

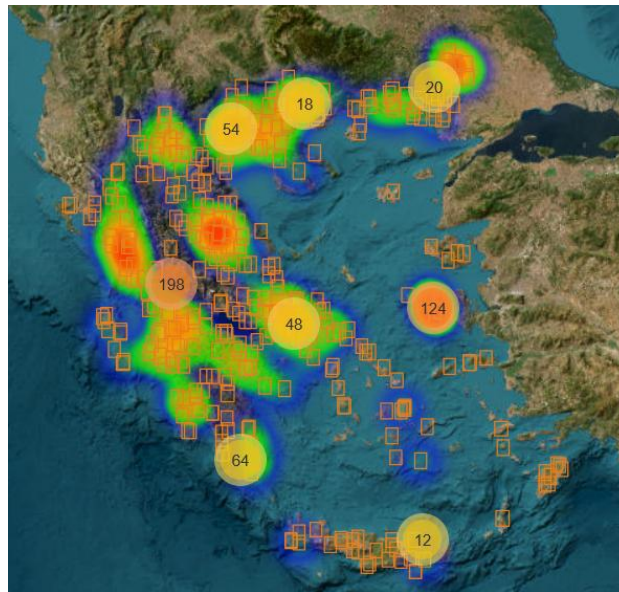


ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Υβριδικό Σύστημα Πρόβλεψης Δασικών Πυρκαγιών
με Χρήση Μηχανικής Μάθησης και Δεδομένων
Τηλεπισκόπησης»



Του Μεταπτυχιακού Φοιτητή
Στρογγύλη Διονύσιου
Αρ. Μητρώου: 24/2023.

Επιβλέπων
Χατζημίσιος Περικλής
Καθηγητής

Ημερομηνία 10 / 02 / 2026

Τίτλος Δ.Ε. Υβριδικό Σύστημα Πρόβλεψης Δασικών Πυρκαγιών με Χρήση Μηχανικής Μάθησης και
Δεδομένων Τηλεπισκόπησης

Κωδικός Δ.Ε. 25309

Όνοματεπώνυμο φοιτητή : Στρογγύλης Διονύσιος

Όνοματεπώνυμο εισηγητή Χατζημίσιος Περικλής

Ημερομηνία ανάληψης Δ.Ε. 30/08/2025

Ημερομηνία περάτωσης Δ.Ε. 10/02/2026

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή Στρογγύλη Διονύση που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

«Αφιέρωση»

*Στην σύζυγό μου Ιωάννα,
στα παιδιά μου Γιώργο και Σταματίνα
και στη μνήμη του πατέρα μου.*

Πρόλογος

Η παρούσα διπλωματική εργασία αποτελεί προϊόν προσωπικής επιθυμίας, σε συνδυασμό με την παρότρυνση του επιβλέποντα καθηγητή μου Δρ. Περικλή Χατζημίσιου, πηγάζοντας από την ανάγκη να συνεισφέρω ουσιαστικά στην προστασία του φυσικού πλούτου της χώρας μας. Κατοικώντας μόνιμα στους πρόποδες του Ολύμπου, η βιωματική ανησυχία για τις καταστροφικές συνέπειες των δασικών πυρκαγιών αποτέλεσε το βασικό κίνητρο για να στραφώ στην έρευνα προηγμένων συστημάτων έγκαιρης προειδοποίησης.

Μέσα από την εκπόνηση της μελέτης, αποκόμισα πολλαπλά οφέλη, καθώς κλήθηκα να συνδυάσω την περιβαλλοντική ευαισθησία με την επιστημονική τεκμηρίωση. Η διαδικασία αυτή μου προσέφερε βαθιά εξειδίκευση σε αλγορίθμους Μηχανικής (ML) και Βαθιάς Μάθησης (DL), ενισχύοντας την ικανότητά μου να επιλύω σύνθετα, πραγματικά προβλήματα. Παράλληλα, η ανάπτυξη και βελτιστοποίηση μοντέλων όπως το XGBoost και τα LSTM για τον Ελλαδικό χώρο, λειτούργησε ως εφαλτήριο για την εξέλιξη της ερευνητικής μου σκέψης, αποδεικνύοντας στην πράξη πώς η τεχνολογία μπορεί να υπηρετήσει το κοινωνικό σύνολο.

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου για την εμπιστοσύνη στο πρόσωπό μου και την αποδοχή της πρότασής μου, καθώς και την οικογένειά μου για την αμέριστη συμπαράσταση.

Περίληψη

Η ραγδαία αύξηση της συχνότητας και της έντασης των δασικών πυρκαγιών στη λεκάνη της Μεσογείου, ως απόρροια της κλιματικής αλλαγής, καθιστά επιτακτική την ανάγκη ανάπτυξης προηγμένων συστημάτων έγκαιρης προειδοποίησης. Οι παραδοσιακοί μετεωρολογικοί δείκτες, όπως το Καναδικό Σύστημα Δείκτη Καιρού Πυρκαγιάς (FWI), αν και αποτελούν θεμελιώδη εργαλεία αναφοράς, συχνά αδυνατούν να αποτυπώσουν τις πολύπλοκες, μη γραμμικές αλληλεπιδράσεις μεταξύ των τοπικών μικροκλιματικών συνθηκών και της συσσωρευμένης εδαφικής ξηρασίας. Η παρούσα διπλωματική εργασία θεραπεύει αυτό το κενό, προτείνοντας ένα καινοτόμο, υβριδικό πλαίσιο πρόβλεψης κινδύνου, το οποίο συνθέτει τη φυσική ερμηνευσιμότητα των καθιερωμένων δεικτών με την προγνωστική ικανότητα προηγμένων αρχιτεκτονικών Μηχανικής (ML) και Βαθιάς Μάθησης (DL).

Για την υλοποίηση της έρευνας αξιοποιήθηκε εκτεταμένο σύνολο ιστορικών και προγνωστικών μετεωρολογικών δεδομένων, τα οποία υποβλήθηκαν σε επεξεργασία μέσω τεχνικής κυλιόμενου παραθύρου (sliding window) διάρκειας δέκα ημερών, με στόχο τη βέλτιστη μοντελοποίηση της απομείωσης της εδαφικής υγρασίας. Πραγματοποιήθηκε συγκριτική αξιολόγηση τεσσάρων διακριτών αλγορίθμων: Random Forest (RF), XGBoost, καθώς και των νευρωνικών δικτύων Long Short-Term Memory (LSTM) και Gated Recurrent Units (GRU). Η επίδοση των μοντέλων εξετάστηκε σε δύο λειτουργικά επίπεδα: την ταξινόμηση (classification) για την εκτίμηση πιθανότητας συμβάντος και την παλινδρόμηση (regression) για την πρόβλεψη της τιμής του δείκτη FWI.

Τα εμπειρικά αποτελέσματα κατέδειξαν ότι ο αλγόριθμος XGBoost επέτυχε βέλτιστη απόδοση στην παλινδρόμηση, σημειώνοντας συντελεστή προσδιορισμού (R^2) άνω του 0.90, ενώ το Random Forest επέδειξε την υψηλότερη σταθερότητα στην ταξινόμηση επιπέδων κινδύνου. Βασιζόμενη στα ευρήματα αυτά, σχεδιάστηκε μια Υβριδική Μηχανή Απόφασης (Hybrid Decision Engine). Το σύστημα αυτό προτάσσει τις προβλέψεις της Τεχνητής Νοημοσύνης για μέγιστη ακρίβεια, ενσωματώνοντας παράλληλα ντετερμινιστικές δικλείδες ασφαλείας (failsafes) βασισμένες στον φυσικό υπολογισμό του FWI, διασφαλίζοντας έτσι την επιχειρησιακή αξιοπιστία υπό ακραίες συνθήκες. Το τελικό μοντέλο ενσωματώθηκε σε διαδραστικό περιβάλλον προσομοίωσης, ειδικά βαθμονομημένο (calibrated) για τα ξηροθερμικά χαρακτηριστικά του Ελληνικού καλοκαιριού, παρέχοντας στις αρχές Πολιτικής Προστασίας ένα αξιόπιστο εργαλείο λήψης αποφάσεων.

Λέξεις-Κλειδιά: Πυρκαγιές, Μηχανική Μάθηση, Βαθιά Μάθηση, Δείκτης FWI, XGBoost, Random Forest, LSTM, Υβριδικά Συστήματα, Πολιτική Προστασία.

«Hybrid Wildfire Prediction System Using Machine Learning and Remote Sensing Data»

«Strongylis Dionysios»

Abstract

The escalating frequency and intensity of wildfires in the Mediterranean basin, driven by climate change, necessitates the development of more sophisticated early warning systems. Traditional meteorological indices, such as the Canadian Fire Weather Index (FWI), serve as a fundamental baseline but often lack the capacity to capture complex, non-linear interactions between local microclimatic conditions and cumulative drought stress. This thesis addresses this limitation by proposing a novel, hybrid fire risk prediction framework that synergizes the physical interpretability of established indices with the predictive capabilities of advanced Machine Learning (ML) and Deep Learning (DL) architectures.

The research utilized an extensive dataset of historical and forecast meteorological parameters, processed through a ten-day sliding window technique to effectively model soil moisture depletion and cumulative dryness. A comparative analysis was conducted across four distinct algorithms: Random Forest (RF), XGBoost, Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRU). The models were evaluated on two functional levels: classification, for assessing fire occurrence probability, and regression, for forecasting the specific FWI value.

Empirical results demonstrated that the XGBoost algorithm achieved superior performance in regression tasks, attaining a coefficient of determination (R^2) exceeding 0.90, while Random Forest provided the most robust stability for risk classification. Building upon these findings, a Hybrid Decision Engine was engineered. This system prioritizes AI-driven predictions for high precision but integrates deterministic failsafe mechanisms based on physical FWI calculations to ensure operational reliability during extreme weather events. The final model was deployed into an interactive simulation environment, specifically calibrated to the xerothermic characteristics of the Greek summer, providing Civil Protection authorities with a reliable, data-driven tool for strategic decision-making.

Keywords: Wildfires, Machine Learning, Deep Learning, FWI Index, XGBoost, Random Forest, LSTM, Hybrid Systems, Civil Protection.

Ευχαριστίες

Η ολοκλήρωση της παρούσας διπλωματικής εργασίας δεν ήταν μια μοναχική διαδρομή. Υπήρξαν άνθρωποι που με καθοδήγησαν με τη γνώση τους και άλλοι που με στήριξαν με την αγάπη τους.

Πρώτα απ' όλα, οφείλω ένα μεγάλο ευχαριστώ στον επιβλέποντα καθηγητή μου, Δρ. Περικλή Χατζημίσιο. Τον ευχαριστώ όχι μόνο για την επιστημονική του καθοδήγηση, αλλά και για την εμπιστοσύνη που έδειξε στο πρόσωπό μου, δίνοντάς μου την ελευθερία και τα εφόδια να προχωρήσω.

Ιδιαίτερη μνεία και βαθιά εκτίμηση οφείλω στον υποψήφιο διδάκτορα, Βασίλη Ξεφτέρη. Η βοήθειά του υπήρξε καταλυτική, ειδικά σε εκείνες τις κρίσιμες καμπές της έρευνας όπου τα εμπόδια έμοιαζαν βουνό. Βασίλη, σε ευχαριστώ που με τη βοήθειά σου έκανες τα δύσκολα να φαίνονται απλά.

Πέρα όμως από την ακαδημαϊκή στήριξη, η καρδιά αυτής της προσπάθειας χτυπάει στο σπίτι μου.

Στην Ιωάννα. Τη γυναίκα που μετουσίωσε την κούρασή μου σε ηρεμία και τις αμφιβολίες μου σε πείσμα. Ιωάννα, σε ευχαριστώ που κράτησες τις ισορροπίες της καθημερινότητάς μας αλώβητες, τις ώρες που εγώ ήμουν απών. Χωρίς τη δική σου γενναιοδωρία ψυχής και την αθόρυβη δύναμή σου, η προσπάθεια δεν θα είχε ολοκληρωθεί.

Στον Γιώργο και τη Σταματίνα, τα παιδιά μου. Είστε το μόνιμο αντίδοτο σε κάθε άγχος και η πιο όμορφη υπενθύμιση του τι πραγματικά αξίζει. Ελπίζω μεγαλώνοντας, αν πιάσετε αυτό το βιβλίο στα χέρια σας, να θυμάστε πως κανένας στόχος δεν είναι ανέφικτος όταν υπάρχει αγάπη, υπομονή και ανθρώπινη στήριξη.

Περιεχόμενα

Πρόλογος.....	ii
Περίληψη.....	iii
Abstract	iv
Ευχαριστίες	v
Περιεχόμενα	vi
Κατάλογος Σχημάτων και Εικόνων.....	x
Κατάλογος Κώδικα	xi
Κατάλογος Πινάκων.....	i
Συνομογραφίες.....	ii
Κεφάλαιο 1ο: Εισαγωγή	1
1.1 Γενικό Πλαίσιο και Κίνητρο	1
1.2 Ορισμός του Προβλήματος	1
1.3 Σκοπός της Διπλωματικής Εργασίας.....	2
1.4 Καινοτομία & Συνεισφορά.....	2
1.5 Δομή της Εργασίας.....	2
Κεφάλαιο 2ο: Θεωρητικό Υπόβαθρο.....	4
2.1 Δασικές Πυρκαγιές & Μετεωρολογία.....	4
2.2 Το Σύστημα FWI (Canadian Forest Fire Weather Index)	4
2.2.1.1 Fine Fuel Moisture Code (FFMC)	5
2.2.1.2 Duff Moisture Code (DMC)	7
2.2.1.3 Drought Code (DC).....	9
2.2.1.4 Initial Spread Index (ISI)	10
2.2.1.5 Buildup Index (BUI):.....	10
2.2.1.6 Fire Weather Index (FWI)	11
2.3 Μηχανική Μάθηση (Machine Learning).....	12
2.3.1 Επιβλεπόμενη Μάθηση και Διαδικασία Εκπαίδευσης (Supervised Learning Workflow)	12
2.3.1.1 Προεπεξεργασία Δεδομένων (Data Preprocessing).....	12
2.3.1.2 Διαχωρισμός Δεδομένων (Splitting):.....	12
2.3.1.3 Επικύρωση K-Fold (K-Fold Cross Validation)	12
2.3.1.4 Βελτιστοποίηση Υπερπαραμέτρων (Hyperparameter Tuning).....	14
2.3.1.5 Κατηγορίες Προβλημάτων & Αλγόριθμοι.....	14

2.3.1.6	Ταξινόμηση vs Παλινδρόμηση (Classification vs Regression)	14
2.3.1.7	Δέντρα Αποφάσεων (Decision Trees)	15
2.3.1.8	Random Forest (Bagging).....	15
2.3.1.9	XGBoost (Gradient Boosting)	16
2.4	Βαθιά Μάθηση & Χρονοσειρές (Deep Learning & Time Series).....	17
2.4.1	Recurrent Neural Networks (RNNs)	18
2.4.1.1	Long Short-Term Memory (LSTM)	18
2.4.2	Gated Recurrent Units (GRU)	19
2.5	Μετρικές Αξιολόγησης (Evaluation Metrics)	20
2.5.1	Πίνακας Σύγχυσης (Confusion Matrix).....	20
2.5.2	Μετρικές Ταξινόμησης (Accuracy, Precision, Recall, F1-Score)	21
2.5.3	ROC Curve & AUC.....	21
2.5.4	Μετρικές Παλινδρόμησης (MAE, RMSE, R ²)	22
Κεφάλαιο 3ο:	Φάση 1 - Μηχανική Δεδομένων (Data Engineering).....	23
3.1	Εισαγωγή.....	23
3.2	Συλλογή Δεδομένων και επεξεργασία αρχικών δεδομένων	24
3.2.1	Δεδομένα Ενεργών Πυρκαγιών (Script 0 και Script 1)	24
3.2.1.1	Καθαρισμός δεδομένων αρχικού dataset	26
3.2.1.2	Μέθοδος Χωρικής Ομαδοποίησης.....	27
3.2.2	Συλλογή Ιστορικών Μετεωρολογικών δεδομένων (Script 2).....	28
3.2.2.1	Ορισμός χρονικού παραθύρου ανάκτησης μετεωρολογικών δεδομένων και επιλογή Μετεωρολογικών Παραμέτρων.....	28
3.2.2.2	Διαδικασία Διασύνδεσης και Λήψης Δεδομένων από το Open-Meteo	29
3.2.2.3	Προετοιμασία δεδομένων για υπολογισμό του FWI	30
3.3	Δημιουργία Συνόλου Εκπαίδευσης	32
3.3.1.1	Δημιουργία Αρνητικών Δειγμάτων (Negative Sampling) κλάσεων (Fire vs No-Fire). Περιπτώσεις 1:1, 1:4, 1:10 και 1:200.	32
3.3.2	Υπολογισμός Δεικτών FWI και δημιουργία τελικού Dataset (Script 4).	33
3.3.2.1	Κλάση FWI_Instant_Calculator και Εποχιακή Αρχικοποίηση.....	33
3.3.2.2	Υλοποίηση των Εξισώσεων Van Wagner (CFFDRS)	34
Κεφάλαιο 4ο:	Φάση 2 - Ανάπτυξη & Αξιολόγηση Μοντέλων.....	37
4.1	Διαχωρισμός Προβλήματος.....	37
4.1.1	Ταξινόμηση (Classification Pipeline).....	37
4.1.2	Παλινδρόμηση (Regression Pipeline).....	38
4.1.3	Επιλογή και Εκπαίδευση Αλγορίθμων	38

4.1.4	Δεδομένα Εισόδου (Input Features)	39
4.1.5	Στρατηγικές Προεπεξεργασίας και Κανονικοποίηση (Scaling)	40
4.1.5.1	Μετασχηματισμός Χρονοσειρών (Sliding Window)	40
4.1.5.2	Αποτελέσματα μετασχηματισμού	41
4.1.6	Αποτελέσματα Μετασχηματισμού Δεδομένων	42
4.2	Πείραμα 1: Ταξινόμηση (Classification Pipeline).....	42
4.2.1	Μοντέλα Βαθιάς Μάθησης (Deep Learning): Αρχιτεκτονική LSTM και GRU	43
4.2.1.1	Αποτελέσματα Εκπαίδευσης Αρχιτεκτονικής LSTM.....	46
4.2.1.2	Αποτελέσματα Αρχιτεκτονικής GRU	48
4.2.1.3	Ανάλυση Σημαντικότητας	50
4.2.2	Μοντέλα Μηχανικής Μάθησης (Machine Learning): Εκπαίδευση XGBoost και Random Forest. 52	
4.2.2.1	Random Forest Classifier (Bagging).....	52
4.2.2.2	XGBoost Classifier (Boosting).....	54
4.2.3	Συζήτηση Αποτελεσμάτων και Επιλογή Βέλτιστου Μοντέλου	58
4.2.3.1	Ανάλυση Επιδόσεων	58
4.2.3.2	Επιλογή βέλτιστου μοντέλου	58
4.3	Πείραμα 2: Παλινδρόμηση (Regression Pipeline)	59
4.3.1	Γραμμική Παλινδρόμηση (Linear Regression)	59
4.3.2	XGBoost Regressor	62
4.3.3	Deep Learning Regression (LSTM & GRU).....	65
4.3.3.1	Μοντέλο LSTM Regression.....	65
4.3.3.2	Μοντέλο GRU Regression.....	68
4.3.4	Συζήτηση Αποτελεσμάτων και Επιλογή Βέλτιστου Μοντέλου	71
4.3.4.1	Συγκριτική Ανάλυση Επιδόσεων	71
4.3.4.2	Επιλογή Βέλτιστου Μοντέλου	72
Κεφάλαιο 5ο:	Φάση 3 – Εφαρμογές & Προσομοίωση	73
5.1	Εισαγωγή: Αρχιτεκτονική Συστήματος (Systems Architecture).....	73
5.1.1	Ανάλυση Κινδύνου Δασικών Πυρκαγιών στην Ελλάδα (Fire Data Analysis without AI) 74	
5.1.1.1	Σκοπός και Λειτουργία	74
5.1.1.2	Τεχνική Υλοποίηση & Εργαλεία	74
5.1.2	Τελική Έκθεση Αξιολόγησης Μοντέλων	77
5.1.3	AI Fire Risk Forecast (Σύστημα Διαχείρισης Κινδύνου Πυρκαγιάς).....	82
5.1.4	AI Wildfire Risk Simulation (Live Πρόγνωση Κινδύνου Πυρκαγιών).....	83

