40.613598
17	0.001	1.1	2.0	64	0.499714	0.741647	0.499714	1.000000	0.666413	0.475813	40.613598
18	0.001	2.0	0.5	16	0.488208	0.735678	0.443098	0.094108	0.155244	0.407197	18.865027
19	0.001	2.0	0.5	32	0.494497	0.715765	0.495913	0.702803	0.581504	0.509036	18.859796
20	0.001	2.0	0.5	64	0.501001	0.765127	0.500409	0.874285	0.636506	0.505273	18.859796
21	0.001	2.0	1.0	16	0.451973	0.735953	0.462947	0.603976	0.524140	0.439630	18.865027
22	0.001	2.0	1.0	32	0.481847	0.771790	0.490321	0.934640	0.643209	0.466756	18.859796
23	0.001	2.0	1.0	64	0.499714	0.778415	0.499714	1.000000	0.666413	0.394210	18.859796

24	0.001	2.0	2.0	16	0.547956	0.694180	0.538901	
0.660755	0.593640	0.564517	18.865027					
25	0.001	2.0	2.0	32	0.501358	0.745141	0.500540	
0.993993	0.665804	0.569623	18.859796					
26	0.001	2.0	2.0	64	0.515723	0.722096	0.509485	
0.829662	0.631298	0.562154	18.859796					
27	0.010	0.5	0.5	16	0.588479	0.688994	0.688339	
0.322511	0.439229	0.637224	123.867840					
28	0.010	0.5	0.5	32	0.527087	0.688128	0.516065	
0.861413	0.645448	0.594980	123.844848					
29	0.010	0.5	0.5	64	0.506861	0.711699	0.676923	
0.025172	0.048538	0.568398	123.844848					
30	0.010	0.5	1.0	16	0.502287	0.718331	0.501997	
0.503432	0.502714	0.511289	123.867840					
31	0.010	0.5	1.0	32	0.518511	0.704819	0.525403	
0.377145	0.439097	0.500403	123.844848					
32	0.010	0.5	1.0	64	0.459620	0.729067	0.471734	
0.679062	0.556722	0.446459	123.844848					
33	0.010	0.5	2.0	16	0.482847	0.704911	0.486304	
0.619565	0.544906	0.513429	123.867840					
34	0.010	0.5	2.0	32	0.501715	0.706872	0.501943	
0.369565	0.425700	0.495528	123.844848					
35	0.010	0.5	2.0	64	0.503431	0.717556	0.504962	
0.320223	0.391913	0.497404	123.844848					
36	0.010	1.1	0.5	16	0.542453	0.684490	0.572804	
0.331951	0.420319	0.582262	40.623567					
37	0.010	1.1	0.5	32	0.566038	0.690054	0.552166	
0.696367	0.615939	0.589539	40.613598					
38	0.010	1.1	0.5	64	0.496069	0.707869	0.497030	
0.706093	0.583397	0.505117	40.613598					
39	0.010	1.1	1.0	16	0.489565	0.740218	0.491552	
0.624142	0.549968	0.471805	40.623567					
40	0.010	1.1	1.0	32	0.487922	0.705296	0.489251	
0.563072	0.523572	0.483481	40.613598					
41	0.010	1.1	1.0	64	0.516152	0.738540	0.563574	
0.140732	0.225223	0.541012	40.613598					
42	0.010	1.1	2.0	16	0.488708	0.856627	0.493599	
0.893307	0.635855	0.502693	40.623567					
43	0.010	1.1	2.0	32	0.490495	0.769729	0.433040	
0.063358	0.110543	0.516085	40.613598					
44	0.010	1.1	2.0	64	0.515509	0.752873	0.547061	
0.177059	0.267531	0.476464	40.613598					
45	0.010	2.0	0.5	16	0.498142	0.746420	0.498735	
0.845824	0.627480	0.553421	18.865027					
46	0.010	2.0	0.5	32	0.499643	0.704979	0.499678	
0.999714	0.666317	0.507017	18.859796					
47	0.010	2.0	0.5	64	0.541238	0.693284	0.530976	
0.702374	0.604766	0.570008	18.859796					

