



Σχολή Θετικών Επιστημών & Τεχνολογίας

Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα

Διπλωματική Εργασία

Ανίχνευση Δασικών πυρκαγιών με εικόνες drones με χρήση
συνελικτικών νευρωνικών δικτύων και μετασχηματιστών όρασης

Ιωάννης Ζαρακάς

Επιβλέπων καθηγητής: Ισίδωρος Περίκος

Πάτρα, Ιούνιος 2026

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή («συγγραφέας/δημιουργός») που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης ο συγγραφέας/δημιουργός εκχωρεί στο ΕΑΠ, μη αποκλειστική άδεια χρήσης του δικαιώματος αναπαραγωγής, προσαρμογής, δημόσιου δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσής τους διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος και για όλο το χρόνο διάρκειας των δικαιωμάτων πνευματικής ιδιοκτησίας. Η ανοικτή πρόσβαση στο πλήρες κείμενο για μελέτη και ανάγνωση δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, αποθήκευση, πώληση, εμπορική χρήση, μετάδοση, διανομή, έκδοση, εκτέλεση, «μεταφόρτωση» (downloading), «ανάρτηση» (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού. Ο συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.

Ανίχνευση Δασικών πυρκαγιών με εικόνες drones με χρήση
συνελικτικών νευρωνικών δικτύων και μετασχηματιστών όρασης

Ιωάννης Ζαρακάς

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής:

Ισίδωρος Περίκος

ΣΕΠ, Ελληνικό Ανοικτό Πανεπιστήμιο

Συν-Επιβλέπουσα Καθηγήτρια:

Αικατερίνη Καμπάση

ΣΕΠ, Ελληνικό Ανοικτό Πανεπιστήμιο

Πάτρα, Ιούνιος 2026

«Στη μνήμη του πατέρα μου»

Περίληψη

Η έγκαιρη ανίχνευση δασικών πυρκαγιών αποτελεί έναν από τους κρίσιμότερους παράγοντες για τον περιορισμό των περιβαλλοντικών, οικονομικών και των ευρύτερων κοινωνικών επιπτώσεών τους. Η ραγδαία εξέλιξη της Τεχνητής Νοημοσύνης και ιδιαίτερα των τεχνικών Βαθιάς Μάθησης έχει οδηγήσει στην ανάπτυξη αποδοτικών συστημάτων αυτόματης αναγνώρισης πυρκαγιών από εικόνες drones, αξιοποιώντας τόσο συνελκτικά νευρωνικά δίκτυα (CNNs) όσο και αρχιτεκτονικές βασισμένες στους Vision Transformers (ViTs). Στην πορεία της προόδου αυτής, ανακύπτει το ερώτημα κατά πόσο η διαδικασία fine-tuning μπορεί να βελτιώσει με συνέπεια την απόδοση διαφορετικών αρχιτεκτονικών σε διαφορετικά σύνολα δεδομένων.

Σκοπός της παρούσας εργασίας είναι η συγκριτική αξιολόγηση έξι σύγχρονων μοντέλων βαθιάς μάθησης (SigLIP2, SegFormer-B0, YOLOv26, EdBianchi ViT, Superflash41 και ConvNeXt) τόσο στην προεκπαιδευμένη (pretrained) όσο και στη fine-tuned εκδοχή τους. Η αξιολόγηση πραγματοποιείται σε τρία ευρέως χρησιμοποιούμενα σύνολα δεδομένων ανίχνευσης πυρκαγιάς (FASDD, FLAME και DFire), χρησιμοποιώντας κοινό πειραματικό πλαίσιο και ίδιες υπερπαραμέτρους εκπαίδευσης, ώστε να διασφαλίζεται η αντικειμενική σύγκριση των αποτελεσμάτων.

Η αξιολόγηση βασίζεται σε ποσοτικές μετρικές (Accuracy, Precision, Recall και F1-score), στους Confusion Matrices, καθώς και σε τεχνικές επεξηγήσιμης τεχνητής νοημοσύνης (XAI), όπως η οπτικοποίηση του χώρου χαρακτηριστικών μέσω t-SNE και η ανάλυση των περιοχών που στρέφουν την προσοχή τους τα μοντέλα μέσω της μεθόδου Grad-CAM. Η συνδυαστική αξιοποίηση των παραπάνω τεχνικών επιτρέπει όχι μόνο την αξιολόγηση της ακρίβειας των μοντέλων αλλά και την ερμηνεία της εσωτερικής λειτουργίας και του τρόπου λήψης των αποφάσεών τους.

Τα πειραματικά αποτελέσματα δείχνουν ότι η επίδραση του fine-tuning διαφέρει σημαντικά μεταξύ των εξεταζόμενων αρχιτεκτονικών και των συνόλων δεδομένων. Μοντέλα όπως το ConvNeXt και το EdBianchi ViT παρουσίασαν ιδιαίτερα σημαντική βελτίωση της απόδοσής τους μετά την προσαρμογή, ενώ σε άλλες περιπτώσεις, όπως το

SegFormer-B0 ή το SigLIP2 σε συγκεκριμένα σύνολα δεδομένων, το fine-tuning δεν οδήγησε σε αντίστοιχη βελτίωση ή ακόμη και υποβάθμισε την απόδοση. Τα ευρήματα της εργασίας αναδεικνύουν ότι η αποτελεσματικότητα του fine-tuning εξαρτάται τόσο από τα χαρακτηριστικά της αρχιτεκτονικής όσο και από τη φύση και τη δομή του συνόλου δεδομένων. Επιβεβαιώνεται έτσι η ανάγκη της πολυδιάστατης αξιολόγησης των μοντέλων μέσω συνδυασμού ποσοτικών μετρικών και τεχνικών επεξηγησιμότητας.

Λέξεις – Κλειδιά

Ανίχνευση δασικών πυρκαγιών, Προσαρμογή μοντέλων, Μετασχηματιστές όρασης, Συνελικτικά νευρωνικά δίκτυα, Επεξηγήσιμη τεχνητή νοημοσύνη

Forest Fire Detection from Drone Images Using Convolutional Neural Networks and Vision Transformers

Ioannis Zarakas

Abstract

Early forest fire detection is one of the most critical factors in mitigating the environmental, economic, and broader societal impacts of wildfires. Recent advances in Artificial Intelligence, and particularly in Deep Learning techniques, have led to the development of highly effective automatic fire detection systems based on drone imagery, employing both Convolutional Neural Networks (CNNs) and Vision Transformer (ViT) architectures. Despite this progress, an important research question remains as to whether the fine-tuning process can consistently improve the performance of different deep learning architectures across different datasets.

The aim of this thesis is to perform a comparative evaluation of six state-of-the-art deep learning models (SigLIP2, SegFormer-B0, YOLOv26, EdBianchi ViT, Superflash41, and ConvNeXt) in both their pretrained and fine-tuned versions. The evaluation is conducted on three widely used fire detection datasets (FASDD, FLAME, and DFire) under a common experimental framework and identical training hyperparameters to ensure a fair and objective comparison of the results.

Model performance is assessed using quantitative evaluation metrics, including Accuracy, Precision, Recall, and F1-score, as well as Confusion Matrices. In addition, Explainable Artificial Intelligence (XAI) techniques are employed to investigate the internal behavior of the models through t-SNE feature space visualization and Grad-CAM-based attention map analysis. The combined use of these evaluation approaches provides not only a quantitative assessment of model performance but also valuable insights into the internal feature representations and decision-making mechanisms of the examined architectures.

The experimental results demonstrate that the effectiveness of fine-tuning varies considerably across both model architectures and datasets. Models such as ConvNeXt and

EdBianchi ViT exhibited substantial performance improvements following fine-tuning, whereas in other cases, including SegFormer-B0 and SigLIP2 on specific datasets, fine-tuning failed to produce comparable improvements or even degraded model performance. These findings indicate that the effectiveness of fine-tuning depends not only on the architectural characteristics of the model but also on the nature and structure of the training dataset. Consequently, the study highlights the importance of multifaceted model evaluation through the combined use of quantitative performance metrics and explainability techniques.

Keywords

Forest fire detection, Fine-tuning, Vision Transformers, Convolutional Neural Networks, Explainable Artificial Intelligence

Περιεχόμενα

Περίληψη.....	2
Abstract	4
Περιεχόμενα.....	6
1. Εισαγωγή.....	9
1.1. Επισκόπηση στο πεδίο των πυρκαγιών.....	9
1.2 Αποτίμηση μεθόδων για early fire detection.....	10
1.3 Τεχνικές τεχνητής νοημοσύνης για fire detection και τεχνικές επεξηγησιμότητας.....	11
1.3.1 Συνελκτικά Νευρωνικά Δίκτυα (CNNs) και Εντοπισμός Αντικειμένων.....	11
1.3.2 Υβριδικά Μοντέλα, Μετασχηματιστές Όρασης και Μηχανισμοί Προσοχής.....	14
1.3.3 Πολυτροπική Σύντηξη Δεδομένων (Multimodal Fusion).....	15
1.3.4 Τεχνικές Επεξηγησιμότητας (Explainable AI - XAI).....	15
1.4 Δομή Εργασίας.....	17
2. Βιβλιογραφικά Δεδομένα.....	18
2.1 Ανασκόπηση Βιβλιογραφίας.....	18
2.2 Σύνολα Δεδομένων.....	28
3. Πειραματική Αξιολόγηση.....	43
3.1 Αξιολόγηση του προεκπαιδευμένου SegFormer-B0.....	50
3.1.1 στο DFire.....	51
3.1.2 στο FLAME.....	54
3.1.3 στο FASDD.....	57
3.2 Fine tuning του SegFormer-B0.....	60
3.2.1 στο DFire.....	60
3.2.2 στο FLAME.....	63
3.2.3 στο FASDD.....	66
3.3 Αξιολόγηση του προεκπαιδευμένου Yolov26.....	69
3.3.1 στο DFire.....	69
3.3.2 στο FLAME.....	72
3.3.3 στο FASDD.....	75
3.4 Fine tuning του Yolov26.....	78
3.4.1 στο DFire.....	78
3.4.2 στο FLAME.....	81
3.4.3 στο FASDD.....	85
3.5 Αξιολόγηση του προεκπαιδευμένου SigLIP2.....	88
3.5.1 στο DFire.....	89
3.5.2 στο FLAME.....	91
3.5.3 στο FASDD.....	94
3.6 Fine tuning του SigLIP2.....	97
3.6.1 στο DFire.....	97
3.6.2 στο FLAME.....	101
3.6.3 στο FASDD.....	104
3.7 Αξιολόγηση του προεκπαιδευμένου EdBianchiVit.....	107

3.7.1	στο DFire.....	107
3.7.2	στο FLAME.....	110
3.7.3	στο FASDD.....	114
3.8	Fine tuning του EdBianchiVit.....	117
3.8.1	στο DFire.....	117
3.8.2	στο FLAME.....	120
3.8.3	στο FASDD.....	123
3.9	Αξιολόγηση του προεκπαιδευμένου Superflash41.....	124
3.9.1	στο DFire.....	127
3.9.2	στο FLAME.....	130
3.9.3	στο FASDD.....	133
3.10	Fine tuning του Superflash41.....	136
3.10.1	στο DFire.....	136
3.10.2	στο FLAME.....	140
3.10.3	στο FASDD.....	143
3.11	Αξιολόγηση του προεκπαιδευμένου ConvNeXt.....	147
3.11.1	στο DFire.....	147
3.11.2	στο FLAME.....	151
3.11.3	στο FASDD.....	154
3.12	Fine tuning του ConvNeXT.....	158
3.12.1	στο DFire.....	158
3.12.2	στο FLAME.....	161
3.12.3	στο FASDD.....	165
4.	Συγκριτική Ανασκόπηση.....	168
4.1.	στο DFire.....	168
4.1.1	Το μοντέλο SegFormer-B0.....	168
4.1.2	Το μοντέλο Yolon26.....	170
4.1.3	Το μοντέλο SigLIP2.....	172
4.1.4	Το μοντέλο EdBianchi.....	174
4.1.5	Το μοντέλο Superflash41.....	175
4.1.6	Το μοντέλο ConvNeXt.....	177
4.2.	στο FLAME.....	178
4.2.1	Το μοντέλο SegFormer-B0.....	179
4.2.2	Το μοντέλο Yolon26.....	180
4.2.3	Το μοντέλο SigLIP2.....	182
4.2.4	Το μοντέλο EdBianchi.....	184
4.2.5	Το μοντέλο Superflash41.....	186
4.2.6	Το μοντέλο ConvNeXt.....	187
4.3.	στο FASDD.....	189
4.3.1	Το μοντέλο SegFormer-B0.....	189
4.3.2	Το μοντέλο Yolon26.....	190
4.3.3	Το μοντέλο SigLIP2.....	192
4.3.4	Το μοντέλο EdBianchi.....	193
4.3.5	Το μοντέλο Superflash41.....	195
4.3.6	Το μοντέλο ConvNeXt.....	196
5.	Συμπεράσματα.....	199

5.1 Γενική Αποτίμηση.....	199
5.2 Επίδραση των Συνόλων Δεδομένων.....	201
5.3 Επίδραση των Αρχιτεκτονικών των Μοντέλων.....	202
Βιβλιογραφία	205

Κεφάλαιο 1. Εισαγωγή

1.1 Επισκόπηση στο πεδίο των πυρκαγιών

Τα δάση αποτελούν έναν από τους πολυτιμότερους φυσικούς πόρους του πλανήτη, διαδραματίζοντας κρίσιμο ρόλο στη ρύθμιση του κλίματος, στον κύκλο του άνθρακα και του νερού, καθώς και στη διατήρηση της δυναμικής των οικοσυστημάτων [12]. Ωστόσο, η αυξανόμενη συχνότητα και η σφοδρότητα των δασικών πυρκαγιών σε παγκόσμια κλίμακα, οι οποίες τροφοδοτούνται τόσο από την κλιματική αλλαγή όσο και από ανθρώπινες δραστηριότητες, έχουν αναδείξει το φαινόμενο αυτό σε μία από τις πλέον καταστροφικές φυσικές καταστροφές της εποχής μας [1]. Πρόσφατες έρευνες, όπως αυτή του Πανεπιστημίου του Maryland, καταδεικνύουν ότι οι δασικές πυρκαγιές προκαλούν πλέον την απώλεια 3 εκατομμυρίων περισσότερων εκταρίων δασοκάλυψης ετησίως σε σύγκριση με το 2001, αντιπροσωπεύοντας περισσότερο από το ένα τέταρτο της συνολικής απώλειας δασοκάλυψης των τελευταίων 20 ετών [13]. Παράλληλα, τα στατιστικά στοιχεία είναι αποκαλυπτικά ως προς τα αίτια: το 2021 καταγράφηκαν 8.186 περιστατικά πυρκαγιών στην Πορτογαλία, εκ των οποίων το 42,8% οφειλόταν σε αμέλεια και το 16% σε εμπρησμό από πρόθεση [13]. Τέτοια δεδομένα υπογραμμίζουν την επιτακτική ανάγκη για προηγμένα συστήματα επιτήρησης.

Το οικολογικό και οικονομικό αποτύπωμα των πυρκαγιών είναι ανυπολόγιστο. Το φαινόμενο προκαλεί μη αναστρέψιμη βλάβη στην οικολογική ισορροπία, απειλώντας άμεσα με εξαφάνιση σπάνια είδη χλωρίδας και πανίδας, ενώ ταυτόχρονα απελευθερώνει τεράστιες ποσότητες μονοξειδίου (CO) και διοξειδίου του άνθρακα (CO₂) στην ατμόσφαιρα, επιβαρύνοντας δραματικά την παγκόσμια ποιότητα του αέρα [1]. Σε κοινωνικοοικονομικό επίπεδο, οι συνέπειες μεταφράζονται σε απώλειες ανθρώπινων ζώων, καταστροφή υποδομών και εκτοπισμό πληθυσμών [23]. Χαρακτηριστικά παραδείγματα αποτελούν η πυρκαγιά «Camp Fire» στην Καλιφόρνια το 2018, η οποία κατέστρεψε πάνω από 18.000 κτίρια και ανάγκασε 50.000 ανθρώπους να εγκαταλείψουν τις εστίες τους, καθώς και οι σαρωτικές πυρκαγιές της Αυστραλίας (2019-2020), οι οποίες

οδήγησαν στον θάνατο ή τον εκτοπισμό σχεδόν τριών δισεκατομμυρίων ζώων [23]. Συνεπώς, η έγκαιρη ανίχνευση και διαχείριση αυτών των καταστροφών κρίνεται επιβεβλημένη για την ελαχιστοποίηση των απωλειών [2].

1.2 Αποτίμηση μεθόδων για early fire detection

Στην πυρόσβεση, ο χρόνος είναι ο κρίσιμότερος παράγοντας. Ένας βασικός κανόνας ορίζει ότι για την κατάσβεση μιας πυρκαγιάς απαιτείται ένα ποτήρι νερό στο πρώτο λεπτό, 100 λίτρα στα πρώτα δύο λεπτά και 1.000 λίτρα στα πρώτα δέκα λεπτά [1]. Η εξέλιξη των μεθόδων έγκαιρης ανίχνευσης (early fire detection) περιλαμβάνει τη μετάβαση από παραδοσιακές, χειροκίνητες μεθόδους σε αυτοματοποιημένα τεχνολογικά συστήματα, καθένα με τα δικά του πλεονεκτήματα και περιορισμούς.

Οι παραδοσιακές μέθοδοι βασίζονται κυρίως στην ανθρώπινη παρατήρηση (πύργοι ελέγχου, πεζές ή εποχούμενες περιπολίες) και σε οπτικά συστήματα ανίχνευσης καπνού. Αν και αποτελούν τη βασική διάσταση της πρόληψης, οι λύσεις αυτές είναι εξαιρετικά δαπανηρές, απαιτούν 24ωρη ανθρώπινη παρουσία, καλύπτουν περιορισμένη γεωγραφική εμβέλεια και είναι επιρρεπείς στο ανθρώπινο λάθος λόγω κόπωσης ή κακής ορατότητας [13]. Επιπλέον, η χρήση σταθερών καμερών (CCTV) περιορίζεται από το στατικό τους οπτικό πεδίο, αδυνατώντας να καλύψουν μεγάλα πεδία όπως οι δασικές εκτάσεις [23].

Στον τομέα των υπολογιστικών συστημάτων όρασης (computer vision), οι παλαιότερες τεχνικές βασίζονταν σε χρωματικούς αλγόριθμους (colour-based approaches). Οι μέθοδοι αυτές προσπαθούσαν να απομονώσουν τα εικονοστοιχεία (pixels) της φωτιάς αναλύοντας τον κορεσμό του κόκκινου χρώματος στο φάσμα RGB ή YCbCr [9]. Ωστόσο, η προσέγγιση αυτή είναι εξαιρετικά ευαίσθητη στις μεταβαλλόμενες συνθήκες φωτισμού (π.χ. αντανάκλαση του ηλίου, δύση, κόκκινα φύλλα) και οδηγεί σε υψηλό αριθμό ψευδώς θετικών συναγερμών (false positives) εξαιτίας αυτού του εγγενούς γνωστικού περιορισμού [9].

Η χρήση Δορυφορικής Τηλεπισκόπησης (Satellite Remote Sensing) προσέφερε δυνατότητες κάλυψης σε παγκόσμια κλίμακα, αξιοποιώντας αισθητήρες (όπως οι MODIS και VIIRS) για τον εντοπισμό θερμικών διαφορών [20,2]. Παρ' όλα αυτά, η δορυφορική παρακολούθηση χαρακτηρίζεται από ένα συμβιβασμό (trade-off) μεταξύ χωρικής και

χρονικής ανάλυσης. Οι δορυφόροι Γεωστατικής Τροχιάς (GEO) παρέχουν συνεχή χρονική παρακολούθηση, αλλά με πολύ χαμηλή χωρική ανάλυση (π.χ. 2km/pixel), καθιστώντας αδύνατο τον εντοπισμό μιας μικρής εστίας στα αρχικά της στάδια. Αντίθετα, οι δορυφόροι Χαμηλής Τροχιάς (LEO) προσφέρουν υψηλότερη χωρική ανάλυση, αλλά η χρονική τους υστέρηση (revisit times) μπορεί να διαρκέσει ημέρες [2]. Επιπρόσθετα, η νέφωση και οι παρεμβολές από τον δασικό θόλο (canopy) συχνά αποκρύπτουν την ορατότητα [20].

Στο πλαίσιο αυτό, τα Μη Επανδρωμένα Αεροσκάφη (UAVs - Drones) έχουν αναδειχθεί ως η απόλυτη τεχνολογία αιχμής [1]. Τα UAVs γεφυρώνουν το χάσμα μεταξύ δορυφόρων και επίγειων σταθμών, προσφέροντας δυνατότητα πτήσης σε χαμηλό υψόμετρο (κάτω από τα νέφη), υψηλή ευελιξία και άμεση ανάπτυξη στο πεδίο [4]. Επιτρέπουν την αδιάλειπτη συλλογή δεδομένων εξαιρετικά υψηλής ανάλυσης, φτάνοντας σε δύσβατα και επικίνδυνα εδάφη, βελτιώνοντας αισθητά την επίγνωση της κατάστασης. Εξοπλισμένα με πολυτροπικούς αισθητήρες, συμπεριλαμβανομένων οπτικών καμερών (RGB) και θερμικών/υπέρυθρων (IR) αισθητήρων, επιτρέπουν την ανίχνευση του θερμικού ίχνους της φωτιάς μέσα από πυκνό καπνό, παρέχοντας κρίσιμα δεδομένα σε πραγματικό χρόνο για την ταχεία επέμβαση των δυνάμεων πυρόσβεσης. Φυσικά, τα UAVs αντιμετωπίζουν προκλήσεις, όπως η περιορισμένη αυτονομία μπαταρίας, η μικρή χωρητικότητα ωφέλιμου φορτίου (payload) και η ανάγκη για επεξεργαστική ισχύ εντός του αεροσκάφους (on-board edge computing) [2].

1.3 Τεχνικές τεχνητής νοημοσύνης για fire detection και τεχνικές επεξηγησιμότητας

Για την υπέρβαση των περιορισμών της ανθρώπινης παρατήρησης και των κλασικών αλγορίθμων, η ενσωμάτωση αλγορίθμων Τεχνητής Νοημοσύνης (AI) και ειδικότερα Βαθιάς Μάθησης (Deep Learning - DL) στα συστήματα UAV έχει ανοίξει νέους ορίζοντες. Τα μοντέλα DL μπορούν να εξάγουν αυτόματα αφηρημένα χωρικά και χρονικά χαρακτηριστικά από τεράστιους όγκους δεδομένων, επιτυγχάνοντας πολύ μεγαλύτερη ακρίβεια και ανθεκτικότητα σε περίπλοκα δασικά περιβάλλοντα[22,2].

1.3.1 Συνελικτικά Νευρωνικά Δίκτυα (CNNs) και Εντοπισμός Αντικειμένων

Η ραχοκοκαλιά της σύγχρονης όρασης υπολογιστών (computer vision) για την ανίχνευση δασικών πυρκαγιών στηρίζεται στα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs). Η αρχιτεκτονική ενός βαθιού CNN για την ταξινόμηση πυρκαγιών αποτελείται δομικά από δύο κύρια τμήματα: το μπλοκ εκμάθησης χαρακτηριστικών (feature learning block) και το μπλοκ ταξινόμησης (feature classification block) [2].

Στο πρώτο τμήμα, τα Επίπεδα Συνέλιξης (Convolutional Layers) αναλαμβάνουν την εξαγωγή χωρικών χαρακτηριστικών εφαρμόζοντας φίλτρα (kernels) με εκπαιδευσιμες παραμέτρους (βάρη) πάνω στα δεδομένα της εικόνας, εντοπίζοντας μοτίβα όπως ακμές, σχήματα και υφές της φλόγας ή του καπνού [2]. Προκειμένου το δίκτυο να μπορεί να προσεγγίσει πολύπλοκες μη-γραμμικές σχέσεις, εφαρμόζονται συναρτήσεις ενεργοποίησης (όπως οι ReLU, Swish και EELU). Στη συνέχεια, τα Επίπεδα Συγκέντρωσης (Pooling Layers), όπως η μέγιστη ή η μέση συγκέντρωση (max/average pooling), αναλαμβάνουν την υποδειγματοληψία των χωρικών διαστάσεων των χαρτών χαρακτηριστικών. Αυτό μειώνει δραστικά τον αριθμό των παραμέτρων του δικτύου, διατηρώντας ταυτόχρονα την πιο κρίσιμη πληροφορία. Στο δεύτερο τμήμα, τα δεδομένα επιπεδοποιούνται (Flattening layer) και τροφοδοτούνται σε πλήρως συνδεδεμένα επίπεδα (Fully Connected layers) για την παραγωγή της τελικής πρόβλεψης [2].

Στο πεδίο της ανίχνευσης πυρκαγιών μέσω UAV, τα CNNs αξιοποιούνται πρακτικά μέσα από τρεις διακριτές προσεγγίσεις, καθεμία με εξειδικευμένες αρχιτεκτονικές.

A) Ταξινόμηση Εικόνας (Image Classification): Στόχος είναι ο καθορισμός της ύπαρξης ή μη φωτιάς στο σύνολο μιας εικόνας. Ιδιαίτερη έμφαση δίνεται σε ελαφριά (lightweight) δίκτυα που μπορούν να λειτουργήσουν στους περιορισμένους υπολογιστικούς πόρους ενός UAV (edge computing). Ένα χαρακτηριστικό παράδειγμα είναι το MobileNetV3, το οποίο αντικαθιστά τις κλασικές συνελίξεις με χωρικά διαχωρίσιμες συνελίξεις βάθους (depth-wise separable convolutions - DWCs), μειώνοντας κατακόρυφα την υπολογιστική πολυπλοκότητα χωρίς να θυσιάζεται η ακρίβεια. Τα μοντέλα αυτά ενσωματώνονται συχνά μέσω τεχνικών Μάθησης Μεταφοράς (Transfer Learning), όπου τα αρχικά επίπεδα του προεκπαιδευμένου δικτύου "παγώνουν" για να διατηρήσουν τη βασική ικανότητα

αναγνώρισης οπτικών χαρακτηριστικών, ενώ τα ανώτερα επίπεδα υφίστανται μικρορύθμιση (fine-tuning) σε δεδομένα δασικών πυρκαγιών [15].

Β) Εντοπισμός Αντικειμένων (Object Detection): Αυτή η προσέγγιση σχεδιάζει πλαίσια οριοθέτησης (bounding boxes) γύρω από τις φλόγες. Κυρίαρχη θέση κατέχει η οικογένεια αλγορίθμων YOLO (You Only Look Once), λόγω της απaráμιλλης ικανότητάς της για επεξεργασία σε πραγματικό χρόνο. Η εξέλιξη από τα YOLOv3/v4 μέχρι τα YOLOv5 και YOLOv8 έχει βελτιώσει σε πολύ μεγάλο βαθμό την ισορροπία ταχύτητας και ακρίβειας. Πρόσφατα, το υπερσύγχρονο YOLOv9 προσέφερε λύσεις σε προβλήματα απώλειας κρίσιμων πληροφοριών ενσωματώνοντας "Προγραμματιζόμενη Πληροφορία Διαβάθμισης" (PGI) και τη δομή GELAN, γεγονός που το καθιστά ικανό να εντοπίζει εξαιρετικά μικρές εστίες φωτιάς ακόμα και από μεγάλο ύψος. Παράλληλα, χρησιμοποιούνται αρχιτεκτονικές δύο σταδίων (two-stage detectors) όπως το Faster R-CNN, καθώς και δίκτυα SSD [12,17].

Γ) Σημασιολογική Τμηματοποίηση (Semantic Segmentation): Παρέχει τον ακριβή διαχωρισμό (pixel-level classification) της περιμέτρου της φωτιάς από το υπόβαθρο, ένα σημείο που έχει ιδιαίτερη σημασία για την εκτίμηση της καμένης έκτασης. Αρχιτεκτονικές όπως το FCN (Fully Convolutional Network) και το U-Net κυριαρχούν. Το U-Net ξεχωρίζει για τη συμμετρική, επεκτεινόμενη δομή του (encoder-decoder) που επιτρέπει την ακριβή τοπικοποίηση. Αντίστοιχα, το DeepLabV3+ αποτελεί μια κορυφαία επιλογή, καθώς ενσωματώνει τη μέθοδο Atrous Spatial Pyramid Pooling (ASPP) για την εξαγωγή χαρακτηριστικών σε πολλαπλές κλίμακες ανάλυσης, επιτρέποντας την ευέλικτη κατανόηση στόχων ανεξαρτήτως μεγέθους [2].

Δ) Χωροχρονικά Υβριδικά Μοντέλα (Spatiotemporal / RNNs): Καθώς η φωτιά είναι ένα δυναμικό φαινόμενο, η ανάλυση μεμονωμένων στατικών καρέ δεν αρκεί. Για να αξιοποιηθεί η χρονική αλληλουχία του βίντεο από ένα UAV, τα CNNs συνενώνονται με Αναδρομικά Νευρωνικά Δίκτυα (RNNs), και συγκεκριμένα με αρχιτεκτονικές LSTM (Long Short-Term Memory) ή GRU (Gated Recurrent Units). Σε μοντέλα όπως τα CNN-RCNN ή τα ConvLSTM, το τμήμα του CNN εξάγει τα χωρικά χαρακτηριστικά (πού βρίσκεται η φωτιά), ενώ το LSTM τμήμα μοντελοποιεί την χρονική τους εξέλιξη ,

βελτιώνοντας την ακρίβεια της ταξινόμησης σε πραγματικές δυναμικές συνθήκες βίντεο [17].

1.3.2 Υβριδικά Μοντέλα, Μετασχηματιστές Όρασης (Vision Transformers) και Μηχανισμοί Προσοχής

Παρά την κυριαρχία τους, τα κλασικά CNNs παρουσιάζουν έναν εγγενή περιορισμό: εστιάζουν υπερβολικά στα τοπικά χαρακτηριστικά και δυσκολεύονται να συλλάβουν το συνολικό, γενικό πλαίσιο (global context) της εικόνας, κάτι που είναι ζωτικής σημασίας σε λήψεις UAV από ψηλά, όπου οι εστίες φωτιάς αλληλεπιδρούν με τη δασική δομή και τα μετεωρολογικά φαινόμενα [23]. Τη λύση σε αυτό το πρόβλημα προσφέρουν οι Μετασχηματιστές Οράσεως (Vision Transformers - ViTs). Τα δίκτυα αυτά, βασισμένα στον μηχανισμό της αυτο-προσοχής (self-attention), αντιμετωπίζουν την εικόνα ως μια ακολουθία από τμήματα (patches), επιτρέποντας στο μοντέλο να αντιλαμβάνεται σχέσεις μεγάλης εμβέλειας (long-range dependencies) μεταξύ απομακρυσμένων περιοχών της εικόνας [23].

Επειδή, ωστόσο, οι παραδοσιακοί Transformers απαιτούν τεράστια υπολογιστική ισχύ και μνήμη (κάτι απαγορευτικό για τα μικροϋπολογιστικά συστήματα των UAVs, όπως το Jetson Nano) [17], η σύγχρονη έρευνα στρέφεται στα ελαφριά υβριδικά μοντέλα που συνδυάζουν ViTs και CNNs (π.χ. αρχιτεκτονικές τύπου MobileViT ή FireViTNet). Τα μοντέλα αυτά αντλούν τα χωρικά χαρακτηριστικά μέσω των CNNs και τη συνολική δομή μέσω των Transformers, μειώνοντας δραστικά τον αριθμό των παραμέτρων χωρίς έκπτωση στην ακρίβεια [9].

Επιπρόσθετα, για τη βελτίωση της ακρίβειας εφαρμόζονται ενσωματωμένοι Μηχανισμοί Προσοχής (Attention Mechanisms), όπως το CBAM (Convolutional Block Attention Module) ή το MSAM (Modified Soft Attention Mechanism) [2,4]. Οι μηχανισμοί αυτοί λειτουργούν ως επιλεκτικά φίλτρα, επιτρέποντας στο μοντέλο να εστιάζει δυναμικά στα κανάλια και τις χωρικές περιοχές της εικόνας που περιέχουν φωτιά, καταστέλλοντας ταυτόχρονα το θόρυβο από το δασικό υπόβαθρο, οδηγώντας σε ανώτερα ποσοστά mAP (Mean Average Precision).

1.3.3 Πολυτροπική Σύντηξη Δεδομένων (Multimodal Fusion)

Δεδομένου ότι το δασικό περιβάλλον υπόκειται σε οπτικές αλλοιώσεις (π.χ. πυκνός καπνός, κακός φωτισμός το βράδυ), τα σύγχρονα συστήματα δεν αρκούνται στην ορατή φασματική ζώνη (RGB). Αντίθετα, εξοπλίζουν τα UAVs με διπλούς αισθητήρες (ορατού και υπέρυθρου/θερμικού φάσματος). Η εικόνα RGB είναι ιδανική για τον εντοπισμό του καπνού και την αποσαφήνιση του περιβάλλοντος χώρου, ενώ η υπέρυθρη (IR) εικόνα διαπερνά τον καπνό και εντοπίζει αμέσως τις θερμικές εστίες. Τεχνικές βαθιάς μάθησης επιτελούν αυτόματη χωρική ευθυγράμμιση (registration) και σύντηξη (fusion) (είτε σε πρώιμο στάδιο "early fusion" είτε σε όψιμο "late fusion") των δύο πηγών, αποτρέποντας τις ψευδείς ανιχνεύσεις (όπως η παρερμηνεία υψηλών θερμοκρασιών το καλοκαίρι από τα IR ή η σύγχυση σύννεφων-καπνού στα RGB) [4].

1.3.4 Τεχνικές Επεξηγησιμότητας (Explainable AI - XAI)

Παρά τις εντυπωσιακές τους επιδόσεις, τα μοντέλα Βαθιάς Μάθησης χαρακτηρίζονται από τεράστια δομική πολυπλοκότητα (συχνά περιέχουν εκατομμύρια παραμέτρους), γεγονός που τα καθιστά «μαύρα κουτιά» (black-boxes). Αυτή η εγγενής έλλειψη διαφάνειας στον τρόπο με τον οποίο ένα νευρωνικό δίκτυο καταλήγει σε μια απόφαση δημιουργεί σημαντικά εμπόδια στην υιοθέτησή τους σε κρίσιμες εφαρμογές, όπως η διαχείριση φυσικών καταστροφών και η ασφάλεια [20].

Η Επεξηγήσιμη Τεχνητή Νοημοσύνη (Explainable AI - XAI) έχει αναδυθεί ως μια κρίσιμη απάντηση σε αυτό το πρόβλημα, αποσκοπώντας στην επεξήγηση της διαδικασίας λήψης αποφάσεων. Ο στόχος της XAI δεν είναι απλώς η τεχνική ερμηνεία, αλλά η παραγωγή «φυσικού νοήματος» (physical meaning) που μπορεί να γίνει κατανοητό από τους ανθρώπους, η οικοδόμηση εμπιστοσύνης μέσω της διαφάνειας, η λογοδοσία (accountability) και η απόδειξη ότι το μοντέλο συμμορφώνεται με επιχειρησιακούς κανόνες ασφαλείας [20,15].

Οι κύριες τεχνικές XAI που εφαρμόζονται στα σύγχρονα συστήματα ανίχνευσης πυρκαγιών κατηγοριοποιούνται ως εξής:

Μέθοδοι Χαρτογράφησης Ενεργοποίησης (Grad-CAM): Η τεχνική Gradient-weighted Class Activation Mapping (Grad-CAM) είναι ίσως η πιο διαδεδομένη μέθοδος για την

οπτική εξήγηση των αποφάσεων ενός CNN. Το Grad-CAM αναλύει τις διαβαθμίσεις (gradients) που ρέουν προς το τελικό συνελκτικό επίπεδο του δικτύου κατά τη διάρκεια της πρόβλεψης. Βάσει αυτών, παράγει έναν θερμικό χάρτη (heatmap) ο οποίος, όταν υπερτεθεί στην αρχική εικόνα, υποδεικνύει οπτικά ποιες ακριβώς περιοχές εικονοστοιχείων (pixels) ήταν οι πλέον καθοριστικές για την ταξινόμηση της εικόνας ως «φωτιά» [22,15]. Αυτό επιτρέπει στους ειδικούς να επιβεβαιώσουν εάν το μοντέλο εντόπισε πράγματι τις φλόγες ή αν ξεγελάστηκε από άσχετα στοιχεία (π.χ. τον ήλιο που δύει ή ένα κόκκινο όχημα).

Ανεξάρτητες από το Μοντέλο Τοπικές Εξηγήσεις (LIME): Η μέθοδος LIME (Local Interpretable Model-agnostic Explanations) προσεγγίζει την τοπική συμπεριφορά του πολύπλοκου "μαύρου κουτιού" κατασκευάζοντας ένα απλούστερο, γραμμικό και ερμηνεύσιμο μοντέλο γύρω από μια συγκεκριμένη (τοπική) πρόβλεψη. Στην ανίχνευση πυρκαγιών, το LIME χωρίζει την εικόνα σε μικρότερα τμήματα (super-pixels) και αξιολογεί πώς η "απόκρυψη" διαφόρων τμημάτων επηρεάζει την τελική πρόβλεψη. Έτσι, εξηγεί ακριβώς ποια γεωμετρικά τμήματα της εικόνας ώθησαν το μοντέλο να ανιχνεύσει καπνό ή φλόγα [15].

Προσέγγιση SHAP (Shapley Additive Explanations): Η μέθοδος SHAP βασίζεται στη θεωρία παιγνίων και υπολογίζει την επιρροή κάθε μεμονωμένου χαρακτηριστικού (feature) στην τελική απόφαση. Μετρά τη «μέση οριακή συνεισφορά» (average marginal contribution) των χαρακτηριστικών εξετάζοντας όλους τους πιθανούς συνδυασμούς τους [20]. Στα πολυτροπικά δεδομένα (π.χ. συνδυασμός οπτικών, υπέρυθρων καμερών και μετεωρολογικών δεδομένων), εργαλεία όπως τα γραφήματα SHAP αναδεικνύουν με μαθηματική ακρίβεια ποιος αισθητήρας ή ποια χωρική δομή συνέβαλε περισσότερο στον εντοπισμό της εστίας.

Η εφαρμογή αυτών των εργαλείων μετατρέπει τα μοντέλα Βαθιάς Μάθησης από απλούς αυτοματισμούς σε «έξυπνους βοηθούς» για τους πυροσβέστες. Οι αναλυτές μπορούν να χρησιμοποιήσουν την XAI (όπως φαίνεται στην απόδοση μοντέλων όπως το FireDetXplainer) για να αντιληφθούν τα συστηματικά σφάλματα του αλγορίθμου, εξασφαλίζοντας ότι η εκμάθηση (feature extraction) εστιάζει στην πραγματική φυσική δομή της φωτιάς και όχι σε εννοιολογικές προκαταλήψεις (bias) των δεδομένων εκπαίδευσης [22,20].

1.4 Δομή Εργασίας

Το σχήμα της δομής της παρούσας εργασίας είναι το ακόλουθο. Στο Κεφάλαιο 2 ανασκοπούμε σύγχρονα ερευνητικά δεδομένα που αναφέρονται στο πεδίο της ανίχνευσης των δασικών πυρκαγιών από εικόνες drones, αλλά και με διάφορους άλλους τρόπους λήψης, με χρήση τεχνικών βαθιάς μάθησης. Συγκεκριμένα στην παράγραφο 2.1 κάνουμε μια περιοδολόγηση εργασιών του πεδίου που αναφέρονται σε σύγχρονες τεχνικές ανίχνευσης πυρκαγιάς και έγκαιρης προειδοποίησης. Σταχυολογούμε τα βασικά ερευνητικά πορίσματα αυτών των εργασιών και αναφερόμαστε στις ερευνητικές μελλοντικές στοχεύσεις που θέτουν. Στην παράγραφο 2.2 κάνουμε μια σύντομη ανάλυση των συνόλων δεδομένων που συναντάμε στις εργασίες που μελετήσαμε. Ακολούθως στο Κεφάλαιο 3 προβαίνουμε στον πειραματικό έλεγχο έξι μοντέλων βαθιάς μάθησης στο έργο της αναγνώρισης-ανίχνευσης φωτιάς σε εικόνα. Για το έργο αυτό κάνουμε χρήση τριών συνόλων δεδομένων. Η αξιολόγηση γίνεται σε πρώτη φάση χρησιμοποιώντας την pretrained εκδοχή των μοντέλων. Έπειτα προχωρούμε σε finetuning για κάθε μοντέλο σε κάθε ένα από τα τρία σύνολα δεδομένων. Στο Κεφάλαιο 4 προβαίνουμε σε μια συγκριτική μελέτη της απόδοσης κάθε μοντέλου στην pretrained και στην finetuned εκδοχή του στα πλαίσια καθενός από τα σύνολα δεδομένων και εκτιμούμε την επίδραση της διαδικασίας finetuning στην όλη συμπεριφορά του κάθε μοντέλου. Η εργασία μας ολοκληρώνεται με το Κεφάλαιο 5 στο οποίο συγκεντρώνονται τα συμπεράσματα από την όλη πειραματική μελέτη.

Κεφάλαιο 2. Βιβλιογραφικά Δεδομένα

2.1 Ανασκόπηση βιβλιογραφίας

Στο [20] ανασκοπείται η εξέλιξη της ανίχνευσης δασικών πυρκαγιών μέσω Τηλεπισκόπησης, αναδεικνύοντας τον ρόλο της Μηχανικής (ML) και Βαθιάς Μάθησης (DL). Τα μέσα συλλογής δεδομένων εξελίχθηκαν από τους δορυφόρους MODIS/VIIRS, σε συστήματα Lidar και SAR (ικανά να διαπερνούν νέφη/καπνό), καθώς και σε γεωστατικούς δορυφόρους (GOES-R, Himawari-8) για παρακολούθηση σε σχεδόν πραγματικό χρόνο. Συμπληρωματικά, τα Μη Επανδρωμένα Αεροσκάφη (UAVs) παρέχουν υψηλής ευκρίνειας πολυφασματικά δεδομένα (π.χ. FLAME, MIVIA), τα οποία είναι κρίσιμα για την εκπαίδευση των μοντέλων. Οι μέθοδοι ML διακρίνονται σε επιβλεπόμενη, μη επιβλεπόμενη (μέσω Αυτοκωδικοποιητών) και ημιεπιβλεπόμενη μάθηση. Κυρίαρχες DL αρχιτεκτονικές αποτελούν τα Συνελκτικά Νευρωνικά Δίκτυα (CNNs) για εξαγωγή χωρικών χαρακτηριστικών, τα U-Nets, τα ConvLSTM για ταυτόχρονη χωροχρονική μοντελοποίηση της εξάπλωσης, και οι Μετασχηματιστές Οράσεως (ViTs) που αξιοποιούν μηχανισμούς προσοχής. Ανάμεσα στις σύγχρονες προκλήσεις οι συγγραφείς περιλαμβάνουν την ανεπαρκή χωροχρονική ανάλυση των δορυφορικών δεδομένων, την ανάγκη μοντελοποίησης σε πραγματικό χρόνο, την αβεβαιότητα μετρήσεων, την ανισορροπία κλάσεων (data imbalance) που προκαλεί υπερπροσαρμογή (overfitting), και την πολυπλοκότητα των περιβαλλοντικών επιπτώσεων. Ως προς τις μελλοντικές ερευνητικές κατευθύνσεις οι συγγραφείς εστιάζουν το ενδιαφέρον τους στην πολυτροπική μάθηση, τη χρήση GANs για αλγοριθμική εξισορρόπηση δεδομένων, και την ενσωμάτωση προχωρημένης χωρικής και περιβαλλοντικής ανάλυσης. Κομβικής σημασίας θεωρούνται η Αυτο-Επιβλεπόμενη Μάθηση, η Επεξηγήσιμη Τεχνητή Νοημοσύνη (XAI – π.χ. SHAP, CAMs) η οποία αυξάνει τη διαφάνεια στη λήψη των αποφάσεων, η σύζευξη αλγορίθμων ML με θερμοδυναμικά/φυσικά μοντέλα και η κατασκευή Ψηφιακών Διδύμων (Digital Twins) σε πλατφόρμες Μεγάλων Δεδομένων.

Στην εργασία [22] παρουσιάζεται μια αρχιτεκτονική βαθιάς μάθησης για την έγκαιρη και ακριβή ανίχνευση πυρκαγιών μέσω συστημάτων όπως τα μη επανδρωμένα αεροσκάφη. Παρόλο που τα μοντέλα βαθιάς μάθησης υπερτερούν έναντι των παραδοσιακών μεθόδων μηχανικής μάθησης, αντιμετωπίζουν σημαντικές προκλήσεις, όπως ο εντοπισμός μικρών εστιών φωτιάς σε εναέριες λήψεις με περίπλοκο υπόβαθρο, η υψηλή υπολογιστική πολυπλοκότητα, καθώς και ο αυξημένος ρυθμός ψευδών συναγερμών (false positive rates). Όσον αφορά την προτεινόμενη μεθοδολογία και την αρχιτεκτονική του μοντέλου για την επίλυση των παραπάνω ζητημάτων, οι συγγραφείς προτείνουν ένα καινοτόμο μοντέλο Συνελικτικού Νευρωνικού Δικτύου (CNN), το οποίο χρησιμοποιεί ως βάση (backbone) μια βελτιστοποιημένη έκδοση της αρχιτεκτονικής MobileNetV3 για την εξαγωγή χαρακτηριστικών. Για να ενισχυθεί η ικανότητα του μοντέλου να εστιάζει σε μικρές περιοχές πυρκαγιάς, ενσωματώθηκε ένας Τροποποιημένος Μηχανισμός Ήπιας Προσοχής (Modified Soft Attention Mechanism - MSAM) σε συνδυασμό με λειτουργίες τρισδιάστατης συνέλιξης (3D convolution). Όσον αφορά στα αποτελέσματα που καταγράφηκαν το μοντέλο αξιολογήθηκε χρησιμοποιώντας τρία σύνολα δεδομένων το FD dataset, το ADSF dataset (δουροφορικές και drone λήψεις) και το DFAN dataset. Τα πειραματικά αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος ξεπερνά τα υφιστάμενα μοντέλα αιχμής (state-of-the-art), επιτυγχάνοντας μέση ακρίβεια (Accuracy) 98,66% στο σύνολο FD, 93,50% στο σύνολο ADSF και 91,20% στο σύνολο DFAN. Παράλληλα, διατήρησε εξαιρετικά υψηλές τιμές στους δείκτες Precision, Recall και F1-score. Για την επαλήθευση του μοντέλου, χρησιμοποιήθηκε η τεχνική Επεξηγήσιμης Τεχνητής Νοημοσύνης (Explainable AI) Grad-CAM.

Στην εργασία [19] οι κύριες κατευθύνσεις του άρθρου μπορούν να συνοψιστούν στα ακόλουθα σημεία (α) παρουσιάζεται μια πρωτότυπη συνδυαστική μέθοδος τμηματοποίησης που ενσωματώνει τις δυνατότητες των Οπτικών Μετατροπέων (Vision Transformers) καθώς και αυτές των συνελικτικών νευρωνικών δικτύων. (β) προκειμένου να βελτιώσουν την απόδοση της τμηματοποίησης οι ερευνητές ενσωματώνουν και συναθροίζουν στο μοντέλο τους διάφορους μηχανισμούς όπως το CBAM μηχανισμό προσοχής, το ASPP και την ομαδοποίηση SP.

Όσον αφορά στο σύνολο των δεδομένων που χρησιμοποιήθηκε σε αυτή την έρευνα, αυτό συνίσταται σε 1748 εικόνες που τραβήχτηκαν από drones και απεικονίζουν πυρκαγιές και

τις οποίες εικόνες συνέλεξαν οι συγγραφείς της μελέτης από το Internet. Προκειμένου το σύνολο των δεδομένων να ανταποκρίνεται όσο γίνεται καλύτερα και πιστότερα στις διαφορετικές συνθήκες που επικρατούν σε πραγματικά πεδία πυρκαγιάς, οι συγγραφείς φιλτράρισαν και επισημείωσαν τα δεδομένα τους. Σε σχέση με την αρχιτεκτονική του μοντέλου και τις μεθόδους που χρησιμοποίησαν οι ερευνητές στη μελέτη τους, βασικό δομικό στοιχείο του μοντέλου αποτελεί το MobileViT (MobileVisualTransformer). Το MobileViT είναι μια εκδοχή Οπτικού Μετατροπέα που ενσωματώνει τεχνικές όπως depth-wise convolution οι οποίες μειώνουν τον αριθμό των απαιτούμενων παραμέτρων και άρα την υπολογιστική πολυπλοκότητα, ενώ συγχρόνως επιδεικνύει μεγάλη ικανότητα στο να συλλαμβάνει το γενικό περίγραμμα των προς ανάλυση εικόνων. Το πρωτότυπο μοντέλο που αναπτύχθηκε από τους ερευνητές και στο οποίο δίνουν το όνομα FireViTNet τίθεται σε σύγκριση με άλλα μοντέλα που βρίσκονται στην αιχμή της τεχνολογίας στο πεδίο της εννοιολογικής τμηματοποίησης (semantic segmentation). Αυτά τα μοντέλα με τα οποία συγκρίνεται το FireViTNet είναι τα FCN, U-Net, Deeplabv3+, PSPNet και το MaskRCNN. Από τη σύγκριση αυτή προκύπτει ότι τα αποτελέσματα που επιτυγχάνει το FireViTNet στις διάφορες μετρικές (81,44% IoU, 85,31% Precision, 92,68% Recall) είναι καλύτερα από όλα τα προαναφερθέντα μοντέλα. Στη δοκιμή του μοντέλου υπό συνθήκες αυξημένης φωτεινότητας καταγράφηκε μια πτώση στους δείκτες. Ενδεικτικά ο δείκτης F1 έπεσε από 88,82% στο αρχικό σύνολο δεδομένων στο 87,20 % στο σύνολο εικόνων που είχαν υποστεί αλλαγή της φωτεινότητάς τους. Επίσης η απόκρυψη κάποιου τμήματος της εικόνας οδήγησε σε περαιτέρω μείωση της μετρικής F1 στο 86,60%. Οι μειώσεις αυτές καταδεικνύουν πως το μοντέλο πρέπει να βελτιωθεί περαιτέρω, όσον αφορά στη διαδικασία της τμηματοποίησης κατά τη φάση της εκπαίδευσης και της δοκιμής.

Στην εργασία [9] οι συγγραφείς προτείνουν δύο αρχιτεκτονικές βαθιάς μάθησης οι οποίες μπορούν να ενσωματωθούν και να λειτουργήσουν στα drones. Οι αρχιτεκτονικές τους βασίζονται στην αρχιτεκτονική του νευρωνικού δικτύου Deeplabv3+ το οποίο χρησιμοποιείται ως βάση για την τμηματοποίηση της εικόνας. Η πρωτοτυπία της ερευνητικής τους συνεισφοράς έγκειται στην ενσωμάτωση οπτικών μετατροπέων και συγκεκριμένα του MobileViT και του MobileNetv3. Κατορθώνουν με αυτό τον τρόπο να συλλαμβάνει το μοντέλο τους τις μακροσκοπικές συνεξαρτήσεις καθώς και τις λεπτές

διαφορές στις εικόνες-δεδομένα, ενώ η εν λόγω επιτυχής λειτουργία του μοντέλου καθίσταται δυνατή με μικρότερο αριθμό παραμέτρων και άρα μικρότερη υπολογιστική πολυπλοκότητα. Τα δύο διακριτά σύνολα δεδομένων-εικόνων που χρησιμοποιήθηκαν στην εργασία, ένα για τμηματοποίηση εικόνων φωτιάς και το άλλο για τμηματοποίηση εικόνων καπνού, βασίστηκαν στα γνωστά σύνολα δεδομένων Corsican, FLAME, SMOKE5K και AI-For-Mankind. Χρησιμοποιήθηκαν πέντε συναρτήσεις απώλειας (lossfunctions) που απαντώνται συχνά σε εργασίες τμηματοποίησης εικόνας. Οι συναρτήσεις αυτές είναι οι: Dice, Tversky, FocalTversky, BinaryCross-Entropy, ComboLoss. Οι δύο αρχιτεκτονικές που προτείνουν οι συγγραφείς τίθενται σε σύγκριση με το μοντέλο ResNet 50 με βάση όλες τις συναρτήσεις απώλειας. Οι διάφορες μετρικές που χρησιμοποιήθηκαν όπως οι MeanPixelAccuracy, F1 score, IntersectionoverUnion (IoU), Recall, Precision, MatthewsCorrelationCoefficient (MCC), καταδεικνύουν το γεγονός ότι το «υβριδικό μοντέλο» σύζευξης νευρωνικών δικτύων και οπτικών μετατροπών που παρουσιάζεται στην εργασία, τα καταφέρνει εξίσου καλά με το συνελκτικό νευρωνικό δίκτυο ResNet-50, ενώ συγχρόνως χρειάζεται λιγότερες παραμέτρους και άρα υπερτερεί σε υπολογιστική ακρίβεια.

Στην εργασία [13] οι συγγραφείς προτείνουν ένα νέο ολοκληρωμένο μοντέλο για την παρακολούθηση των δασών, το οποίο συνίσταται από τα ακόλουθα μέρη (α) μετάδοση των εικόνων που τραβήχτηκαν από drone στο πεδίο παρατήρησης (β) ανάλυση των εικόνων από ένα μοντέλο βαθιάς μάθησης (γ) μετάδοση των αποτελεσμάτων της ανάλυσης σε μια πλατφόρμα web και δυνατότητα παρατήρησης των αποτελεσμάτων αυτών από οποιαδήποτε συσκευή οπουδήποτε στον κόσμο. Για την φάση της ανάλυσης των εικόνων οι ερευνητές χρησιμοποιούν ως βάση την αρχιτεκτονική του μοντέλου YOLOv5. Η καινοτομία την οποία εισάγουν οι συγγραφείς αυτής του άρθρου είναι πως προκειμένου να χρησιμοποιήσουν αποδοτικότερα την πληροφορία από την εικόνα του υπέρυθρου φάσματος, προσθέτουν στο φίλτρο εισαγωγής δεδομένων της αρχιτεκτονικής του μοντέλου YOLOv5 ένα τέταρτο κανάλι (το ονομάζουν κανάλι Alpha). Τα δεδομένα εισαγωγής καθίστανται έτσι εικόνες στο «φάσμα RGBA». Προκειμένου να εκπαιδεύσουν το μοντέλο τους οι ερευνητές ανέπτυξαν εκ νέου ένα σύνολο δεδομένων –εικόνων που τραβήχτηκαν από drone σε αυθεντικά δασικά περιβάλλοντα στο ορατό και στο υπέρυθρο φάσμα. Στη σύγκριση του μοντέλου τους με άλλα μοντέλα στην αιχμή της τεχνολογίας

που επεξεργάζονται όμως εικόνες μόνο από το RGB φάσμα, διαπιστώθηκε ότι το μοντέλο της εργασίας αυτής κατέγραψε καλύτερα αποτελέσματα στις μετρικές της ακρίβειας που χρησιμοποιήθηκαν για αυτό το σκοπό, όπως Precision, RecallIoU, AveragePrecision και MeanAveragePrecision. Η καλύτερη απόδοση του μοντέλου στους δείκτες που προαναφέρθηκαν, πρέπει να εκτιμηθεί περαιτέρω δεδομένου ότι τα αποτελέσματα αυτά καταγράφηκαν σε εικόνες που χαρακτηρίζονταν από ελλιπή ορατότητα και από μερική απόκρυψη των προς καταγραφή αντικειμένων λόγω καπνού ή άλλων οπτικών εμποδίων.

Η εργασία [11] στοχεύει στην ανάπτυξη μιας αξιόπιστης μεθόδου για την πρόωμη ανίχνευση δασικών πυρκαγιών μέσω Μη Επανδρωμένων Αεροσκαφών (UAVs) η οποία μέθοδος βασίζεται στη σύζευξη των εικόνων του ορατού και του υπέρυθρου φάσματος οι οποίες και θα λαμβάνονται από τις κάμερες του drone. Σημειώνει πως η κάθε μία κατηγορία των εικόνων ενέχουν τους δικούς τους λειτουργικούς περιορισμούς. Οι εικόνες στο ορατό φάσμα είναι κατάλληλες για εργασίες αναγνώρισης καπνού όμως οι λήψεις τους παρεμποδίζονται από τυχόν ελλιπή φωτισμό ή συνθήκες χαμηλής ορατότητας. Από την άλλη πλευρά οι εικόνες στο υπέρυθρο φάσμα τείνουν να δίνουν ψευδώς θετικά αποτελέσματα (false positives) σε περιβάλλοντα με υψηλές θερμοκρασίες. Όσον αφορά τις μεθόδους που χρησιμοποιούνται στο άρθρο, η προσέγγιση των συγγραφέων ενσωματώνει προηγμένες τεχνικές Βαθιάς Μάθησης. Η ανίχνευση υλοποιείται μέσω του προτεινόμενου μοντέλου FFDM-F (Forest Fire Detection Model-Fusion), το οποίο βασίζεται στο YOLOv5s. Αρχιτεκτονικά, βελτιώνεται με ένα επίπεδο ανίχνευσης μικρών στόχων (small target detection layer), μια αποσυζευγμένη κεφαλή (decoupled head) για τον ανεξάρτητο υπολογισμό ταξινόμησης και χωρικού εντοπισμού, και τον ελαφρύ μηχανισμό προσοχής καναλιού ECA (Efficient Channel Attention). Ως προς το σύνολο δεδομένων που χρησιμοποιούν, αντιμετωπίζοντας την έλλειψη ανοιχτών δεδομένων, οι ερευνητές δημιούργησαν ένα σύνολο 2.752 συγχρονισμένων ζευγών εικόνων στο ορατό και υπέρυθρο φάσμα. Οι εικόνες ελήφθησαν από διάφορα ύψη πτήσης, γωνίες και τοπία, και υπεβλήθησαν σε τεχνικές επαύξησης. Το μοντέλο του άρθρου κατέγραψε εξαιρετική απόδοση, σημειώνοντας Ακρίβεια 0.952, Ανάκληση 0.943 και mAP@0.5 0.954. Στη συγκριτική αξιολόγηση με το αρχικό YOLOv5s, το FFDM-F αύξησε την ακρίβεια εντοπισμού μικρών εστιών κατά τουλάχιστον 10%. Παράλληλα, υπερέτρησε απέναντι σε προγενέστερα ή νεότερα μοντέλα της σειράς (όπως τα YOLOv7 και YOLOv8) στους

δείκτες Precision και Recall, επιβεβαιώνοντας τη δραστική μείωση των ψευδών συναγεμίων.

Στην εργασία [15] προτείνεται ένα μοντέλο βαθιάς μάθησης για ανίχνευση πυρκαγιών που οι ερευνητές ονομάζουν FireDetXplainer (FDX). Με το μοντέλο αυτό οι συγγραφείς επιχειρούν να προσφέρουν όχι μόνο υψηλή ακρίβεια ταξινόμησης, αλλά και διαφάνεια ως προς τον τρόπο λήψης των αποφάσεων του μοντέλου, ενισχύοντας με αυτό τον τρόπο την αξιοπιστία του συστήματος. Η αρχιτεκτονική του FDX βασίζεται στο προεκπαιδευμένο μοντέλο MobileNetV3, αξιοποιώντας μεθοδολογίες Μεταφοράς Μάθησης και επιλεκτικής Βελτιστοποίησης (Fine-Tuning) στα ανώτερα επίπεδά του μοντέλου. Κεντρικό ρόλο διαδραματίζει η εφαρμογή του πλαισίου «Μάθησης Χωρίς Λήθη» (Learning without Forgetting - LwF), το οποίο αποτρέπει την καταστροφική λήθη των ήδη αποκτηθέντων χαρακτηριστικών κατά την εκπαίδευση σε νέα δεδομένα. Επιπλέον, το μοντέλο ενσωματώνει εργαλεία Εξηγήσιμης Τεχνητής Νοημοσύνης (Explainable AI - XAI), και συγκεκριμένα τις μεθόδους Grad-CAM (για την παραγωγή θερμικών χαρτών ενεργοποίησης κλάσης) και LIME (για ανεξάρτητες από το μοντέλο τοπικές εξηγήσεις). Σχετικά με το σύνολο δεδομένων που χρησιμοποιήθηκε το μοντέλο εκπαιδεύτηκε και αξιολογήθηκε σε ένα ετερογενές, ισορροπημένο σύνολο 3.520 εικόνων (2.216 με παρουσία φωτιάς και 1.304 χωρίς), προερχόμενο από πλατφόρμες όπως αυτές των Kaggle και Mendeley. Τα δεδομένα περιλαμβάνουν ποικίλα περιβάλλοντα (πυκνά δάση, ανοιχτές εκτάσεις, αστικές παρυφές) και λήφθηκαν τόσο κατά τη διάρκεια της ημέρας όσο και της νύχτας. Ως προς την απόδοση του μοντέλου το FDX κατέγραψε εξαιρετικές επιδόσεις, επιτυγχάνοντας συνολικά Accuracy 99.91%, Precision 99.88%, Recall 99.97% και F1-Score 99.93%. Παρά τον όγκο των 6.5 εκατομμυρίων παραμέτρων του, διατήρησε πολύ χαμηλές απώλειες επικύρωσης (validation loss 0.0084), αποφεύγοντας την υπερπροσαρμογή (overfitting). Στη συγκριτική ανάλυση, το FDX ξεπέρασε κορυφαία προγενέστερα μοντέλα της βιβλιογραφίας (όπως τα ShuffleNetV2-OnFire, LW-FIRE και FireXplainer) σε όλους τους βασικούς δείκτες.

Η μελέτη [2] αποτελεί ένα εκτενές άρθρο επισκόπησης. Παρέχει μια εις βάθος και ολοκληρωμένη χαρτογράφηση της ενσωμάτωσης Μη Επανδρωμένων Αεροσκαφών (UAVs) και Τεχνητής Νοημοσύνης στη διαχείριση των δασικών πυρκαγιών. Η ανασκόπηση αυτή εξετάζει τον ρόλο των αλγορίθμων σε ολόκληρο τον κύκλο ζωής μιας

πυρκαγιάς: από την πρόληψη, στην ενεργή καταστολή και τέλος στη διαχείριση μετά την κατάσβεση. Κατά την ενεργή πυρκαγιά, επιστρατεύονται αρχιτεκτονικές Βαθιάς Μάθησης. Όσον αφορά στις αρχιτεκτονικές που υπάρχουν στο πεδίο προκειμένου για την εργασία της ταξινόμησης και ανίχνευσης αντικειμένων, επισημαίνεται ότι υπερέχουν τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs), όπως τα ResNet, MobileNet, EfficientNet, και κυρίως η οικογένεια αλγορίθμων YOLO (YOLOv3 έως YOLOv8). Για τη σημασιολογική τμηματοποίηση, αναλύονται οι αρχιτεκτονικές U-Net, DeepLabV3+ και Vision Transformers. Αξίζει να σημειωθεί ότι το άρθρο αποτελεί την πρώτη ανασκόπηση που αξιολογεί διεξοδικά τους αλγόριθμους Ενισχυτικής Μάθησης για τον συντονισμό πολλαπλών πρακτόρων. Σε σχέση με τα σύνολα δεδομένων αξιολογείται η χρήση πολυάριθμων ανοιχτών συνόλων, όπως τα Flame1, Flame2, Corsican, Kaggle, DeepFire και Foggia. Στο σημείο αυτό σημειώνεται από τους μελετητές το μεγάλο κενό στη διαθεσιμότητα πλούσιων, ετερογενών και πολυτροπικών οπτικών και θερμικών δεδομένων τα οποία να συνδυάζουν συγχρονισμένα δορυφορικές ροές, εναέρια δεδομένα και καιρικές συνθήκες. Επιπρόσθετα καταγράφεται μια συγκριτική αποτίμηση διάφορων μοντέλων. Βάσει της αποτίμησης αυτής καταδεικνύεται ότι τα ελαφριά δίκτυα όπως τα MobileNetV3 και SqueezeNet ενδείκνυνται για επεξεργασία στο άκρο εντός των περιορισμένων υπολογιστικά UAVs, διατηρώντας F1-Scores γύρω στο 95-98%. Η αρχιτεκτονική YOLO (ιδίως το YOLOv5) υπερέχει στην ταχύτητα ανίχνευσης πραγματικού χρόνου, αξιοποιώντας μηχανισμούς όπως το Spatial Pyramid Pooling (SPP). Στην τμηματοποίηση, ο συνδυασμός του ResNet με το U-Net καταγράφει τις υψηλότερες μετρικές ακρίβειας (έως 99%).

Η εργασία [12] στοχεύει στην ανάπτυξη ενός ολοκληρωμένου συστήματος για την ανίχνευση και τον ακριβή χωρικό εντοπισμό δασικών πυρκαγιών σε αρχικό στάδιο. Το σύστημα βασίζεται σε Μη Επανδρωμένα Αεροσκάφη εξοπλισμένα με τη συσκευή OAK-D (που περιλαμβάνει κάμερες με διόπτρες και κάμερα RGB), με σκοπό την παροχή δεδομένων σε πραγματικό χρόνο για την άμεση υποστήριξη των πυροσβεστικών δυνάμεων. Το ποσοστό επιτυχούς εντοπισμού των πλαισίων με φλόγες ανέρχεται σε 89,34% σε συνθήκες πραγματικού εντοπισμού. Όσον αφορά την ακρίβεια του χωρικού εντοπισμού, το μέσο απόλυτο σφάλμα ήταν ελάχιστο. Μόλις $7.341 \cdot 10^{-5}$ μοίρες για το γεωγραφικό μήκος, $3.664 \cdot 10^{-5}$ μοίρες για το πλάτος και 0.12 μέτρα για το υψόμετρο.

Σε συγκριτικές δοκιμές, το NanoDet υπερτέρησε απέναντι σε άλλα μοντέλα όπως το YOLOv5-s (βελτίωση 1.6% στη μετρική mAP). Σημαντικό σημείο είναι ότι το σύστημα που προτείνεται έχει το πλεονέκτημα της υψηλής υπολογιστικής ισχύος σε συνδυασμό με χαμηλή κατανάλωση ενέργειας. Συνελέχθησαν δεδομένα τα οποία τροποποιήθηκαν κατάλληλα για να αυξηθεί η απόδοση του συστήματος. Ειδικότερα για την εκπαίδευση του αλγορίθμου NanoDet, συγκεντρώθηκαν πάνω από 10.000 εικόνες από ανοιχτές βάσεις δεδομένων (BowFire, FiSmo, FireSense, VisiFire, ForestryImages). Επιπρόσθετα τραβήχτηκαν για τους σκοπούς της μελέτης 1400 εικόνες από UAV's. Ένα μειονέκτημα που εντοπίζεται από τους συγγραφείς και χρήζει περαιτέρω μελέτης είναι ότι το σύστημα αποδεικνύεται ασθενές στο εντοπισμό μικρών εστιών φωτιάς.

Το άρθρο [1] διερευνά την αποτελεσματικότητα του υπερσύγχρονου αλγορίθμου YOLOv9 στην ανίχνευση δασικών πυρκαγιών σε πραγματικό χρόνο, χρησιμοποιώντας εικόνες που συλλέγονται από Μη Επανδρωμένα Αεροσκάφη. Αφού πρώτα το άρθρο κάνει μια σύντομη ανασκόπηση στα βασικά χαρακτηριστικά των αλγορίθμων YOLOv5 και YOLOv8, ακολούθως εστιάζει την προσοχή του στον αλγόριθμο YOLOv9. Ο αλγόριθμος YOLOv9 ο οποίος έκανε το ντεμπούτο του στις 21-2-2024 είναι ένας αλγόριθμος ανίχνευσης αντικειμένων ενός σταδίου. Χρησιμοποιεί το Spatial Pyramid Pooling Network (SPPNet) ούτως ώστε να επεξεργαστεί ολόκληρη την εικόνα μία φορά με σκοπό την ανίχνευση αντικειμένων. Η αρχιτεκτονική ανίχνευσης η οποία βασίζεται στο μοντέλο YOLOv9, εισάγει τρεις κρίσιμες δομικές αναβαθμίσεις για την επίλυση προβλημάτων σύγκλισης που παρατηρούνταν σε παλαιότερες εκδόσεις. Πρώτον, ενσωματώνει την Προγραμματιζόμενη Πληροφορία Διαβάθμισης, η οποία δημιουργεί ασφαλείς διαβαθμίσεις μέσω ενός αναστρέψιμου κλάδου, αποτρέποντας την απώλεια κρίσιμων χαρακτηριστικών κατά τη μάθηση. Δεύτερον, υιοθετεί τον σχεδιασμό GELAN (Generalized Efficient Layer Aggregation Network), ο οποίος συγχωνεύει τα δίκτυα CSPNet και ELAN, δημιουργώντας μια πιο ελαφριά και υπολογιστικά αποδοτική αρχιτεκτονική νευρωνικού δικτύου. Επίσης μια τρίτη καινοτομία της αρχιτεκτονικής του YOLOv9 είναι η χρήση διακριτών πυραμίδων χαρακτηριστικών ειδικά προσαρμοσμένων για την ανίχνευση μικρών αντικειμένων. Tashallow features επιστρατεύονται για τον εντοπισμό μικρών, αρχικών εστιών φωτιάς, ωθώντας τα μεγαλύτερα αντικείμενα του υποβάθρου σε δεύτερο πλάνο, επιτρέποντας έτσι εξαιρετικά ταχύτερα και ακριβέστερα

συμπεράσματα. Λόγω της εγγενούς δυσκολίας συλλογής στιγμιαίων εικόνων πραγματικού χρόνου κατά την έναρξη πυρκαγιών, οι ερευνητές συγκέντρωσαν ένα προσαρμοσμένο σύνολο δεδομένων από διαδικτυακές πλατφόρμες. Σε αντίθεση με προγενέστερες μελέτες που βασίζονται αυστηρά σε ποσοτικές μετρικές (όπως mAP ή F1-score), η παρούσα μελέτη πραγματοποιεί μια ποιοτική αξιολόγηση των αποτελεσμάτων. Μέσα από συγκεκριμένα οπτικά παραδείγματα η μελέτη αποδεικνύει ότι το αλγόριθμος μπορεί να οριοθετήσει με εξαιρετική ευαισθησία και ακρίβεια, μέσω των πλαισίων οριοθέτησης, τις περιοχές που έχουν πληγεί από τη φωτιά. Ιδιαίτερη έμφαση δίνεται στην προσαρμοστικότητα του αλγορίθμου σε διαφορετικές περιβαλλοντικές συνθήκες και εντάσεις πυρκαγιάς, καθώς και στην ικανότητά του να διακρίνει τη φωτιά μέσα από δύσκολα σενάρια που αφορούν τις καιρικές συνθήκες και την τοπογραφία του πεδίου. Επιπρόσθετα η ποιοτική αξιολόγηση μελετά τις περιπτώσεις όπου το μοντέλο δυσκολεύεται ή αποτυγχάνει.