5.1.4.1	Διαδραστικός Χάρτης.....	85
5.1.4.2	Επιλογή Λειτουργίας Συστήματος.....	85
5.1.4.3	Παράμετροι Μοντέλου	86
5.1.4.4	Προβολή Αναφορών (Reports).....	87
5.1.4.5	Επίδειξη λειτουργίας Συστήματος AI Wildfire Risk Simulation.....	87
Κεφάλαιο 6ο:	Συμπεράσματα, Περιορισμοί και Μελλοντικές Επεκτάσεις.....	93
6.1	Διερεύνηση της Επίδρασης Ανισορροπίας Κλάσεων.....	93
6.2	Συμπεράσματα.....	96
6.3	Περιορισμοί της Έρευνας.....	96
6.4	Μελλοντικές Επεκτάσεις.....	97
ΒΙΒΛΙΟΓΡΑΦΙΑ.....		98

Κατάλογος Σχημάτων και Εικόνων

Εικόνα 2.1: Δομή και Συνιστώσες του Συστήματος FWI [6] [15].....	5
Εικόνα 2.2: Διαχωρισμός δεδομένων με τη μέθοδο k-Fold Cross-Validation.....	13
Εικόνα 2.3: Αρχιτεκτονική του αλγορίθμου Random Forest.....	16
Εικόνα 2.4: Αρχιτεκτονική του αλγορίθμου XGBoost.....	17
Εικόνα 2.5: Δομή Αλγορίθμου LSTM/GRU [24].....	19
Εικόνα 2.6: Πίνακας Σύγχυσης (Confusion Matrix).....	20
Εικόνα 3.1: Αναλυτικό διάγραμμα ροής εργασιών (Workflow).....	24
Εικόνα 3.2: Ροή επεξεργασίας δεδομένων ενεργών πυρκαγιών.....	28
Εικόνα 3.3: Ροή υπολογισμού δεικτών FWI.....	36
Εικόνα 4.1: Διαχωρισμός προβλήματος.....	37
Εικόνα 4.2: Αρχιτεκτονική πειράματος Ταξινόμησης.....	38
Εικόνα 4.3: Αρχιτεκτονική πειράματος Παλινδρόμησης.....	38
Εικόνα 4.4: Classification.....	42
Εικόνα 4.5: Regression.....	42
Εικόνα 4.6: Καμπύλες Loss και Accuracy κατά την εκπαίδευση του LSTM.....	46
Εικόνα 4.7: Πίνακας Σύγχυσης του μοντέλου LSTM για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire)	47
Εικόνα 4.8: Καμπύλες Loss και Accuracy κατά την εκπαίδευση του GRU ($lr=0.001$).....	48
Εικόνα 4.9: Καμπύλες Loss και Accuracy κατά την εκπαίδευση του GRU ($lr=0.0001$).....	48
Εικόνα 4.10: Πίνακας Σύγχυσης του μοντέλου GRU για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire)	49
Εικόνα 4.11: Features Importance του μοντέλου LSTM.....	51
Εικόνα 4.12: Features Importance του μοντέλου GRU.....	51
Εικόνα 4.13: Πίνακας Σύγχυσης του αλγορίθμου Random Forest για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire).....	53
Εικόνα 4.14: Σημαντικότητα Χαρακτηριστικών στο Random Forest.....	54
Εικόνα 4.15: Καμπύλες Μάθησης (Learning Curves).....	56
Εικόνα 4.16: Πίνακας Σύγχυσης του αλγορίθμου XGBoost για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire).....	56
Εικόνα 4.17: Σημαντικότητα Χαρακτηριστικών στο XGBoost.....	57
Εικόνα 4.18: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI.....	61
Εικόνα 4.19: Σημαντικότητα Χαρακτηριστικών στο Linear Regression.....	62
Εικόνα 4.20: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για XGBoost.....	64
Εικόνα 4.21: Σημαντικότητα Χαρακτηριστικών στο XGBoost.....	65
Εικόνα 4.22: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για LSTM.....	67
Εικόνα 4.23: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για LSTM.....	67
Εικόνα 4.24: Σημαντικότητα Χαρακτηριστικών στο LSTM (Regression).....	68
Εικόνα 4.25: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για GRU.....	70
Εικόνα 4.26: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για GRU.....	70
Εικόνα 4.27: Σημαντικότητα Χαρακτηριστικών στο GRU.....	71
Εικόνα 5.1: Τελικά συστήματα του Project.....	73
Εικόνα 5.2: Πλαϊνή μπάρα επιλογών Χάρτη.....	74
Εικόνα 5.3: Χάρτης Συμβάντων.....	75
Εικόνα 5.4: Σύνολο Συμβάντων ανά Grid.....	76

Εικόνα 5.5: Σημείο πυρκαγιάς	76
Εικόνα 5.6: TOP 50 Μοναδικά Συμβάντα (Highest FWI).....	77
Εικόνα 5.7: Χρονική εξέλιξη περιστατικών πυρκαγιάς (Έτη, Μήνες) και συσχέτιση με τον δείκτη κινδύνου FWI.....	77
Εικόνα 5.8: Αποτελέσματα Classification / Regression και κατάταξη μοντέλων.....	78
Εικόνα 5.9: Καμπύλες εκμάθησης (Μοντέλα DL).....	78
Εικόνα 5.10: Πίνακες Σύγχυσης.....	79
Εικόνα 5.11: Διάγραμμα Διασποράς Παλινδρόμησης	80
Εικόνα 5.12: Ανάλυση Σημαντικότητας Χαρακτηριστικών (Classification Models).....	81
Εικόνα 5.13: Ανάλυση Σημαντικότητας Χαρακτηριστικών (Regression Models).....	82
Εικόνα 5.14: Σύστημα Διαχείρισης Κινδύνου Πυρκαγιάς	83
Εικόνα 5.15: Αρχική οθόνη AI Wildfire Risk Simulator	84
Εικόνα 5.16: Επιλογή σημείου στο Χάρτη.....	85
Εικόνα 5.17: Επιλογές λειτουργίας συστήματος.....	85
Εικόνα 5.18: Εμφάνιση καιρικών δεδομένων για το όρος Όλυμπος (default σημείο).....	86
Εικόνα 5.19: Τροποποίηση τιμών μετεωρολογικών δεικτών.....	86
Εικόνα 5.20: Dropdown Menu (Προβολή Αναφορών).....	87
Εικόνα 5.21: Πρόβλεψη για Χωρίο Πεύκη (περιφέρεια Θεσσαλίας)	88
Εικόνα 5.22: Κύριοι Δείκτες Κινδύνου και Μπάρα Συναγερμού	88
Εικόνα 5.23: Αναφορά Συνθηκών.....	89
Εικόνα 5.24: Πίνακας Κλίμακα Επικινδυνότητας των δεικτών του CFFDRS [49].....	90
Εικόνα 5.25: Ιστορικά Δεδομένα Καιρού - Δεδομένα Εισόδου.....	90
Εικόνα 5.26: Μάτι Αττικής (23/7/2018)	91
Εικόνα 5.27: Δείκτες Σεναρίου - Μάτι Αττικής.....	92

Κατάλογος Κώδικα

Κώδικας 3.1: Επιλεγμένες μεταβλητές αρχικών δεδομένων.....	26
Κώδικας 3.2: Τεχνική Spatial Binning.....	27
Κώδικας 3.3: Ορισμός πρώτης και τελευταίας ημερομηνίας ανάκτησης μετεωρολογικών δεδομένων.....	29
Κώδικας 3.4: Ορισμός μεταβλητών από Open-Meteo Api.....	29
Κώδικας 3.5: Επιλογή μεταβλητών από Open-Meteo Api.....	31
Κώδικας 3.6: Δειγματοληψία Αρνητικών Δειγμάτων.....	32
Κώδικας 3.7: Υπολογισμός του πλήθους των αρνητικών δειγμάτων.	33
Κώδικας 3.8: Στιγμαία προσέγγιση υπολογισμού με βάση την εποχή.....	34
Κώδικας 3.9: Υπολογισμός Δείκτη FFMC (εκδικητικός κώδικας).....	35
Κώδικας 3.10: Υπολογισμός Δείκτη FWI.....	36
Κώδικας 4.1: Μεταβλητές Εισόδου	40
Κώδικας 4.2: Κανονικοποίηση Δεδομένων	41
Κώδικας 4.3: Αρχιτεκτονική LSTM και GRU.....	44
Κώδικας 4.4: Υπολογισμός Βαρών Κλάσεων και Ρύθμιση Μοντέλου.....	44
Κώδικας 4.5: Έλεγχος Εκπαίδευσης (Callbacks).....	45
Κώδικας 4.6: Υλοποίηση Permutation Importance.....	50
Κώδικας 4.7: Παράμετροι Random Forest.....	52

Κώδικας 4.8: Βασικές Ρυθμίσεις XGBoost.....	55
Κώδικας 4.9: Προετοιμασία και Εκπαίδευση Linear Regration	60
Κώδικας 4.10: Ρύθμιση Υπερπαραμέτρων XGBoost Regressor.....	63
Κώδικας 4.11: Αρχιτεκτονική LSTM (Regression Mode).....	66
Κώδικας 4.12: Αρχιτεκτονική GRU (Regression Mode)	69
Κώδικας 5.1: Ιστορικές Πυρκαγιές στην Ελλάδα	91

Κατάλογος Πινάκων

Πίνακας 2.1: Διάρκεια ημέρας ανά μήνα σε ώρες (L_e) [10].....	8
Πίνακας 2.2: συντελεστής διάρκειας ημέρας L_f	9
Πίνακας 3.1: Ενδεικτική μορφή δεδομένων αρχικού συνόλου δεδομένων.....	25
Πίνακας 3.2: Περιγραφή μεταβλητών αρχείου [30] [31]	25
Πίνακας 3.3: Πίνακας με ID των grid cells και οι συντεταγμένες του κέντρου του grid cell.....	27
Πίνακας 3.4: Δομή του Αρχείου Ιστορικών Μετεωρολογικών Δεδομένων ανά Grid Cell.....	31
Πίνακας 3.5: Δείγματα με βάση την αναλογία.....	33
Πίνακας 3.6: Τελικό dataset	36
Πίνακας 4.1: Δείγματα Χαρακτηριστικών Εισόδου (Features).....	39
Πίνακας 4.2: Υπερπαράμετροι Εκπαίδευσης Μοντέλων LSTM/GRU	45
Πίνακας 4.3: Επιδόσεις Μοντέλου LSTM	47
Πίνακας 4.4: Επιδόσεις Μοντέλου GRU	49
Πίνακας 4.5: Επιδόσεις Random Forest.....	53
Πίνακας 4.6: Επιδόσεις XGBoost	57
Πίνακας 4.7: Συγκεντρωτική αξιολόγηση Classification.....	58
Πίνακας 4.8: Επιδόσεις Linear Regration	60
Πίνακας 4.9: Επιδόσεις XGBoost	63
Πίνακας 4.10: Επιδόσεις LSTM Regression	66
Πίνακας 4.11: Επιδόσεις GRU Regression	69
Πίνακας 4.12: Συγκεντρωτική αξιολόγηση Regression	71
Πίνακας 6.1: Αποτελέσματα Περίπτωσης 1:1.....	93
Πίνακας 6.2: Αποτελέσματα Περίπτωσης 1:4.....	94
Πίνακας 6.3: Αποτελέσματα Περίπτωσης 1:10.....	94
Πίνακας 6.4: Αποτελέσματα Περίπτωσης 1:200.....	95

Συντομογραφίες

ARIMA	Autoregressive Integrated Moving Average
AUC	Area Under the Curve
BUI	Buildup Index (Δείκτης Συσσώρευσης)
CFFDRS	Canadian Forest Fire Danger Rating System
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
DC	Drought Code (Κωδικός Ξηρασίας)
DL	Deep Learning (Βαθιά Μάθηση)
DMC	Duff Moisture Code (Κωδικός Υγρασίας Duff)
DT	Decision Trees (Δέντρα Αποφάσεων)
EMS	Emergency Management Service
FFMC	Fine Fuel Moisture Code (Κωδικός Υγρασίας Λεπτών Καυσίμων)
FN	False Negative (Ψευδώς Αρνητικό)
FP	False Positive (Ψευδώς Θετικό)
FWI	Fire Weather Index (Δείκτης Καιρού Πυρκαγιάς)
GRU	Gated Recurrent Unit
IQR	Interquartile Range (Ενδοτεταρτημοριακό Εύρος)
ISI	Initial Spread Index (Δείκτης Αρχικής Εξάπλωσης)
KNN	K-Nearest Neighbors
LR	Linear Regression
LSTM	Long Short-Term Memory
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error (Μέσο Απόλυτο Σφάλμα)
MAPE	Mean Absolute Percentage Error
ML	Machine Learning (Μηχανική Μάθηση)
MODIS	Moderate Resolution Imaging Spectroradiometer (Φασματοραδιόμετρο Απεικόνισης Μέσης Ανάλυσης)
MSE	Mean Squared Error (Μέσο Τετραγωνικό Σφάλμα)
NDVI	Normalized Difference Vegetation Index
NDVI	δείκτη βλάστησης
RF	Random Forest
RH	Relative Humidity (Σχετική Υγρασία)
RMSE	Root Mean Squared Error (Ρίζα Μέσου Τετραγωνικού Σφάλματος)
RNN	Recurrent Neural Networks (Αναδρομικά Νευρωνικά Δίκτυα)
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
T	Temperature (Θερμοκρασία)
TN	True Negative (Σωστός Αρνητικό)
TP	True Positive (Σωστός Θετικό)
UI	User Interface (διεπαφή χρήστη)
W	Wind Speed (Ταχύτητα Ανέμου)
XGBoost	Extreme Gradient Boosting

Κεφάλαιο 1ο: Εισαγωγή

1.1 Γενικό Πλαίσιο και Κίνητρο

Η κλιματική αλλαγή μεταμόρφωσε ριζικά το πρόβλημα των δασικών πυρκαγιών παγκοσμίως και ειδικά στη Μεσόγειο, με την Ελλάδα να βρίσκεται στην πρώτη γραμμή. Οι αυξανόμενες θερμοκρασίες, η μείωση βροχοπτώσεων και οι παρατεταμένοι καύσωνες δημιούργησαν συνθήκες ιδανικές για "mega-fires" – πυρκαγιές που καλύπτουν πάνω από 5.000 στρέμματα και είναι σχεδόν αδύνατο να ελεγχθούν με συμβατικά μέσα. Σύμφωνα με μελέτες του Πανεπιστημίου Πατρών, τέτοια φαινόμενα που συνέβαιναν μία φορά κάθε 5 χρόνια στη δεκαετία του 1990, έχουν γίνει **συχνά ετήσια γεγονότα** τα τελευταία χρόνια.[1]

Η Ελλάδα κατέγραψε πρωτοφανείς καταστροφές: Η φωτιά στο **Μάτι (2018)** με 102 νεκρούς, η πυρκαγιά στην **Εύβοια (2021)** που κατέκαψε 45.000 εκτάρια, και η καταστροφική φωτιά στον **Έβρο-Δαδιά (2023)** με πάνω από 90.000 εκτάρια, επηρεάζοντας πάνω από το 50% του Εθνικού Πάρκου Δαδιάς. Αυτές οι πυρκαγιές δεν περιορίζονται πλέον στην καλοκαιρινή περίοδο Ιούνιο-Σεπτέμβριο, αλλά εμφανίζονται όλο και νωρίτερα (Μάιος) και αργότερα (Νοέμβριος), δυσκολεύοντας σημαντικά το συντονισμό των δυνάμεων πυρόσβεσης.[2]

Στην Ελλάδα, το πρόβλημα είναι ιδιαίτερα έντονο, με εκτεταμένες πυρκαγιές που συμβαίνουν κάθε καλοκαίρι, όπως αποδεικνύεται από τις καταστροφικές πυρκαγιές στην Εύβοια (2021), στην Αττική (2024) και στον Έβρο (2023), όπου κάηκαν περίπου 90.000 εκτάρια, επηρεάζοντας πάνω από το 50% του Εθνικού Πάρκου Δαδιάς[2][3].

Τα παραδοσιακά συστήματα πρόληψης φτάνουν στα όριά τους. Οι πυροσβεστικές ζώνες, τα εναέρια μέσα και οι μετεωρολογικές προβλέψεις από μόνες τους δεν επαρκούν πλέον για να αντιμετωπίσουν την **κλιματική επιθετικότητα** των mega-fires. Η ανάγκη για **προληπτική πρόβλεψη υψηλής ακρίβειας 24-48 ώρες νωρίτερα** έχει γίνει επιτακτική, ιδιαίτερα σε μια χώρα με περιορισμένους πόρους και εκτεταμένες δασικές εκτάσεις.[3]

Η Τεχνητή Νοημοσύνη και η Μηχανική Μάθηση προσφέρουν πρωτοποριακή λύση. Σε αντίθεση με τους παραδοσιακούς μετεωρολόγους, τα προηγμένα αλγόριθμοι μπορούν να [4]:

- Επεξεργάζονται **τεράστιους όγκους δεδομένων** (big data) από δορυφόρους, μετεωρολογικούς σταθμούς και δείκτες FWI σε πραγματικό χρόνο.
- **Μαθαίνουν από το ιστορικό** δεκαετιών πυρκαγιών, αναγνωρίζοντας μοτίβα που διαφεύγουν της ανθρώπινης αντίληψης.
- **Προβλέπουν δυναμικά**, προσαρμόζοντας τις προβλέψεις τους σε πραγματικό χρόνο καθώς εξελίσσονται οι καιρικές συνθήκες.

Η παρούσα εργασία γεφυρώνει αυτό το κενό, συνδυάζοντας εμπειρική μετεωρολογική γνώση (FWI) με τις δυνατότητες της TN για να δημιουργήσει ένα **υβριδικό σύστημα πρόβλεψης** που θα επιτρέψει στην Πολιτική Προστασία να δράσει **προληπτικά** και **στοχευμένα**, σώζοντας ζωές, περιουσίες και φυσικό περιβάλλον.

1.2 Ορισμός του Προβλήματος

Παρόλο που υπάρχουν καθιερωμένα συστήματα εκτίμησης κινδύνου, όπως το Καναδικό Σύστημα FWI (Fire Weather Index), αυτά παρουσιάζουν εγγενείς περιορισμούς. Οι κλασικοί δείκτες είναι εκ φύσεως **στατικοί**: υπολογίζουν τον κίνδυνο βασιζόμενοι μόνο στις τρέχουσες συνθήκες, χωρίς να έχουν τη

δυνατότητα να «μάθουν» από το ιστορικό παρελθόν ή να προσαρμοστούν στα ιδιαίτερα χαρακτηριστικά κάθε μικροκλίματος.

Επιπλέον, η πρόβλεψη πυρκαγιών αποτελεί ένα εξαιρετικά δύσκολο υπολογιστικό πρόβλημα λόγω της φύσης των δεδομένων:

- **Ανισοβαρής Κατανομή (Class Imbalance):** Τα δεδομένα χαρακτηρίζονται από τεράστια ανισορροπία, καθώς οι ημέρες με πυρκαγιά είναι ελάχιστες συγκριτικά με τις ημέρες ηρεμίας, δυσκολεύοντας την εκπαίδευση των αλγορίθμων.
- **Πολυπλοκότητα:** Η αλληλεπίδραση μετεωρολογικών παραγόντων (θερμοκρασία, υγρασία, άνεμος) είναι χαοτική και μη γραμμική.