48	0.010	2.0	1.0	16	0.478988	0.733402	0.482917
0.602403	0.536082	0.486703	18.865027				
49	0.010	2.0	1.0	32	0.451043	0.765095	0.403149
0.205092	0.271874	0.390221	18.859796				
50	0.010	2.0	1.0	64	0.508862	0.727771	0.505872
0.739273	0.600697	0.540085	18.859796				
51	0.010	2.0	2.0	16	0.497784	0.795204	0.498740
0.990990	0.663538	0.498161	18.865027				
52	0.010	2.0	2.0	32	0.514008	0.811025	0.517285
0.410898	0.457995	0.510723	18.859796				
53	0.010	2.0	2.0	64	0.480274	0.895896	0.427536
0.118135	0.185119	0.412992	18.859796				
54	0.100	0.5	0.5	16	0.611206	0.659466	0.583621
0.774600	0.665683	0.668512	123.867840				
55	0.100	0.5	0.5	32	0.596698	0.672994	0.632906
0.459382	0.532361	0.623803	123.844848				
56	0.100	0.5	0.5	64	0.606990	0.671634	0.659202
0.442077	0.529236	0.646889	123.844848				
57	0.100	0.5	1.0	16	0.559963	0.722312	0.613174
0.323513	0.423556	0.601577	123.867840				
58	0.100	0.5	1.0	32	0.573756	0.702871	0.567489
0.618135	0.591731	0.602205	123.844848				
59	0.100	0.5	1.0	64	0.553388	0.720240	0.531979
0.883867	0.664195	0.603677	123.844848				
60	0.100	0.5	2.0	16	0.529588	0.741005	0.538563
0.409468	0.465226	0.498323	123.867840				
61	0.100	0.5	2.0	32	0.499857	1.105471	0.499671
0.651459	0.565557	0.509884	123.844848				
62	0.100	0.5	2.0	64	0.424528	0.871729	0.429876
0.464674	0.446598	0.412876	123.844848				
63	0.100	1.1	0.5	16	0.562536	0.702901	0.542409
0.796625	0.645386	0.619012	40.623567				
64	0.100	1.1	0.5	32	0.535520	0.722514	0.572097
0.279748	0.375756	0.551329	40.613598				
65	0.100	1.1	0.5	64	0.555174	0.710981	0.545541
0.657895	0.596473	0.573483	40.613598				
66	0.100	1.1	1.0	16	0.509005	1.531376	0.508062
0.549771	0.528095	0.512111	40.623567				
67	0.100	1.1	1.0	32	0.536164	1.585183	0.548456
0.406322	0.466809	0.553102	40.613598				
68	0.100	1.1	1.0	64	0.503073	1.355342	0.502169
0.645595	0.564921	0.526349	40.613598				
69	0.100	1.1	2.0	16	0.532804	5.611526	0.559137
0.307637	0.396900	0.554546	40.623567				
70	0.100	1.1	2.0	32	0.563751	4.999872	0.564273
0.557494	0.560863	0.591052	40.613598				
71	0.100	1.1	2.0	64	0.526444	4.870373	0.552556
0.275172	0.367386	0.510684	40.613598				

72 0.100 2.0 0.5 16 0.476058 1.435914 0.475782 0.476259 0.476020 0.476243 18.865027
73 0.100 2.0 0.5 32 0.560106 0.961906 0.587902 0.400315 0.476304 0.567716 18.859796
74 0.100 2.0 0.5 64 0.494640 1.290916 0.494851 0.542906 0.517766 0.494744 18.859796
75 0.100 2.0 1.0 16 0.527802 10.037296 0.559988 0.257008 0.352318 0.534352 18.865027
76 0.100 2.0 1.0 32 0.522084 5.655947 0.514211 0.789188 0.622694 0.552437 18.859796
77 0.100 2.0 1.0 64 0.511649 4.934181 0.523198 0.256436 0.344179 0.501989 18.859796
78 0.100 2.0 2.0 16 0.529874 60.478683 0.535751 0.443650 0.485370 0.528935 18.865027
79 0.100 2.0 2.0 32 0.555460 28.591934 0.591167 0.357981 0.445929 0.561395 18.859796
80 0.100 2.0 2.0 64 0.491852 53.096287 0.494568 0.768307 0.601770 0.488193 18.859796