Στην εργασία [17] παρουσιάζεται ο σχεδιασμός και η ανάπτυξη ενός ολοκληρωμένου, αυτόνομου συστήματος επιτήρησης για την έγκαιρη ανίχνευση δασικών πυρκαγιών. Ο βασικός στόχος είναι η ενσωμάτωση αλγορίθμων Βαθιάς Μάθησης απευθείας σε ένα Μη Επανδρωμένο Αεροσκάφος, επιτρέποντας την επεξεργασία δεδομένων στο άκρο για την αποστολή οπτικών δεδομένων και συντεταγμένων σε πραγματικό χρόνο σε έναν επίγειο σταθμό ελέγχου. Σε επίπεδο λογισμικού, η μελέτη συγκρίνει δύο προσεγγίσεις: την ταξινόμηση εικόνας και την ανίχνευση αντικειμένων. Για την ταξινόμηση, αναπτύχθηκε ένα υβριδικό μοντέλο CNN-RCNN, το οποίο συνδυάζει συνελκτικά επίπεδα για την εξαγωγή χωρικών χαρακτηριστικών με αμφίδρομα αναδρομικά επίπεδα (Bidirectional LSTM και GRU) για την επεξεργασία της χρονικής ακολουθίας. Για την ανίχνευση αντικειμένων, αξιολογήθηκαν οι αλγόριθμοι YOLOv5 και YOLOv8. Εξαιρετικό ενδιαφέρον παρουσιάζει η άμεση συγκριτική αξιολόγηση μεταξύ μοντέλων ταξινόμησης (όπως το προτεινόμενο CNN-RCNN) και μοντέλων ανίχνευσης αντικειμένων (YOLO) ως προς τις υπολογιστικές τους απαιτήσεις και την απόδοσή τους σε περιβάλλοντα με περιορισμένους πόρους επεξεργασίας, όπως ο μικροϋπολογιστής Jetson Nano. Για την εκπαίδευση και επικύρωση των μοντέλων, συγκροτήθηκε ένα ετερογενές σύνολο δεδομένων 2.947 εικόνων, εκ των οποίων 1.672 περιείχαν φωτιά και 1.275 απεικόνιζαν φυσικό περιβάλλον χωρίς φωτιά. Τα δεδομένα αντλήθηκαν από την πλατφόρμα Mendeley

καθώς και από προσωπικές συλλογές των ερευνητών. Κατά τη δοκιμαστική αξιολόγηση, τόσο το προτεινόμενο υβριδικό μοντέλο CNN-RCNN όσο και το μοντέλο ταξινόμησης του YOLOv8 σημείωσαν εξαιρετικές επιδόσεις, επιτυγχάνοντας Ακρίβεια (Accuracy) 96%. Αντίθετα, τα μοντέλα ανίχνευσης αντικειμένων YOLOv5n και YOLOv8n κατέγραψαν ελαφρώς χαμηλότερη ακρίβεια, φτάνοντας το 89%. Το πιο κρίσιμο συμπέρασμα της μελέτης ήταν πως, κατά την εφαρμογή σε πραγματικό χρόνο, τα μοντέλα ταξινόμησης αποδείχθηκαν υπολογιστικά πολύ πιο αποδοτικά από τα μοντέλα ανίχνευσης αντικειμένων, καθιστώντας τα ιδανικά για το συγκεκριμένο περιβάλλον λειτουργίας.

Στο άρθρο [4] εισάγεται το FLAME2, ένα σύνολο δεδομένων το οποίο αποτελείται από ταυτόχρονες λήψεις εικόνων στο ορατό (RGB) και υπέρυθρο (IR) φάσμα, συλλεγμένες από Μη Επανδρωμένα Αεροσκάφη κατά τη διάρκεια προδιαγεγραμμένων (ελεγχόμενων) καύσεων. Κεντρικός στόχος της έρευνας είναι η ανάπτυξη και η συγκριτική αξιολόγηση αρχιτεκτονικών Βαθιάς Μάθησης που αξιοποιούν τη συγχώνευση αυτών των πολυτροπικών δεδομένων για την ταχεία ταξινόμηση των εικόνων και τον ακριβή χωρικό εντοπισμό της πυρκαγιάς. Όσον αφορά το σύνολο δεδομένων στα πλαίσια της παρούσας έρευνας δημιουργήθηκε το ελεύθερα διαθέσιμο FLAME2, που περιλαμβάνει 53.451 ζεύγη καρέ υψηλής και μειωμένης ανάλυσης (254x254) για ταχύτερη επεξεργασία. Το άρθρο παρέχει επίσης ένα πολύτιμο συμπληρωματικό σύνολο δεδομένων, το οποίο ενσωματώνει προ-πυρικά νέφη σημείων LiDAR, Ψηφιακά Μοντέλα Εδάφους και αναλυτικά μετεωρολογικά στοιχεία της περιοχής καύσης. Προκειμένου για την αξιολόγηση των μοντέλων επέλεξαν μετρικές μακρο-επιπέδου (Macro F1-score, Macro Recall, Macro Precision), οι οποίες υπολογίζουν τον μέσο όρο των μετρικών για κάθε κλάση ανεξάρτητα από το μέγεθός της. Στα πλαίσια της μελέτης συγκρίθηκαν κλασικά προεκπαιδευμένα μοντέλα (όπως τα ResNet18, VGG16, MobileNetV2) με προσαρμοσμένα δίκτυα (όπως τα Logistic Regression, LeNet5 και το δικό τους Flame net) τα οποία εκπαιδεύτηκαν από την αρχή πάνω στο σύνολο δεδομένων FLAME2. Μοντέλα όπως το ResNet18 και το MobileNetV2 υπερέτρησαν σημαντικά των υπολοίπων. Λόγω της προεκπαίδευσής τους σε τεράστια σύνολα δεδομένων (όπως το ImageNet), διέθεταν ήδη πλούσια εμπειρία στην κατανόηση οπτικών χαρακτηριστικών. Συγκεκριμένα, το ResNet18 με όψιμη συγχώνευση σημείωσε την υψηλότερη απόδοση. Αν και τα

προεκπαιδευμένα μοντέλα πέτυχαν την υψηλότερη ακρίβεια, οι συγγραφείς τονίζουν ότι το προσαρμοσμένο "Flame net" έχει ένα μεγάλο πρακτικό πλεονέκτημα: διαθέτει λιγότερες από 1.000 εκπαιδύσιμες παραμέτρους. Αυτό συνεπάγεται μικρότερο χρόνο εκπαίδευσης και υπολογιστικό κόστος, καθιστώντας το ιδανικό για συστήματα με περιορισμένους πόρους. Παρόλοτιν εξαιρετικά υψηλή απόδοση, οι συγγραφείς αναγνωρίζουν τον περιορισμό της γενίκευσης, καθώς η εκπαίδευση με υψηλή ακρίβεια αφορούσε ένα συγκεκριμένο σύνολο.

2.2 Σύνολα Δεδομένων

Στην παρούσα παράγραφο γίνεται μια σύντομη ανασκόπηση των βασικότερων συνόλων δεδομένων που αναφέρονται στις εργασίες που μελετήσαμε. Μετά από μια σύντομη ανάλυση για κάθε σύνολο, ακολουθώς παρατίθενται κάποια χαρακτηριστικά instances του συνόλου.

1) FD (FireDetection)dataset[1]

Το FD dataset αποτελεί ένα δημόσια διαθέσιμο σύνολο δεδομένων μεγάλης κλίμακας για την ανίχνευση πυρκαγιάς μέσω τεχνικών βαθιάς μάθησης και υπολογιστικής όρασης. Έχει σχεδιαστεί κυρίως για εφαρμογές ταξινόμησης εικόνων. Το σύνολο περιλαμβάνει περίπου 50.000 έγχρωμες εικόνες RGB, με περίπου 25.000 εικόνες που περιέχουν φωτιά και άλλες τόσες χωρίς φωτιά. Το FD dataset δημιουργήθηκε μέσω της ενοποίησης εικόνων από δύο μεγάλα σύνολα δεδομένων ανίχνευσης πυρκαγιάς, τα BoWFire και Foggia Fire Dataset. Η συγχώνευση αυτή είχε ως στόχο τη δημιουργία ενός μεγαλύτερου και περισσότερο ετερογενούς συνόλου, ικανού να υποστηρίξει την εκπαίδευση βαθιών νευρωνικών δικτύων. Οι εικόνες συνοδεύονται από ετικέτες σε επίπεδο εικόνας και ταξινομούνται αποκλειστικά στις δύο κλάσεις Fire και No Fire. Το γεγονός αυτό καθιστά το σύνολο κατάλληλο για δυαδική ταξινόμηση εικόνων. Οι εικόνες προέρχονται κυρίως από επίγειες λήψεις και απεικονίζουν ποικίλα περιβάλλοντα, όπως αστικές περιοχές, αγροτικές εκτάσεις, δασικά τοπία καθώς και περιστατικά πυρκαγιάς σε οχήματα ή κτίρια. Η συλλογή τους πραγματοποιήθηκε από δημόσια διαθέσιμες διαδικτυακές πηγές, και οπτικό υλικό επιτήρησης. Ως συμπληρωματικά δεδομένα, το FD Dataset παρέχει τις αντίστοιχες ετικέτες κατηγορίας για κάθε εικόνα, χωρίς επιπλέον χωρική πληροφορία. Η μεγάλη

κλίμακα του συνόλου, η ισορροπημένη κατανομή των κλάσεων και η ποικιλία των σκηνών το καθιστούν ιδιαίτερα κατάλληλο για την αξιολόγηση μοντέλων βαθιάς μάθησης, όπως CNNs, Vision Transformers και υβριδικών αρχιτεκτονικών για εφαρμογές αυτόματης ανίχνευσης πυρκαγιάς.



Εικόνα 2-1 Στιγμιότυπα από το FDDataset

2) DFAN (Dual Fire Attention Network) dataset [22]

Το DFAN δημιουργήθηκε με σκοπό την αντιμετώπιση ενός σημαντικού περιορισμού των περισσότερων διαθέσιμων συνόλων δεδομένων ανίχνευσης πυρκαγιάς, τα οποία συνήθως περιορίζονται σε δυαδική ταξινόμηση μεταξύ των κατηγοριών Fire και No Fire. Σε αντίθεση με τα κλασικά fire datasets, το DFAN εισάγει ένα πρόβλημα ταξινόμησης πολλαπλών κλάσεων, αποτελούμενο από δώδεκα διαφορετικές κατηγορίες. Η σχεδίαση αυτή επιτρέπει όχι μόνο την αναγνώριση της παρουσίας φωτιάς αλλά και την καλύτερη κατανόηση των αντικειμένων ή των σεναρίων που σχετίζονται με το περιστατικό πυρκαγιάς. Το DFAN Dataset αποτελείται από 3.804 εικόνες RGB που προέρχονται κυρίως από επίγειες λήψεις και περιλαμβάνουν σκηνές από διαφορετικά περιβάλλοντα και συνθήκες φωτισμού. Οι εικόνες συνοδεύονται από ετικέτες σε επίπεδο εικόνας, χωρίς χωρικούς σχολιασμούς όπως bounding boxes ή segmentation masks. Έτσι, το σύνολο είναι κατάλληλο για εργασίες image classification. Η συλλογή των εικόνων πραγματοποιήθηκε από δημόσια διαθέσιμες διαδικτυακές πηγές και πραγματικά περιστατικά πυρκαγιάς.



Εικόνα 2-2 Στιγμιότυπα από το DFANdataset

3) ADSF dataset

Το ADSF (AdvancedDeepSupervisionFire) παρουσιάστηκε το 2023 στο [23] με στόχο την αντιμετώπιση ενός σημαντικού περιορισμού των υπαρχόντων συνόλων δεδομένων ανίχνευσης πυρκαγιάς. Τα περισσότερα διαθέσιμα datasets βασίζονταν σε εικόνες που προέρχονται από σταθερές κάμερες επιτήρησης (CCTV), οι οποίες όπως είναι φυσικό καλύπτουν ένα περιορισμένο οπτικό πεδίο και δεν παρέχουν δυνατότητα μεταβολής της γωνίας λήψης ή παρακολούθησης εκτεταμένων περιοχών. Για τον λόγο αυτό οι δημιουργοί του ADSF συνέλεξαν δεδομένα από εναέριες πλατφόρμες και δορυφορικές εικόνες, δημιουργώντας ένα πιο αντιπροσωπευτικό σύνολο δεδομένων για εφαρμογές έγκαιρης ανίχνευσης πυρκαγιάς. Το σύνολο περιλαμβάνει συνολικά 6000 εικόνες RGB, εκ των οποίων 3000 περιέχουν περιστατικά πυρκαγιάς και 3000 δεν περιέχουν φωτιά. Το σύνολο δεδομένων είναι κατάλληλο για εργασίες δυαδικής ταξινόμησης εικόνων. Οι εικόνες συνοδεύονται από ετικέτες σε επίπεδο εικόνας και δεν παρέχουν χωρικούς σχολιασμούς όπως bounding boxes ή segmentation masks. Οι εικόνες προέρχονται από συνδυασμό εναέριων λήψεων μέσω μη επανδρωμένων αεροσκαφών και δορυφορικών εικόνων. Για τη συλλογή των δεδομένων από drone χρησιμοποιήθηκε το DJI Spark, μέσω του οποίου καταγράφηκαν βίντεο πραγματικών περιστατικών πυρκαγιάς και εξήχθησαν αντιπροσωπευτικά καρέ για τη δημιουργία του συνόλου. Η χρήση δεδομένων από διαφορετικά ύψη και γωνίες λήψης καθιστά το ADSF ιδιαίτερα χρήσιμο για την ανάπτυξη συστημάτων έγκαιρης ανίχνευσης δασικών πυρκαγιών μέσω drones και δορυφορικής παρατήρησης.



Εικόνα 2-3 Στιγμιότυπα από το ADSFdataset

4) α)FLAME 1

Το FLAME (FireLuminosityAirborne-basedMachineLearningEvaluation) dataset αποτελεί ένα δημόσια διαθέσιμο σύνολο δεδομένων που σχεδιάστηκε για την ανάπτυξη και αξιολόγηση αλγορίθμων ανίχνευσης και εντοπισμού πυρκαγιάς σε εναέριας εικόνες. Το σύνολο παρουσιάστηκε το 2021 στο [16] με στόχο την υποστήριξη εφαρμογών έγκαιρης ανίχνευσης δασικών πυρκαγιών μέσω μη επανδρωμένων αεροσκαφών (UAVs). Το FLAME επικεντρώνεται αποκλειστικά σε εναέριας λήψεις, προσομοιώνοντας ρεαλιστικά σενάρια επιχειρησιακής παρακολούθησης από drones. Το σύνολο περιλαμβάνει συνολικά περίπου 48.000 εικόνες RGB, οι οποίες προέρχονται από βίντεο που καταγράφηκαν με UAVs . Οι εικόνες ταξινομούνται σε δύο κατηγορίες (Fire και No Fire) και συνοδεύονται από ετικέτες σε επίπεδο εικόνας για εφαρμογές ταξινόμησης. Επιπλέον, το FLAME παρέχει pixel-level segmentation masks, γεγονός που το καθιστά κατάλληλο και για εργασίες semantic segmentation, όπου απαιτείται ο ακριβής εντοπισμός της περιοχής της φωτιάς μέσα στην εικόνα. Πιο συγκεκριμένα το σύνολο περιέχει περίπου 39.375 training frames, 8.617 test frames για classification και 2.003 εικόνες με αντίστοιχες segmentation masks. Οι εικόνες προέρχονται κυρίως από αγροτικά και δασικά περιβάλλοντα, καθώς και περιοχές ελεγχόμενων καύσεων (rural and prescribed burns), παρουσιάζοντας σημαντική ποικιλία ως προς το μέγεθος της φωτιάς, την πυκνότητα του καπνού και τις συνθήκες φωτισμού.



Εικόνα 2-4 Στιγμιότυπα από το FLAME 1 dataset

β) FLAME 2

Το FLAME 2 (Fire detection and modeLing: Aerial Multi-spectral image Dataset) αποτελεί τη δεύτερη γενιά της οικογένειας FLAME και παρουσιάστηκε το 2022 με στόχο την υποστήριξη της έρευνας στην εναέρια ανίχνευση και παρακολούθηση δασικών πυρκαγιών μέσω μη επανδρωμένων αεροσκαφών (UAVs). Το σύνολο δημιουργήθηκε για να αντιμετωπίσει τους περιορισμούς των παραδοσιακών fire datasets, τα οποία συνήθως περιέχουν μόνο εικόνες RGB και περιορισμένη πληροφορία σχετικά με τη θερμική συμπεριφορά της φωτιάς. Το FLAME 2 αποτελεί ένα πολυφασματικό (multi-spectral) σύνολο δεδομένων και περιλαμβάνει συγχρονισμένα ζεύγη εικόνων και βίντεο από δύο διαφορετικούς αισθητήρες: έναν αισθητήρα ορατού φάσματος (RGB) και έναν θερμικό υπέρυθρο αισθητήρα (Infrared – IR). Τα δεδομένα συλλέχθηκαν μέσω UAV κατά τη διάρκεια ελεγχόμενης καύσης (prescribed burn) σε πευκοδάσος ανοικτής κόμης στη Βόρεια Αριζόνα το 2021. Το σύνολο προορίζεται κυρίως για εργασίες ταξινόμησης εικόνων, ανίχνευσης πυρκαγιάς και semantic segmentation. Οι εικόνες συνοδεύονται από ετικέτες σε επίπεδο εικόνας που δηλώνουν την παρουσία ή απουσία φωτιάς, ενώ η ταυτόχρονη διαθεσιμότητα RGB και IR δεδομένων επιτρέπει τη μελέτη μοντέλων πολυτροπικής μάθησης. Σε αντίθεση με το αρχικό FLAME, το FLAME 2 δίνει ιδιαίτερη έμφαση στην αξιοποίηση θερμικής πληροφορίας για την ανίχνευση φλόγας ακόμη και όταν αυτή είναι μερικώς καλυμμένη από καπνό ή βλάστηση. Οι εικόνες προέρχονται αποκλειστικά από εναέριες λήψεις, γεγονός που καθιστά το dataset ιδιαίτερα κατάλληλο για εφαρμογές ανίχνευσης πυρκαγιάς με χρήση UAVs.



Εικόνα 2-5 Στιγμιότυπα από το FLAME 2 dataset

5) CorsicanFire Database

Το Corsican Fire Database , γνωστό και ως Corsican dataset, αποτελεί ένα από τα πλέον καθιερωμένα σύνολα δεδομένων για την έρευνα στην ανίχνευση και τμηματοποίηση δασικών πυρκαγιών μέσω τεχνικών υπολογιστικής όρασης. Αναπτύχθηκε από το Πανεπιστήμιο της Κορσικής (Université de Corse Pascal Paoli) με σκοπό τη δημιουργία ενός κοινού σημείου αναφοράς (benchmark) για την αξιολόγηση αλγορίθμων ανίχνευσης φωτιάς σε πραγματικές συνθήκες. Η βάση δεδομένων περιλαμβάνει περίπου 500 εικόνες RGB, 100 ζεύγη συγχρονισμένων εικόνων RGB–NIR (Near Infrared) καθώς και 5 ακολουθίες πολυφασματικών εικόνων, οι οποίες καταγράφηκαν σε διαφορετικές συνθήκες καύσης, τύπους βλάστησης, αποστάσεις λήψης, υπό διαφορετικό φωτισμό και σε ποικίλες καιρικές συνθήκες. Οι εικόνες προέρχονται κυρίως από επίγειες λήψεις πραγματικών ή ελεγχόμενων πυρκαγιών στην περιοχή της Κορσικής. Η συλλογή πραγματοποιήθηκε μέσω ειδικών οπτικών και υπέρυθρων αισθητήρων που κατέγραφαν ταυτόχρονα το ορατό και το υπέρυθρο-φάσμα. Το Corsican dataset χρησιμοποιείται κυρίως για εργασίες fire detection, semantic segmentation και multispectral image analysis. Σε αντίθεση με πολλά σύγχρονα datasets ταξινόμησης, παρέχει αναλυτικά δεδομένα σε επίπεδο εικονοστοιχείου. Για κάθε εικόνα διατίθενται χειροκίνητα παραγόμενες μάσκες όπου επισημαίνονται τα pixels που αντιστοιχούν σε φωτιά, καθώς και πρόσθετες πληροφορίες όπως το ποσοστό των fire pixels, ο βαθμός κάλυψης από καπνό, η κυρίαρχη χρωματική σύσταση της φλόγας και χαρακτηριστικά υφής της περιοχής καύσης. Η από κοινού ύπαρξη όλων αυτών των στοιχείων καθιστά το Corsican ένα από τα πιο χρήσιμα benchmarks για έρευνα στην αυτόματη ανίχνευση και τμηματοποίηση δασικών πυρκαγιών.



Εικόνα 2-6 Στιγμιότυπα από το Corsicandataset

6) SMOKE5K

Το SMOKE5K αποτελεί ένα σύγχρονο και υψηλής ποιότητας σύνολο δεδομένων που παρουσιάστηκε στη βιβλιογραφία το 2023 [20]. Σχεδιάστηκε για την ανάπτυξη και αξιολόγηση αλγορίθμων τμηματοποίησης καπνού (smoke segmentation). Σε αντίθεση με πολλά προγενέστερα datasets που επικεντρώνονται κυρίως στη φωτιά, το SMOKE5K εστιάζει αποκλειστικά στον καπνό, ο οποίος αποτελεί κρίσιμο δείκτη έγκαιρης ανίχνευσης πυρκαγιάς. Το σύνολο περιλαμβάνει συνολικά περίπου 5.400 εικόνες, αποτελούμενες από 1.400 πραγματικές εικόνες καπνού και 4.000 συνθετικές εικόνες, οι οποίες δημιουργήθηκαν για να αυξήσουν την ποικιλία των συνθηκών και των σχημάτων καπνού. Όλες οι εικόνες συνοδεύονται από pixel-level annotations, δηλαδή μάσκες τμηματοποίησης που προσδιορίζουν με ακρίβεια τα εικονοστοιχεία που αντιστοιχούν στον καπνό. Η ύπαρξη αυτών των αναλυτικών σχολιασμών καθιστά το SMOKE5K κατάλληλο κυρίως για εργασίες semantic segmentation και όχι για απλή ταξινόμηση εικόνων. Οι εικόνες είναι αποκλειστικά RGB και προέρχονται κυρίως από επίγειες λήψεις πραγματικών περιστατικών καπνού σε εξωτερικά περιβάλλοντα. Το dataset περιλαμβάνει μεγάλο εύρος συνθηκών φωτισμού, διαφορετικές πυκνότητες καπνού, ποικίλες αποστάσεις λήψης και σύνθετα υπόβαθρα. Ως συμπληρωματικά δεδομένα, το SMOKE5K παρέχει αναλυτικές μάσκες τμηματοποίησης για κάθε εικόνα. Η υψηλή ποιότητα των σχολιασμών και η συνδυασμένη χρήση πραγματικών και συνθετικών δεδομένων έχουν καταστήσει το SMOKE5K ένα από τα πιο διαδεδομένα benchmarks για έρευνα στη semantic segmentation καπνού.



Εικόνα 2-7 Στιγμιότυπα από το SMOKE5Kdataset

7) AI-For-Mankind Wildfire Smoke Dataset

Το AI For Mankind Wildfire Smoke dataset αποτελεί ένα δημόσια διαθέσιμο σύνολο δεδομένων που δημιουργήθηκε με σκοπό την ανάπτυξη συστημάτων έγκαιρης ανίχνευσης καπνού από δασικές πυρκαγιές μέσω τεχνικών βαθιάς μάθησης και υπολογιστικής όρασης. Το dataset αναπτύχθηκε από τον οργανισμό AI For Mankind σε συνεργασία με το πρόγραμμα HPWREN (High Performance Wireless Research and Education Network) του Πανεπιστημίου της Καλιφόρνιας, στο πλαίσιο ερευνητικών δράσεων για την πρόληψη και την έγκαιρη αντιμετώπιση δασικών πυρκαγιών. Η βασική φιλοσοφία του συνόλου είναι η ανίχνευση καπνού στα πρώτα στάδια εκδήλωσης μιας πυρκαγιάς, πριν η φωτιά εξελιχθεί σε μεγάλο μέτωπο. Το σύνολο περιλαμβάνει περίπου 737 εικόνες καπνού, οι οποίες προέρχονται από κάμερες παρακολούθησης βουνοκορφών του δικτύου HPWREN στην Καλιφόρνια. Οι εικόνες συνοδεύονται από bounding-box annotations, τα οποία επισημαίνουν τη θέση του καπνού μέσα στην εικόνα. Ως εκ τούτου, το dataset προορίζεται κυρίως για εργασίες object detection. Η μοναδική κλάση που επισημαίνεται είναι η κλάση Smoke, ενώ οι περιοχές εκτός των σχολιασμένων πλαισίων θεωρούνται υπόβαθρο. Οι εικόνες είναι αποκλειστικά RGB και προέρχονται από επίγειες λήψεις σταθερών καμερών επιτήρησης μεγάλης εμβέλειας. Το περιβάλλον λήψης περιλαμβάνει κυρίως ορεινές και δασικές περιοχές της Καλιφόρνιας, με σημαντικές μεταβολές στις καιρικές συνθήκες, τον φωτισμό και την απόσταση του καπνού από την κάμερα. Η συλλογή των δεδομένων πραγματοποιήθηκε από πραγματικά περιστατικά πυρκαγιάς που

καταγράφηκαν από το δίκτυο HPWREN και σχολιάστηκαν χειροκίνητα από εθελοντές του AI For Mankind.



Εικόνα 2-8 Στιγμιότυπα από το AI For Mankind Wildfire Smoke dataset

8) COCO dataset

Το COCO dataset παρουσιάστηκε το 2014 στο [11] με στόχο την προώθηση της έρευνας στην αναγνώριση αντικειμένων μέσα σε πολύπλοκες καθημερινές σκηνές. Το COCO δεν περιέχει απλώς μεμονωμένα αντικείμενα αλλά τα τοποθετεί μέσα στο φυσικό τους περιβάλλον, επιτρέποντας τη μελέτη του αντικειμένου σε συνάρτηση με το περιβάλλον και τα υπόλοιπα αντικείμενα της σκηνής. Το dataset περιλαμβάνει περίπου 330.000 εικόνες, εκ των οποίων περισσότερες από 200.000 είναι πλήρως σχολιασμένες, με πάνω από 1,5 εκατομμύριο επισημασμένα αντικείμενα. Η αρχική έκδοση περιγράφει 91 κατηγορίες αντικειμένων, όπως ουρανός, δρόμος, γρασίδι, νερό. Οι εικόνες είναι μόνο RGB και το dataset περιλαμβάνει αστικά, αγροτικά, εσωτερικά και εξωτερικά περιβάλλοντα, προσφέροντας μεγάλη ποικιλία σκηνών. Ένα από τα σημαντικότερα χαρακτηριστικά του COCO είναι ο πλούσιος σχολιασμός του. Ενδεικτικά, μεταξύ άλλων, παρέχονται boundingboxes για object detection, instance segmentation masks και image captions (πέντε περιγραφές ανά εικόνα).



Εικόνα 2-9 Στιγμιότυπα από το COCO dataset

9) DeepFire

Το σύνολο δεδομένων DeepFire παρουσιάστηκε το 2022 στην εργασία [7] και αποτελεί ένα από τα πρώτα datasets που δημιουργήθηκαν αποκλειστικά για δασικές πυρκαγιές (forest fire classification). Οι δημιουργοί του επισήμαναν ότι πολλά διαθέσιμα datasets της εποχής είτε ήταν μικρά είτε περιλάμβαναν γενικές εικόνες φωτιάς χωρίς να εστιάζουν αποκλειστικά σε δασικά περιβάλλοντα. Για τον λόγο αυτό συνέλεξαν και επιμελήθηκαν ένα νέο σύνολο δεδομένων ειδικά προσανατολισμένο στη διάκριση εικόνων δασικών περιοχών με και χωρίς φωτιά. Το dataset περιλαμβάνει συνολικά 1.900 εικόνες, κατανεμημένες ισόρροπα στις κατηγορίες Fire (950 εικόνες) και No Fire (950 εικόνες). Οι εικόνες συλλέχθηκαν από διάφορες διαδικτυακές πηγές και στη συνέχεια πραγματοποιήθηκε χειροκίνητος καθαρισμός και περικοπή των εικόνων ώστε να αφαιρεθούν τυχόν μη σχετικά αντικείμενα (άνθρωποι, οχήματα πυρόσβεσης κ.λπ.) και να διατηρηθεί μόνο η σχετική δασική περιοχή ενδιαφέροντος. Το DeepFire είναι σχεδιασμένο αποκλειστικά για binary image classification μιας και οι σχολιασμοί του βρίσκονται σε επίπεδο εικόνας. Οι εικόνες είναι μόνο στο φάσμα RGB, προέρχονται κυρίως από επίγειες λήψεις και εστιάζουν σε δασικά και ορεινά τοπία.



Εικόνα 2-10 Στιγμιότυπα από το DeepFire dataset

10) Foggia (ή MIVIA)

Πρωτοπαρουσιάστηκε το 2015 στο [6]. Στη βιβλιογραφία συχνά αναφέρεται ως MIVIA Fire Detection Dataset επειδή δημιουργήθηκε από το εργαστήριο MIVIA του Πανεπιστημίου του Salerno υπό την καθοδήγηση του P. Foggia. Δημιουργήθηκε με στόχο την αξιολόγηση αλγορίθμων ανίχνευσης φωτιάς σε πραγματικά συστήματα

βιντεοεπιτήρησης και αποτέλεσε ένα από τα πρώτα δημόσια benchmarks για video-based fire detection. Αποτελείται από 31 βίντεο, εκ των οποίων 14 περιέχουν περιστατικά φωτιάς και 17 αντιστοιχούν σε σκηνές χωρίς φωτιά. Οι σκηνές προέρχονται τόσο από πραγματικά περιβάλλοντα όσο και από υλικό που συλλέχθηκε από το διαδίκτυο. Τα δεδομένα αποτελούνται από RGB βίντεο που καταγράφηκαν από σταθερές κάμερες επιτήρησης (CCTV). Οι σκηνές περιλαμβάνουν εσωτερικούς και εξωτερικούς χώρους, βιομηχανικά περιβάλλοντα, αστικές περιοχές και υπαίθριες σκηνές. Ως συμπληρωματικά δεδομένα (supplemental data), το dataset παρέχει τα ίδια τα βίντεο και την αντίστοιχη ετικέτα fire/non-fire για κάθε ακολουθία. Λόγω της φύσης των δεδομένων του το σύνολο Foggia είναι ένα από τα πιο συχνά χρησιμοποιούμενα benchmarks για αξιολόγηση αλγορίθμων ανίχνευσης φωτιάς σε βίντεο επιτήρησης.

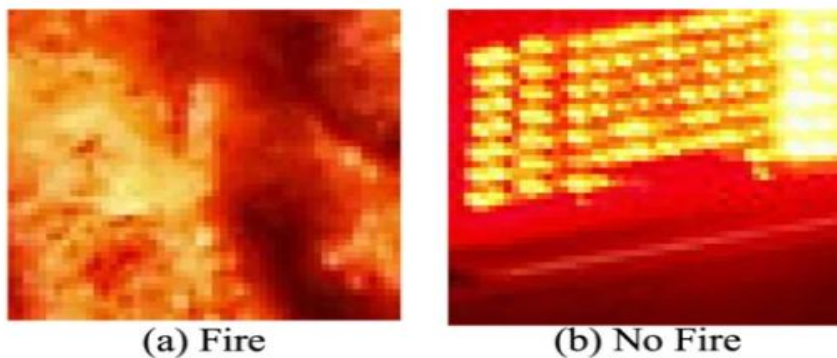


Εικόνα 2-11 Στιγμιότυπα από το Foggiadataset

11) BoWFire

Το BoWFire παρουσιάστηκε στο [5] το 2015. Ο βασικός στόχος του dataset ήταν η ανάπτυξη και αξιολόγηση αλγορίθμων ανίχνευσης φωτιάς σε στατικές εικόνες με έμφαση στη μείωση των ψευδών θετικών ανιχνεύσεων που προκαλούνται από αντικείμενα με χρώματα παρόμοια με αυτά της φλόγας. Ήταν από τα πρώτα datasets που έδειξαν ότι η χρωματική πληροφορία από μόνη της δεν αρκεί για την αξιόπιστη ανίχνευση φωτιάς. Η ενσωμάτωση χαρακτηριστικών υφής (texture features) αποτέλεσε μία από τις πρώτες επιτυχημένες προσπάθειες μείωσης των ψευδών συναγερμών. Το BoWFire σχεδιάστηκε πρωτίστως για fire segmentation και pixel-level fire detection. Το σύνολο περιλαμβάνει 226 εικόνες διαφορετικών αναλύσεων, εκ των οποίων 119 περιέχουν φωτιά και 107 δεν περιέχουν φωτιά. Οι εικόνες παρουσιάζουν μεγάλη ποικιλία σε περιβάλλοντα

και συνθήκες λήψης, όπως δασικές πυρκαγιές, πυρκαγιές σε κτίρια, βιομηχανικά σκηνικά, τροχαία ατυχήματα. Παράλληλα περιλαμβάνονται παραδείγματα εικόνων όπως ηλιοβασιλέματα, φωτεινές επιγραφές και αντικείμενα με κόκκινες ή πορτοκαλί αποχρώσεις. Οι εικόνες είναι RGB προερχόμενες κατά το μεγαλύτερο μέρος τους από επίγειες λήψεις. Η συλλογή των εικόνων πραγματοποιήθηκε από δημόσια διαθέσιμες διαδικτυακές πηγές και εικόνες πραγματικών περιστατικών έκτακτης ανάγκης. Το BoWFire παρέχει χειροκίνητα επισημασμένες περιοχές φωτιάς σε επίπεδο pixel, επιτρέποντας την αξιολόγηση αλγορίθμων segmentation και όχι μόνο classification. Δεν παρέχονται bounding boxes, αλλά διατίθενται δυαδικές μάσκες όπου τα fire pixels επισημαίνονται ξεχωριστά από το υπόβαθρο. Η ύπαρξη των масκών κάνει το dataset ιδιαίτερα χρήσιμο για την αξιολόγηση semantic segmentation μοντέλων.



Εικόνα 2-12 Στιγμιότυπα από το BoWFire dataset

12) FiSmo

Το FiSmo δεν είναι ένα μεμονωμένο dataset, αλλά μία συλλογή πολλών επιμέρους datasets για φωτιά και καπνό που δημιουργήθηκε από το Πανεπιστήμιο του São Paulo με στόχο να παρέχει ένα κοινό αποθετήριο για διαφορετικά fire/smoke tasks. Παρουσιάστηκε το 2017 στην εργασία [3]. Η συλλογή περιλαμβάνει συνολικά 4 image datasets και 2 video datasets τα οποία καλύπτουν διαφορετικές εργασίες υπολογιστικής όρασης σχετικές με καταστάσεις έκτακτης ανάγκης. Ένα από τα σημαντικότερα χαρακτηριστικά του FiSmo είναι ότι περιλαμβάνει τόσο εικόνες RGB όσο και βίντεο, επιτρέποντας την αξιολόγηση αλγορίθμων σε διαφορετικά επίπεδα πολυπλοκότητας. Τα

δεδομένα προέρχονται από πραγματικά περιστατικά έκτακτης ανάγκης και περιλαμβάνουν φωτιά, καπνό και εκρήξεις σε αστικά και βιομηχανικά περιβάλλοντα. Το FiSmo έχει ιδιαίτερη αξία επειδή συγκεντρώνει σε μία ενιαία πλατφόρμα δεδομένα για πολλές διαφορετικές εργασίες, επιτρέποντας τη σύγκριση μεθόδων classification, segmentation, retrieval και video analysis.



Εικόνα 2-13 Στιγμιότυπα από το FiSmodataset

13) VisiFire

Το VisiFire αναπτύχθηκε με στόχο την αξιολόγηση συστημάτων έγκαιρης ανίχνευσης φωτιάς και καπνού μέσω ανάλυσης βίντεο. Σε αντίθεση με datasets όπως το FLAME ή το DeepFire που βασίζονται κυρίως σε στατικές εικόνες, το VisiFire επικεντρώνεται στην εκμετάλλευση της χρονικής πληροφορίας (temporal information) που περιέχεται σε ακολουθίες βίντεο. Για τον λόγο αυτό χρησιμοποιείται ευρέως στην έρευνα για CNN-LSTM, 3D CNNs, ConvLSTM, Vision Transformers για βίντεο και γενικότερα σε μεθόδους που αξιοποιούν τη δυναμική συμπεριφορά της φλόγας και του καπνού. Το dataset αποτελείται από 57 βίντεο χωρισμένα σε τέσσερις βασικές κατηγορίες: 13 videos με φλόγα, 21 videos με καπνό, 21 videos με δασικό καπνό και 2 videos άλλων φαινομένων. Οι λήψεις προέρχονται κατά κύριο λόγο από RGB βίντεο πραγματικών περιστατικών, καθώς και από κάμερες επιτήρησης και από διαδικτυακές πηγές. Οι προοπτικές λήψεις είναι κυρίως επόψεις, αν και ορισμένα βίντεο παρουσιάζουν πιο απομακρυσμένες λήψεις δασικών περιοχών. Το dataset περιλαμβάνει τόσο αστικά όσο και φυσικά περιβάλλοντα, με διαφορετικές αποστάσεις από τη φωτιά, ποικίλες καιρικές συνθήκες και διαφορετικό βαθμό ορατότητας του καπνού. Αρχικά το VisiFire παρείχε

μόνο labels σε επίπεδο βίντεο (π.χ. fire/smoke). Ωστόσο, σε μεταγενέστερες εργασίες δημιουργήθηκαν pixel-level segmentation annotations για ένα μέρος του dataset.



Εικόνα 2-14 Στιγμιότυπα από το VisiFiredataset

14) Forestry Images

Το Forestry Images dataset είναι ένα από τα datasets που εμφανίζονται συχνά στη βιβλιογραφία της ταξινόμησης δασικών πυρκαγιών, ιδιαίτερα σε εργασίες transfer learning και CNN-based classification. Το Forestry Images επικεντρώνεται κυρίως στο πρόβλημα της δυαδικής ταξινόμησης εικόνων (binary image classification). Το σύνολο περιλαμβάνει περίπου 5.000 εικόνες, κατανεμημένες στις δύο βασικές κατηγορίες Fire και Non-Fire. Οι εικόνες συλλέχθηκαν από διαδικτυακές πηγές και δημόσια διαθέσιμα αποθετήρια εικόνων, με στόχο να καλυφθεί μια μεγάλη ποικιλία δασικών τοπίων διαφορετικών εποχών και με διαφορετικές συνθήκες φωτισμού, καθώς και εικόνες με διαφορετικά μεγέθη πυρκαγιών. Οι δημιουργοί έδωσαν ιδιαίτερη έμφαση στη συμπερίληψη δύσκολων αρνητικών παραδειγμάτων, όπως ηλιοβασιλέματα, σύννεφα, ομίχλη και φωτεινές αντανάκλασεις, με στόχο να μειωθεί ο κίνδυνος ψευδών συναγερμών. Οι εικόνες είναι RGB και το dataset περιλαμβάνει τόσο επίγειες λήψεις από φωτογραφικές μηχανές και κινητά τηλέφωνα όσο και εναέριες λήψεις που προέρχονται από drones ή άλλες πλατφόρμες. Οι σκηνές αφορούν κυρίως δασικά και αγροδασικά περιβάλλοντα, με παρουσία φλόγας, καπνού ή συνδυασμού των δύο.



Εικόνα 2-15 Στιγμιότυπα από το ForestryImagesdataset

Κεφάλαιο 3. Πειραματική Αξιολόγηση

Στο παρόν κεφάλαιο παρουσιάζουμε έξι μοντέλα βαθιάς μάθησης που επιλέξαμε για τη συγκριτική αξιολόγηση της ανίχνευσης δασικών πυρκαγιών από εναέριες εικόνες, καθώς και τα τρία σύνολα δεδομένων που χρησιμοποιήθηκαν κατά την πειραματική διαδικασία. Η επιλογή των μοντέλων πραγματοποιήθηκε με στόχο να καλυφθεί ένα ευρύ φάσμα σύγχρονων αρχιτεκτονικών υπολογιστικής όρασης, συμπεριλαμβανομένων αμιγώς μετασχηματιστικών μοντέλων (Vision Transformers), συνελκτικών νευρωνικών δικτύων (CNN), υβριδικών αρχιτεκτονικών καθώς και μοντέλων που έχουν σχεδιαστεί αρχικά για εργασίες ανίχνευσης αντικειμένων ή σημασιολογικής τμηματοποίησης. Με τον τρόπο αυτό καθίσταται δυνατή η διερεύνηση της συμπεριφοράς διαφορετικών αρχιτεκτονικών προσεγγίσεων στο ίδιο πρόβλημα δυαδικής ταξινόμησης εικόνων πυρκαγιάς.

Για κάθε μοντέλο πραγματοποιούμε αρχικά αξιολόγηση της προεκπαιδευμένης (pretrained) έκδοσής του (παράγραφος 3.1) και στη συνέχεια εφαρμόζεται διαδικασία fine-tuning χρησιμοποιώντας κοινές υπερπαραμέτρους εκπαίδευσης (παράγραφος 3.2). Οι κοινές αυτές υπερπαραμέτροι εκπαίδευσης του fine-tuning των μοντέλων συγκεντρώνονται στον ακόλουθο πίνακα.

Optimizer	Adam
Learning Rate	$5 \cdot 10^{-5}$
Batch Size	32
Epochs	3

Η σύγκριση των δύο εκδόσεων θα μας επιτρέψει ακολούθως την ποσοτική και ποιοτική διερεύνηση της επίδρασης του fine tuning στην απόδοση κάθε αρχιτεκτονικής (βλ. Κεφάλαιο 4). Η αξιολόγηση βασίζεται σε ποσοτικές μετρικές (Accuracy, Precision, Recall και F1-score), στους Confusion Matrices, στην ανάλυση του χώρου χαρακτηριστικών μέσω διαγραμμάτων t-SNE καθώς και στην ερμηνεία των αποφάσεων των μοντέλων με χρήση της τεχνικής Grad-CAM. Η συνδυαστική αυτή προσέγγιση μας επιτρέπει την εξαγωγή ασφαλέστερων

συμπερασμάτων σχετικά με την αποτελεσματικότητα κάθε μοντέλου και τον τρόπο με τον οποίο το fine-tuning επηρεάζει τις εσωτερικές αναπαραστάσεις και τη διαδικασία λήψης αποφάσεων.

Το SigLIP2 (Sigmoid Loss for Language-Image Pretraining 2) αποτελεί ένα σύγχρονο πολυτροπικό μοντέλο όρασης και γλώσσας που αναπτύχθηκε από την Google ως εξέλιξη της οικογένειας SigLIP. Βασίζεται στην αρχιτεκτονική Vision Transformer (ViT) για την εξαγωγή οπτικών χαρακτηριστικών και έχει εκπαιδευτεί με στόχο την ευθυγράμμιση εικόνων και φυσικής γλώσσας μέσω αντίθετης (contrastive) μάθησης χρησιμοποιώντας sigmoid loss αντί της κλασικής softmax loss. Η προεκπαίδευση του μοντέλου πραγματοποιήθηκε σε πολύ μεγάλης κλίμακας ζεύγη εικόνων–κειμένου που συλλέχθηκαν από το διαδίκτυο, επιτρέποντας στο μοντέλο να μάθει γενικεύσιμες οπτικές αναπαραστάσεις. Αν και δεν έχει σχεδιαστεί ειδικά για την ανίχνευση πυρκαγιών, η υψηλή ποιότητα των χαρακτηριστικών που εξάγει το καθιστά ιδιαίτερα κατάλληλο για εργασίες ταξινόμησης εικόνων μέσω διαδικασίας fine-tuning.

Το SegFormer-B0 αποτελεί τη μικρότερη έκδοση της οικογένειας SegFormer, μιας αρχιτεκτονικής Transformer που αναπτύχθηκε από τη NVIDIA για εργασίες σημασιολογικής τμηματοποίησης (semantic segmentation). Η αρχιτεκτονική συνδυάζει έναν ιεραρχικό Transformer encoder με έναν ελαφρύ αποκωδικοποιητή πολλαπλών επιπέδων, επιτυγχάνοντας υψηλή ακρίβεια με σχετικά μικρό υπολογιστικό κόστος. Στην παρούσα εργασία χρησιμοποιήθηκε η έκδοση millionhz/segformer-b0-finetuned-flame, η οποία έχει ήδη προσαρμοστεί στο σύνολο δεδομένων FLAME για την ανίχνευση περιοχών πυρκαγιάς. Παρότι η αρχική σχεδίαση του μοντέλου αφορά την τμηματοποίηση εικόνων, στην εργασία αξιοποιείται ως δυαδικός ταξινομητής εικόνων μετά από κατάλληλη προσαρμογή.

Το YoloV26 ανήκει στην οικογένεια των μοντέλων You Only Look Once (YOLO) και έχει σχεδιαστεί πρωτίστως για εργασίες ανίχνευσης αντικειμένων (object detection) σε πραγματικό χρόνο. Η συγκεκριμένη έκδοση (SalahALHaismawi/yolov26-fire-detection) που θα χρησιμοποιήσουμε έχει

προεκπαιδευτεί ειδικά για την ανίχνευση πυρκαγιών σε εικόνες και αποσκοπεί στην ταυτόχρονη ανίχνευση της πυρκαγιάς και στον εντοπισμό της θέσης των εστιών της πυρκαγιάς. Παρά το γεγονός ότι η αρχική του λειτουργία αφορά object detection, στην παρούσα εργασία προσαρμόζεται ώστε να χρησιμοποιηθεί ως μοντέλο δυαδικής ταξινόμησης εικόνων, επιτρέποντας τη σύγκρισή του με τις υπόλοιπες αρχιτεκτονικές.

Το EdBianchi ViT βασίζεται στην αρχιτεκτονική Vision Transformer (ViT) και έχει αναπτυχθεί ειδικά για εφαρμογές ανίχνευσης πυρκαγιάς. Το μοντέλο αξιοποιεί τον μηχανισμό αυτοπροσοχής (self-attention) των Transformers ώστε να συλλαμβάνει μακρινές χωρικές εξαρτήσεις μέσα στην εικόνα, γεγονός που επιτρέπει την αποτελεσματική αναγνώριση των οπτικών χαρακτηριστικών που σχετίζονται με την παρουσία φλόγας και καπνού. Η προεκπαιδευμένη έκδοση έχει εκπαιδευτεί σε σύνολα δεδομένων εικόνων πυρκαγιάς και προορίζεται για εργασίες δυαδικής ταξινόμησης.

Το Fire-Chad Detector αποτελεί ένα υβριδικό μοντέλο που βασίζεται σε ensemble αρχιτεκτονική CNN. Συνδυάζει τρία ανεξάρτητα συνελκτικά νευρωνικά δίκτυα (DenseNet121, ResNet50 και Xception), των οποίων τα χαρακτηριστικά συγχωνεύονται σε κοινό επίπεδο ταξινόμησης. Η προσέγγιση αυτή επιτρέπει την αξιοποίηση των πλεονεκτημάτων διαφορετικών CNN αρχιτεκτονικών, οδηγώντας σε περισσότερο ανθεκτικές και γενικεύσιμες αναπαραστάσεις χαρακτηριστικών. Το μοντέλο έχει αναπτυχθεί ειδικά για την αναγνώριση πυρκαγιών από εικόνες και η προεκπαίδευσή του έχει γίνει στο σύνολο FLAME.

Το ConvnexT-Tiny αποτελεί μία σύγχρονη αρχιτεκτονική συνελκτικών νευρωνικών δικτύων (CNN) που αναπτύχθηκε από τη Meta ως επανασχεδιασμός των κλασικών ResNet, ενσωματώνοντας σχεδιαστικές αρχές που προέκυψαν από την εξέλιξη των Vision Transformers. Παρότι βασίζεται αποκλειστικά σε συνελκτικά φίλτρα, επιτυγχάνει επιδόσεις συγκρίσιμες με σύγχρονους Transformers σε πλήθος εργασιών υπολογιστικής όρασης. Η προεκπαιδευμένη έκδοση έχει εκπαιδευτεί στο σύνολο δεδομένων ImageNet-1K.

Στον πίνακα που ακολουθεί συγκεντρώνουμε τα βασικά στοιχεία για τα μοντέλα που χρησιμοποιούμε στις πειραματικές μας δοκιμές.

Πίνακας 3-1 Πληροφορίες για τα Μοντέλα

Μοντέλο	Τύπος	Αρχιτεκτονική	Εργασία για την οποία σχεδιάστηκε	Σύνολο Προεκπαίδευσης
SigLIP2	Vision-Language	Vision Transformer	Image Classification/ Vision-Language	Web image-text pairs
SegFormer-B0	Transformer	Hierarchical Transformer	Semantic Segmentation	FLAME
Yolov26	CNN Detector	YOLO	Object Detection	Fire Datasets
EdBianchi Vit	Transformer	ViT	Fire Classification	Web Fire Images
Superflash 41	CNN Ensemble	DenseNet+ResNet+Xception	Fire Classification	Fire Datasets
Convnext	CNN	ConvNext	Image Classification	ImageNet-1K

Προκειμένου για τα τρία σύνολα που θα χρησιμοποιήσουμε, μπορούμε σε αδρές γραμμές να σημειώσουμε τα ακόλουθα στοιχεία.

Το σύνολο δεδομένων DFire είναι ένα μεγάλης κλίμακας σύνολο δεδομένων εικόνων (dataset), ειδικά σχεδιασμένο για την εκπαίδευση και αξιολόγηση αλγορίθμων μηχανικής μάθησης και υπολογιστικής όρασης στην ανίχνευση φωτιάς και καπνού. Περιλαμβάνει 21.527 εικόνες μορφής RGB με ανάλυση στα 416×416 pixels. Οι εικόνες που περιλαμβάνει το DFire διακρίνονται σε 4 κατηγορίες. «Μόνο φωτιά» (OnlyFire) με 1.164 εικόνες, «Μόνο καπνός» (Only smoke) με 5.867 εικόνες, την κατηγορία «Φωτιά και καπνός» (Fire and smoke) με 4.658 εικόνες και τέλος την κατηγορία «Κανένα» (None) με 9.838 εικόνες (οι

εικόνες της κατηγορίας “None” λειτουργούν ως αρνητικά δείγματα για την αποφυγή ψευδών συναγερμών). Αξίζει να σημειωθεί πως ανάμεσα στις εικόνες του συνόλου δεδομένων περιλαμβάνονται ρεαλιστικές και δύσκολες συνθήκες (όπως π.χ. παρουσία σύννεφων ή ομίχλης), βοηθώντας έτσι το μοντέλο να ξεχωρίζει τη φωτιά από αντικείμενα ή φαινόμενα που της μοιάζουν στο χρώμα και τη φωτεινότητα.

Στο σύνολο δεδομένων FLAME αναφερθήκαμε στην προηγούμενη παράγραφο (βλ. 2.2 σημείο (4)). Στα πλαίσια της πειραματικής μας εργασίας χρησιμοποιούμε εικόνες από το έβδομο και όγδοο αποθετήριο του FLAME1. Συγκεκριμένα χρησιμοποιούμε 47.992 εικόνες που συγκροτούν το training και test set. Από αυτές 30.155 είναι εικόνες με φωτιά ενώ οι υπόλοιπες 17.837 είναι εικόνες χωρίς φωτιά. Το σύνολο δεδομένων FASDD (Flame and Smoke Detection Dataset) είναι ένα από τα μεγαλύτερα, πιο ολοκληρωμένα και πολυσυλλεκτικά open-access σύνολα δεδομένων για την εκπαίδευση μοντέλων βαθιάς μάθησης (Deep Learning) στην ανίχνευση φλόγας και καπνού. Το FASDD έχει σχεδιαστεί για τη δημιουργία συστημάτων επιτήρησης σε δίκτυα που συνδυάζουν επίγεια μέσα, drones και δορυφόρους. Περιλαμβάνει πάνω από 122.000 εικόνες και διαθέτει περισσότερα από 140.000 χειροκίνητα οριοθετημένα αντικείμενα (περισσότερες από 82.000 «φλόγες» και 57.000 «καπνούς»). Το FASDD παρουσιάζει μια πολυσχιδή δομή εικόνων τα οποία προέρχονται τόσο από επίγεια μέσα όσο και από εικόνες drones και δορυφορικά δεδομένα. Ως επακόλουθο επιτρέπει στους ερευνητές να εκπαιδεύσουν μοντέλα που δεν περιορίζονται μόνο σε μία κάμερα, αλλά μπορούν να αναγνωρίσουν μια πυρκαγιά είτε αυτή φαίνεται από μια κάμερα δρόμου, είτε από ένα drone που πετάει, είτε από έναν δορυφόρο. Στο σημείο αυτό θα πρέπει να σημειώσουμε ότι στην παρούσα εργασία Στην παρούσα εργασία δεν χρησιμοποιήθηκε το πλήρες σύνολο δεδομένων FASDD, αλλά ένα αντιπροσωπευτικό υποσύνολό του, το οποίο περιλαμβάνει συνολικά 651 εικόνες, εκ των οποίων 110 ανήκουν στην κατηγορία Fire και 541 στην κατηγορία No Fire. Η επιλογή μας αυτή πραγματοποιήθηκε για λόγους διαχείρισης υπολογιστικών πόρων ώστε να τηρηθεί ένας λογικός χρόνος εκπαίδευσης για όλα τα εξεταζόμενα μοντέλα. Πρέπει να σημειώσουμε πως παρατηρούμενη ανισορροπία μεταξύ των

δύο κλάσεων στο υποσύνολο του FASDD, καθιστά την αξιολόγηση ιδιαίτερα απαιτητική, καθώς μετρικές όπως η Accuracy ενδέχεται να οδηγήσουν σε παραπλανητικά συμπεράσματα όταν εξετάζονται μεμονωμένα. Για τον λόγο αυτό, στην παρούσα εργασία η αξιολόγηση των μοντέλων βασίζεται στον συνδυασμό των μετρικών Precision, Recall και F1-score, καθώς και στους πίνακες σύγχυσης, ώστε να αποτυπωθεί με μεγαλύτερη αξιοπιστία η πραγματική ικανότητα των μοντέλων να ανιχνεύουν εικόνες πυρκαγιάς. Εξάλλου πιστεύουμε ότι η μελέτη του τρόπου που τα μοντέλα θα συμπεριφερθούν στην εργασία τους για ανίχνευση πυρκαγιάς κινούμενα στα πλαίσια ενός τέτοιου «προκαταλημμένου υπέρ των εικόνων που δεν έχουν φωτιά» συνόλου, αποτελεί μια ενδιαφέρουσα ερευνητική πρόκληση.

Στον πίνακα που ακολουθεί συγκεντρώνουμε τα δεδομένα για τα σύνολα μας.

Πίνακας 3-2 Πληροφορίες για τα Datasets

Σύνολο Δεδομένων	Training Set	Fire	No Fire	Test set	Ισορροπία Κλάσεων	Προέλευση εικόνων	Άλλα στοιχεία
FASDD (υποσύνολο)	520	110	541	131	Ανισορροπία κλάσεων	Επίγειες εικόνες πυρκαγιών	
FLAME	39.375	30.155	17.837	8.617	Μέτριος βαθμός ισορροπίας	Εικόνες από drones	Διαθέτει Μάσκες για segmentation tasks
DFire	21.527	11.000	10.527	4.306	Ισορροπημένο	Πολλαπλές πηγές	Μεγάλη ποικιλία υπόβαθρου και φωτισμού

Σε σχέση δε με τον ορισμό της θετικής κλάσης, τη διαχείριση του καπνού καθώς και τον εκάστοτε κανόνα θετικής πρόβλεψης που χρησιμοποιείται στα πειράματα, ο ακόλουθος πίνακας συγκεντρώνει τις σχετικές πληροφορίες για τα διάφορα μοντέλα στην pretrained και fine tuned εκδοχή τους.

Μοντέλο	Pretrained θετική πρόβλεψη	Fine tuned θετική πρόβλεψη	Διαχείριση εικόνων καπνού
EdBianchi	Πιθανότητα της κλάσης Fire $\geq 0,5$	Επιλογή της εξόδου Fire από τον δυαδικό ταξινομητή	Στην pretrained εκδοχή η πιθανότητα Smoke δεν συνυπολογίζεται. Στη fine-tuned εκδοχή η διαχείριση καθορίζεται από τις ετικέτες εκπαίδευσης.
SigLip2	Μεγαλύτερη ομοιότητα με την περιγραφή «a photo of fire» έναντι της «a photo without fire»	Επιλογή της θετικής κλάσης από τον δυαδικό ταξινομητή	Δεν υπάρχει ξεχωριστή περιγραφή Smoke στην pretrained αξιολόγηση. Μετά το fine-tuning, η διαχείριση εξαρτάται από τις ετικέτες εκπαίδευσης.
Superflash41	Έξοδος μοντέλου $\geq 0,5$, η οποία αντιμετωπίζεται από τον κώδικα ως Fire	Έξοδος μοντέλου $\geq 0,5$, με θετική ετικέτα εκπαίδευσης τη Fire	Δεν εφαρμόζεται ξεχωριστός κανόνας για Smoke. Η πραγματική ετικέτα προκύπτει από τις επισημειώσεις ή τον φάκελο προέλευσης.
ConvNext	Επιλογή εξόδου από τυχαία αρχικοποιημένη δυαδική κεφαλή, χωρίς εκπαίδευσή της	Επιλογή της αντίστοιχης κλάσης από τον εκπαιδευμένο δυαδικό ταξινομητή	Δεν υπάρχει ξεχωριστή έξοδος Smoke. Στη fine-tuned εκδοχή η διαχείριση καθορίζεται από τις ετικέτες εκπαίδευσης.

Segformer-B0	Τουλάχιστον 50 pixels στην κατηγορία που ο κώδικας αντιμετωπίζει ως Fire	DFire / FASDD: δυαδικός ταξινομητής. FLAME: πρόβλεψη μέσω τμηματοποίησης με κατώφλι 50 pixel	Δεν συνυπολογίζεται ξεχωριστή πρόβλεψη Smoke. Η πραγματική ετικέτα εξαρτάται από το σύνολο και το υποσύνολο αξιολόγησης.
YOLOv26	DFire / FASDD: οποιοδήποτε bounding box με confidence $\geq 0,25$. FLAME: μόνο bounding box Fire	DFire / FASDD: οποιοδήποτε bounding box με confidence $\geq 0,25$. FLAME: μόνο bounding box Fire	Ο αρχικός κανόνας δέχεται ανιχνεύσεις Smoke και, στην pretrained εκδοχή, Other. Ο διορθωμένος κανόνας του FLAME δέχεται αποκλειστικά Fire.

Κατά την παρουσίαση των αποτελεσμάτων, οι μετρικές Precision, Recall και F1 ερμηνεύονται ως μετρικές της σημαίνουσας θετικής κατηγορίας (ετικέτα 1). Στα κελιά δυαδικής αξιολόγησης, η ετικέτα 1 αντιστοιχεί στη θετική κατηγορία που ονομάζεται Fire. Ειδικά στο DFire, όμως, η κατηγορία αυτή περιλαμβάνει και εικόνες μόνο με καπνό, καθώς οι ετικέτες δημιουργήθηκαν από την ύπαρξη μη κενού αρχείου επισημειώσεων.

3.1 Αξιολόγηση του προεκπαιδευμένου SegFormer-B0

Αρχικά θα ασχοληθούμε με το μοντέλο “Segformer-B0-finetuned-flame” (σε ότι ακολουθεί θα αναφερόμαστε σε αυτό ως SegFormer-B0). Πρόκειται για ένα μοντέλο που εκτελεί semanticsegmentation οπότε και χρειάζεται «μάσκες» (masks) σε επίπεδο pixel για να κάνει την εκτίμησή του. Τέτοιες μάσκες διατίθενται στο σύνολο δεδομένων FLAME. Από την άλλη μεριά επειδή τα άλλα δύο σύνολα δεδομένων δεν διαθέτουν αυτές τις μάσκες, χρειάστηκε μια προσαρμογή του μοντέλου SegFormer-B0 ώστε να εκτελέσει δυαδική αναγνώριση εικόνας (φωτιά ή μη-φωτιά) στο DFire και στο FASDD. Στην

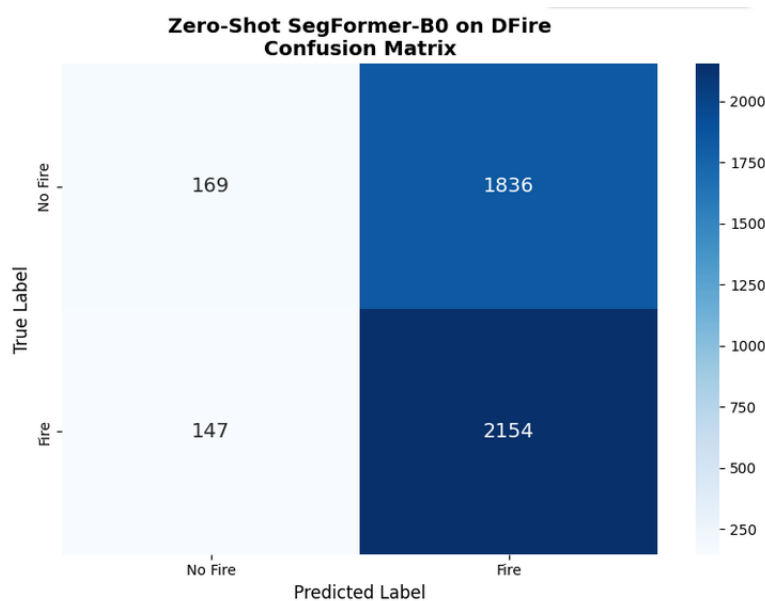
παρούσα παράγραφο θα δούμε τη συμπεριφορά του μοντέλου SegFormer-B0 στα τρία σύνολα δεδομένων DFire, FLAME και FASDD αντίστοιχα, χωρίς να προβούμε σε περαιτέρω εκπαίδευση του μοντέλου.

3.1.1 στο DFire

Αρχικά θα χρησιμοποιήσουμε το προεκπαιδευμένο μοντέλο SegFormer-B0 (Pretrained ή Zero-Shot SegFormer-B0) στο σύνολο δεδομένων DFire. Η απόδοση του μοντέλου καταγράφεται στον ακόλουθο πίνακα

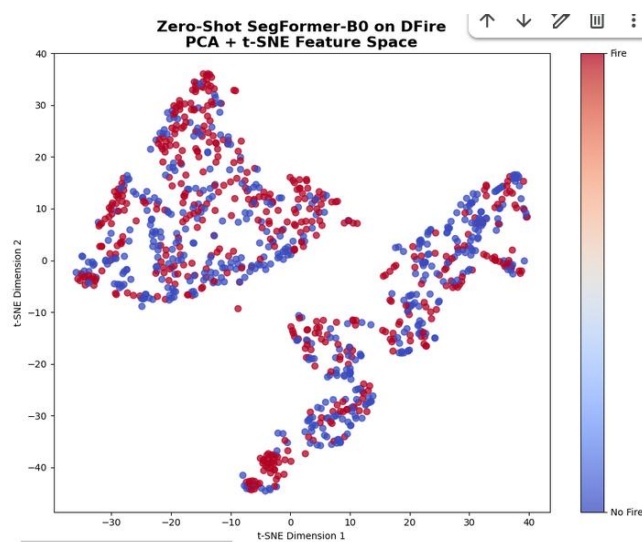
Πίνακας 3-3 Μετρικές του pretrained SegFormer στο DFire

Pretrained SegFormer-B0 στο DFire	
Accuracy	0.5395
Precision	0.5398
Recall	0.9361
F1-score	0.6848



Εικόνα 3-1 Confusion Matrix του pretrained SegFormer στο DFire

Το προεκπαιδευμένο SegFormer-B0 καταγράφει Accuracy 53,95%, Precision 53,98%, Recall 93,61% και F1-score 68,48%. Το Recall είναι υψηλό, όμως οι υπόλοιπες μετρικές παραμένουν σε σχετικά χαμηλά επίπεδα, άρα το μοντέλο δεν επιτυγχάνει ισορροπημένη απόδοση στο σύνολο δεδομένων DFire. Η εικόνα συμπληρώνεται από την ανάγνωση του confusion matrix. Από τα 2.301 πραγματικά δείγματα Fire, το μοντέλο αναγνώρισε σωστά τα 2.154 και αστόχησε σε μόλις 147 περιπτώσεις (False Negatives), γεγονός που εξηγεί το πολύ υψηλό Recall. Ωστόσο, από τα 2.005 δείγματα No Fire, μόνο 169 ταξινομήθηκαν σωστά, ενώ τα 1.836 χαρακτηρίστηκαν λανθασμένα ως Fire (False Positives). Επομένως, το μοντέλο τείνει να προβλέπει την παρουσία φωτιάς σχεδόν σε κάθε εικόνα, οδηγώντας σε μεγάλο αριθμό ψευδώς θετικών προβλέψεων και, κατά συνέπεια, σε χαμηλή ακρίβεια και μέτρια συνολική επίδοση. Θα μπορούσαμε να πούμε συμπερασματικά ότι το SegFormer-B0 διαθέτει υψηλή ευαισθησία στην ανίχνευση πυρκαγιάς, όμως δεν γενικεύει ικανοποιητικά στο σύνολο δεδομένων DFire χωρίς fine-tuning.

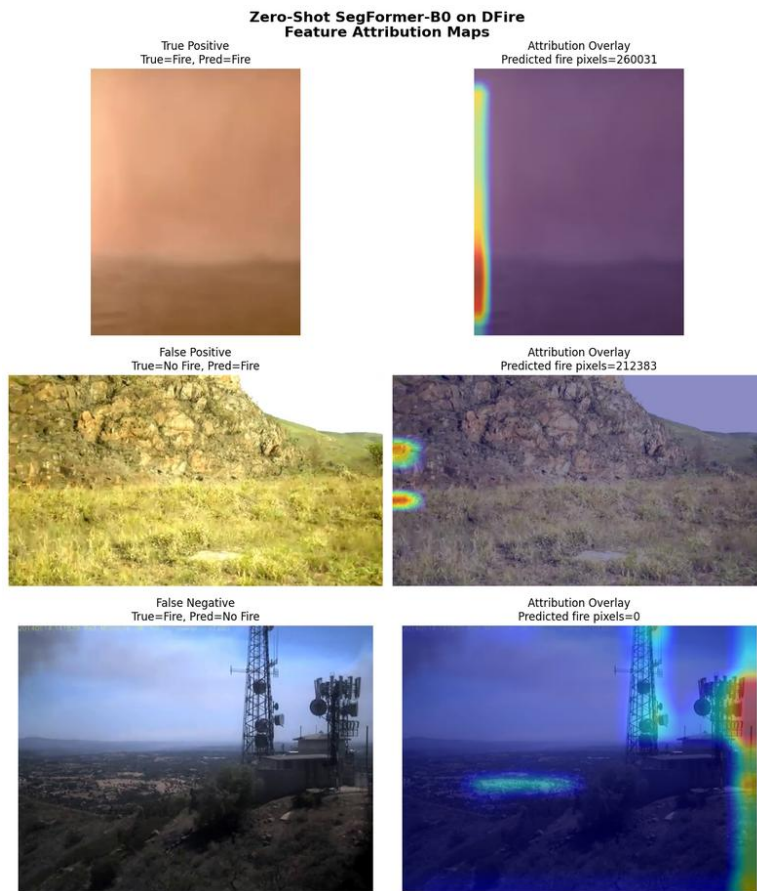


Εικόνα3-2 t-SNE του pretrained SegFormer στο DFire

Η απεικόνιση t-SNE δείχνει ότι τα χαρακτηριστικά που εξάγει το προεκπαιδευμένο SegFormer-B0 δεν οδηγούν σε σαφή διαχωρισμό μεταξύ των κλάσεων Fire και No Fire. Τα σημεία των δύο κατηγοριών εμφανίζουν έντονη

αλληλοεπικάλυψη σχεδόν σε όλες τις τοπικές συστάδες, χωρίς να σχηματίζονται διακριτές περιοχές που να αντιστοιχούν αποκλειστικά σε μία από τις δύο κλάσεις. Η έντονη αυτή επικάλυψη είναι συνεπής με τα αποτελέσματα της αξιολόγησης και του confusionmatrix. Παρότι το μοντέλο εμφανίζει πολύ υψηλό Recall, τείνει να ταξινομεί μεγάλο αριθμό εικόνων ως Fire, γεγονός που οδηγεί σε πλήθος ψευδώς θετικών προβλέψεων.