1.3 Σκοπός της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία στοχεύει στην ανάπτυξη ενός ολοκληρωμένου συστήματος πρόγνωσης που υπερβαίνει τους παραπάνω περιορισμούς. Οι επιμέρους στόχοι της έρευνας είναι:

- **Συλλογή και Επεξεργασία Δεδομένων:** Η δημιουργία ενός αξιόπιστου συνόλου δεδομένων που συνδυάζει ιστορικά μετεωρολογικά στοιχεία και καταγραφές πυρκαγιών από την ελληνική επικράτεια.
- **Ανάπτυξη και Σύγκριση Μοντέλων:** Η εκπαίδευση και αξιολόγηση διαφορετικών αρχιτεκτονικών, από αλγορίθμους Μηχανικής Μάθησης (Random Forest, XGBoost) έως προηγμένα δίκτυα Βαθιάς Μάθησης (LSTM, GRU).
- **Δημιουργία Υβριδικού Συστήματος (Hybrid System):** Η ανάπτυξη μιας καινοτόμου μεθοδολογίας που συνθέτει την εμπειρική εγκυρότητα του δείκτη FWI με την προβλεπτική ισχύ του AI.
- **Επιχειρησιακή Εφαρμογή:** Η υλοποίηση μιας διαδικτυακής πλατφόρμας (Dashboard) που επιτρέπει την προσομοίωση και την πρόγνωση κινδύνου σε πραγματικό χρόνο.

1.4 Καινοτομία & Συνεισφορά

Η βασική καινοτομία της εργασίας έγκειται στην εισαγωγή της **Υβριδικής Λογικής**. Σε αντίθεση με προσεγγίσεις που βασίζονται τυφλά στα μοντέλα Μηχανικής και Βαθιάς Μάθησης, το σύστημα που αναπτύξαμε ενσωματώνει δικλείδες ασφαλείας: όταν οι φυσικοί δείκτες (FWI) υποδεικνύουν ακραίο κίνδυνο, το σύστημα έχει τη δυνατότητα να παρακάμψει τυχόν λανθασμένες προβλέψεις του νευρωνικού δικτύου, διασφαλίζοντας την επιχειρησιακή αξιοπιστία.

Επιπρόσθετα, η χρήση δικτύων LSTM/GRU με χρονικό παράθυρο (Lookback window) εισάγει την έννοια της "μνήμης" στο σύστημα, επιτρέποντας την κατανόηση της συσσώρευσης ξηρασίας σε βάθος χρόνου, κάτι που απουσιάζει από τις στατικές προσεγγίσεις.

1.5 Δομή της Εργασίας

Η παρούσα διπλωματική εργασία οργανώθηκε σε έξι κεφάλαια, τα οποία περιγράφονται συνοπτικά στη συνέχεια:

- Στο **Κεφάλαιο 2** αναλύεται το θεωρητικό υπόβαθρο γύρω από τους δείκτες FWI, τις αρχιτεκτονικές Μηχανικής (ML) και Βαθιάς (DL) Μάθησης, καθώς τις έννοιες των Μετρικών Αξιολόγησης.
- Στο **Κεφάλαιο 3** παρουσιάζεται η μεθοδολογία προ-επεξεργασίας των δεδομένων, η άντληση ιστορικών μετεωρολογικών δεδομένων από το API της Open-Meteo, δημιουργία Grid Cells και negative samples (No-Fire), η δημιουργία του FWI Generator και η εξαγωγή ενός big dataset.
- Στο **Κεφάλαιο 4** περιγράφεται η εκπαίδευση των μοντέλων και η αξιολόγησή τους.

- Στο **Κεφάλαιο 5** αναλύεται η υλοποίηση του Υβριδικού Συστήματος και η λειτουργία του Dashboard, καθώς και η λειτουργία των υπο-συστημάτων αυτού.
- Τέλος, στο **Κεφάλαιο 6** συγκρίνουμε τα αποτελέσματα των πειραματικών δοκιμών, συνοψίζονται τα συμπεράσματα και προτείνονται κατευθύνσεις για μελλοντική έρευνα.

Κεφάλαιο 2ο: Θεωρητικό Υπόβαθρο

2.1 Δασικές Πυρκαγιές & Μετεωρολογία

Οι δασικές πυρκαγιές αποτελούν πολύπλοκα φαινόμενα που καθοδηγούνται από έναν συνδυασμό ανθρωπογενών και φυσικών παραγόντων και, εάν δεν αντιμετωπιστούν έγκαιρα, μπορούν να αποτελέσουν σοβαρές απειλές για τα οικοσυστήματα, τις υποδομές και την ανθρώπινη ζωή [5] [6] [20] [27]. Η ανάφλεξη και η εξάπλωσή τους διέπονται από την αλληλεπίδραση των μετεωρολογικών συνθηκών, των χαρακτηριστικών της καύσιμης ύλης και των τοπογραφικών χαρακτηριστικών, καθιστώντας τη συμπεριφορά τους δύσκολο να προβλεφθεί και να ελεγχθεί [5] [6] [16]. Παρόλο που η ανθρώπινη δραστηριότητα αποτελεί κύριο οδηγό, ιδιαίτερα στη Μεσόγειο όπου η καύση βιομάζας ανθρωπογενούς προέλευσης ευθύνεται για το 90–95% των περιστατικών, η κλιματική αλλαγή έχει αναδειχθεί τις τελευταίες δεκαετίες ως σημαντικός συντελεστής στην αυξανόμενη συχνότητα και ένταση των πυρκαγιών παγκοσμίως. Οι μακροπρόθεσμες κλιματικές τάσεις, όπως η άνοδος της θερμοκρασίας, η παρατεταμένη ξηρασία και η αλλαγή στα πρότυπα βροχοπτώσεων, δημιουργούν συνθήκες που ευνοούν την ανάφλεξη και την ταχεία εξάπλωση της φωτιάς [6] [19] [23] [24].

Συγκεκριμένα, οι μετεωρολογικές μεταβλητές παίζουν καθοριστικό ρόλο στη διαμόρφωση του κινδύνου πυρκαγιάς. Οι υψηλότερες θερμοκρασίες αυξάνουν τους ρυθμούς εξάτμισης και ξηραίνουν την καύσιμη ύλη, αυξάνοντας έτσι την πιθανότητα ανάφλεξης [5] [29]. Τα χαμηλότερα επίπεδα σχετικής υγρασίας μειώνουν την περιεκτικότητα σε υγρασία της βλάστησης, καθιστώντας την πιο ευάλωτη στη φωτιά, ενώ οι ισχυροί άνεμοι επιταχύνουν την εξάπλωση και επηρεάζουν σημαντικά τη συμπεριφορά του μετάπου της πυρκαγιάς [5] [6]. Επιπρόσθετα, οι πρόσφατες βροχοπτώσεις επηρεάζουν την περιεκτικότητα σε υγρασία των καυσίμων, ενώ παρατεταμένες ξηρές περιόδους αυξάνουν δραματικά τον κίνδυνο [5] [29]. Στην Ελλάδα, για παράδειγμα, η κλιματολογική ανάλυση για την περίοδο 1965–2023 αποκαλύπτει μια σταθερή στροφή προς θερμότερες και ξηρότερες συνθήκες, ιδιαίτερα κατά τους μήνες Ιούλιο και Αύγουστο, με αυξανόμενες ανωμαλίες θερμοκρασίας και επίμονα χαμηλή σχετική υγρασία. Αυτές οι αλλαγές ενισχύουν την ξηρότητα των καυσίμων και το δυναμικό ανάφλεξης, ευθυγραμμίζοντας τα εποχιακά κλιματικά πρότυπα με συνθήκες που ευνοούν την ανάπτυξη πυρκαγιών, ενώ ταυτόχρονα η βροχόπτωση παραμένει εξαιρετικά μεταβλητή χωρίς σαφή τάση. Και γενικότερα στη νοτιοανατολική Ευρώπη, η άνοδος της θερμοκρασίας κατά 2.5°C έως 4.4°C προβλέπεται να αυξήσει τη συχνότητα των πυρκαγιών κατά 14% έως 30% [5] [6].

2.2 Το Σύστημα FWI (Canadian Forest Fire Weather Index)

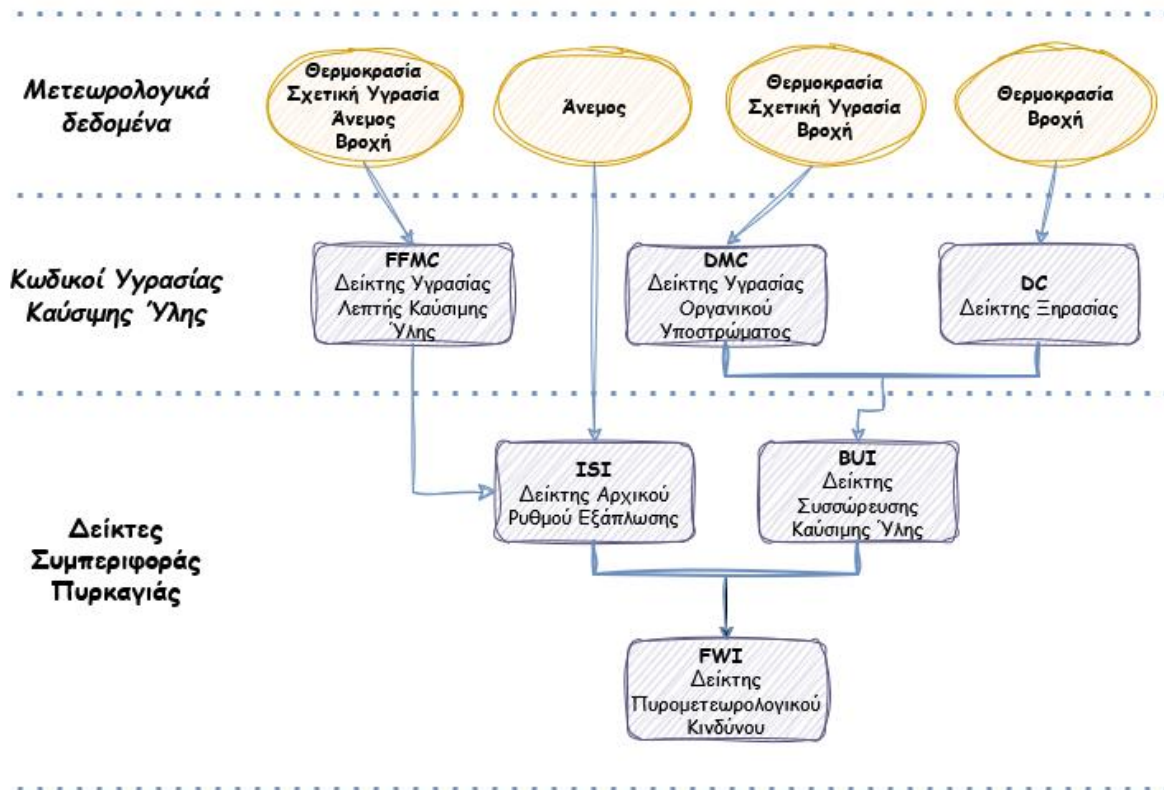
Το Καναδικό Σύστημα Δείκτη Καιρού Δασικών Πυρκαγιών (FWI) αποτελεί ένα από τα πιο διαδεδομένα και επιχειρησιακά επικυρωμένα εργαλεία παγκοσμίως για την εκτίμηση του κινδύνου πυρκαγιάς. Αν και αναπτύχθηκε αρχικά για τα δασικά οικοσυστήματα του Καναδά (**Van Wagner, 1987**), έχει υιοθετηθεί ευρέως σε χώρες της Μεσογείου λόγω της ικανότητάς του να αποτυπώνει τις πυρο-μετεωρολογικές συνθήκες που επικρατούν σε ξηρά και θερμά κλίματα. Πρόκειται για ένα εμπειρικό σύστημα που βασίζεται αποκλειστικά σε μετεωρολογικές παρατηρήσεις για να ποσοτικοποιήσει την πιθανότητα εκδήλωσης και τη συμπεριφορά της φωτιάς [5-7] [15].

Λεδομένα Εισόδου (Input Data) Για τον υπολογισμό των δεικτών του συστήματος απαιτούνται τέσσερις βασικές μετεωρολογικές μεταβλητές, οι οποίες λαμβάνονται καθημερινά το μεσημέρι (συνήθως 12:00 UTC):

- **Θερμοκρασία αέρα (Temperature):** Επηρεάζει την ξήρανση της καύσιμης ύλης [5]
- **Σχετική Υγρασία (Relative Humidity):** Καθορίζει την υγρασία της νεκρής βλάστησης [5] [6]

- **Ταχύτητα Ανέμου (Wind Speed):** Επηρεάζει την εξάπλωση της φωτιάς και την ξήρανση [5]
- **Βροχόπτωση 24ώρου (Rainfall):** Καθορίζει την περιεκτικότητα σε υγρασία των καυσίμων [5] [2]

Δομή και Συνιστώσες του Συστήματος Το σύστημα FWI αποτελείται από έξι (6) επιμέρους δείκτες, οι οποίοι χωρίζονται σε δύο κατηγορίες: τους Κώδικες Υγρασίας Καυσίμων (Fuel Moisture Codes) και τους Δείκτες Συμπεριφοράς Πυρκαγιάς (Fire Behavior Indices) (**Εικόνα 2.1**) [6] [8]:



Εικόνα 2.1: Δομή και Συνιστώσες του Συστήματος FWI [6] [15]

Κώδικες Υγρασίας Καυσίμων (Fuel Moisture Codes)

Αυτοί οι δείκτες παρακολουθούν την περιεκτικότητα σε υγρασία διαφορετικών στρωμάτων της δασικής καύσιμης ύλης και διαθέτουν «μνήμη» (time lag), δηλαδή επηρεάζονται από τις καιρικές συνθήκες των προηγούμενων ημερών:

2.2.1.1 Fine Fuel Moisture Code (FFMC)

Αντιπροσωπεύει την υγρασία της λεπτής, νεκρής καύσιμης ύλης (π.χ. πευκοβελόνες, ξερά χόρτα) στην επιφάνεια του εδάφους. Έχει πολύ μικρή χρονική υστέρηση (μόλις 16 ώρες), που σημαίνει ότι αντιδρά άμεσα στις αλλαγές του καιρού (π.χ. βροχή ή αέρας) [6]. Υπολογίζεται ημερήσια, με είσοδο T_{12} , H_{12} , U_{12} και βροχόπτωση P [9]. Παρακάτω αναλυτικά θα δείξουμε τις μαθηματικές σχέσεις ώστε να βγει το τελικό αποτέλεσμα.

Εξισώσεις και βήματα υπολογισμού του FFMC:

1. **Αρχικοποίηση** με την τιμή της προηγούμενης μέρας $FFMC_{t-1}$ [9]

2. **Μετατροπή FFMC σε υγρασία λεπτής ύλης m_{t-1}** (Εξίσωση 2.2) [9]:

$$m_{t-1} = 147.2 \cdot \frac{101 - FFMC_{t-1}}{59.5 + FFMC_{t-1}} \quad (2.1)$$

3. **Φάση βροχής** (αν $P > 0.5\text{mm}$) [9]:

Effective rainfall (Εξίσωση 2.2):

$$P_f = P - 0.5, \text{για } P > 0.5 \quad (2.2)$$

Υγρασία μετά τη βροχή m_{rt} :

- Για $m_{t-1} \leq 150$ (Εξίσωση 2.3):

$$m_{rt} = m_{t-1} + 42.5 \cdot P_f \cdot \left(\frac{-100}{e^{251-m_{t-1}}} \right) \cdot \left(1 - e^{\frac{-6.93}{P_f}} \right) \quad (2.3)$$

- Για $m_{t-1} > 150$ (Εξίσωση 2.4):

$$m_{rt} = m_{t-1} + 42.5 \cdot P_f \cdot \left(\frac{-100}{e^{251-m_{t-1}}} \right) \cdot \left(1 - e^{\frac{-6.93}{P_f}} \right) + 0.0015 \cdot ((m_{t-1} - 150)^2 \cdot P_f^{0.5}) \quad (2.4)$$

Αν $m_{rt} > 250$, τότε θέτουμε $m_{rt} = 250$ [9]

Αν $P \leq 0.5\text{mm}$, παραλείπεται η βροχή και m_{t-1} παραμένει ως έχει.

4. **Υπολογισμός καμπύλης ισορροπίας για drying** [9]:

$$E_d = 0.942 \cdot H_{12}^{0.679} + 11 \cdot e^{\frac{H_{12}-100}{10}} + 0.18 \cdot (21.1 - T_{12}) \cdot (1 - e^{-0.115 \cdot H_{12}}) \quad (2.5)$$

- Αν $E_d < m_{t-1} \Rightarrow$ φάση drying:

Υπολογισμός log drying rate k_d :

$$k_0 = 0.424 \cdot \left(1 - \left(\frac{H_{12}}{100} \right)^{1.7} \right) + 0.0694 \cdot U_{12}^{0.5} \cdot \left(1 - \left(\frac{H_{12}}{100} \right)^8 \right) \quad (2.6)$$

$$k_d = k_0 \cdot 0.581 \cdot e^{0.0365 \cdot T_{12}} \quad (2.7)$$

Νέα υγρασία:

$$m_t = E_d + (m_{t-1} - E_d) \cdot 10^{-k_d} \quad (2.8)$$

- Αν $E_d \geq m_{t-1} \Rightarrow$ φάση wetting:

Ισορροπία υγρής φάσης:

$$E_w = 0.618 \cdot H_{12}^{0.753} + 10 \cdot e^{\frac{H_{12}-100}{10}} + 0.18 \cdot (21.1 - T_{12}) \cdot (1 - e^{-0.115 \cdot H_{12}}) \quad (2.9)$$

Log wetting rate:

$$k_1 = 0.424 \cdot \left(1 - \left(\frac{100 - H_{12}}{100}\right)^{1.7}\right) + 0.0694 \cdot U_{12}^{0.5} \cdot \left(1 - \left(\frac{100 - H_{12}}{100}\right)^8\right) \quad (2.10)$$

$$k_w = k_1 \cdot 0.581 \cdot e^{0.0365 \cdot T_{12}} \quad (2.11)$$

Νέα υγρασία:

$$m_t = E_w - (E_w - m_{t-1}) \cdot 10^{-k_w} \quad (2.12)$$

Αν $E_w \leq m_{t-1} \leq E_d$, τότε $m_t = m_{t-1}$ [9].

5. Μετατροπή πίσω σε δείκτη FFMC (Εξίσωση 2.13) [9]

$$FFMC_t = 59.5 \cdot \frac{250 - m_t}{147.2 + m_t} \quad (2.13)$$

Συμπέρασμα: Ο FFMC χρησιμοποιεί διπλό μηχανισμό (wetting/drying) με ξεχωριστά ισορροπα υγρασίας E_d, E_w και λογαριθμικούς συντελεστές ταχύτητας στεγνώματος. **Έξυπνη λεπτομέρεια:** αγνοεί ελαφρύ νετό <0.5mm και περιορίζει την υγρασία 0-250. **Ρεαλιστικός αλγόριθμος** που μιμείται φυσικά τη φυσιολογία λεπτής καύσιμης ύλης [6] [10].

2.2.1.2 Duff Moisture Code (DMC)

Αφορά την υγρασία των οργανικών υλικών μέτριου βάθους. Έχει χρονική υστέρηση 12 ημερών, αντικατοπτρίζοντας μεσοπρόθεσμες συνθήκες ξηρασίας [6]. Εξαρτάται από $T_{12}, H_{12},$ βροχή P [10].