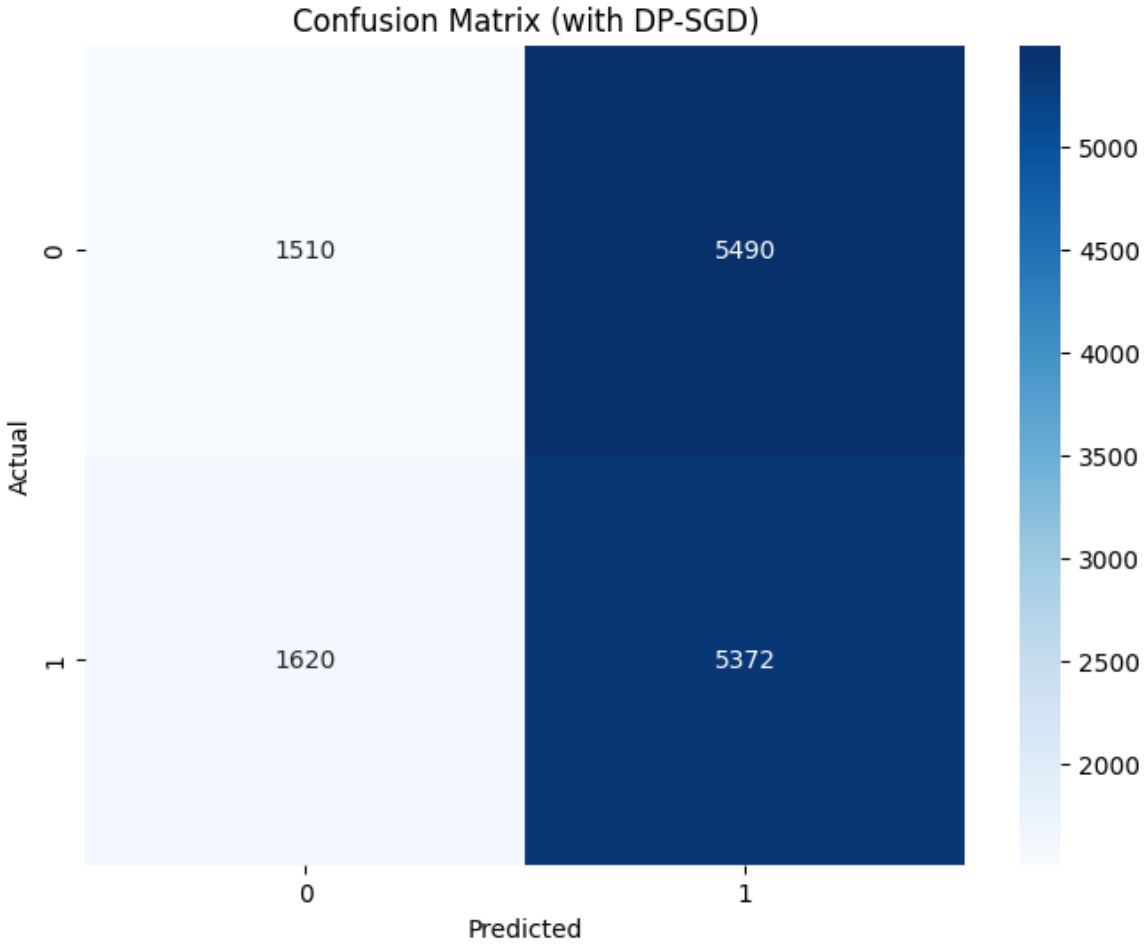


Neos kodikas
438/438 [==========] - 1s 2ms/step
438/438 [==========] - 1s 1ms/step