Οι χάρτες Grad-CAM που φαίνονται στη σειρά των εικόνων που ακολουθούν μας δίνουν μια οπτική ερμηνεία των περιοχών της εικόνας στις οποίες βασίστηκε το προεκπαιδευμένο SegFormer-B0 για τη λήψη των αποφάσεών του. Στην περίπτωση του True Positive, το μοντέλο ενεργοποιείται κυρίως σε περιοχές με έντονα χαρακτηριστικά καπνού. Στο False Positive παρατηρείται ενεργοποίηση σε περιοχές του βραχώδους ανάγλυφου που βρίσκονται στα όρια της εικόνας και οι οποίες αυτές περιοχές δεν σχετίζονται με φωτιά. Στην περίπτωση του False Negative, το μοντέλο εστιάζει κυρίως στην κεραία και στο δεξιό άκρο της εικόνας, ενώ αγνοεί τις περιοχές όπου εμφανίζονται τα πραγματικά ίχνη καπνού. Η συμπεριφορά αυτή υπαινίσσεται την αδυναμία του μοντέλου να αναγνωρίζει μοτίβα πυρκαγιάς. Ένα στοιχείο που αξίζει να αναφέρουμε είναι πως και στις τρεις περιπτώσεις υπάρχουν πολύ έντονες ενεργοποιήσεις στα όρια της εικόνας.



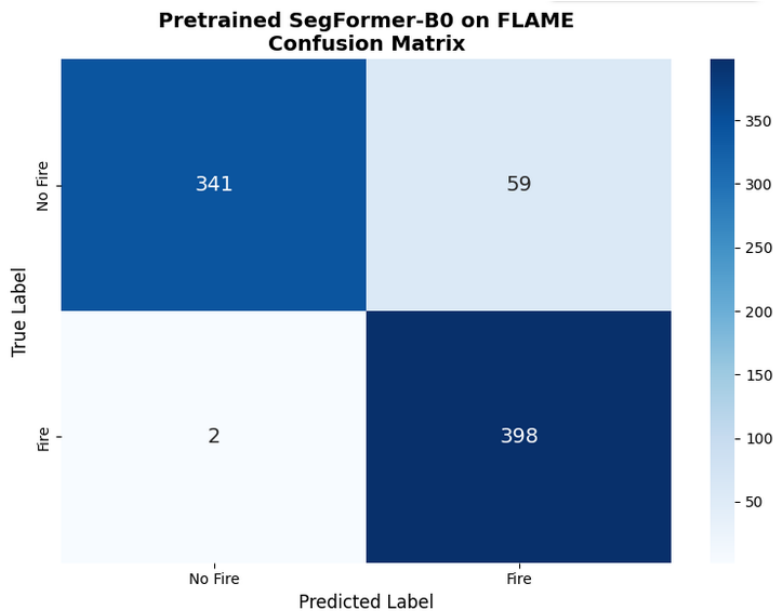
Εικόνα3-3Χάρτες Grad-CAM γιατο pretrained SegFormer στο DFire

3.1.2 στο FLAME

Οι τιμές των μετρικών που αναφέρονται στην απόδοση του μοντέλου ως προς την εργασία της ταξινόμησης εικόνας είναι οι παρακάτω

Πίνακας 3-4 Μετρικέςτου pretrained SegFormer στο FLAME

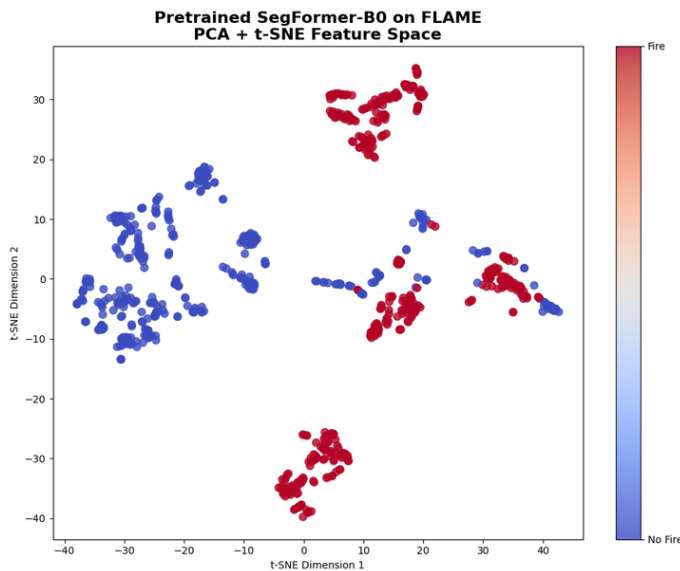
Pretrained SegFormer-B0 στο FLAME	
Accuracy	0.9237
Precision	0.8709
Recall	0.9950
F1-score	0.9288



Εικόνα3-4Confusion Matrix του pretrained SegFormer στοFLAME

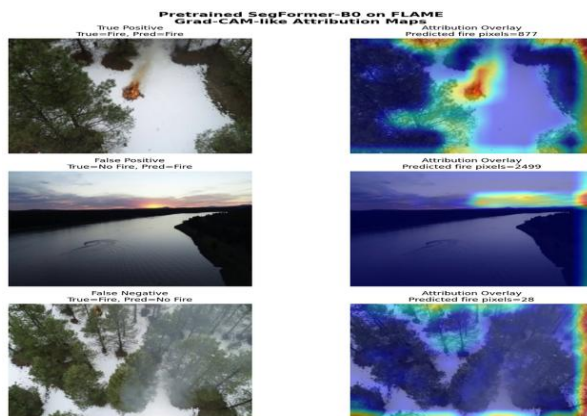
Παρατηρούμε ότι το προεκπαιδευμένο SegFormer-B0 πέτυχεσυνολικά πολύ υψηλή απόδοση στο σύνολο δεδομένων FLAME. Η σχεδόν τέλεια τιμή του Recall (99,50%) μας δηλώνει ότι το μοντέλο αναγνωρίζει σχεδόν όλες τις εικόνες που περιέχουν πυρκαγιά, ενώ οι υψηλές τιμές του Accuracy και του F1-score επιβεβαιώνουν την αποτελεσματική ισορροπία μεταξύ ανίχνευσης και συνολικής ταξινόμησης. Από τον confusionmatrixβλέπουμε ότι από τις 400εικόνες με πυρκαγιά, το μοντέλο ταξινόμησε σωστά τις 398, ενώ μόνο 2 εικόνες χαρακτηρίστηκαν λανθασμένα ως No Fire (False Negatives). Αντίστοιχα, από τις 400 εικόνες χωρίς πυρκαγιά, οι 341 ταξινομήθηκαν σωστά, ενώ 59 χαρακτηρίστηκαν λανθασμένα ως Fire (False Positives). Επομένως, το μοντέλο εμφανίζει εξαιρετικά υψηλή ευαισθησία στην ανίχνευση πυρκαγιάς, με το κύριο σφάλμα του να αφορά έναν σχετικά περιορισμένο αριθμό ψευδώς θετικών προβλέψεων. Αξίζει βέβαια να αναφέρουμε ότι τα πολύ καλά αυτά αποτελέσματα ήταν εν μέρει αναμενόμενα καθώς η προεκπαίδευση του SegFormer-B0 έχει γίνει στο ίδιο σύνολο δεδομένων. Η πολύ υψηλή τιμή του Recall και ο εξαιρετικά μικρός αριθμός False Negatives καθιστούν το μοντέλο ιδιαίτερα αξιόπιστο για

εφαρμογές ανίχνευσης πυρκαγιάς. Βέβαια, ως σημειωθεί πως η ύπαρξη ορισμένων False Positives δείχνει ότι υπάρχει ακόμη περιθώριο βελτίωσης στις προβλέψεις του μοντέλου.



Εικόνα 3-5 t-SNE του pretrained SegFormer στο FLAME

Το διάγραμμα t-SNE του μοντέλου SegFormer-B0 στο FLAME αποκαλύπτει μια εξαιρετικά δομημένη αναπαράσταση του λανθάνοντος χώρου, με σαφή διαχωρισμό μεταξύ των κλάσεων Fire και No Fire. Εμφανίζονται αρκετές συμπαγείς κόκκινες περιοχές, γεγονός που υποδηλώνει ότι το μοντέλο αναγνωρίζει διαφορετικά μοτίβα πυρκαγιάς. Επισημαίνεται μόνο μια περιορισμένη περιοχή επικάλυψης, γεγονός που συμφωνεί με τον μικρό αριθμό σφαλμάτων ταξινόμησης που αναφέρθηκαν στον πίνακα σύγχυσης (59 ψευδώς θετικά και 2 ψευδώς αρνητικά).



Εικόνα 3-6 Χάρτες Grad-CAM για το pretrained SegFormer στο FLAME

Στην σειρά των χαρτών Grad-CAM παραπάνω, παρατηρούμε τα εξής σημεία.

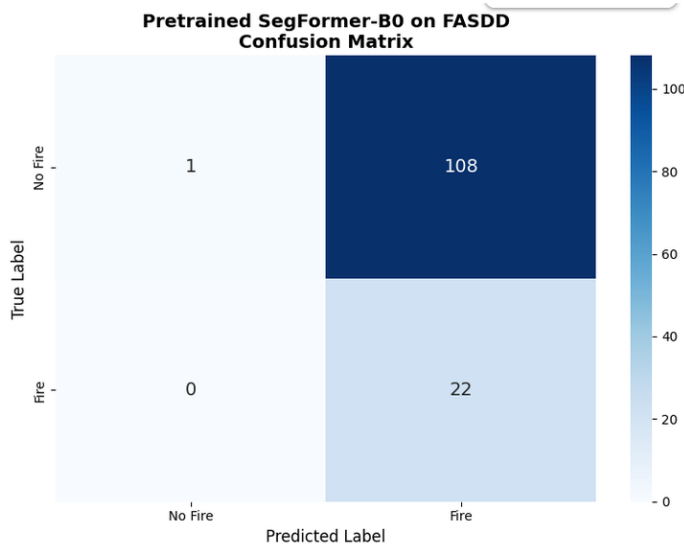
Στο True Positive, η ενεργοποίηση είναι εντοπισμένη γύρω από τη φλόγα, κάτι που δείχνει ότι το μοντέλο «κοιτάει» σημασιολογικά σχετικές περιοχές της εικόνας. Στο False Positive, η ενεργοποίηση εξαπλώνεται σε μια φωτεινή περιοχή του ουρανού κάτι που μας λέει ότι έντονα φωτεινά σημεία ή αντανakλάσεις που μοιάζουν χρωματικά με τη φωτιά μπορεί να μπερδέψουν το μοντέλο. Στο False Negative, η σχεδόν πλήρης απουσία ενεργοποίησης στην περιοχή του καπνού υποδηλώνει ότι οι συγκεκριμένες οπτικές συνθήκες (λεπτός ή διάχυτος καπνός, πιθανή ομίχλη κ.α.) δεν ενεργοποίησαν τα χαρακτηριστικά που έχει μάθει το μοντέλο.

3.1.3 στο FASDD

Η απόδοση του προεκπαιδευμένου μοντέλου στο FASDD είναι η ακόλουθη

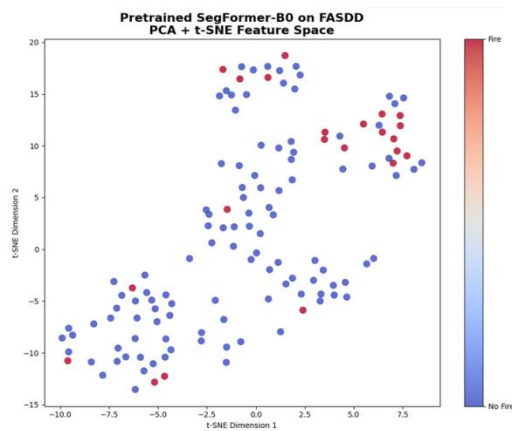
Πίνακας 3-5 Μετρικές του pretrained SegFormer στο FASDD

Pretrained SegFormer-B0 στο FASDD	
Accuracy	0.1756
Precision	0.1692
Recall	1.0000
F1-score	0.2895



Εικόνα 3-7 ConfusionMatrix του pretrained SegFormer στο FASDD

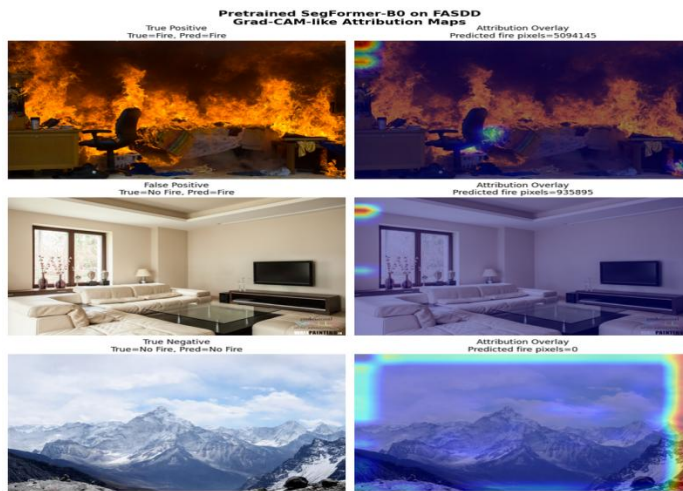
Από τον πίνακα των αποτελεσμάτων φαίνεται πως το προεκπαιδευμένο SegFormer-B0 παρουσιάζει ιδιαίτερα χαμηλή συνολική απόδοση στο σύνολο δεδομένων FASDD. Παρότι το μοντέλο ανίχνευσε επιτυχώς όλες τις εικόνες που περιείχαν πυρκαγιά (Recall = 100%), η πολύ χαμηλή τιμή του Precision και του Accuracy υποδηλώνει ότι ταξινομήσε λανθασμένα τη μεγάλη πλειονότητα των εικόνων χωρίς πυρκαγιά ως Fire. Ο confusionmatrixέρχεται προς επίρρωση των μετρικών. Συγκεκριμένα, και οι 22 εικόνες με πυρκαγιά ταξινομήθηκαν σωστά (False Negatives = 0), ενώ από τις 109 εικόνες χωρίς πυρκαγιά μόνο 1 αναγνωρίστηκε σωστά και οι υπόλοιπες 108 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Το αποτέλεσμα αυτό δείχνει ότι το μοντέλο παρουσιάζει εξαιρετικά έντονη τάση να προβλέπει την παρουσία πυρκαγιάς σχεδόν σε κάθε εικόνα, γεγονός που οδηγεί σε πολύ μεγάλο αριθμό ψευδώς θετικών προβλέψεων.



Εικόνα 3-8 t-SNE του pretrained SegFormer στο FASDD

Στο διάγραμμα t-SNE στην παραπάνω εικόνα βλέπουμε ότι σχηματίζονται ορισμένες τοπικές συστάδες δειγμάτων, ωστόσο οι δύο κλάσεις (Fire και No Fire) δεν παρουσιάζουν σαφή διαχωρισμό. Τα περισσότερα δείγματα της κλάσης Fire βρίσκονται αναμειγμένα με δείγματα της κλάσης No Fire, γεγονός που υποδηλώνει ότι τα χαρακτηριστικά που εξάγει το μοντέλο δεν είναι επαρκώς διακριτικά για το συγκεκριμένο σύνολο δεδομένων. Η έντονη αυτή επικάλυψη μας εξηγεί την αδυναμία του μοντέλου να διαχωρίσει αποτελεσματικά τις δύο κατηγορίες και είναι συνεπής με τα αποτελέσματα της αξιολόγησης, όπου παρατηρήθηκε πολύ μεγάλος αριθμός False Positives και ιδιαίτερα χαμηλή συνολική ακρίβεια. Συνεπώς, και με το t-SNE συμπεραίνουμε ότι το προεκπαιδευμένο SegFormer-B0 δεν έχει προσαρμοστεί στα χαρακτηριστικά του FASDD.

Όσον αφορά τώρα στη σειρά των χαρτών Grad-CAM που παρατίθενται ακολούθως, ένα σημείο που μπορούμε να σημειώσουμε είναι ότι και στις τρεις περιπτώσεις (TruePositive, FalsePositive, TrueNegative) η εστίαση του μοντέλου επικεντρώνεται στα άκρα της εικόνας. Το σημείο αυτό ουσιαστικά δείχνει ότι το μοντέλο δεν βασίζεται σε σημασιολογικά σημαντικές περιοχές, αλλά επηρεάζεται από περιφερειακά χαρακτηριστικά ή στοιχεία του υποβάθρου.



Εικόνα 3-9 Χάρτες Grad-CAM του pretrained SegFormer στο FASDD

3.2 Finetuning του SegFormer-B0

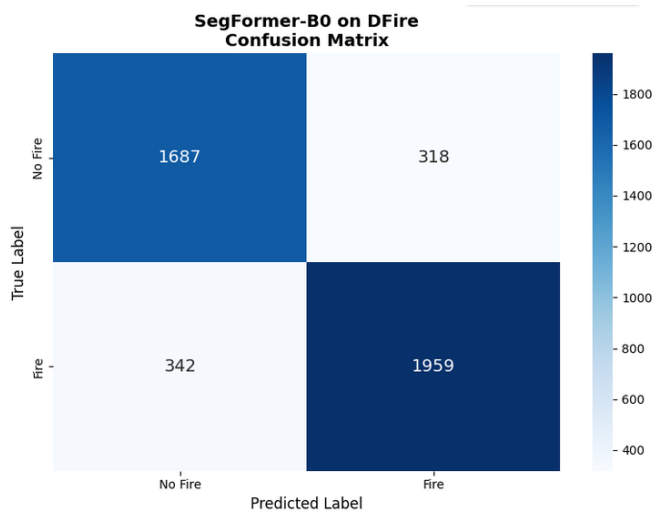
Στην παράγραφο αυτή προβαίνουμε σε finetuning του μοντέλου SegFormer στα τρία σύνολα δεδομένων DFire, FLAME και FASDD.

3.2.1 στο DFire

Μετά την εκπαίδευση του μοντέλου για 3 epochs, η αξιολόγηση του στο testset έδωσε τα ακόλουθα αποτελέσματα.

Πίνακας 3-6 Μετρικές του fine-tuned SegFormer στο DFire

Fine Tuned SegFormer-B0 στο DFire	
Accuracy	0.8509
Precision	0.8738
Recall	0.8555
F1-score	0.8645



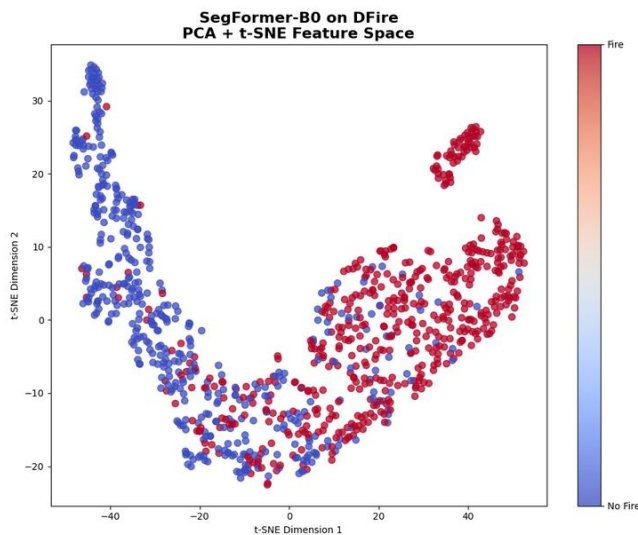
Εικόνα 3-10 Confusion Matrix του fine-tuned SegFormer στο DFire

Παρατηρούμε ότι μετά τη διαδικασία fine-tuning, το SegFormer-B0 παρουσίασε σημαντική βελτίωση στην απόδοσή του στο σύνολο δεδομένων DFire, επιτυγχάνοντας Accuracy 85,09%, Precision 87,38%, Recall 85,55% και F1-score 86,45%. Οι τιμές αυτές μας δείχνουν ότι το μοντέλο απέκτησε μεγαλύτερη ικανότητα διάκρισης μεταξύ των εικόνων με και χωρίς πυρκαγιά, επιτυγχάνοντας υψηλή ισορροπία μεταξύ ακρίβειας και ευαισθησίας. Διαβάζοντας το confusion matrix σημειώνουμε ότι από τις 2.005 εικόνες χωρίς πυρκαγιά, οι 1.687 ταξινομήθηκαν σωστά, ενώ μόνο 318 χαρακτηρίστηκαν λανθασμένα ως Fire (False Positives). Αντίστοιχα, από τις 2.301 εικόνες με πυρκαγιά, το μοντέλο αναγνώρισε σωστά τις 1.959, ενώ 342 εικόνες ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Εξάγουμε λοιπόν το συμπέρασμα ότι το μοντέλο διατηρεί υψηλή ικανότητα ανίχνευσης πυρκαγιάς, ενώ παράλληλα περιορίζει σημαντικά τον αριθμό των ψευδώς θετικών προβλέψεων. Η σημαντική αύξηση της ακρίβειας και του F1-score αποδεικνύει ότι το μοντέλο προσαρμόστηκε αποτελεσματικά στα χαρακτηριστικά του νέου συνόλου δεδομένων.

Περνώντας τώρα στην ανάγνωση του t-SNE που παρατίθεται ακολούθως, σε αντίθεση με την προεκπαιδευμένη έκδοση του μοντέλου, παρατηρούμε σαφώς καλύτερο διαχωρισμό μεταξύ των δύο κλάσεων (Fire και No Fire). Εξακολουθούν να υπάρχουν ορισμένες περιοχές επικάλυψης μεταξύ των δύο κλάσεων, όμως

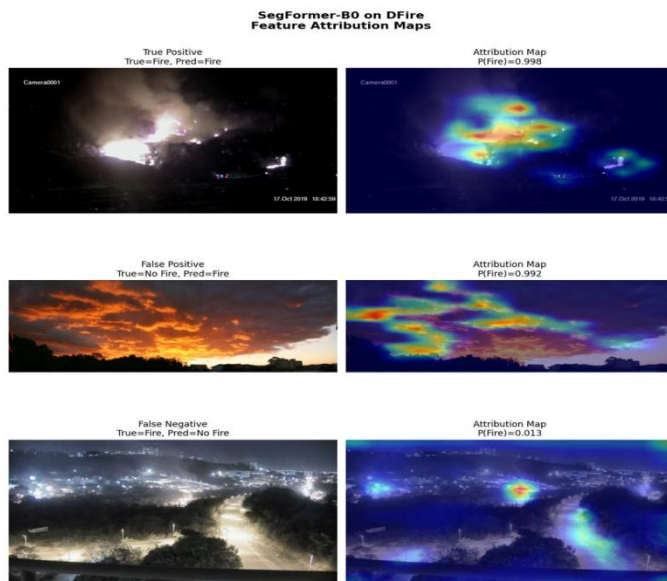
αυτές παρατηρούμε ότι περιορίζονται κυρίως στη μεταβατική περιοχή μεταξύ των δύο κύριων συστάδων. Η εικόνα αυτή υποδηλώνει ότι το fine-tuning οδήγησε στην εκμάθηση περισσότερο διακριτικών χαρακτηριστικών, επιτρέποντας στο μοντέλο

να διαχωρίζει αποτελεσματικότερα τις δύο κατηγορίες.



Εικόνα 3-11 t-SNE του fine-tuned SegFormer στο DFire

Στην ακόλουθη σειρά εικόνων βλέπουμε τους χάρτες Grad-CAM για τρία instances από το σύνολο δεδομένων. Αν συγκρίνουμε αυτά τα Grad-CAM με εκείνα του pretrained SegFormer στο DFire, μπορούμε να διακρίνουμε μια σαφή διαφορά. Στην προεκπαιδευμένη περίπτωση οι ενεργοποιήσεις στους χάρτες Grad-CAM ήταν συχνά διάχυτες ή μετατοπισμένες προς τα άκρα της εικόνας και σε περιοχές χωρίς σημασιολογική σχέση με την πυρκαγιά. Μετά το fine-tuning, οι ενεργοποιήσεις των χαρτών Grad-CAM είναι περισσότερες εστιασμένες και σωστά τοποθετημένες σε σχέση με τις πραγματικές περιοχές φωτιάς και καπνού, ιδιαίτερα στις σωστά ταξινομημένες εικόνες. Η παρατήρηση αυτή μας δείχνει ότι το μοντέλο δεν βελτίωσε μόνο την ικανότητά του να ταξινομεί σωστά αλλά και την ποιότητα των χαρακτηριστικών στα οποία βασίζει τις αποφάσεις του.



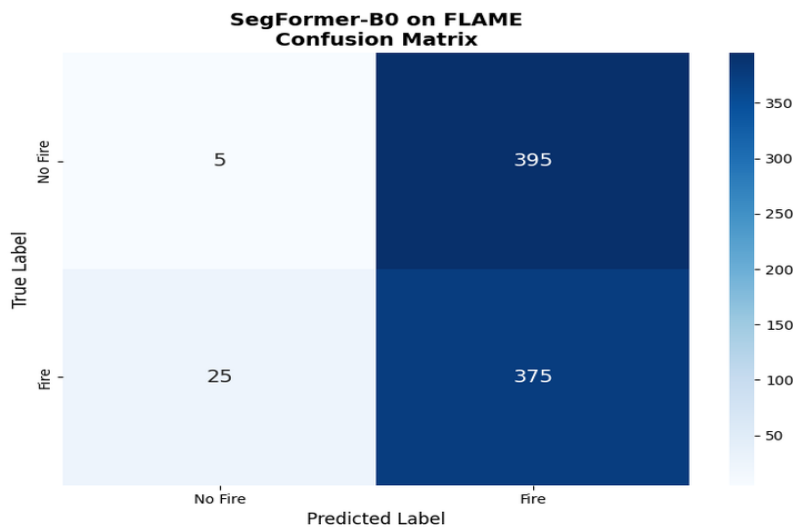
Εικόνα 3-12 Χάρτες Grad-CAM του fine-tuned SegFormer στο DFire

3.2.2 στο FLAME

Μετά από την εκπαίδευση του μοντέλου SegFormer-B0 στο FLAME, το μοντέλο αξιολογήθηκε στο testset και πήραμε τα ακόλουθα αποτελέσματα στις διάφορες μετρικές .

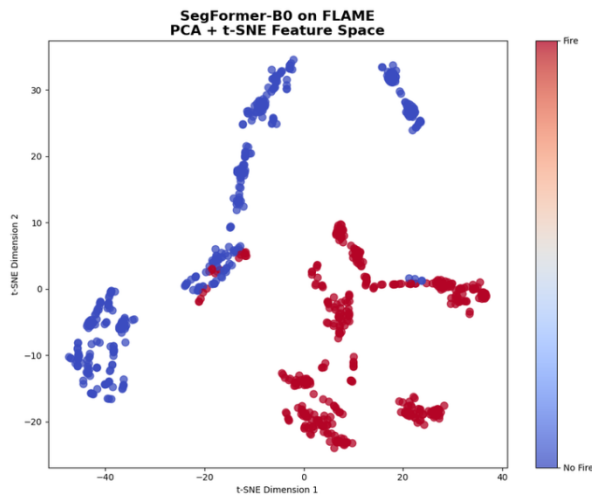
Πίνακας 3-7 Μετρικές του fine-tuned SegFormer στο FLAME

Fine Tuned SegFormer-B0 στο FLAME	
Accuracy	0.4750
Precision	0.4870
Recall	0.9375
F1-score	0.6410



Εικόνα 3-13 Confusion Matrix του fine-tuned SegFormer στοFLAME

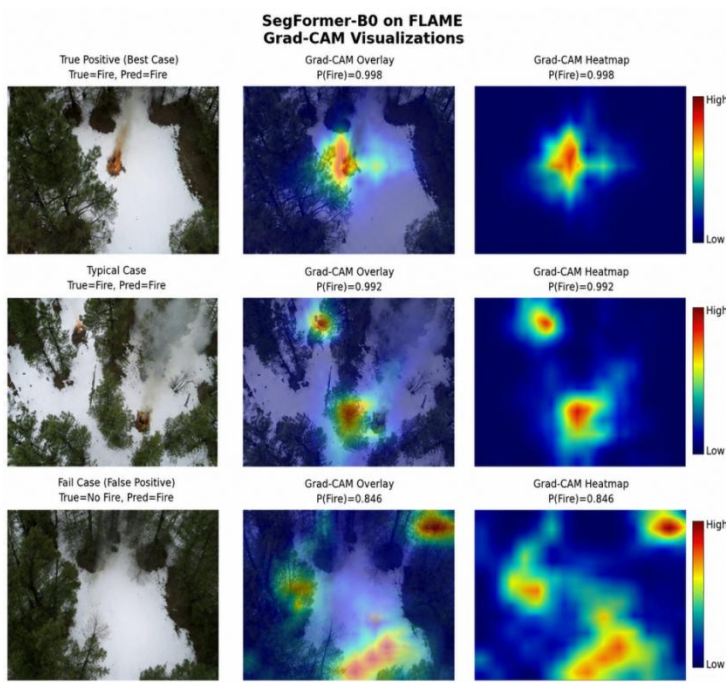
Μετά τη διαδικασία fine-tuning, βλέπουμε πως το SegFormer-B0 παρουσίασε σημαντική υποβάθμιση της απόδοσής του στο σύνολο δεδομένων FLAME, καταγράφοντας Accuracy 47,50%, Precision 48,70%, Recall 93,75% και F1-score 64,10%. Μπορούμε να σχολιάσουμε ότι ναι μεν το μοντέλο εξακολουθεί να διατηρεί υψηλή ευαισθησία στην ανίχνευση εικόνων με πυρκαγιά αλλά οπωσδήποτε η συνολική ικανότητα του για διάκριση μεταξύ των δύο κλάσεων μειώθηκε αισθητά. Σχετικά με τον confusionmatrix παρατηρούμε ότι από τις 400 εικόνες χωρίς πυρκαγιά μόνο 5 ταξινομήθηκαν σωστά, ενώ οι υπόλοιπες 395 χαρακτηρίστηκαν λανθασμένα ως Fire (False Positives). Αντίθετα, από τις 400 εικόνες με πυρκαγιά αναγνωρίστηκαν σωστά οι 375, ενώ μόλις 25 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Η συμπεριφορά αυτή υπογραμμίζει ότι το μοντέλο εμφανίζει έντονη μεροληψία προς την πρόβλεψη της κλάσης Fire, δημιουργώντας έτσι μεγάλο αριθμό ψευδώς θετικών προβλέψεων.



Εικόνα 3-14 t-SNE του fine-tuned SegFormer στο FLAME

Στο διάγραμμα t-SNE παρατηρούμε ότι τα δείγματα των δύο κλάσεων (Fire και No Fire) σχηματίζουν σαφώς διακριτές συστάδες, με περιορισμένη μόνο επικάλυψη σε ορισμένες περιοχές. Η εικόνα αυτή υποδηλώνει ότι το μοντέλο εξακολουθεί να εξάγει ιδιαίτερα διακριτικά χαρακτηριστικά για τις δύο κατηγορίες, διατηρώντας έναν καλά οργανωμένο χώρο χαρακτηριστικών. Βέβαια ο ικανοποιητικός αυτός διαχωρισμός, δεν συνοδεύεται όπως είδαμε και από την ίδια καλή εικόνα ως προς τις μετρικές αξιολόγησης οι οποίες έδειξαν σημαντική υποβάθμιση της απόδοσης μετά το fine-tuning. Αυτό το σημείο θα μπορούσε να σημαίνει ότι η υποβάθμιση της απόδοσης του μοντέλου σχετίζεται με μεταβολή του μηχανισμού λήψης απόφασης και δεν έχει να κάνει με την ποιότητα των εξαγόμενων χαρακτηριστικών.

Από τη σειρά των χαρτών Grad-CAM που ακολουθούν σημειώνουμε ότι παρά τη μείωση της συνολικής απόδοσης μετά το fine-tuning, το μοντέλο εξακολουθεί να εστιάζει σε περιοχές με έντονα οπτικά χαρακτηριστικά που σχετίζονται με τη φωτιά. Όμως, σε σύνθετες σκηνές ή σε εικόνες με έντονες μεταβολές στο φως, η προσοχή του επεκτείνεται και σε περιοχές που δεν αντιστοιχούν σε πραγματικές εστίες πυρκαγιάς. Φαίνεται λοιπόν ότι παρά τις χαμηλές τιμές στις μετρικές το μοντέλο δεν έχει χάσει την ικανότητά του στο να εντοπίζει τα χαρακτηριστικά της πυρκαγιάς.



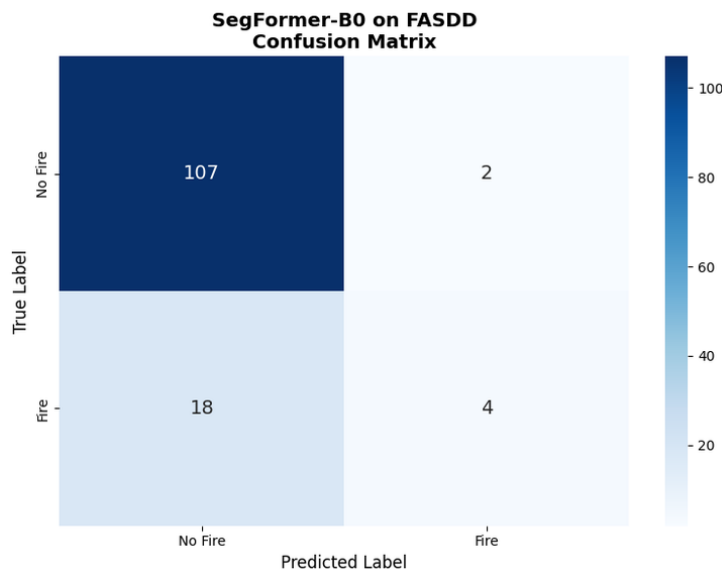
Εικόνα 3-15 Χάρτες Grad-CAM του fine-tuned SegFormer στο FLAME

3.2.3 στο FASDD

Η αξιολόγηση του fine-tuned μοντέλου στο test set του FASDD κατέγραψε τα ακόλουθα αποτελέσματα

Πίνακας 3-8 Μετρικές του fine-tuned SegFormer στο FASDD

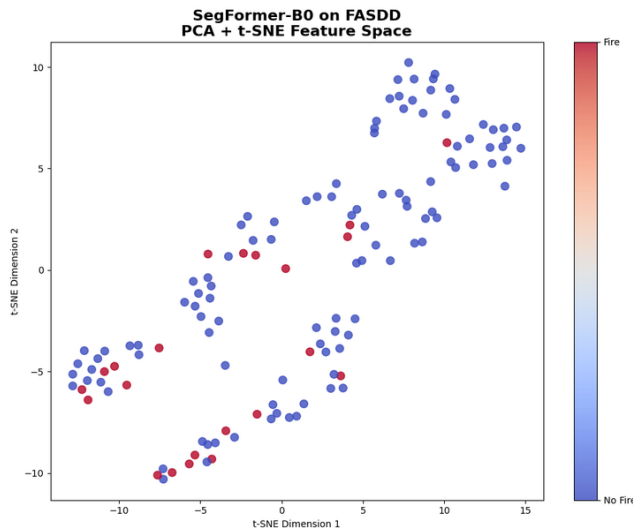
Fine Tuned SegFormer-B0 στο FASDD	
Accuracy	0.8473
Precision	0.6667
Recall	0.1818
F1-score	0.2857



Εικόνα 3-16 Confusion Matrix του fine-tuned SegFormer στο FASDD

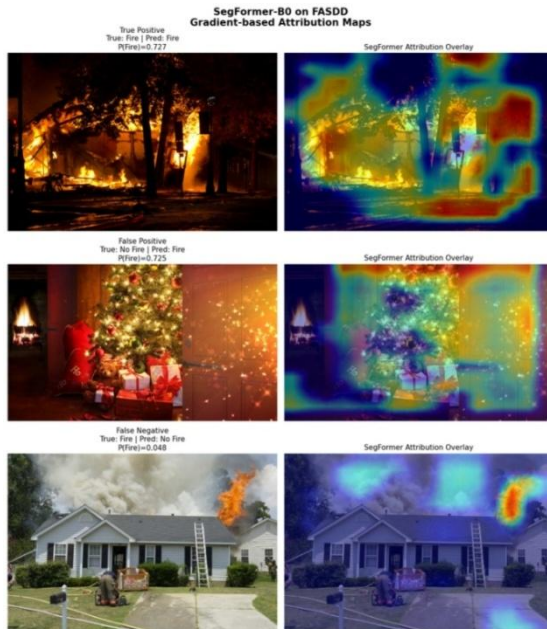
Από τον παραπάνω πίνακα επισημαίνουμε ότι οι χαμηλές τιμές του Recall και του F1-score δείχνουν ότι το μοντέλο δυσκολεύεται να αναγνωρίσει αποτελεσματικά τις εικόνες που περιέχουν πυρκαγιά. Ο confusionmatrix επιβεβαιώνει την παραπάνω σημείωση. Συγκεκριμένα, από τις 109 εικόνες χωρίς πυρκαγιά, οι 107 ταξινομήθηκαν σωστά και μόνο 2 χαρακτηρίστηκαν λανθασμένα ως Fire (False Positives). Αντίθετα, από τις 22 εικόνες με πυρκαγιά, το μοντέλο αναγνώρισε σωστά μόλις 4, ενώ οι υπόλοιπες 18 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Επομένως, το μοντέλο εμφανίζει πολύ υψηλή ικανότητα αναγνώρισης της κλάσης No Fire, αλλά ιδιαίτερα χαμηλή ευαισθησία στην ανίχνευση πραγματικών περιστατικών πυρκαγιάς. Σε σχέση με την προεκπαιδευμένη έκδοση του μοντέλου, μπορούμε να διαπιστώσουμε ότι η διαδικασία fine-tuning οδήγησε σε σημαντική μεταβολή της συμπεριφοράς του. Ενώ το προεκπαιδευμένο μοντέλο παρουσίαζε έντονη τάση να ταξινομεί σχεδόν όλες τις εικόνες ως Fire, μετά το fine-tuning η τάση αυτή αντιστράφηκε και το μοντέλο προβλέπει κυρίως την κλάση No Fire. Οι παραπάνω παρατηρήσεις ενισχύονται περαιτέρω και με τη μελέτη του t-SNE. Βλέπουμε ότι στο διάγραμμα t-SNE τα δείγματα των δύο κλάσεων (Fire και No Fire) εμφανίζουν σημαντική επικάλυψη. Τα περισσότερα δείγματα της κλάσης Fire είναι διάσπαρτα ανάμεσα στα δείγματα της κλάσης No Fire, χωρίς να σχηματίζουν έναν σαφώς διακριτό χώρο χαρακτηριστικών. Η

εικόνα αυτή υποδηλώνει ότι, παρά την προσαρμογή του μοντέλου μέσω fine-tuning, τα εξαγόμενα χαρακτηριστικά δεν είναι επαρκή ώστε να επιτρέψουν στο μοντέλο ένα αξιόπιστο διαχωρισμό μεταξύ των δύο κλάσεων στο FASDD.



Εικόνα 3-17 t-SNE του fine-tuned SegFormer στο FASDD

Προχωρώντας στο σχολιασμό των χαρτών Grad-CAM, βλέπουμε πρώτα πως στην περίπτωση του True Positive, οι ενεργοποιήσεις εντοπίζονται κυρίως στις περιοχές όπου εμφανίζονται οι φλόγες και οι έντονες φωτεινές εστίες της πυρκαγιάς, γεγονός που υποδηλώνει ότι το μοντέλο μπορεί να αναγνωρίσει σωστά τα χαρακτηριστικά της φωτιάς όταν αυτά είναι ιδιαίτερα εμφανή. Στην περίπτωση του False Positive, το μοντέλο ενεργοποιείται σε φωτεινά αντικείμενα της εικόνας, όπως τα διακοσμητικά φώτα και ο φωτισμός του χριστουγεννιάτικου δέντρου, οδηγώντας σε λανθασμένη πρόβλεψη για παρουσία πυρκαγιάς. Αντίθετα, στην περίπτωση του False Negative, οι ενεργοποιήσεις εντοπίζονται κυρίως σε περιοχές του ουρανού και σε δευτερεύοντα στοιχεία της σκηνής, ενώ η πραγματική περιοχή της φλόγας λαμβάνει περιορισμένη προσοχή. Γενικά μπορούμε να πούμε πως οι χάρτες Grad-CAM δείχνουν ότι μετά το fine-tuning το μοντέλο έχει γίνει περισσότερο επιλεκτικό στις προβλέψεις του, αλλά η προσοχή του δεν κατευθύνεται πάντοτε στα πλέον αντιπροσωπευτικά χαρακτηριστικά της πυρκαγιάς.



Εικόνα 3-18 Χάρτες Grad-CAM του fine-tuned SegFormer στο FASDD

3.3 Αξιολόγηση του προεκπαιδευμένου Yolov26

Στην παρούσα παράγραφο 3.3 θα ασχοληθούμε με το μοντέλο yolov26-fire-detection –v1.0 (εφεξής αναφερόμενο ως Yolov26) στην προεκπαιδευμένη εκδοχή του και θα δούμε την απόδοσή του στα τρία σύνολα δεδομένων τα οποία και χρησιμοποιούμε σε όλο το Κεφάλαιο 3.

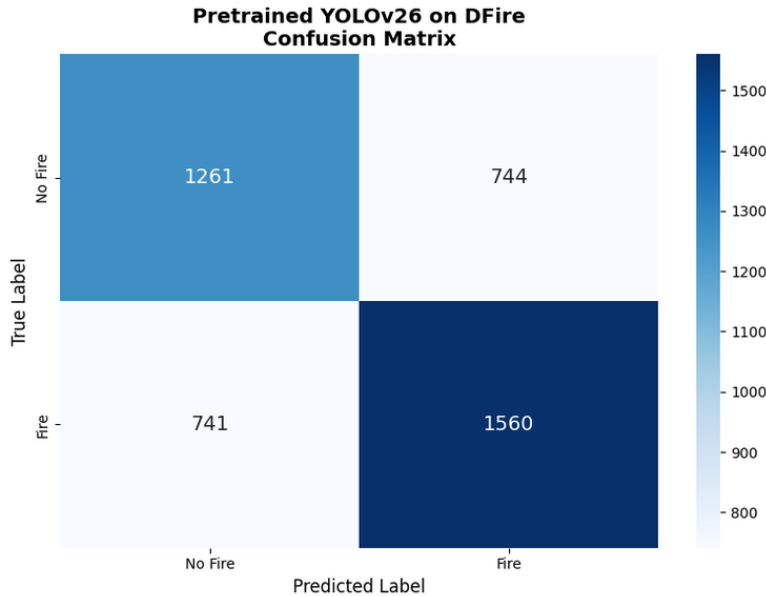
3.3.1 στο DFire

Τα αποτελέσματα που πήραμε από την αξιολόγηση του προεκπαιδευμένου YOLOv26 στο DFire είναι τα ακόλουθα .

Πίνακας 3-9 Μετρικές του pretrained Yolov26 στο DFire

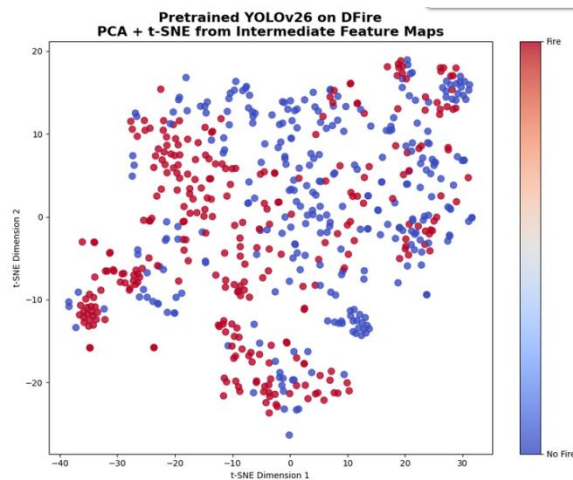
Pretrained Yolov26 on DFire	
Accuracy	0.6551
Precision	0.6771
Recall	0.6780

F1-score	0.6775
----------	--------



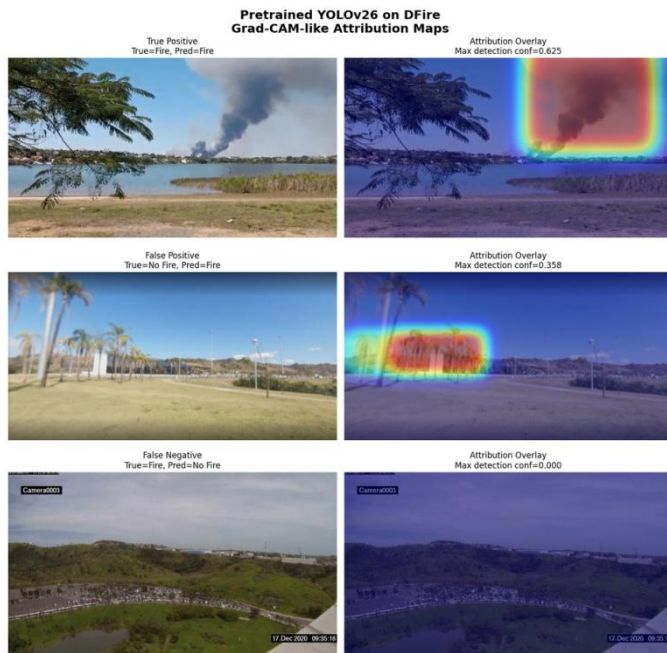
Εικόνα 3-19 Confusion Matrix του pretrained YOLOv26 στο DFire

Από τον confusion matrix του προεκπαιδευμένου μοντέλου YOLOv26 στο σύνολο δεδομένων DFire παρατηρούμε ότι το μοντέλο δείχνει μια μέτρια ικανότητα διάκρισης μεταξύ εικόνων με και χωρίς φωτιά. Συγκεκριμένα, το μοντέλο ταξινομήσε σωστά 1.560 εικόνες της κλάσης Fire και 1.261 εικόνες της κλάσης No Fire, ενώ παρήγαγε 741 ψευδώς αρνητικές (False Negatives) και 744 ψευδώς θετικές (False Positives) προβλέψεις. Η σχεδόν ισοδύναμη κατανομή των δύο τύπων σφαλμάτων υποδηλώνει ότι το μοντέλο δυσκολεύεται τόσο να ανιχνεύσει ορισμένα περιστατικά φωτιάς όσο και να απορρίψει σκηνές χωρίς φωτιά που παρουσιάζουν οπτικά χαρακτηριστικά παρόμοια με καπνό ή φλόγες.



Εικόνα 3-20 t-SNE του pretrained YOLOv26 στο DFire

Το διάγραμμα t-SNE για το προ-εκπαιδευμένο μοντέλο YOLOv26 στο σύνολο δεδομένων DFire αποκαλύπτει μια μερική διαχωριστικότητα μεταξύ των εικόνων Fire και NoFire. Παρόλο που παρατηρούνται αρκετές τοπικές συστάδες (clusters) που αντιστοιχούν κατά κύριο λόγο σε μία κλάση, εξακολουθεί να υπάρχει σημαντική αλληλοεπικάλυψη μεταξύ των δύο κατηγοριών. Αυτή η συμπεριφορά συμφωνεί με τα αποτελέσματα της ποσοτικής αξιολόγησης, όπου το μοντέλο πέτυχε F1-score 67.7%. Η οπτικοποίηση του t-SNE μας υποδεικνύει ότι το προ-εκπαιδευμένο μοντέλο συλλαμβάνει ορισμένα οπτικά μοτίβα που σχετίζονται με τη φωτιά, αλλά δεν καταφέρνει να μάθει αρκετά ώστε να επιτύχει ένα πιο εύρωστο διαχωρισμό ανάμεσα στις εικόνες με και χωρίς φωτιά στο DFire.



Εικόνα 3-21ΧάρτεςGrad-CAM τουpretrainedYolov26 στοDFire

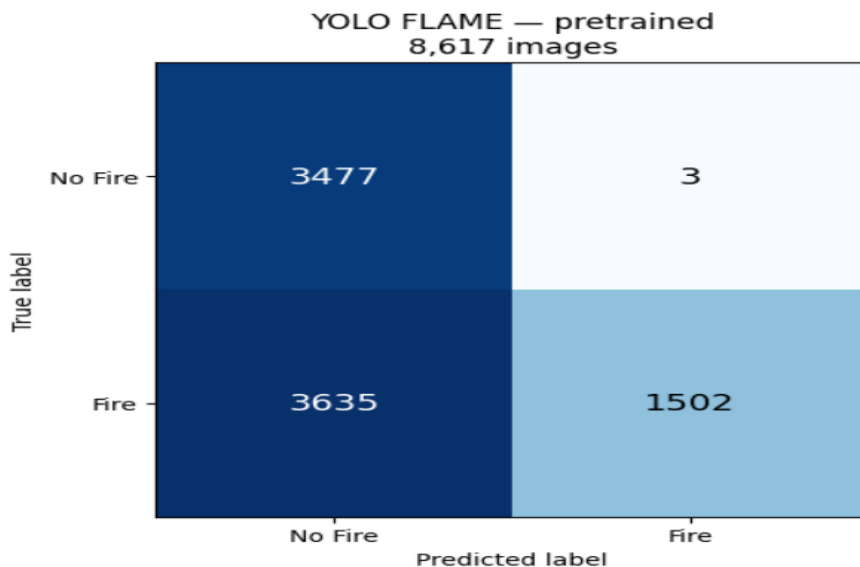
Στα τρία παραπάνω ζεύγη εικόνων χαρτών Grad-CAM, παρατηρούμε πως στο μεν παράδειγμα του True Positive, το μοντέλο εστιάζει σωστά στη στήλη καπνού, αποδεικνύοντας ότι έχει μάθει ουσιαστικά οπτικά χαρακτηριστικά που σχετίζονται με τη φωτιά. Αντίθετα, το παράδειγμα του False Positive αποκαλύπτει έντονες ενεργοποιήσεις πάνω από τη βλάστηση και τις δομές των δέντρων, υποδηλώνοντας ότι ορισμένα μοτίβα του υπόβαθρου συσχετίζονται εσφαλμένα με ενδείξεις φωτιάς. Τέλος, το παράδειγμα του False Negative εμφανίζει σχεδόν μηδενική ενεργοποίηση, υποδεικνύοντας ότι το μοντέλο αποτυγχάνει να εντοπίσει επαρκώς διακριτικά χαρακτηριστικά φωτιάς σε απαιτητικές σκηνές όπου η εστία της φωτιάς δεν είναι εμφανής.

3.3.2 στο FLAME

Τα αποτελέσματα που πήραμε από την αξιολόγηση του προεκπαιδευμένου YOLOv26 στο FLAME είναι τα ακόλουθα .

Πίνακας 3-10 Μετρικές του pretrained Yolon26 στο FLAME

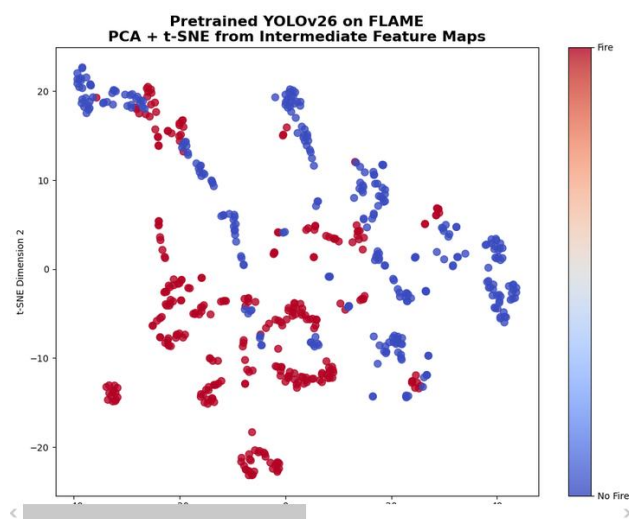
Pretrained Yolon26 on FLAME	
Accuracy	0.5778
Precision	0.9980
Recall	0.2924
F1-score	0.4523



Εικόνα 3-22 Confusion Matrix του pretrained Yolon26 στο FLAME

Το confusion matrix του προεκπαιδευμένου μοντέλου YOLOv26 στο σύνολο δεδομένων FLAME αποκαλύπτει σημαντικές δυσκολίες στην ανίχνευση περιστατικών πυρκαγιάς. Συγκεκριμένα, βλέπουμε ότι το μοντέλο ταξινόμησε σωστά 3.477 εικόνες της κλάσης No Fire και 1.502 εικόνες της κλάσης Fire, ενώ παρήγαγε 3 ψευδώς θετικές και 3.635 ψευδώς αρνητικές προβλέψεις. Ο ιδιαίτερα υψηλός αριθμός ψευδώς αρνητικών δειγμάτων αντανακλάται στη χαμηλή τιμή ανάκλησης (Recall = 29,24%), γεγονός που σηματοδοτεί ότι το μοντέλο μας αδυνατεί να εντοπίσει μεγάλο ποσοστό των πραγματικών περιστατικών πυρκαγιάς. Αντίθετα, η σχετικά υψηλή ακρίβεια θετικών προβλέψεων (Precision = 99,8%)

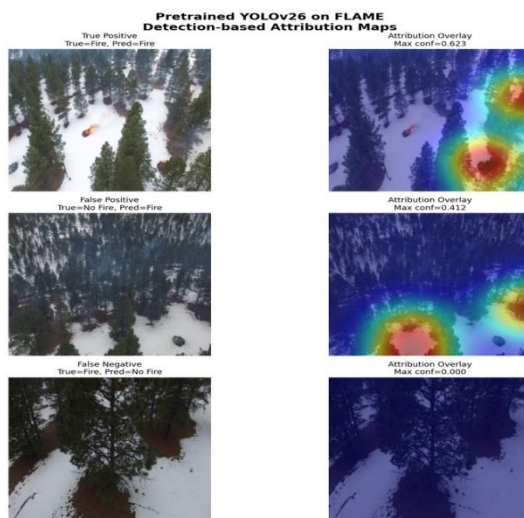
δείχνει ότι όταν το μοντέλο αποφασίζει πως υπάρχει φωτιά, η πρόβλεψη του είναι σχεδόν πάντα σωστή. Μια πιθανή εξήγηση για τη χαμηλή απόδοση του μοντέλου είναι ότι το προεκπαιδευμένο YOLOv26 εκπαιδεύτηκε ως detector για συγκεκριμένα firepatterns. Το σύνολο FLAME περιέχει πολλές εικόνες με μεγάλες περιοχές καπνού, μικρές εστίες φωτιάς, καθώς και λήψεις UAV από μεγάλη απόσταση. Ενδέχεται λοιπόν το μοντέλο δυσκολευόμενο να ανταποκριθεί στα χαρακτηριστικά του FLAME και προσπαθώντας να αποφύγει τα false alarms, να αστοχεί στην ορθή εκτίμηση πολλών πραγματικών πυρκαγιών.



Εικόνα 3-23 t-SNE του pretrained YOLOv26 στο FLAME

Η οπτικοποίηση t-SNE των ενδιάμεσων χαρακτηριστικών του προεκπαιδευμένου YOLOv26 στο FLAME αποκαλύπτει την ύπαρξη αρκετών διακριτών ομάδων εικόνων Fire και No Fire, υποδηλώνοντας ότι το backbone του μοντέλου έχει μάθει χρήσιμες οπτικές αναπαραστάσεις των δύο κατηγοριών. Παρά τη σχετικά καλή διαχωρισιμότητα που παρατηρείται στον χώρο χαρακτηριστικών, οι ποσοτικές μετρικές αξιολόγησης παραμένουν χαμηλές (Accuracy = 57,78%, Recall = 29,24%). Το γεγονός αυτό υποδηλώνει ότι η κύρια πηγή σφάλματος δεν εντοπίζεται στην εξαγωγή χαρακτηριστικών αλλά στο στάδιο της ανίχνευσης, όπου το detection head αποτυγχάνει να ενεργοποιηθεί σε μεγάλο αριθμό πραγματικών περιστατικών πυρκαγιάς.

Οι χάρτες Grad-CAM αποκαλύπτουν ότι το μοντέλο εστιάζει σωστά στις περιοχές φλόγας στις περιπτώσεις επιτυχούς ανίχνευσης (True Positives), γεγονός που επιβεβαιώνει ότι έχει μάθει κατάλληλα χαρακτηριστικά πυρκαγιάς. Ωστόσο, στις περιπτώσεις ψευδώς θετικών προβλέψεων παρατηρούμε ενεργοποίηση σε περιοχές με υψηλή φωτεινότητα. Επιπλέον, στις ψευδώς αρνητικές περιπτώσεις το μοντέλο δεν εμφανίζει ουσιαστική ενεργοποίηση, υποδηλώνοντας αδυναμία εντοπισμού μικρών, απομακρυσμένων ή οπτικά δυσδιάκριτων εστιών φωτιάς.



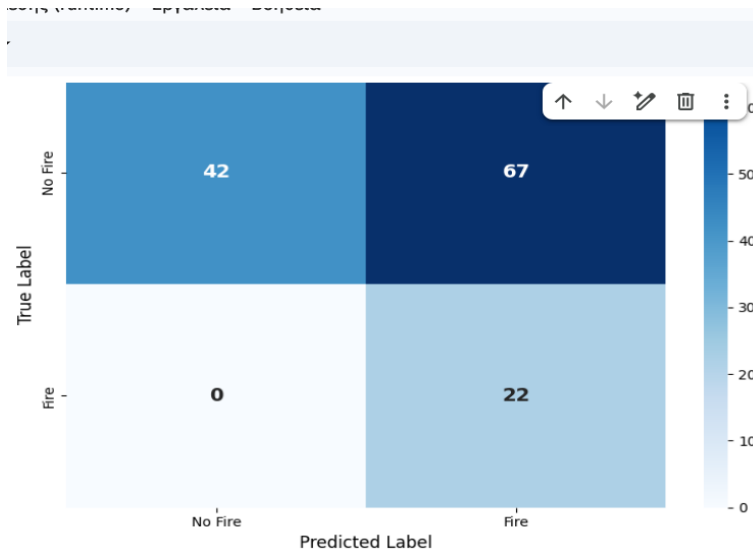
Εικόνα 3-24 Χάρτες Grad-CAM του pretrained YOLOv26 στο FLAME

3.3.3 στο FASDD

Τα αποτελέσματα που πήραμε από το προεκπαιδευμένο YOLOv26 στο FASDD είναι τα ακόλουθα

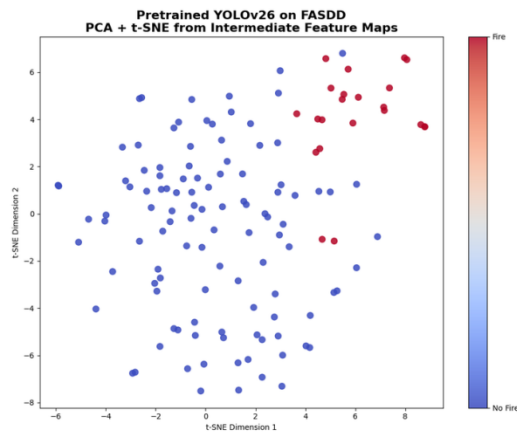
Πίνακας 3-11 Μετρικές του pretrained YOLOv26 στο FASDD

Pretrained YOLOv26 on FASDD	
Accuracy	0.4885
Precision	0.2472
Recall	1.000
F1-score	0.3964



Εικόνα 3-25 Confusion Matrix του pretrained YOLOv26 στο FASDD

Παρατηρούμε ότι η αξιολόγηση του προεκπαιδευμένου YOLOv26 στο FASDD δείχνει έντονα στοιχεία ασυμμετρίας. Το μοντέλο εντόπισε σωστά όλες τις εικόνες της κλάσης Fire, επιτυγχάνοντας $Recall = 1.000$, γεγονός που υποδηλώνει υψηλή ευαισθησία απέναντι σε οπτικά μοτίβα φωτιάς. Ωστόσο, η απόδοση αυτή συνοδεύεται από μεγάλο αριθμό ψευδώς θετικών προβλέψεων, καθώς 67 από τις 109 εικόνες No Fire ταξινομήθηκαν λανθασμένα ως Fire. Αυτό οδηγεί σε χαμηλή $Precision = 0.247$ και συνολικό $F1-score = 0.396$. Συνεπώς προκύπτει το γεγονός ότι το μοντέλο προτιμά να προβλέπει πιθανή φωτιά ακόμη και σε αμφίβολες σκηνές που μπορεί να περιέχουν χαρακτηριστικά που μοιάζουν με φωτιά.



Εικόνα 3-26 t-SNE του pretrained YOLOv26 στο FASDD

Όσον αφορά στο t-SNE διάγραμμα αυτό αποκαλύπτει ένα σχετικά καλό διαχωρισμό μεταξύ των κλάσεων Fire και No Fire. Ειδικότερα παρατηρούμε ότι οι περισσότερες εικόνες πυρκαγιάς σχηματίζουν ένα διακριτό cluster στον χώρο χαρακτηριστικών, γεγονός που υποδηλώνει ότι το backbone του μοντέλου έχει μάθει χρήσιμες αναπαραστάσεις των οπτικών χαρακτηριστικών της φωτιάς. Ωστόσο, ο σχετικά καλός διαχωρισμός δεν μεταφράζεται σε υψηλή επίδοση στις μετρικές ταξινόμησης, καθώς το confusion matrix μας έδειξε μεγάλο αριθμό ψευδώς θετικών προβλέψεων. Το εύρημα αυτό ενδεχομένως να υποδηλώνει ότι το πρόβλημα δεν εντοπίζεται στην εξαγωγή χαρακτηριστικών αλλά στον τρόπο με τον οποίο το detection head λαμβάνει την τελική απόφαση, οδηγώντας το μοντέλο σε μία απόδοση με πολύ υψηλό Recall αλλά χαμηλή Precision.

Από τους χάρτες Grad-Cam των τριών instances του προεκπαιδευμένου YOLOv26 στο FASDD φαίνεται ότι το μοντέλο εστιάζει σωστά στις περιοχές φλόγας στις περιπτώσεις επιτυχούς ανίχνευσης (πρώτη σειρά στις παραπάνω εικόνες), επιτυγχάνοντας πολύ υψηλό confidence στις εικόνες πραγματικής πυρκαγιάς. Ωστόσο, στις ψευδώς θετικές περιπτώσεις παρατηρείται έντονη ενεργοποίηση σε φωτεινά αστικά αντικείμενα και τεχνητές πηγές φωτισμού, γεγονός που υποδηλώνει ότι το μοντέλο μπερδεύει χαρακτηριστικά υψηλής φωτεινότητας με την παρουσία φωτιάς.



Εικόνα 3-27 Χάρτες Grad-CAM του pretrained Yolon26 στο FASDD

3.4 Finetuning του μοντέλου Yolon26.

Μετά από την εκτίμηση της απόδοσης της προεκπαιδευμένης εκδοχής του μοντέλου Yolon26, προχωρούμε στην παρούσα παράγραφο στη διαδικασία finetuning για το μοντέλο με τη σειρά σε καθένα από τα τρία σύνολα δεδομένων DFire, FLAME και FASDD.

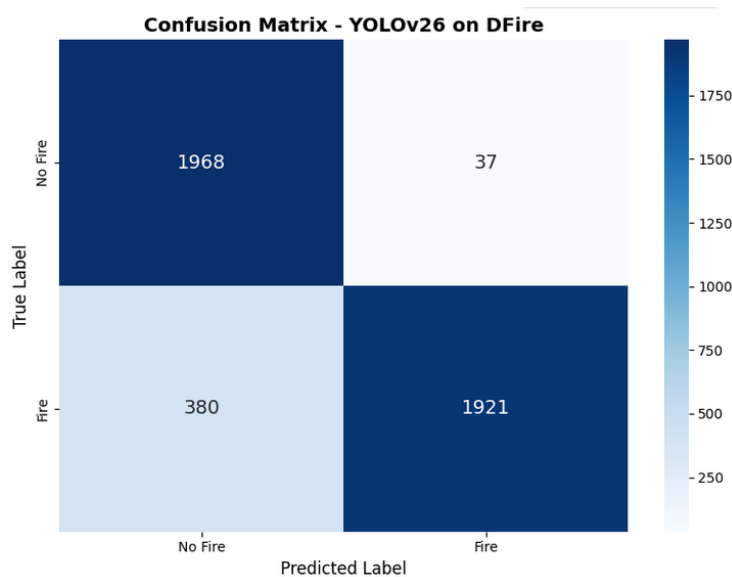
3.4.1 στο DFire

Ύστερα από το fine-tuning στο DFire, το μοντέλο Yolon26 κατά την δοκιμή του στο testset μας έδωσε τα ακόλουθα αποτελέσματα στις μετρικές που αναφέρονται στο imageclassification.

Πίνακας 3-12 Μετρικές του fine-tuned Yolon26 στο DFire

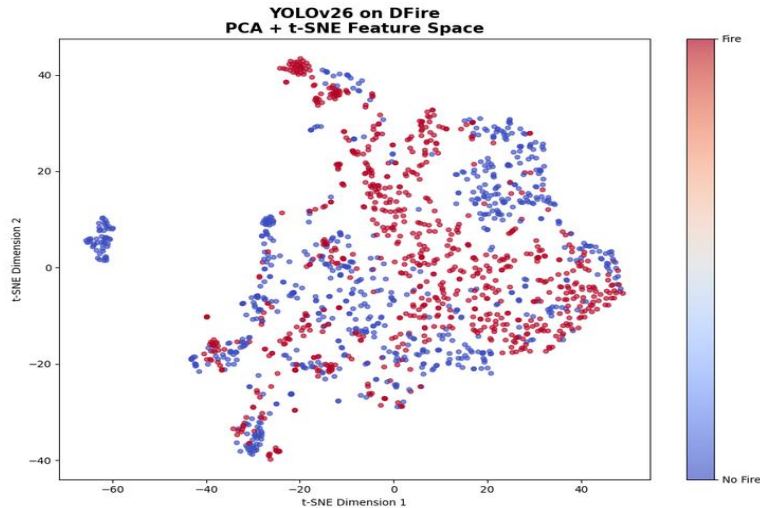
Fine Tuned Yolon26 on DFire	
Accuracy	0.9032

Precision	0.9811
Recall	0.8349
F1-score	0.9021



Εικόνα 3-28 Confusion Matrix του fine-tuned Yolov26 στο DFire

Με βάση το πίνακα με τις μετρικές και τον confusionmatrix το YOLOv26 μετά από το finetuning πέτυχε Accuracy 90,32%, Precision 98,11%, Recall 83,49% και F1-score 90,21% στο σύνολο δεδομένων DFire, σημειώνοντας συνολικά μια πολύ υψηλή απόδοση στην ανίχνευση πυρκαγιάς. Η ιδιαίτερα υψηλή τιμή του Precision υποδηλώνει ότι οι προβλέψεις της κλάσης Fire είναι εξαιρετικά αξιόπιστες, ενώ η υψηλή τιμή του F1-score καταδεικνύει καλή ισορροπία μεταξύ ακρίβειας και ευαισθησίας. Η πολύ καλή αυτή επίδοση μας δείχνει ότι το μοντέλο μετά από το finetuning προσαρμόστηκε ικανοποιητικά στα χαρακτηριστικά του συνόλου δεδομένων DFire. Δεδομένων των ποσοτικών αποτελεσμάτων το μοντέλο παρουσιάζεται ως μια ιδιαίτερα αξιόπιστη λύση για εφαρμογές αυτόματης ανίχνευσης δασικών πυρκαγιών.