Εξισώσεις και βήματα υπολογισμού του FFMC:

1. **Αρχικοποίηση** με την τιμή της προηγούμενης μέρας DMC_{t-1} [10]
2. **Φάση βροχής** (αν $P > 1.5\text{mm}$) [10]:

Effective rainfall:

$$P_e = 0.92 \cdot P - 1.27, \text{για } P > 1.5 \quad (2.14)$$

Υγρασία duff πριν τη βροχή:

$$M_{t-1} = 20 + e^{5.6348 - \frac{DMC_{t-1}}{43.43}} \quad (2.15)$$

Παράμετρος κλίσης b :

$$b = \begin{cases} \frac{100}{0.5 + 0.3 \cdot DMC_{t-1}}, & DMC_{t-1} \leq 33 \\ 14 - 1.3 \cdot \ln(DMC_{t-1}), & 33 < DMC_{t-1} \leq 65 \\ 6.2 \cdot \ln(DMC_{t-1}) - 17.2, & DMC_{t-1} > 65 \end{cases} \quad (2.16)$$

Υγρασία μετά τη βροχή:

$$M_{r_t} = M_{t-1} + \frac{1000 \cdot P_e}{48.77 + b \cdot P_e}$$

Μετατροπή σε DMC:

$$DMC_{r_t} = 244.72 - 43.43 \cdot \ln(M_{r_t} - 20) \quad (2.17)$$

Αν $DMC_{r_t} < 0$, θέτουμε $DMC_{r_t} = 0$ [10]

Αν $P \leq 1.5\text{mm}$, η βροχή παραλείπεται [10]

3. Ξήρανση (log drying rate K) [10]:

$$K = 1.894 \cdot (T_{12} + 1.1) \cdot (100 - H_{12}) \cdot L_e \cdot 10^{-6} \quad (2.18)$$

Αν $T_{12} < -1.1$, θέτουμε $T_{12} = -1.1$ [10]

Ο L_e είναι «effective day length» ανά μήνα (πίνακας στο WikiFire / Van Wagner). [wikifire.wsl](http://wikifire.wsl.ch/)

Πίνακας 2.1: Διάρκεια ημέρας ανά μήνα σε ώρες (L_e) [10]

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
L_e	6.5	7.5	9.0	12.8	13.9	13.9	12.4	10.9	9.4	8.0	7.0	6.0

4. Τελικός DMC [10]:

$$DMC_t = \begin{cases} DMC_{t-1} + 100 \cdot K, & P \leq 1.5 \\ DMC_{r_t} + 100 \cdot K, & P > 1.5 \end{cases} \quad (2.19)$$

Ο DMC (Duff Moisture Code) μετρά υγρασία σε βαθύτερα οργανικά στρώματα (duff ~5kg/m²) με διπλή φάση βρεξίματος/στεγνώματος. [10]

Κλειδί: Εξαρτάται από μήνα (ημερήσια διάρκεια L_e), αγνοεί ελαφρύς υετό $\leq 1.5\text{mm}$, **κομμάτι-σοφός** slope b για βροχή, **λογαριθμική** μετατροπή $M \leftrightarrow DMC$. [10]

Ρεαλιστικός για μεσαίες duff & ξυλώδη υλικά μεσημεριανές ώρες. [10]

2.2.1.3 Drought Code (DC)

Αντιπροσωπεύει την υγρασία των βαθύτερων στρωμάτων του εδάφους (οργανικά υλικά σε βάθος). Έχει τη μεγαλύτερη χρονική υστέρηση και χρησιμοποιείται για την παρακολούθηση της μακροχρόνιας ξηρασίας [6] Εξαρτάτε από T_{12} , βροχή P , μήνας [11].

- Βροχή (αν $P > 2.8$ mm):

$$P_d = 0.83P - 1.27 \quad (2.20)$$

$$Q_{t-1} = 800e^{-\frac{DC_{t-1}}{400}} \quad (2.21)$$

$$Q_{r_t} = Q_{t-1} + 3.937P_d \quad (2.22)$$

$$DC_{r_t} = 400 * \ln\left(\frac{800}{Q_{r_t}}\right), \quad \text{if } DC_{r_t} < 0, \text{ then } DC_{r_t} = 0 \quad (2.23)$$

Βροχή → νερό στο έδαφος → μειώνεται ξηρασία (DC) [11]

- Drying factor [11]:

$$V = 0.36(T_{12} + 2.8) + L_f, \quad (\text{αν } T_{12} < -2.8 \rightarrow -2.8, V \geq 0) \quad (2.24)$$

Υπολογίζεται drying factor ως συνάρτηση θερμοκρασίας και daylength με βάση τον Πίνακα 2.2

Πίνακας 2.2: συντελεστής διάρκειας ημέρας L_f

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
L_e	-1.6	-1.6	-1.6	0.9	3.8	5.8	6.4	5.0	2.4	0.4	-1.6	-1.6

- Τελικός DC [11]:

$$DC_t = \begin{cases} DC_{t-1} + 0.5V & P \leq 2.8 \\ DC_{r_t} + 0.5V & P > 2.8 \end{cases} \quad (2.25)$$

Όπου V είναι ο ρυθμός ξήρανσης.

Ο DC προσομοιώνει εποχική ξηρασία σε βαθιά duff στρώματα ($\sim 25\text{kg/m}^2$) με **διπλή φάση** βροχής/εξάτμισης. **Έξυπνα:** Εκθετική μετατροπή σε «υδατικό ισοδύναμο» Q , μηνιαίος παράγοντας

ηλίου L_f , threshold 2.8mm βροχής. **Ιδανικός για μεγάλες πυρκαγιές** βαθέων καυσίμων στις μεσημεριανές ώρες με ηλιοφάνεια [11].

Δείκτες Συμπεριφοράς Πυρκαγιάς (Fire Behavior Indices) Οι δείκτες αυτοί προκύπτουν από τον συνδυασμό των κωδικών υγρασίας και του ανέμου [6].

2.2.1.4 Initial Spread Index (ISI)

Συνδυάζει τον κώδικα FFMC (λεπτά καύσιμα) και την ταχύτητα του ανέμου για να εκτιμήσει την αναμενόμενη ταχύτητα εξάπλωσης της φωτιάς (Rate of Spread) αμέσως μετά την ανάφλεξη. Εξαρτάται από την ταχύτητα ανέμου στις 12:00 (U_{12} σε km/h) και του FFMC [6] [8]. Υπολογίζει:

- Συνάρτηση ανέμου $f(U)$ [12]:

$$f(U) = e^{0.05039 \cdot U_{12}} \quad (2.26)$$

- Συνάρτηση υγρασίας $f(F)$ [12]:

$$f(F) = (91.9 \cdot e^{-0.01386 \cdot m}) \cdot \left(1 + \frac{m^{5.31}}{4.93 \cdot 10^7}\right) \quad (2.27)$$

- Υγρασία λεπτής ύλης (%) [12]:

$$m = 147.2 \cdot \frac{101 - FFMC}{59.5 + FFMC} \quad (2.28)$$

όπου m είναι η περιεκτικότητα σε υγρασία καυσίμου [%].

- Τελικός υπολογισμός ISI [12]:

$$ISI = 0.208 \cdot f(U) \cdot f(F) \quad (2.29)$$

Με απλά λόγια, Άνεμος \times Στεγνή λεπτής ύλης = αρχική ταχύτητα εξάπλωσης πυρκαγιάς.

2.2.1.5 Buildup Index (BUI):

Συνδυάζει τους κώδικες DMC και DC για να μετρήσει τη συνολική διαθέσιμη καύσιμη ύλη που μπορεί να καταναλωθεί από τη φωτιά. Είναι δείκτης της θερμότητας που θα παραχθεί [6] [8] [13].

- $Av \text{ DMC} \leq 0.4 \cdot DC$ [9]:

$$BUI = 0.8 \cdot \frac{DMC \cdot DC}{DMC + 0.4 \cdot DC} \quad (2.30)$$

- Αν $DMC > 0.4 \cdot DC$ [13]:

$$BUI = DMC - \left(1 - \frac{0.8 \cdot DC}{DMC + 0.4 \cdot DC}\right) \cdot [0.92 + (0.0114 \cdot DMC)^{1.7}] \quad (2.31)$$

Ο συνδιασμός DMC και DC μας αποτυπώνει την διαθέσιμη καύσιμη ύλη για να έχει διάρκεια η πυρκαγιά [13].

2.2.1.6 Fire Weather Index (FWI)

Ο τελικός δείκτης που συνδυάζει το ISI και το BUI. Αντιπροσωπεύει τη γενική ένταση της φωτιάς ως έναν μοναδικό αριθμό. Όσο υψηλότερη είναι η τιμή του, τόσο πιο ευνοϊκές είναι οι συνθήκες για την εκδήλωση και εξάπλωση πυρκαγιάς [6] [8] [14].

- Μετατροπή του BUI σε βάρος καμένου καυσίμου [14]

$$f(D) = \begin{cases} 0.626 \cdot BUI^{0.809} + 2, & BUI \leq 80 \\ \frac{1000}{25 + 108.64 \cdot e^{-0.023 \cdot BUI}}, & BUI > 80 \end{cases} \quad (2.32)$$

- B-scale FWI (ενδιάμεση κλίμακα)

$$B = 0.1 \cdot ISI \cdot f(D) \quad (2.33)$$

- S-scale FWI (τελική κλίμακα FWI)

$$S = \begin{cases} e^{2.72 \cdot (0.434 \cdot \ln B)^{0.647}}, & B > 1 \\ B, & B \leq 1 \end{cases} \quad (2.34)$$

Στη βιβλιογραφία, το S είναι το «FWI» που χρησιμοποιούμε συνήθως ως τελική τιμή [14].

Οι τιμές του FWI κατηγοριοποιούνται σε επίπεδα κινδύνου: πολύ χαμηλός (< 5.2), χαμηλός (5.2–11.2), μέτριος (11.2–21.3), υψηλός (21.3–38.0), πολύ υψηλός (38.0–50.0) και ακραίος (> 50.0) [5].

Υπολογιστική Διαδικασία: Ο υπολογισμός των δεικτών γίνεται αλυσιδωτά. Οι κώδικες υγρασίας (FFMC, DMC, DC) υπολογίζονται πρώτοι, χρησιμοποιώντας τις τιμές της προηγούμενης ημέρας ως σημείο εκκίνησης, γεγονός που καθιστά το σύστημα δυναμικό [8]. Για παράδειγμα, οι τυπικές τιμές εκκίνησης κατά την έναρξη της αντιπυρικής περιόδου είναι FFMC=85, DMC=6 και DC=15. Στη συνέχεια, τα αποτελέσματα αυτά τροφοδοτούν τις εξισώσεις για τους δείκτες ISI και BUI, οι οποίοι τελικά δίνουν την τιμή του FWI [8].

2.3 Μηχανική Μάθηση (Machine Learning)

Η Μηχανική Μάθηση (ML) έχει εξελιχθεί σε κρίσιμο εργαλείο για την πρόβλεψη δασικών πυρκαγιών, προσφέροντας λύσεις εκεί όπου τα παραδοσιακά φυσικά και μαθηματικά μοντέλα προσομοίωσης αντιμετωπίζουν δυσκολίες λόγω κόστους, πολυπλοκότητας και χρόνου εκτέλεσης. Οι αλγόριθμοι ML έχουν την ικανότητα να ανακαλύπτουν μοτίβα σε τεράστιους όγκους δεδομένων, μια διαδικασία γνωστή ως Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων (Knowledge Discovery in Database - KDD), βελτιώνοντας σημαντικά τη λήψη αποφάσεων [6]. Σε σύγκριση με τις παραδοσιακές στατιστικές μεθοδολογίες, οι τεχνικές ML έχουν αποδείξει ότι μπορούν να προσφέρουν ανώτερα αποτελέσματα στη γεωχωρική πρόβλεψη της ευπάθεια σε πυρκαγιές [16]. Η διαδικασία περιλαμβάνει τη συλλογή δεδομένων από μετεωρολογικούς σταθμούς και δορυφόρους, τα οποία περιλαμβάνουν κρίσιμες μεταβλητές όπως η θερμοκρασία, η σχετική υγρασία, ο άνεμος και η βροχόπτωση, καθώς αυτοί οι παράγοντες επηρεάζουν άμεσα την ανάφλεξη και την εξάπλωση της φωτιάς [6].

2.3.1 Επιβλεπόμενη Μάθηση και Διαδικασία Εκπαίδευσης (Supervised Learning Workflow)

Στην παρούσα εργασία δουλέψαμε με την μέθοδο της επιβλεπόμενης μάθησης. Στην επιβλεπόμενη μάθηση, τα μοντέλα εκπαιδεύονται σε δεδομένα που περιέχουν τόσο τα χαρακτηριστικά εισόδου (features) όσο και την επιθυμητή έξοδο (target), όπως το αν συνέβη πυρκαγιά ή την επιθυμητή τιμή εξόδου (στην προκυμμένη περίπτωση, την τιμή του FWI). Για την επίτευξη υψηλής ακρίβειας, ακολουθείται μια αυστηρή μεθοδολογία:

2.3.1.1 Προεπεξεργασία Δεδομένων (Data Preprocessing)

Τα ακατέργαστα δεδομένα συχνά απαιτούν καθαρισμό και μετασχηματισμό. Τεχνικές όπως η κωδικοποίηση ετικετών (Label Encoding) χρησιμοποιούνται για τη μετατροπή κατηγορικών δεδομένων (π.χ. μήνες, ημέρες) σε αριθμητικές τιμές, ώστε να είναι κατανοητά από τις μηχανές. Επίσης, εφαρμόζεται κανονικοποίηση (Normalization), όπως η κλιμάκωση Min-Max στο εύρος, για τη μείωση της επίδρασης των ακραίων τιμών (outliers) [6] [16].

2.3.1.2 Διαχωρισμός Δεδομένων (Splitting):

Τα δεδομένα χωρίζονται συνήθως σε σύνολο εκπαίδευσης (training set) και σύνολο ελέγχου (testing set). Μια συνήθης αναλογία είναι 70-80% για εκπαίδευση και 20-30% για έλεγχο, επιτρέποντας στο μοντέλο να μάθει από το μεγαλύτερο μέρος των δεδομένων και να αξιολογηθεί σε άγνωστα δεδομένα [6] [16] [17].

2.3.1.3 Επικύρωση K-Fold (K-Fold Cross Validation)

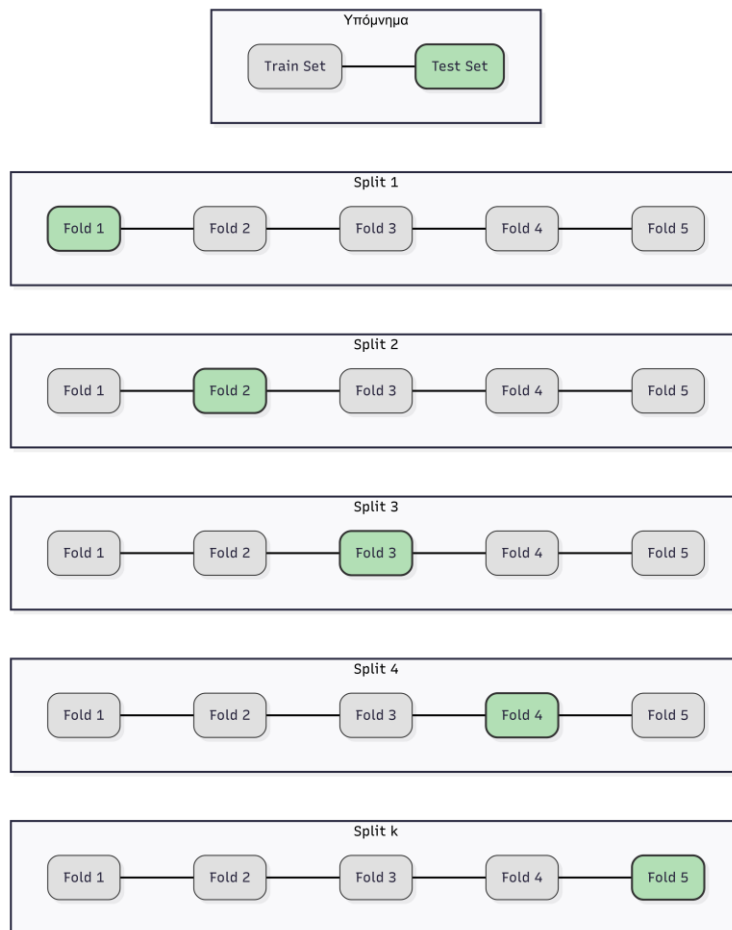
Για την αποφυγή της υπερ-προσαρμογής (overfitting) και την εξαγωγή πιο αξιόπιστων αποτελεσμάτων, χρησιμοποιείται η τεχνική της διασταυρούμενης επικύρωσης K-fold. Η μέθοδος αυτή αποτελεί συχνά την **πρώτη φάση** στη μεθοδολογία ανάπτυξης μοντέλων, πριν ακόμα εφαρμοστεί η ρύθμιση υπερπαραμέτρων (hyperparameter tuning), λειτουργώντας ως μηχανισμός πιστοποίησης της σταθερότητας του μοντέλου (**Εικόνα 2.2**) [17].

Η διαδικασία λειτουργεί ως εξής:

- **Διαχωρισμός:** Τα δεδομένα χωρίζονται σε K ισομερή υποσύνολα.

- **Επανάληψη:** Η διαδικασία εκπαίδευσης επαναλαμβάνεται **k** φορές. Σε κάθε επανάληψη, ένα διαφορετικό υποσύνολο χρησιμοποιείται για έλεγχο (validation) και τα υπόλοιπα $K-1$ για εκπαίδευση [17].
- **Επιλογή του K:** Συνήθης επιλογή είναι το **K=10**, καθώς έχει παρατηρηθεί ότι προσφέρει ακριβέστερα αποτελέσματα στη γενίκευση του μοντέλου. Ωστόσο, σε άλλες μελέτες, όπως για την πρόβλεψη πυρκαγιών στο πάρκο Montesinho με χρήση μοντέλων XGBoost και Random Forest, έχει εφαρμοστεί και η πρακτική της ιεραρχικής διασταυρούμενης επικύρωσης 5 πτυχών (**5-fold cross-validation**) [16].

Η αξία της μεθόδου φαίνεται στη σύγκριση των αποτελεσμάτων. Για παράδειγμα, σε πειράματα με τον αλγόριθμο **Random Forest**, η ακρίβεια που μετρήθηκε μέσω 10-fold cross-validation (96.00%) ήταν πολύ κοντά στην ακρίβεια ελέγχου (test accuracy 95.12%), επιβεβαιώνοντας ότι το μοντέλο δεν «απομνημόνευσε» απλώς τα δεδομένα εκπαίδευσης αλλά έμαθε να γενικεύει σωστά. Αντίθετα, μοντέλα όπως τα Support Vector Classifiers (SVC) μπορεί να εμφανίσουν χαμηλότερη απόδοση στο K-fold (π.χ. 81.27%) σε σχέση με τα βελτιστοποιημένα μοντέλα, υποδεικνύοντας την ανάγκη για περαιτέρω ρύθμιση [17].



Εικόνα 2.2: Διαχωρισμός δεδομένων με τη μέθοδο k-Fold Cross-Validation

2.3.1.4 Βελτιστοποίηση Υπερπαραμέτρων (Hyperparameter Tuning)

Για την εύρεση της βέλτιστης απόδοσης και την αποφυγή υπερ-προσαρμογής (overfitting), εφαρμόζονται τεχνικές βελτιστοποίησης όπως η αναζήτηση πλέγματος (Grid Search). Η διαδικασία αυτή αποτελεί τη δεύτερη φάση μετά την επικύρωση (π.χ. K-fold cross-validation) και στοχεύει στον εντοπισμό του ιδανικού συνδυασμού παραμέτρων που ενισχύουν την προβλεπτική ικανότητα του μοντέλου. Συγκεκριμένα, δημιουργείται ένα πλέγμα (grid) πιθανών τιμών για τις παραμέτρους και το σύστημα εκτελεί επαναληπτικές δοκιμές όλων των συνδυασμών μέχρι να καταλήξει στο βέλτιστο μοντέλο [17].

Η εφαρμογή της μεθόδου αυτής έχει αποδειχθεί κρίσιμη για την αύξηση της ακρίβειας σε μοντέλα πρόβλεψης δασικών πυρκαγιών. Χαρακτηριστικά παραδείγματα βελτιστοποίησης από τη βιβλιογραφία περιλαμβάνουν:

- **Support Vector Classifier (SVC):** Πριν τη βελτιστοποίηση, το μοντέλο παρουσίαζε ακρίβεια περίπου 91.50% (με χρήση k-fold). Μέσω του Grid Search, εντοπίστηκαν οι βέλτιστες παράμετροι ($C=50$, $\gamma=scale$, $kernel=rbf$), οι οποίες εκτόξευσαν την ακρίβεια στο **96.88%**, καθιστώντας το SVC την αποδοτικότερη στρατηγική για το συγκεκριμένο σύνολο δεδομένων [17].
- **Random Forest (RF):** Σε διαφορετικό σύνολο δεδομένων (Cortez and Morais), η βελτιστοποίηση οδήγησε σε ακρίβεια **90.48%**. Οι βέλτιστες παράμετροι που επιλέχθηκαν ήταν $max_features=sqrt$ και $n_estimators=100$ (αριθμός δέντρων) [17].
- **Βαθιά Μάθηση (Deep Learning - LSTM):** Στα νευρωνικά δίκτυα, η ρύθμιση αφορά την αρχιτεκτονική και τη διαδικασία εκπαίδευσης. Για παράδειγμα, σε μοντέλα LSTM για πρόβλεψη πυρκαγιών, επιλέγονται παράμετροι όπως ο αριθμός των κρυφών νευρώνων (π.χ. 100 hidden neurons), η πιθανότητα Dropout (π.χ. 0.5) για την αντιμετώπιση του overfitting, το μέγεθος του MiniBatch (π.χ. 12) και ο αριθμός των εποχών εκπαίδευσης (MaxEpochs = 1000) ώστε να ελαχιστοποιηθεί το σφάλμα RMSE [6].