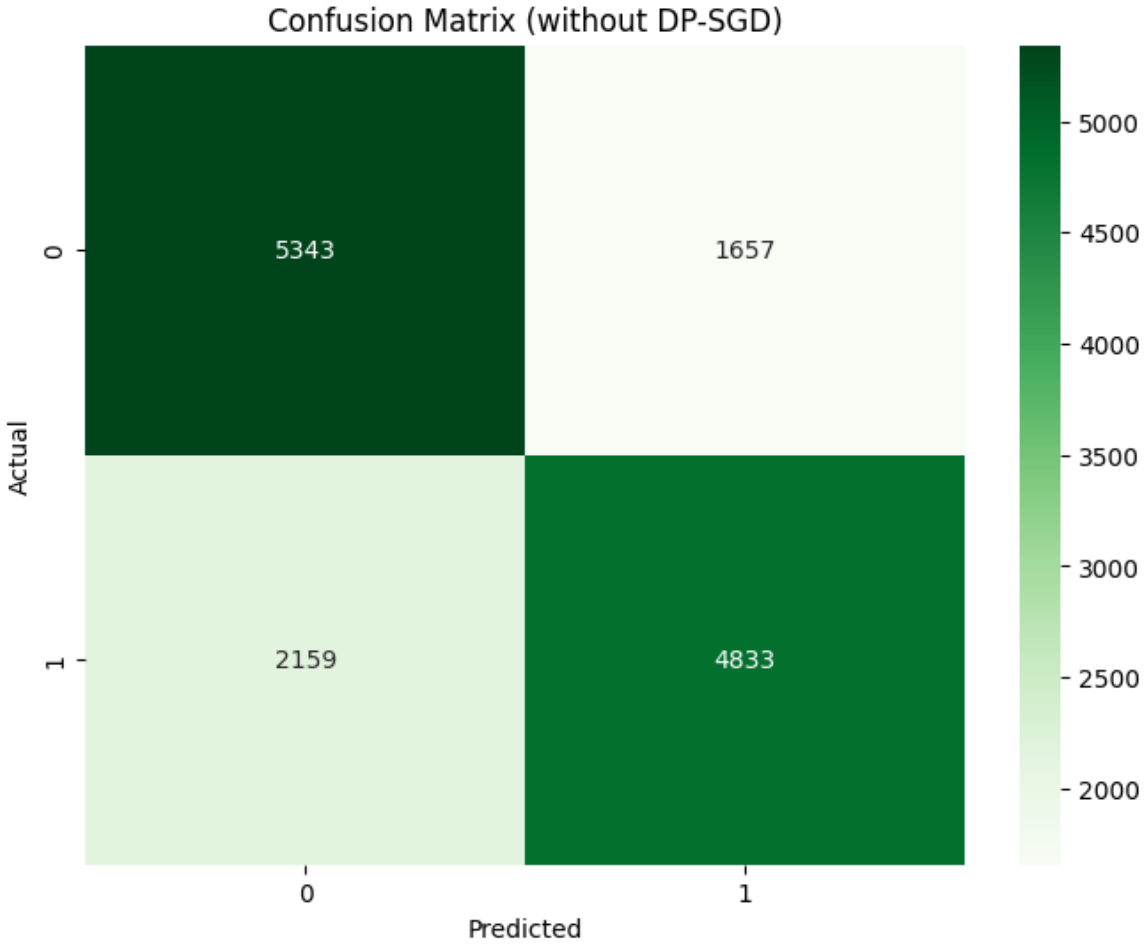
Classification Report (Cardiovascular dataset with DP-SGD):

	precision	recall	f1-score	support
0	0.48	0.22	0.30	7000
1	0.49	0.77	0.60	6992
accuracy			0.49	13992
macro avg	0.49	0.49	0.45	13992
weighted avg	0.49	0.49	0.45	13992

ROC-AUC Score: 0.4882



438/438 [==============================] - 1s 1ms/step



WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
WARNING:absl:`compute_dp_sgd_privacy` is deprecated. It does not account for doubling of sensitivity with microbatching, and assumes Poisson subsampling, which is rarely used in practice. Please use `compute_dp_sgd_privacy_statement`, which provides appropriate context for the guarantee. To compute epsilon under different assumptions than those in `compute_dp_sgd_privacy_statement`, call the `dp_accounting` libraries directly.
False Positives (with DP-SGD): 5490, False Negatives (with DP-SGD): 1620
False Positives (no DP-SGD): 1657, False Negatives (no DP-SGD): 2159

Statistical Analysis of Results:

Mean Accuracy: 0.5139, Std Accuracy: 0.0353
Mean Loss: 2.9024, Std Loss: 9.2849

Delta (δ) used for experiments: 0.00000008
The delta value ensures theoretical guarantees within differential privacy context.

Relationship between privacy budget (ε) and noise multiplier:
Noise Multiplier: 0.5 => Epsilon (ε): 7.3373
Noise Multiplier: 1.1 => Epsilon (ε): 0.8057
Noise Multiplier: 2.0 => Epsilon (ε): 0.2588
[ ]:

Υπεύθυνη Δήλωση Συγγραφέα:

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.