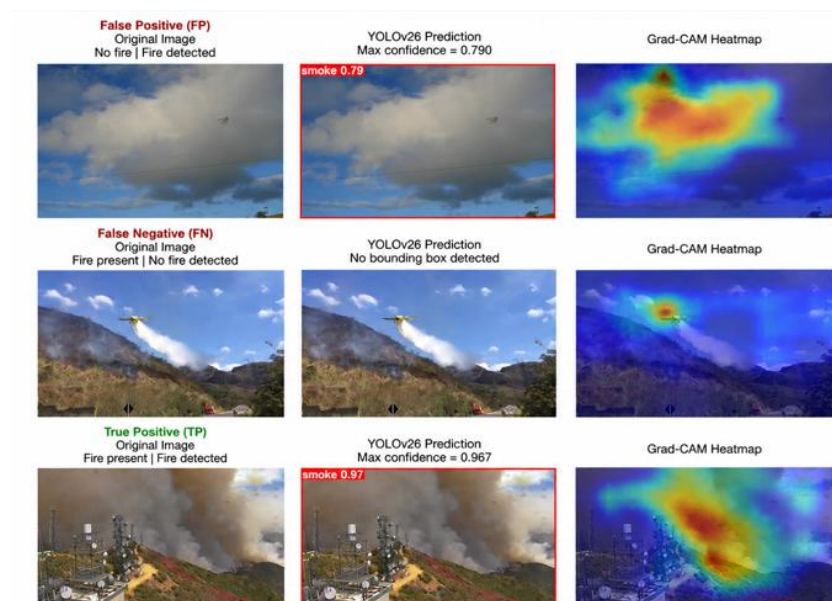


Εικόνα 3-29 t-SNE του fine-tuned Yolov26 στοDFire

Αυτό που παρατηρούμε στο διάγραμμα t-SNE είναι ότι τα δείγματα των δύο κλάσεων (Fire και No Fire) σχηματίζουν αρκετές τοπικές συστάδες, ωστόσο εξακολουθεί να υπάρχει σημαντική επικάλυψη μεταξύ τους σε αρκετές περιοχές του χώρου χαρακτηριστικών. Παρά την επικάλυψη αυτή, διακρίνονται και επιμέρους περιοχές στις οποίες κυριαρχεί μία από τις δύο κλάσεις, γεγονός που υποδηλώνει ότι το μοντέλο έχει μάθει να εξάγει χαρακτηριστικά με ικανοποιητική διακριτική ικανότητα, χωρίς όμως να επιτυγχάνει πλήρη διαχωρισμό των δύο κατηγοριών. Με μια πρώτη ματιά η εικόνα του t-SNE ενδεχομένως να φαίνεται σε αναντιστοιχία με τις υψηλές τιμές των αξιολογικών μετρικών του μοντέλου. Βάσει αυτών των τιμών ίσως να περιμέναμε ένα t-SNE στο οποίο οι κλάσεις θα συγκροτούσαν πιο συμπαγή και πιο καλά διαχωρισμένα clusters. Μια πιθανή αιτία για αυτό ίσως θα πρέπει να αναζητηθεί στη φύση του ίδιου του μοντέλου. Θυμόμαστε ότι το YOLOv26 είναι ένα μοντέλο σχεδιασμένο για object detection, επομένως οι αναπαραστάσεις που παράγει είναι βελτιστοποιημένες κυρίως για τον εντοπισμό και την αναγνώριση αντικειμένων, και όχι αποκλειστικά και μονομερώς για τη διάκριση ολόκληρων εικόνων σε δύο κλάσεις.

Στην ακόλουθη σειρά των χαρτών Grad-CAM-like δίπλα σε κάθε εικόνα έχουμε το boundingbox που δημιουργεί το μοντέλο και ακολούθως το grad-CAM heatmap

που μας υποδεικνύει σε ποια περιοχή εστιάζει την προσοχή του το μοντέλο. Ένα συνολικό σχόλιο που θα μπορούσαμε να κάνουμε από τις εικόνες αυτές είναι πως το YOLOv26 βασίζει τις αποφάσεις του κυρίως σε χαρακτηριστικά όπως ο πυκνός καπνός και οι έντονες φωτεινές περιοχές που συνοδεύουν μια πυρκαγιά. Παράλληλα, κρίνοντας από την 1^η σειρά εικόνων που απεικονίζει μια FalsePositive περίπτωση, μπορούμε να εικάσουμε πως αιτίες λανθασμένων προβλέψεων σχετίζονται είτε με σκηνές που περιλαμβάνουν στοιχεία οπτικά παρόμοια με καπνό (εδώ το νερό που ρίχνει το αεροπλάνο), είτε με περιπτώσεις όπου η πραγματική περιοχή της πυρκαγιάς είναι μικρή ή αποσπάται από άλλα κυρίαρχα αντικείμενα της εικόνας (εδώ το αεροπλάνο).



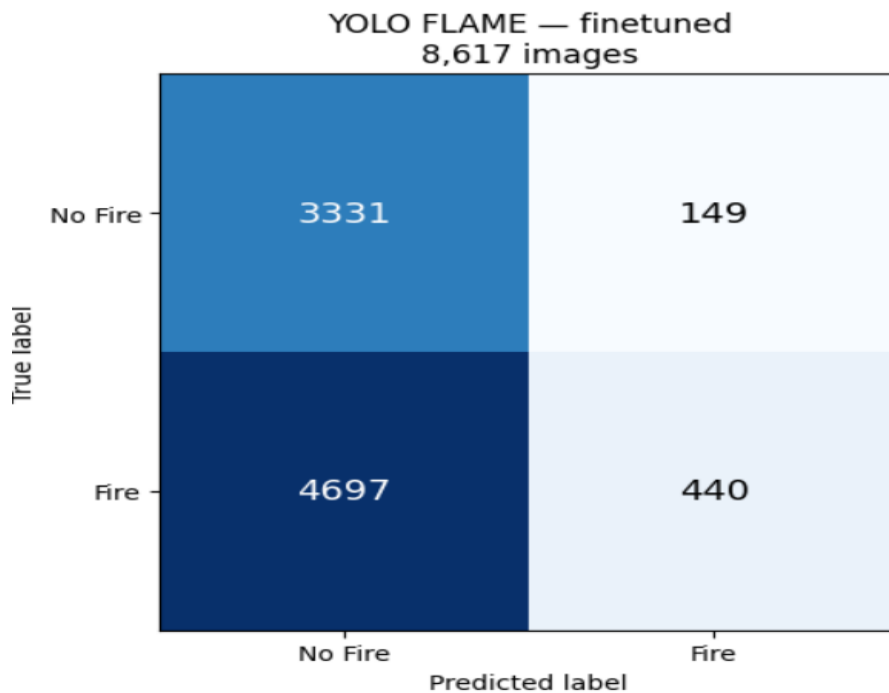
Εικόνα 3-30 Χάρτες Grad-CAM του fine-tuned YOLOv26 στο DFire

3.4.2 στο FLAME

Η εκπαίδευση του μοντέλου YOLOv26 στο FLAME μας έδωσε τα παρακάτω αποτελέσματα

Πίνακας 3-13 Μετρικές του fine-tunedYolov26 στο FLAME

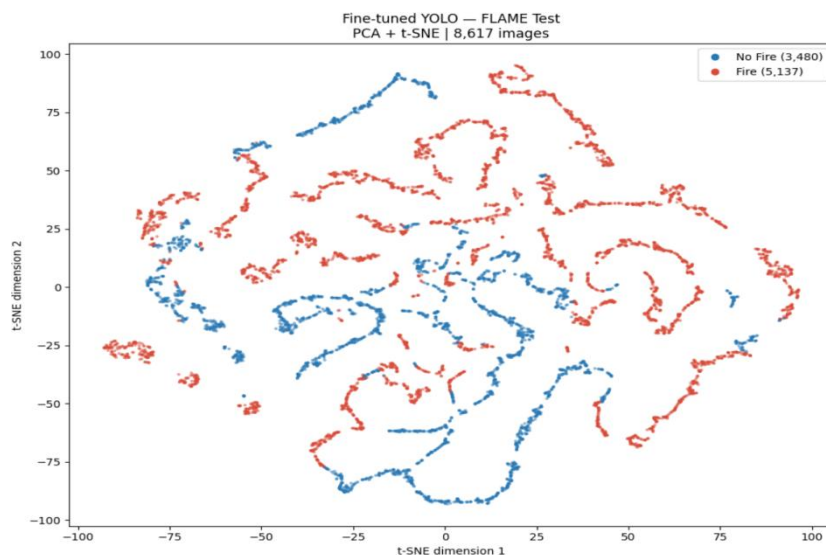
Fine Tuned Yolov26 στο FLAME	
Accuracy	0.4376
Precision	0.7470
Recall	0.0857
F1-score	0.1537



Εικόνα 3-31Confusion Matrix του fine-tuned Yolov26 στο FLAME

Διαβάζοντας κανείς τον παραπάνω πίνακα μετρικών ένα σημείο που τραβάει το ενδιαφέρον είναι η μεγάλη διαφορά μεταξύ της Precision και της Recall. Με βάση την τιμή της Precision βλέπουμε ότι περίπου τρεις στις τέσσερις προβλέψεις της κλάσης Fire είναι σωστές. Συγχρόνως όμως η ιδιαίτερη χαμηλή τιμή της Recall δείχνει ότι οι περισσότερες εικόνες με πυρκαγιά δεν αναγνωρίζονται σωστά. Συμπερασματικά το μοντέλο εμφανίζει έντονα συντηρητική συμπεριφορά ως προς την απόδοση της κλάσης Fire, με αποτέλεσμα μεγάλο αριθμό ψευδώς αρνητικών

προβλέψεων. Σχετικά με τον επιμερισμό των προβλέψεων από τις 3.840 εικόνες χωρίς πυρκαγιά οι 3.331 ταξινομήθηκαν σωστά ενώ 149 χαρακτηρίστηκαν λανθασμένα ως Fire. Από την άλλη πλευρά, από τις 5.317 εικόνες με φωτιά μόλις 440 εντοπίστηκαν ορθά ενώ 4.697 ταξινομήθηκαν λανθασμένα ως No Fire. Τα αποτελέσματα αυτά υπογραμμίζουν τη σημαντική αδυναμία της συγκεκριμένης εκδοχής του fine tuning μοντέλου να εντοπίσει τις εικόνες πυρκαγιάς στο σύνολο FLAME test. Ωστόσο στη συνολική ερμηνεία της επίδοσης θα πρέπει να συνεκτιμήσουμε ένα πρόβλημα που προέκυψε στην προετοιμασία των δεδομένων. Συγκεκριμένα υπήρξε ένα πρόβλημα αντιστοίχισης εικόνων-μασκών segmentation. Λόγω αυτού του προβλήματος αντιστοίχισης τα bounding boxes που προέκυψαν από τις μάσκες δύνανται να τοποθετούνται σε περιοχές διαφορετικές από τις πραγματικές εστίες φωτιάς των εικόνων εκπαίδευσης. Η δυσκολία αυτή ενδέχεται να επηρέασε την επίδοση του μοντέλου, η ενδεχόμενη όμως αυτή επιρροή μένει να αποτυπωθεί περαιτέρω πειραματικά.

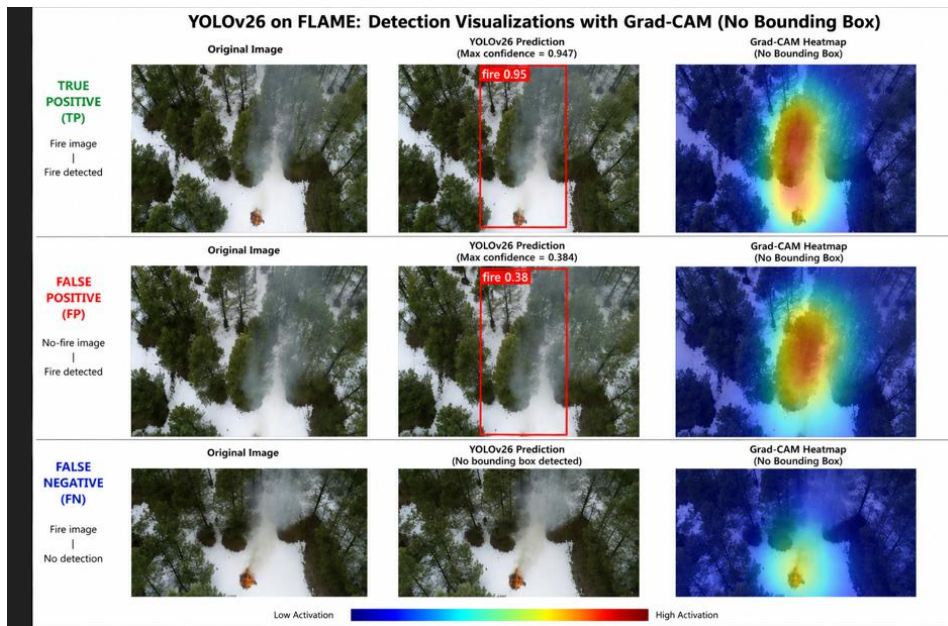


Εικόνα 3-32 t-SNE του fine-tuned YOLOv26 στο FLAME

Η παραπάνω εικόνα του διαγράμματος t-SNE μας δείχνει με καθαρό τρόπο ότι οι κλάσεις Fire και No Fire δεν οργανώνονται σε δύο ενιαίες και σαφώς διαχωρισμένες συστάδες. Αντιθέτως σχηματίζουν πολλές επιμέρους ομάδες με

περιοχές όπου οι δύο κλάσεις διαχωρίζονται τοπικά αλλά και άλλες όπου γειτνιάζουν ή αναμειγνύονται. Η δισδιάστατη αυτή εικόνα του χώρου των χαρακτηριστικών μας υποδεικνύει ότι τα χαρακτηριστικά του μοντέλου διατηρούν ορισμένες διαφορές μεταξύ εικόνων με και χωρίς φωτιά χωρίς να προκύπτει πλήρης διαχωρισμός.

Στην ακόλουθη ομάδα εικόνων έχουμε σε κάθε σειρά ένα instance από το σύνολο δεδομένων το οποίο έχει δίπλα του το detection visualization με το bounding box που σκισάρει το μοντέλο και αυτό με τη σειρά του ακολουθείται από το Grad-CAM χάρτη. Στην περίπτωση του True Positive, το μοντέλο εντοπίζει επιτυχώς την περιοχή όπου εμφανίζονται η φλόγα και ο καπνός, δημιουργώντας πλαίσιο ανίχνευσης (bounding box) με υψηλή τιμή εμπιστοσύνης (confidence = 0.947). Στο False Positive, το Yolov26 ενεργοποιείται σε περιοχή όπου κυριαρχεί ο καπνός, παρά την απουσία πραγματικής φλόγας, παράγοντας πρόβλεψη για την κλάση Fire. Παρ'όλα αυτά η σχετική του πρόβλεψη γίνεται με αρκετά χαμηλή εμπιστοσύνη (confidence = 0.384). Η χαμηλότερη τιμή confidence σε σχέση με το True Positive μας ενημερώνει ότι το μοντέλο εμφανίζει μεγαλύτερη αβεβαιότητα στις λανθασμένες προβλέψεις του. Τέλος στην περίπτωση του False Negative, το μοντέλο δεν παράγει κανένα πλαίσιο ανίχνευσης, παρά την παρουσία πυρκαγιάς στην εικόνα. Ενδεχομένως αυτό να σχετίζεται με το ότι τα χαρακτηριστικά της φωτιάς είναι περιορισμένα ή δεν ξεχωρίζει επαρκώς η εστία της φωτιάς από το περιβάλλοντα χώρο.



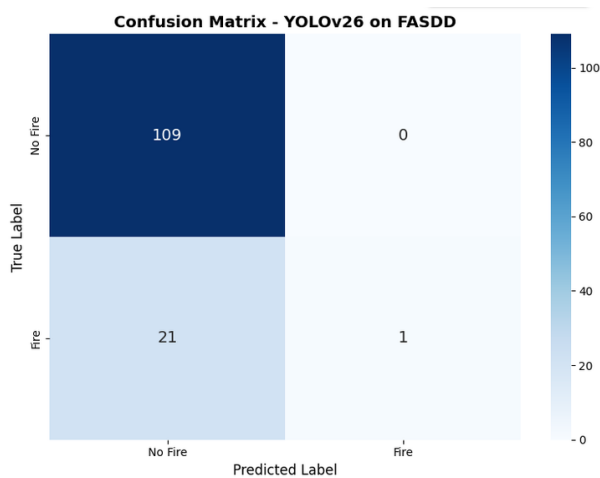
Εικόνα 3-33 Χάρτες Grad-CAM του fine-tuned YOLOv26 στο FLAME

3.4.3 στο FASDD

Μετά από το fine-tuning του μοντέλου YOLOv26 στο FASDD, η απόδοσή του στο testset ως προς τις γνωστές μας μετρικές σχημάτισε τον ακόλουθο πίνακα.

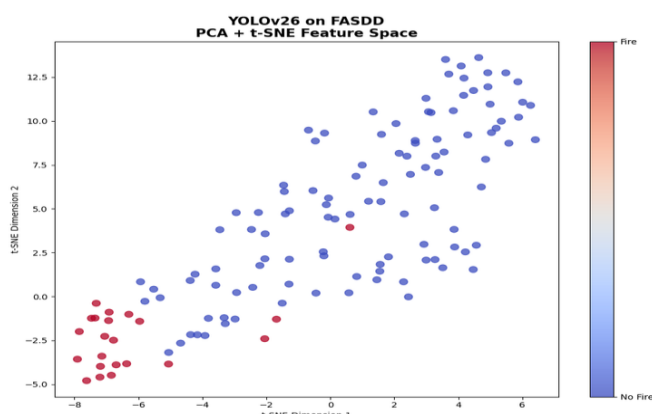
Πίνακας 3-14 Μετρικές του fine-tuned YOLOv26 στο FASDD

Fine Tuned YOLOv26 στο FASDD	
Accuracy	0.8397
Precision	1.000
Recall	0.0455
F1-score	0.0870



Εικόνα 3-34 Confusion Matrix του fine-tuned YOLOv26 στο FASDD

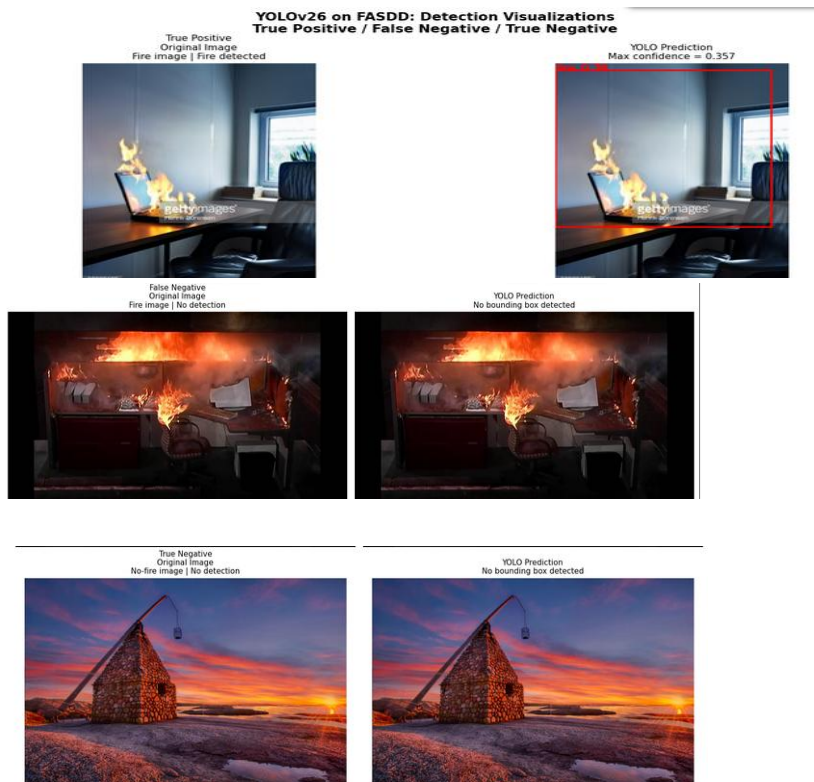
Μετά από τη διαδικασία fine-tuning, το YOLOv26 πέτυχε Accuracy 83,97%, Precision 100,00%, Recall 4,55% και F1-score 8,70% στο σύνολο δεδομένων FASDD. Από τη μια πλευρά το απόλυτο στην τιμή Precision (100%) σημαίνει ότι κάθε πρόβλεψη της κλάσης Fire ήταν σωστή. Όμως η τιμή αυτή θα πρέπει να διαβαστεί και να εκτιμηθεί σε συνάρτηση με το πολύ χαμηλό Recall (4,55%). Το χαμηλό Recall προκύπτει από το γεγονός ότι το μοντέλο πραγματοποίησε μόλις μία σωστή ανίχνευση πυρκαγιάς. Συνεπώς, η πολύ υψηλή Precision δεν συνδυάζεται με μια υψηλή συνολική απόδοση. Τελικά η ιδιαίτερα συντηρητική στρατηγική του μοντέλου, το οποίο αποφεύγει σχεδόν πλήρως να χαρακτηρίσει μια εικόνα ως Fire, την οποία στρατηγική παρατηρήσαμε ως συμπεριφορά και στο FLAME, είναι αυτή που οδηγεί το μοντέλο στο τόσο χαμηλό σκορ σωστών ανιχνεύσεων.



Εικόνα 3-35 t-SNE του fine-tuned Yolov26 στο FASDD

Το διάγραμμα t-SNE μας διαφωτίζει ως προς την οργάνωση του χώρου των χαρακτηριστικών. Παρατηρούμε ότι τα περισσότερα δείγματα της κλάσης No Fire σχηματίζουν μια εκτεταμένη συστάδα, ενώ τα δείγματα της κλάσης Fire εμφανίζονται συγκεντρωμένα σε μικρότερη περιοχή, με αρκετές όμως περιπτώσεις να βρίσκονται κοντά ή ακόμη και εντός της περιοχής της κλάσης No Fire. Αυτό μας επιβεβαιώνει ότι, παρότι το μοντέλο έχει μάθει σε μεγάλο βαθμό τα χαρακτηριστικά της κλάσης No Fire, δεν έχει επιτύχει αντίστοιχα διακριτή αναπαράσταση για τις εικόνες που περιέχουν πυρκαγιά.

Η φτωχή εν γένει απόδοση του μοντέλου αντανακλάται και στην ακόλουθη σειρά εικόνων στην οποία έχουμε τα detection visualizations από τρία instances. Βλέπουμε ότι στην μεν True Positive περίπτωση το μοντέλο εντοπίζει επιτυχώς την περιοχή της φωτιάς δημιουργώντας πλαίσιο ανίχνευσης (bounding box), αν και με σχετικά χαμηλή τιμή εμπιστοσύνης (confidence = 0.357). Παρότι η πρόβλεψη είναι σωστή, η χαμηλή εμπιστοσύνη υποδηλώνει ότι ακόμη και στις επιτυχημένες ανιχνεύσεις το μοντέλο εμφανίζει σημαντική επιφύλαξη. Στη δε περίπτωση του False Negative, είναι αξιοσημείωτο πως παρά την εμφανή παρουσία εκτεταμένων φλογών στην εικόνα, το μοντέλο δεν δημιουργεί κανένα πλαίσιο ανίχνευσης. Το αποτέλεσμα αυτό δείχνει ότι το μοντέλο αποτυγχάνει να ενεργοποιηθεί ακόμη και σε ορισμένες ιδιαίτερα εμφανείς περιπτώσεις πυρκαγιάς, γεγονός που εξηγεί τον μεγάλο αριθμό False Negatives που σημειώθηκαν στην αξιολόγηση.



Εικόνα 3-36 Χάρτες Grad-CAM του fine-tuned YOLOv26 στο FASDD

3.5 Η αξιολόγηση του προεκπαιδευμένου μοντέλου “Fire-Detection-Siglip2”.

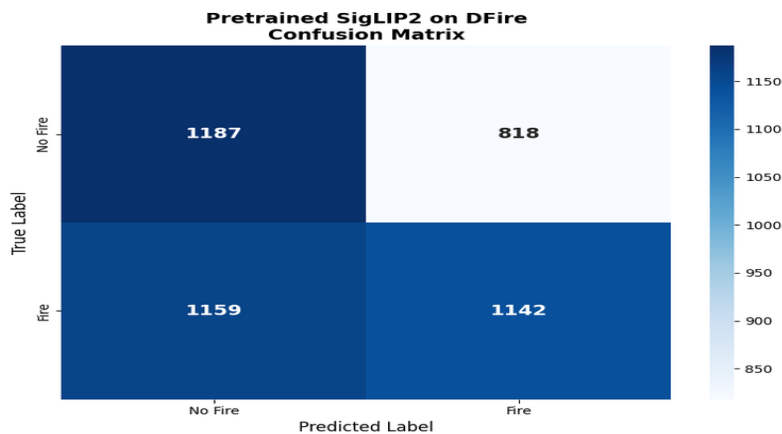
Στην παρούσα παράγραφο πρόκειται να ασχοληθούμε με το προεκπαιδευμένο μοντέλο “Fire-Detection-Siglip2” (στο εξής αναφερόμενο ως SigLIP2). Πρέπει σε αυτό το σημείο να αναφέρουμε πως το SigLIP2 δεν διαθέτει εξειδικευμένη κεφαλή ταξινόμησης για την δυαδική εργασία ταξινόμησης εικόνων σε κλάσεις Fire/No Fire. Επομένως, η αξιολόγησή του πραγματοποιήθηκε σε zero-shot πλαίσιο μέσω text prompts, συγκρίνοντας την ομοιότητα των image embeddings με τις λεκτικές περιγραφές των δύο κλάσεων. Προκειμένου να υπάρχει ομοιομορφία στις αξιολογήσεις του προεκπαιδευμένου μοντέλου (3.5.1- 3.5.3) χρησιμοποιήθηκαν σε όλες τις περιπτώσεις δύο απλά λεκτικά prompts: "a photo of fire" και "a photo without fire".

3.5.1 στο DFire

Για το προεκπαιδευμένο μοντέλο SigLIP2 τα αποτελέσματα των μετρικών Accuracy, Precision, Recall και F1-score στο Dfire συγκεντρώνονται στον ακόλουθο πίνακα

Πίνακας 3-15 Μετρικές του pretrained SigLIP2 στο DFire

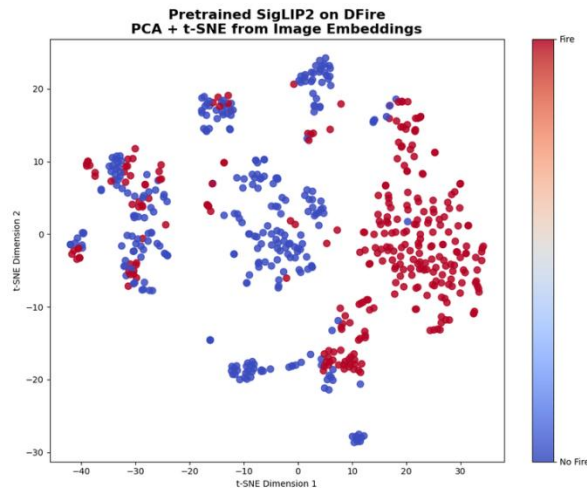
Pretrained SigLIP2 on DFire	
Accuracy	0.5409
Precision	0.5827
Recall	0.4963
F1-score	0.5360



Εικόνα 3-37 Confusion Matrix του pretrained SigLIP2 στο DFire

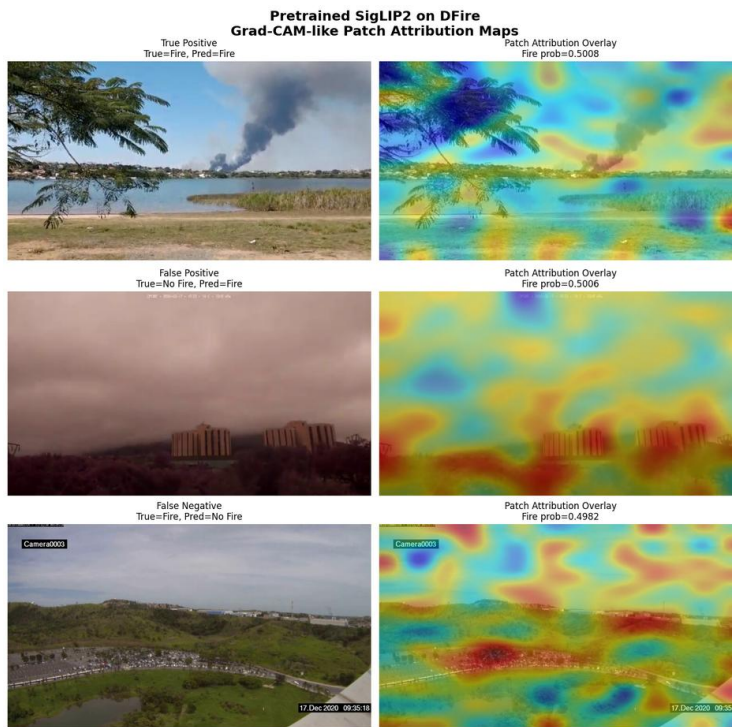
Η αξιολόγηση του προεκπαιδευμένου SigLIP2 στο DFire οδήγησε σε φτωχά αποτελέσματα, με Accuracy = 54,1% και F1-score = 53,6%. Το confusion matrix δείχνει σημαντική σύγχυση και προς τις δύο κατευθύνσεις, καθώς το μοντέλο παράγει 818 ψευδώς θετικές και 1.159 ψευδώς αρνητικές προβλέψεις. Το γεγονός αυτό υποδηλώνει ότι, χωρίς fine-tuning, το SigLIP2 δεν παρουσιάζει ένα αξιόπιστο υποψήφιο μοντέλο ως προς το να αποφασίζει ορθά για την παρουσία

φωτιάς στο DFire. Συνεπώς, η χρήση του προεκπαιδευμένου μοντέλου δεν επαρκεί για αξιόπιστη αναγνώριση πυρκαγιάς στο συγκεκριμένο dataset.



Εικόνα 3-38 t-SNE του pretrained SigLIP2 στο DFire

Από το t-SNE του προεκπαιδευμένου SigLIP2 στο DFire παρατηρούμε την ύπαρξη αρκετών διακριτών clusters που αντιστοιχούν στις κλάσεις Fire και No Fire. Το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο εξάγει χρήσιμες και διαχωρίσιμες αναπαραστάσεις των εικόνων. Ωστόσο, η σχετικά χαμηλή επίδοση που παρατηρήθηκε κατά την αξιολόγηση ως προς τις μετρικές ταξινόμησης δείχνει ότι η πληροφορία αυτή δεν μετατρέπεται αποτελεσματικά σε σωστές τελικές προβλέψεις μέσω των απλών text prompts που χρησιμοποιήθηκαν (στην περίπτωση μας όπως προείπαμε στην εισαγωγή της παραγράφου 3.5 έγινε χρήση μόνο δύο απλών prompts "a photo of fire" και "a photo without fire"). Συνεπώς, τα αποτελέσματα υποδεικνύουν ότι το βασικό περιοριστικό στοιχείο δεν είναι η ποιότητα των οπτικών χαρακτηριστικών αλλά η ευθυγράμμιση μεταξύ του οπτικού και του γλωσσικού χώρου αναπαράστασης κατά τη διαδικασία της αξιολόγησης του προεκπαιδευμένου μοντέλου.



Εικόνα 3-39 Χάρτες Grad-CAM του pretrained SigLIP2 στο DFire

Όπως φαίνεται και από τα Grad-CAM-like maps στο μοντέλο SigLIP2 παρουσιάζει ελλιπή κατανόηση των σκηνών φωτιάς. Αν και μπορεί να εντοπίσει ευρείες περιοχές που σχετίζονται με τη φωτιά, βασίζεται συχνά σε γενικά οπτικά χαρακτηριστικά, όπως η φωτεινότητα και η κατανομή των θερμών χρωμάτων. Αυτή η συμπεριφορά οδηγεί σε σημαντική επικάλυψη μεταξύ δειγμάτων με φωτιά και χωρίς φωτιά, με αποτέλεσμα πολύ μέτρια απόδοση ταξινόμησης στο σύνολο δεδομένων DFire.

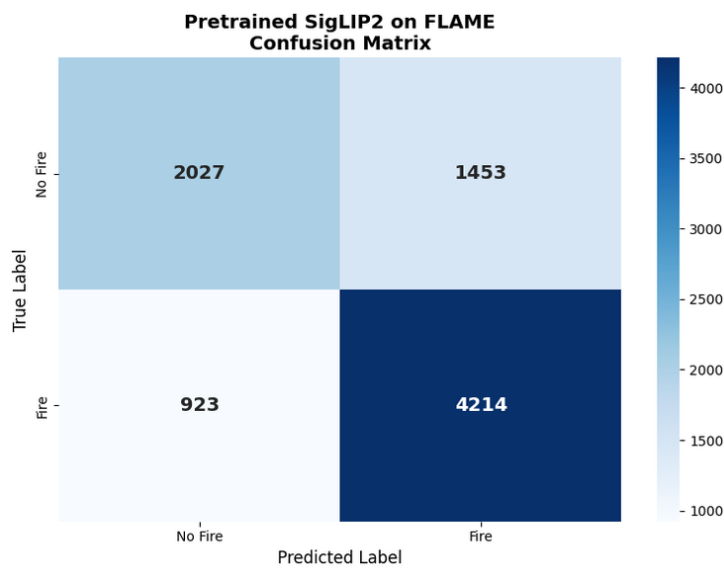
3.5.2 στο FLAME

Για το προεκπαιδευμένο μοντέλο SIGLIP2 τα αποτελέσματα των μετρικών Accuracy, Precision, Recall και F1-score στο FLAME δίνονται στον ακόλουθο πίνακα

Πίνακας 3-16 Μετρικές του pretrainedSigLIP2 στο FLAME

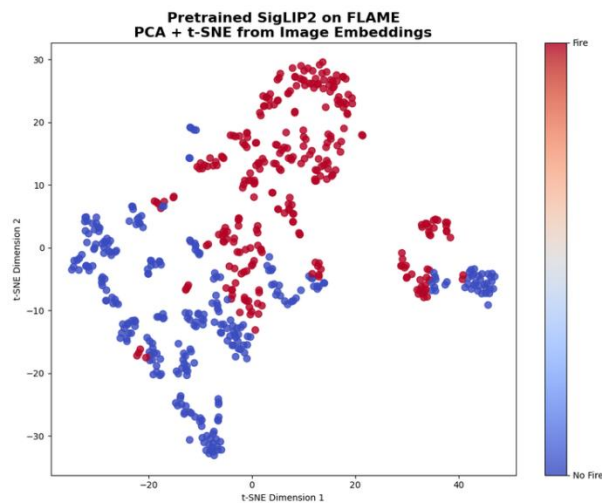
Pretrained SigLIP2 στο FLAME

Accuracy	0.7243
Precision	0.7436
Recall	0.8203
F1-score	0.7801



Εικόνα 3-40 Confusion Matrix του pretrained SigLIP2 στο FLAME

Η ανάγνωση του confusion matrix δείχνει ότι το SigLIP2 αναγνωρίζει αρκετά καλά τις εικόνες Fire του FLAME, με 4214 σωστές Fire προβλέψεις και σχετικά υψηλό Recall = 82.0%. Ωστόσο, υπάρχουν αρκετά False Positives (1453), πράγμα που δείχνει ότι το μοντέλο τείνει να χαρακτηρίζει ως Fire αρκετές εικόνες χωρίς φωτιά, πιθανώς επειδή περιέχουν οπτικά στοιχεία που μοιάζουν με καπνό, θερμές αποχρώσεις ή φωτεινές περιοχές.

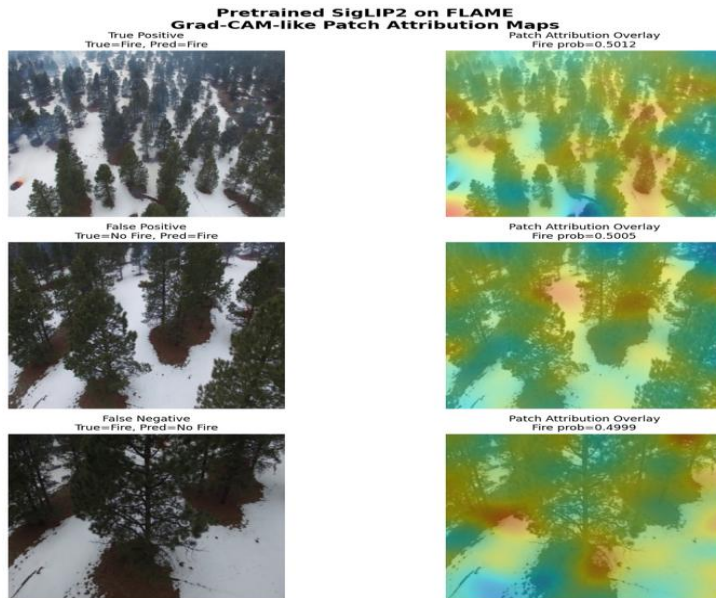


Εικόνα 3-41 t-SNE του pretrained SigLIP2 στο FLAME

Όπως φαίνεται από το t-SNE διάγραμμα οι δύο κατηγορίες σχηματίζουν διακριτά clusters με περιορισμένη μόνο επικάλυψη στο κέντρο, γεγονός που υποδηλώνει ότι ο vision encoder εξάγει σχετικά διακριτές αναπαραστάσεις για τις εικόνες του FLAME. Η παρατηρούμενη δομή του χώρου χαρακτηριστικών είναι συνεπής με τη σημαντικά υψηλότερη ταξινομητική επίδοση του μοντέλου (Accuracy 72.4%, F1-score 78.0%) σε σχέση με αυτή του ίδιου προεκπαιδευμένου μοντέλου στο DFire. Οι περιοχές στις οποίες οι δύο κλάσεις αλληλοεπικαλύπτονται αντιστοιχούν στα σφάλματα που καταγράφονται στον confusion matrix (1453 False Positives και 923 False Negatives). Παρ'όλα αυτά συνολικά το embedding space εμφανίζει σχετικά καλή διαχωριστική ικανότητα, γεγονός που σηματοδοτεί ότι οι εικόνες του FLAME είναι περισσότερο ευθυγραμμισμένες με τις γενικές οπτικές έννοιες που έχει μάθει το προεκπαιδευμένο SigLIP2.

Τα Grad-CAM-like patch attribution maps του προεκπαιδευμένου SigLIP2 στο FLAME δείχνουν ότι το μοντέλο δεν εστιάζει αποκλειστικά στις περιοχές όπου εμφανίζεται η φωτιά, αλλά κατανέμει την προσοχή του σε ένα μεγάλο μέρος της εικόνας. Οι ενεργοποιήσεις είναι περισσότερο διάχυτες και αποτυπώνουν μια ολιστική αναπαράσταση της σκηνής, γεγονός που είναι χαρακτηριστικό των vision-language transformer μοντέλων. Παρατηρούμε ότι το False Negative παρουσιάζει παρόμοια χωρική κατανομή ενεργοποιήσεων με το True Positive,

γεγονός που υποδηλώνει ότι οι λανθασμένες αποφάσεις οφείλονται μάλλον περισσότερο στις πολύ μικρές διαφορές της ομοιότητας μεταξύ των image και text embeddings παρά στην αποτυχία εξαγωγής οπτικών χαρακτηριστικών.



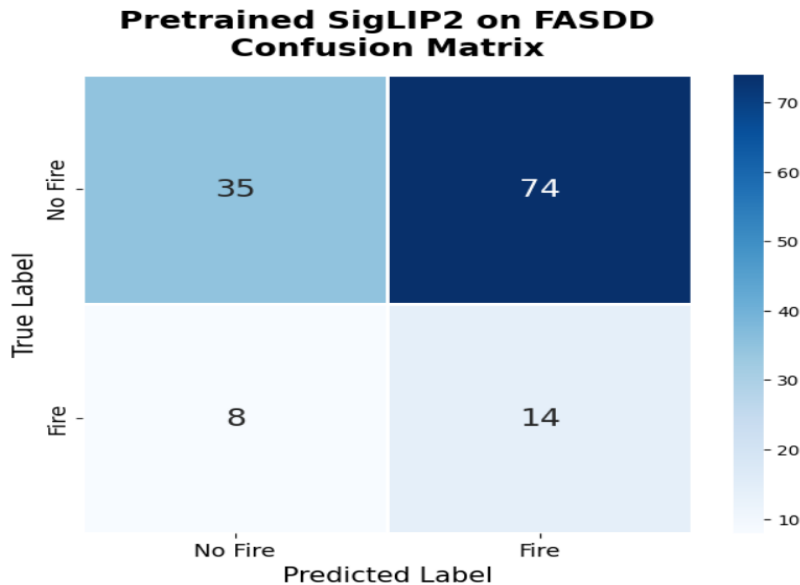
Εικόνα 3-42 Χάρτες Grad-CAM του pretrained SigLIP2 στο FLAME

3.5.3 στο FASDD

Για το προεκπαιδευμένο μοντέλο SigLIP2 τα αποτελέσματα των μετρικών Accuracy, Precision, Recall και F1-score στο FASDD

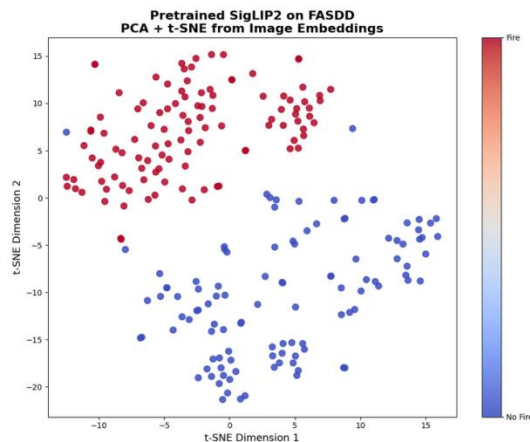
Πίνακας 3-17 Μετρικές του pretrained SigLIP2 στο FASDD

Pretrained SigLIP2 στο FASDD	
Accuracy	0.3740
Precision	0.1591
Recall	0.6364
F1-score	0.2545



Εικόνα 3-43 Confusion Matrix του pretrained SigLIP2 στο FASDD

Η αξιολόγηση του προεκπαιδευμένου SigLIP2 στο FASDD οδήγησε σε χαμηλή συνολική απόδοση, με Accuracy = 37,4% και F1-score = 25,5%. Χαρακτηριστικά παρατηρούμε ότι το μοντέλο προβλέπει πολύ συχνά την κλάση Fire, καθώς 74 από τις 109 εικόνες No Fire ταξινομήθηκαν λανθασμένα ως Fire, αριθμός που είναι πολύ μεγάλος για validation set 131 εικόνων. Παρότι το Recall για την κλάση Fire ήταν σχετικά υψηλότερο (63,6%), η Precision παρέμεινε ιδιαίτερα χαμηλή (15,9%), γεγονός που εξηγείται από τον μεγάλο αριθμό ψευδών συναγερμών. Επιπλέον, οι πιθανότητες της κλάσης Fire κυμάνθηκαν σε εξαιρετικά μικρό εύρος γύρω από το 0,50, στοιχείο που δείχνει ότι το μοντέλο δεν σχηματίζει ισχυρή απόφαση μεταξύ των δύο prompts. Τα αποτελέσματα αυτά υποδεικνύουν ότι, χωρίς περαιτέρω fine-tuning, το SigLIP2 δεν αξιοποιεί αποτελεσματικά τις οπτικές αναπαραστάσεις του για αξιόπιστη διάκριση Fire/No Fire στο FASDD.



Εικόνα 3-44 t-SNE του pretrained SigLIP2 στο FASDD

Το διάγραμμα t-SNE των image embeddings του προεκπαιδευμένου SigLIP2 στο FASDD παρουσιάζει έναν αρκετά καθαρό διαχωρισμό μεταξύ των εικόνων Fire και No Fire.. Η εικόνα αυτή μας δείχνει ότι ο vision encoder του SigLIP2 εξάγει διακριτές οπτικές αναπαραστάσεις ακόμη και για το μικρότερο και περισσότερο ανισόρροπο σύνολο δεδομένων FASDD. Ο σχετικά καλός διαχωρισμός των embeddings φαίνεται να έρχεται σε αντίθεση με τη συνολική απόδοση του μοντέλου η οποία είναι χαμηλή (Accuracy = 37.4%, F1-score = 25.5%). Το γεγονός αυτό υποδηλώνει ότι το πρόβλημα δεν βρίσκεται στην ποιότητα των image embeddings, αλλά στο τελικό στάδιο της ταξινόμησης, όπου κάθε embedding συγκρίνεται αποκλειστικά με τα δύο κείμενα-prompts ("a photo without fire" και "a photo of fire"). Όπως έδειξαν και οι πιθανότητες πρόβλεψης, οι τιμές similarity μεταξύ εικόνων και prompts διαφέρουν ελάχιστα και βρίσκονται πολύ κοντά στο 0.50 (TP=0.5003, FP = 0.5014, TN = 0.4997) με αποτέλεσμα ακόμη και μικρές μεταβολές να οδηγούν σε λανθασμένες ταξινομήσεις. Συνεπώς, το t-SNE δείχνει ότι ο vision encoder έχει μάθει μια χρήσιμη αναπαράσταση των εικόνων, όμως η απλή αντιστοίχιση με δύο γενικά prompts δεν είναι επαρκής για να αξιοποιήσει πλήρως αυτή την πληροφορία στο FASDD.

Τα Grad-CAM-like patch attribution maps του προεκπαιδευμένου SigLIP2 στο FASDD δείχνουν ότι το μοντέλο βασίζει την απόφασή του σε μια ολιστική αναπαράσταση της εικόνας και όχι αποκλειστικά στις περιοχές όπου εμφανίζονται οι φλόγες. Στο παράδειγμα True Positive οι ενεργοποιήσεις επικεντρώνονται τόσο

στη φωτιά όσο και στα υπόλοιπα στοιχεία της σκηνής, ενώ στο False Positive το μοντέλο αποδίδει αυξημένη σημασία σε φωτεινές επιφάνειες, όπως το παράθυρο και το φωτιστικό, οδηγώντας σε λανθασμένη πρόβλεψη της κλάσης Fire. Αντίθετα, στο True Negative οι ενεργοποιήσεις παραμένουν περισσότερο διάχυτες και χαμηλότερης έντασης, οδηγώντας στη σωστή ταξινόμηση της εικόνας ως No Fire. Τα αποτελέσματα υποδεικνύουν ότι οι λανθασμένες ταξινομήσεις οφείλονται κυρίως στην οριακή διαδικασία αντιστοίχισης εικόνας-κειμένου και όχι σε αποτυχία του vision encoder να εξαγάγει χρήσιμα οπτικά χαρακτηριστικά



Εικόνα 3-45 Χάρτες Grad-CAM του pretrained SigLIP2 στο FASDD

3.6 Fine tuning του μοντέλου SigLIP2

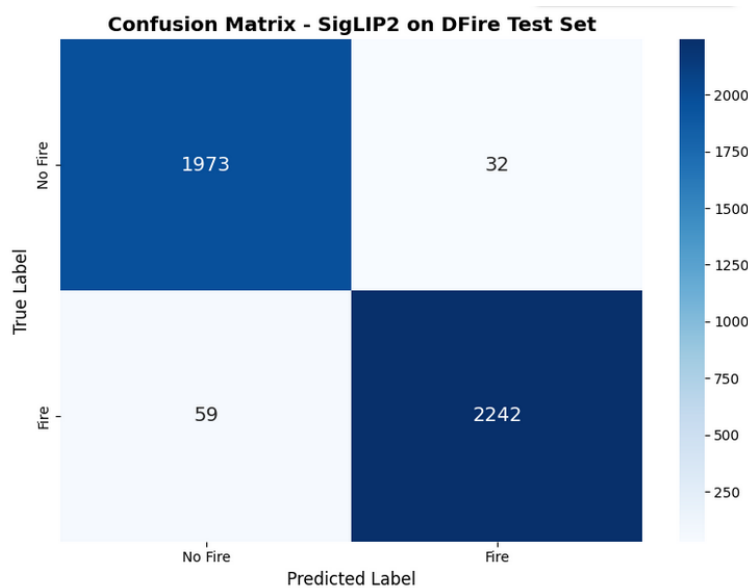
Στην παρούσα παράγραφο προχωρούμε στην εκπαίδευση του μοντέλου SigLIP2 στα τρία σύνολα δεδομένων. Σημειώνουμε ότι το fine-tuned SigLIP2 αξιολογείται ως δυαδικός ταξινομητής εικόνων μέσω ενός task-specific classification head.

3.6.1 στο DFire

Ξεκινάμε με την εκπαίδευση του μοντέλου Siglip2 στο DFire. Μετά και την αξιολόγησή του στο testset κατεγράφησαν οι ακόλουθες μετρικές.

Πίνακας 3-18 Μετρικές του fine-tuned SigLIP2 στο DFire

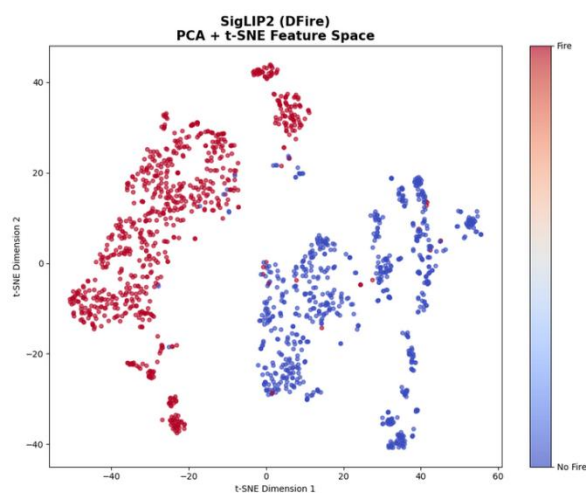
Fine Tuned SigLIP2 στο DFire	
Accuracy	0,9789
Precision	0,9859
Recall	0,9744
F1-score	0.9801



Εικόνα 3-46 Confusion Matrix του fine-tuned SigLIP2 στο DFire

Η διαδικασία fine-tuning, αποδεικνύεται ιδιαίτερα ενισχυτική για το SigLIP2. Μετά το finetuning το μοντέλο κατέγραψε Accuracy 97,89%, Precision 98,59%, Recall 97,44% και F1-score 98,01% στο σύνολο δεδομένων DFire. Γενικά παρουσιάζει μια εξαιρετική συνολική απόδοση στην αναγνώριση εικόνων με και χωρίς πυρκαγιά. Το finetuned SigLIP2 εμφανίζει εξαιρετική ισορροπία μεταξύ Precision και Recall, γεγονός που σημαίνει ότι όχι μόνο αποφεύγει τους ψευδείς συναγερμούς, αλλά ταυτόχρονα ανιχνεύει σχεδόν όλες τις πραγματικές περιπτώσεις πυρκαγιάς. Ο confusion matrix μας δείχνει ότι από τις 2.005 εικόνες

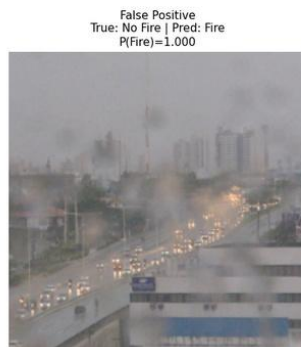
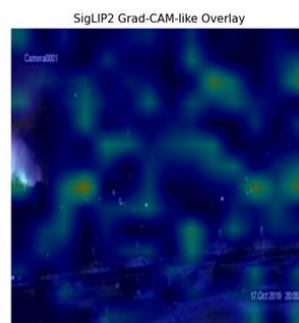
χωρίς πυρκαγιά, οι 1.973 ταξινομήθηκαν σωστά, ενώ μόλις 32 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Αντίστοιχα, από τις 2.301 εικόνες με πυρκαγιά, οι 2.242 αναγνωρίστηκαν σωστά, ενώ μόνο 59 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Ο μικρός αριθμός σφαλμάτων και στις δύο κλάσεις αποδεικνύει ότι το μοντέλο επιτυγχάνει ιδιαίτερα αξιόπιστη και ισορροπημένη ταξινόμηση.



Εικόνα 3-47 t-SNE του fine-tuned SigLIP2 στοDFire

Στο διάγραμμα t-SNE παρατηρούμε ένα σαφή διαχωρισμό μεταξύ των δύο κλάσεων (Fire και No Fire), με τα δείγματα να οργανώνονται σε δύο ευδιάκριτες και σχετικά συμπαγείς περιοχές του χώρου χαρακτηριστικών. Η επικάλυψη μεταξύ των δύο κλάσεων είναι ελάχιστη. Ένα πολύ ενδιαφέρον στοιχείο που παρατηρούμε στο t-SNE είναι ότι και οι δύο κύριες κλάσεις χωρίζονται σε μικρότερες υποσυστάδες. Αυτό δείχνει ότι το SigLIP2 δεν αντιμετωπίζει όλες τις εικόνες Fire ή No Fire ως ομοιογενείς ομάδες, αλλά μαθαίνει διαφορετικές υποκατηγορίες μέσα σε κάθε κλάση. Οι υποσυστάδες αυτές θα μπορούσαμε να εικαστούμε ότι αντιστοιχούν σε διαφορετικούς τύπους σκηνών, όπως διαφορετικές εντάσεις στη φωτιά, διαφορετική ποσότητα καπνού ή συνθήκες φωτισμού. Η ικανότητα αυτή του μοντέλου να κατανοεί και τα διαφορετικά μοτίβα πυρκαγιάς αποτελεί ισχυρή ένδειξη πλούσιων και καλά οργανωμένων αναπαραστάσεων χαρακτηριστικών.

Στην ακόλουθη σειρά των χαρτών Grad-CAM βλέπουμε πως στην περίπτωση του True Positive, το μοντέλο ενεργοποιείται κυρίως στην περιοχή όπου εμφανίζεται η φωτιά και ο καπνός, αποδίδοντας πιθανότητα $P(\text{Fire})=1.000$. Παρότι οι ενεργοποιήσεις είναι σχετικά διάχυτες, η μεγαλύτερη ένταση συγκεντρώνεται στην περιοχή της πυρκαγιάς. Ως προς την περίπτωση του False Positive, το μοντέλο αποδίδει επίσης πιθανότητα $P(\text{Fire})=1.000$, παρότι η εικόνα δεν περιέχει πραγματική πυρκαγιά. Οι ενεργοποιήσεις κατανέμονται σε περιοχές με έντονη θόλωση και μειωμένη ορατότητα, γεγονός που υποδηλώνει ότι το μοντέλο συγχέει οπτικά χαρακτηριστικά, όπως η ομίχλη ή βροχή, με τον καπνό που συνοδεύει μια πυρκαγιά. Τέλος στην περίπτωση του False Negative, όπου η εικόνα περιέχει καπνό αλλά όχι εμφανείς φλόγες, το μοντέλο αποδίδει πιθανότητα $P(\text{Fire})=0.000$ και οι ενεργοποιήσεις εμφανίζονται ιδιαίτερα ασθενείς και διάσπαρτες. Η συμπεριφορά αυτή δείχνει ότι το μοντέλο δυσκολεύεται να αναγνωρίσει περιπτώσεις στις οποίες η πυρκαγιά εκδηλώνεται κυρίως μέσω του καπνού, χωρίς έντονα χαρακτηριστικά φλόγας.





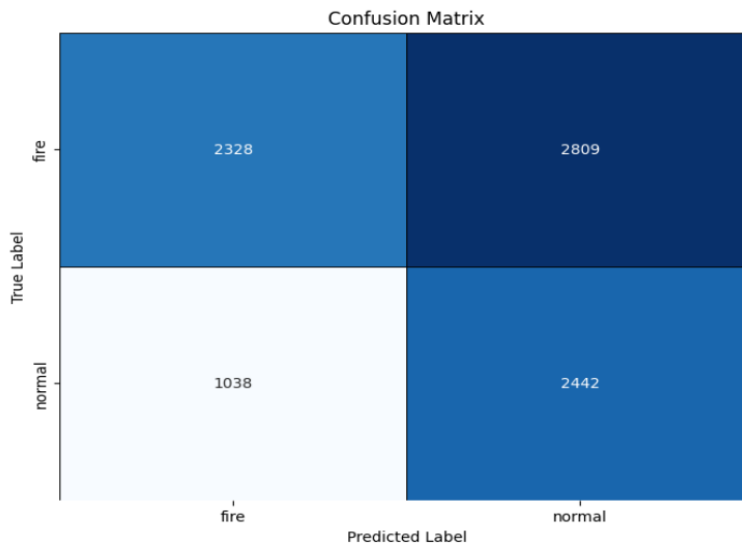
Εικόνα 3-48 Χάρτες Grad-CAM του fine-tuned SigLIP2 στο DFire

3.6.2 στο FLAME

Συνεχίζουμε με την εκπαίδευση του μοντέλου SigLIP2 στο FLAME. Μετά το πέρας της εκπαίδευσης και της δοκιμής του μοντέλου στο test set προέκυψαν οι ακόλουθες μετρικές.

Πίνακας 3-19 Μετρικές του fine-tuned SigLIP2 στο FLAME

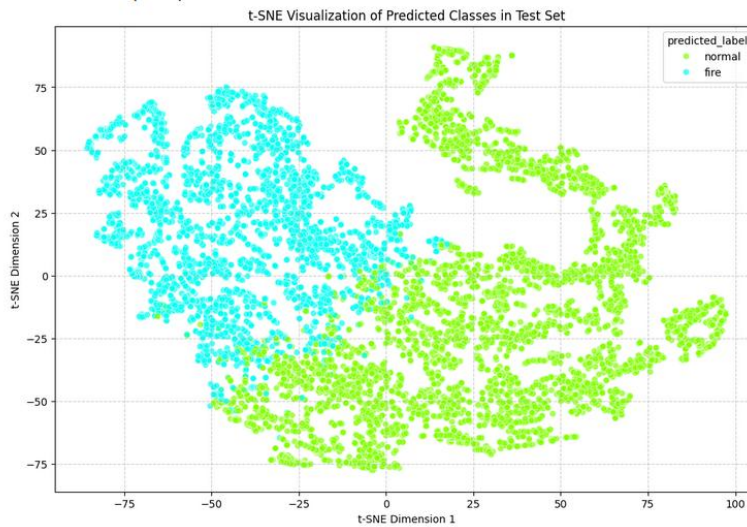
Fine Tuned SigLIP2 στο FLAME	
Accuracy	0,5536
Precision	0,6916
Recall	0,4532
F1-score	0.5746



Εικόνα 3-49 Confusion Matrix του fine-tuned SigLIP2 στοFLAME

Οι τιμές στον πίνακα των μετρικών δείχνουν ότι το μοντέλο παρουσιάζει περιορισμένη ικανότητα διάκρισης μεταξύ των εικόνων με και χωρίς πυρκαγιά, επιτυγχάνοντας απόδοση που υπερβαίνει οριακά την τυχαία ταξινόμηση.

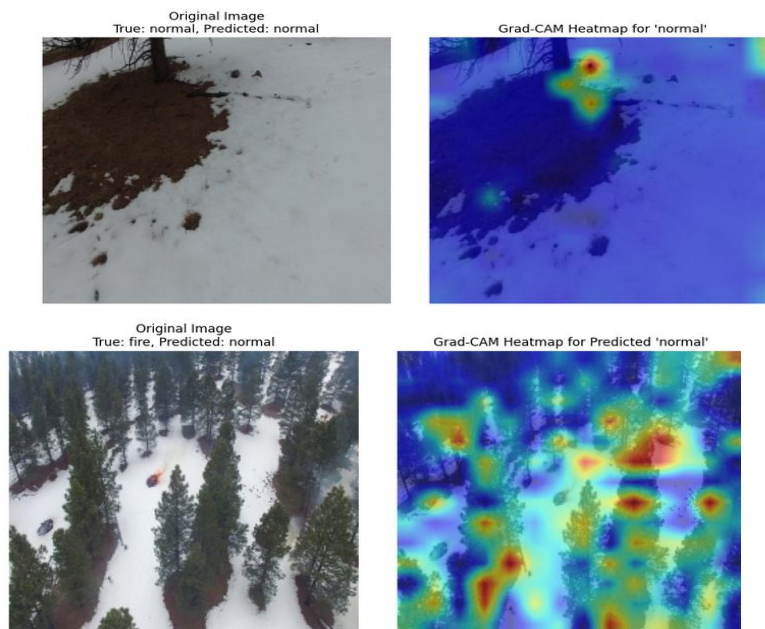
Σημειώνουμε στο σημείο αυτό ότι στον confusionmatrix που παρατίθεται παραπάνω με το labelNormal νοείται η κλάση NoFire. Με βάση τον confusionmatrix φαίνεται ότι το μοντέλο έχει ένα υψηλό αριθμό λανθασμένων αναγνωρίσεων εικόνων που δεν περιέχουν φωτιά. Το τελευταίο αυτό σημείο μπορεί να καταδεικνύει την ανάγκη αύξησης του αριθμού των εικόνων χωρίς φωτιά στην φάση της εκπαίδευσης του μοντέλου. Συγκεκριμένα, από τις 5.137 εικόνες με πυρκαγιά, οι 2.328 ταξινομήθηκαν σωστά, ενώ 2.809 ταξινομήθηκαν λανθασμένα ως Normal (False Negatives). Αντίστοιχα, από τις 3.480 εικόνες χωρίς πυρκαγιά, οι 2.442 αναγνωρίστηκαν σωστά, ενώ 1.038 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Το μοντέλο παρουσιάζει σημαντικό αριθμό σφαλμάτων και στις δύο κατηγορίες, γεγονός που εξηγεί τις σχετικά χαμηλές τιμές όλων των μετρικών αξιολόγησης.



Εικόνα 3-50 t-SNE του fine-tuned SigLIP2 στοFLAME

Από το παραπάνω διάγραμμα διασποράς t-SNE καθίσταται εμφανής η διάκριση δύο μεγάλων συστάδων. Η διάκριση όμως αυτή δεν είναι τελείως ξεκάθαρη αφού φαίνεται στο t-SNE διάγραμμα ότι υπάρχει μια κεντρική περιοχή όπου οι κλάσεις εμφανίζονται συγκεχυμένες και όχι σαφώς διαχωρισμένες, κάτι που αντανακλάται και στα αποτελέσματα των μετρικών ακρίβειας. Πιο συγκεκριμένα η αισθητή επικάλυψη μεταξύ των δύο κλάσεων που φαίνεται στο κέντρο του διαγράμματος μας δείχνει ότι αρκετά δείγματα παρουσιάζουν παρόμοιες αναπαραστάσεις χαρακτηριστικών. Η επικάλυψη αυτή δυσκολεύει τη σαφή διάκριση μεταξύ εικόνων με και χωρίς πυρκαγιά. Επίσης παρατηρούμε ότι οι περισσότερες λανθασμένες ταξινομήσεις συγκεντρώνονται στη μεταβατική περιοχή μεταξύ των δύο κύριων συστάδων και όχι στο εσωτερικό τους. Αυτό σημαίνει ότι το μοντέλο διαχωρίζει αποτελεσματικά τις «τυπικές» εικόνες των δύο κλάσεων, αλλά δυσκολεύεται στις εικόνες που παρουσιάζουν αμφίσημα ή κοινά οπτικά χαρακτηριστικά. Η συμπεριφορά αυτή συμφωνεί και με τα ευρήματα του confusion matrix, όπου παρατηρείται μεγάλος αριθμός τόσο False Positives όσο και False Negatives, χωρίς εμφανή μεροληψία προς κάποια από τις δύο κλάσεις. Στην πρώτη σειρά από τις εικόνες με τους Grad-CAM χάρτες που ακολουθούν η εικόνα δεν περιέχει φωτιά και ορθά εκτιμάται από το μοντέλο ως τέτοια ενώ στο θερμικό χάρτη φαίνονται οι περιοχές της εικόνας όπου το μοντέλο εστίασε την

προσοχή του για να λάβει την απόφασή του. Παρατηρούμε ότι μια κύρια περιοχή εστίασης του ενδιαφέροντος του μοντέλου είναι η περιοχή κοντά στη βάση του δέντρου. Στο επόμενο σετ εικόνων, της πρωτότυπης και του θερμικού χάρτη της όπως αυτός παράγεται από το Grad-CAM, βλέπουμε μια εικόνα στην οποία υπάρχει μια μικρή εστία φωτιάς σε ένα άνοιγμα μεταξύ των δέντρων. Την εστία αυτή δεν μπόρεσε να εντοπίσει το μοντέλο κατηγοριοποιώντας λάθος την εικόνα αυτή ως Normal. Στο heatmap δεξιά βλέπουμε ότι το μοντέλο εστίασε την προσοχή του στις περιοχές των δέντρων και όχι στην περιοχή ανάμεσα από αυτά στην οποία βρίσκεται η μικρή εστία φωτιάς.



Εικόνα 3-51 Χάρτες Grad-CAM του fine-tuned SigLIP2 στο FLAME

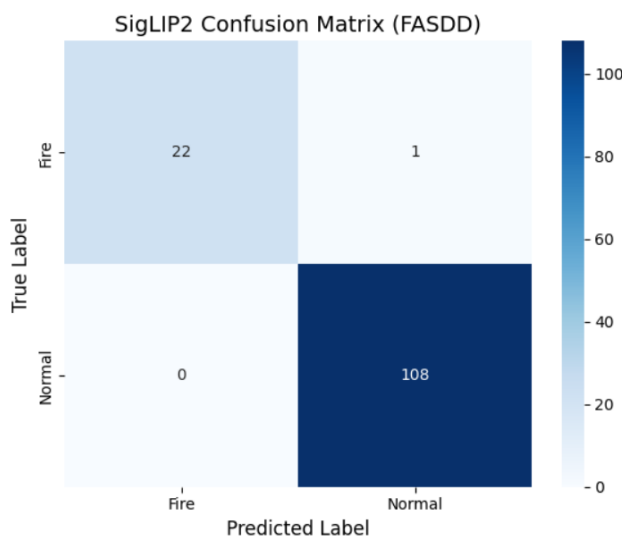
3.6.3 στο FASDD

Μετά το finetuning του μοντέλου SigLIP2 στο σύνολο FASDD και την αξιολόγηση του μοντέλου στο test set προέκυψαν τα ακόλουθα αποτελέσματα.

Πίνακας 3-20 Μετρικές του fine-tuned SigLIP2 στο FASDD

Fine Tuned SigLIP2 στο FASDD

Accuracy	0,9924
Precision	0,9954
Recall	0,9783
F1-score	0.9866

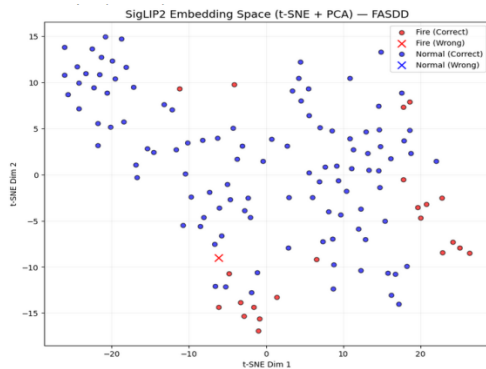


Εικόνα 3-52 Confusion Matrix του fine-tuned SigLIP2 στο FASDD

Τα αποτελέσματα του πίνακα των μετρικών καθώς και αυτά στο confusionmatrix συνηγορούν στο ότι το finetunedSigLIP2 προσαρμόστηκε εξαιρετικά στις αναπαραστάσεις των εικόνων του FASDD. Ο confusionmatrix επικυρώνει την εξαιρετική αυτή συμπεριφορά. Συγκεκριμένα, από τις 23 εικόνες με πυρκαγιά, οι 22 ταξινομήθηκαν σωστά, ενώ μόνο μία εικόνα ταξινομήθηκε λανθασμένα ως Normal (False Negative). Παράλληλα, και οι 108 εικόνες χωρίς πυρκαγιά αναγνωρίστηκαν σωστά, χωρίς καμία ψευδώς θετική πρόβλεψη (False Positives = 0). Τα αποτελέσματα αυτό δείχνουν ότι το μοντέλο διακρίνει με πολύ μεγάλη αξιοπιστία τις δύο κατηγορίες Fire και Normal, διατηρώντας ταυτόχρονα εξαιρετικά χαμηλό ποσοστό σφαλμάτων.