2.3.1.5 Κατηγορίες Προβλημάτων & Αλγόριθμοι

Η επιλογή της κατάλληλης μεθοδολογίας Μηχανικής Μάθησης εξαρτάται άμεσα από τη φύση της μεταβλητής στόχου (target variable) και το είδος του προβλήματος που καλείται να επιλύσει το μοντέλο. Στο πλαίσιο της πρόβλεψης δασικών πυρκαγιών, τα προβλήματα διακρίνονται κυρίως σε δύο κατηγορίες: την Ταξινόμηση και την Παλινδρόμηση. Στην παρούσα μελέτη, αξιοποιούνται και οι δύο προσεγγίσεις για την εξαγωγή ολοκληρωμένων συμπερασμάτων, κάνοντας χρήση ισχυρών αλγορίθμων όπως τα Δέντρα Απόφασης, τα Random Forest και το XGBoost.

2.3.1.6 Ταξινόμηση vs Παλινδρόμηση (Classification vs Regression)

Η **Ταξινόμηση (Classification)** εφαρμόζεται όταν το ζητούμενο είναι η πρόβλεψη διακριτών κλάσεων ή κατηγοριών. Στην περίπτωση των δασικών πυρκαγιών, η ταξινόμηση χρησιμοποιείται για να διαχωρίσει τα δεδομένα σε καταστάσεις όπως «Πυρκαγιά» και «Μη Πυρκαγιά» (Binary Classification) ή για να κατηγοριοποιήσει τον κίνδυνο σε επίπεδα, όπως «Χαμηλός», «Μέτριος», «Υψηλός» και «Πολύ Υψηλός» [17] [18]. Ο στόχος είναι η εκμάθηση μιας συνάρτησης που αντιστοιχίζει τα μετεωρολογικά και περιβαλλοντικά δεδομένα εισόδου (π.χ. θερμοκρασία, υγρασία, άνεμος) σε μια συγκεκριμένη ετικέτα κλάσης (class label), επιτρέποντας στις αρχές να λαμβάνουν αποφάσεις τύπου "ναι/όχι" για την κινητοποίηση πόρων [18] [19]. Η αξιολόγηση αυτών των μοντέλων βασίζεται στην ικανότητά τους να διαχωρίζουν σωστά τις κλάσεις, ελαχιστοποιώντας τα ψευδώς θετικά ή αρνητικά αποτελέσματα [17].

Αντίθετα, η **Παλινδρόμηση (Regression)** χρησιμοποιείται όταν η μεταβλητή στόχος είναι συνεχής και αριθμητική. Στο πλαίσιο της παρούσας διατριβής, η παλινδρόμηση εφαρμόζεται για την πρόβλεψη συγκεκριμένων τιμών, όπως η τιμή του δείκτη FWI (Fire Weather Index) [16] [20]. Τα μοντέλα παλινδρόμησης προσπαθούν να προσαρμόσουν μια γραμμή ή μια καμπύλη στα δεδομένα, ώστε να ελαχιστοποιήσουν το σφάλμα μεταξύ της προβλεπόμενης και της πραγματικής τιμής. Για παράδειγμα, η χρήση μετεωρολογικών δεδομένων για την πρόβλεψη της έντασης μιας πυρκαγιάς απαιτεί μοντέλα παλινδρόμησης, καθώς η έξοδος είναι μια ποσότητα και όχι μια κατηγορία. Η διάκριση αυτή είναι κρίσιμη, καθώς καθορίζει τόσο την αρχιτεκτονική του μοντέλου όσο και τις μετρικές που θα χρησιμοποιηθούν για την αξιολόγησή του [17] [18].

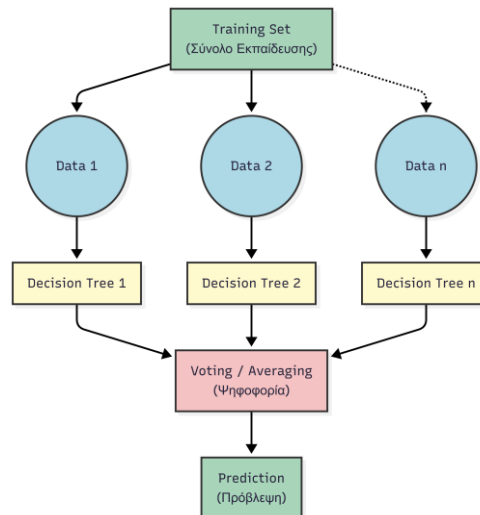
2.3.1.7 Δέντρα Αποφάσεων (Decision Trees)

Τα **Δέντρα Αποφάσεων (Decision Trees - DT)** αποτελούν έναν από τους πιο θεμελιώδεις αλγόριθμους στη Μηχανική Μάθηση, λειτουργώντας ως βάση για πιο σύνθετες μεθόδους συνόλου (ensemble methods). Πρόκειται για ιεραρχικά μοντέλα που αποτελούνται από μια «ρίζα» (root), εσωτερικούς κόμβους (internal nodes) και «φύλλα» (leaves) [21]. Η διαδικασία μάθησης περιλαμβάνει τον αναδρομικό διαχωρισμό του συνόλου δεδομένων σε μικρότερα υποσύνολα, με βάση κανόνες απόφασης που εξάγονται από τα χαρακτηριστικά των δεδομένων (π.χ. «αν η Θερμοκρασία > 30°C»). Κάθε εσωτερικός κόμβος αντιπροσωπεύει έναν έλεγχο σε ένα χαρακτηριστικό, κάθε κλαδί αντιπροσωπεύει το αποτέλεσμα του ελέγχου και κάθε φύλλο αντιπροσωπεύει την τελική ετικέτα κλάσης (για ταξινόμηση) ή την τελική τιμή (για παλινδρόμηση) [18] [22].

Το κύριο πλεονέκτημα των Δέντρων Απόφασης είναι η ερμηνευσιμότητά τους, καθώς η δομή τους μοιάζει με ανθρώπινη λογική λήψης αποφάσεων και επιτρέπει την εύκολη κατανόηση της συμβολής κάθε μεταβλητής στο αποτέλεσμα [21]. Σε μελέτες πρόβλεψης πυρκαγιών, τα DT έχουν χρησιμοποιηθεί για την ταξινόμηση περιοχών σε ζώνες κινδύνου, επιτυγχάνοντας σε ορισμένες περιπτώσεις εξαιρετικά υψηλή ακρίβεια (έως και 98%) και τιμές AUC κοντά στο 0.99 [18]. Ωστόσο, τα μεμονωμένα δέντρα είναι επιρρεπή στο πρόβλημα της υπερ-προσαρμογής (overfitting), ειδικά όταν το δέντρο γίνεται πολύ βαθύ και πολύπλοκο, μαθαίνοντας θόρυβο αντί για πραγματικά μοτίβα. Για την αντιμετώπιση αυτού, χρησιμοποιούνται τεχνικές «κλαδέματος» (pruning) ή, πιο αποτελεσματικά, μέθοδοι συνόλου όπως τα Random Forests [21].

2.3.1.8 Random Forest (Bagging)

Ο αλγόριθμος **Random Forest (RF)** είναι μια μέθοδος συνόλου (ensemble learning method) που βασίζεται στην τεχνική **Bagging (Bootstrap Aggregating)** για να βελτιώσει την ακρίβεια και τη σταθερότητα των προβλέψεων, ξεπερνώντας τους περιορισμούς των απλών Δέντρων Απόφασης. Η βασική ιδέα πίσω από το Random Forest είναι η κατασκευή ενός μεγάλου αριθμού δέντρων απόφασης κατά την εκπαίδευση, όπου το καθένα εκπαιδεύεται σε ένα τυχαίο υποσύνολο των δεδομένων (bootstrap sample) και χρησιμοποιεί ένα τυχαίο υποσύνολο χαρακτηριστικών για κάθε διάσπαση κόμβου [22]. Για προβλήματα ταξινόμησης, η τελική πρόβλεψη προκύπτει από την πλειοψηφία των "ψήφων" των δέντρων (majority voting), ενώ για παλινδρόμηση, προκύπτει από τον μέσο όρο των προβλέψεων (**Εικόνα 2.3**) [21].

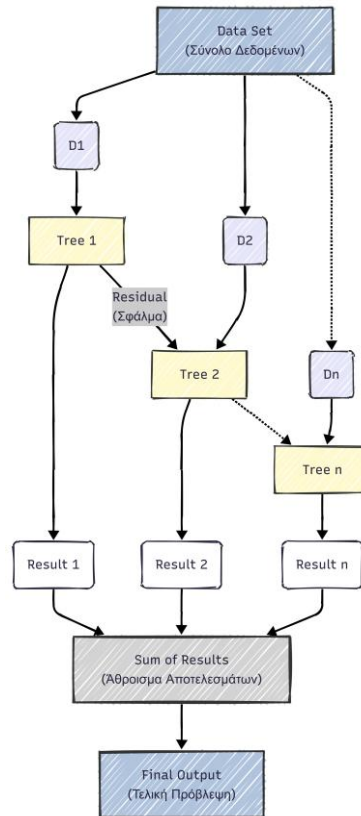


Εικόνα 2.3: Αρχιτεκτονική του αλγορίθμου Random Forest

Στην παρούσα μελέτη, το Random Forest αναδείχθηκε ως το μοντέλο με την υψηλότερη απόδοση στις μετρικές αξιολόγησης, εύρημα που συμφωνεί με τη διεθνή βιβλιογραφία. Για παράδειγμα, σε συγκριτικές μελέτες σε δεδομένα από την Πορτογαλία (Montesinho Natural Park) και την Ινδία, το Random Forest πέτυχε ακρίβεια έως 90.24% και 95.12% αντίστοιχα, ξεπερνώντας αλγορίθμους όπως τα SVM και KNN [16] [17]. Η υπεροχή του οφείλεται στην ικανότητά του να διαχειρίζεται πολύπλοκες, μη γραμμικές σχέσεις μεταξύ των μετεωρολογικών μεταβλητών και να αντιστέκεται στην υπερπροσαρμογή μέσω της τυχαιότητας που εισάγει στη διαδικασία εκπαίδευσης. Επιπλέον, παρέχει σημαντικές πληροφορίες για τη σπουδαιότητα των χαρακτηριστικών (feature importance), βοηθώντας στον εντοπισμό των πιο κρίσιμων παραγόντων, όπως η θερμοκρασία και η υγρασία, που οδηγούν σε πυρκαγιές [16] [23].

2.3.1.9 XGBoost (Gradient Boosting)

Ο αλγόριθμος **XGBoost (Extreme Gradient Boosting)** αποτελεί μια προηγμένη και εξαιρετικά αποδοτική υλοποίηση της τεχνικής **Gradient Boosting**. Σε αντίθεση με το Random Forest που κατασκευάζει δέντρα ανεξάρτητα (παράλληλα), το XGBoost λειτουργεί ακολουθιακά (sequentially). Η λογική του Boosting βασίζεται στην ιδέα ότι κάθε νέο δέντρο που προστίθεται στο μοντέλο στοχεύει να διορθώσει τα σφάλματα (residuals) που έκαναν τα προηγούμενα δέντρα. Συνδυάζοντας πολλούς «ασθενείς» εκμάθητες (weak learners) – δηλαδή ρηγά δέντρα απόφασης – δημιουργεί έναν ισχυρό προβλεπτικό μηχανισμό που μειώνει τόσο την προκατάληψη (bias) όσο και τη διακύμανση (variance) (Εικόνα 2.4) [22].



Εικόνα 2.4: Αρχιτεκτονική του αλγορίθμου XGBoost

Το XGBoost θεωρείται σήμερα state-of-the-art αλγόριθμος για δομημένα/πίνακες δεδομένων και έχει χρησιμοποιηθεί ευρέως για την πρόβλεψη της χωροχρονικής ευπάθειας σε πυρκαγιές. Μελέτες έχουν δείξει ότι το XGBoost μπορεί να επιτύχει ανώτερη απόδοση σε σχέση με προηγούμενα μοντέλα, με βελτίωση της ακρίβειας έως και 11.37% σε συγκεκριμένα datasets, και υψηλά σκορ σε μετρικές όπως το AUC (0.8052) [16]. Η αποτελεσματικότητά του οφείλεται στην ενσωματωμένη τακτοποίηση (regularization) που αποτρέπει την υπερ-προσαρμογή, στην ικανότητά του να διαχειρίζεται ελλιπή δεδομένα και στην ταχύτητα εκτέλεσης. Στο πλαίσιο της πρόβλεψης δασικών πυρκαγιών, το XGBoost έχει αποδειχθεί βέλτιστο στην αναγνώριση των πολύπλοκων μοτίβων που διέπουν τη συμπεριφορά της φωτιάς κάτω από διαφορετικές μετεωρολογικές συνθήκες [19].

2.4 Βαθιά Μάθηση & Χρονοσειρές (Deep Learning & Time Series)

Η Βαθιά Μηχανική Μάθηση (Deep Learning - DL), ως υποκατηγορία της τεχνητής νοημοσύνης, έχει φέρει επανάσταση στην ανάλυση περιβαλλοντικών δεδομένων και ειδικότερα στην πρόβλεψη φυσικών καταστροφών, όπως οι δασικές πυρκαγιές. Σε αντίθεση με τις παραδοσιακές μεθόδους Μηχανικής Μάθησης (όπως π.χ. τα Random Forests) που συχνά απαιτούν χειροκίνητη επιλογή χαρακτηριστικών, τα μοντέλα DL έχουν την ικανότητα να μαθαίνουν αυτόματα πολύπλοκες, μη γραμμικές συσχετίσεις απευθείας από μεγάλα σύνολα δεδομένων [6] [23]. Στο πλαίσιο των πυρκαγιών, τα δεδομένα είναι κυρίως χωροχρονικά, καθώς οι μετεωρολογικές συνθήκες (θερμοκρασία, υγρασία, άνεμος) και η κατάσταση της βλάστησης εξελίσσονται δυναμικά στον χρόνο και στον χώρο [19].

Πρόσφατες μελέτες έχουν δείξει ότι τα μοντέλα DL μπορούν να ξεπεράσουν σε απόδοση τους παραδοσιακούς δείκτες κινδύνου, όπως το Fire Weather Index (FWI), καθώς δεν βασίζονται μόνο σε

στατικές εμπειρικές εξισώσεις αλλά προσαρμόζονται στα πραγματικά ιστορικά δεδομένα. Για παράδειγμα, η χρήση κύβων δεδομένων (datacubes) που περιλαμβάνουν μεταβλητές όπως η θερμοκρασία επιφάνειας εδάφους (LST), ο δείκτης βλάστησης (NDVI) και η υγρασία εδάφους, επιτρέπει στα μοντέλα να εντοπίζουν κρυμμένα μοτίβα που οδηγούν σε ανάφλεξη, τα οποία οι γραμμικές μέθοδοι αδυνατούν να συλλάβουν [19].

2.4.1 Recurrent Neural Networks (RNNs)

Τα **Επαναλαμβανόμενα Νευρωνικά Δίκτυα** (Recurrent Neural Networks - RNNs) αποτελούν τη θεμελιώδη αρχιτεκτονική για την επεξεργασία ακολουθιακών δεδομένων (sequential data), όπως οι χρονοσειρές μετεωρολογικών μετρήσεων. Η βασική καινοτομία των RNN είναι η ύπαρξη βρόχων ανάδρασης, οι οποίοι επιτρέπουν στο δίκτυο να διατηρεί μια μορφή «μνήμης» από προηγούμενα χρονικά βήματα. Σε κάθε χρονική στιγμή, το δίκτυο λαμβάνει ως είσοδο τα τρέχοντα δεδομένα X_t και την «κρυφή κατάσταση» (hidden state) h_{t-1} από το προηγούμενο βήμα, συνδυάζοντας έτσι την τρέχουσα πληροφορία με το ιστορικό πλαίσιο [24].

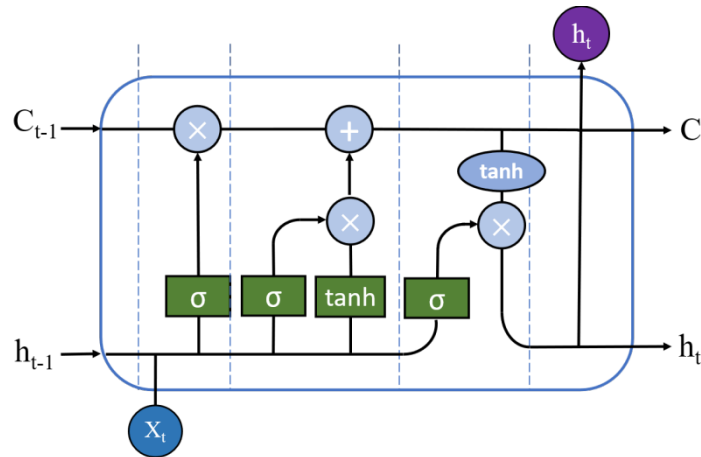
Ωστόσο, τα παραδοσιακά RNN αντιμετωπίζουν σημαντικούς περιορισμούς κατά την εκπαίδευσή τους σε μακρές χρονοσειρές. Το κυριότερο πρόβλημα είναι η «εξάλειψη της βαθμίδας» (vanishing gradient problem), όπου η πληροφορία από τα αρχικά βήματα της ακολουθίας εξασθενεί εκθετικά καθώς διαδίδεται προς τα πίσω στον χρόνο (backpropagation through time) [6] [24]. Αυτό καθιστά τα απλά RNN ανίκανα να μάθουν μακροχρόνιες εξαρτήσεις (long-term dependencies) [16], οι οποίες είναι κρίσιμες για την πρόβλεψη πυρκαγιών, καθώς η ξηρασία που προηγείται μιας πυρκαγιάς μπορεί να έχει συσσωρευτεί για εβδομάδες ή μήνες. Για την αντιμετώπιση αυτού του προβλήματος, αναπτύχθηκαν πιο εξελιγμένες αρχιτεκτονικές όπως τα LSTM και GRU.

2.4.1.1 Long Short-Term Memory (LSTM)

Τα δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory - LSTM), που εισήχθησαν από τους Hochreiter και Schmidhuber [6], αποτελούν σήμερα την πιο διαδεδομένη λύση για την πρόβλεψη χρονοσειρών στον τομέα των πυρκαγιών. Η αρχιτεκτονική τους έχει σχεδιαστεί ειδικά για να επιλύει το πρόβλημα της εξάλειψης της βαθμίδας και να διαχειρίζεται αποτελεσματικά τις χρονικές υστερήσεις (time lags) [16] [26].

Δομή και Λειτουργία (Εικόνα 2.5): Το κελί μνήμης του LSTM είναι πιο σύνθετο από αυτό του RNN και περιλαμβάνει τρεις «πύλες» (gates) που ελέγχουν τη ροή της πληροφορίας:

1. **Πύλη Λήθης (Forget Gate):** Αποφασίζει ποιες πληροφορίες από την προηγούμενη κατάσταση είναι πλέον άχρηστες και πρέπει να διαγραφούν [6].
2. **Πύλη Εισόδου (Input Gate):** Καθορίζει ποιες νέες πληροφορίες θα αποθηκευτούν στην κατάσταση του κελιού [6].
3. **Πύλη Εξόδου (Output Gate):** Ελέγχει την τελική έξοδο που θα περάσει στο επόμενο χρονικό βήμα [6].