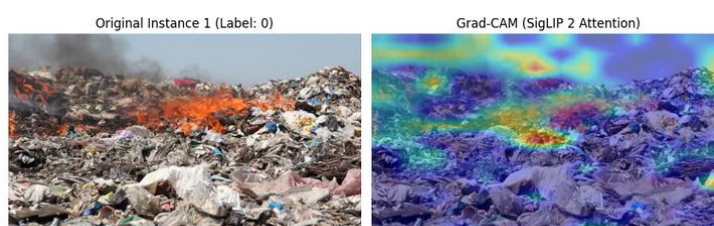
Παρατηρώντας το διάγραμμα t-SNE παρακάτω βλέπουμε πως η μεγάλη πλειονότητα των δειγμάτων ταξινομείται σωστά, γεγονός που αποτυπώνεται από

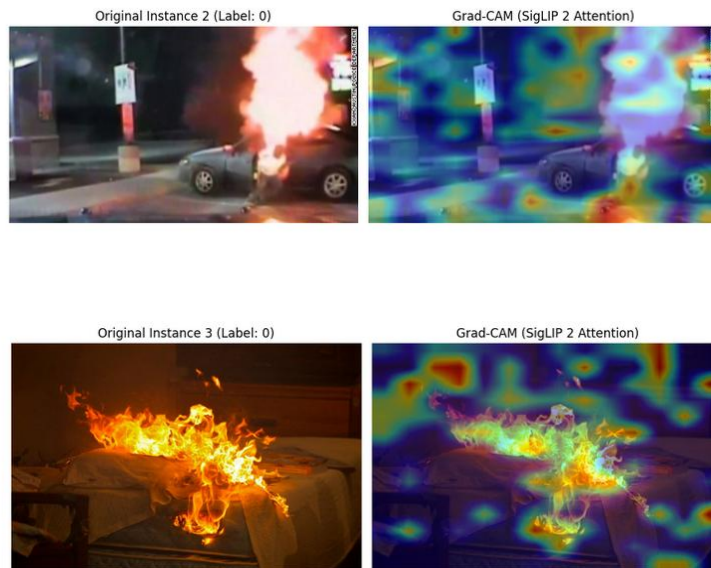
την επικράτηση των κυκλικών σημείων (σωστές ταξινομήσεις), ενώ εμφανίζεται μόλις μία λανθασμένη ταξινόμηση (σημειωμένη με σύμβολο X). Η μοναδική λανθασμένη ταξινόμηση εντοπίζεται κοντά στα όρια των δύο περιοχών του χώρου χαρακτηριστικών, γεγονός που μας κάνει να σκεφτούμε ότι μάλλον πρόκειται για δείγμα με χαρακτηριστικά που παρουσιάζουν ομοιότητες και με τις δύο κλάσεις, Αξίζει επίσης να σημειώσουμε ότι στο διάγραμμα t-SNE δεν παρατηρείται μόνο διαχωρισμός των δύο κλάσεων, αλλά και πολύ μικρός αριθμός σημείων κοντά στη μεταβατική περιοχή όπου οι δύο συστάδες παρουσιάζουν τη μεγαλύτερη εγγύτητα στον χώρο χαρακτηριστικών. Αυτή η «γεωμετρική αναπαράσταση» απηχεί το γεγονός ότι το μοντέλο παρουσίασε σχεδόν μηδενικό αριθμό λαθών στον confusionmatrix.



Εικόνα 3-53 t-SNE του fine-tuned SigLIP2 στοFASDD

Στα ακόλουθα τρία ζεύγη χαρτών Grad-CAMβλέπουμε τρεις περιπτώσεις TruePositives. Από τους θερμικούς χάρτες φαίνεται πως το SigLIP2 εστιάζει με συνέπεια στα πραγματικά χαρακτηριστικά της πυρκαγιάς και δεν βασίζεται σε άσχετα στοιχεία του περιβάλλοντος για τη λήψη των αποφάσεών του. Οι ισχυρότερες ενεργοποιήσεις εντοπίζονται στις περιοχές όπου εμφανίζονται οι φλόγες και ο καπνός.





Εικόνα 3-54 Χάρτες Grad-CAM του fine-tuned SigLIP2 στο FASDD

3.7 Το προεκπαιδευμένο μοντέλο “EdBianchiVit-fire-detection”

Το μοντέλο που χρησιμοποιούμε σε αυτή την παράγραφο είναι το "EdBianchi/vit-fire-detection" (στα ακόλουθα αναφερόμαστε σε αυτό ως EdBianchiVit). Είναι και αυτό ένα μοντέλο Οπτικού Κωδικοποιητή χρήσιμο για εργασίες αναγνώρισης εικόνας. Συγκεκριμένα το μοντέλο αυτό έχει εκπαιδευθεί για αναγνώριση εικόνων φωτιάς. Στην παρούσα παράγραφο ακολουθεί η πειραματική δοκιμή του προεκπαιδευμένου μοντέλου στα τρία σύνολα δεδομένων.

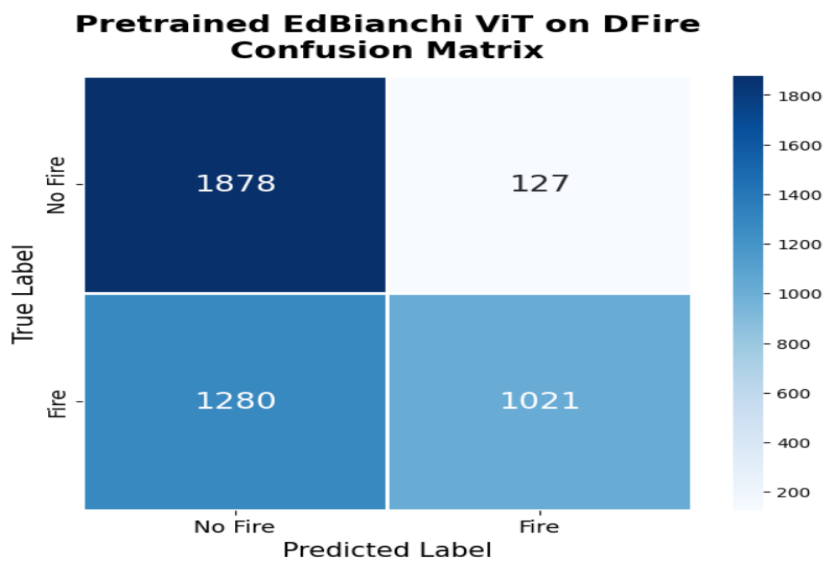
3.7.1 στο DFire

Από την αξιολόγηση του προεκπαιδευμένου EdBianchiVit στο σύνολο Dfire πήραμε τα ακόλουθα αποτελέσματα

Πίνακας 3-21 Μετρικές του pretrained EdBianchiVit στο DFire

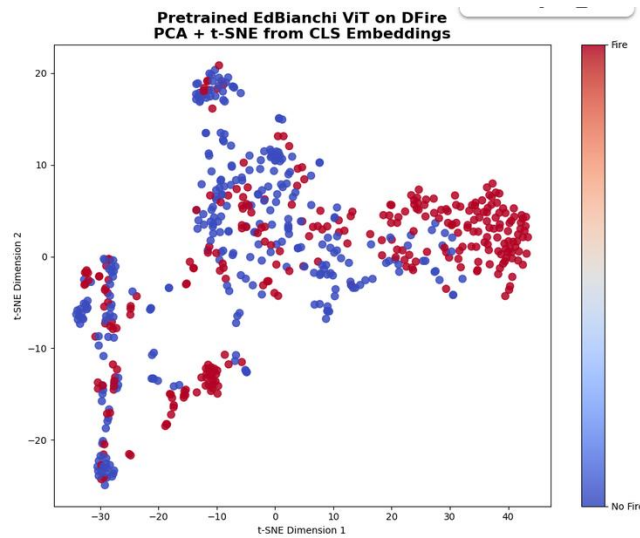
Pretrained EdBianchiVit στο DFire

Accuracy	0.6732
Precision	0.8894
Recall	0.4437
F1-score	0.5921



Εικόνα 3-55 Confusion Matrix του pretrained EdBianchi Vit στο DFire

Η αξιολόγηση του προεκπαιδευμένου ViT Fire Detection στο DFire έδειξε συνολική Accuracy = 67,3% και F1-score = 59,2%. Το μοντέλο παρουσίασε ιδιαίτερα υψηλή Precision για την κλάση Fire (88,9%), γεγονός που υποδηλώνει ότι οι προβλέψεις Fire είναι κατά κανόνα αξιόπιστες. Ωστόσο, το Recall παρέμεινε χαμηλό (44,4%), καθώς 1.280 εικόνες Fire ταξινομήθηκαν λανθασμένα ως No Fire. Η συμπεριφορά αυτή δείχνει ότι το μοντέλο λειτουργεί συντηρητικά, αποφεύγοντας τους ψευδείς συναγερμούς αλλά από την άλλη πλευρά χάνει σημαντικό αριθμό πραγματικών περιπτώσεων πυρκαγιάς.

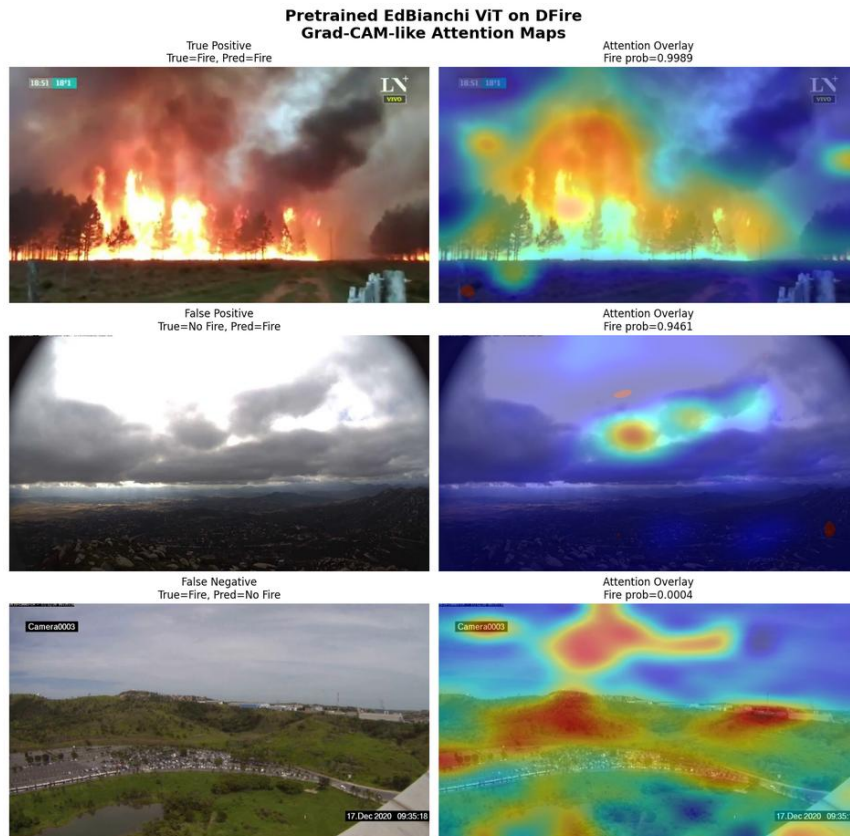


Εικόνα 3-56 t-SNE του pretrained EdBianchi Vit στο DFire

Το διάγραμμα t-SNE του προεκπαιδευμένου EdBianchi ViT στο DFire παρουσιάζει έναν μερικό αλλά σαφή διαχωρισμό μεταξύ των εικόνων Fire και No Fire. Παράλληλα, στο κεντρικό τμήμα του χώρου παρατηρείται σημαντική επικάλυψη μεταξύ των δύο κλάσεων, γεγονός που υποδηλώνει ότι ορισμένες εικόνες διαθέτουν παρόμοια χαρακτηριστικά ανεξάρτητα από την πραγματική τους ετικέτα. Η εικόνα αυτή συμφωνεί με τις ποσοτικές μετρικές του μοντέλου. Η ύπαρξη ενός αρκετά διακριτού cluster για την κλάση Fire εξηγεί την υψηλή Precision (88.9%), καθώς όταν μια εικόνα τοποθετείται μέσα σε αυτή την περιοχή, το μοντέλο την ταξινομεί με μεγάλη αξιοπιστία ως Fire. Ωστόσο, η σημαντική επικάλυψη των δύο κλάσεων στο κεντρικό τμήμα του χώρου χαρακτηριστικών οδηγεί πολλές πραγματικές εικόνες Fire να βρίσκονται κοντά στις αναπαραστάσεις της κλάσης No Fire, γεγονός που εξηγεί το σχετικά χαμηλό Recall (44.4%) και τον μεγάλο αριθμό False Negatives που παρατηρήθηκε στο confusion matrix.

Τα Grad-CAM-like attention maps δείχνουν ότι το μοντέλο εντοπίζει αποτελεσματικά έντονες πυρκαγιές, όπως φαίνεται στο True Positive παράδειγμα όπου η ενεργοποίηση συμπίπτει με τις φλόγες και τον καπνό. Ωστόσο, στο False Positive παράδειγμα το μοντέλο αποδίδει υψηλή πιθανότητα Fire σε σκοτεινά νέφη και φωτεινές περιοχές του ουρανού, υποδηλώνοντας πιθανή σύγχυση μεταξύ καπνού και νεφώσεων. Αντίθετα, στο False Negative παράδειγμα η προσοχή διαχέεται στο τοπίο χωρίς να εντοπίζεται σαφής περιοχή φωτιάς, οδηγώντας σε

πολύ χαμηλή πιθανότητα Fire. Από την τελευταία αυτή περίπτωση καταλαβαίνουμε ότι το μοντέλο χάνει μικρές και αδύναμες εστίες φωτιάς.



Εικόνα 3-57 Χάρτες Grad-CAM του pretrained EdBianchi ViT στο DFire

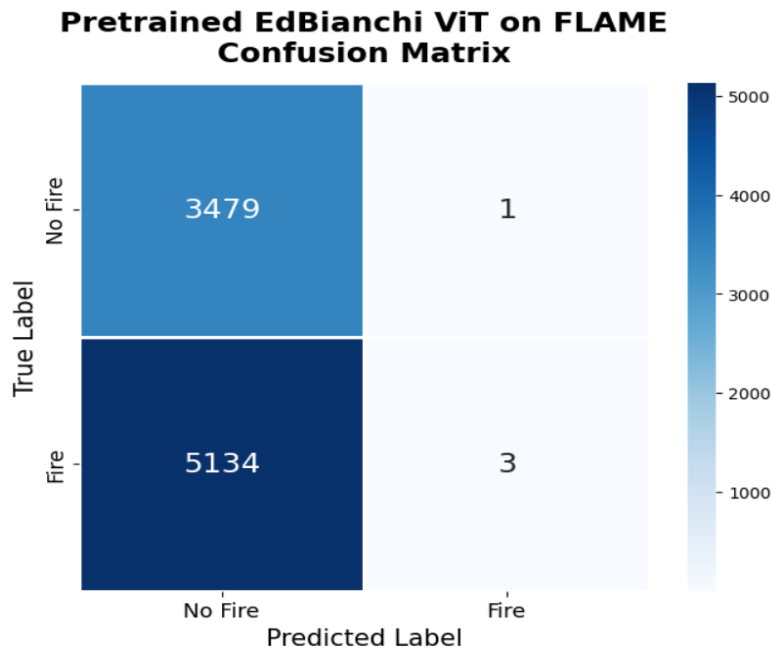
3.7.2 στο FLAME

Από την αξιολόγηση του προεκπαιδευμένου EdBianchiViT στο σύνολο FLAME πήραμε τα ακόλουθα αποτελέσματα

Πίνακας 3-22 Μετρικές του pretrained EdBianchiViT στο FLAME

Pretrained EdBianchiViT στο FLAME	
Accuracy	0.4041
Precision	0.7500
Recall	0.0006

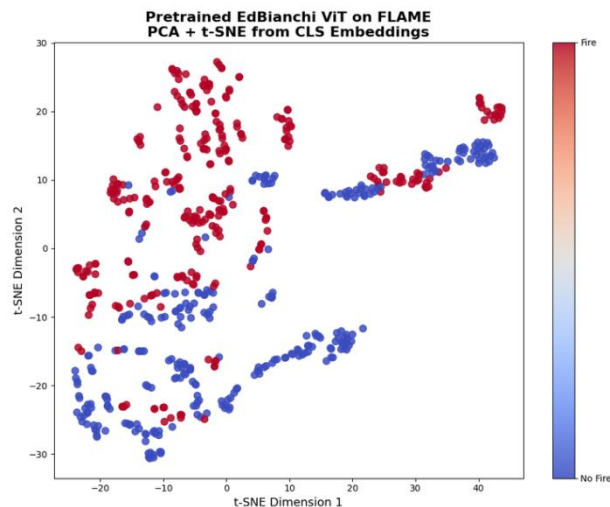
F1-score	0.0012
----------	--------



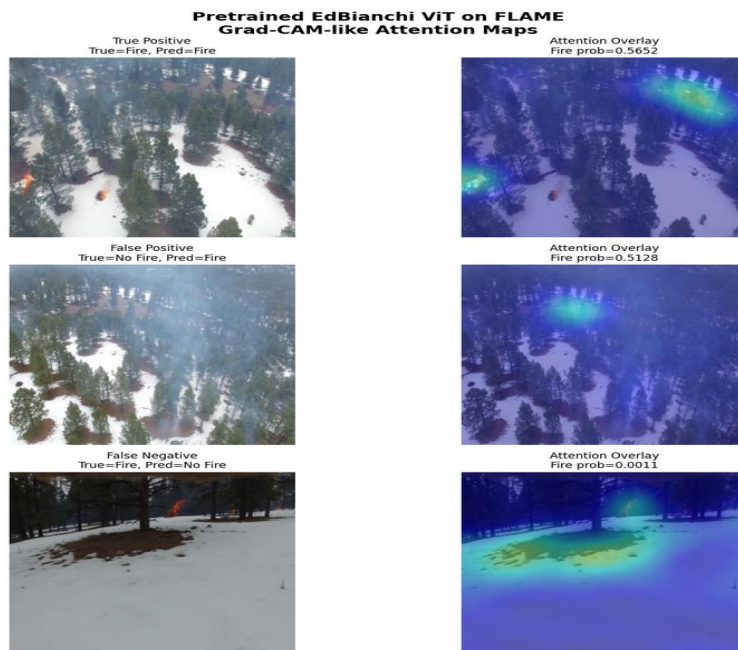
Εικόνα3-58 Confusion Matrix του pretrained EdBianchi Vit στο FLAME

Η αξιολόγηση του προεκπαιδευμένου EdBianchiViT στο σύνολο FLAME ανέδειξε έντονη αδυναμία γενίκευσης σε εικόνες εναέριας λήψης. Παρότι το μοντέλο πέτυχε υψηλή Precision (75,0%), το Recall περιορίστηκε μόλις στο 0,06%, οδηγώντας σε F1-score 0,12%. Το confusion matrix έδειξε ότι από τις 5.137 εικόνες Fire ταξινομήθηκαν σωστά μόνο τρεις, ενώ σχεδόν όλες οι υπόλοιπες χαρακτηρίστηκαν ως No Fire. Αυτό πιθανότατα σχετίζεται με το ίδιο το dataset. Το EdBianchiViT εκπαιδεύτηκε σε ένα dataset που περιέχει κυρίως κοντινές λήψεις φωτιάς, έντονες φλόγες, μεγάλες θερμές περιοχές και εμφανή καπνό. Το FLAME όμως περιέχει εναέριες λήψεις από drone, πολύ μικρές εστίες φωτιάς, μεγάλες δασικές εκτάσεις και πολύ μεγάλο ποσοστό υπόβαθρου. Το μοντέλο φαίνεται έτσι ότι έχει μάθει να αναγνωρίζει "τυπικές" εικόνες φωτιάς και όχι μικρές δασικές εστίες από μεγάλο ύψος.

Το t-SNE που ακολουθεί μας δείχνει ότι το προεκπαιδευμένο μοντέλο διατηρεί έναν ικανοποιητικό βαθμό διαχωρισμού μεταξύ των κλάσεων Fire και No Fire στο σύνολο FLAME, καθώς σχηματίζονται αρκετές διακριτές συστάδες χαρακτηριστικών. Ωστόσο, παρατηρούνται και περιοχές σημαντικής επικάλυψης, γεγονός που αντανακλά την αυξημένη οπτική πολυπλοκότητα των εναέριων εικόνων του FLAME. Παρά τον σχετικά καλό διαχωρισμό στο embedding space, η ταξινόμηση αποτυγχάνει σχεδόν πλήρως, με Recall μόλις 0.06%. Η ασυμφωνία αυτή υποδηλώνει ότι το πρόβλημα δεν εντοπίζεται αποκλειστικά στις αναπαραστάσεις που εξάγει ο Vision Transformer, αλλά κυρίως στο τελικό στάδιο λήψης απόφασης, το οποίο εφαρμόζει ένα ιδιαίτερα αυστηρό όριο για την πρόβλεψη της κλάσης Fire.



Εικόνα 3-59 t-SNE του pretrained EdBianchi ViT στο FLAME



Εικόνα 3-60 Χάρτες Grad-CAM του pretrained EdBianchi Vit στο FLAME

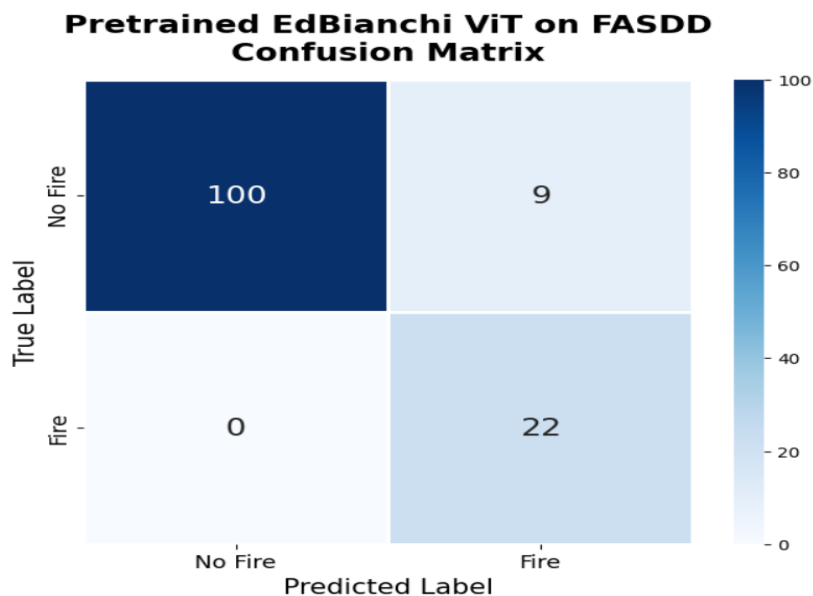
Τα Grad-CAM-like attention maps του προεκπαιδευμένου EdBianchiViT στο FLAME αποκαλύπτουν ότι το μοντέλο δεν βασίζεται αποκλειστικά στις περιοχές όπου εμφανίζεται η φωτιά, αλλά συχνά αξιοποιεί ευρύτερα χαρακτηριστικά της σκηνής. Στο παράδειγμα του True Positive, η προσοχή συγκεντρώνεται τόσο κοντά στην εστία της πυρκαγιάς όσο και σε γειτονικές περιοχές της εικόνας, οδηγώντας σε σωστή πρόβλεψη. Αντίθετα, στο False Positive η ενεργοποίηση εντοπίζεται σε περιοχές του δάσους χωρίς εμφανή φωτιά, γεγονός που υποδηλώνει πιθανή σύγχυση με χαρακτηριστικά της βλάστησης ή του φωτισμού. Στο False Negative, παρότι η πραγματική εστία πυρκαγιάς βρίσκεται εντός της εικόνας, το attention map εμφανίζει περιορισμένη ενεργοποίηση γύρω από αυτή και σημαντικό μέρος της προσοχής κατευθύνεται προς το υπόβαθρο. Η συμπεριφορά αυτή βρίσκεται σε αντιστοιχία και με το ιδιαίτερα χαμηλό Recall του μοντέλου και επιβεβαιώνει ότι το προεκπαιδευμένο μοντέλο δυσκολεύεται να αναγνωρίσει μικρές ή απομακρυσμένες δασικές πυρκαγιές σε εναέριες εικόνες.

3.7.3 στο FASDD

Από την αξιολόγηση του προεκπαιδευμένου EdBianchiVitστο σύνολο FASDD πήραμε τα ακόλουθα αποτελέσματα

Πίνακας 3-23 Μετρικές του pretrainedEdBianchiVitστο FASDD

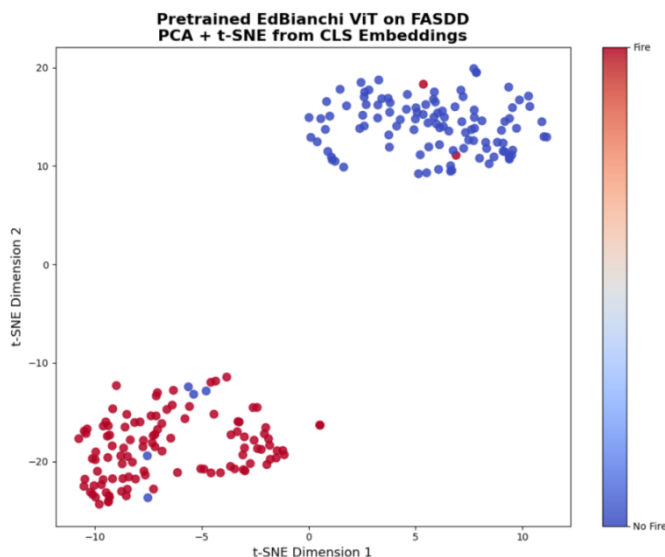
Pretrained EdBianchiVitστο FASDD	
Accuracy	0.9313
Precision	0.7097
Recall	1.000
F1-score	0.8302



Εικόνα 3-61 Confusion Matrix του pretrained EdBianchi Vit στο FASDD

Η αξιολόγηση του προεκπαιδευμένου EdBianchiVitστο σύνολο FASDD κατέγραψε μια ιδιαίτερα υψηλή απόδοση, με Accuracy 93.13%, Recall 100% και F1-score 83.02%. Το μοντέλο αναγνώρισε σωστά όλες τις εικόνες που περιείχαν φωτιά, χωρίς κανένα False Negative, ενώ μόνο εννέα εικόνες χωρίς φωτιά

ταξινομήθηκαν λανθασμένα ως Fire. Η συμπεριφορά αυτή διαφέρει σημαντικά από τα αποτελέσματα που παρατηρήθηκαν στα σύνολα DFire και ιδιαίτερα στο FLAME. Αυτό ίσως υποδηλώνει ότι τα χαρακτηριστικά των εικόνων του FASDD είναι περισσότερο συμβατά με το domain στο οποίο εκπαιδεύτηκε αρχικά το μοντέλο. Συνεπώς, φαίνεται πως το μοντέλο EdBianchiVit εμφανίζει εξαιρετική ικανότητα αναγνώρισης επίγειων εικόνων πυρκαγιάς, αλλά σημαντική μείωση της απόδοσής του όταν εφαρμόζεται σε εναέρια δεδομένα με διαφορετικής κατανομή μεταξύ φωτιάς και υπόβαθρου.



Εικόνα 3-62 t-SNE του pretrainedEdBianchiVit στο FASDD

Η προβολή των classification embeddings μέσω t-SNE για το προεκπαιδευμένο EdBianchiVit στο σύνολο FASDD παρουσιάζει σχεδόν πλήρη διαχωρισμό μεταξύ των δύο κλάσεων. Οι εικόνες Fire και No Fire σχηματίζουν δύο σαφώς διακριτά και ιδιαίτερα συμπαγή clusters, με ελάχιστη μόνο επικάλυψη. Η παρατήρηση αυτή βρίσκεται σε πλήρη συμφωνία με τα αποτελέσματα της ποσοτικής αξιολόγησης (Accuracy 93.13%, Recall 100%), επιβεβαιώνοντας ότι το μοντέλο έχει μάθει μια ιδιαίτερα διακριτική αναπαράσταση των χαρακτηριστικών του συγκεκριμένου συνόλου δεδομένων. Τα ελάχιστα σημεία που εμφανίζονται μέσα στο cluster της αντίθετης κλάσης αντιστοιχούν στις λίγες λανθασμένες ταξινομήσεις που παρατηρήθηκαν στον confusion matrix.



Εικόνα 3-63 Χάρτες Grad-CAM του pretrained EdBianchiViT στο FASDD

Τα Grad-CAM-like attention maps του προεκπαιδευμένου EdBianchi ViT στο σύνολο FASDD επιβεβαιώνουν την ιδιαίτερα υψηλή απόδοση του μοντέλου. Στο παράδειγμα του True Positive, η προσοχή συγκεντρώνεται στις περιοχές όπου εμφανίζονται οι φλόγες και ο καπνός, οδηγώντας σε πολύ υψηλή πιθανότητα ταξινόμησης ως Fire. Στο False Positive, η ενεργοποίηση εντοπίζεται κυρίως σε ξύλινες επιφάνειες και φωτεινές περιοχές του εσωτερικού χώρου, γεγονός που υποδηλώνει ότι το μοντέλο συγχέει ορισμένα θερμά οπτικά χαρακτηριστικά με ενδείξεις πυρκαγιάς. Αντίθετα, στο True Negative η προσοχή κατανέμεται ομαλά στην αστική σκηνή χωρίς έντονες ενεργοποιήσεις σε περιοχές που θα μπορούσαν να υποδηλώνουν φωτιά, οδηγώντας σε εξαιρετικά χαμηλή πιθανότητα Fire. Η οπτική αυτή συμπεριφορά συμφωνεί πλήρως με τις ποσοτικές μετρικές και τον confusionmatrix, επιβεβαιώνοντας ότι το μοντέλο διακρίνει αποτελεσματικά τις εικόνες του FASDD.

3.8 Fine tuning του μοντέλου EdBianchi ViT

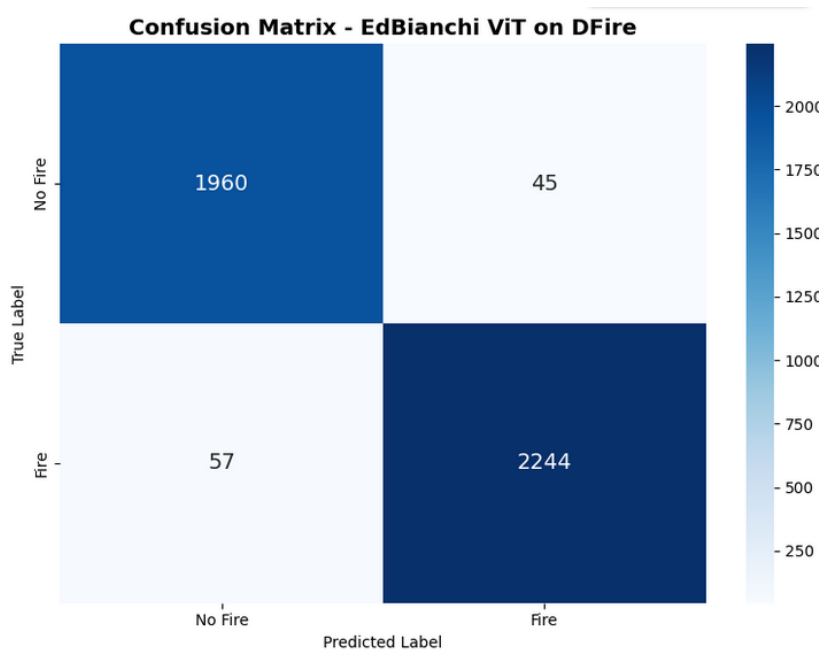
Στην παράγραφο αυτή θα ακολουθήσουμε την διαδικασία του finetuning για το μοντέλο EdBianchiViT.

3.8.1 στο DFire

Μετά από finetuning του EdBianchiViT στο σύνολο DFire και την αξιολόγηση του μοντέλου στο testset τα αποτελέσματα που συγκεντρώσαμε είναι τα εξής

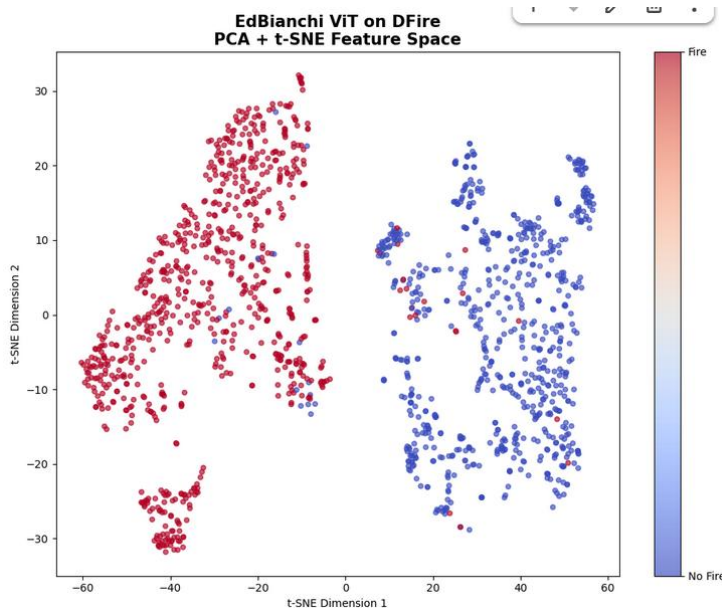
Πίνακας 3-24 Μετρικές του fine-tuned EdBianchiViT στο DFire

Fine tuned EdBianchi ViT στο DFire	
Accuracy	0.9763
Precision	0.9803
Recall	0.9752
F1-score	0.9778



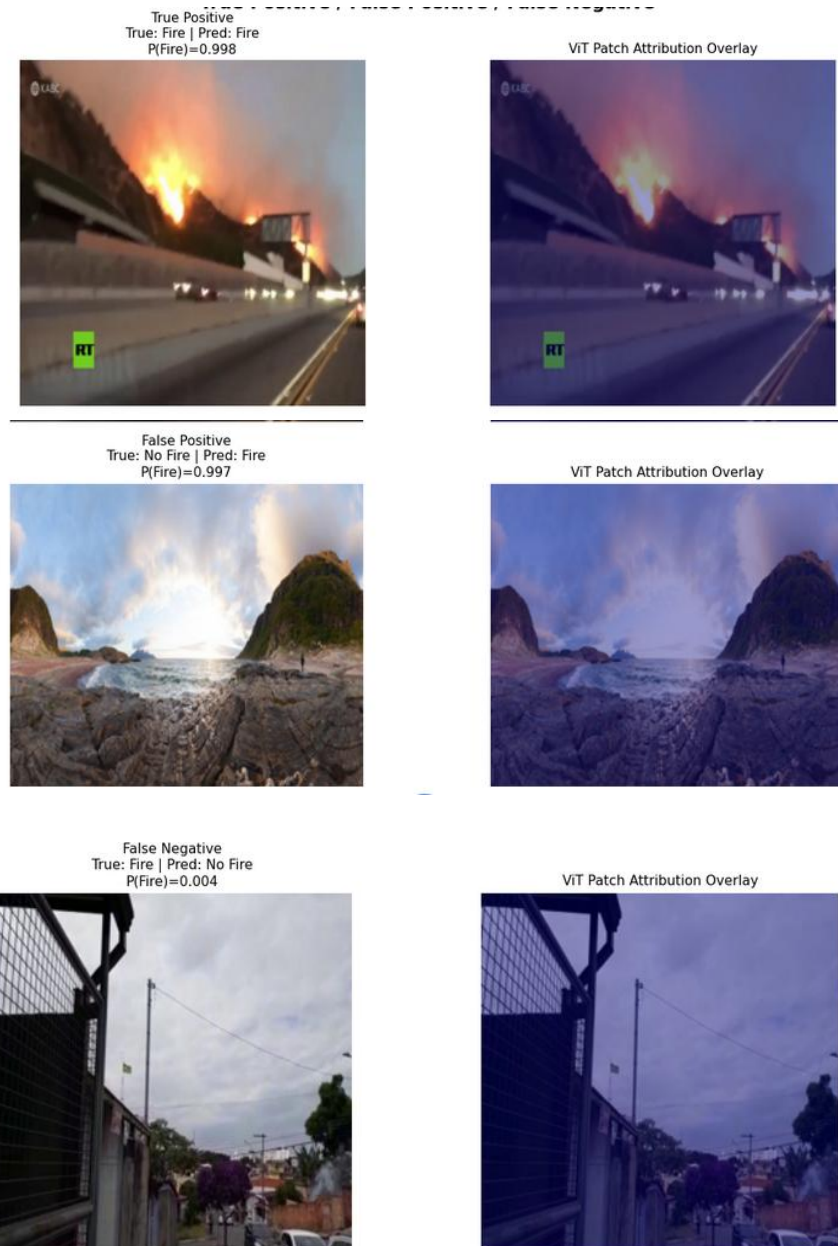
Εικόνα 3-64 Confusion Matrix του fine-tuned EdBianchiVit στο DFire

Η εικόνα που παίρνουμε από τον πίνακα των μετρικών είναι αυτή ενός μοντέλου με εξαιρετική απόδοση. Ιδιαίτερα αξίζει να σημειωθεί η σχεδόν απόλυτη ισορροπία μεταξύ Precision και Recall, όπου Precision = 98,03% και Recall = 97,52% . Το σημείο αυτό μας δείχνει ότι το μοντέλο δεν παρουσιάζει εμφανή «προτίμηση» προς κάποια από τις δύο κλάσεις. Δεν είναι ούτε υπερβολικά «ευαίσθητο», παράγοντας πολλούς ψευδείς συναγερμούς, ούτε υπερβολικά «συντηρητικό», χάνοντας πραγματικές πυρκαγιές. Οπωσδήποτε μια τέτοια συμπεριφορά κρίνεται ως ιδιαίτερα επιθυμητή σε εφαρμογές αυτόματης ανίχνευσης πυρκαγιάς. Από τον confusionmatrix ενημερωνόμαστε ότι από τις 2.005 εικόνες χωρίς πυρκαγιά, οι 1.960 ταξινομήθηκαν σωστά, ενώ μόλις 45 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Αντίστοιχα, από τις 2.301 εικόνες με πυρκαγιά, οι 2.244 αναγνωρίστηκαν σωστά, ενώ μόνο 57 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Ο μικρός αριθμός σφαλμάτων και στις δύο κλάσεις καταδεικνύει την υψηλή ικανότητα γενίκευσης του μοντέλου.



Εικόνα 3-65 t-SNE του fine-tuned EdBianchiVit στο DFire

Από την παρατήρηση του t-SNE βλέπουμε πως η κλάση Fire συγκεντρώνεται σχεδόν αποκλειστικά στο αριστερό τμήμα του διαγράμματος, ενώ η κλάση No Fire καταλαμβάνει το δεξί τμήμα. Ανάμεσα στις δύο κύριες περιοχές υπάρχει ένα σχετικά μικρό μεταβατικό τμήμα, όπου εμφανίζονται λίγα μόνο δείγματα των δύο κατηγοριών. Από την εικόνα αυτή αντιλαμβανόμαστε ότι το μοντέλο έχει μάθει έναν ιδιαίτερα καλά οργανωμένο χώρο χαρακτηριστικών, στον οποίο οι περισσότερες εικόνες ταξινομούνται με μεγάλη βεβαιότητα. Ένα ακόμα ενδιαφέρον χαρακτηριστικό του διαγράμματος είναι η παρουσία υποσυστάδων και στις δύο κλάσεις. Οι υποσυστάδες αυτές πιθανότατα αντιστοιχούν σε διαφορετικούς τύπους πυρκαγιών. Το γεγονός ότι οι υποσυστάδες παραμένουν σαφώς διαχωρισμένες από εκείνες της αντίθετης κλάσης υποδηλώνει ότι το μοντέλο δεν μαθαίνει μόνο έναν γενικό διαχωρισμό μεταξύ Fire και No Fire, αλλά και πιο εξειδικευμένες αναπαραστάσεις που βελτιώνουν την ικανότητα γενίκευσης.



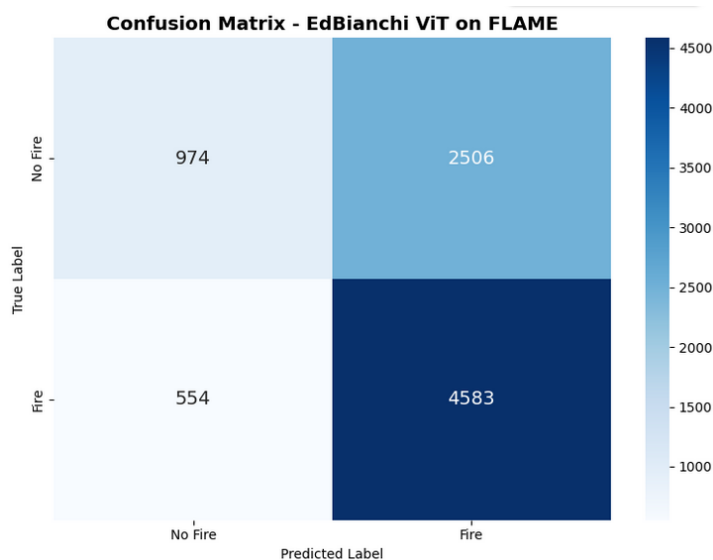
Εικόνα 3-66 Χάρτες Grad-CAM του fine-tuned EdBianchiViT στο DFire

3.8.2 στο FLAME

Από το finetuning του EdBianchiViT στο σύνολο FLAME και την ακόλουθη αξιολόγηση του fine-tuned μοντέλου στο testset του FLAME τα αποτελέσματα φαίνονται στον ακόλουθο πίνακα

Πίνακας 3-25 Μετρικές του fine-tune dEdBianchiVit στο FLAME

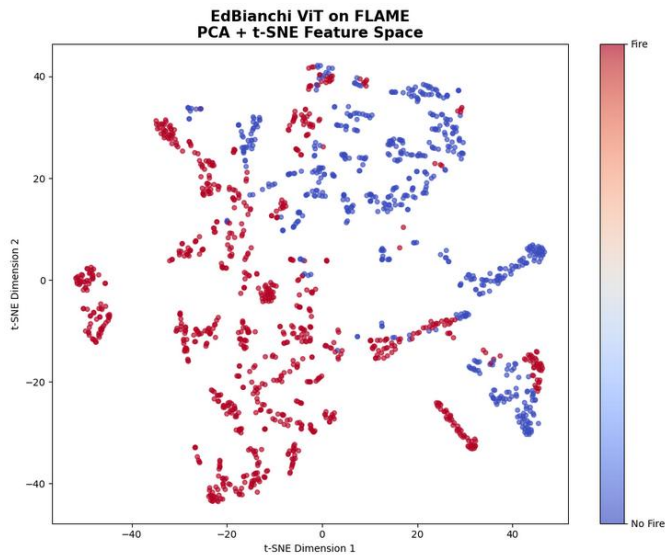
Fine tuned EdBianchi Vit στο FLAME	
Accuracy	0.6449
Precision	0.6465
Recall	0.8922
F1-score	0.7497



Εικόνα 3-67 Confusion Matrix του fine-tuned EdBianchiVit στο FLAME

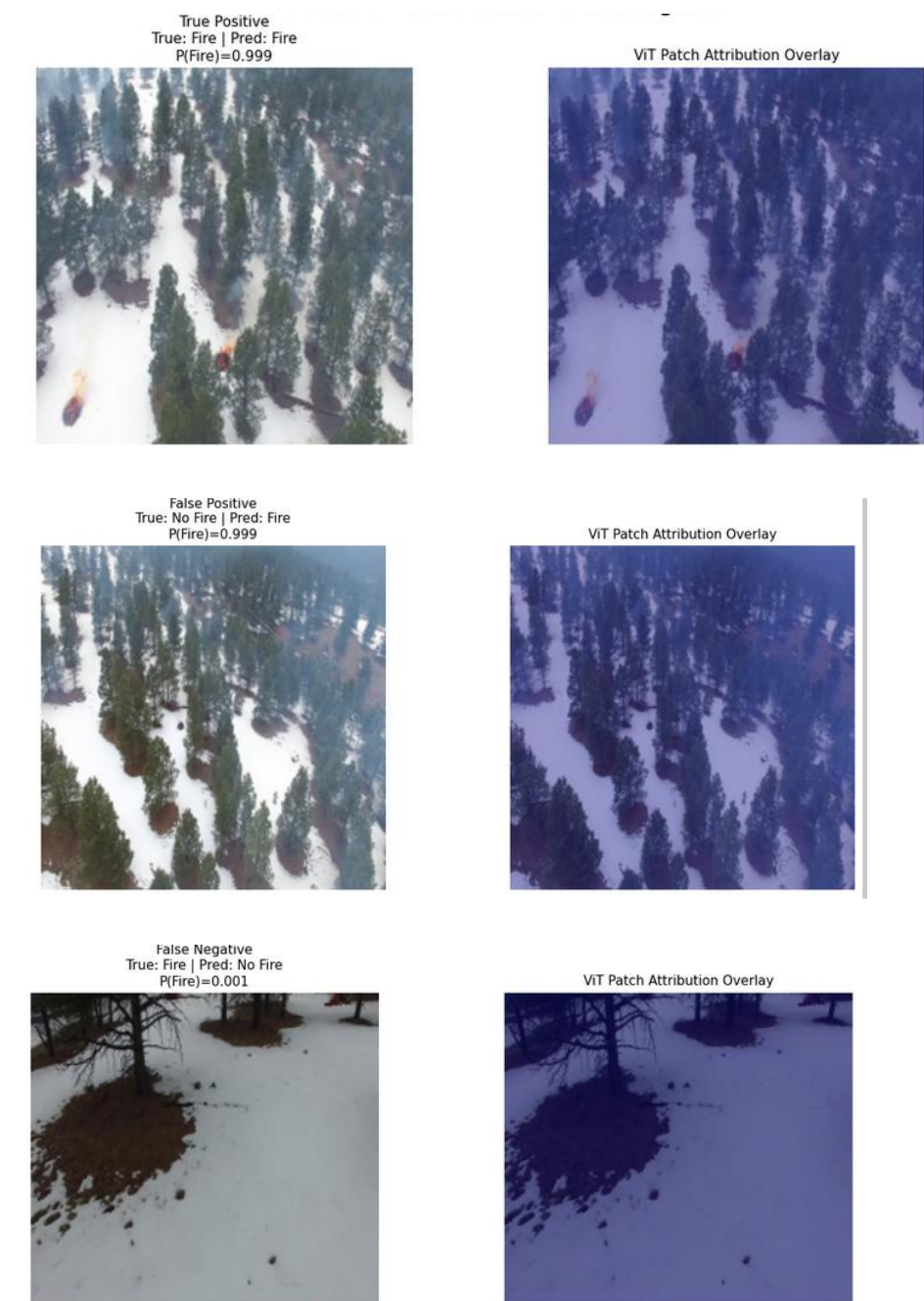
Από τον confusion matrix καθώς και τον πίνακα με τις μετρικές, ένα χαρακτηριστικό γνώρισμα του μοντέλου που μπορούμε να υπογραμμίσουμε είναι η έντονη προδιάθεση του μοντέλου προς την πρόβλεψη της κλάσης Fire. Αυτό φαίνεται ξεκάθαρα από τον confusionmatrixόπου έχουμε 4.583 True Positives αλλά και 2.506 False Positives. Δηλαδή, το μοντέλο προτιμά να χαρακτηρίσει μια εικόνα ως Fire, ακόμη και όταν αυτό δεν ισχύει πραγματικά. Πρόκειται για μια σχετικά "ευαίσθητη στη φωτιά" στρατηγική ταξινόμησης, η οποία αυξάνει την πιθανότητα εντοπισμού πραγματικών πυρκαγιών, αλλά ταυτόχρονα οδηγεί σε αυξημένο αριθμό ψευδών συναγερμών. Θα μπορούσαμε να πούμε πως μετά το

finetuningτο μοντέλο έμαθε να δίνει μεγαλύτερη βαρύτητα σε οπτικά χαρακτηριστικά που σχετίζονται με την παρουσία πυρκαγιάς, ακόμη και όταν αυτά εμφανίζονται σε εικόνες χωρίς πραγματική φωτιά (π.χ. καπνός).



Εικόνα 3-68 t-SNE του fine-tuned EdBianchiVit στο FLAME

Με βάση την οπτικοποίηση του t-SNE οι δύο κλάσεις (Fire και No Fire) δεν εμφανίζουν πλήρη διαχωρισμό, αλλά παρουσιάζουν αρκετές περιοχές επικάλυψης και γειτνίασης. Μπορούμε να διακρίνουμε αρκετές υποσυστάδες που αντιστοιχούν στις δύο κατηγορίες, οι υποσυστάδες όμως αυτές βρίσκονται πολύ κοντά μεταξύ τους ή εμφανίζουν μερική επικάλυψη. Το γεγονός αυτό μας υποδεικνύει ότι ορισμένες εικόνες Fire και No Fire διαθέτουν παρόμοια οπτικά χαρακτηριστικά. Το στοιχείο αυτό μπορεί να εξηγεί εν μέρει τον αυξημένο αριθμό False Positives που παρατηρήθηκε στον confusionmatrix.



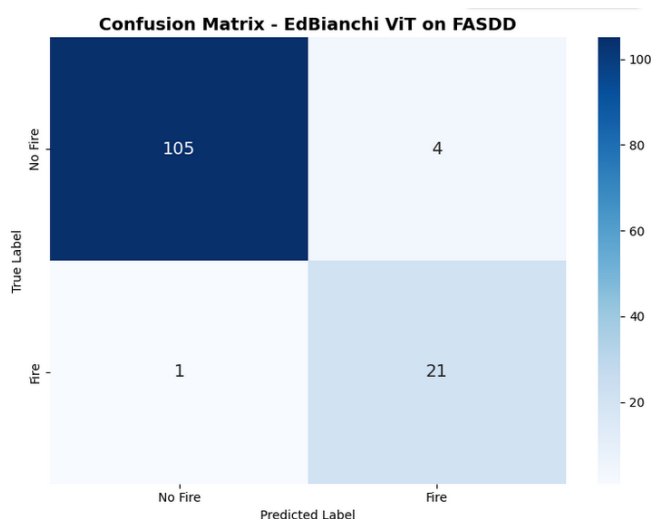
Εικόνα 3-69 ΧάρτεςGrad-CAM του fine-tuned EdBianchiVit στο FLAME

3.8.3 στο FASDD

Από το finetuning τουEdBianchiVit στο σύνολο FASDD αποκομίσαμε τα αποτελέσματα που συγκεντρώνονται στον ακόλουθο πίνακα

Πίνακας 3-26 Μετρικές του fine-tuned EdBianchiVit στο FASDD

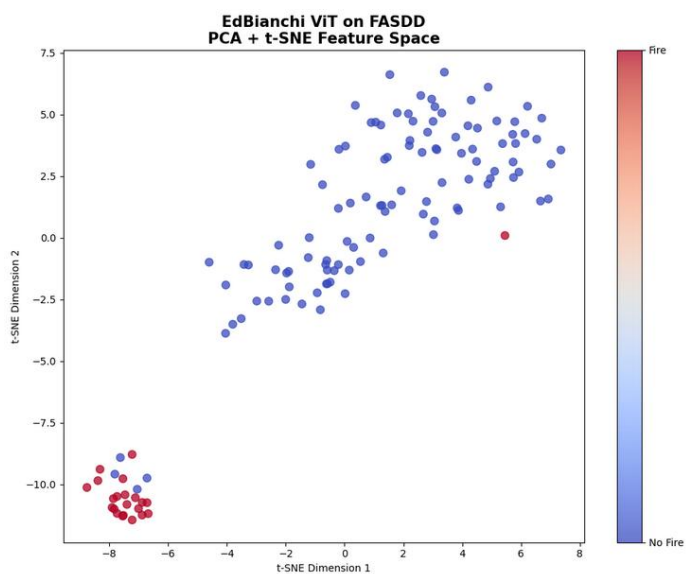
Fine tuned EdBianchi Vit στο FASDD	
Accuracy	0.9618
Precision	0.8400
Recall	0.9545
F1-score	0.8936

**Εικόνα 3-70 Confusion Matrix του fine-tuned EdBianchiVit στο FASDD**

Σημειώνουμε ότι μετά από την διαδικασία finetuning το μοντέλο EdBianchiVit παρουσιάζει μια πολύ θετική εικόνα σε σχέση με την ποσοτική απόδοσή του ως προς τις διάφορες μετρικές. Συγκεκριμένα πέτυχε Accuracy 96,18%, Precision 84,00%, Recall 95,45% και F1-score 89,36% στο σύνολο δεδομένων FASDD. Η υψηλή τιμή του Recall μας πληροφορεί ότι το μοντέλο εντοπίζει σχεδόν όλες τις πραγματικές περιπτώσεις πυρκαγιάς, ενώ από την Precision (84,00%) σημειώνουμε ότι οι περισσότερες προβλέψεις της κλάσης Fire είναι ορθές. Το F1-score (89,36%) αποτυπώνει την καλή ισορροπία μεταξύ των δύο μετρικών (Precision και Recall) και επικυρώνει τη συνολική αξιοπιστία του μοντέλου. Σύμφωνα με τον confusion matrix από τις 109 εικόνες χωρίς πυρκαγιά, οι 105

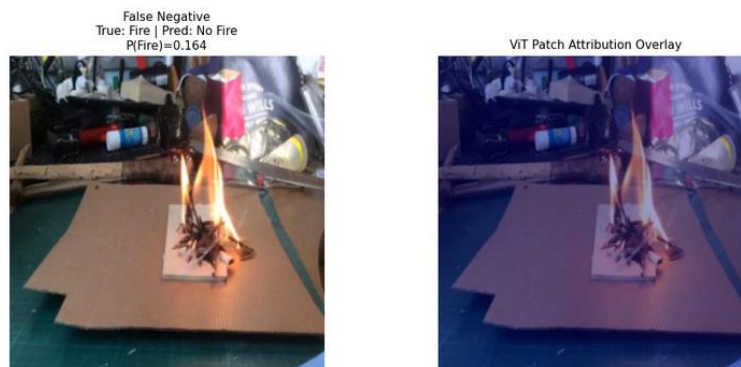
ταξινομήθηκαν σωστά και μόλις 4 από αυτές ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Αντίστοιχα, από τις 22 εικόνες με πυρκαγιά, οι 21 αναγνωρίστηκαν σωστά, απολήγοντας σε μόλις ένα False Negative. Οι 4 FalsePositiveπεριπτώσεις δείχνουν ότι το μοντέλο παραμένει προσεκτικό, χωρίς να παράγει μεγάλο αριθμό ψευδών συναγερμών.

Το διάγραμμα t-SNE μας δίνει μια ξεκάθαρη εικόνα. Οι δύο κλάσεις (Fire και No Fire) διαχωρίζονται σαφώς μεταξύ τους, με τις περισσότερες εικόνες πυρκαγιάς να σχηματίζουν μία συμπαγή συστάδα, ενώ οι εικόνες χωρίς πυρκαγιά οργανώνονται σε μία δεύτερη, διακριτή περιοχή του χώρου χαρακτηριστικών. Η επικάλυψη μεταξύ των δύο κλάσεων είναι εξαιρετικά περιορισμένη και αφορά μόνο μεμονωμένα δείγματα. Με βάση αυτή την εικόνα εξάγουμε το συμπέρασμα ότι το μοντέλο αναπτύσσει αναπαραστάσεις χαρακτηριστικών που διαχωρίζουν αποτελεσματικά τις δύο κλάσεις. Ένα άλλο γεγονός που πρέπει να υπογραμμίσουμε είναι ότι η κλάση Fire εμφανίζει μικρότερη εσωτερική διασπορά σε σχέση με την κλάση No Fire. Συνακόλουθα αυτό μας οδηγεί στο συμπέρασμα ότι οι αναπαραστάσεις των εικόνων πυρκαγιάς είναι περισσότερο ομοιογενείς, γεγονός που πιθανόν εξηγεί και τη μεγάλη τιμή του Recall. Αντίθετα, στην κλάση NoFire βλέπουμε μεγάλη διασπορά μεταξύ των δειγμάτων κάτι που απηχεί ότι οι εικόνες No Fire εμφανίζουν μεγαλύτερη ποικιλία οπτικών χαρακτηριστικών. Σε κάθε περίπτωση παρόλο τη μεγαλύτερη αυτή διασπορά των εικόνων NoFire, το μοντέλο εξακολουθεί να διατηρεί σαφή διαχωρισμό μεταξύ των δύο κλάσεων.



Εικόνα 3-71 t-SNE του fine-tuned EdBianchi ViT στο FASDD





Εικόνα 3-72 Χάρτες Grad-CAM του fine-tuned EdBianchiVit στο FASDD

3.9 Το προεκπαιδευμένο μοντέλο “Superflash 41/ fire-chad-detector-v1.0”

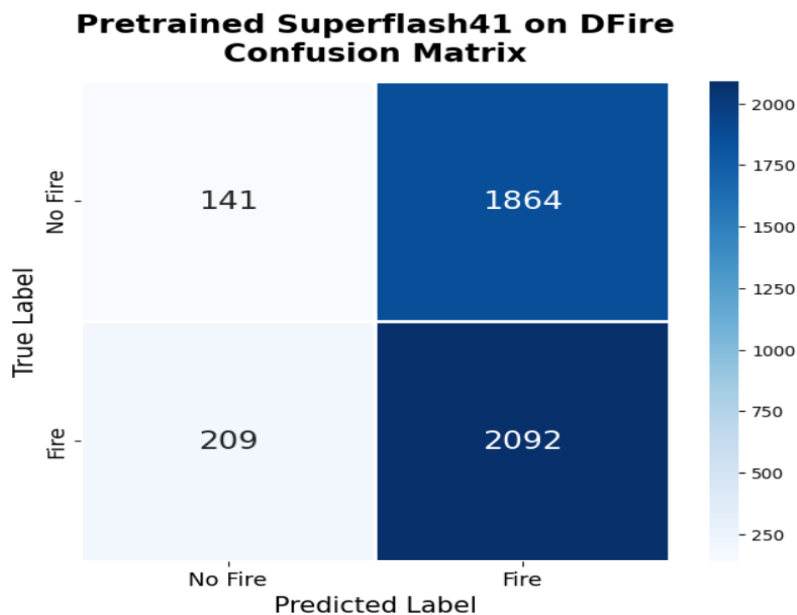
Στην παράγραφο αυτή επιχειρούμε την αποτίμηση του προεκπαιδευμένου μοντέλου “superflash 4.1/ fire-chad-detector-v1.0” (στα ακόλουθα θα αναφερόμαστε σε αυτό ως Superflash41) στα σύνολα δεδομένων DFire, FLAME και FASDD.

3.9.1 στο DFire

Όσον αφορά το προεκπαιδευμένο μοντέλο Superflash41 και τον υπολογισμό των ανάλογων μετρικών στο DFire dataset, παίρνουμε τα ακόλουθα αποτελέσματα.

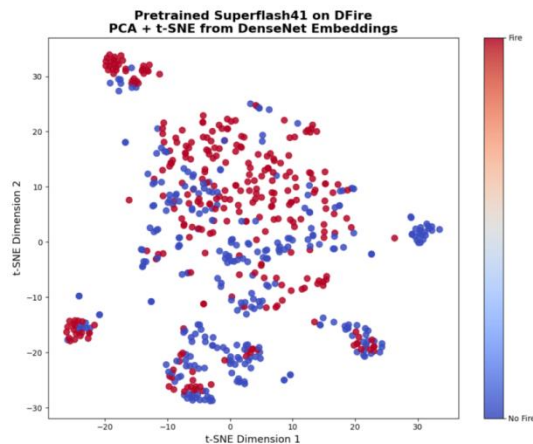
Πίνακας 3-27 Μετρικές του pretrained Superflash41 στο DFire

Pretrained Superflash41 στο DFire	
Accuracy	0.5186
Precision	0.5288
Recall	0.9092
F1-score	0.6687



Εικόνα 3-73 Confusion Matrix του pretrained Superflash41 στοDFire

Η αξιολόγηση του προεκπαιδευμένου Superflash41 στο DFire έδειξε Accuracy 51.86% και F1-score 66.87% για την κλάση Fire. Το μοντέλο πέτυχε υψηλό Recall (90.92%), εντοπίζοντας τις περισσότερες εικόνες που περιείχαν φωτιά. Ωστόσο, η Precision παρέμεινε περιορισμένη (52.88%), καθώς 1.864 εικόνες No Fire ταξινομήθηκαν λανθασμένα ως Fire. Η κατανομή των πιθανοτήτων επιβεβαιώνει αυτή τη συμπεριφορά, με διάμεση πιθανότητα Fire 0.9936, γεγονός που δείχνει έντονη προκατάληψη του μοντέλου προς την πρόβλεψη της κλάσης Fire. Συμπερασματικά, θα μπορούσαμε να πούμε πως το Superflash41 στο DFire λειτουργεί ως ένας ιδιαίτερα ευαίσθητος ως προς τις εικόνες φωτιάς αλλά λιγότερο επιλεκτικός ταξινομητής, προτιμώντας να ελαχιστοποιεί τα False Negatives εις βάρος της αύξησης των False Positives.



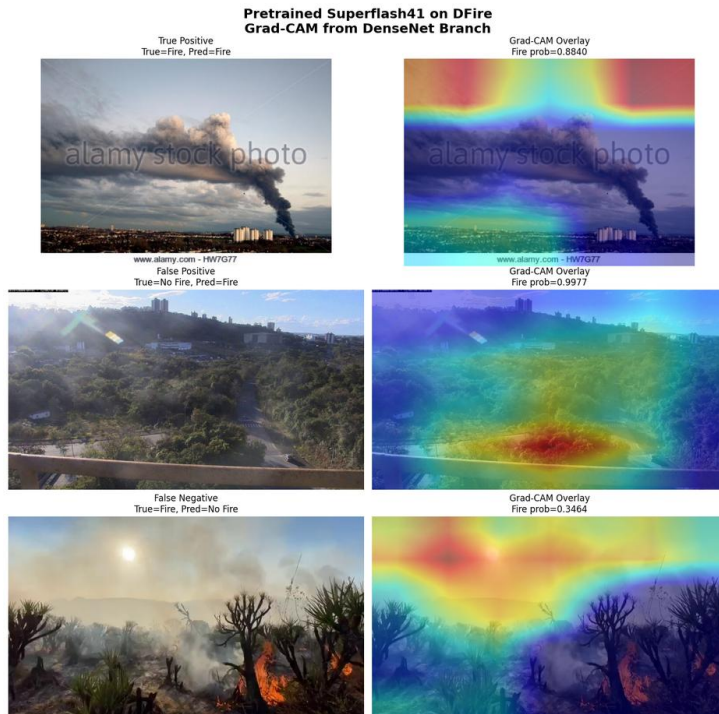
Εικόνα 3-74 t-SNE του pretrained Superflash41 στοDFire

Παρατηρούμε στο t-SNEδιάγραμμα ότι οι δύο κλάσεις δεν διαχωρίζονται καθαρά στον χώρο των embeddings. Στο κεντρικό τμήμα του διαγράμματος υπάρχει έντονη επικάλυψη μεταξύ των δύο κατηγοριών. Πολλά σημεία No Fire βρίσκονται αναμειγμένα με σημεία Fire, γεγονός που δείχνει ότι τα χαρακτηριστικά που εξάγει το μοντέλο δεν είναι επαρκώς διακριτά για το DFire.

Παρ'όλα αυτά, διακρίνονται και ορισμένες τοπικές συμπαγείς συστάδες, ιδιαίτερα στα άκρα του διαγράμματος. Αυτές αντιστοιχούν πιθανότατα σε εικόνες με πολύ χαρακτηριστικά μοτίβα (π.χ. έντονες φλόγες ή πολύ καθαρές εικόνες χωρίς φωτιά), τις οποίες το μοντέλο αναγνωρίζει με μεγαλύτερη βεβαιότητα.

Η εικόνα αυτή συμφωνεί με τον confusionmatrix και τις ποσοτικές μετρικές, όπου το μοντέλο επιτυγχάνει πολύ υψηλό Recall αλλά χαμηλή Precision εξαιτίας του μεγάλου αριθμού False Positives.

Ανασκοπώντας τις εικόνες με τους θερμικούς χάρτες Grad-CAMπου παρατίθενται παρακάτω, βλέπουμε ότι στην περίπτωση του TruePositiveto μοντέλο χρησιμοποιεί κυρίως τα μοτίβα καπνού για να αποφασίσει ότι υπάρχει φωτιά. Στο FalsePositiveto μοντέλο εστιάζει το ενδιαφέρον του σε ορισμένες έντονες φωτεινές περιοχές ή αντιθέσεις που θυμίζουν χαρακτηριστικά πυρκαγιάς. Τέλος στο FalseNegativeto μοντέλο χάνει το μικρό μέτωπο της φωτιάς επειδή η εικόνα κυριαρχείται από τον έντονα φωτισμένο ουρανό στον οποίο και εστιάζει την προσοχή του.



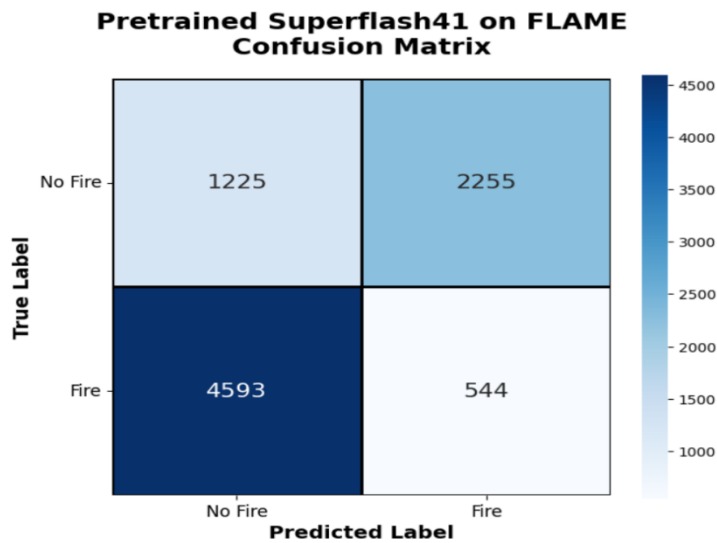
Εικόνα 3-75 Χάρτες Grad-CAM του pretrained Superflash41 στο DFire

3.9.2 στο FLAME

Από την αποτίμηση του προεκπαιδευμένου Superflash41 στο σύνολο FLAME πήραμε τα ακόλουθα αποτελέσματα

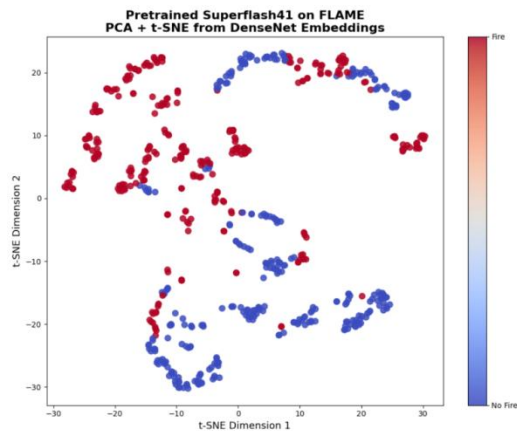
Πίνακας 3-28 Μετρικές του pretrained Superflash41 στο FLAME

Pretrained Superflash41 στο FLAME	
Accuracy	0.2053
Precision	0.1944
Recall	0.1059
F1-score	0.1371



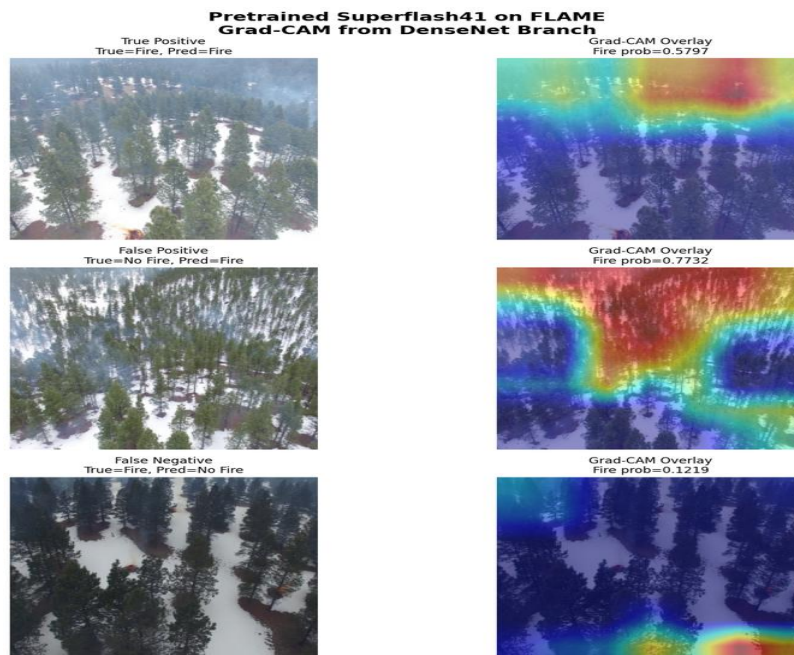
Εικόνα 3-76 Confusion Matrix του pretrained Superflash41 στο FLAME

Η αξιολόγηση του προεκπαιδευμένου Superflash41 στο FLAME δείχνει σημαντική υποβάθμιση της απόδοσης σε σχέση με το DFire. Το μοντέλο πέτυχε Accuracy 20.53% και F1-score 13.71%, ενώ το Recall της κλάσης Fire περιορίστηκε μόλις στο 10.59%. Ο confusionmatrix έχει πολύ μεγάλο αριθμό False Negatives (4.593), γεγονός που σημαίνει ότι το μοντέλο απέτυχε να αναγνωρίσει τη μεγάλη πλειονότητα των πραγματικών περιστατικών πυρκαγιάς. Η κατανομή των πιθανοτήτων εξόδου (διάμεση πιθανότητα Fire 0.225) δείχνει ότι το μοντέλο αποδίδει γενικά χαμηλές πιθανότητες στην κλάση Fire, γεγονός που οδηγεί σε συστηματική υποεκτίμηση των εικόνων πυρκαγιάς.