Εικόνα 2.5: Δομή Αλγορίθμου LSTM/GRU [24]

Εφαρμογές και Απόδοση: Στη βιβλιογραφία, τα LSTM έχουν χρησιμοποιηθεί ευρέως για την πρόβλεψη «θερμών σημείων» (hotspots) και **κινδύνου πυρκαγιάς**.

- Σύγκριση με Παραδοσιακές Μεθόδους: Σε συγκριτικές μελέτες, τα LSTM υπερέχουν έναντι στατιστικών μοντέλων όπως το ARIMA και το SARIMA. Για παράδειγμα, σε πειράματα πρόβλεψης, το LSTM επέτυχε χαμηλότερο σφάλμα (RMSE 3.09) σε σχέση με το ARIMA (RMSE 3.22) και το SARIMA (RMSE 3.43), αποδεικνύοντας την ικανότητά του να διαχειρίζεται τη μη γραμμικότητα των μετεωρολογικών δεδομένων [16].
- Πρόβλεψη Κινδύνου στην Ελλάδα: Σε πρόσφατη μελέτη για την Ελλάδα (Kondylatos et al., 2022), ένα μοντέλο LSTM εκπαιδεύτηκε σε κύβο δεδομένων (datacube) και συγκρίθηκε με τον δείκτη FWI. Το LSTM επέτυχε σημαντικά υψηλότερη τιμή AUC (0.886 ± 0.011) σε σχέση με το FWI (0.693 ± 0.007), αποδεικνύοντας ότι μπορεί να γενικεύσει ακόμα και σε ακραίες αντιπυρικές περιόδους, όπως το καλοκαίρι του 2021 [19].
- Συνδυαστικά Μοντέλα: Πιο προηγμένες προσεγγίσεις, όπως το LSTNet, συνδυάζουν τα LSTM με Συνελκτικά Νευρωνικά Δίκτυα (CNN) για να εξάγουν ταυτόχρονα βραχυπρόθεσμα τοπικά χαρακτηριστικά και μακροπρόθεσμες χρονικές τάσεις, βελτιώνοντας περαιτέρω την ακρίβεια πρόβλεψης [24].

2.4.2 Gated Recurrent Units (GRU)

Τα Gated Recurrent Units (GRU) αποτελούν μια νεότερη παραλλαγή των LSTM, η οποία προτάθηκε για να απλοποιήσει την αρχιτεκτονική μειώνοντας τον υπολογιστικό φόρτο, χωρίς σημαντική απώλεια στην ακρίβεια.

Διαφορές από το LSTM: Σε αντίθεση με το LSTM που διαθέτει τρεις πύλες και ξεχωριστή κατάσταση κελιού (cell state), το GRU συνδυάζει την κρυφή κατάσταση και την κατάσταση κελιού σε μία ενιαία δομή και χρησιμοποιεί μόνο δύο πύλες:

1. **Πύλη Ενημέρωσης (Update Gate):** Καθορίζει πόση από την προηγούμενη πληροφορία θα διατηρηθεί (συνδυάζοντας λειτουργικά τις πύλες εισόδου και λήθης του LSTM) [16].
2. **Πύλη Επαναφοράς (Reset Gate):** Αποφασίζει πόσο από το παρελθόν θα αγνοηθεί για τον υπολογισμό της νέας υποψήφιας κατάστασης [16].

Απόδοση και Χρήση: Τα GRU θεωρούνται πιο αποδοτικά από άποψη ταχύτητας εκπαίδευσης λόγω των λιγότερων παραμέτρων τους. Στον τομέα των δασικών πυρκαγιών, μελέτες έχουν δείξει ότι τα GRU μπορούν να επιτύχουν αποτελέσματα συγκρίσιμα ή και ελαφρώς καλύτερα από τα LSTM σε συγκεκριμένα datasets [16].

- **Ακρίβεια:** Σε μελέτη σύγκρισης για την πρόβλεψη πυρκαγιών, το μοντέλο GRU επέτυχε το χαμηλότερο σφάλμα RMSE (2.95) σε σχέση με το LSTM (3.09) και το RNN (3.15), υποδεικνύοντας ότι η απλοποιημένη δομή του μπορεί να είναι πιο αποτελεσματική όταν τα δεδομένα δεν απαιτούν την εξαιρετικά μακροχρόνια μνήμη του LSTM [16].
- **Εφαρμογή:** Λόγω της ταχύτητάς τους, τα GRU προτιμώνται συχνά σε εφαρμογές που απαιτούν ταχεία επεξεργασία ή όταν οι διαθέσιμοι υπολογιστικοί πόροι είναι περιορισμένοι, προσφέροντας μια ισορροπημένη λύση μεταξύ πολυπλοκότητας και απόδοσης [16].

2.5 Μετρικές Αξιολόγησης (Evaluation Metrics)

Η επιλογή των κατάλληλων μετρικών είναι καθοριστική για την αντικειμενική αξιολόγηση της απόδοσης των μοντέλων Μηχανικής Μάθησης. Οι μετρικές αυτές παρέχουν ποσοτικά στοιχεία για το πόσο καλά το μοντέλο μπορεί να γενικεύσει σε άγνωστα δεδομένα και να διαχειριστεί τις ιδιαιτερότητες του προβλήματος, όπως η ανισορροπία των κλάσεων (class imbalance).

2.5.1 Πίνακας Σύγχυσης (Confusion Matrix)

Ο Πίνακας Σύγχυσης είναι το θεμελιώδες εργαλείο για την αξιολόγηση μοντέλων ταξινόμησης. Πρόκειται για έναν πίνακα που συνοψίζει τα αποτελέσματα των προβλέψεων συγκρίνοντάς τα με τις πραγματικές τιμές. Αποτελείται από τέσσερις βασικές συνιστώσες [17] [24]:

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Εικόνα 2.6: Πίνακας Σύγχυσης (Confusion Matrix)

- **True Positives (TP - Αληθώς Θετικά):** Ο αριθμός των περιπτώσεων που το μοντέλο πρόβλεψε σωστά ότι υπάρχει πυρκαγιά (ή κίνδυνος).
- **True Negatives (TN - Αληθώς Αρνητικά):** Ο αριθμός των περιπτώσεων που το μοντέλο πρόβλεψε σωστά ότι δεν υπάρχει πυρκαγιά (ασφάλεια).
- **False Positives (FP - Ψευδώς Θετικά):** Γνωστά και ως "Type I error" ή "ψευδής συναγερμός". Το μοντέλο προέβλεψε πυρκαγιά, ενώ στην πραγματικότητα δεν υπήρξε.

- **False Negatives (FN - Ψευδώς Αρνητικά):** Γνωστά και ως "Type II error". Το μοντέλο προέβλεψε ότι δεν υπάρχει κίνδυνος, ενώ στην πραγματικότητα εκδηλώθηκε πυρκαγιά. Αυτό είναι το πιο κρίσιμο σφάλμα σε συστήματα ασφαλείας.

Οπτικά παραδείγματα πινάκων σύγχυσης για αλγορίθμους όπως RF, SVM και KNN παρουσιάζονται συχνά στη βιβλιογραφία για να δείξουν την κατανομή των σφαλμάτων [24].

2.5.2 Μετρικές Ταξινόμησης (Accuracy, Precision, Recall, F1-Score)

Με βάση τον Πίνακα Σύγχυσης, υπολογίζονται οι εξής μετρικές:

- **Accuracy (Ακρίβεια):** Το ποσοστό των σωστών προβλέψεων στο σύνολο των δειγμάτων (Εξίσωση 2.35). Αν και δημοφιλής, μπορεί να είναι παραπλανητική σε ανισοβαρείς κατηγορίες (π.χ. όταν οι ημέρες πυρκαγιάς είναι πολύ λιγότερες από τις ημέρες χωρίς πυρκαγιά) [17].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.35)$$

- **Precision (Ακρίβεια Θετικής Πρόβλεψης):** Το ποσοστό των πραγματικών πυρκαγιών στο σύνολο των προβλέψεων για πυρκαγιά (Εξίσωση 2.36). Υψηλή τιμή σημαίνει λίγους ψευδείς συναγεμμούς [17].

$$Precision = \frac{TP}{TP + FP} \quad (2.36)$$

- **Recall (Ανάκληση ή Ευαισθησία):** Το ποσοστό των πυρκαγιών που εντοπίστηκαν σωστά στο σύνολο των πραγματικών πυρκαγιών (Εξίσωση 2.37). Για την πρόβλεψη πυρκαγιών, το Recall είναι υψίστης σημασίας, καθώς στόχος είναι η ελαχιστοποίηση των False Negatives (FN). Το να χάσει το σύστημα ένα πραγματικό συμβάν μπορεί να έχει καταστροφικές συνέπειες [17-19].

$$Recall = \frac{TP}{TP + FN} \quad (2.37)$$

- **F1-Score:** Ο αρμονικός μέσος της Precision και του Recall (Εξίσωση 2.38). Είναι η πιο αξιόπιστη μετρική όταν υπάρχει ανισορροπία στα δεδομένα, καθώς τιμωρεί τις ακραίες τιμές και εξασφαλίζει ότι το μοντέλο έχει καλή απόδοση και στις δύο πτυχές (εντοπισμός πυρκαγιών και αποφυγή ψευδών συναγεμώσεων) [17] [19].

$$F1 - Score = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (2.38)$$

2.5.3 ROC Curve & AUC

Η καμπύλη ROC (Receiver Operating Characteristic) είναι ένα γράφημα που απεικονίζει την απόδοση ενός ταξινομητή σε όλα τα κατώφλια ταξινόμησης, σχεδιάζοντας το True Positive Rate (Recall) έναντι του False Positive Rate. Το AUC (Area Under the Curve) μετρά την περιοχή κάτω από την καμπύλη ROC και παρέχει μια συνολική μέτρηση της ικανότητας του μοντέλου να διαχωρίζει τις κλάσεις [17] [21].

- Η τιμή του AUC κυμαίνεται από 0 έως 1.
- Μια τιμή AUC = 1.0 υποδηλώνει τέλει διαχωρισμό (το μοντέλο προβλέπει σωστά όλα τα θετικά και αρνητικά δείγματα).

- Μια τιμή $AUC = 0.5$ υποδηλώνει τυχαία μαντεψιά. Στη βιβλιογραφία, μοντέλα όπως το Random Forest και το XGBoost συχνά επιτυγχάνουν υψηλά scores AUC (π.χ. >0.90), αποδεικνύοντας την υπεροχή τους έναντι παραδοσιακών δεικτών όπως το FWI [16] [18] [21].

2.5.4 Μετρικές Παλινδρόμησης (MAE, RMSE, R²)

Για τα προβλήματα παλινδρόμησης (π.χ. πρόβλεψη δείκτη FWI), χρησιμοποιούνται οι εξής μετρικές:

- **MAE (Mean Absolute Error):** Το μέσο απόλυτο σφάλμα (Εξίσωση 2.39). Εκφράζει το μέσο μέγεθος του σφάλματος στις ίδιες μονάδες με την προβλεπόμενη μεταβλητή και είναι εύκολα ερμηνεύσιμο [24] [25].

$$MAE = \frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}| \quad (2.39)$$

- **RMSE (Root Mean Square Error):** Η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος. Επειδή υψώνει τα σφάλματα στο τετράγωνο πριν τον μέσο όρο, το RMSE «τιμωρεί» τα μεγάλα σφάλματα περισσότερο από το MAE (Εξίσωση 2.40). Αυτό είναι επιθυμητό στην πρόβλεψη πυρκαγιών, όπου μεγάλες αποκλίσεις στην πρόβλεψη κινδύνου πρέπει να αποφεύγονται πάση θυσία [24] [25].

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}|^2} \quad (2.40)$$

- **R² (Coefficient of Determination - Συντελεστής Προσδιορισμού):** Μετρά το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο. Παίρνει τιμές από 0 έως 1 (ή και αρνητικές για πολύ κακά μοντέλα). Μια τιμή κοντά στο 1 υποδηλώνει ότι το μοντέλο προσαρμόζεται εξαιρετικά στα δεδομένα και εξηγεί τη μεταβλητότητα του φαινομένου. Στη μελέτη μας, υψηλές τιμές R² (π.χ. >0.90) σε μοντέλα Gradient Boosting υποδεικνύουν ισχυρή συσχέτιση μεταξύ των προβλέψεων και των πραγματικών τιμών [20] [25].

Κεφάλαιο 3ο: Φάση 1 - Μηχανική Δεδομένων (Data Engineering)

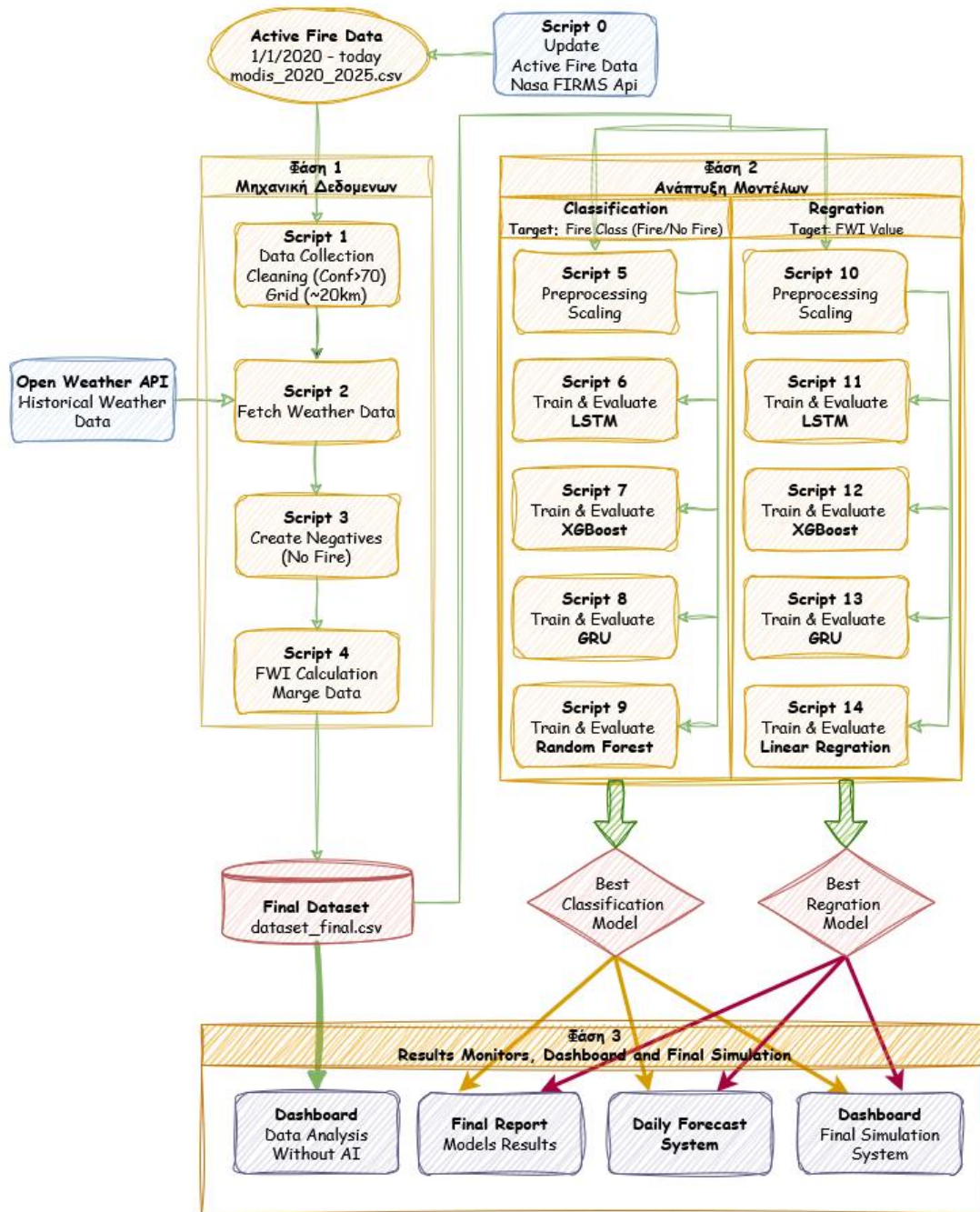
3.1 Εισαγωγή

Η αρχιτεκτονική που υιοθετήθηκε στην παρούσα διπλωματική εργασία, αφορά την συλλογή τηλεσκοπικών δορυφορικών δεδομένων από αισθητήρες που υποδηλώνουν ενεργές εστίες φωτιάς (Hotspots). Τα δεδομένα αυτά ως ακατέργαστα και με την βοήθεια των μετεωρολογικών δεδομένων για όλα τα hotspots της συλλογής, μετατράπηκαν σε αξιοποιήσιμη πληροφορία για την ανάλυση κινδύνου δασικών πυρκαγιών στον Ελλαδικό χώρο. Το σύστημα κατασκευάστηκε σε κώδικα **Python**, αξιοποιώντας βιβλιοθήκες ανάλυσης δεδομένων (**Pandas**), Μηχανικής Μάθησης (**Scikit-Learn**, **XGBoost**, **TensorFlow**), διεπαφής χρήστη UI (**Streamlit**), γεωγραφικής απεικόνισης (**Folium**, **Geopy**), καθώς δημιουργίας γραφικών και διαγραμμάτων (**Matplotlib**, **Plotly**, **Folium**).

Η ροή εργασίας (workflow) που ακολουθήσαμε αποτελούνται από τρεις αλληλένδετες φάσεις :

1. **Φάση 1:** Συλλογή δεδομένων, καθαρισμός, επεξεργασία και δημιουργία τελικού συνόλου δεδομένων.
2. **Φάση 2:** Προ - επεξεργασία δεδομένων και εκπαίδευση μοντέλων
3. **Φάση 3:** Αξιολόγηση βέλτιστων μοντέλων και ανάπτυξη συστήματος υβριδικής λήψης αποφάσεων

Στο **Σχήμα 3.1** που ακολουθεί, παρουσιάζουμε γραφικά όλα τα βήματα υλοποίησης της κάθε φάσης. Στην συνέχεια, αναλύουμε την περιγραφή των διαδικασιών που υλοποιούνται σε κάθε φάση.



Εικόνα 3.1: Αναλυτικό διάγραμμα ροής εργασιών (Workflow)

3.2 Συλλογή Δεδομένων και επεξεργασία αρχικών δεδομένων

Στην παρούσα ενότητα θα αναλύσουμε την διαδικασία συλλογής και επεξεργασίας του αρχικού dataset και την προετοιμασία αυτού για την συλλογή μετεωρολογικών δεδομένων για τις περιοχές ενδιαφέροντος.

3.2.1 Δεδομένα Ενεργών Πυρκαγιών (Script 0 και Script 1)

Ο σημαντικότερος παράγοντας για την σωστή δημιουργία του συστήματος, είναι η άντληση δεδομένων από αξιόπιστες πηγές. Τα δεδομένα περιέχουν ιστορικές καταγραφές ενεργών πυρκαγιών στον

Ελλαδικό χώρο των τελευταίων πέντε (5) ετών (2020 – Σήμερα). Η άντληση των δεδομένων αυτών έγινε από την πλατφόρμα του **Copernicus EMS**. [30]

Το αρχείο με τα αρχικά δεδομένα ονομάστηκε **modis_2020_2025.csv**. Στον **Πίνακα 3.1** φαίνεται ενδεικτικά η μορφή των δεδομένων ενώ στον **Πίνακα 3.2** παρουσιάζεται η αναλυτική περιγραφή της κάθε μεταβλητής.

Ωστόσο από το σύνολο των δεδομένων που αντλήθηκαν για τα τελευταία 5 χρόνια, κάναμε χρήση συγκεκριμένες εγγραφές. Επίσης από το σύνολο των μεταβλητών επιλέχθηκαν αυτές που θα ήταν χρήσιμες για τον υπολογισμό τον χαρακτηριστικών εισόδου των μοντέλων που θα αναλύσουμε παρακάτω.