Εικόνα 3-77 t-SNE του pretrained Superflash41 στο FLAME

Το διάγραμμα t-SNE των embeddings του προεκπαιδευμένου μοντέλου στο FLAME παρουσιάζει έναν σχετικά σαφή διαχωρισμό μεταξύ των κλάσεων Fire και No Fire, με περιορισμένη επικάλυψη στις μεταβατικές περιοχές. Παρά τη χαμηλή ταξινομητική επίδοση του μοντέλου στο FLAME, οι εξαγόμενες αναπαραστάσεις εμφανίζονται οργανωμένες και διακριτές, γεγονός που υποδηλώνει ότι το υποδίκτυο εξαγωγής χαρακτηριστικών έχει μάθει χρήσιμες οπτικές πληροφορίες. Η αντίφαση μεταξύ του σχετικά καλού διαχωρισμού των embeddings και των χαμηλών τιμών Accuracy και Recall ενδεχομένως να σημαίνει ότι η κύρια αιτία της μειωμένης απόδοσης εντοπίζεται στο στάδιο της τελικής ταξινόμησης και όχι στην ποιότητα των εξαγόμενων χαρακτηριστικών.



Εικόνα 3-78 Χάρτες Grad-CAM του pretrained Superflash41 στο FLAME

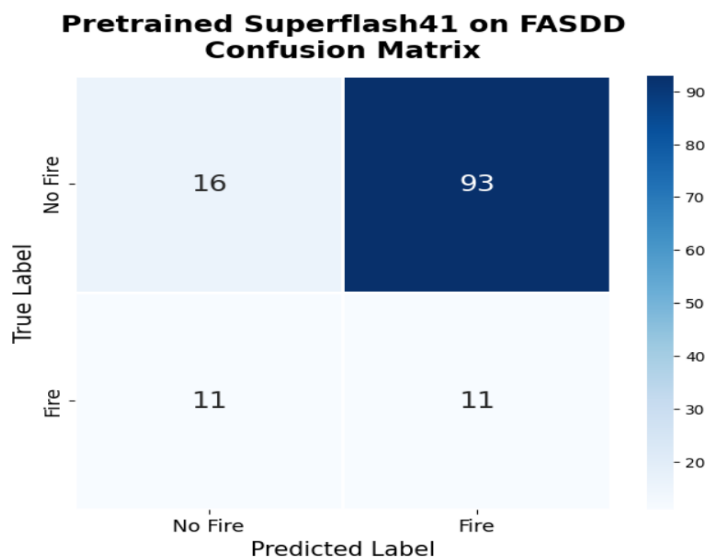
Σε σχέση με τους χάρτες Grad-Cam, βλέπουμε ενδεικτικά ότι στο παράδειγμα του True Positive, η ενεργοποίηση συγκεντρώνεται κυρίως γύρω από την περιοχή της φωτιάς και του καπνού, γεγονός που υποδηλώνει ότι το μοντέλο βασίζεται σε οπτικά χαρακτηριστικά που σχετίζονται άμεσα με την παρουσία πυρκαγιάς. Αντίθετα, στο False Positive παρατηρείται ότι το μοντέλο ενεργοποιεί περιοχές με έντονη φωτεινότητα, ανακλάσεις ή χαρακτηριστικά της βλάστησης, χωρίς να υπάρχει πραγματική εστία φωτιάς, γεγονός που εξηγεί την εσφαλμένη ταξινόμηση. Στην περίπτωση του False Negative, η ενεργοποίηση εντοπίζεται κυρίως σε εκτεταμένες περιοχές του φόντου και του φυσικού τοπίου, ενώ η πραγματική εστία της φωτιάς λαμβάνει περιορισμένη προσοχή.

3.9.3 στο FASDD

Από την αξιολόγηση του προεκπαιδευμένου Superflash41 στο σύνολο FASDD πήραμε τα ακόλουθα αποτελέσματα

Πίνακας 3-29 Μετρικές του pretrained Superflash41 στο FASDD

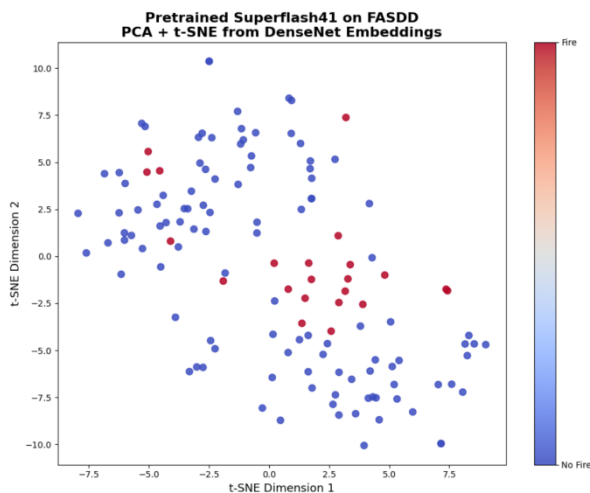
Pretrained Superflash41 στο FASDD	
Accuracy	0.2061
Precision	0.1058
Recall	0.5000
F1-score	0.1746



Εικόνα 3-79 Confusion Matrix του pretrained Superflash41 στο FASDD

Η αξιολόγηση του προεκπαιδευμένου Superflash41 μοντέλου στο σύνολο FASDD έδειξε ιδιαίτερα περιορισμένη ικανότητα γενίκευσης. Το μοντέλο πέτυχε Accuracy 20.61%, Precision 10.58%, Recall 50.00% και F1-score 17.46%. Ο confusion matrix έδειξε ότι από τις 109 εικόνες χωρίς φωτιά, μόλις 16 ταξινομήθηκαν σωστά, ενώ 93 χαρακτηρίστηκαν λανθασμένα ως Fire, γεγονός που οδήγησε σε πολύ μεγάλο αριθμό False Positives. Παράλληλα, από τις 22 πραγματικές εικόνες φωτιάς, οι 11 ανιχνεύθηκαν επιτυχώς και οι υπόλοιπες 11 ταξινομήθηκαν ως No Fire. Συνεπώς, το προεκπαιδευμένο μοντέλο εμφανίζει έντονη προκατάληψη προς την πρόβλεψη της κλάσης Fire στο FASDD, γεγονός που οδηγεί σε σημαντική αύξηση των ψευδώς θετικών ταξινομήσεων και συνολικά χαμηλή ταξινομητική επίδοση.

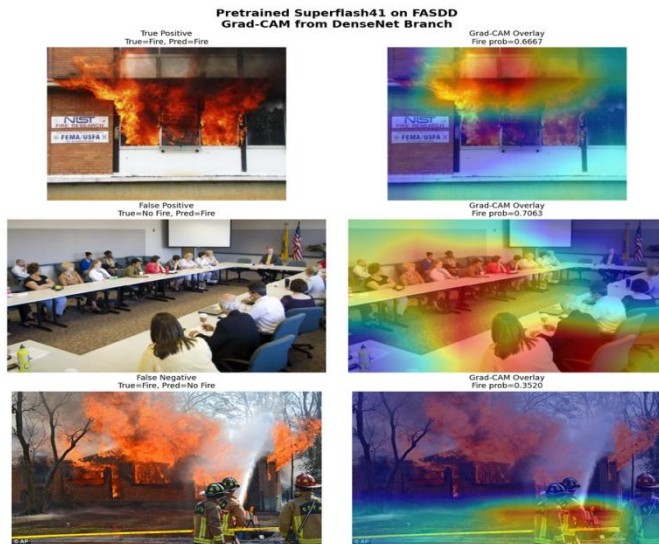
Όπως βλέπουμε από το διάγραμμα t-SNE υπάρχει σημαντική επικάλυψη μεταξύ των κλάσεων Fire και NoFire, χωρίς να σχηματίζονται δύο σαφώς διακριτές συστάδες. Οι εικόνες της κλάσης No Fire εμφανίζονται οργανωμένες σε περισσότερες υποσυστάδες, ενώ τα δείγματα Fire παρουσιάζουν μεγαλύτερη διασπορά και αλληλεπικάλυψη με την αντίθετη κλάση. Η εικόνα αυτή απηχεί τις χαμηλές τιμές Accuracy (20.61%) και F1-score (17.46%) καθώς και τον μεγάλο αριθμό False Positives που παρατηρήθηκε στον confusionmatrix. Τα αποτελέσματα υποδηλώνουν ότι το προεκπαιδευμένο μοντέλο δεν έχει υψηλή διακριτική ικανότητα στη μάθηση χαρακτηριστικών για τις εικόνες του FASDD, γεγονός που επηρεάζει αρνητικά την τελική ταξινόμηση.



Εικόνα 3-80 t-SNE του pretrained Superflash41 στο FASDD

Τα Grad-Cam δείχνουν να επιβεβαιώνουν τα ποσοτικά αποτελέσματα. Στην τρίτη σειρά εικόνων (FalseNegative) βλέπουμε ότι ενώ η φωτιά καταλαμβάνει μεγάλο μέρος της εικόνας, το μοντέλο δίνει περισσότερη προσοχή σε άλλα στοιχεία όπως οι πυροσβέστες, ή το έδαφος και τελικά η εικόνα ταξινομείται λανθασμένα ως No Fire. Στη δεύτερη σειρά εικόνων το μοντέλο σχεδόν δεν εστιάζει σε κάποιο συγκεκριμένο αντικείμενο, η εστίαση του θα μπορούσαμε να πούμε ότι είναι διάχυτη. Η πρώτη σειρά εικόνων μας δείχνει ένα TruePositive.

Φαίνεται πως όταν η φωτιά είναι έντονη και καταλαμβάνει μεγάλο μέρος της εικόνας, το μοντέλο χρησιμοποιεί τα κατάλληλα χαρακτηριστικά.



Εικόνα 3-81 Χάρτες Grad-CAM του pretrained Superflash41 στο FASDD

3.10 Finetuning του μοντέλου Superflash41

Στην παρούσα παράγραφο προχωράμε στη διαδικασία finetuning του μοντέλου Superflash41 στα σύνολα δεδομένων DFire, FLAME και FASDD αντίστοιχα.

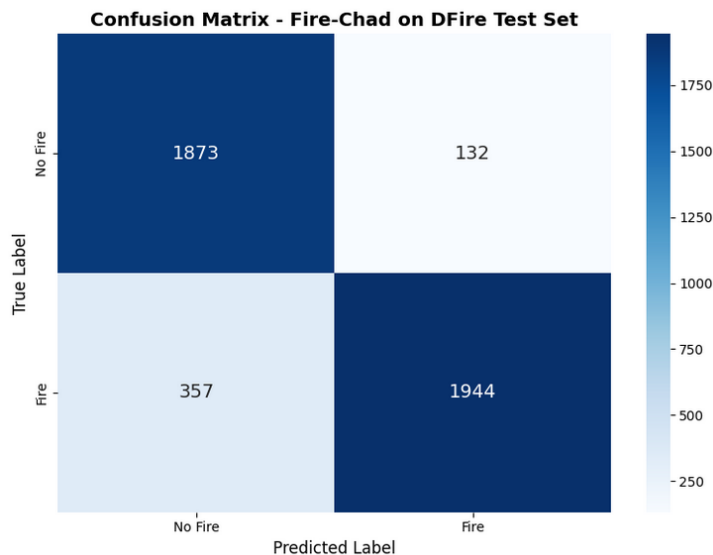
3.10.1 στο DFire

Από το fine tuning του μοντέλου Superflash41 στο DFire και κατόπιν του ελέγχου του finetuned μοντέλου στο test set τα αποτελέσματα που πήραμε στις μετρικές που χρησιμοποιούμε είναι τα ακόλουθα

Πίνακας 3-30 Μετρικές του fine-tuned Superflash41 στο DFire

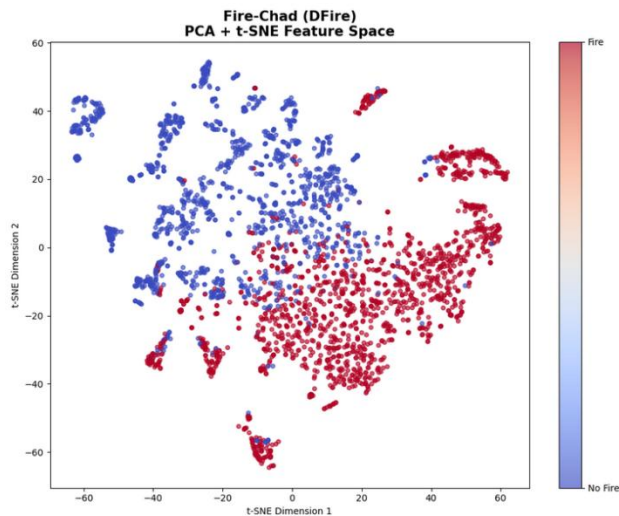
Fine tuned Superflash41 στο DFire	
Accuracy	0.8864
Precision	0.9364

Recall	0.8449
F1-score	0.8883



Εικόνα 3-82 Confusion Matrix του fine-tuned Superflash41 στοDFire

Το πιο χαρακτηριστικό στοιχείο των αποτελεσμάτων είναι ότι το μοντέλο παρουσιάζει υψηλότερη Precision από Recall. Αυτό πρακτικά σημαίνει ότι το μοντέλο είναι σχετικά συντηρητικό στις προβλέψεις του υπό την έννοια ότι προτιμά να χαρακτηρίσει μια εικόνα ως Fire μόνο όταν έχει υψηλή βεβαιότητα. Με αυτό τον τρόπο μειώνονται οι ψευδείς συναγερμοί. Συγχρόνως όμως αυτή η στρατηγική οδηγεί και σε μεγαλύτερο αριθμό False Negatives, δηλαδή σε πραγματικές πυρκαγιές που δεν ανιχνεύονται. Εν προκειμένω αντλώντας τα στοιχεία από τον confusionmatrix βλέπουμε ότι από τις 2.005 εικόνες χωρίς πυρκαγιά, οι 1.873 ταξινομήθηκαν σωστά, ενώ 132 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Αντίστοιχα, από τις 2.301 εικόνες με πυρκαγιά, οι 1.944 αναγνωρίστηκαν σωστά, ενώ 357 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Τα αποτελέσματα αυτά δείχνουν ότι το μοντέλο επιτυγχάνει υψηλή ακρίβεια στις θετικές προβλέψεις, αλλά εξακολουθεί να χάνει ένα μέρος των πραγματικών περιστατικών πυρκαγιάς.

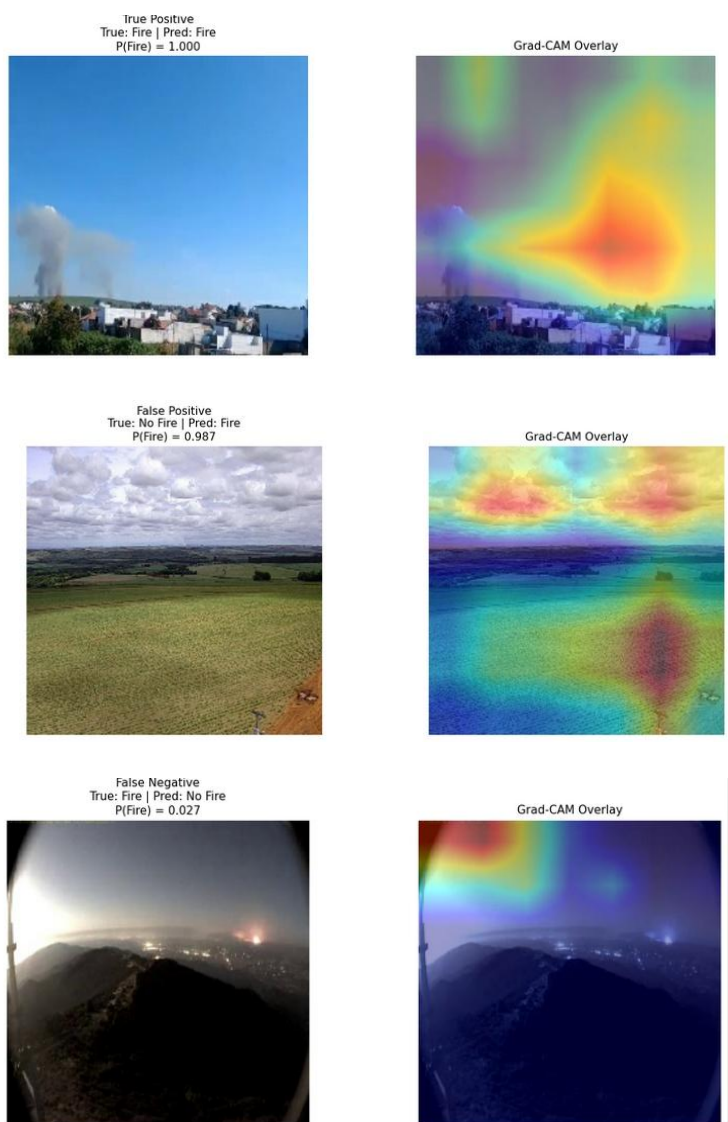


Εικόνα 3-83 t-SNE του fine-tunedSuperflash41 στο DFire

Στο διάγραμμα t-SNE παρουσιάζεται ο χώρος χαρακτηριστικών που εξάγει το Superflash41 μετά το fine-tuning στο σύνολο δεδομένων DFire. Διαπιστώνουμε ότι οι δύο κλάσεις (Fire και No Fire) σχηματίζουν δύο κύριες περιοχές στον χώρο χαρακτηριστικών. Αυτό σημαίνει ότι το μοντέλο έχει μάθει να διαχωρίζει σε ένα ικανοποιητικό βαθμό τις δύο κατηγορίες. Ωστόσο εμφανίζεται μια όχι αμελητέα επικάλυψη μεταξύ των δύο κλάσεων, ιδιαίτερα στην κεντρική περιοχή του διαγράμματος. Η επικάλυψη αυτή αποτελεί ένδειξη ότι ορισμένες εικόνες των δύο κλάσεων διαθέτουν παρόμοια χαρακτηριστικά, στοιχείο που δυσκολεύει τον τελικό διαχωρισμό. Αξίζει να παρατηρηθεί το γεγονός ότι στο διάγραμμα t-SNE δημιουργούνται αρκετές υποσυστάδες και από τις δύο κλάσεις. Αυτό σημαίνει ότι το μοντέλο αναγνωρίζει διαφορετικούς τύπους εικόνων μέσα σε κάθε κατηγορία, όπως θα μπορούσαν να είναι μεταβολές στην ένταση της φωτιάς, στην ποσότητα του καπνού ή στις συνθήκες φωτισμού. Ωστόσο, αρκετές από αυτές τις υποσυστάδες των δύο κλάσεων βρίσκονται ανανεμιγμένες μεταξύ τους, στοιχείο δηλωτικό της αβεβαιότητας του μοντέλου κατά την ταξινόμηση ορισμένων δειγμάτων.

Από τους χάρτες Grad-CAM φαίνεται ότι το Superflash41 εστιάζει κυρίως στις περιοχές της εικόνας που σχετίζονται με την παρουσία καπνού και φωτιάς, γεγονός που εξηγεί τη σχετικά υψηλή Precision (93,64%) και τη συνολικά καλή απόδοσή

του στο σύνολο δεδομένων DFire. Οι περιπτώσεις FalsePositive και FalseNegative μας υποδεικνύει ότι το μοντέλο επηρεάζεται από μεγάλης κλίμακας φωτεινά ή ατμοσφαιρικά χαρακτηριστικά της σκηνής (νέφη, λάμπες), τα οποία σε ορισμένες περιπτώσεις μπορεί να μοιάζουν οπτικά με καπνό ή φωτιά.



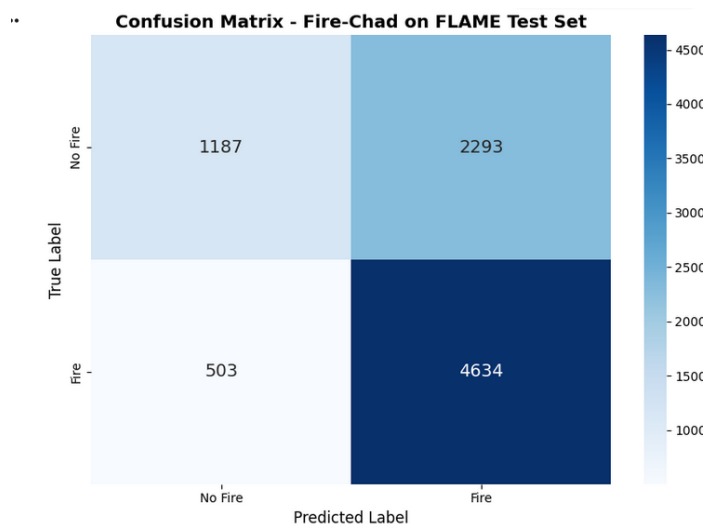
Εικόνα 3-84ΧάρτεςGrad-CAM του fine-tuned Superflash41 στοDFire

3.10.2 στο FLAME

Από το finetuning του μοντέλου Superflash41 στο FLAME έχουμε τον ακόλουθο συγκεντρωτικό πίνακα απόδοσης ως προς τις διάφορες μετρικές.

Πίνακας 3-31 Μετρικές του fine-tuned Superflash41 στο FLAME

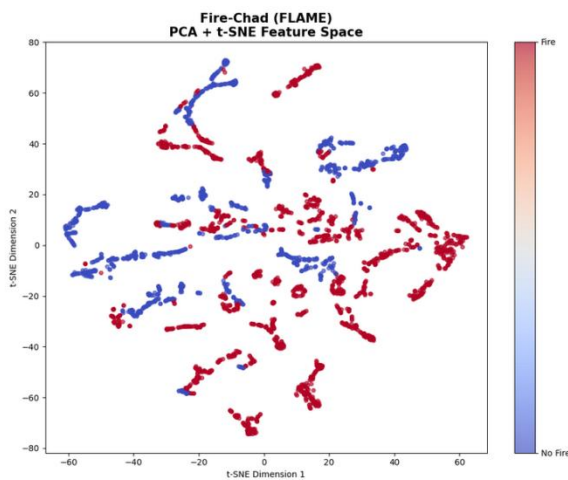
Fine tuned Superflash41στο FLAME	
Accuracy	0.6755
Precision	0.6690
Recall	0.9021
F1-score	0.7682



Εικόνα 3-85 Confusion Matrix του fine-tuned Superflash41 στοFLAME

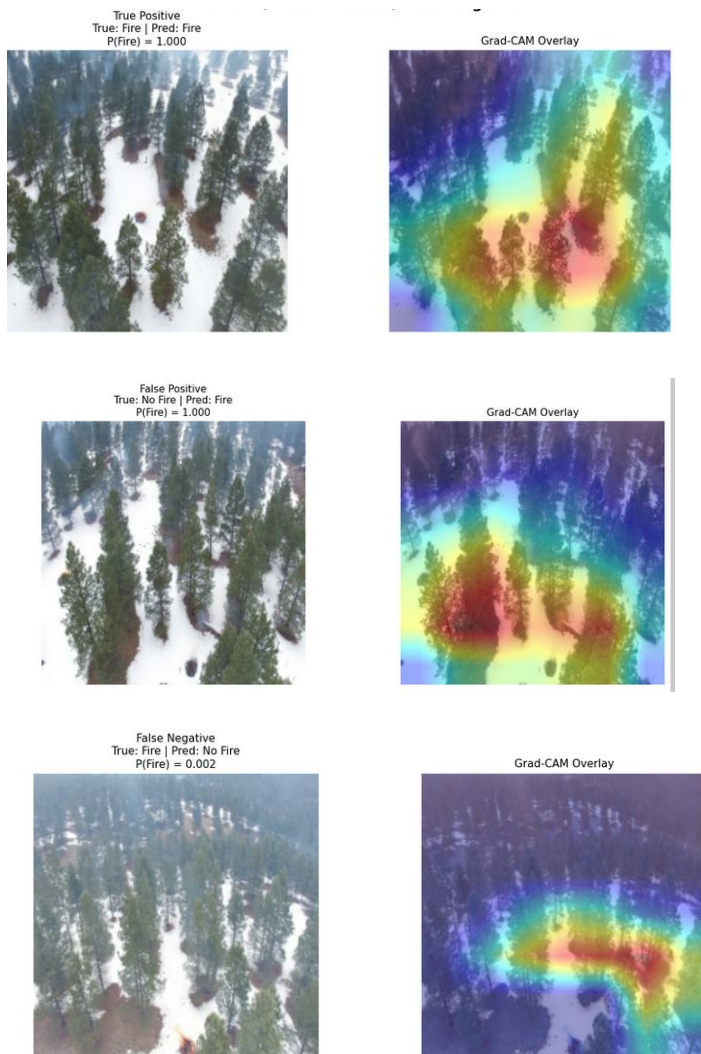
Με βάση τις τιμές των μετρικών παρατηρούμε ότι το μοντέλο παρουσιάζει ένα υψηλό Recall 90,2% , που σημαίνει ότι το μοντέλο εντοπίζει σωστά το 90.2% των πυρκαγιών. Παρ'όλα αυτά η μέτρια τιμή του Precision 0.6690 δείχνει πως το μοντέλο χαρακτηρίζει ως «φωτιά» πολλές εικόνες που δεν έχουν φωτιά. Το F1-score (76,82%) αντανakλά τον συμβιβασμό μεταξύ της υψηλής ευαισθησίας του μοντέλου ως προς την αναγνώριση της φωτιάς και της μειωμένης ικανότητάς του να αποφεύγει ψευδείς συναγερμούς. Ο confusionmatrix μας πληροφορεί ότι από

τις 5.137 εικόνες με πυρκαγιά, οι 4.634 ταξινομήθηκαν σωστά, ενώ 503 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Αντίστοιχα, από τις 3.480 εικόνες χωρίς πυρκαγιά, μόνο 1.187 αναγνωρίστηκαν σωστά, ενώ 2.293 ταξινομήθηκαν λανθασμένα ως Fire (False Positives).



Εικόνα 3-86 t-SNE του fine-tuned Superflash41 στο FLAME

Η οπτικοποίηση του t-SNE αποκαλύπτει τον σχηματισμό πολλών τοπικών συστάδων που αντιστοιχούν σε πρότυπα εικόνων με φωτιά και χωρίς φωτιά. Παρόλο που δεν παρατηρείται πλήρης διαχωρισμός των κλάσεων, πολλές περιοχές κυριαρχούνται από μία μόνο κλάση, γεγονός που υποδηλώνει ότι το μοντέλο έχει μάθει ορισμένες αντιπροσωπευτικές οπτικές αναπαραστάσεις. Η παρουσία επικαλυπτόμενων περιοχών μεταξύ δειγμάτων φωτιάς και μη φωτιάς βρίσκεται σε συμφωνία με τα αποτελέσματα του confusion matrix, ιδιαίτερα με τον σχετικά υψηλό αριθμό ψευδώς θετικών προβλέψεων (false positives).



Εικόνα 3-87 Χάρτες Grad-CAM του fine-tuned Superflash41 στο FLAME

Οι παραπάνω Grad-CAM οπτικοποιήσεις μας δείχνουν ότι το Superflash41 αξιοποιεί τόσο τα χαρακτηριστικά της φωτιάς όσο και το ευρύτερο περιβάλλον της σκηνής κατά τη διαδικασία λήψης απόφασης. Θα λέγαμε ότι οι χάρτες του Superflash41 εμφανίζονται πιο διάχυτοι και καλύπτουν μεγαλύτερες περιοχές της εικόνας. Αυτό υποδηλώνει ότι το μοντέλο βασίζεται περισσότερο στο γενικότερο πλαίσιο της εικόνας και λιγότερο σε αυστηρά τοπικά χαρακτηριστικά της φωτιάς. Παρακολουθώντας τη σειρά των παραπάνω εικόνων βλέπουμε ότι στην περίπτωση του False Positive το μοντέλο παρερμηνεύει ως φωτιά στοιχεία του τοπίου ή της βλάστησης. Επιπλέον, στην περίπτωση του False Negative όταν η εστία της πυρκαγιάς είναι μικρή ή λιγότερο έντονη, το μοντέλο ενδέχεται να εστιάσει σε

δευτερεύοντα χαρακτηριστικά της εικόνας και να χάσει έτσι τη σωστή πρόβλεψη της φωτιάς.

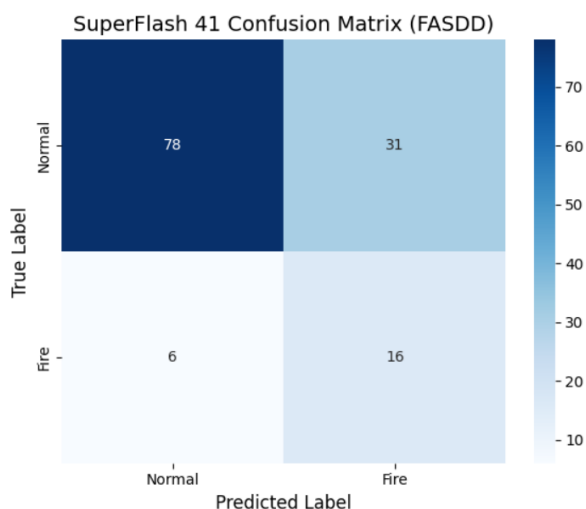
3.10.3 στο FASDD

Από το finetuning του μοντέλου Superflash41 στο FASDD και μετά από την δοκιμή του μοντέλου στο testset συλλέξαμε τα ακόλουθα αποτελέσματα

Πίνακας 3-32 Μετρικές του fine-tuned Superflash41 στο FASDD

Fine tuned Superflash41 στο FASDD	
Accuracy	0.7176
Precision	0.3404
Recall	0.7273
F1-score	0.4638

Στο σημείο αυτό αξίζει να σημειώσουμε ότι κατά τη διάρκεια του finetuning σε 3 epochs παρατηρήθηκε μια πτωτική τάση στο validation loss (από 0.848 σε 0.681 και μετά στην 3^η epoch σε 0.597) στοιχείο που δείχνει προς την κατεύθυνση της αύξησης της μάθησης του μοντέλου. Με βάση το στοιχείο αυτό θα μπορούσαμε να επιχειρηματολογήσουμε πως μια συνέχιση της διαδικασίας του fine-tuning για περισσότερες epochs θα ήταν δυνατό να βελτιώσει την απόδοση του μοντέλου, κάτι που μένει να επιβεβαιωθεί βεβαίως πειραματικά.

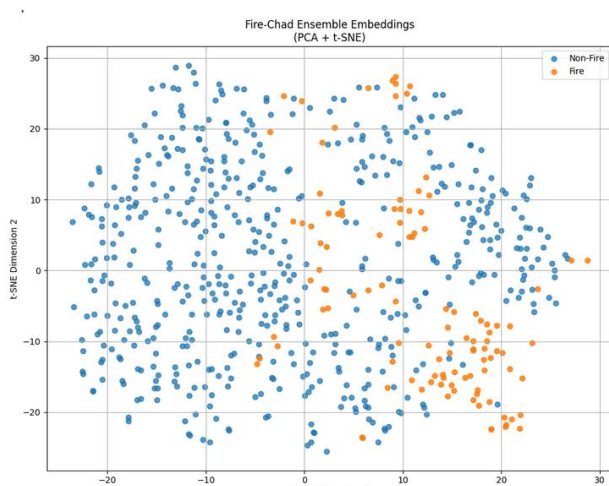


Εικόνα 3-88 Confusion Matrix του fine-tuned Superflash41 στοFASDD

Σε σχέση με την απόδοση του μοντέλου σημειώνουμε ότι η ιδιαίτερα χαμηλή Precision (34,04%) δείχνει ότι μεγάλο ποσοστό των εικόνων που προβλέπονται ως Fire αντιστοιχεί στην πραγματικότητα σε εικόνες χωρίς πυρκαγιά. Συγχρόνως, το Recall (72,73%) μας πληροφορεί ότι το μοντέλο εντοπίζει σχεδόν τρεις στις τέσσερις πραγματικές περιπτώσεις πυρκαγιάς. Ως συνδυαστικός δείκτης των Precision και Recall το F1-score (46,38%) παραμένει σχετικά χαμηλό, αντανακλώντας τη μειωμένη συνολική αξιοπιστία του μοντέλου στο συγκεκριμένο dataset. Ο confusion matrix συμπληρώνει την εικόνα της ελλιπούς απόδοσης του μοντέλου δείχνοντάς μας πως από τις 109 εικόνες χωρίς πυρκαγιά, οι 78 ταξινομήθηκαν σωστά, ενώ 31 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Αντίστοιχα, από τις 22 εικόνες με πυρκαγιά, μόνο 16 αναγνωρίστηκαν σωστά, ενώ 6 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Παρόλο που το μοντέλο αναγνωρίζει την πλειοψηφία των εικόνων χωρίς πυρκαγιά, εμφανίζει σημαντικό αριθμό σφαλμάτων και στις δύο κατηγορίες. Μια πιθανή αιτία της ελλιπούς απόδοσης του fine-tuned μοντέλου θα μπορούσε να είναι πως το Superflash 41 έχει εκπαιδευτεί στο FLAME ενώ το fine-tuning γίνεται στο υποσύνολο του FASDD στο οποίο οι εικόνες με τις πυρκαγιές μπορεί να διαφέρουν αρκετά από αυτές στις οποίες έχει γίνει η αρχική εκπαίδευση του μοντέλου. Επίσης το μέγεθος του υποσυνόλου του FASDD που χρησιμοποιούμε

για το fine-tuning είναι μικρό και αυτό μπορεί να δυσκολεύει το μοντέλο να ανταπεξέλθει στην αναγνώριση μοτίβων

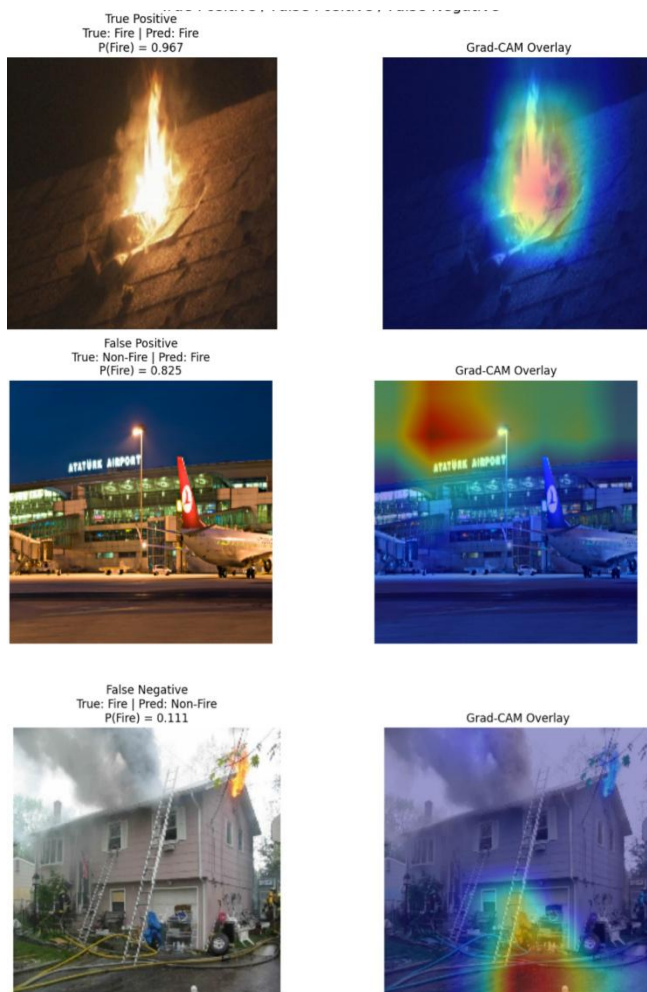
Το διάγραμμα t-SNE δείχνει ένα μερικό διαχωρισμό μεταξύ των κατηγοριών Fire και Non-Fire. Ειδικότερα ένα σημείο που αξίζει να σημειώσουμε είναι η δημιουργία τοπικών συστάδων που αντιστοιχούν κυρίως σε εικόνες φωτιάς στην δεξιά κάτω γωνία του διαγράμματος, γεγονός που υποδηλώνει ότι το μοντέλο έχει μάθει κάποια διακριτά χαρακτηριστικά για ορισμένες περιπτώσεις φωτιάς. Ωστόσο στη συνολική εικόνα η σημαντική επικάλυψη μεταξύ των δύο κατηγοριών μας οδηγεί στο συμπέρασμα ότι ο διαχωρισμός δεν είναι πλήρης, κάτι που συμφωνεί με τις μέτριες τιμές των μετρικών ταξινόμησης.



Εικόνα 3-89 t-SNE του fine-tuned Superflash41 στο FASDD

Εξετάζοντας τις Grad-CAM οπτικοποιήσεις που φαίνονται παρακάτω εξάγουμε το συμπέρασμα ότι το Superflash41 μπορεί να εντοπίσει σωστά την πυρκαγιά όταν αυτή είναι έντονη και κυριαρχεί σε επίπεδο εικόνας (1^η σειρά εικόνων, περίπτωση TruePositive). Ωστόσο, σε πιο σύνθετες σκηνές η προσοχή του μοντέλου ενδέχεται να διαχέεται σε φωτεινά αντικείμενα, τεχνητές πηγές φωτισμού ή άλλα στοιχεία του περιβάλλοντος. Αυτό συμβαίνει στη 2^η σειρά εικόνων, περίπτωση FalsePositive, όπου η προσοχή του μοντέλου συγκεντρώνεται κυρίως γύρω από τον έντονα φωτισμένο στύλο και τη φωτεινή περιοχή του ουρανού, με αποτέλεσμα το μοντέλο λανθασμένα να ερμηνεύει τα στοιχεία αυτά ως φωτιά. Τέλος η

περίπτωση του FalseNegative (3^η σειρά εικόνων) είναι ενδεικτική ως προς το ότι ενώ στη σκηνή υπάρχει μια σχετικά ευδιάκριτη εστία φωτιάς αλλά και πυκνός καπνός που βγαίνει από τη στέγη του σπιτιού το μοντέλο στρέφει την προσοχή του στο κάτω μέρος της εικόνας, κοντά στο έδαφος και στον εξοπλισμό των πυροσβεστών. Η αποτυχία αυτή είναι εμφαντική ως προς την ελλιπή μάθηση από το μοντέλο των βασικών χαρακτηριστικών φωτιάς.



Εικόνα 3-90 Χάρτες Grad-CAM του fine-tuned Superflash41 στο FASDD

3.11 Το προεκπαιδευμένο μοντέλο Convnext

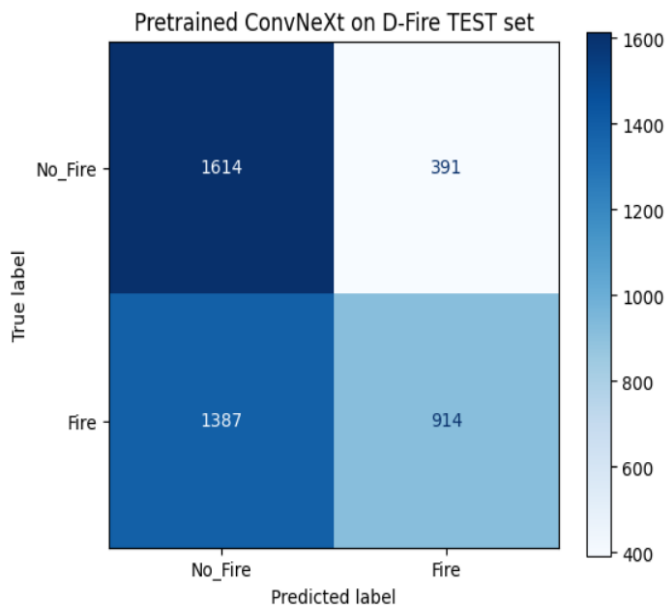
Στην παράγραφο αυτή θα ασχοληθούμε με το προεκπαιδευμένο μοντέλο “convnext.tiny.fbinc22kftin 1k”, το οποίο σε ότι ακολουθεί θα αναφέρεται ως Convnext. Η καταγραφή της απόδοσης του μοντέλου γίνεται στα σύνολα DFire, FLAME και FASDD στις υποπαραγράφους που ακολουθούν.

3.11.1 στο DFire

Θα πρέπει εδώ να σημειώσουμε πως δεδομένου ότι το μοντέλο Convnext είναι ένα μοντέλο που έχει κατασκευαστεί για να κάνει ταξινόμηση εικόνας και το σύνολο δεδομένων D-Fire είναι προσανατολισμένο στην εργασία της αναγνώρισης αντικειμένου, χρειάστηκε η προσαρμογή των εικόνων του D-fire προκειμένου το μοντέλο να μπορεί να ανταποκριθεί στο έργο της αναγνώρισης εικόνας. Έτσι προκειμένου για το D-Fire, οι ετικέτες YOLO μετατράπηκαν σε δυαδική ταξινόμηση εικόνας. Με αυτό τον τρόπο κάθε εικόνα με μη κενό αρχείο ετικετών θεωρήθηκε θετικό δείγμα, καθώς περιείχε ένδειξη πυρκαγιάς ή καπνού. Αντίθετα, εικόνες με κενό αρχείο ετικετών θεωρήθηκαν αρνητικά δείγματα. Η ίδια διαδικασία χρησιμοποιήθηκε τόσο στην αξιολόγηση του pretrained Convnext όσο και στο fine-tuning, ώστε τα αποτελέσματα να είναι άμεσα συγκρίσιμα. Η απόδοση του προεκπαιδευμένου Convnext στο σύνολο DFire συνίσταται στα ακόλουθα αποτελέσματα.

Πίνακας 3-33 Μετρικές του pretrained Convnext στο DFire

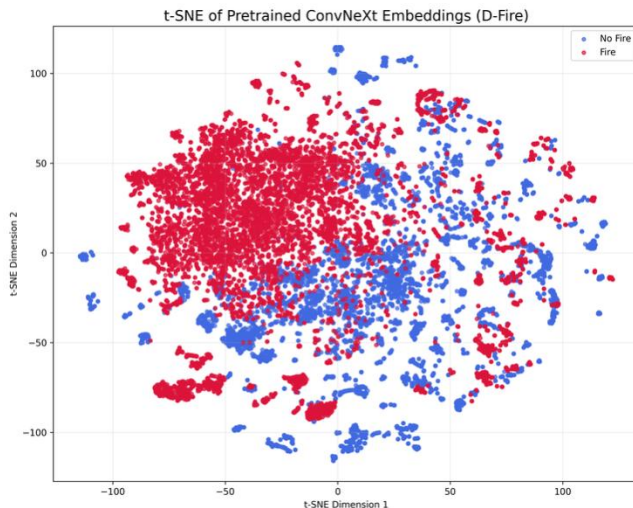
Pretrained Convnext στο DFire	
Accuracy	0.5871
Precision	0.7004
Recall	0.3972
F1-score	0.5069



Εικόνα 3-91 Confusion Matrix του pretrained Convnext στο D-Fire

Η αξιολόγηση του pretrained Convnext στο σύνολο δεδομένων D-Fire απέδωσε Accuracy 58.71%, Precision 70.04%, Recall 39,72% και F1-score 50,69%. Η συνολική ακρίβεια υποδηλώνει ότι το μοντέλο μπορεί να διακρίνει σε κάποιο βαθμό εικόνες με και χωρίς ένδειξη πυρκαγιάς, παρά το γεγονός ότι δεν έχει προσαρμοστεί ειδικά στο συγκεκριμένο πρόβλημα. Ιδιαίτερα υψηλή είναι η τιμή της Precision (70.04%), γεγονός που δείχνει ότι όταν το μοντέλο προβλέπει την παρουσία πυρκαγιάς, η πρόβλεψή του είναι συνήθως σωστή και εμφανίζει σχετικά χαμηλό ποσοστό ψευδώς θετικών (False Positives). Αντίθετα, η Recall (39.72%) είναι αισθητά χαμηλότερη, γεγονός που υποδηλώνει ότι σημαντικός αριθμός πραγματικών περιστατικών πυρκαγιάς δεν ανιχνεύεται και ταξινομείται λανθασμένα ως No Fire (False Negatives). Ως αποτέλεσμα, το F1-score (50.69%), το οποίο συνδυάζει Precision και Recall, παραμένει σε μέτρια επίπεδα, αντανακλώντας την ανισορροπία μεταξύ των δύο μετρικών. Θα πρέπει εδώ να τονίσουμε ότι τα αποτελέσματα αυτά είναι σε ένα βαθμό αναμενόμενα, καθώς το μοντέλο χρησιμοποιήθηκε αποκλειστικά με τα προεκπαιδευμένα βάρη του backbone και μία μη εκπαιδευμένη δυαδική κεφαλή ταξινόμησης, χωρίς καμία διαδικασία fine-tuning στο D-Fire. Παρόλα αυτά, η επίδοση είναι αισθητά καλύτερη από μια τυχαία πρόβλεψη, γεγονός που υποδηλώνει ότι τα

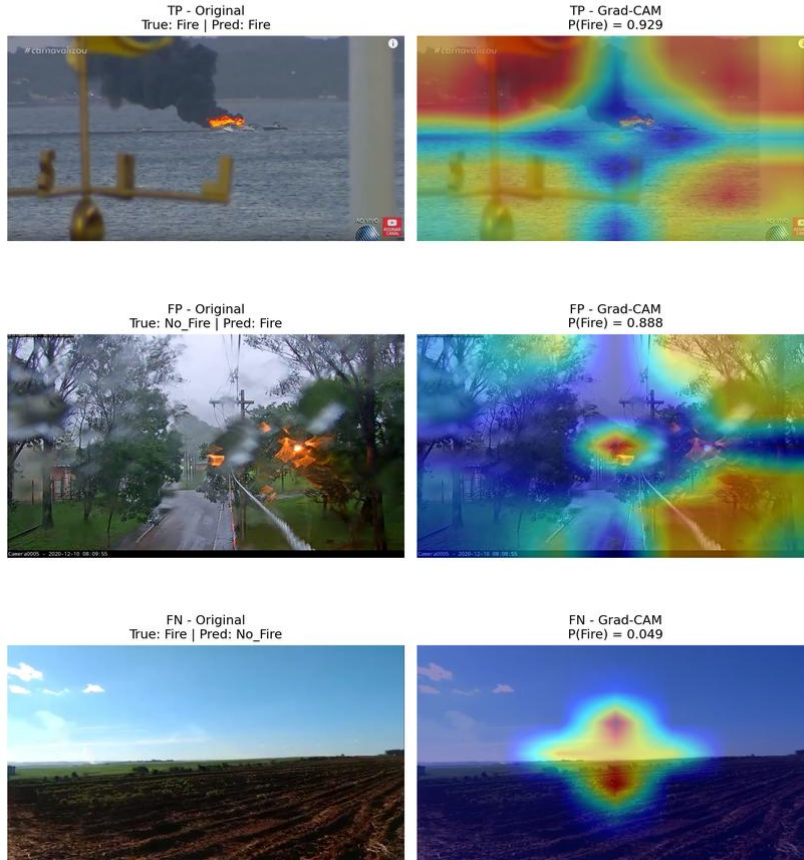
χαρακτηριστικά που έχει μάθει το Convnext κατά την προεκπαίδευση του στο ImageNet περιέχουν χρήσιμη γενική οπτική πληροφορία.



Εικόνα 3-92 t-SNE του pretrained Convnext στο DFire

Το διάγραμμα t-SNE των χαρακτηριστικών που εξήχθησαν από το προεκπαιδευμένο Convnext στο σύνολο δεδομένων D-Fire παρουσιάζει μια μερική αλλά όχι πλήρη διάκριση μεταξύ των δύο κλάσεων (Fire και No Fire). Παρατηρείται ότι μεγάλο μέρος των δειγμάτων της κλάσης Fire συγκεντρώνεται κυρίως στο αριστερό τμήμα του διαγράμματος, σχηματίζοντας μια σχετικά συνεκτική περιοχή, ενώ σημαντικός αριθμός δειγμάτων της κλάσης No Fire εντοπίζεται κυρίως στο κεντρικό και δεξιό τμήμα. Στις περιοχές μετάβασης μεταξύ των δύο ομάδων παρατηρείται έντονη αλληλοεπικάλυψη, γεγονός που υποδηλώνει ότι οι αναπαραστάσεις χαρακτηριστικών που παράγει το pretrained μοντέλο δεν επαρκούν για πλήρη διαχωρισμό των δύο κατηγοριών. Συνεπώς, το t-SNE επιβεβαιώνει ότι, αν και το προεκπαιδευμένο Convnext έχει ήδη μάθει γενικά οπτικά χαρακτηριστικά που επιτρέπουν έναν αρχικό διαχωρισμό των δύο κατηγοριών, οι αναπαραστάσεις αυτές δεν είναι ακόμη επαρκώς εξειδικευμένες για το πρόβλημα της ανίχνευσης δασικών πυρκαγιών.

Grad-CAM: Pretrained ConvNeXt on D-Fire



Εικόνα 3-93 Χάρτες Grad-CAM του pretrained Convnext στο DFire

Όπως βλέπουμε από τα τρία ενδεικτικά ζεύγη εικόνων παραπάνω, στην περίπτωση του True Positive, όπου η εικόνα ταξινομήθηκε σωστά ως Fire με υψηλή πιθανότητα (0.929), ο χάρτης ενεργοποίησης εστιάζει κυρίως στην περιοχή όπου εμφανίζονται οι φλόγες και ο πυκνός καπνός. Η συμπεριφορά αυτή υποδηλώνει ότι, ακόμη και χωρίς fine-tuning, το προεκπαιδευμένο Convnext έχει μάθει να αναγνωρίζει βασικά οπτικά χαρακτηριστικά που σχετίζονται με περιστατικά πυρκαγιάς. Στην περίπτωση του False Positive, το μοντέλο ταξινόμησε λανθασμένα την εικόνα ως Fire με πιθανότητα 0.888. Ο χάρτης Grad-CAM δείχνει ότι η προσοχή του μοντέλου συγκεντρώνεται σε φωτεινές περιοχές του τοπίου και ιδιαίτερα σε αντικείμενα με έντονη χρωματική αντίθεση. Αυτό υποδηλώνει ότι το μοντέλο επηρεάζεται από οπτικά μοτίβα τα οποία παρουσιάζουν ομοιότητες με χαρακτηριστικά πυρκαγιάς. Τέλος, στην περίπτωση του False Negative, ο χάρτης

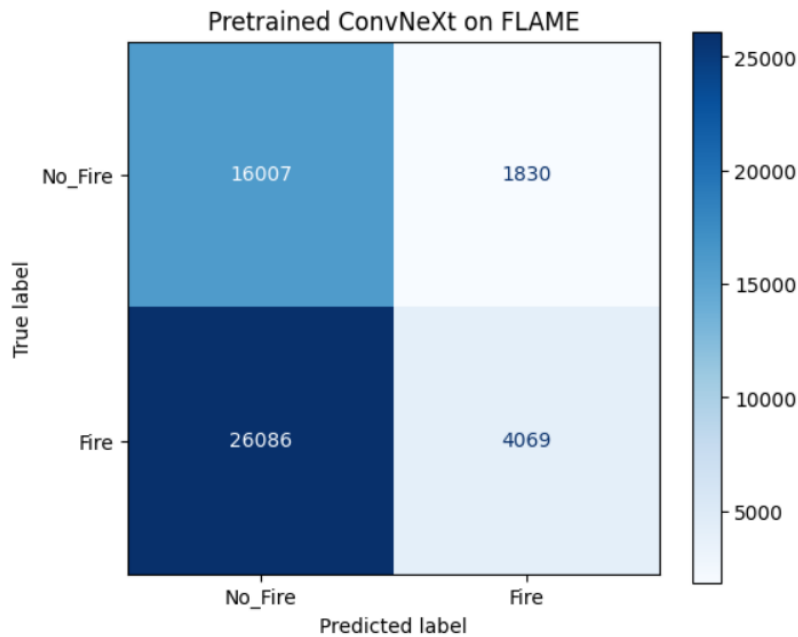
ενεργοποίησης δεν εστιάζει στην περιοχή όπου βρίσκεται η μικρή εστία φωτιάς, αλλά κατανέμει την προσοχή του κυρίως στον ουρανό και στον ορίζοντα. Καταλαβαίνουμε λοιπόν πως το μοντέλο δυσκολεύεται να αναγνωρίσει μικρές ή απομακρυσμένες εστίες πυρκαγιάς, όταν αυτές καταλαμβάνουν πολύ μικρό ποσοστό της εικόνας.

3.11.2 στο FLAME

Η απόδοση του προεκπαιδευμένου μοντέλου Convnext στο σύνολο FLAME μας έδωσε τα ακόλουθα αποτελέσματα.

Πίνακας 3-34 Μετρικές του pretrained Convnext στο FLAME

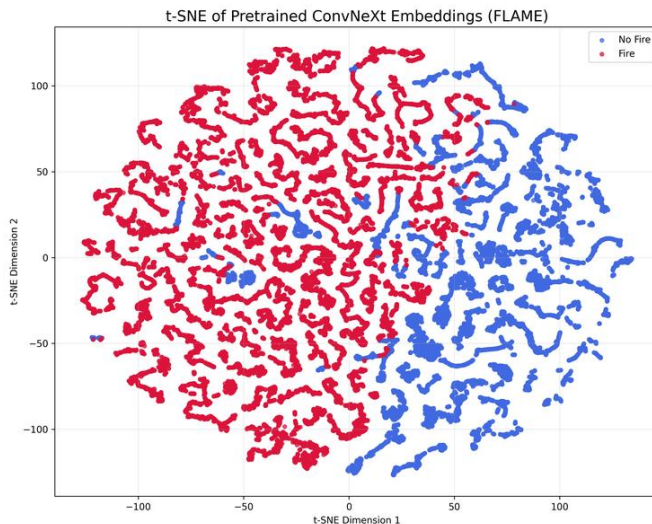
Pretrained Convnext στο FLAME	
Accuracy	0.4183
Precision	0.6898
Recall	0.1349
F1-score	0.2257



Εικόνα 3-94 Confusion Matrix του pretrained Convnext στο FLAME

Από την αξιολόγηση του προεκπαιδευμένου Convnext στο σύνολο δεδομένων FLAME προέκυψαν οι τιμές Accuracy 41.83%, Precision 68.98%, Recall 13.49% και F1-score 22.57%. Παρότι η τιμή της Precision παραμένει σχετικά υψηλή, γεγονός που υποδηλώνει ότι όταν το μοντέλο προβλέπει την παρουσία πυρκαγιάς οι προβλέψεις του είναι συχνά ορθές, η ιδιαίτερα χαμηλή τιμή της Recall αποδεικνύει ότι το μεγαλύτερο ποσοστό των πραγματικών περιστατικών πυρκαγιάς δεν ανιχνεύεται. Το αποτέλεσμα αυτό αντανακλάται και στο χαμηλό F1-score, το οποίο υποδεικνύει περιορισμένη συνολική ικανότητα του μοντέλου να ισορροπήσει μεταξύ σωστών θετικών προβλέψεων και ανίχνευσης όλων των πραγματικών περιστατικών πυρκαγιάς. Η εικόνα αυτή επιβεβαιώνεται και από τον confusionmatrix, όπου παρατηρούνται 26.086 False Negatives έναντι μόλις 4.069 True Positives. Αντίθετα, η κλάση No Fire αναγνωρίζεται με σημαντικά μεγαλύτερη επιτυχία, καθώς 16.007 από τις 17.837 εικόνες ταξινομήθηκαν σωστά. Το γεγονός αυτό δείχνει ότι το προεκπαιδευμένο μοντέλο παρουσιάζει έντονη τάση να χαρακτηρίζει μεγάλο αριθμό εικόνων ως No Fire, ακόμη και όταν αυτές περιέχουν πραγματικές εστίες πυρκαγιάς. Η συμπεριφορά αυτή κρίνεται οπωσδήποτε ιδιαίτερα ανεπιθύμητη σε εφαρμογές έγκαιρης ανίχνευσης δασικών

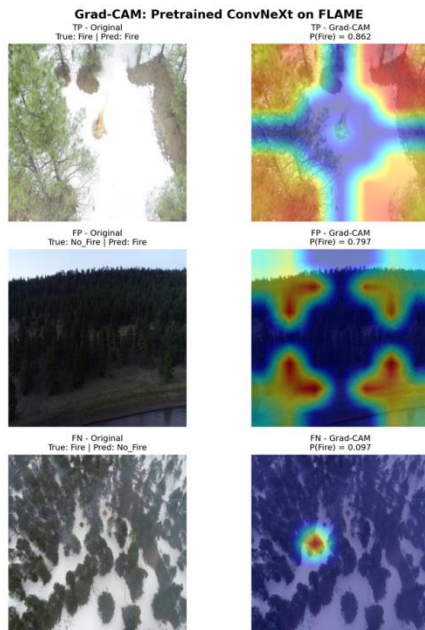
πυρκαγιών, όπου η αποτυχία εντοπισμού μιας πραγματικής πυρκαγιάς μπορεί να έχει σημαντικές επιχειρησιακές συνέπειες.



Εικόνα 3-95 t-SNE του pretrained Convnext στοFLAME

Το διάγραμμα t-SNE παρουσιάζει έναν σχετικά σαφή αλλά όχι πλήρη διαχωρισμό μεταξύ των δύο κλάσεων (Fire και No Fire). Παρότι οι δύο κλάσεις σχηματίζουν διακριτές περιοχές, εξακολουθεί να υπάρχει σημαντική ζώνη αλληλοεπικάλυψης στο κεντρικό τμήμα του διαγράμματος, όπου δείγματα των δύο κατηγοριών εμφανίζονται αναμειγμένα. Η παρατήρηση αυτή υποδηλώνει ότι οι αναπαραστάσεις που παράγει το προεκπαιδευμένο Convnext διαθέτουν ήδη κάποια διακριτική ικανότητα, χωρίς όμως να είναι επαρκώς εξειδικευμένες ώστε να επιτυγχάνουν αξιόπιστο διαχωρισμό των εικόνων με και χωρίς πυρκαγιά. Με την βοήθεια των χαρτών Grad-CAM επιβεβαιώνουμε ότι το προεκπαιδευμένο Convnext διαθέτει ήδη τη δυνατότητα να εξάγει γενικές οπτικές αναπαραστάσεις που σχετίζονται με την παρουσία φωτιάς, όμως η προσοχή του παραμένει αρκετά διάχυτη και επηρεάζεται από χαρακτηριστικά του φυσικού περιβάλλοντος που δεν αποτελούν πραγματικές εστίες πυρκαγιάς. Ειδικότερα η περίπτωση του False Negative (τρίτη στην σειρά εικόνων παρακάτω) είναι χαρακτηριστική. Παρότι η εικόνα περιέχει πραγματική εστία πυρκαγιάς, το μοντέλο την ταξινομεί ως No Fire. Ο χάρτης ενεργοποίησης παρουσιάζει μια μικρή και σχετικά αδύναμη

συγκέντρωση προσοχής γύρω από την περιοχή της φωτιάς, χωρίς όμως να αναδεικνύει επαρκώς την πραγματική εστία. Το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο δυσκολεύεται να αναγνωρίσει μικρές ή απομακρυσμένες πυρκαγιές, ιδιαίτερα όταν αυτές καταλαμβάνουν περιορισμένο τμήμα της εικόνας.



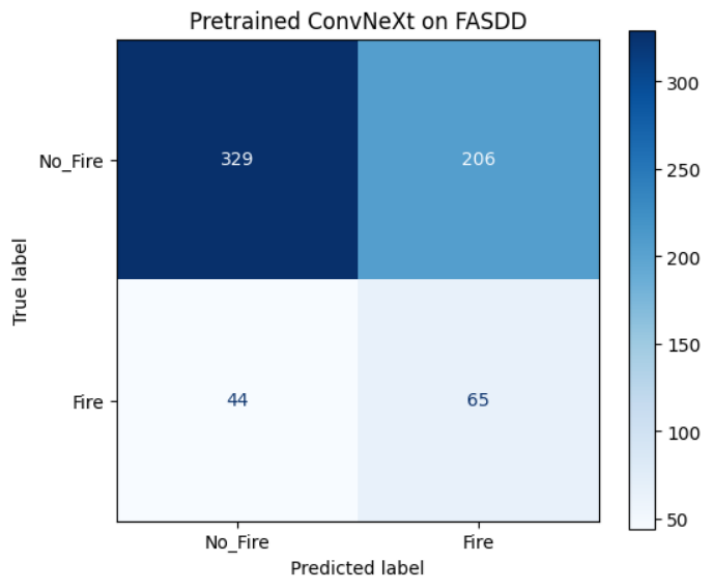
Εικόνα 3-96 Χάρτες Grad-CAM του pretrained Convnext στο FLAME

3.11.3 στο FASDD

Η απόδοση του προεκπαιδευμένου μοντέλου Convnext στο σύνολο FASDD μας έδωσε τα ακόλουθα αποτελέσματα.

Πίνακας 3-35 Μετρικές του pretrained Convnext στο FASDD

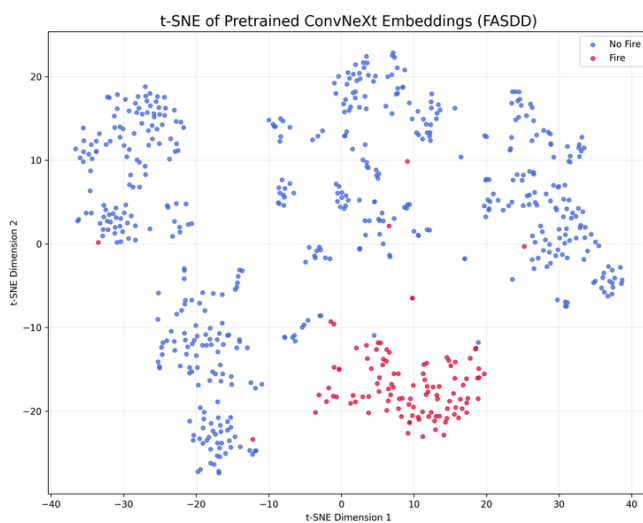
Pretrained Convnext στο FASDD	
Accuracy	0.6118
Precision	0.2399
Recall	0.5963
F1-score	0.3421



Εικόνα 3-97 Confusion Matrix του pretrained Convnext στο FASDD

Τα ποσοστά των μετρικών είναι Accuracy 61.18%, Precision 23.99%, Recall 59.63% και F1-score 34.21%. Η συνολική ακρίβεια είναι η υψηλότερη από αυτή που παρατηρήθηκε στα τρία σύνολα δεδομένων που αξιολογήθηκαν με το pretrained μοντέλο Convnext, γεγονός που υποδηλώνει ότι πιθανώς οι γενικές οπτικές αναπαραστάσεις που έχει μάθει το Convnext από το ImageNet είναι περισσότερο συμβατές με τα χαρακτηριστικά των εικόνων του FASDD. Ωστόσο, η ιδιαίτερα χαμηλή Precision (23.99%) δείχνει ότι μεγάλο ποσοστό των εικόνων που προβλέπονται ως Fire είναι στην πραγματικότητα εικόνες χωρίς πυρκαγιά, οδηγώντας σε σημαντικό αριθμό ψευδώς θετικών προβλέψεων. Όσον αφορά στον confusion matrix, παρατηρούμε 65 True Positives, 44 False Negatives, 329 True Negatives και 206 False Positives. Ειδικότερα, το σχετικά μεγάλο πλήθος ψευδώς θετικών προβλέψεων εξηγεί τη χαμηλή τιμή της Precision, ενώ ο μικρότερος αριθμός ψευδώς αρνητικών συγκριτικά με τα άλλα datasets οδηγεί στην αυξημένη Recall. Το αποτέλεσμα αυτό υποδηλώνει ότι το pretrained μοντέλο παρουσιάζει μεγαλύτερη ευαισθησία στην ανίχνευση πιθανών εστιών πυρκαγιάς, ακόμη και αν αυτό συνεπάγεται αυξημένο αριθμό λανθασμένων συναγερμών. Από πρακτικής άποψης, η συμπεριφορά αυτή είναι προτιμότερη από την αντίθετη περίπτωση, καθώς στις εφαρμογές έγκαιρης ανίχνευσης δασικών πυρκαγιών η απώλεια μιας

πραγματικής πυρκαγιάς (False Negative) θεωρείται συνήθως πιο κρίσιμη από την εμφάνιση ενός ψευδούς συναγερμού. Συνολικά, τα αποτελέσματα δείχνουν ότι αναπαραστάσεις που έχει μάθει κατά την προεκπαίδευση είναι ικανές για να αναγνωρίζει το μοντέλο αρκετά από τα οπτικά μοτίβα που σχετίζονται με την παρουσία πυρκαγιάς αλλά δεν είναι επαρκώς εξειδικευμένες για μια αξιόπιστη δυαδική ταξινόμηση.

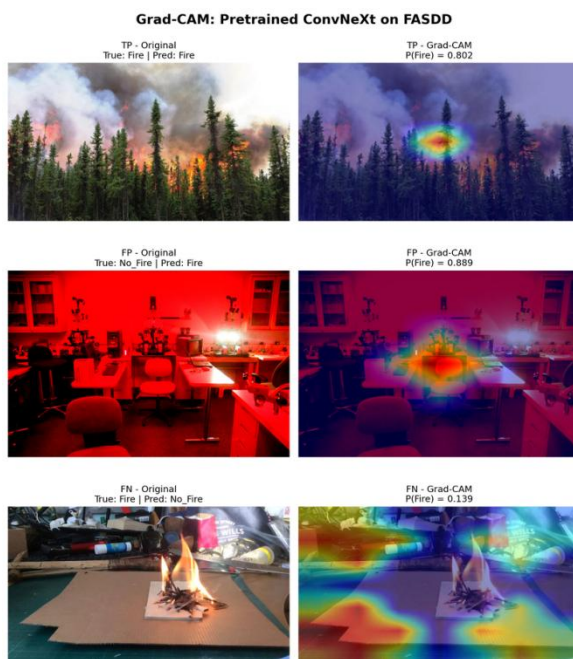


Εικόνα 3-98 t-SNE του pretrained Convnext στοFASDD

Το διάγραμμα t-SNE παρουσιάζει έναν μερικό αλλά αρκετά διακριτό διαχωρισμό μεταξύ των δύο κλάσεων. Παρατηρείται ότι οι περισσότερες εικόνες της κλάσης Fire συγκεντρώνονται σε μια σχετικά συμπαγή περιοχή στο κάτω κεντρικό τμήμα του χώρου χαρακτηριστικών, ενώ οι εικόνες της κλάσης No Fire κατανέμονται σε πολλαπλές διακριτές συστάδες που καταλαμβάνουν το μεγαλύτερο μέρος του υπόλοιπου χώρου. Παρά τον ικανοποιητικό συνολικό διαχωρισμό, εμφανίζονται ορισμένα δείγματα της κλάσης Fire μέσα στις περιοχές της κλάσης No Fire, καθώς και λίγες περιοχές αλληλοεπικάλυψης μεταξύ των δύο κατηγοριών. Η εικόνα αυτή υποδηλώνει ότι το προεκπαιδευμένο Convnext έχει ήδη μάθει γενικά χαρακτηριστικά που επιτρέπουν έναν ικανοποιητικό διαχωρισμό των δύο κλάσεων, χωρίς όμως να επιτυγχάνει πλήρη διάκριση όλων των δειγμάτων. Η οπτική αυτή αναπαράσταση συμφωνεί με τις ποσοτικές μετρικές αξιολόγησης. Η σχετικά καλή τιμή του Recall (59.63%) δείχνει ότι το μοντέλο καταφέρνει να

εντοπίζει έναποσοστό των πραγματικών εικόνων πυρκαγιάς, γεγονός που αντανακλάται στη δημιουργία μιας σχετικά συμπαγούς συστάδας για την κλάση Fire. Ωστόσο, η ιδιαίτερα χαμηλή Precision (23.99%) υποδηλώνει ότι αρκετές εικόνες χωρίς πυρκαγιά εμφανίζουν χαρακτηριστικά παρόμοια με εκείνα της κλάσης Fire, οδηγώντας σε σημαντικό αριθμό ψευδώς θετικών προβλέψεων.

Οι χάρτες Grad-CAM επιβεβαιώνουν ότι το προεκπαιδευμένο Convnext διαθέτει ήδη τη δυνατότητα να αναγνωρίζει γενικά οπτικά χαρακτηριστικά που σχετίζονται με την παρουσία πυρκαγιάς, όπως φλόγες, καπνός καθώς και περιοχές υψηλής φωτεινότητας. Ωστόσο, η προσοχή του μοντέλου δεν είναι πάντοτε επαρκώς σωστή, καθώς επηρεάζεται και από στοιχεία του περιβάλλοντος που παρουσιάζουν παρόμοια με τη φωτιά οπτικά χαρακτηριστικά. Η παρατήρηση αυτή βρίσκεται σε αντιστοιχία με τις ποσοτικές μετρικές αξιολόγησης, όπου καταγράφηκε σχετικά υψηλή Recall (59.63%), αλλά ιδιαίτερα χαμηλή Precision (23.99%), πράγμα που σημαίνει ότι το μοντέλο ανιχνεύει αρκετές πραγματικές πυρκαγιές, αλλά παράλληλα ενεργοποιείται και σε εικόνες χωρίς φωτιά.