Η αρχική εκκαθάριση των δεδομένων βάζοντας αυστηρό φίλτρο εμπιστοσύνης (Confidence). Έτσι λοιπόν πήραμε όλες τις εγγραφές που είχαν ποσοστό εμπιστοσύνης **μεγαλύτερο του 70%**. Αυτό έγινε ώστε να διασφαλίσουμε την εγκυρότητα των δεδομένων αποφεύγοντας ψευδούς συναγεμούς από μικρές αγροτικές φωτιές ή ενδεχομένως ανανακλάσεις από φωτοβολταϊκά πάνελ.

Βάζοντας αυτό το σκληρό φίλτρο (Confidence >70%) σε δεδομένα που συλλέχθηκαν από το Φασματοραδιόμετρο MODIS που είναι εγκατεστημένο στους δορυφόρους **Terra** και **Aqua** της NASA, Εξασφαλίζοντας χωρική ανάλυση 250 m [32], Διασφαλίζουμε ότι τα μοντέλα θα εκπαιδευτούν πάνω σε πραγματικά συμβάντα πυρκαγιών στον Ελλαδικό χώρο.

Πίνακας 3.1: Ενδεικτική μορφή δεδομένων αρχικού συνόλου δεδομένων

acq_at	lat	lon	frp	confidence	satellite	admin1
2020-01-02 09:30	38.841	22.482	90.0	53	T	GRC.8_1
2020-01-02 11:10	41.704	26.268	1279.0	100	A	GRC.6_1

Πίνακας 3.2: Περιγραφή μεταβλητών αρχείου [30] [31]

Μεταβλητή	Τύπος	Περιγραφή
X / Y	Float	Προβαλλόμενες συντεταγμένες (Projected Coordinates) του σημείου ανάγνωσης.
acq_at	Datetime	Ημερομηνία και ώρα λήψης δεδομένων.
lat	Float	Γεωγραφικό Πλάτος (Latitude)
lon	Float	Γεωγραφικό Μήκος (Longitude)
frp	Float	Fire Radiative Power (MW): Δείκτης της έντασης της φωτιάς σε MW.
confidence	Integer	Επίπεδο Εμπιστοσύνης (%): Ποσοστό βεβαιότητας για την ύπαρξη φωτιάς (0-100).
night	Boolean	0 = Ημέρα, 1 = Νύχτα
satellite	Char	Ο δορυφόρος λήψης: T = Terra, A = Aqua.
scan	Float	Το πλάτος του εικονοστοιχείου (pixel) κατά τη σάρωση (scan direction).

track	Float	Το μήκος του εικονοστοιχείου (pixel) κατά την τροχιά (track direction).
ver	String	Έκδοση του αλγορίθμου επεξεργασίας της NASA (π.χ. 6.0 NRT - Near Real Time).
bright_mir	Float	Θερμοκρασία Λαμπρότητας (Channel 21/22): Μετρημένη στο Μεσαίο Υπέρυθρο (Mid-Infrared). Χρησιμοποιείται για τον εντοπισμό θερμών σημείων. (σε Kelvin).
bright_tir	Float	Θερμοκρασία Λαμπρότητας (Channel 31): Μετρημένη στο Θερμικό Υπέρυθρο (Thermal Infrared). (Μονάδα: Kelvin).
ndvi	Float	Normalized Difference Vegetation Index: Δείκτης βλάστησης.
admin0	String	Κωδικός Χώρας (π.χ. GRC για Ελλάδα).
admin1 - admin3	String	Διοικητικές υποδιαιρέσεις (Περιφέρεια, Δήμος, Δημοτική Ενότητα)

Μπαίνοντας στο πρακτικό κομμάτι της εργασίας, θα αναλύσουμε το πρώτο script που κάνει το αρχικό ξεκαθάρισμα των δεδομένων (Script 1 - **1_01_data_collection.py**).

Έχοντας λοιπόν φόρτωση τα αρχικά ακατέργαστα δεδομένα τα βήματα που ακολουθούμε είναι τα εξής: Παρακάτω θα δούμε τα στάδια υλοποίησης του 1^{ου} Script.

3.2.1.1 Καθαρισμός δεδομένων αρχικού dataset

Έχοντας λοιπόν φόρτωση τα αρχικά ακατέργαστα δεδομένα, κάναμε μια αρχική επιλογή των σημαντικότερων μεταβλητών για την δημιουργία χαρακτηριστικών εισόδου (input features) στα μοντέλα (Σε παρακάτω στάδια). Έτσι λοιπόν επιλέχθηκαν οι παρακάτω μεταβλητές που φαίνονται στο κομμάτι του κώδικα **Κώδικα 3.1**.

```
features = [
    "lat", "lon",      # Συντεταγμένες σημείου
    "acq_at",         # Πότε πάρθηκε το δείγμα
    "frp",            # Ένταση φωτιάς
    "confidence",     # Δείκτης Εμπιστοσύνης
    "ndvi",           # Δείκτης Βλάστησης (Καύσιμη Υλη)
    "admin2"          # Αναγνωριστικό Περιοχής
]
```

Κώδικας 3.1: Επιλεγμένες μεταβλητές αρχικών δεδομένων

Μια πολύ σημαντική διαδικασία για τη δημιουργία ενός αξιόπιστου αρχείου δεδομένων (εκτός της απομάκρυνσης των εγγράφων που έχουν ελλιπή στοιχεία – NaN values) ήταν και ο καθαρισμός του

σκληρού φίλτρου εμπιστοσύνης. Έτσι λοιπόν επιλέχθηκαν οι εγγραφές είχαν ποσοστό εμπιστοσύνης μεγαλύτερο ή ίσο του 70% (Confidence \geq 70%).

3.2.1.2 Μέθοδος Χωρικής Ομαδοποίησης

Για μεγαλύτερη εξοικονόμηση υπολογιστικής ισχύος και την αποδοτική διαχείριση των καιρικών μετεωρολογικών δεδομένων, εφαρμόστηκε η τεχνική της χωρικής ομαδοποίησης (**grid-based spatial binning**). Στόχος της μεθόδου ήταν στην μείωση της υπολογιστικής πολυπλοκότητας. Ομαδοποιήσαμε ουσιαστικά τα γειτονικά hotspot σε ένα σημείο, ώστε να μειωθούν αισθητά τα δείγματα των ενεργών πυρκαγιών, χωρίς όμως να αλλοιωθεί η ακρίβεια αυτών.

Συγκεκριμένα ο γεωγραφικός χώρος της Ελλάδας με βάση τα σημεία (lat, lon) που αντλήσαμε από το αρχικό dataset, χωρίστηκε σε γεωγραφικό πλέγμα με βήμα 0,2°. Υπολογίστηκε αλγοριθμικά οι συντεταγμένες κάθε σημείου στο κοντινότερο κέντρο του κελιού Grid [4]. Στον **Κώδικα 3.2** φαίνεται ο υπολογισμός για lat και lon με βήμα 0.2° και η δημιουργία μοναδικού ID για κάθε Grid που δημιουργήθηκε.

```
step = 0.2 # Βήμα για spatial binning (~ 20km)

# Τεχνική Spatial Binning (Ανάλυση ~20km)
df_clean["lat_grid"] = (df_clean["lat"] / step).round() * step
df_clean["lon_grid"] = (df_clean["lon"] / step).round() * step

# Φτιάχνουμε ένα μοναδικό ID για κάθε Grid (π.χ. "Grid_37.8_23.4")
df_clean["admin2"] = "Grid_" + df_clean["lat_grid"].map("{:.1f}".format) + "_" +
df_clean["lon_grid"].map("{:.1f}".format)
```

Κώδικας 3.2: Τεχνική Spatial Binning

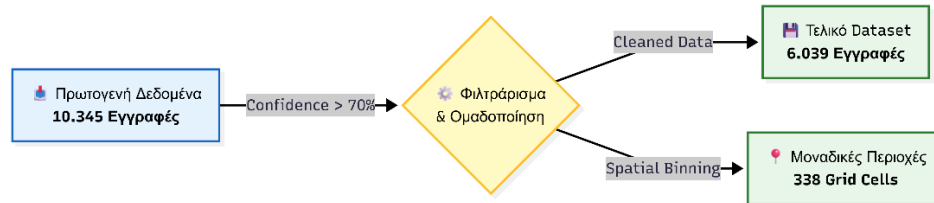
Η επιλογή του βήματος 0.2° αντιστοιχεί σε χωρική ανάλυση ~ 20km. Έγινε και δοκιμή και με βήμα 0.1° αλλά υπήρξε μεγάλη καθυστέρηση στην λήψη των καιρικών δεδομένων για όλα τα σημεία. [34]

Η διαδικασία ολοκληρώνετε με την εξαγωγή του αρχείου **unique_regions_for_weather.csv** που περιέχει ένα σύνολο από 338 εγγραφές, που η κάθε γραμμή του αντιπροσωπεύει την ονομασία του Grid και τις συντεταγμένες (lat,lon) με το κέντρο του κάθε grid cell. Είναι καθοριστικό για την λειτουργία του **Script 2** που θα περιγράψουμε στην επόμενη παράγραφο. Στον **Πίνακα 3.3** παρουσιάζουμε ενδεικτικά τις πρώτες 3 γραμμές του πίνακα.

Πίνακας 3.3: Πίνακας με ID των grid cells και οι συντεταγμένες του κέντρου του grid cell

admin2	lat	lon
Grid_34.8_25.6	34.872	25.6958
Grid_34.8_25.8	34.8708	25.704
Grid_35.0_24.8	34.956	24.888

Στην **Εικόνα 3.2** παρουσιάζουμε την ροή επεξεργασίας του αρχικού μας πίνακα δεδομένων. Σε αυτό φαίνεται ότι από το σύνολο των **10.345** αρχικών εγγραφών, διατηρήθηκαν οι **6.039** υψηλής αξιοπιστίας (Confidence $\geq 70\%$), οι οποίες μετατράπηκαν σε **338** εγγραφές με τα μοναδικά χωρικά κελιά (Grid Cells).



Εικόνα 3.2: Ροή επεξεργασίας δεδομένων ενεργών πυρκαγιών

Τέλος κατασκευάσαμε και ένα βοηθητικό Script (**1_00_update_fire_data_nasa.py**) το οποίο ανανεώνει τον αρχικό πίνακα (modis_2020_2025.csv) με νέα δεδομένα ενεργών πυρκαγιών (αυτόματα από το NASA FIRMS Api [35]), ώστε να μην χρειάζεται να κάνουμε συνέχεια Request στην πλατφόρμα του Copernicus EMS.

3.2.2 Συλλογή Ιστορικών Μετεωρολογικών δεδομένων (Script 2)

Σε αυτή την ενότητα θα περιγράψουμε την διαδικασία ανάκτησης μετεωρολογικών δεδομένων για κάθε ένα Grid Cell ξεχωριστά που απαιτούνται για την εκπαίδευση των μοντέλων. Το σκεπτικό της διαδικασίας αυτής ήταν να πάρουμε συγκεκριμένες μεταβλητές ποια κάθε ένα ξεχωριστά για όλη την πενταετία του αρχικού συνόλου δεδομένων. Παρακάτω θα αναλύσουμε και τη διαδικασία δημιουργίας Αρνητικών Δειγμάτων (Negative Sampling) μέσα στα ήδη υπάρχοντα Grid Cells. Έτσι λοιπόν, αφού έχουμε ανακτήσει δεδομένα για όλες τις ημέρες κάθε έτους της τελευταίας πενταετίας (2020-Σήμερα), Δεν θα χρειαστεί να τρέξουμε ξεχωριστά ξανά το script ανάκτησης ιστορικών μετεωρολογικών δεδομένων και για τα αρνητικά δείγματα (No-Fires). Η ανάκτηση των μετεωρολογικών δεδομένων γίνεται με διασύνδεση του Api της Open-Meteo [36].

3.2.2.1 Ορισμός χρονικού παραθύρου ανάκτησης μετεωρολογικών δεδομένων και επιλογή Μετεωρολογικών Παραμέτρων

Αρχικά ορίζουμε το χρονικό εύρος για το οποίο θα ανακτήσουμε τα δεδομένα. Αυτό γίνεται με δυναμικό τρόπο, περνώντας ουσιαστικά την πρώτη ημερομηνία του αρχικού dataset και ορίζεται σαν μεταβλητή (START_DATE_FIXED). Στη συνέχεια σαν ημερομηνία λήξης της διαδικασίας ορίζουμε την χθεσινή και την ονομάζουμε στον κώδικα σαν (END_DATE_FIXED). Αυτό γίνεται για την διασφάλιση της ακεραιότητας των δεδομένων, διότι για να νικήσουμε αργότερα τους δείκτες του FWI χρειάζεται ημερήσια καταγραφή (24 ώρες) όλων των δεικτών ώστε να παρθεί σωστά ο μέσος όρος της ημέρας. Παρακάτω στο κομμάτι **Κώδικα 3.3** παρουσιάζεται ενδεικτικά η παραπάνω διαδικασία.

```
# Ορισμός της πρώτης ημερομηνίας με βάση τη στήλη 'acq_at'
first_date = df_date['acq_at'].min()
START_DATE_FIXED = first_date.strftime('%Y-%m-%d')

# Ορισμός της τελευταίας ημερομηνίας
end_date_obj = datetime.now() - timedelta(days=1)
END_DATE_FIXED = end_date_obj.strftime("%Y-%m-%d")
```

Κώδικας 3.3: Ορισμός πρώτης και τελευταίας ημερομηνίας ανάκτησης μετεωρολογικών δεδομένων

Με τον παραπάνω τρόπο εξασφαλίζουμε ότι τα δεδομένα μας θα είναι πάντα ενημερωμένα.

Στο επόμενο κομμάτι του κώδικα φαίνεται οι επιλογές συγκεκριμένων μεταβλητών κατά την κλήση στο api (**Κώδικας 3.4**).

```
# Οι μεταβλητές που ζητάμε από το API
HOURLY_VARS = "temperature_2m,relative_humidity_2m,windspeed_10m,precipitation"
API = "https://archive-api.open-meteo.com/v1/archive"
```

Κώδικας 3.4: Ορισμός μεταβλητών από Open-Meteo Api

Η επιλογή των συγκεκριμένων παραμέτρων δεν έγινε καθόλου τυχαία. Είναι ακριβώς οι 4 παράμετροι που απαιτούνται για τον υπολογισμό των υπόδεικτων του Καναδικού Συστήματος Κινδύνου Πυρκαγιάς (Canadian Forest Fire Weather Index System - CFFDRS) ώστε να υπολογιστεί ο τελικός και πιο καθοριστικός δείκτης που είναι ο FWI. Οι 4 αυτοί μετεωρολογικοί παράγοντες ουσιαστικά επηρεάζουν το πόσο εύφλεκτα μπορούν να γίνουν τα υλικά της δασικής καύσιμης ύλης, καθώς και πόσο γρήγορα μπορεί να εξαπλωθεί μια φωτιά. Οι 4 μετεωρολογικοί δείκτες που ανακτούνται από το Api της Open – Meteo είναι:

- **Θερμοκρασία Αέρα (temperature_2m)**
- **Σχετική Υγρασία (relative_humidity_2m)**
- **Ταχύτητα Ανέμου (windspeed_10m)**
- **Βροχόπτωση (precipitation):** Εδώ παίρνουμε στην συνολική βροχόπτωση όλου του 24ώρου και με αυτό υπολογίζουμε την μέση τιμή.

Η “καρδιά” όμως της διαδικασίας αυτής (Script 2), παρουσιάζεται στις επόμενες 2 συναρτήσεις.

- **fetch_full_history_safe**
- **hourly_to_daily_agg**

3.2.2.2 Διαδικασία Διασύνδεσης και Λήψης Δεδομένων από το Open-Meteo

Στην παράγραφο αυτή θα αναλύσουμε τους περιορισμούς καθώς και τις λύσεις που βρέθηκαν για τη διαδικασία ανάκτησης μετεωρολογικών δεικτών.

Κατά το πειραματικό στάδιο, διαπιστώσαμε ότι υπήρχε σοβαρό πρόβλημα στην ανάκτηση μετεωρολογικών δεδομένων για ένα τεράστιο dataset πριν η διαδικασία της χωρικής ομαδοποίησης. Έτσι λοιπόν εφαρμόσαμε έναν αλγόριθμο ο οποίος θα έκανε όσες κλήσεις επιτρέπει το σύστημα και από κει και πέρα βάζουμε το σύστημα σε αδράνεια για κάποια δευτερόλεπτα ώστε να ξεκινήσει και

πάλι τις κλήσεις για τις επόμενες περιοχές. Η συνάρτηση **fetch_full_history_safe** έκανε όλη αυτή τη δουλειά. Παρακάτω θα παρουσιάσουμε βήμα βήμα τον αλγόριθμο της συνάρτησης αυτής.

- **Βήμα 1ο:** Αρχικά δημιουργούμε ένα λεξικό με τις απαραίτητες ρυθμίσεις που θα χρειαστεί για την κλήση στο Api. Αυτές είναι το χρονικό παράθυρο, οι συντεταγμένες του σημείου και οι μετεωρολογικοί δείκτες που θα χρειαστούμε.
- **Βήμα 2ο:** Για να εξασφαλίσουμε την ανθεκτικότητα του συστήματος και την εγκυρότητα των δεδομένων, έχουμε ορίσει μια μεταβλητή την `max_retries = 10`, η οποία σε περίπτωση αποτυχία σύνδεσης του script με το api, ο βρόχος της διαδικασίας εκτελείται 10 φορές.
- **Βήμα 3ο:** Άλλο ένα σημαντικό πρόβλημα που αντιμετωπίσαμε, ήταν η υπερφόρτωση του διακομιστή λόγω πολλών αιτημάτων το οποίο οδηγούσε σε σφάλμα 429 (Too Many Requests). Μια τυπική πρακτική είναι η χρήση τυχαίας καθυστέρησης, γνωστή ως Jitter, μεταξύ των διαδοχικών αιτημάτων. Η τεχνική αυτή αποτρέπει το φαινόμενο «thundering herd», Σταματώντας τον συγχρονισμό των κλήσεων, αποτρέποντάς την ενεργοποίηση των μηχανισμών περιορισμού του Api [8]. Πιο συγκεκριμένα, πριν από κάθε κλήση το σύστημα περιμένει τυχαία 2 και 4 δευτερόλεπτα χρησιμοποιώντας τυχαία κατανομή `time.sleep(random.uniform(2, 4))`.
- **Βήμα 4ο:** Εφόσον το αίτημα της κλήσης είναι επιτυχής τα δεδομένα που λαμβάνουμε είναι σε μορφή JSON. Έτσι λοιπόν με τη βοήθεια της βιβλιοθήκης Pandas μετατρέπουμε τα δεδομένα σε ένα δομημένο DataFrame προσθέτοντας την ημερομηνία και το όνομα της περιοχής.

3.2.2.3 Προετοιμασία δεδομένων για υπολογισμό του FWI

Στο σημείο αυτό και αφού έχουμε πάρει τα μετεωρολογικά δεδομένα μέσω της επικοινωνίας με το Api της Open-Meteo, η συνάρτηση **hourly_to_daily_agg** (Κώδικας 3.5) αναλαμβάνει να μετατρέψει τα ημερήσια δεδομένα (24 μετρήσεις – 1 ανά ώρα) Ναι για μια συγκεκριμένη γεωγραφική περιοχή, σε μια εγγραφή για κάθε ημέρα ώστε να συμβαδίζουμε στις συγκεκριμένες μετρήσεις αναφοράς για τον υπολογισμό των δεικτών του FWI.

Για να το πετύχουμε αυτό και με βάση τον κανόνα του CFFDRS :

- Για θερμοκρασία, υγρασία και ταχύτητα ανέμου κρατάμε μόνο την τιμή που προέκυψε στις 12:00 το μεσημέρι (12η εγγραφή - `iloc[12]`). Η χρονική στιγμή αυτή θεωρείται διεθνώς ως πρότυπο αναφοράς, καθώς αντιπροσωπεύει Οι συνθήκες μέγιστης ηλιακής έκθεσης, με την ελάχιστη σχετική υγρασία που αυτό έχει σαν αποτέλεσμα τη μέγιστη ευφλεκτότητα της δασικής καύσιμης ύλης.
- Για την τιμή της βροχής διαφοροποιείται λίγο η κατάσταση. Στην περίπτωση αυτή, μας ενδιαφέρει το συνολικό ύψος νετού για το 24ωρο (συνολική ποσότητα νερού), καθορίζοντας με αυτό τον τρόπο το πόσο εύφλεκτη θα είναι η δασική καύσιμη ύλη.