Εικόνα 3-99 Χάρτες Grad-CAM του pretrained Convnext στο FASDD

3.12 Finetuning του μοντέλου Convnext

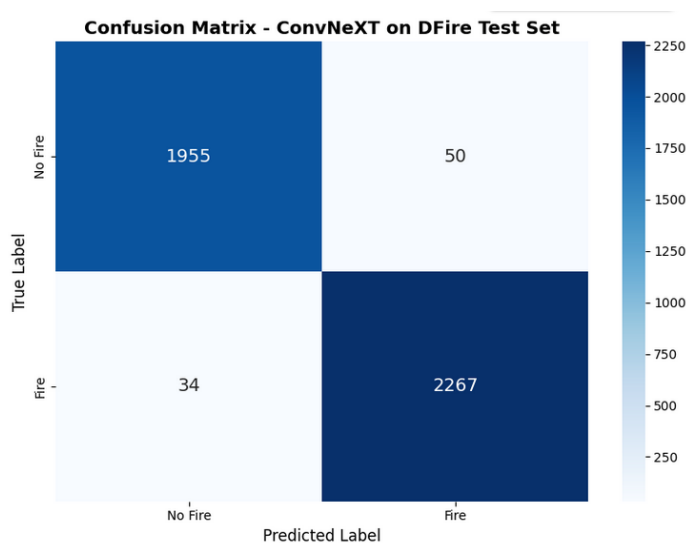
Στα πλαίσια της παραγράφου αυτής προχωρούμε στη διαδικασία finetuning για το μοντέλο Convnext.

3.12.1 στο DFire

Από το finetuning του μοντέλου Convnext στο DFire τα αποτελέσματα που πήραμε είναι τα εξής.

Πίνακας 3-36 Μετρικές του fine-tuned Convnext στο DFire

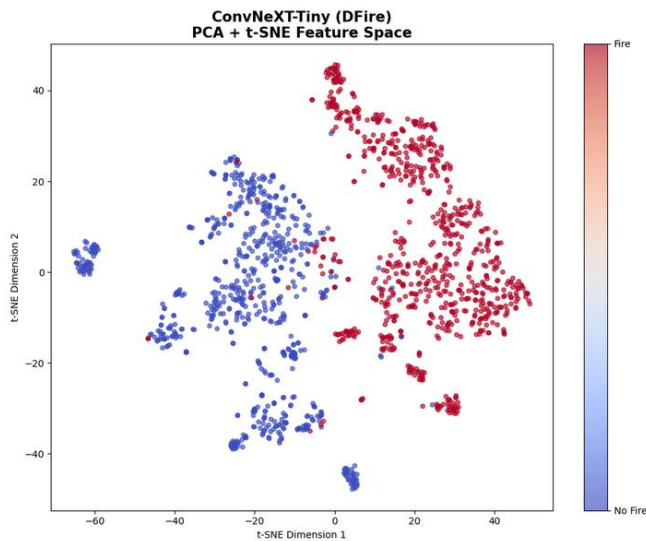
Fine tuned Convnext στο DFire	
Accuracy	0.9805
Precision	0.9784
Recall	0.9852
F1-score	0.9818



Εικόνα 3-100 Confusion Matrix του fine-tuned Convnext στο DFire

Οι πολύ καλές και ισορροπημένες τιμές όλων των μετρικών στον παραπάνω πίνακα πιστοποιούν ότι το μοντέλο διακρίνει με πολύ μεγάλη αξιοπιστία τις εικόνες με και χωρίς πυρκαγιά, παρουσιάζοντας εξαιρετική συνολική απόδοση. Αξιοσημείωτη θα λέγαμε πως είναι η σχεδόν ιδανική ισορροπία μεταξύ Precision και Recall. Η διαφορά μεταξύ των δύο μετρικών είναι μικρότερη από μία ποσοστιαία μονάδα, κάτι που δηλώνει ότι το μοντέλο δεν παρουσιάζει εμφανή μεροληψία προς κάποια από τις δύο κλάσεις. Αντιθέτως, επιτυγχάνει ταυτόχρονα πολύ χαμηλό αριθμό τόσο False Positives όσο και False Negatives. Πιο συγκεκριμένα ο confusionmatrix μας πληροφορεί ότι από τις 2.005 εικόνες χωρίς πυρκαγιά, οι 1.955 ταξινομήθηκαν σωστά, ενώ μόλις 50 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Αντίστοιχα, από τις 2.301 εικόνες με πυρκαγιά, οι 2.267 αναγνωρίστηκαν σωστά και μόνο 34 ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives).

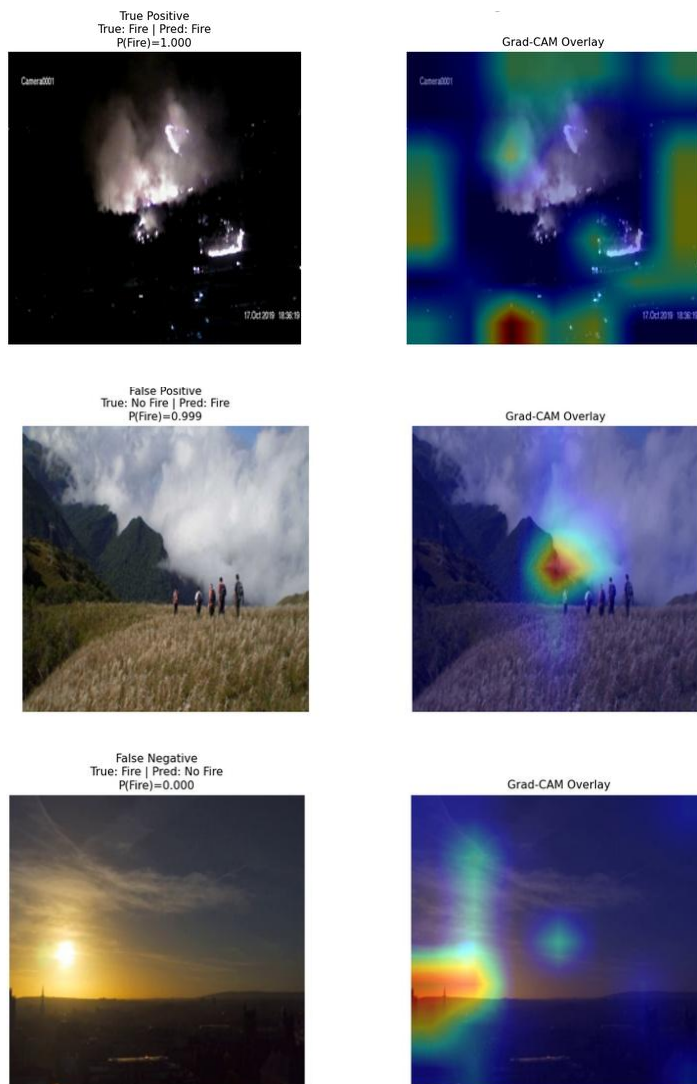
Από την ανασκόπηση του t-SNE βγάζουμε το συμπέρασμα πως το Convnext έχει μάθει έναν ιδιαίτερα καλά οργανωμένο χώρο χαρακτηριστικών, στον οποίο οι δύο κλάσεις (Fire και No Fire) διαχωρίζονται με μεγάλη σαφήνεια. Οι περισσότερες εικόνες συγκεντρώνονται σε δύο διακριτές ομάδες με περιορισμένη μόνο επικάλυψη, γεγονός που υποδηλώνει ότι το μοντέλο έχει αναπτύξει αναπαραστάσεις χαρακτηριστικών με υψηλή διακριτική ικανότητα. Η εικόνα αυτή βρίσκεται σε άμεση συνάφεια με τις πολύ υψηλές τιμές Accuracy, Precision, Recall και F1-score που καταγράψαμε κατά την αξιολόγηση του μοντέλου.



Εικόνα 3-101 t-SNE του fine-tunedConvnext στο DFire

Κοιτώντας τις ακόλουθες Grad-CAM ενεργοποιήσεις μπορούμε να πούμε πως το Convnext πως στις λίγες περιπτώσεις που σφάλει η ενεργοποίηση του εστιάζει σε κάποιο στοιχείο που παρουσιάζει μεγάλο βαθμό ομοιότητας με καπνό (όπως τα πυκνά και συμπαγή σύννεφα στη 2^η σειρά εικόνων-περίπτωση FalsePositive). Επίσης στην 3^η σειρά εικόνων (περίπτωση FalseNegative) η έντονη λάμψη του ήλιου είναι αυτή που κερδίζει την προσοχή του μοντέλου χάνοντας έτσι από το «οπτικό πεδίο ενδιαφέροντός του» τη δυσδιάκριτη μικρή εστία καπνού. Ιδιαίτερο ενδιαφέρον παρουσιάζει και η περίπτωση του TruePositive (1^η σειρά εικόνων), όπου ενώ στην εικόνα βλέπουμε την εκδήλωση μιας μεγάλης δασικής πυρκαγιάς η οποία και προβλέπεται από το μοντέλο και μάλιστα με απόλυτη πιθανότητα ($P(\text{Fire})=1.000$) ο Grad-Cam χάρτης δείχνει πως η προσοχή του μοντέλου εστιάζεται περισσότερο στον σκοτεινό περιβάλλοντα γύρω από τη φωτιά χώρο. Το φαινόμενο αυτό ίσως έχει την εξήγησή του στην ίδια την αρχιτεκτονική του μοντέλου και τον τρόπο που έχει αυτό μάθει να χρησιμοποιεί και να ερμηνεύει το ευρύτερο χωρικό πλαίσιο μιας πυρκαγιάς. Ενδέχεται δηλαδή το μοντέλο να έχει μάθει να συνδέει μια φωτιά με πληροφορίες του περιβάλλοντος αυτής χώρου όπως τα όρια του καπνού, σκιές, εξάπλωση της φωτεινότητας των φλογών, αντιθέσεις με το σκοτεινό υπόβαθρο. Αν κάτι τέτοιο ισχύει θα εξηγούσε και γιατί η θερμική

περιοχή του Grad-CAM στο TruePositive ενεργοποιείται κυρίως γύρω από τη μεγάλη φωτιά.



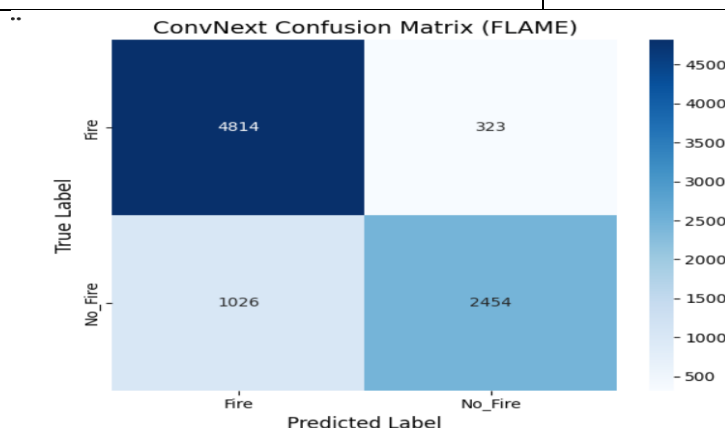
Εικόνα 3-102Χάρτες Grad-CAM του fine-tuned Convnext στο DFire

3.12.2 στο FLAME

Από το finetuning του μοντέλου στο FLAME τα αποτελέσματα στις μετρικές για συγκεντρώνονται στον πίνακα

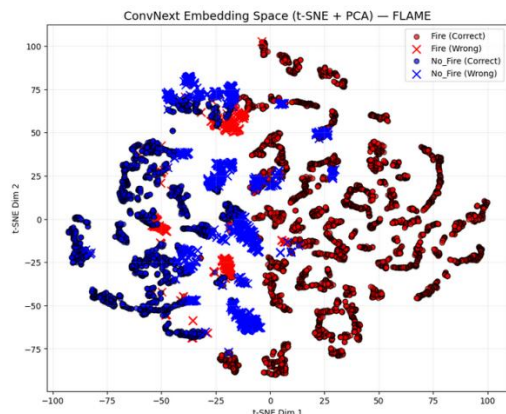
Πίνακας 3-37 Μετρικές του fine-tuned Convnext στο FLAME

Fine tuned Convnext στο FLAME	
Accuracy	0.8434
Precision	0.8243
Recall	0.9371
F1-score	0.8771



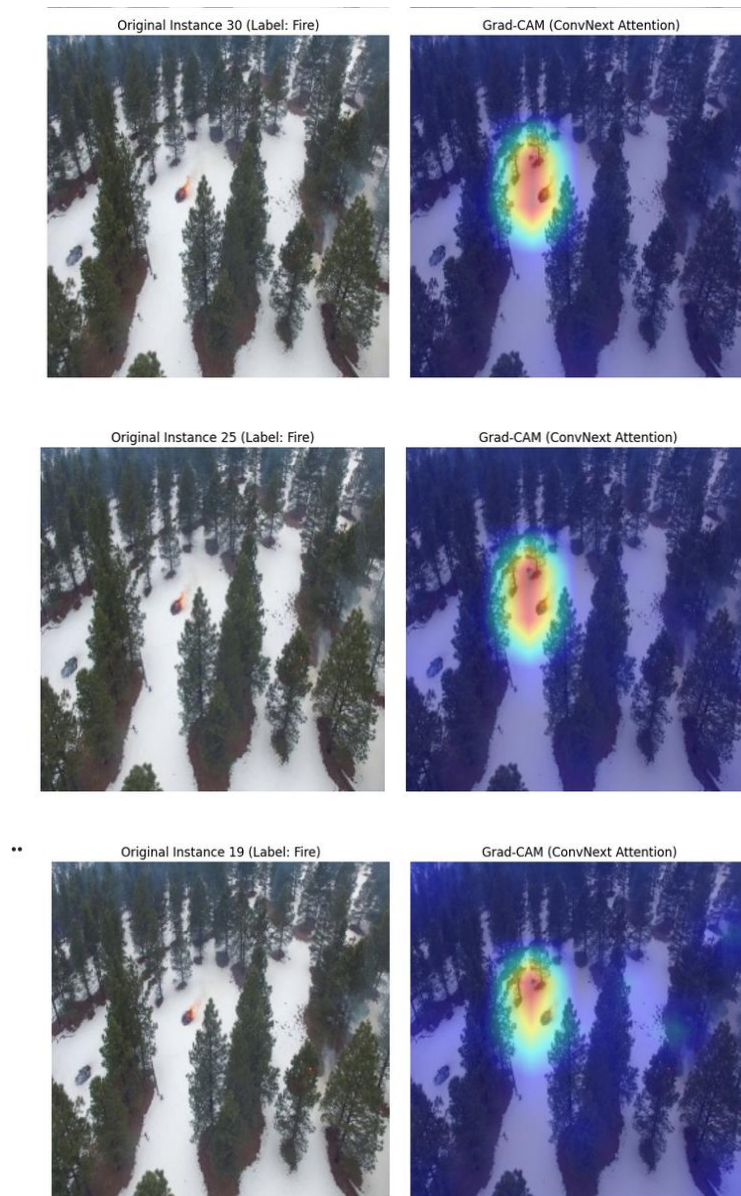
Εικόνα 3-103 Confusion Matrix του fine-tuned Convnext στο FLAME

Περιδιαβάζοντας τα αποτελέσματα των μετρικών στον παραπάνω πίνακα βλέπουμε ότι το μοντέλο διατηρεί πολύ υψηλό Recall χωρίς να θυσιάζει σημαντικά την Precision. Στην πράξη αυτό σημαίνει ότι το μοντέλο προτιμά να χαρακτηρίζει μια αμφίβολη εικόνα ως Fire, μειώνοντας έτσι σημαντικά τις χαμένες πυρκαγιές (False Negatives), ακόμη και αν αυτό συνεπάγεται μια μικρή αύξηση των False Positives. Ειδικότερα παρατηρώντας το confusionmatrixέχουμε ότι από τις 5.137 εικόνες με πυρκαγιά, οι 4.814 ταξινομήθηκαν σωστά, ενώ μόλις 323ταξινομήθηκαν λανθασμένα ως No Fire (False Negatives). Αντίστοιχα, από τις 3.480 εικόνες χωρίς πυρκαγιά, οι2.454 ταξινομήθηκαν σωστά, ενώ 1.026 ταξινομήθηκαν λανθασμένα ως Fire (False Positives). Το αποτέλεσμα αυτό δείχνει ότι το μοντέλο παρουσιάζει πολύ υψηλή ικανότητα εντοπισμού της φωτιάς, με το κύριο σημείο αδυναμίας του να εντοπίζεται στους σχετικά αυξημένους ψευδείς συναγερμούς.



Εικόνα 3-104 t-SNE του fine-tuned Convnext στο FLAME

Η εικόνα που μας παρέχει το διάγραμμα t-SNEως προς την οργάνωση των χαρακτηριστικών που εξάγει το μοντέλο είναι αυτή ενός εμφανούς διαχωρισμού μεταξύ των κλάσεων. Οι εικόνες Fire (κόκκινα σημεία) καταλαμβάνουν κυρίως το δεξί τμήμα του χώρου χαρακτηριστικών, ενώ οι εικόνες No Fire (μπλε σημεία) συγκεντρώνονται κυρίως στο αριστερό τμήμα. Ο γενικός αυτός διαχωρισμός δείχνει ότι το μοντέλο έχει αναπτύξει αναπαραστάσεις με υψηλή διακριτική ικανότητα, ικανές να διαχωρίζουν αποτελεσματικά τις δύο βασικές κατηγορίες. Επιπλέον με το σύμβολο “ X” σημειώνονται τα λανθασμένα ταξινομημένα δείγματα. Στο διάγραμμα t-SNE παρατηρούμε ότι εμφανίζονται αρκετές περιοχές όπου κόκκινα και μπλε σημεία συνυπάρχουν. Οι περισσότερες από αυτές τις περιοχές συνοδεύονται από σύμβολα “X”. Αυτό μας δηλώνει ότι υπάρχουν εικόνες των οποίων τα χαρακτηριστικά βρίσκονται πολύ κοντά στο όριο μεταξύ των δύο κλάσεων. Θα ήταν πιθανό να εικάσουμε ότι οι εικόνες αυτές περιλαμβάνουν σκηνές με αραιό καπνό, μικρές εστίες φωτιάς, ή φωτισμούς και αντανάκλασεις, στοιχεία που συνήθως δυσκολεύουν την ταξινόμηση στο FLAME.



Εικόνα 3-105 Χάρτες Grad-CAM του fine-tuned Convnext στο FLAME

Από την ανάγνωση των θερμικών χαρτών Grad-CAM παραπάνω βλέπουμε ότι οι απεικονίσεις Grad-CAM μας δείχνουν ότι το μοντέλο εστιάζει με ιδιαίτερα ακρίβεια στις περιοχές όπου βρίσκεται η πραγματική εστία της πυρκαγιάς. Παρακολουθούμε πως και στις τρεις εικόνες η περιοχή μέγιστης ενεργοποίησης συμπίπτει σχεδόν πλήρως με τις φλόγες και το άμεσο περιβάλλον τους, ενώ οι υπόλοιπες περιοχές της εικόνας παρουσιάζουν σημαντικά χαμηλότερη ένταση ενεργοποίησης. Να επισημάνουμε σε αυτό το σημείο πως και τα τρία instances με

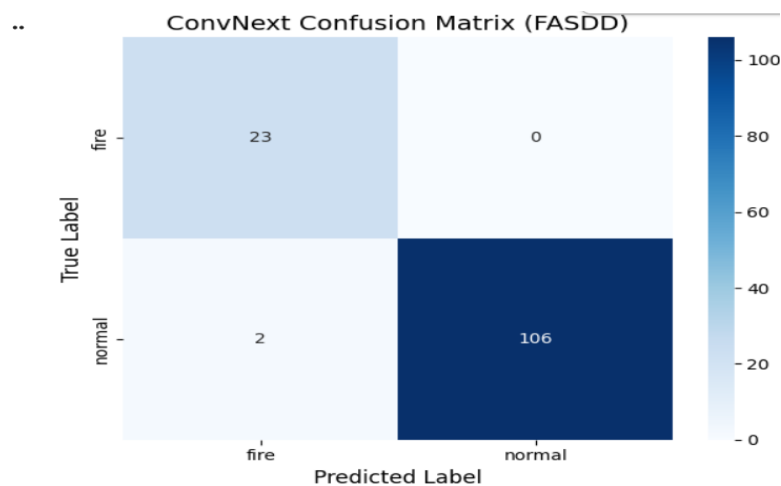
τα Grad-CAM visualizations που τα συνοδεύουν είναι σωστές προβλέψεις του μοντέλου (TruePositives).

3.12.3 στο FASDD

Από το finetuning του μοντέλου Convnext στο FASDD τα αποτελέσματα στις μετρικές μετά την αξιολόγηση του μοντέλου στο testset είναι τα εξής:

Πίνακας 3-38 Μετρικές του fine-tuned Convnext στο FASDD

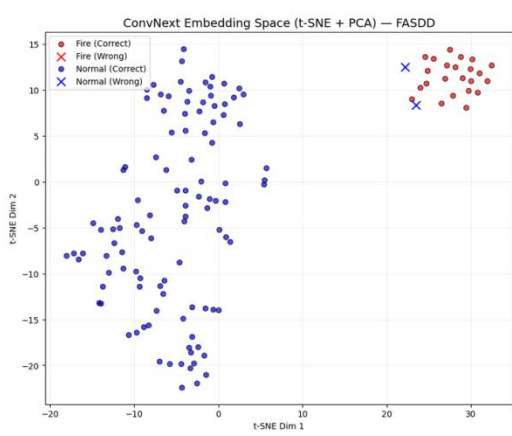
Fine tuned Convnext στο FASDD	
Accuracy	0.9847
Precision	0.9200
Recall	1.0000
F1-score	0.9583



Εικόνα 3-106 Confusion Matrix του fine-tuned Convnext στο FASDD

Ένα σημείο που κάνει εντύπωση στην ανάγνωση του παραπάνω πίνακα είναι ο απόλυτος βαθμός στο Recall (100%). Το στοιχείο αυτό μας λέει ότι το μοντέλο δεν παρέλειψε καμία πραγματική περίπτωση πυρκαγιάς. Συγχρόνως, η τιμή της μετρικής Precision 92,00% μας δείχνει ότι η μεγάλη πλειονότητα των εικόνων που

ταξινομήθηκαν ως Fire αντιστοιχούσε πράγματι σε πραγματικές πυρκαγιές, με ελάχιστους ψευδείς συναγερμούς. Ο confusionmatrix μας διαφωτίζει περαιτέρω ως προς την αμιγώς ποσοτική διάσταση της ταξινόμησης του μοντέλου. Συγκεκριμένα και οι 23 εικόνες πυρκαγιάς ταξινομήθηκαν σωστά ως Fire, χωρίς να καταγραφεί καμία λανθασμένη αρνητική πρόβλεψη (False Negative = 0). Από τις 108 εικόνες χωρίς πυρκαγιά, οι 106 ταξινομήθηκαν σωστά ως Normal (δηλαδή No Fire), ενώ μόλις 2 εικόνες ταξινομήθηκαν λανθασμένα ως Fire (False Positives).



Εικόνα 3-107 t-SNE του fine-tuned Convnext στο FASDD

Το διάγραμμα t-SNE μας δίνει μια πολύ καθαρή εικόνα του χώρου των χαρακτηριστικών και επιβεβαιώνει ότι το Convnext αναπτύσσει ιδιαίτερα διακριτικές και σταθερές αναπαραστάσεις χαρακτηριστικών για το σύνολο δεδομένων FASDD. Εν προκειμένω εδώ έχουμε σχεδόν πλήρη διαχωρισμό των δύο κλάσεων με ελάχιστες λανθασμένες ταξινομήσεις. Συγκεκριμένα οι λανθασμένες ταξινομήσεις περιορίζονται σε μόλις δύο εικόνες της κλάσης Normal οι οποίες σημειώνονται με το σύμβολο “ X “ στο διάγραμμα t-SNE. Οι περιπτώσεις των λανθασμένων αυτών ταξινομήσεων βλέπουμε ότι βρίσκονται κοντά στη συστάδα της κλάσης Fire, κάτι που ενδεχομένως υποδηλώνει ότι οι εικόνες αυτές παρουσιάζουν οπτικά χαρακτηριστικά παρόμοια με αυτά μιας πραγματικής πυρκαγιάς. Επιπρόσθετα παρατηρούμε την πολύ μικρή διασπορά που παρουσιάζει η συστάδα των κόκκινων σημείων (κλάση Fire). Αυτή η γεωμετρική πληροφορία μας επισημαίνει ότι ανεξάρτητα από τις διαφορές μεταξύ

των εικόνων (μέγεθος φλόγας, φωτισμός, γωνία λήψης), το μοντέλο τις αναπαριστά με παρόμοιο τρόπο στον χώρο χαρακτηριστικών.



Εικόνα 3-108 Χάρτες Grad-CAM του fine-tuned Convnext στο FASDD

Στις παραπάνω τρεις περιπτώσεις TruePositive παρατηρούμε ότι το Grad-CAM ενεργοποιεί περιοχές που είναι άμεσα συναρτώμενες με τις εκδηλωθείσες πυρκαγιές. Βλεπουμε ότι το μοντέλο ενεργοποιεί την προσοχή του και σε μέρος των εστιών φωτιάς και του καπνού αλλά και περιμετρικά της υπό εξέλιξης φωτιάς. Το τελευταίο αυτό σημείο, της εν γένει προσοχής του μοντέλου στο συνολικό χώρο της εικόνας το είχαμε παρατηρήσει ξανά στις Grad-CAM ενεργοποιήσεις του μοντέλου στο DFire (βλ. παράγραφος 3.12.1).

Κεφάλαιο 4 Συγκριτική Ανασκόπηση

Στο κεφάλαιο αυτό θα προσπαθήσουμε να εξάγουμε κάποια συμπεράσματα από τη σύγκριση της απόδοσης του κάθε μοντέλου στην προεκπαιδευμένη και τη finetuned του εκδοχή.

4.1 στο DFire

Αρχικά θα δούμε τη σύγκριση κάθε μοντέλου, στη προεκπαιδευμένη και τη finetuned μορφή του στο σύνολο DFire.

4.1.1 Το μοντέλο SegFormer-B0

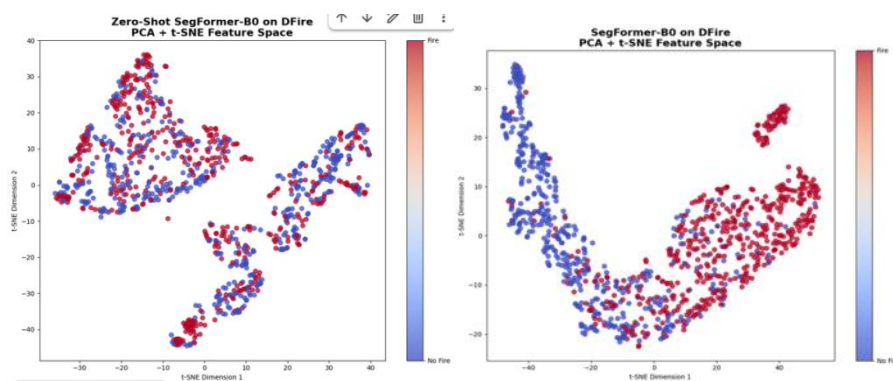
Η απόδοση του SegFormer-B0 στο DFire στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-1 Σύγκριση Μετρικών SegFormer στο DFire

Μετρική	Pretrained SegFormer-B0	Fine tuned SegFormer-B0	Μεταβολή
Accuracy	0.5395	0.8509	+0.3114
Precision	0.5398	0.8738	+0.3340
Recall	0.9361	0.8555	-0.0806
F1-score	0.6848	0.8645	+0.1797

Τα αποτελέσματα του Πίνακα δείχνουν ότι η προσαρμογή του μοντέλου SegFormer στα χαρακτηριστικά του DFire βελτίωσε ουσιαστικά την ικανότητά του να διακρίνει μεταξύ εικόνων με και χωρίς πυρκαγιά. Η Accuracy αυξήθηκε από 53,95% σε 85,09%, παρουσιάζοντας ιδιαίτερα σημαντική βελτίωση της τάξης των 31,14 ποσοστιαίων μονάδων. Μετά το fine tuning το μοντέλο ταξινομεί σωστά πολύ μεγαλύτερο ποσοστό των εικόνων του συνόλου δοκιμής, μειώνοντας σημαντικά τον συνολικό αριθμό λανθασμένων προβλέψεων. Ακόμη μεγαλύτερη είναι η βελτίωση της Precision, η οποία αυξήθηκε από 53,98% σε 87,38%. Η συγκεκριμένη αύξηση υποδηλώνει ότι το fine-tuned μοντέλο παράγει πολύ

λιγότερους ψευδείς συναγερμούς, καθώς η πλειονότητα των εικόνων που χαρακτηρίζονται ως Fire αντιστοιχεί πλέον σε πραγματικά περιστατικά πυρκαγιάς. Η μοναδική μετρική που παρουσίασε μικρή μείωση είναι η Recall, η οποία μειώθηκε από 93,61% σε 85,55%. Παρότι η μείωση αυτή δείχνει ότι το μοντέλο χάνει ελαφρώς περισσότερες πραγματικές περιπτώσεις πυρκαγιάς σε σχέση με την pretrained έκδοση, η τιμή της Recall εξακολουθεί να παραμένει υψηλή και υποδηλώνει πολύ καλή ικανότητα εντοπισμού των περισσότερων περιστατικών πυρκαγιάς. Ως συμπερασματικό σχόλιο σε σχέση με τις αλλαγές στις μετρικές θα μπορούσαμε να πούμε ότι οι μεταβολές των μετρικών δείχνουν ότι το fine tuning μετέβαλε ουσιαστικά τη στρατηγική λήψης αποφάσεων του SegFormer-B0. Ενώ η pretrained έκδοση ήταν ιδιαίτερα προσανατολισμένη στην πρόβλεψη της κλάσης Fire, επιτυγχάνοντας πολύ υψηλή Recall αλλά αρκετά χαμηλή Precision, μετά το fine tuning το μοντέλο έγινε σημαντικά πιο ισορροπημένο. Η μεγάλη αύξηση της Accuracy, της Precision και του F1-score υπερκαλύπτει τη σχετικά μικρή μείωση της Recall, οδηγώντας το μοντέλο σε σαφώς καλύτερη συνολική απόδοση στο σύνολο δεδομένων DFire.



Εικόνα 4-1 Σύγκριση t-SNE SegFormer στο DFire

Σε σχέση με τις οπτικοποιήσεις των t-SNE βλέπουμε πως στην pretrained έκδοση παρατηρείται εκτεταμένη επικάλυψη μεταξύ των δύο κλάσεων (Fire και No Fire). Μετά τη διαδικασία fine tuning, ο χώρος χαρακτηριστικών μεταβάλλεται ουσιαστικά. Οι δύο κλάσεις οργανώνονται πλέον σε δύο μεγάλες και σαφώς διαχωρισμένες περιοχές. Η επικάλυψη μεταξύ των δύο ομάδων περιορίζεται

σημαντικά, ενώ κάθε κλάση παρουσιάζει μεγαλύτερη εσωτερική συνοχή και μικρότερη διασπορά. Βέβαια εξακολουθούν να υπάρχουν λίγα δείγματα κοντά στα όρια μεταξύ των δύο συστάδων, όμως ο συνολικός διαχωρισμός είναι εμφανώς σαφέστερος σε σχέση με την pretrained έκδοση. Επιπλέον θα πρέπει να επισημάνουμε πως από την pretrained στη finetuned έκδοχή τα αντίστοιχα t-SNE διαγράμματα δεν δείχνουν απλώς μια καλύτερη ομαδοποίηση των συστάδων αλλά και μια ριζική αλλαγή στη «γεωμετρία» της αναπαράστασης. Το τελευταίο αυτό σημείο είναι ίσως δηλωτικό της μάθησης που απέκτησε το μοντέλο πάνω στα χαρακτηριστικά του συνόλου με τη διαδικασία finetuning.

4.1.2 Το μοντέλο Yolov26

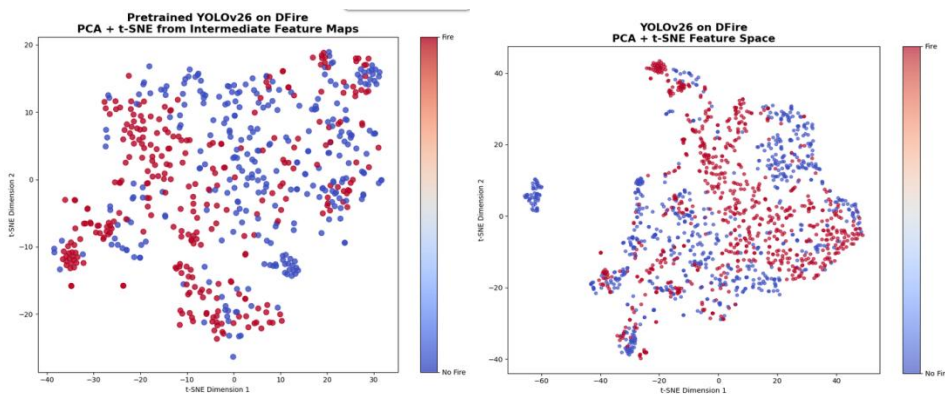
Η απόδοση του Yolov26 στο DFire στην προεκπαιδευμένη και τη finetuned έκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-2 Σύγκριση Μετρικών Yolov26 στο DFire

Μετρική	Pretrained Yolov26	Fine tuned Yolov26	Μεταβολή
Accuracy	0.6551	0.9032	+0.2481
Precision	0.6771	0.9811	+0.3040
Recall	0.6780	0.8349	+0.1569
F1-score	0.6775	0.9021	+0.2246

Με βάση τα αποτελέσματα του πίνακα το Yolov26 παρουσιάζει πολύ σημαντική βελτίωση μετά το fine-tuning. Το μοντέλο μετά από το finetuning πέτυχε ταυτόχρονη βελτίωση σε Accuracy, Precision, Recall και F1-score, κάτι που σημαίνει ότι βελτίωσε ουσιαστικά τόσο την ικανότητα του να εντοπίζει τα πραγματικά περιστατικά πυρκαγιάς όσο και την αξιοπιστία των προβλέψεών του. Η μεγαλύτερη αύξηση εμφανίζεται στην Precision και είναι της τάξης του 30,4%. Η αύξηση αυτή δείχνει ότι μετά την προσαρμογή του στο DFire το Yolov26 παράγει πολύ λιγότερες ψευδώς θετικές προβλέψεις, καθώς σχεδόν όλες οι εικόνες που ταξινομούνται ως Fire αντιστοιχούν σε πραγματικά περιστατικά πυρκαγιάς. Ένα ιδιαίτερο σημείο το οποίο πρέπει να επισημάνουμε είναι πως το Yolov26 είναι object detector, δηλαδή έχει σχεδιαστεί πρωτίστως για εντοπισμό αντικειμένων και όχι

αποκλειστικά για ταξινόμηση εικόνων. Παρά το γεγονός αυτό, βλέπουμε ότι το fine tuning οδήγησε σε πολύ σημαντική βελτίωση όλων των μετρικών, γεγονός που δείχνει ότι οι αναπαραστάσεις που έχει μάθει το μοντέλο μπορούν να προσαρμοστούν αποτελεσματικά και σε ένα πρόβλημα δυαδικής ταξινόμησης.



Εικόνα 4-2 Σύγκριση t-SNE YOLOv26 στο DFire

Παρατηρώντας τη γεωμετρία των διαγραμμάτων t-SNE βλέπουμε ότι μετά τη διαδικασία fine tuning, ο χώρος χαρακτηριστικών παρουσιάζει μια σαφή βελτίωση ως προς την οργάνωση των κλάσεων. Παρότι εξακολουθούν να υπάρχουν περιοχές στις οποίες οι δύο κλάσεις συνυπάρχουν, παρατηρείται μια τάση ομαδοποίησης των δειγμάτων της κλάσης Fire κυρίως στο κεντρικό και δεξιό τμήμα του διαγράμματος, ενώ αρκετά δείγματα της κλάσης No Fire συγκεντρώνονται σε διακριτές περιοχές του χώρου χαρακτηριστικών. Η εσωτερική συνοχή των δύο ομάδων είναι μεγαλύτερη σε σχέση με την pretrained έκδοση και η επικάλυψη μεταξύ τους περιορίζεται αισθητά. Ένα σημείο που αξίζει να τονίσουμε είναι ότι ο διαχωρισμός των δύο κλάσεων στο finetuned μοντέλο δεν είναι τόσο καθαρός όσο σε μοντέλα αμιγώς προσανατολισμένα στην ταξινόμηση (όπως τα SigLIP2, Convnext, SegFormer). Το YOLOv26 εκπαιδεύεται να συγκεντρώνει στοιχεία γύρω από το πού βρίσκεται το αντικείμενο, ποια τα χαρακτηριστικά του και ποια είναι η κλάση του. Άρα τα χαρακτηριστικά που παράγει το μοντέλο είναι πλουσιότερα, αλλά και λιγότερο "διχοτομημένα- διακριτά" σε επίπεδο «γεωμετρίας» στο διάγραμμα t-SNE.

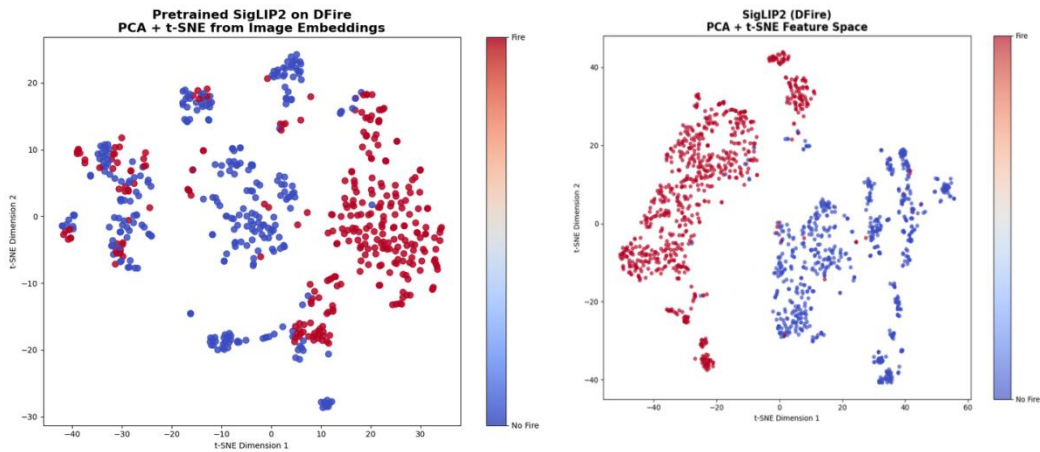
4.1.3 Το μοντέλο SigLIP2

Η απόδοση του SigLIP2 στο DFire στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-3 Σύγκριση Μετρικών SigLIP2 στο DFire

Μετρική	Pretrained SigLIP2	Fine tuned SigLIP2	Μεταβολή
Accuracy	0.8235	0.9789	+0.1554
Precision	0.8188	0.9859	+0.1671
Recall	0.8401	0.9744	+0.1343
F1-score	0.8293	0.9801	+0.1508

Η σύγκριση των αποτελεσμάτων μας δείχνει ότι η προσαρμογή του μοντέλου στα χαρακτηριστικά του συγκεκριμένου συνόλου δεδομένων οδήγησε σε βελτίωση όλων των μετρικών αξιολόγησης. Διαπιστώνουμε έτσι ότι το μοντέλο κατάφερε να προσαρμόσει αποτελεσματικά τις αναπαραστάσεις χαρακτηριστικών του στο πρόβλημα της ανίχνευσης δασικών πυρκαγιών. Πιο συγκεκριμένα η Accuracy αυξήθηκε από 82,35% σε 97,89%, παρουσιάζοντας βελτίωση περίπου 15,5 ποσοστιαίων μονάδων, γεγονός που υποδηλώνει σημαντική μείωση του συνολικού αριθμού λανθασμένων ταξινομήσεων. Παράλληλα, η Precision αυξήθηκε από 81,88% σε 98,59%, δείχνοντας ότι μετά το fine-tuning το μοντέλο παράγει αισθητά λιγότερες ψευδώς θετικές προβλέψεις. Αντίστοιχα, η Recall αυξήθηκε από 84,01% σε 97,44%, γεγονός που σημαίνει ότι το μοντέλο εντοπίζει πλέον σχεδόν όλες τις πραγματικές περιπτώσεις πυρκαγιάς. Τέλος, η αύξηση του F1-score από 82,93% σε 98,01% επιβεβαιώνει ότι η βελτίωση αφορά συνολικά την ικανότητα ταξινόμησης και όχι μόνο μία επιμέρους πτυχή της απόδοσης του μοντέλου. Συνολικά, θα σχολιάζαμε πως οι μεταβολές των τεσσάρων μετρικών υποδεικνύουν ότι η διαδικασία του fine-tuning βελτίωσε ουσιαστικά την ικανότητα του SigLIP2 να διακρίνει μεταξύ εικόνων με και χωρίς πυρκαγιά στο σύνολο δεδομένων DFire, οδηγώντας σε σημαντική αύξηση τόσο της αξιοπιστίας των θετικών προβλέψεων όσο και της ικανότητας εντοπισμού των πραγματικών περιστατικών πυρκαγιάς.



Εικόνα 4-3 Σύγκριση t-SNE SigLIP2 στο DFire

Από τη συγκριτική ανασκόπηση των δύο t-SNE βλέπουμε ότι το pretrained SigLIP2 παρουσιάζει αρκετές μικρές απομονωμένες συστάδες και αρκετά δείγματα των δύο κλάσεων διάσπαρτα ανάμεσά τους. Μετά το fine-tuning, παρατηρούμε ότι αυτές οι μικρές συστάδες αναδιοργανώνονται σε δύο μεγάλες, συνεκτικές περιοχές που αντιστοιχούν στις δύο κλάσεις. Αυτό μας δείχνει ότι το fine-tuning δεν βελτίωσε μόνο τον τοπικό διαχωρισμό των δειγμάτων, αλλά αναδιαμόρφωσε συνολικά τη γεωμετρία του χώρου χαρακτηριστικών. Δηλαδή πρακτικά αυτό σημαίνει πως το μοντέλο έμαθε μια νέα αναπαράσταση των εικόνων που είναι πολύ καλύτερα προσαρμοσμένη στις ιδιαιτερότητες του DFire. Συμπερασματικά διαπιστώνουμε ότι η σύγκριση μεταξύ της pretrained και της fine-tuned έκδοσης του SigLIP2 δείχνει ότι η διαδικασία fine-tuning υπήρξε ιδιαίτερα αποτελεσματική για το σύνολο δεδομένων DFire. Η σημαντική αύξηση όλων των μετρικών αξιολόγησης συνοδεύεται από σαφέστερο διαχωρισμό των δύο κλάσεων στον χώρο χαρακτηριστικών. Το μοντέλο προσαρμόστηκε επιτυχώς στις ιδιαιτερότητες του συγκεκριμένου συνόλου δεδομένων

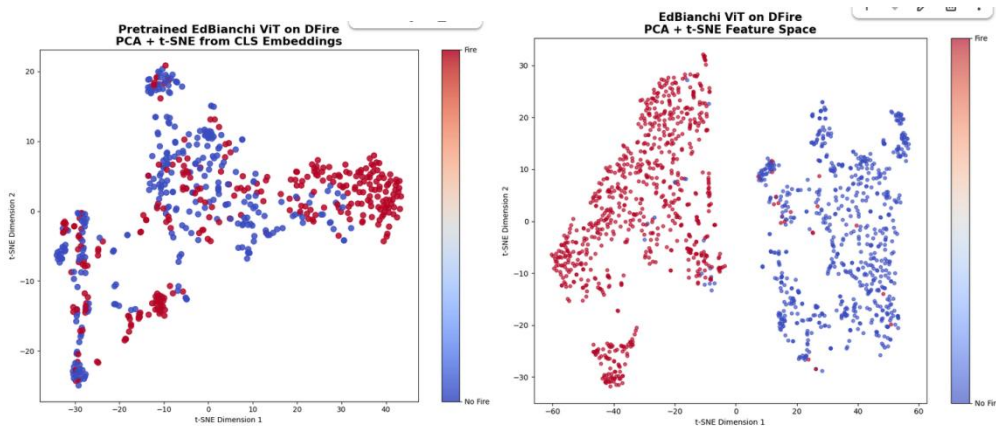
4.1.4 Το μοντέλο EdBianchiVit

Η απόδοση του EdBianchiVit στο DFire στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-4 Σύγκριση Μετρικών EdBianchiVit στο DFire

Μετρική	Pretrained EdBianchi	Fine tuned EdBianchi	Μεταβολή
Accuracy	0.6732	0.9763	+0.3031
Precision	0.8894	0.9803	+0.0909
Recall	0.4437	0.9752	+0.5315
F1-score	0.5921	0.9778	+0.3857

Παρατηρώντας τα συγκεντρωτικά στοιχεία του Πίνακα βλέπουμε ότι στο pretrained μοντέλο η Precision είναι ήδη υψηλή (88,94%), ενώ η Recall είναι χαμηλή (44,37%). Αυτό σημαίνει ότι το μοντέλο ήταν ιδιαίτερα προσεκτικό στις προβλέψεις του για την κλάση Fire. Στην πράξη προέβλεπε μια εικόνα ως Fire μόνο όταν ήταν σχεδόν απόλυτα βέβαιο, με αποτέλεσμα να κάνει λίγα False Positives, αλλά από την άλλη πλευρά να χάνει μεγάλο αριθμό πραγματικών περιστατικών πυρκαγιάς (False Negatives). Μετά το fine tuning διαπιστώνουμε ότι η συμπεριφορά αυτή αλλάζει εντυπωσιακά. Η Recall αυξάνεται κατά περισσότερες από 53 ποσοστιαίες μονάδες, ενώ η Precision όχι μόνο δεν μειώνεται, αλλά αυξάνεται ακόμη περισσότερο, φτάνοντας το 98,03%. Το αποτέλεσμα αυτό μας δείχνει ότι το μοντέλο έμαθε να αναγνωρίζει αποτελεσματικότερα τα πραγματικά χαρακτηριστικά της πυρκαγιάς, διατηρώντας ταυτόχρονα εξαιρετικά υψηλή αξιοπιστία στις θετικές προβλέψεις.



Εικόνα 4-4 Σύγκριση t-SNE EdBianchiViT στο DFire

Η αλματώδης αύξηση των τιμών των μετρικών που επέφερε στο μοντέλο το finetuning απηχείται και στα διαγράμματα t-SNE. Στο pretrained μοντέλο παρόλο που διαμορφώνονται ορισμένες περιοχές στις οποίες κυριαρχεί μία από τις δύο κλάσεις, υπάρχει σημαντική επικάλυψη μεταξύ των δειγμάτων Fire και No Fire στο κεντρικό τμήμα του χώρου χαρακτηριστικών. Μετά τη διαδικασία fine tuning, η εικόνα μεταβάλλεται δραστικά. Οι δύο κλάσεις οργανώνονται πλέον σε δύο σαφώς ευδιάκριτες περιοχές του χώρου χαρακτηριστικών. Σχολιάζοντας και τη γεωμετρία των t-SNE διαγραμμάτων η μετάβαση από έναν έντονα επικαλυπτόμενο χώρο χαρακτηριστικών (pretrained μοντέλο) σε έναν σχεδόν συμμετρικά διαχωρισμένο χώρο (finetuned μοντέλο) αποτελεί ίσως μία από τις πιο ευανάγνωστες ενδείξεις ότι το fine tuning βελτίωσε ουσιαστικά τις εσωτερικές αναπαραστάσεις του μοντέλου.

4.1.5 Το μοντέλο Superflash41

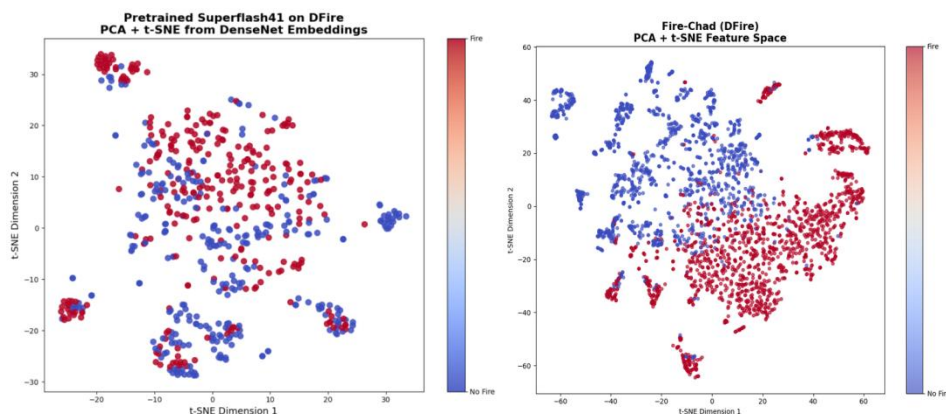
Η απόδοση του Superflash41 στο DFire στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-5 Σύγκριση Μετρικών Superflash41 στο DFire

Μετρική	Pretrained Superflash41	Fine tuned Superflash41	Μεταβολή
Accuracy	0.5186	0.8864	+0.3678
Precision	0.5288	0.9364	+0.4076

Recall	0.9092	0.8449	-0.0643
F1-score	0.6687	0.8883	+0.2196

Με βάση τον Πίνακα παρατηρούμε ότι το pretrained μοντέλο λειτουργούσε με πολύ προκατειλημμένο τρόπο στις προβλέψεις του υπό την έννοια ότι έδειχνε προτίμηση στο να χαρακτηρίζει πολλές εικόνες ως Fire, ώστε να μην χάσει πραγματικές πυρκαγιές. Η στρατηγική αυτή εξηγεί την πολύ υψηλή Recall (90,92%) του pretrained μοντέλου. Ταυτόχρονα όμως η ίδια αυτή στρατηγική οδηγούσε το μοντέλο σε μεγάλο αριθμό ψευδώς θετικών προβλέψεων, γεγονός που αποτυπώνεται στη χαμηλή Precision (52,88%) και στη χαμηλή Accuracy (51,86%). Μετά το fine tuning, το μοντέλο έγινε αισθητά πιο ισορροπημένο. Έμαθε με ένα τρόπο το μοντέλο να διαχωρίζει αποτελεσματικότερα τις δύο κλάσεις. Ως αποτέλεσμα, η Precision αυξήθηκε κατά περισσότερες από 40 ποσοστιαίες μονάδες, ενώ η Recall μειώθηκε μόνο κατά περίπου 6 ποσοστιαίες μονάδες. Ως ένα γενικό σχόλιο θα μπορούσαμε να πούμε ότι εδώ βρίσκουμε ένα ένα χαρακτηριστικό παράδειγμα του γνωστού Precision–Recall trade-off. Στην pretrained έκδοση το μοντέλο ήταν υπερβολικά ευαίσθητο, επιτυγχάνοντας υψηλή Recall εις βάρος της Precision. Το fine tuning δεν επιδίωξε να μεγιστοποιήσει μία μόνο μετρική, αλλά να βρει μια καλύτερη ισορροπία μεταξύ των δύο. Αυτό αποδεικνύεται και από τη θεαματική αύξηση του F1-score που αποτελεί την μετρική που ενσωματώνει και το Precision και το Recall και θα μπορούσε να χαρακτηριστεί ως μια μετρική δείκτης της ισορροπίας του μοντέλου.



Εικόνα 4-5 Σύγκριση t-SNE Superflash41 στο DFire

Από τη σύγκριση των διαγραμμάτων t-SNE σημειώνουμε ότι μετά από τη διαδικασία fine tuning, ο χώρος χαρακτηριστικών εμφανίζει σημαντική αναδιοργάνωση. Τα δείγματα της κλάσης No Fire συγκεντρώνονται κυρίως στο αριστερό τμήμα του διαγράμματος, ενώ τα δείγματα της κλάσης Fire οργανώνονται σε μία εκτεταμένη περιοχή στο δεξί και κάτω τμήμα του χώρου χαρακτηριστικών. Παρ'όλο που εξακολουθεί να υπάρχει περιορισμένη επικάλυψη στην κεντρική περιοχή, η συνολική κατανομή των δύο κλάσεων είναι εμφανώς καλύτερα οργανωμένη και παρουσιάζει μεγαλύτερη εσωτερική συνοχή. Η σχετική επικάλυψη των δύο κλάσεων στη κεντρική περιοχή του t-SNE διαγράμματος του finetuned μοντέλου συνεκτιμώμενη με το γεγονός ότι έχουμε μια σημαντική αύξηση της ταξινομητικής απόδοσης του μοντέλου σε επίπεδο μετρικών απόδοσης μας οδηγεί στο να συναγάγουμε ότι το μοντέλο δεν χρειάστηκε να επιτύχει απόλυτο διαχωρισμό των χαρακτηριστικών για να αυξήσει σημαντικά την απόδοσή του.

4.1.6 Το μοντέλο Convnext

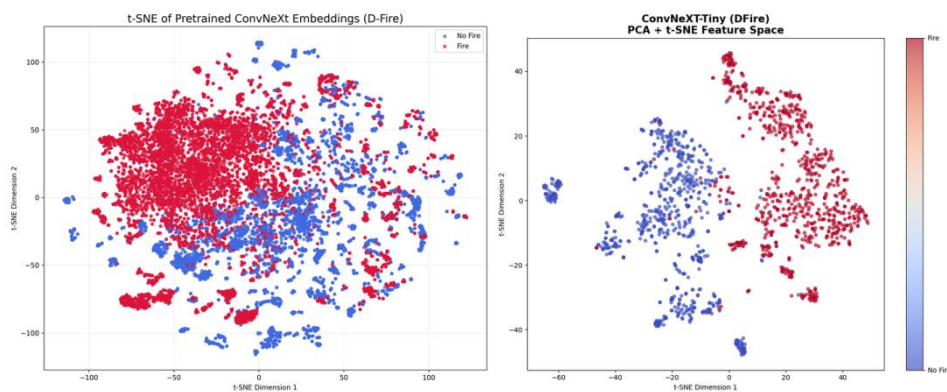
Η απόδοση του Convnext στο DFire στην προεκπαιδευμένη και τη finetuned έκδοσή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-6 Σύγκριση Μετρικών Convnext στο DFire

Μετρική	Pretrained Convnext	Fine tuned Convnext	Μεταβολή
Accuracy	0.5941	0.9805	+0.3864
Precision	0.7189	0.9784	+0.2595
Recall	0.4145	0.9852	+0.5707
F1-score	0.5258	0.9818	+0.4560

Τα αποτελέσματα του Πίνακα καταδεικνύουν ότι η προσαρμογή του μοντέλου στο συγκεκριμένο σύνολο δεδομένων επέφερε θεαματική βελτίωση της συνολικής του απόδοσης, οδηγώντας σε εξαιρετικά υψηλές τιμές σε όλες τις μετρικές. Η μεγαλύτερη μεταβολή παρατηρείται στη Recall, η οποία αυξήθηκε από 41,45% σε 98,52%, παρουσιάζοντας βελτίωση μεγαλύτερη των 57 ποσοστιαίων μονάδων. Η pretrained έκδοση του μοντέλου αναγνώριζε λιγότερες από τις μισές πραγματικές περιπτώσεις πυρκαγιάς, ενώ μετά το fine-tuning, το μοντέλο αναγνωρίζει πλέον σχεδόν όλες τις εικόνες πυρκαγιάς του

συνόλου δοκιμής. Ένα ακόμα ιδιαίτερα σημαντικό στοιχείο που πρέπει να υπογραμμίσουμε είναι ότι στην περίπτωση αυτή δεν παρατηρείται το συνηθισμένο trade-off μεταξύ Precision και Recall αλλά αντιθέτως αυξάνονται και οι δύο μετρικές ταυτόχρονα σε πολύ υψηλά επίπεδα, απολήγοντας εντέλει σε F1-score 98,18%.



Εικόνα 4-6 Σύγκριση t-SNE Convnext στο DFire

Μελετώντας συγκριτικά τα t-SNE διαγράμματα βλέπουμε ότι στην pretrained περίπτωση παρατηρείται εκτεταμένη επικάλυψη μεταξύ των δύο κλάσεων, ενώ μετά το finetuningη αναδιοργάνωση του χώρου χαρακτηριστικών είναι εντυπωσιακή. Όπως βλέπουμε από το t-SNE της finetuned περίπτωσης οι δύο κλάσεις εμφανίζονται σχεδόν πλήρως διαχωρισμένες. Συμπερασματικά, τα διαγράμματα t-SNE επιβεβαιώνουν ότι το fine tuning δεν βελτίωσε μόνο την τελική ταξινόμηση του Convnext, αλλά αναδιαμόρφωσε ουσιαστικά τον εσωτερικό χώρο αναπαράστασης των χαρακτηριστικών. Το μοντέλο μέσω του finetuningκατόρθωσε να μάθει πως να αναπαριστά αποτελεσματικότερα τις οπτικές ιδιότητες που χαρακτηρίζουν τις εικόνες πυρκαγιάς και τις εικόνες χωρίς πυρκαγιά.

4.2 στο FLAME

Ακολουθώντας στην παράγραφο αυτή θα προχωρήσουμε σε μια συγκριτική ανασκόπηση ως προς την απόδοσή κάθε ενός από τα έξι μοντέλα, στη προεκπαιδευμένη και τη finetuned εκδοχή τους, στο σύνολο FLAME.

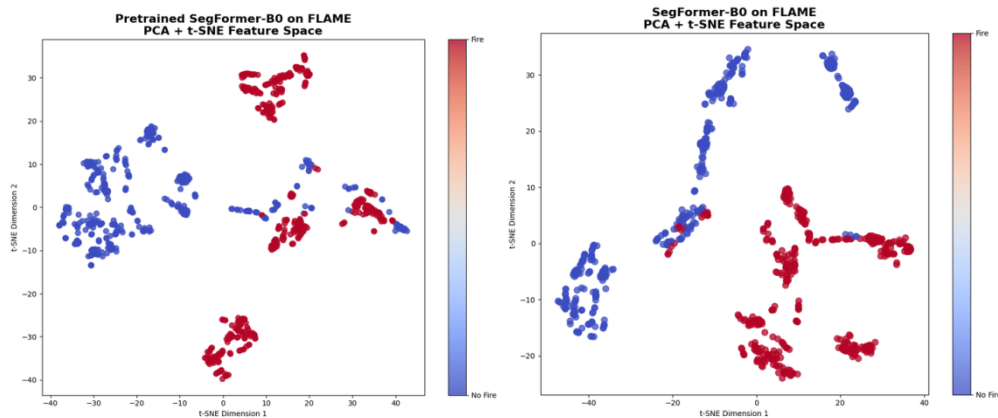
4.2.1 Το μοντέλο SegFormer

Η απόδοση του SegFormer-B0 στο FLAME στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-7 Σύγκριση Μετρικών SegFormer στο FLAME

Μετρική	Pretrained SegFormer-B0	Fine tuned SegFormer-B0	Μεταβολή
Accuracy	0.9237	0.4750	-0.4487
Precision	0.8709	0.4870	-0.3839
Recall	0.9950	0.9375	-0.0575
F1-score	0.9288	0.6410	-0.2878

Είναι αξιοσημείωτο ότι σε αντίθεση με τα περισσότερα μοντέλα της παρούσας μελέτης, όπου το fine-tuning οδήγησε σε σημαντική βελτίωση της απόδοσης, στην περίπτωση του SegFormer-B0 παρατηρείται μείωση όλων των μετρικών αξιολόγησης. Ως πιθανές αιτίες θα μπορούσαμε να σημειώσουμε τις εξής. α) Το SegFormer-B0 είχε ήδη προεκπαιδευτεί πάνω σε δεδομένα που σχετίζονται με πυρκαγιές (συγκεκριμένα στο ίδιο το σύνολο FLAME). Επομένως, οι αναπαραστάσεις του ήταν ήδη εξαιρετικά κατάλληλες για αυτό το πρόβλημα. Επιπρόσθετα οι τιμές των μετρικών του pretrained μοντέλου ήταν ήδη πολύ υψηλές. Αυτό μπορεί να είναι ένδειξη ότι υπάρχει ελάχιστο περιθώριο περαιτέρω βελτίωσης. Σε τέτοιες περιπτώσεις, ακόμη και μικρές μεταβολές των βαρών μπορεί να οδηγήσουν σε χειρότερα αποτελέσματα αντί για καλύτερα κάτι που καταγράψαμε μετά το fine-tuning. β) Το μοντέλο κατά τη διάρκεια του finetuning μπορεί να απώλεσε μέρος της πολύ καλής γνώσης που είχε αποκτήσει κατά την αρχική προεκπαίδευσή του πάνω στο FLAME και να αντικατέστησε αυτές τις γενικευμένες αναπαραστάσεις με περισσότερο εξειδικευμένες, αλλά λιγότερο αποτελεσματικές για το test set (το φαινόμενο catastrophic forgetting). γ) Με μόνο 3 epochs στο finetuning και σχετικά χαμηλό learning rate, το μοντέλο ενδέχεται να άρχισε να προσαρμόζεται υπερβολικά στα συγκεκριμένα δείγματα του training set, χάνοντας μέρος της γενικευτικής του ικανότητας (φαινόμενο overfitting).



Εικόνα 4-7 Σύγκριση t-SNE SegFormer στο FLAME

Η ανασκόπηση των διαγραμμάτων-t-SNE μας επιφυλάσσει μια έκπληξη σε αυτό το σημείο. Ενώ όπως είδαμε οι μετρικές καταρρέουν μετά το fine-tuning, ο οπτικός διαχωρισμός των δύο κλάσεων παραμένει πολύ καλός, ίσως και ελαφρώς καλύτερος σε ορισμένα σημεία. Αν καλούμασταν να δώσουμε μια πιθανή εξήγηση αυτού του εν πρώτοις αντιφατικού στοιχείου θα λέγαμε ότι αυτό μπορεί να συμβαίνει γιατί το t-SNE απεικονίζει μόνο τη γεωμετρία των χαρακτηριστικών και όχι τη διαδικασία λήψης της τελικής απόφασης. Δηλαδή το finetuned μοντέλο μπορεί να εξακολουθεί να παράγει καλά χαρακτηριστικά όπως και το pretrained (και οι δύο χώροι χαρακτηριστικών είναι καλά οργανωμένοι) αλλά ο τελικός classifier του finetuned μοντέλου να τα αξιοποιεί λιγότερο αποτελεσματικά πιθανώς λόγω αλλαγών στα όρια απόφασης.

4.2.2 Το μοντέλο Yolon26

Η απόδοση του Yolon26 στο FLAME στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

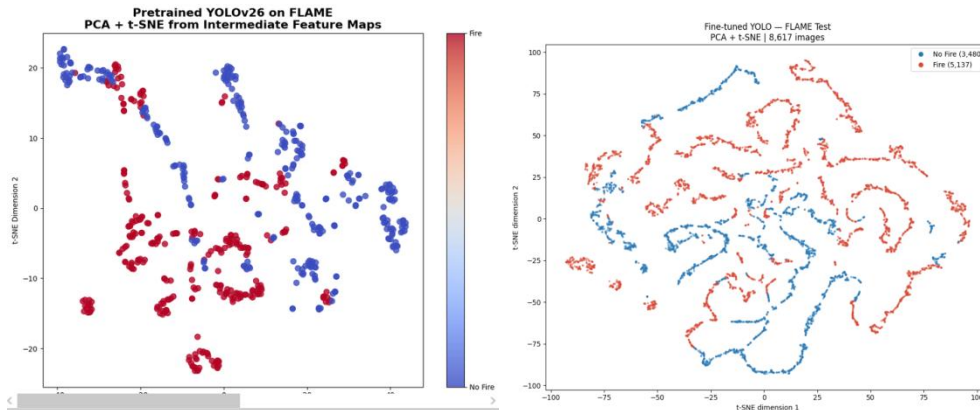
Πίνακας 4-8 Σύγκριση Μετρικών Yolon26 στο FLAME

Μετρική	Pretrained Yolon26	Fine tuned Yolon26	Μεταβολή
Accuracy	0.5778	0.4376	-0.1402

Precision	0.9980	0.7470	-0.2510
Recall	0.2924	0.0857	-0.2067
F1-score	0.4523	0.1537	-0.2986

Επισκοπώντας τα δεδομένα του Πίνακα 4-8 παρατηρούμε ότι το fine tuning δεν βελτίωσε την απόδοση του Yolon26 στο σύνολο δεδομένων FLAME. Στην κοινή αξιολόγηση των 8.617 εικόνων του Test set, η Accuracy μειώθηκε από 57,78% σε 43,76%, ενώ η Precision υποχώρησε από 99,80% σε 74,70%. Η μεγαλύτερη δυσκολία εντοπίζεται στην αναγνώριση των πραγματικών περιστατικών πυρκαγιάς: η Recall μειώθηκε από 29,24% σε μόλις 8,57%, ενώ το F1-score από 0,4523 σε 0,1537. Συγκεκριμένα, το fine-tuned μοντέλο αναγνώρισε σωστά μόνο 440 από τις 5.137 εικόνες Fire, ενώ οι υπόλοιπες 4.697 ταξινομήθηκαν ως No Fire. Παράλληλα, χαρακτήρισε λανθασμένα ως Fire 149 από τις 3.480 εικόνες χωρίς πυρκαγιά. Επομένως, η Precision 74,70% σημαίνει ότι περίπου τρεις στις τέσσερις θετικές προβλέψεις ήταν σωστές, αλλά η πολύ χαμηλή Recall δείχνει ότι η μεγάλη πλειονότητα των πραγματικών περιστατικών πυρκαγιάς δεν εντοπίστηκε. Κατά την ερμηνεία αυτής της υποχώρησης πρέπει να ληφθεί υπόψη το πρόβλημα προετοιμασίας των δεδομένων της αρχικής εκπαίδευσης. Οι εικόνες Fire και οι μάσκες αντιστοιχίστηκαν με βάση τον αριθμό του ονόματός τους, παρότι προέρχονταν από διαφορετικές συλλογές εικόνων του FLAME. Η αριθμητική σύμπτωση δεν εξασφάλιζε ότι η μάσκα περιέγραφε την αντίστοιχη εικόνα. Κατά συνέπεια, τα πλαίσια εκπαίδευσης που προέκυψαν από τις μάσκες δεν είχαν εξασφαλισμένη χωρική αντιστοίχιση με τις πραγματικές περιοχές φωτιάς. Το πρόβλημα αυτό αποτελεί πιθανό παράγοντα υποβάθμισης της απόδοσης, χωρίς τα διαθέσιμα αποτελέσματα να επιτρέπουν την απομόνωση της επίδρασής του. Επιπλέον, στην αξιολόγηση σε επίπεδο εικόνας, η θετική απόφαση προϋποθέτει την παραγωγή τουλάχιστον ενός πλαισίου της κλάσης Fire με confidence τουλάχιστον 0,25. Εάν δεν παραχθεί τέτοια ανίχνευση, η εικόνα χαρακτηρίζεται ως No Fire, ακόμη και όταν περιέχει πραγματική πυρκαγιά. Η εξάρτηση αυτή εξηγεί πώς οι αποτυχίες εντοπισμού μεταφέρονται στις ψευδώς αρνητικές

προβλέψεις της δυαδικής αξιολόγησης, χωρίς να προσδιορίζει από μόνη της την αιτία τους.



Εικόνα 4-8 Σύγκριση t-SNE Yolov26 στο FLAME

Ως προς τα διαγράμματα t-SNE, παρατηρούνται πολλές επιμέρους συστάδες και τοπικές συγκεντρώσεις εικόνων της ίδιας κλάσης. Στη fine-tuned εκδοχή εμφανίζονται επιμήκεις σχηματισμοί, με περιοχές τοπικού διαχωρισμού αλλά και γειτνίασης των δύο κατηγοριών. Δεν προκύπτουν δύο ενιαίες, σαφώς απομονωμένες περιοχές που να αντιστοιχούν αποκλειστικά στις Fire και No Fire. Επιπλέον, η διαφορετική πυκνότητα των δύο απεικονίσεων περιορίζει την άμεση σύγκρισή τους και υπογραμμίζει τη χρεία περαιτέρω επιβεβαίωσης ότι χρησιμοποιήθηκαν οι ίδιες εικόνες και οι αντίστοιχες ρυθμίσεις στα πειράματα της pretrained και finetuned εκδοχής. Τα διαγράμματα παρέχουν ποιοτική εικόνα της οργάνωσης των χαρακτηριστικών, αλλά δεν αποδεικνύουν βελτίωση της ταξινόμησης ή της γενίκευσης.

4.2.3 Το μοντέλο SigLIP2

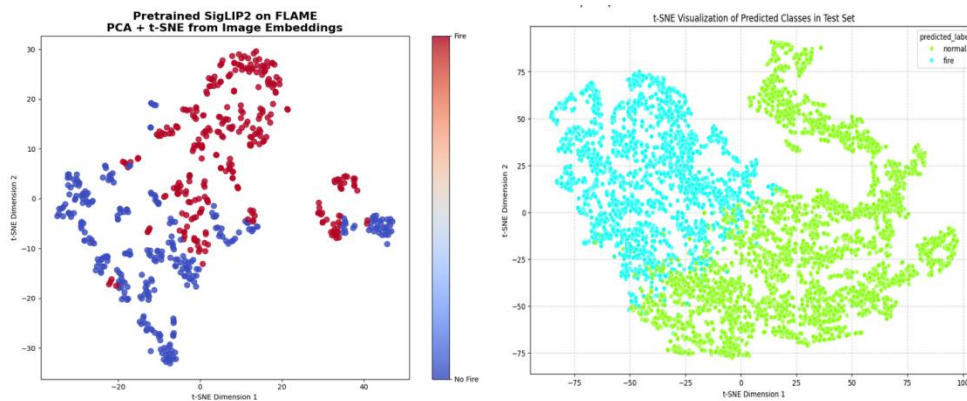
Η απόδοση του SigLIP2 στο FLAME στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-9 Σύγκριση Μετρικών SigLIP2 στο FLAME

Μετρική	Pretrained SigLIP2	Fine tuned SigLIP2	Μεταβολή
---------	--------------------	--------------------	----------

Accuracy	0.7243	0.5536	-0.1707
Precision	0.7436	0.6916	-0.0520
Recall	0.8203	0.4532	-0.3671
F1-score	0.7801	0.5476	-0.2325

Από τη σύγκριση των τιμών των μετρικών στον Πίνακα προκύπτει ότι η σημαντικότερη μεταβολή παρατηρείται στη Recall, η οποία μειώθηκε από 82,03% σε 45,32%. Η πτώση αυτή υποδηλώνει ότι μετά το fine tuning το μοντέλο αδυνατεί να εντοπίσει μεγάλο αριθμό πραγματικών περιστατικών πυρκαγιάς, οδηγώντας σε σημαντική αύξηση των ψευδώς αρνητικών προβλέψεων. Η συγκεκριμένη μεταβολή αποτελεί τον βασικό παράγοντα που επηρεάζει αρνητικά τη συνολική απόδοση του μοντέλου. Η συμπεριφορά του SigLIP2 μετά από finetuning στο FLAME παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς διαπιστώνουμε ότι διαφοροποιείται αισθητά από εκείνη που παρατηρήθηκε στα σύνολα δεδομένων FASDD και DFire. Προσπαθώντας να κάνουμε μια εκτίμηση του τι μπορεί να συμβαίνει σημειώνουμε ότι η σχετικά μικρή μεταβολή της Precision, σε συνδυασμό με τη μεγάλη μείωση της Recall, ίσως να υποδηλώνει ότι το πρόβλημα δεν οφείλεται στην αδυναμία του μοντέλου να αναγνωρίζει τα χαρακτηριστικά της πυρκαγιάς, αλλά κυρίως στη μεταβολή του τρόπου με τον οποίο το μοντέλο λαμβάνει την τελική του απόφαση. Με άλλα λόγια, το μοντέλο φαίνεται να έγινε πιο επιφυλακτικό πριν αποδώσει την ετικέτα Fire, χάνοντας έτσι σημαντικό αριθμό πραγματικών περιστατικών. Μία πιθανή εξήγηση είναι ότι οι προεκπαιδευμένες αναπαραστάσεις του SigLIP2 ήταν ήδη αρκετά κατάλληλες για το συγκεκριμένο πρόβλημα και ότι το επιπλέον fine tuning οδήγησε σε overfitting στα δεδομένα εκπαίδευσης ή σε μερική αλλοίωση των αρχικών χαρακτηριστικών που είχε μάθει το μοντέλο κατά την προεκπαίδευσή του.



Εικόνα 4-9 Σύγκριση t-SNE SigLIP2 στο FLAME

Αυτό που παρατηρούμε από τα t-SNE διαγράμματα είναι πως μετά τη διαδικασία fine tuning, η γεωμετρία του χώρου χαρακτηριστικών μεταβάλλεται αισθητά. Οι δύο κλάσεις εξακολουθούν να εμφανίζουν έναν γενικό διαχωρισμό, ωστόσο η εσωτερική τους δομή γίνεται σημαντικά πιο διάχυτη και λιγότερο συνεκτική. Παρατηρείται μεγαλύτερη διασπορά των δειγμάτων, ενώ στις περιοχές όπου συναντώνται οι δύο κλάσεις εμφανίζεται αυξημένη ανάμειξη χαρακτηριστικών. Η εικόνα αυτή σηματοδοτεί ότι το μοντέλο έχασε μέρος της σταθερότητας των αναπαραστάσεων που διέθετε πριν από το fine tuning. Συμπερασματικά θα μπορούσαμε να πούμε πως στο SigLIP2 μετά από το finetuning παρατηρείται αισθητή αποδιοργάνωση των ίδιων των αναπαραστάσεων χαρακτηριστικών, γεγονός που συγχρόνως συνοδεύεται και από την σημαντική πτώση των μετρικών Recall και F1-score.

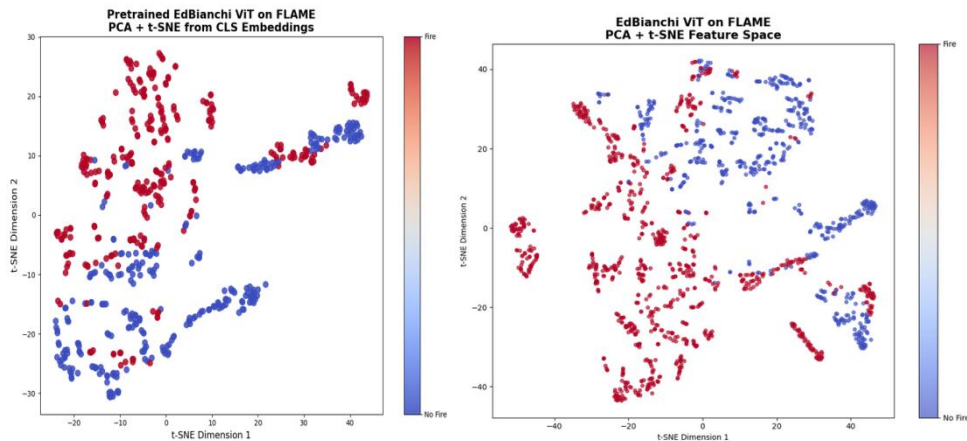
4.2.4 Το μοντέλο EdBianchiVit

Η απόδοση του EdBianchiVit στο FLAME στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-10 Σύγκριση Μετρικών EdBianchi στο FLAME

Μετρική	Pretrained EdBianchi	Fine tuned EdBianchi	Μεταβολή
Accuracy	0.4041	0.6449	+0.2408
Precision	0.7500	0.6465	-0.1035
Recall	0.0006	0.8922	+0.8916
F1-score	0.0012	0.7497	+0.7485

Ένα σημείο που τραβάει αμέσως την προσοχή από την ανάγνωση του παραπάνω Πίνακα είναι η εντυπωσιακή μεταβολή στο Recall, η οποία μετρική αυξήθηκε από 0,06% σε 89,22%. Η pretrained έκδοση του μοντέλου αναγνώριζε ουσιαστικά μηδενικό αριθμό πραγματικών περιστατικών πυρκαγιάς, γεγονός που καθιστούσε το μοντέλο πρακτικά μη λειτουργικό για την ανίχνευση πυρκαγιών. Μετά το fine tuning, το μοντέλο εντοπίζει πλέον σχεδόν εννέα στις δέκα πραγματικές εικόνες πυρκαγιάς, παρουσιάζοντας θεαματική βελτίωση της ευαισθησίας του. Συνοπτικά τα αποτελέσματα καταδεικνύουν ότι η διαδικασία fine tuning ήταν καθοριστική για την αύξηση της απόδοσης του EdBianchi ViT στο σύνολο δεδομένων FLAME. Επιπρόσθετα σημειώνουμε ότι το υψηλό F1-score της fine-tuned έκδοσης αποδεικνύει ότι η μικρή μείωση της Precision αντισταθμίζεται πλήρως από τη θεαματική αύξηση της Recall. Επομένως, η συνολική απόδοση του μοντέλου βελτιώνεται σημαντικά και το καθιστά πλέον ουσιαστικά αξιοποιήσιμο για εφαρμογές ανίχνευσης πυρκαγιάς.



Εικόνα 4-10 Σύγκριση t-SNE EdBianchi στο FLAME

Εξετάζοντας τα t-SNE διαγράμματα βλέπουμε πως εδώ είμαστε σε μια περίπτωση όπου η βελτίωση των μετρικών συνοδεύεται και από θεαματική βελτίωση της δομής του χώρου χαρακτηριστικών. Στην pretrained έκδοση παρατηρείται έντονη επικάλυψη μεταξύ των δειγμάτων των κλάσεων Fire και No Fire. Αν και μπορούμε να διακρίνουμε επιμέρους τοπικές συστάδες, μεγάλο μέρος των δύο κατηγοριών συνυπάρχει στις ίδιες περιοχές του χώρου χαρακτηριστικών. Μετά τη διαδικασία fine-tuning, η δομή του χώρου χαρακτηριστικών μεταβάλλεται σημαντικά. Οι εικόνες Fire συγκεντρώνονται κυρίως στο

αριστερό και στο κάτω τμήμα του διαγράμματος, ενώ οι εικόνες No Fire οργανώνονται κυρίως στο άνω δεξιό τμήμα. Η επικάλυψη μεταξύ των δύο κλάσεων περιορίζεται αισθητά, αν και ασφαλώς ξακολουθούν να υπάρχουν ορισμένα σημεία στις περιοχές επαφής των δύο κλάσεων, γεγονός που εξηγεί γιατί οι μετρικές Accuracy και η Precision δεν φθάνουν σε υψηλά επίπεδα τιμών.

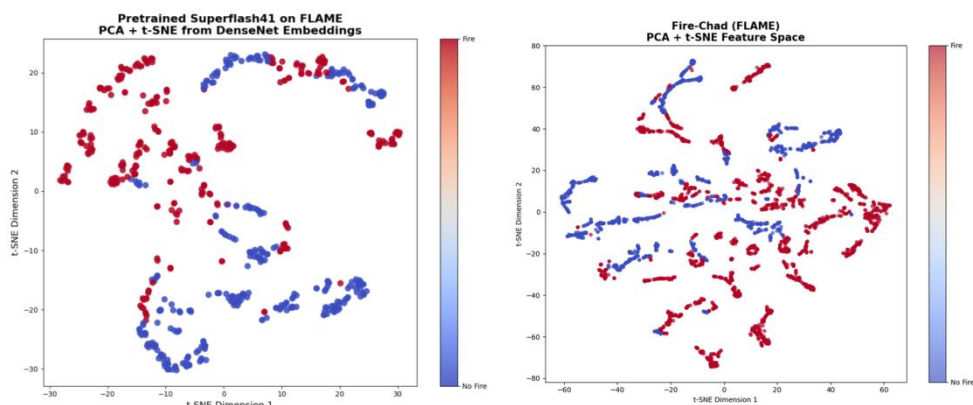
4.2.5 Το μοντέλο Superflash41

Η απόδοση του Superflash41 στο FLAME στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-11 Σύγκριση Μετρικών Superflash41 στο FLAME

Μετρική	Pretrained Superflash41	Fine tuned Superflash41	Μεταβολή
Accuracy	0.2053	0.6755	+0.4702
Precision	0.1944	0.6690	+0.4746
Recall	0.1059	0.9021	+0.7962
F1-score	0.1371	0.7682	+0.6311

Με βάση τον Πίνακα παρατηρούμε ότι με το finetuning βελτιώθηκαν σημαντικότερα όλες οι μετρικές. Η μεγαλύτερη μεταβολή καταγράφεται στη Recall, η οποία αυξήθηκε από 10,59% σε 90,21%. Συμπερασματικά μετά το fine tuning, η συμπεριφορά του μοντέλου μεταβάλλεται ουσιαστικά. Η πολύ μεγάλη αύξηση της Recall δείχνει ότι το μοντέλο απέκτησε την ικανότητα να αναγνωρίζει αποτελεσματικά τις περισσότερες εικόνες πυρκαγιάς, ενώ η παράλληλη αύξηση της Precision υποδηλώνει ότι η βελτίωση αυτή δεν συνοδεύτηκε από αύξηση των ψευδώς θετικών προβλέψεων. Η σημαντική αύξηση του F1-score επιβεβαιώνει ότι το fine-tuning οδήγησε σε πολύ καλύτερη ισορροπία μεταξύ των δύο αυτών χαρακτηριστικών. Γενικά θα μπορούσαμε να ισχυριστούμε ότι η διαδικασία fine tuning μετέτρεψε το Superflash41 από ένα μοντέλο με ιδιαίτερα περιορισμένη απόδοση σε έναν σαφώς αποτελεσματικότερο ταξινομητή εικόνων πυρκαγιάς.



Εικόνα 4-11 Σύγκριση t-SNE Superflash41 στο FLAME

Αυτό που παρατηρούμε από τα t-SNE είναι ότι μετά το finetuning η τοπική επικάλυψη μειώνεται σε ορισμένες περιοχές, αλλά η συνολική γεωμετρία του χώρου χαρακτηριστικών παραμένει παρόμοια. Θα μπορούσαμε να ισχυριστούμε πως η συνολική δομή του χώρου χαρακτηριστικών παραμένει σχετικά παρόμοια πριν και μετά το fine-tuning. Οι σημαντικές βελτιώσεις των μετρικών φαίνεται να προκύπτουν κυρίως από την καλύτερη αξιοποίηση των ήδη υπάρχοντων χαρακτηριστικών καθώς και από μια αποτελεσματικότερη διάκριση των ορίων μεταξύ των δύο κατηγοριών.

4.2.6 Το μοντέλο Convnext

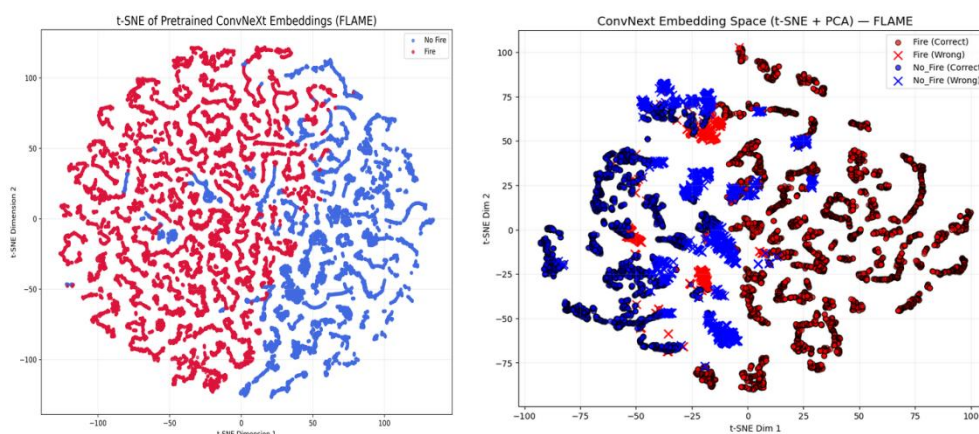
Η απόδοση του Convnext στο FLAME στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-12 Σύγκριση Μετρικών Convnext στο FLAME

Μετρική	Pretrained Convnext	Fine tuned Convnext	Μεταβολή
Accuracy	0.4183	0.8434	+0.4251
Precision	0.6898	0.8243	+0.1345
Recall	0.1349	0.9371	+0.8022
F1-score	0.2257	0.8771	+0.6514

Όπως βλέπουμε από τον Πίνακα των μετρικών το fine tuning βελτίωσε θεαματικά την απόδοση του Convnext. Μετά το fine-tuning, αυξήθηκαν όλες οι μετρικές. Η θεαματική αύξηση της Recall δείχνει ότι το μοντέλο απέκτησε πολύ μεγαλύτερη ικανότητα

αναγνώρισης των εικόνων πυρκαγιάς, ενώ η παράλληλη αύξηση της Precision υποδηλώνει ότι η βελτίωση αυτή δεν απέφερε ως παράπλευρη συνέπεια την αύξηση των ψευδώς θετικών ταξινομήσεων. Το μοντέλο καθίσταται εύρωστο και σταθερό στην ταξινόμησή του. Το υψηλό F1-score, λειτουργεί ως ένας ακόμα δείκτης ότι το fine-tuning οδήγησε σε εξαιρετική ισορροπία μεταξύ των δύο βασικών χαρακτηριστικών αξιολόγησης.



Εικόνα 4-12 Σύγκριση t-SNE Convnext στο FLAME

Περνώντας τώρα στη συγκριτική μελέτη των t-SNE διαγραμμάτων βλέπουμε ότι στην pretrained έκδοση παρατηρείται ήδη ένας γενικός διαχωρισμός μεταξύ των δύο κλάσεων, γεγονός που υποδηλώνει ότι οι προεκπαιδευμένες αναπαραστάσεις περιέχουν χρήσιμη πληροφορία για τη διάκριση των εικόνων Fire και No Fire. Ωστόσο, η σημαντική επικάλυψη που εμφανίζεται στην κεντρική περιοχή του διαγράμματος δημιουργεί αβεβαιότητα κατά τη διαδικασία ταξινόμησης, γεγονός που εξηγεί τη χαμηλή Recall της pretrained έκδοσης. Μετά το fine-tuning, η γεωμετρία του χώρου χαρακτηριστικών γίνεται σαφώς πιο οργανωμένη. Οι δύο κλάσεις καταλαμβάνουν περισσότερο διακριτές περιοχές, ενώ οι συστάδες εμφανίζουν αυξημένη εσωτερική συνοχή. Η μεταβολή αυτή μειώνει σημαντικά την αβεβαιότητα στα όρια μεταξύ των δύο κατηγοριών και επιτρέπει στο μοντέλο να αναγνωρίζει πολύ μεγαλύτερο ποσοστό πραγματικών περιστατικών πυρκαγιάς, όπως αποτυπώνεται στη θεαματική αύξηση της Recall και του F1-score.

4.3 στο FASDD

Συνεχίζοντας το έργο της σύγκρισης των δύο εκδοχών (pretrained και finetuned) των μοντέλων, περνάμε στη παρούσα παράγραφο στην ανασκόπηση της σύγκρισης αυτής στα πλαίσια του συνόλου FASDD.

4.3.1 Το μοντέλο SegFormer

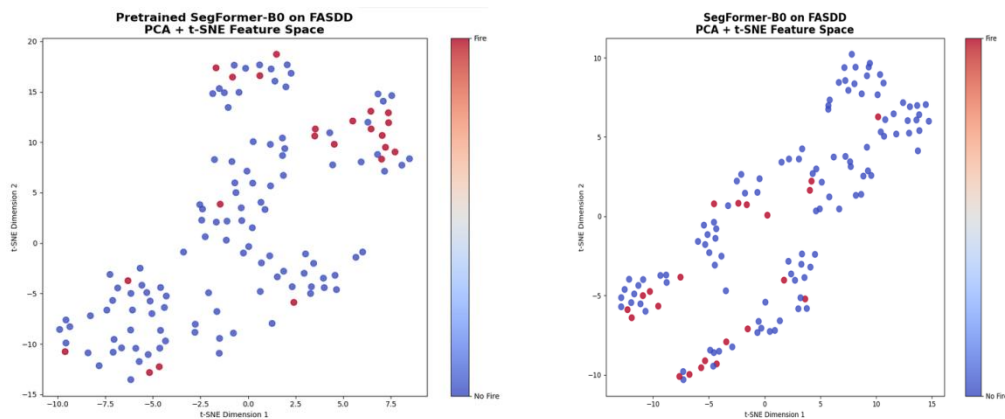
Η απόδοση του SegFormer-B0 στο FASDD στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-13 Σύγκριση Μετρικών SegFormer στο FASDD

Μετρική	Pretrained SegFormer-B0	Fine tuned SegFormer-B0	Μεταβολή
Accuracy	0.1756	0.8473	+0.6717
Precision	0.1692	0.6667	+0.4975
Recall	1.0000	0.1818	-0.8182
F1-score	0.2895	0.2857	-0.0038

Από την ανάγνωση των συγκριτικών αποτελεσμάτων των μετρικών στον Πίνακα βλέπουμε ότι η Accuracy αυξήθηκε εντυπωσιακά από 17,56% σε 84,73%, γεγονός που υποδηλώνει ότι το μοντέλο ταξινομεί σωστά πολύ μεγαλύτερο ποσοστό εικόνων μετά το fine tuning. Ωστόσο, η Recall παρουσίασε θα λέγαμε κατακόρυφη πτώση, από 100,00% σε μόλις 18,18%. Το αποτέλεσμα αυτό σημαίνει ότι το μοντέλο μετά το fine tuning αποτυγχάνει να εντοπίσει τη μεγάλη πλειονότητα των πραγματικών εικόνων πυρκαγιάς, αναγνωρίζοντας μόνο ένα μικρό ποσοστό αυτών. Στην περίπτωση του SegFormer-B0 στο FASDD το fine-tuning φαίνεται να μετέβαλε σημαντικά το όριο απόφασης του μοντέλου. Η pretrained έκδοση χαρακτηριζόταν από εξαιρετικά υψηλή Recall, γεγονός που δείχνει ότι το μοντέλο ταξινομούσε σχεδόν όλες τις εικόνες πυρκαγιάς ως θετικές, με κόστος όμως μεγάλο αριθμό ψευδώς θετικών προβλέψεων και πολύ χαμηλή Precision. Μετά το fine-tuning, διαπιστώνουμε ότι η συμπεριφορά αντιστράφηκε. Το μοντέλο έγινε

πολύ πιο επιλεκτικό στις θετικές προβλέψεις, βελτιώνοντας σημαντικά την Precision, αλλά χάνοντας μεγάλο αριθμό πραγματικών εικόνων πυρκαγιάς.



Εικόνα 4-13 Σύγκριση t-SNE SegFormer στο FASDD

Η σύγκριση των δύο t-SNE απεικονίσεων μας δείχνει ότι το fine-tuning επέφερε ορισμένες αλλαγές στην οργάνωση των χαρακτηριστικών, χωρίς όμως να δημιουργήσει σαφή και πλήρως διαχωρισμένο χώρο μεταξύ των δύο κατηγοριών. Μετά τη διαδικασία fine tuning, ο χώρος χαρακτηριστικών εμφανίζεται ελαφρώς περισσότερο οργανωμένος. Παρ'όλα αυτά η μικρή βελτίωση της οργάνωσης του χώρου χαρακτηριστικών δεν συνοδεύτηκε όπως σημειώσαμε στην προηγούμενη παράγραφο από αντίστοιχη βελτίωση της συνολικής ικανότητας του μοντέλου να διαχωρίζει αποτελεσματικά τις δύο κατηγορίες.

4.3.2 Το μοντέλο Yolov26

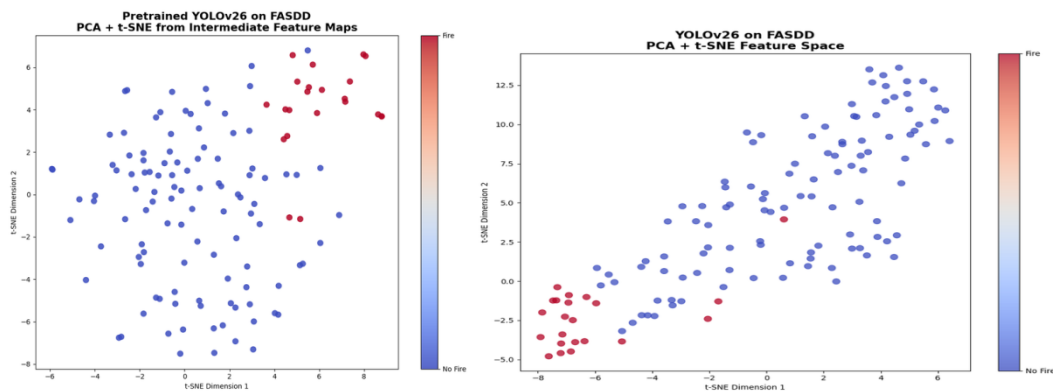
Η απόδοση του Yolov26 στο FASDD στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-14 Σύγκριση Μετρικών Yolov26 στο FASDD

Μετρική	Pretrained Yolov26	Fine tuned Yolov26	Μεταβολή
Accuracy	0.4885	0.8397	+0.3512
Precision	0.2472	1.0000	+0.7528

Recall	1.0000	0.0455	-0.9545
F1-score	0.3964	0.0870	-0.3094

Στην περίπτωση αυτή τα αποτελέσματα δείχνουν ότι το fine tuning δεν οδήγησε σε ουσιαστική βελτίωση του Yolov26 στο σύνολο δεδομένων FASDD. Αντίθετα, το μοντέλο ανέπτυξε έντονη μεροληψία υπέρ της κλάσης No Fire, γεγονός που περιορίσε δραστικά την ικανότητά του να ανιχνεύει πραγματικά περιστατικά πυρκαγιάς. Η συγκεκριμένη συμπεριφορά πιθανόν να συνδέεται με το γεγονός ότι το Yolov26 έχει σχεδιαστεί πρωτίστως ως μοντέλο ανίχνευσης αντικειμένων (object detection) . Η προσαρμογή ενός τέτοιου μοντέλου σε ένα πρόβλημα δυαδικής ταξινόμησης με σχετικά μικρό και ανισόρροπο σύνολο δεδομένων όπως το FASDD ενδέχεται να οδήγησε σε υπερπροσαρμογή των ορίων απόφασης και σε έντονη μεροληψία υπέρ της κυρίαρχης κλάσης.



Εικόνα 4-14 Σύγκριση t-SNE Yolov26 στο FASDD

Η συγκριτική ανάγνωση των διαγραμμάτων t-SNE παρουσιάζει ένα πολύ ενδιαφέρον σημείο. Στο pretrained μοντέλο παρατηρείται ότι οι εικόνες της κλάσης Fire σχηματίζουν μία σχετικά διακριτή ομάδα στο πάνω δεξί τμήμα του χώρου χαρακτηριστικών, ενώ οι εικόνες No Fire καταλαμβάνουν το μεγαλύτερο μέρος του διαγράμματος. Μετά το fine-tuning, το διάγραμμα εμφανίζει έναν ακόμη πιο καθαρό γεωμετρικό διαχωρισμό μεταξύ των δύο κλάσεων. Ωστόσο, η βελτίωση αυτή δεν μεταφράστηκε όπως είδαμε σε καλύτερη συμπεριφορά ταξινόμησης. Το γεγονός αυτό πιθανώς να δείχνει ότι, παρότι το μοντέλο έμαθε περισσότερο διακριτές αναπαραστάσεις, το τελικό όριο απόφασης που χρησιμοποιεί για την ταξινόμηση μετατοπίστηκε υπερβολικά προς την κλάση No Fire, με αποτέλεσμα να απορρίπτει σχεδόν όλες τις πραγματικές εικόνες πυρκαγιάς.

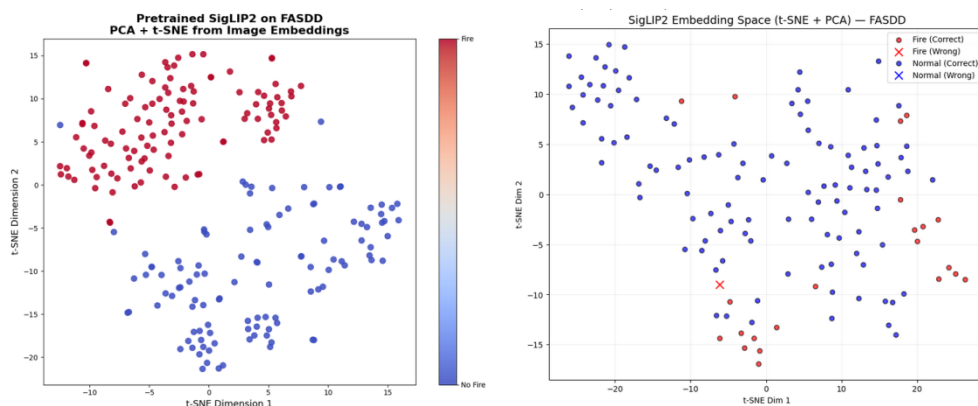
4.3.3 Το μοντέλο SigLIP2

Η απόδοση του SigLIP2 στο FASDD στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-15 Σύγκριση Μετρικών SigLIP2 στο FASDD

Μετρική	Pretrained SigLIP2	Fine tuned SigLIP2	Μεταβολή
Accuracy	0.3740	0.9924	+0.6184
Precision	0.1591	0.9954	+0.8363
Recall	0.6364	0.9783	+0.3419
F1-score	0.2545	0.9866	+0.7321

Τα αποτελέσματα του Πίνακα δείχνουν ότι η προσαρμογή του μοντέλου στο συγκεκριμένο σύνολο δεδομένων οδήγησε σε θεαματική βελτίωση όλων των βασικών μετρικών αξιολόγησης. Αξίζει στο σημείο αυτό να σημειώσουμε πως το ίδιο μοντέλο εμφάνισε την αντίθετη συμπεριφορά στο σύνολο δεδομένων FLAME, όπου το fine-tuning οδήγησε σε υποβάθμιση της απόδοσης του. Ένα ενδεχόμενο συμπέρασμα που θα μπορούσαμε να εξάγουμε από την παρατήρηση αυτή είναι ότι η αποτελεσματικότητα της διαδικασίας fine-tuning εξαρτάται σε μεγάλο βαθμό από τα χαρακτηριστικά και την κατανομή των δεδομένων εκπαίδευσης, καθώς και από τον βαθμό συμβατότητας των προεκπαιδευμένων αναπαραστάσεων με το εκάστοτε σύνολο δεδομένων.



Εικόνα 4-15 Σύγκριση t-SNE SigLIP2 στο FASDD

Η συγκριτική μελέτη τώρα των δύο t-SNE είναι θα λέγαμε εν πρώτοις αντιφατική. Στην pretrained έκδοση παρατηρούνται δύο σχετικά συμπαγείς και καλά διαχωρισμένες συστάδες. Η εικόνα αυτή υποδηλώνει ότι το μοντέλο παράγει ήδη χρήσιμες αναπαραστάσεις χαρακτηριστικών. Μετά τη διαδικασία fine tuning, ο χώρος χαρακτηριστικών εμφανίζεται περισσότερο διάχυτος και λιγότερο οργανωμένος. Βέβαια παρά τη μειωμένη οπτική συνοχή των συστάδων, η μεγάλη πλειονότητα των δειγμάτων ταξινομείται σωστά από το finetuned μοντέλο, ενώ μόνο ελάχιστα σημεία εμφανίζονται ως λανθασμένες ταξινομήσεις. Τα συγκεκριμένα διαγράμματα t-SNE μας δείχνουν ότι η ποιότητα ενός χώρου χαρακτηριστικών δεν μπορεί να αξιολογείται αποκλειστικά από τον βαθμό οπτικού διαχωρισμού των συστάδων. Στην περίπτωση του SigLIP2, η σημαντική βελτίωση της ταξινόμησης φαίνεται πως προκύπτει από την αποτελεσματικότερη αξιοποίηση των χαρακτηριστικών και όχι από τη δημιουργία περισσότερο συμπαγών ή καλύτερα διαχωρισμένων συστάδων.

4.3.4 Το μοντέλο EdBianchiVit

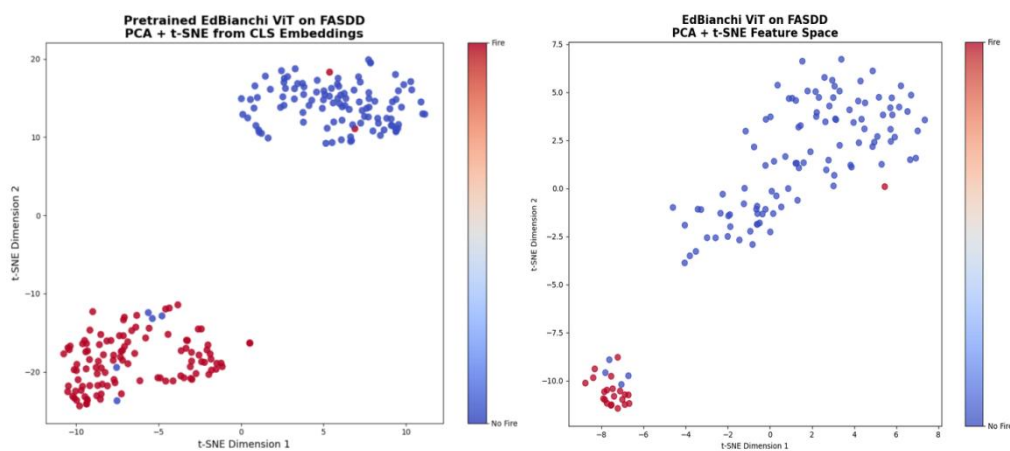
Η απόδοση του EdBianchiVit στο FASDD στην προεκπαιδευμένη και τη finetuned έκδοσή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-16 Σύγκριση Μετρικών EdBianchiVit στο FASDD

Μετρική	Pretrained EdBianchi	Fine tuned EdBianchi	Μεταβολή
Accuracy	0.9313	0.9618	+0.0305
Precision	0.7097	0.8400	+0.1303
Recall	1.0000	0.9545	-0.0455
F1-score	0.8302	0.8936	+0.0634

Με βάση τον Πίνακα βλέπουμε πως μετά το fine tuning, το μοντέλο υιοθετεί μια περισσότερο ισορροπημένη στρατηγική ταξινόμησης. Η σημαντική αύξηση της Precision δείχνει ότι μειώθηκαν οι λανθασμένες θετικές προβλέψεις, ενώ η μικρή μείωση της Recall υποδηλώνει ότι το μοντέλο έγινε ελαφρώς πιο συντηρητικό στην αναγνώριση εικόνων πυρκαγιάς. Η νέα αυτή ισορροπία οδηγεί σε υψηλότερο F1-score, αποδεικνύοντας ότι η

συνολική ποιότητα της ταξινόμησης βελτιώθηκε. Θα πρέπει βέβαια να σημειώσουμε ότι η pretrained έκδοση του μοντέλου εμφανίζει ήδη ιδιαίτερα υψηλή απόδοση, γεγονός που περιορίζει τα περιθώρια βελτίωσης μέσω της διαδικασίας fine-tuning. Έτσι θα μπορούσαμε να πούμε ότι το fine tuning λειτούργησε κυρίως ως διαδικασία εξισορρόπησης της συμπεριφοράς του μοντέλου, βελτιώνοντας την ακρίβεια των θετικών προβλέψεων χωρίς να επηρεάσει σημαντικά τη συνολική ικανότητα ανίχνευσης πυρκαγιάς.



Εικόνα 4-16 Σύγκριση t-SNE EdBianchiVit στο FASDD

Η εικόνα της σχεδόν ισορροπίας όσον αφορά τα ποσοτικά αποτελέσματα των μετρικών της pretrained και finetuned εκδοχής του μοντέλου φαίνεται να διατηρείται και στο χώρο της αναπαράστασης των χαρακτηριστικών όπως μας δηλώνουν οι γεωμετρίες των t-SNE διαγραμμάτων. Συμπερασματικά το fine tuning δεν άλλαξε ιδιαίτερα έναν ήδη πολύ καλό χώρο χαρακτηριστικών, αλλά τον βελτίωσε ποιοτικά. Η ποιοτική αυτή βελτίωση συνίσταται στο ότι στην finetuned εκδοχή οι εικόνες της κατηγορίας Fire εμφανίζονται περισσότερο συγκεντρωμένες, ενώ οι εικόνες No Fire οργανώνονται με μεγαλύτερη ομοιογένεια στον χώρο χαρακτηριστικών. Παράλληλα, σημειώνεται σχεδόν πλήρης απουσία επικάλυψης μεταξύ των δύο κατηγοριών, με εξαίρεση ελάχιστα μεμονωμένα δείγματα που βρίσκονται κοντά στα όρια των δύο συστάδων.

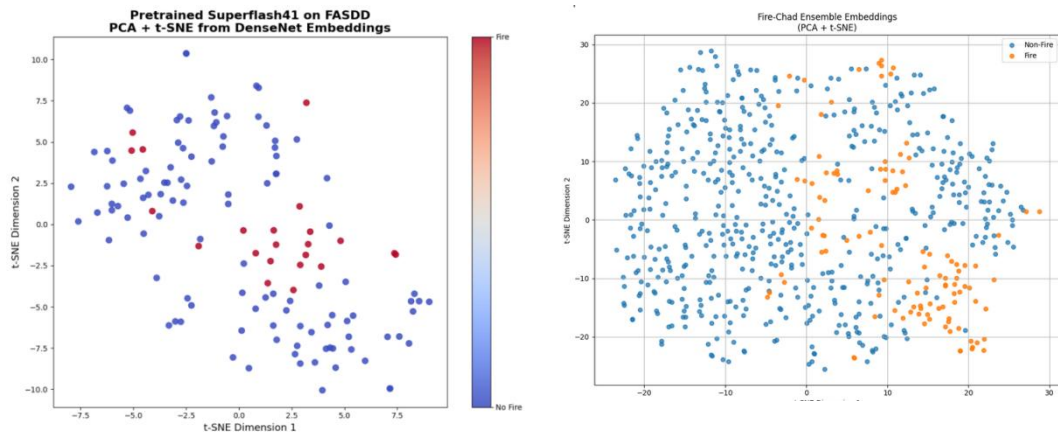
4.3.5 Το μοντέλο Superflash41

Η απόδοση του Superflash41 στο FASDD στην προεκπαιδευμένη και τη finetuned εκδοχή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-17 Σύγκριση Μετρικών Superflash41 στο FASDD

Μετρική	Pretrained Superflash41	Fine tuned Superflash41	Μεταβολή
Accuracy	0.2061	0.7176	+0.5115
Precision	0.1058	0.3404	+0.2346
Recall	0.5000	0.7273	+0.2273
F1-score	0.1746	0.4638	+0.2892

Παρατηρούμε ότι το fine-tuning βελτίωσε την απόδοση του Superflash41 στο FASDD, με την Accuracy να αυξάνεται από 20,61% σε 71,76% και το F1-score από 0,1746 σε 0,4638. Βεβαίως, η τελική τιμή του F1-score παραμένει σχετικά χαμηλή. Η Recall αυξήθηκε από 50,00% σε 72,73%, γεγονός που δείχνει ότι το μοντέλο αναγνωρίζει πλέον περίπου τρεις στις τέσσερις πραγματικές εικόνες πυρκαγιάς, έναντι των μισών στην pretrained εκδοχή. Παράλληλα, η Precision βελτιώθηκε από 10,58% σε 34,04%, παραμένει όμως περιορισμένη: περίπου μία στις τρεις θετικές προβλέψεις αντιστοιχεί πράγματι σε εικόνα πυρκαγιάς. Επομένως, παρά τη βελτίωση της ανίχνευσης των θετικών περιστατικών, οι ψευδώς θετικές προβλέψεις εξακολουθούν να αποτελούν σημαντική αδυναμία του μοντέλου. Μία πιθανή εξήγηση είναι ότι το σχετικά μικρό μέγεθος του συνόλου δεδομένων FASDD ενδεχομένως δεν παρείχε επαρκή ποικιλία δειγμάτων ώστε να προσαρμοστούν αποτελεσματικά οι τρεις CNN υπομονάδες (DenseNet, ResNet και Xception) που συνθέτουν το Superflash41. Ως αποτέλεσμα, το fine-tuning φαίνεται να βελτίωσε κυρίως τα όρια απόφασης του τελικού ταξινομητή, χωρίς να οδηγήσει σε ουσιαστική αναβάθμιση των εσωτερικών αναπαραστάσεων χαρακτηριστικών. Το τελευταίο αυτό συμπέρασμα περί της μη ουσιαστικής αναβάθμισης των εσωτερικών αναπαραστάσεων χαρακτηριστικών φαίνεται πως έρχονται να μας το επιβεβαιώσουν και τα t-SNE διαγράμματα.



Εικόνα 4-17 Σύγκριση t-SNE Superflash41 στο FASDD

Η σημαντική επικάλυψη που εξακολουθεί να παρατηρείται μεταξύ των δύο κλάσεων μετά το fine tuning μας υποβάλει την ιδέα ότι το μοντέλο δεν κατάφερε να μάθει περισσότερο διακριτές αναπαραστάσεις των εικόνων πυρκαγιάς. Αντίθετα, η βελτίωση της Accuracy και της Precision φαίνεται να οφείλεται κυρίως σε καλύτερη προσαρμογή του τελικού μηχανισμού ταξινόμησης και όχι σε ουσιαστική μεταβολή των χαρακτηριστικών που εξάγονται από τις εικόνες.

4.3.6 Το μοντέλο Convnext

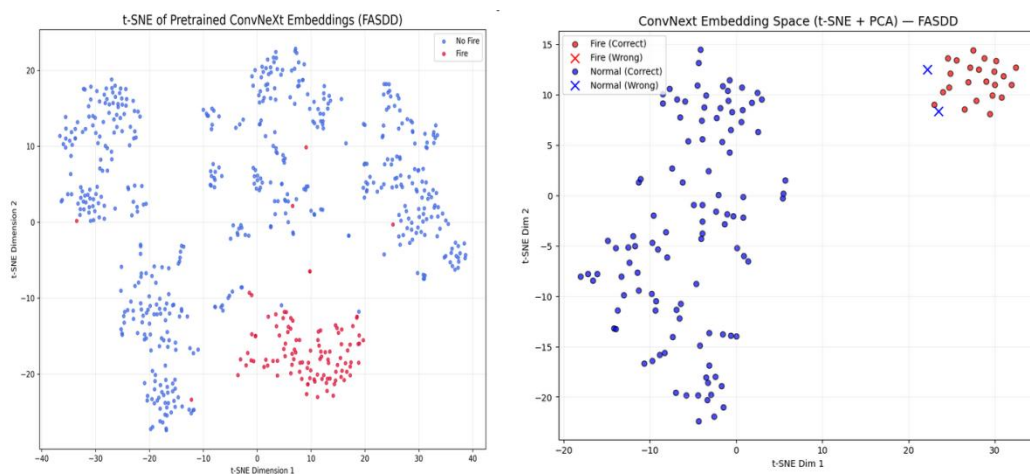
Η απόδοση του Convnext στο FASDD στην προεκπαιδευμένη και τη finetuned έκδοσή του μοντέλου ως προς τις διάφορες μετρικές καταγράφεται στον ακόλουθο πίνακα.

Πίνακας 4-18 Σύγκριση Μετρικών Convnext στο FASDD

Μετρική	Pretrained Convnext	Fine tuned Convnext	Μεταβολή
Accuracy	0.6118	0.9847	+0.3729
Precision	0.2399	0.9200	+0.6801
Recall	0.5963	1.0000	+0.4037
F1-score	0.3421	0.9583	+0.6162

Αυτό που παρατηρούμε από τον Πίνακα είναι ότι η pretrained έκδοση του Convnext εμφανίζει σχετικά χαμηλή συνολική απόδοση, κυρίως λόγω της πολύ χαμηλής Precision. Η συμπεριφορά αυτή δείχνει ότι το μοντέλο είχε την τάση να

χαρακτηρίζει αρκετές εικόνες ως Fire, ακόμη και όταν αυτές ανήκαν στην κατηγορία No Fire, οδηγώντας σε αυξημένο αριθμό ψευδώς θετικών προβλέψεων. Μετά το fine tuning, η εικόνα μεταβάλλεται πλήρως. Η σημαντική αύξηση της Precision δείχνει ότι το μοντέλο απέκτησε πολύ καλύτερη ικανότητα διάκρισης μεταξύ των δύο κατηγοριών, μειώνοντας δραστικά τα ψευδώς θετικά αποτελέσματα. Παράλληλα, η επίτευξη Recall ίσης με 100% δηλώνει ότι όλες οι πραγματικές εικόνες πυρκαγιάς αναγνωρίστηκαν επιτυχώς. Η ιδιαίτερα καλή απόδοση του finetuned μοντέλου θα πρέπει να συνεκτιμηθεί και με το γεγονός ότι το σύνολο FASDD είναι το μικρότερο από τα τρία σύνολα δεδομένων που χρησιμοποιούμε. Συνήθως, τα μικρότερα σύνολα δεδομένων αυξάνουν τον κίνδυνο υπερπροσαρμογής κατά το fine-tuning, κάτι που δεν ίσχυσε για το Convnext.



Εικόνα 4-18 Σύγκριση t-SNE Convnext στο FASDD

Τα διαγράμματα t-SNE επιβεβαιώνουν ότι το fine tuning οδήγησε σε ουσιαστική βελτίωση της ποιότητας των χαρακτηριστικών που εξάγει το Convnext, δημιουργώντας έναν ιδιαίτερα διακριτό χώρο αναπαραστάσεων, ο οποίος εξηγεί την σχεδόν άριστη απόδοση του μοντέλου στο σύνολο δεδομένων FASDD. Στη fine tuned εκδοχή, οι αναπαραστάσεις των δύο κατηγοριών αποκτούν πολύ μεγαλύτερη συνοχή και οργανώνονται σε δύο σαφώς διακριτές περιοχές του χώρου χαρακτηριστικών. Η σημαντική αυτή αναδιοργάνωση μειώνει ουσιαστικά την αβεβαιότητα στα όρια μεταξύ των δύο κλάσεων και επιτρέπει στο μοντέλο

να ταξινομεί σχεδόν όλες τις εικόνες σωστά, όπως είδαμε και από τις εξαιρετικές τιμές των μετρικών του finetuned μοντέλου.

Κεφάλαιο 5 Συμπεράσματα

Στο παρόν Κεφάλαιο αρχικά προβαίνουμε σε μια περίληψη της ερευνητικής πορείας της εργασίας μας. Ακολούθως παρουσιάζουμε μια γενικότερη αποτίμηση των αποτελεσμάτων των πειραμάτων που εκτελέσαμε και συναγάγουμε τα τελικά συμπεράσματα της εργασίας.

5.1 Γενική Αποτίμηση

Η συνεχής αύξηση της συχνότητας και της έντασης των δασικών πυρκαγιών τα τελευταία χρόνια σε συνδυασμό με τις ευρύτερες επιχειρησιακές δυνατότητες που προσφέρουν τα drones, έχουν δημιουργήσει αυξημένο ερευνητικό ενδιαφέρον για την ανάπτυξη αυτοματοποιημένων συστημάτων αναγνώρισης πυρκαγιών από εναέριας εικόνες. Συγχρόνως η αλματώδης εξέλιξη των αλγορίθμων βαθιάς μάθησης έχει οδηγήσει στην ανάπτυξη πληθώρας διαφορετικών αρχιτεκτονικών που διαφέρουν μεταξύ τους ως προς τον τρόπο εξαγωγής χαρακτηριστικών όσο και ως προς τη δυνατότητα γενίκευσης σε νέα δεδομένα.

Σε αυτό το πλαίσιο, βασικός σκοπός της παρούσας εργασίας ήταν η συστηματική διερεύνηση της αποτελεσματικότητας του fine-tuning σε διαφορετικές αρχιτεκτονικές βαθιάς μάθησης για την ανίχνευση δασικών πυρκαγιών από εικόνες drones. Για αυτό το σκοπό συγκρίναμε έξι διαφορετικά μοντέλα βαθιάς μάθησης, τόσο στην pretrained όσο και στην fine-tuned εκδοχή τους, σε τρία διαφορετικά σύνολα δεδομένων με διαφορετικά χαρακτηριστικά.

Τα μοντέλα που αξιολογήσαμε αντιπροσωπεύουν διαφορετικές αρχιτεκτονικές. Συγκεκριμένα, εξετάστηκαν δύο αρχιτεκτονικές βασισμένες στους Vision Transformers (SigLIP2 και EdBianchi ViT), μία σύγχρονη αρχιτεκτονική Συνελκτικών Νευρωνικών Δικτύων (ConvNeXt), ένα ensemble CNN (Superflash41), μία αρχιτεκτονική ανίχνευσης αντικειμένων (YOLOv26) και μία αρχιτεκτονική σημασιολογικής τμηματοποίησης (SegFormer-B0). Η επιλογή αυτή επέτρεψε την αξιολόγηση όχι μόνο διαφορετικών μοντέλων, αλλά και διαφορετικών οικογενειών αλγορίθμων υπολογιστικής όρασης.

Αντίστοιχα, χρησιμοποιήθηκαν τρία σύνολα δεδομένων με διαφορετική φιλοσοφία σχεδίασης και διαφορετικό βαθμό δυσκολίας. Το FLAME αποτέλεσε το μεγαλύτερο σύνολο δεδομένων της μελέτης και περιλάμβανε εικόνες από drones με διαθέσιμες μάσκες

σημασιολογικής τμηματοποίησης. Το DFire χαρακτηριζόταν από μεγάλη ποικιλία σκηνών, φωτιστικών συνθηκών και πραγματικών περιστατικών πυρκαγιάς, αποτελώντας ένα ιδιαίτερα απαιτητικό σύνολο αξιολόγησης της ικανότητας γενίκευσης των μοντέλων. Τέλος, χρησιμοποιήθηκε ένα αντιπροσωπευτικό υποσύνολο του FASDD, σημαντικά μικρότερο σε μέγεθος και με έντονη ανισορροπία μεταξύ των δύο κλάσεων, γεγονός που επέτρεψε τη διερεύνηση της συμπεριφοράς των μοντέλων σε συνθήκες περιορισμένων και μη ισορροπημένων δεδομένων.

Για τη διασφάλιση της αντικειμενικότητας της συγκριτικής αξιολόγησης εφαρμόσαμε κοινή πειραματική διαδικασία για όλα τα μοντέλα. Η διαδικασία του fine-tuning πραγματοποιήθηκε με τις ίδιες υπερπαραμέτρους εκπαίδευσης, ενώ η αξιολόγηση βασίστηκε στις ίδιες ποσοτικές μετρικές (Accuracy, Precision, Recall και F1-score). Η ποιοτική διάσταση της αξιολόγησης συμπληρώθηκε με την εξέταση του χώρου χαρακτηριστικών μέσω t-SNE καθώς και με την ερμηνεία των αποφάσεων των μοντέλων με τη μέθοδο Grad-CAM. Η πολυδιάστατη αυτή διαδικασία αξιολόγησης μας επέτρεψε όχι μόνο τη μέτρηση της τελικής απόδοσης αλλά και τη βαθύτερη κατανόηση των μηχανισμών που οδηγούν κάθε μοντέλο στις αντίστοιχες αποφάσεις.

Η συνολική εικόνα των αποτελεσμάτων ανέδειξε ένα ιδιαίτερα ενδιαφέρον συμπέρασμα. Παρά την κοινή διαδικασία προσαρμογής, τα έξι μοντέλα δεν παρουσίασαν ενιαία συμπεριφορά. Σε ορισμένες περιπτώσεις, όπως το ConvNeXt και το EdBianchi ViT, το fine-tuning οδήγησε σε εντυπωσιακή βελτίωση της απόδοσης στα περισσότερα σύνολα δεδομένων. Αντίθετα, σε άλλες περιπτώσεις, όπως το SegFormer-B0, το SigLIP2 ή το YOLOv26 σε ορισμένα datasets, η προσαρμογή δεν απέφερε τα αναμενόμενα αποτελέσματα ή ακόμη και υποβάθμισε ορισμένες κρίσιμες μετρικές. Το εύρημα αυτό υποδηλώνει ότι η αποτελεσματικότητα του fine-tuning δεν μπορεί να θεωρείται δεδομένη, αλλά εξαρτάται από μία σύνθετη αλληλεπίδραση μεταξύ της αρχιτεκτονικής του μοντέλου, της φύσης των διαθέσιμων δεδομένων και του τρόπου με τον οποίο αυτά αξιοποιούνται κατά την εκπαίδευση.

5.2 Επίδραση των Συνόλων Δεδομένων

Ένα από τα σημαντικότερα συμπεράσματα που προέκυψαν από την παρούσα εργασία αφορά την καθοριστική επίδραση των χαρακτηριστικών των συνόλων δεδομένων στη συμπεριφορά των εξεταζόμενων μοντέλων. Όσον αφορά το FASDD η έντονη αυτή ανισορροπία των κλάσεων του δημιούργησε ένα ιδιαίτερα απαιτητικό περιβάλλον αξιολόγησης, στο οποίο αρκετά μοντέλα εμφάνισαν δυσκολία προσαρμογής. Τα αποτελέσματα έδειξαν ότι το μικρό μέγεθος του συνόλου δεδομένων αύξησε σημαντικά την ευαισθησία των μοντέλων στη διαδικασία fine-tuning. Ορισμένες αρχιτεκτονικές, όπως το ConvNeXt και το SigLIP2, κατάφεραν να αξιοποιήσουν αποτελεσματικά τα διαθέσιμα δεδομένα, παρουσιάζοντας θεαματική βελτίωση της απόδοσής τους. Αντίθετα, άλλα μοντέλα, όπως το YOLOv26 και το SegFormer-B0, παρουσίασαν έντονη μεταβολή της συμπεριφοράς τους, με σημαντική μείωση της Recall και περιορισμένη συνολική βελτίωση του F1-score. Η συμπεριφορά αυτή υποδηλώνει ότι η περιορισμένη διαθεσιμότητα θετικών δειγμάτων μπορεί να οδηγήσει ορισμένες αρχιτεκτονικές στην υιοθέτηση υπερβολικά συντηρητικών στρατηγικών ταξινόμησης, με αποτέλεσμα την αύξηση των ψευδώς αρνητικών προβλέψεων.

Σε σχέση με το FLAME οπωσδήποτε ο μεγάλο αριθμός εικόνων έδωσε τη δυνατότητα σε αρκετές αρχιτεκτονικές να αξιοποιήσουν αποτελεσματικότερα τη διαδικασία προσαρμογής. Το ConvNeXt, το EdBianchi ViT και το Superflash41 παρουσίασαν σημαντική βελτίωση όλων σχεδόν των ποσοτικών μετρικών τους, γεγονός που υποδηλώνει ότι η επάρκεια εκπαιδευτικών δεδομένων διευκολύνει την εκμάθηση περισσότερο αντιπροσωπευτικών χαρακτηριστικών.

Παράλληλα όμως, το FLAME ανέδειξε και μία ιδιαίτερα ενδιαφέρουσα παρατήρηση. Παρά το μεγάλο μέγεθός του, δεν επωφελήθηκαν όλες οι αρχιτεκτονικές από τη διαδικασία fine-tuning. Το SigLIP2 και ιδιαίτερα τα SegFormer-B0 και YOLOv26 παρουσίασαν αισθητή υποβάθμιση της απόδοσής τους, γεγονός που δείχνει ότι η ύπαρξη μεγάλου αριθμού εικόνων δεν αποτελεί από μόνη της εγγύηση επιτυχούς προσαρμογής. Προκύπτει λοιπόν το συμπέρασμα ότι η αποτελεσματικότητα του fine-tuning εξαρτάται και από τον βαθμό συμβατότητας μεταξύ της αρχικής προεκπαίδευσης του μοντέλου και των ιδιαίτερων χαρακτηριστικών του συνόλου δεδομένων.

Τέλος όσον αφορά στο σύνολο DFire αποτελέσε το σύνολο δεδομένων της παρούσας μελέτης με την μεγαλύτερη ποικιλία των εικόνων. Σε αντίθεση με το FLAME, οι εικόνες του DFire προέρχονται από διαφορετικές πηγές και καλύπτουν μεγάλο εύρος συνθηκών φωτισμού, και μεγεθών και εντάσεων πυρκαγιάς, καθώς και μεγάλη ποικιλία φυσικών τοπίων και αντικειμένων μέσα στις εικόνες.

Παρά την αυξημένη δυσκολία του, λόγω της ετερογένειας των δεδομένων του τα περισσότερα μοντέλα παρουσίασαν ιδιαίτερα σημαντική βελτίωση στο DFire μετά το fine-tuning. Το γεγονός αυτό υποδηλώνει ότι η ποικιλία των διαθέσιμων εικόνων λειτούργησε ευεργετικά για την εκμάθηση περισσότερο γενικεύσιμων χαρακτηριστικών, επιτρέποντας στα μοντέλα να προσαρμοστούν αποτελεσματικότερα στις πραγματικές διαφοροποιήσεις των δεδομένων. Μπορούμε να σημειώσουμε πως ιδιαίτερα εντυπωσιακή ήταν η βελτίωση των Convnext, EdBianchi ViT, SigLIP2 και YOLOv26, τα οποία εμφάνισαν σημαντική αύξηση τόσο της Accuracy όσο και του F1-score.

5.3 Επίδραση των Αρχιτεκτονικών των Μοντέλων

Ένα βασικό συμπέρασμα της παρούσας εργασίας αφορά τον καθοριστικό ρόλο της αρχιτεκτονικής των μοντέλων στη διαδικασία αναγνώρισης δασικών πυρκαγιών. Παρότι όλα τα εξεταζόμενα μοντέλα επιλύουν το ίδιο πρόβλημα ταξινόμησης, βασίζονται σε διαφορετικές φιλοσοφίες επεξεργασίας της εικόνας, γεγονός που επηρεάζει τόσο τον τρόπο εκμάθησης των χαρακτηριστικών όσο και την ικανότητα προσαρμογής τους σε νέα σύνολα δεδομένων. Η συγκριτική αξιολόγηση που διεξάγαμε έδειξε ότι η επιλογή της αρχιτεκτονικής επηρεάζει τον τρόπο με τον οποίο κάθε μοντέλο ανταποκρίνεται στη διαδικασία fine-tuning. Αποδείχθηκε ότι διαφορετικές οικογένειες μοντέλων παρουσιάζουν διαφορετικό βαθμό ευαισθησίας στα χαρακτηριστικά των συνόλων δεδομένων, γεγονός που καθιστά αδύνατη την ανάδειξη μιας καθολικά βέλτιστης αρχιτεκτονικής για κάθε εφαρμογή ανίχνευσης πυρκαγιάς.

Αρχικά ως προς τα Συνελκτικά Νευρωνικά Δίκτυα η πλέον αντιπροσωπευτική περίπτωση της μελέτης ήταν το Convnext, το οποίο ανήκει στη σύγχρονη γενιά των Συνελκτικών Νευρωνικών Δικτύων. Παρά το γεγονός ότι η αρχιτεκτονική του ενσωματώνει αρκετές σχεδιαστικές επιλογές εμπνευσμένες από τους Vision Transformers, εξακολουθεί να

βασίζεται στη θεμελιώδη αρχή της συνελκτικής επεξεργασίας της εικόνας. Τα αποτελέσματα της εργασίας έδειξαν ότι το ConvNeXt παρουσίασε την πιο σταθερή συμπεριφορά μεταξύ όλων των εξεταζόμενων μοντέλων. Η διαδικασία fine-tuning οδήγησε σε σημαντική βελτίωση και στα τρία σύνολα δεδομένων, με ιδιαίτερα υψηλές τιμές Accuracy και F1-score. Η συμπεριφορά αυτή υποδηλώνει ότι η αρχιτεκτονική του διαθέτει υψηλή ικανότητα προσαρμογής ακόμη και όταν τα διαθέσιμα δεδομένα παρουσιάζουν σημαντικές διαφοροποιήσεις ως προς το μέγεθος ή την κατανομή των κλάσεων. Επίσης θα πρέπει να σημειώσουμε ότι η βελτίωση των ποσοτικών μετρικών συνοδεύτηκε και από σαφέστερο διαχωρισμό των κλάσεων στα διαγράμματα t-SNE, καθώς και από περισσότερο στοχευμένη εστίαση στις περιοχές της πυρκαγιάς στις απεικονίσεις Grad-CAM.

Η δεύτερη μεγάλη οικογένεια αρχιτεκτονικών που εξετάστηκε στην εργασία ήταν οι Μετασχηματιστές Όρασης (Vision Transformers), εκπροσωπούμενοι από τα μοντέλα SigLIP2 και EdBianchi ViT. Ένα ιδιαίτερα ενδιαφέρον εύρημα της μελέτης ήταν ότι, παρά το γεγονός ότι και τα δύο μοντέλα ανήκουν στην ίδια οικογένεια αρχιτεκτονικών, παρουσίασαν σημαντικά διαφορετική συμπεριφορά. Το αποτέλεσμα αυτό καταδεικνύει ότι η συνολική απόδοση ενός Vision Transformer δεν εξαρτάται αποκλειστικά από την ίδια την αρχιτεκτονική Transformer, αλλά επηρεάζεται καθοριστικά από τον τρόπο προεκπαίδευσης, το μέγεθος και την ποιότητα των δεδομένων προεκπαίδευσης, καθώς και από τον σχεδιασμό της τελικής διαδικασίας προσαρμογής. Το EdBianchi ViT παρουσίασε σταθερή και ιδιαίτερα υψηλή απόδοση σχεδόν σε όλα τα σύνολα δεδομένων. Ακόμη και στην προεκπαιδευμένη μορφή του εμφάνιζε ικανοποιητική ικανότητα γενίκευσης, ενώ μετά τη διαδικασία fine-tuning κατέγραψε σημαντικές βελτιώσεις στις περισσότερες μετρικές αξιολόγησης. Η συμπεριφορά αυτή υποδηλώνει ότι οι αναπαραστάσεις που είχαν αποκτηθεί κατά την προεκπαίδευση ήταν ιδιαίτερα κατάλληλες για το πρόβλημα της ανίχνευσης πυρκαγιάς. Αντίθετα, το SigLIP2 παρουσίασε μεγαλύτερη μεταβλητότητα μεταξύ των τριών συνόλων δεδομένων. Σε ορισμένες περιπτώσεις, όπως στο FASDD και στο DFire, η διαδικασία fine-tuning οδήγησε σε θεαματική βελτίωση της απόδοσης. Ωστόσο, στο FLAME παρατηρήθηκε αισθητή υποβάθμιση των αποτελεσμάτων μετά την προσαρμογή.

Στην τρίτη οικογένεια αρχιτεκτονικών βρίσκουμε τα Ensemble Συνελικτικά Δίκτυα. Το Superflash41 ως εκπρόσωπος αυτής της οικογένειας βασίζει την σχεδιάσή του στον συνδυασμό τριών ανεξάρτητων συνελικτικών δικτύων (DenseNet, ResNet και Xception). Η λογική εξάλλου των ensemble μοντέλων είναι ότι ο συνδυασμός διαφορετικών εξαγωγέων χαρακτηριστικών μπορεί να οδηγήσει σε μεγαλύτερη ανθεκτικότητα απέναντι στις ιδιαιτερότητες των δεδομένων. Τα αποτελέσματα επιβεβαίωσαν αλλά μόνο εν μέρει την υπόθεση αυτή. Το μοντέλο παρουσίασε αξιοσημείωτη βελτίωση μετά το fine-tuning, ιδιαίτερα στα μεγαλύτερα σύνολα δεδομένων. Ωστόσο, η βελτίωσή του δεν ήταν αντίστοιχη εκείνης του Convnex. Η συμπεριφορά αυτή πιθανόν οφείλεται στη μεγαλύτερη πολυπλοκότητα της αρχιτεκτονικής, η οποία απαιτεί μεγαλύτερο πλήθος δεδομένων ώστε να μπορέσουν να αξιοποιηθούν αποτελεσματικά όλα τα επιμέρους δίκτυα του.

Τέλος στην κατηγορία των μοντέλων ανίχνευσης αντικειμένων και σημασιολογικής τμηματοποίησης ιδιαίτερο ενδιαφέρον παρουσίασε η αξιολόγηση των μοντέλων YOLOv26 και SegFormer-B0. Πρόκειται για δύο μοντέλα των οποίων οι αρχιτεκτονικές που δεν σχεδιάστηκαν αρχικώς για δυαδική ταξινόμηση εικόνων. Η προσαρμογή των δύο αυτών αρχιτεκτονικών σε ένα πρόβλημα καθαρής ταξινόμησης παρουσίασε ενδιαφέροντα αποτελέσματα. Παρότι σε ορισμένα datasets καταγράφηκαν σημαντικές βελτιώσεις, η συνολική συμπεριφορά τους ήταν λιγότερο σταθερή σε σχέση με τις αρχιτεκτονικές που είχαν εξ αρχής σχεδιαστεί για image classification. Η παρατήρηση αυτή είναι αναμενόμενη, καθώς μέρος της υπολογιστικής ισχύος των συγκεκριμένων μοντέλων παραμένει προσανατολισμένο στην εκτέλεση διαφορετικών εργασιών υπολογιστικής όρασης.

Ως καταληκτικό συμπέρασμα θα μπορούσαμε να πούμε πως η επιτυχία ενός μοντέλου στην ανίχνευση δασικών πυρκαγιών δεν εξαρτάται αποκλειστικά από το αν βασίζεται σε Συνελικτικά Νευρωνικά Δίκτυα ή σε Μετασχηματιστές Όρασης. Αντίθετα, φαίνεται να αποτελεί αποτέλεσμα της αλληλεπίδρασης μεταξύ της αρχιτεκτονικής, της στρατηγικής προεκπαίδευσης, της διαδικασίας fine-tuning και των χαρακτηριστικών του συνόλου δεδομένων.

Βιβλιογραφία

1. Bakirci, M., & Bayraktar, I. (2024). Harnessing UAV technology and YOLOv9 algorithm for real-time forest fire detection. *2024 International Russian Automation Conference (RusAutoCon)*, 95-100.
2. Boroujeni, S. P. H., Razi, A., Khoshdel, S., Afghah, F., Coen, J. L., O'Neill, L., Fulé, P., Watts, A., Kokolakis, N.-M. T., & Vamvoudakis, K. G. (2024). A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management. *Information Fusion*, 108, 102369.
3. Cazzolato, M. T., et al. (2017). FiSmo: A compilation of datasets from emergency situations for fire and smoke analysis. *Brazilian Symposium on Databases (SBBD)*.
4. Chen, X., Hopkins, B., Wang, H., O'Neill, L., Afghah, F., Razi, A., Fulé, P., Coen, J., Rowell, E., & Watts, A. (2022). Wildland fire detection and monitoring using a drone-collected RGB/IR image dataset. *IEEE Access*, 10, 121301-121317.
5. Chino, D. Y. T., Avalhais, L. P. S., Rodrigues, J. F., Jr., & Traina, A. J. M. (2015). BoWFire: Detection of fire in still images by integrating pixel color and texture analysis. *Proceedings of the 28th SIBGRAPI Conference on Graphics, Patterns and Images*.
6. Foggia, P., Saggese, A., & Vento, M. (2015). Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(9), 1545–1556.
7. Khan, A., Hassan, B., Khan, S., Ahmed, R., & Adnan, A. (2022). DeepFire: A novel dataset and deep transfer learning benchmark for forest fire detection. *Mobile Information Systems*, 2022.
8. Koutsikakis, J., Chalkidis, I., Malakasiotis, P., & Androutsopoulos, I. (2020). Greek-BERT: The Greeks visiting Sesame Street. <https://arxiv.org/pdf/2008.12014>.

9. Kumar, A., Perrusquía, A., Al-Rubaye, S., & Guo, W. (2024). Wildfire and smoke early detection for drone applications: A light-weight deep learning approach. *Engineering Applications of Artificial Intelligence*, 136, 108977.
10. Li, S., Yan, Q., & Liu, P. (2020). An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism. *IEEE Transactions on Image Processing*, 29, 8467–8475.
11. Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). Microsoft COCO: Common objects in context. *European Conference on Computer Vision (ECCV)*.
12. Lu, K., Xu, R., Li, J., Lv, Y., Lin, H., & Liu, Y. (2022). A vision-based detection and spatial localization scheme for forest fire inspection from UAV. *Forests*, 13(3), 383.
13. Marques, T., Carreira, S., Miragaia, R., Ramos, J., & Pereira, A. (2024). Applying deep learning to real-time UAV-based forest monitoring: Leveraging multi-sensor imagery for improved results. *Expert Systems With Applications*, 245, 123107.
14. Niu, K., Wang, C., Xu, J., Zhou, X., Wen, K., & Yang, C. (2025). Early Forest Fire Detection with UAV Image Fusion. *IEEE*, 18, 6617-6628.
15. Rubab, S. F., Ghaffar, A. A., & Choi, G. S. (2024). FireDetXplainer: Decoding wildfire detection with transparency and explainable AI insights. *IEEE Access*.
16. Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P. Z., & Blasch, E. (2021). Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Computer Networks*, 193, 108001.
17. Shamta, I., & Demir, B. E. (2024). Development of a deep learning-based surveillance system for forest fire detection and monitoring using UAV. *PLoS ONE*, 19(3), e0299058.

18. Tschannen, M., et al. (2025). SigLIP 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features. *Google DeepMind*.
19. Wang, G., Bai, D., Lin, H., Zhou, H., & Qian, J. (2024). FireViTNet: A hybrid model integrating ViT and CNNs for forest fire segmentation. *Computers and Electronics in Agriculture*, 218, 108722.
20. Yan, S., Zhang, J., & Barnes, N. (2022). Transmission-guided Bayesian generative model for smoke segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(3), 2707–2715.
21. Yar, H., et al. (2022). Optimized dual fire attention network and medium-scale fire classification benchmark. *IEEE Transactions on Image Processing*, 31, 6331–6343.
22. Yar, H., et al. (2023). An effective attention-based CNN model for fire detection in adverse weather conditions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206, 335–346.
23. Yar, H., Khan, Z. A., Rida, I., Ullah, W., Kim, M. J., & Baik, S. W. (2024). An efficient deep learning architecture for effective fire detection in smart surveillance. *Image and Vision Computing*, 145, 104989.

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.