Αξίζει να σημειωθεί ότι επιλέξαμε να έχουμε και μια δικλείδα ασφαλείας στις την περίπτωση έλλειψης δεδομένων όπου η ωριαία μέτρηση δεν διαθέτει την δωδέκατη εγγραφή. Σε μια τέτοια περίπτωση ο αλγόριθμος λειτουργεί παίρνοντας το χειρότερο σενάριο :

- Για τη θερμοκρασία παίρνουμε τη μέγιστη ημερήσια τιμή
- Για τη σχετική υγρασία παίρνουμε την ελάχιστη ημερήσια τιμή
- και για την ταχύτητα του ανέμου παίρνουμε το μέσο όρο όλης της ημέρας

```
def hourly_to_daily_agg(hr_df):
    """
    Μετατροπή των ωριαίων μετρήσεων σε μια εγγραφή για τον υπολογισμό του FWI.
    """
    daily = hr_df.groupby(["admin2", "date"]).agg(
        # Παίρνουμε την 12η εγγραφή (12:00 με βάση το FWI)
        temp_noon=("temperature_2m", lambda x: x.iloc[12] if len(x) > 12 else x.max()),
        temp_max=("temperature_2m", "max"),
        rh_noon=("relative_humidity_2m", lambda x: x.iloc[12] if len(x) > 12 else x.min()),
        wind_noon=("windspeed_10m", lambda x: x.iloc[12] if len(x) > 12 else x.mean()),
        rain_24h=("precipitation", "sum")
    ).reset_index()
    return daily
```

Κώδικας 3.5: Επιλογή μεταβλητών από Open-Meteo Api

Η διαδικασία ολοκληρώνεται με τα κύρια συνάρτηση (**main**) που καλείται να συντονίσει Όλη τη διαδικασία της λήψης των δεδομένων από το εξωτερικό Api, βάζοντας στη διαδικασία σε σωστή σειρά.

- Πριν την έναρξη, το σύστημα ελέγχει τις εγγραφές του αρχείου εξόδου και εντοπίζει για ποιες περιοχές έχει αντλήσει μετεωρολογικά δεδομένα. Συγκρίνοντας με το αρχείο εισόδου δημιουργείται μια λίστα με τις περιοχές οι οποίες δεν έχουν δεδομένα (`to_download`), ώστε το σύστημα να ενεργοποιηθεί μόνο για αυτές τις περιοχές εξοικονομώντας χρόνο και υπολογιστική ισχύ.
- Στη συνέχεια τα δεδομένα Αφού επεξεργαστούν, αποθηκεύονται στην προσωρινή για κάθε ξεχωριστό Grid Space και στο τέλος της διαδικασίας προστίθεται στο τελικό αρχείο (`weather_history_by_region.csv`) το οποίο αποτελεί απαραίτητη βάση για τη διαδικασία για την επίλυση των μαθηματικών εξισώσεων του συστήματος FWI (FFMC, DMC, DC, ISI, BUI). Στο αρχείο αυτό, που με τα τωρινά δεδομένα απαριθμεί 756.827 εγγραφές, τα δεδομένα είναι πλήρως ενημερωμένα για κάθε περιοχή και ενδεικτικά η μορφή του φαίνεται στον Πίνακα 3.4.

Πίνακας 3.4: Δομή του Αρχείου Ιστορικών Μετεωρολογικών Δεδομένων ανά Grid Cell

admin2	date	temp_noon	temp_max	rh_noon	wind_noon	rain_24h
GRC.1.1_1	1/1/2020	10.5	11.1	68	21.3	0.1
GRC.1.1_1	2/1/2020	8.5	9.1	73	32.0	5.7
GRC.1.1_1	3/1/2020	9.1	9.5	70	24.1	0
GRC.1.1_1	4/1/2020	12.7	12.9	42	10.6	0
GRC.1.1_1	5/1/2020	13.6	14.1	69	8.4	5.4

3.3 Δημιουργία Συνόλου Εκπαίδευσης

Έχοντας λοιπόν όλα τα ιστορικά δεδομένα καιρού για όλες τις εγγεγραμμένες πυρκαγιές της τελευταίας πενταετίας, η παρούσα ενότητα εστιάζει στη δημιουργία του τελικού συνόλου δεδομένων (dataset) με τη δημιουργία των χαρακτηριστικών εισόδου (input features) στα μοντέλα πρόβλεψης. Η διαδικασία αυτή χωρίζεται σε 2 κόμματα. Το **πρώτο** αφορά τη δημιουργία αρνητικών εγγραφών (negative sampling) Για την παραγωγή δειγμάτων “No Fire”, εξετάζοντας περιπτώσεις αναλογίας κλάσεων 1:1, 1:4, 1:10 και 1:200. Το δεύτερο αφορά τον υπολογισμό του αλγόριθμου των δεικτών του συστήματος FWI (FFMC, DMC, DC, ISI, BUI), συγχωνεύοντάς τους με τα χωρικά δεδομένα, παράγοντας με αυτό τον τρόπο το τελικό dataset εκπαίδευσης.

3.3.1.1 Δημιουργία Αρνητικών Δειγμάτων (Negative Sampling) κλάσεων (Fire vs No-Fire). Περιπτώσεις 1:1, 1:4, 1:10 και 1:200.

Σε αυτό το στάδιο ήρθαμε αντιμέτωποι με το εξής παράδοξο. Όλα τα δεδομένα που έχουμε συλλέξει, διαμορφώσει, αφορούν αποκλειστικά και μόνο περιπτώσεις που αφορούν αποπληκτικά δεδομένα πυρκαγιών. Αν συνεχίζαμε έτσι και τροφοδοτούσαμε τα μοντέλα με αυτά τα δεδομένα, οι προβλέψεις θα ήταν συνεχώς «Fire», αφού δεν θα είχε εκπαιδευτεί ποτέ σε δεδομένα «No Fire».

Λαμβάνοντας υπόψη την σχετική σπανιότητα των περιστατικών πυρκαγιές σε σχέση με τα περιστατικά μη πυρκαγιές, προκύπτει μια ανισορροπία στα δεδομένα. Για τη διερεύνηση της επίδρασης αυτής αποφασίστηκε πειραματικά να αξιολογήσουμε διάφορες ανισορροπίες ώστε να καταλήξουμε στη βέλτιστη λύση μέγιστης γενίκευσης, ώστε να αποφύγουμε την υπερ-προσαρμογή στη μία ή την άλλη κλάση. Τα αποτελέσματα θα τα παρουσιάσουμε κατά την εκπαίδευση των μοντέλων.

Για να λύσουμε το πρόβλημα αυτό, δημιουργήσαμε τεχνητά **Αρνητικά Δείγματα (Negative Samples)**. Ουσιαστικά αυτά είναι δείγματα τα οποία αντιπροσωπεύουν απουσία πυρκαγιές (No Fire). Η διαδικασία αυτή ακολουθεί συγκεκριμένα βήματα τα οποία τα εξηγούμε παρακάτω:

- **Βήμα 1ο :** Γίνονται τυχαίες επιλογές στον χάρτη, για τυχαίες ημερομηνίες μέσα στην πενταετία (2020 – Σήμερα), για τα οποία δεν υπήρχε καταγραφή πυρκαγιές (1 ή Fire). Σε αυτές τις περιπτώσεις δώσαμε την σήμανση No Fire (0). Επιπλέον και πάλι με την μέθοδο της τυχαίας επιλογής, παράγουμε και τιμή στον δείκτη βλάστησης (NDVI), για τον συνοχή και τον ρεαλισμό των δειγμάτων (**Κώδικας 3.6**).

```
# Τυχαία επιλογή περιοχής (admin2) και ημερομηνίας
rand_admin = random.choice(unique_admins)
rand_date = np.random.choice(dates_pool)

# Δημιουργία της εγγραφής με σήμανση "No Fire" (0)
negative_sample = {
    "admin2": rand_admin,
    "acq_at": pd.to_datetime(rand_date),
    "confidence": 0, # Μηδενική βεβαιότητα
    "target": 0     # <--- Η ετικέτα της κλάσης (No Fire)
}
```

Κώδικας 3.6: Δειματοληψία Αρνητικών Δειγμάτων

- **Βήμα 2ο:** Ένα σημείο που μας προβλημάτισε πολύ, ήταν ο καθορισμός της βέλτιστης αναλογίας κλάσεων. Αποφασίσαμε να δώσουμε αρκετές εγγραφές «No Fire» (1:4), ώστε τα μοντέλα κατά την εκπαίδευση «να καταλάβουν» ότι δεν είναι φυσιολογικό το να καίγεται το δάσος. Ταυτόχρονα όμως δώσαμε και αρκετές περιπτώσεις «Fire» ώστε να μπορεί να διακρίνει το σύστημα τις μικρές διαφορές στις καιρικές συνθήκες που μετατρέπουν μια απλή καλοκαιρινή μέρα σε μέρα υψηλού κινδύνου.

Στον **Κώδικα 3.7** παρουσιάζουμε τον υπολογισμό των δειγμάτων και στον **Πίνακα 3.5** παρουσιάζουμε το πλήθος των δειγμάτων με βάση την αναλογία.

```
# Επιλογή αναλογίας 1:4 (1 Πυρκαγιά προς 4 Μη-Πυρκαγιές)
RATIO = 4

# Υπολογισμός του συνολικού πλήθους αρνητικών δειγμάτων
# με βάση το πλήθος των πραγματικών πυρκαγιών
num_negatives = len(df_pos) * RATIO
```

Κώδικας 3.7: Υπολογισμός του πλήθους των αρνητικών δειγμάτων.

Πίνακας 3.5: Δείγματα με βάση την αναλογία

Αναλογία	Φωτιές (1)	Μη-Φωτιές (0)	Σύνολο Δειγμάτων
1	6041	6041	12082
4	6041	24164	30205
10	6041	60410	66451
100	6041	604100	610141
200	6041	1208200	1214241

Τέλος, τα παραγόμενα αρνητικά δείγματα ενσωματώνονται με τις πραγματικές εγγραφές, και το dataset ταξινομείται χρονικά, δημιουργώντας dataset έτοιμο για το επόμενο στάδιο, αυτό του υπολογισμού των δεικτών του FWI.

3.3.2 Υπολογισμός Δεικτών FWI και δημιουργία τελικού Dataset (Script 4).

Σε αυτό το στάδιο, μπαίνουμε στην «καρδιά» του προβλήματος, υπολογίζοντας του επιμέρους δείκτες επικινδυνότητας πυρκαγιάς με βάση του Καναδικό Σύστημα Fire Weather Index (CFFDRS). Στόχος του αλγορίθμου είναι η μετατροπή ιστορικών μετεωρολογικών δεδομένων ανά περιοχή και ανά ημέρα, σε δείκτες επικινδυνότητας. Η διαδικασία βασίστηκε στις εξισώσεις του Van Wagner και οργανώθηκε γύρω από την κλάση **FWI_Instant_Calculator**, η οποία εφαρμόζεται γραμμή – γραμμή στο αρχείο **weather_history_by_region.csv**.

3.3.2.1 Κλάση FWI_Instant_Calculator και Εποχιακή Αρχικοποίηση

Όπως αναφερθήκαμε και Θεωρητικό κομμάτι της εργασίας, οι δείκτες του συστήματος FWI είναι αθροιστικοί. Αυτό σημαίνει πως η σημερινή τιμή για κάθε δείκτη εξαρτάται από την τιμή τις προηγούμενης ημέρας και ούτω καθεξής. Αυτό προϋποθέτει συνεχόμενη χρονοσειρά για κάθε σημείο υπολογισμού. Όμως στην παρούσα εργασία, τα δεδομένα αποτελούνται από χωρικά και χρονικά

διάσπαρτα δεδομένα. Οπότε ο σωστός υπολογισμός των δεικτών θα ήταν αδύνατος. Για την αποφυγή του προβλήματος αυτού σχεδιάστηκε μια στιγμιαία προσέγγιση υπολογισμού με βάση την εποχή όπου οι κώδικες υγρασίας της καύσιμης ύλης (FFMC, DMC, DC) δεν παίρνουν τιμή από την προηγούμενη ημέρα αλλά την αρχική τιμή εκκίνησης (startup value) που δώσαμε με βάση τον μήνα και φυσικά το κλίμα της Ελλάδας [38].

Γίνεται δηλαδή εκτίμηση της κατάστασης του εδάφους ανάλογα με τον μήνα (Κώδικας 3.7):

- Ιούλιος – Αύγουστος (Υψηλός Κίνδυνος): Πολύ ξηρό έδαφος.
- Μάιος, Οκτώβριος: Μέτριες συνθήκες
- Χειμώνας: Υγρό έδαφος.

```
# Στιγμιαία προσέγγιση υπολογισμού με βάση την εποχή
if month in [7, 8]:      # Ιούλιος, Αύγουστος (High Risk)
    prev_ffmc = 92.0    # Πολύ ξηρή λεπτή καύσιμη ύλη
    prev_dmc = 60.0    # Πολύ ξηρό οργανικό στρώμα
    prev_dc = 500.0    # Πολύ ξηρό υπέδαφος
elif month == 6:       # Ιούνιος: Αρχή καλοκαιριού
    prev_ffmc = 88.0
    prev_dmc = 35.0
    prev_dc = 250.0
elif month == 9:       # Σεπτέμβριος (συσσωρευμένη ξηρασία)
    prev_ffmc = 90.0
    prev_dmc = 50.0
    prev_dc = 600.0
elif month in [5, 10]: # Μάιος, Οκτώβριος (Μέτριες Συνθήκες)
    prev_ffmc = 85.0
    prev_dmc = 20.0
    prev_dc = 100.0
else:                  # Χειμώνας
    prev_ffmc = 80.0
    prev_dmc = 10.0
    prev_dc = 15.0
```

Κώδικας 3.8: Στιγμιαία προσέγγιση υπολογισμού με βάση την εποχή

3.3.2.2 Υλοποίηση των Εξισώσεων Van Wagner (CFFDRS)

Αφού ορίσαμε τις αρχικές τιμές μπαίνουμε στο στάδιο της επίλυσης των εξισώσεων του Καναδικού Συστήματος (CFFDRS - Van Wagner 1987) ώστε να υπολογιστούν οι 6 βασικοί δείκτες. Οι δείκτες υπολογίζονται αλυσιδωτά με συγκεκριμένη σειρά καθώς ο ένας τροφοδοτεί τον άλλον.

Η σειρά υλοποίησης παρουσιάζεται παρακάτω:

1. **FFMC:** Εκφράζει την υγρασία των υλικών στο πάνω στρώμα του εδάφους (ξερόφυλλα και ξερόχορτα). Βασικοί παράγοντες του δείκτη αυτού είναι η ποσότητα της βροχής, η ταχύτητα του ανέμου και η υγρασία.

Πολύ σημαντικός παράγοντας στον υπολογισμό του δείκτη αυτού, είναι η μεταβλητή **mo**, η οποία αντιπροσωπεύει την περιεκτικότητα της υγρασίας (%) στο έδαφος. Πιο συγκεκριμένα η μεταβλητή **mo** λειτουργεί σαν αλυσίδα μεταξύ των μαθηματικών πράξεων και της φυσικής. Δεδομένου ότι ο δείκτης FFMC είναι απλώς μια εννοιολογική αξιολόγηση κινδύνου, δεν επιτρέπει την άμεση εφαρμογή φυσικών μεγεθών όπως χλιοστά βροχής, ταχύτητα ανέμου. Έτσι με βάση τη μεταβλητή **mo**

υπολογίζουμε πρώτα το πραγματικό ποσοστό νερού που έχει το έδαφος αυτή τη στιγμή. Και με βάση το ποσοστό αυτό γίνονται οι φυσικοί υπολογισμοί και το τελικό αποτέλεσμα αυτού μεταφράζεται σαν δείκτης FFMC.

Στον **Κώδικας 3.9** παρουσιάζονται ενδεικτικά τα βήματα υλοποίησης σε κώδικα.

```
# Μετατροπή δείκτη σε φυσική υγρασία (%)
mo = (147.2 * (101.0 - prev_ffmc)) / (59.5 + prev_ffmc)

# Εφαρμογή Φυσικών Φαινομένων
if rain > 0.5: # Αν έβρεξε...
    mo += ... # (Εφαρμογή εξίσωσης)

# Υπολογισμός της τελικής υγρασίας 'm' καθώς αυτή μειώνεται (εξατμίζεται) τείνοντας
m = ed + (mo - ed) * (10.0**(-kd))

# Μετατροπή τελικής υγρασίας πίσω σε δείκτη FFMC
ffmc = (59.5 * (250.0 - m)) / (147.2 + m)
```

Κώδικας 3.9: Υπολογισμός Δείκτη FFMC (εκδικητικός κώδικας)

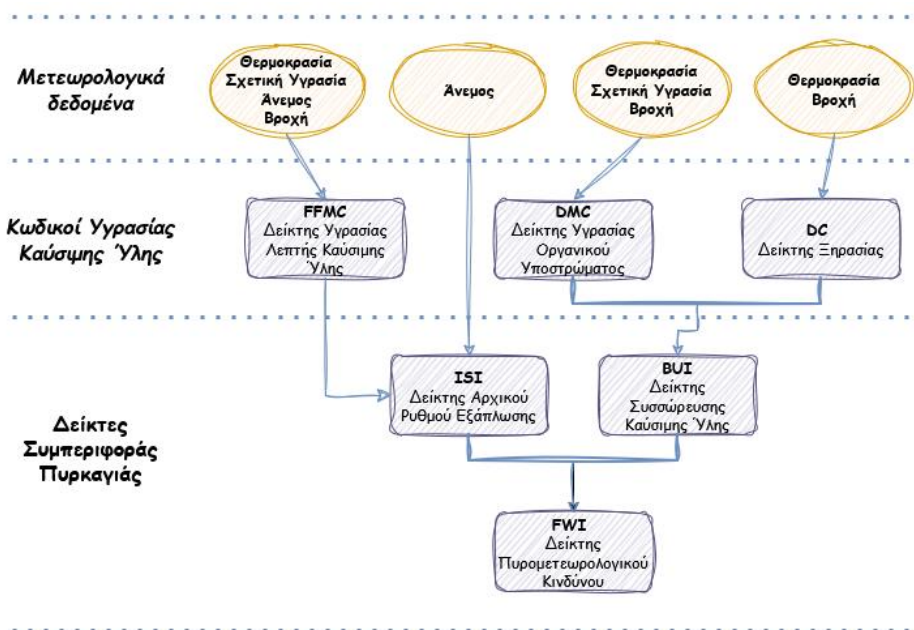
2. Οι δείκτες **DMC** και **DC** υπολογίζουν την ξηρασία που εφαρμόζεται σε βαθύτερα στρώματα του εδάφους όπως αυτά είναι ρίζες ή η σάρκα των φύλλων. Και εδώ οι δείκτες αυτοί βασίζονται στη θερμοκρασία της προηγούμενης κατάστασης.
3. Μπαίνοντας στους δείκτες συμπεριφοράς της πυρκαγιάς, πρώτος στην αλυσίδα είναι ο δείκτης **ISI**. Συνδυάζει τον δείκτη **FFMC** και την ταχύτητα του ανέμου υπολογίζοντας το πόσο εύκολα μπορεί να εξαπλωθεί μια φωτιά.
4. Ο τελευταίος δείκτης πριν τον τελικό υπολογισμό του δείκτη στόχου **FWI**, είναι ο **BUI**. Για τον υπολογισμό του δείκτη αυτού χρειαζόμαστε τις τιμές των δεικτών **DMC** και **DC** Και υπολογίζει την ποσότητα καύσιμης ύλης που είναι διαθέσιμη.
5. Ο τελικός Δείκτης **FWI** προκύπτει από έναν συνδυασμό με τη χρήση του αρχικού δείκτη εξαπλώσεως (**ISI**) και τον πολλαπλασιασμό με μια μη γραμμική συνάρτηση του **BUI**, δίνοντας ένα ενδιάμεσο μέγεθος έντασης **bb**. Για χαμηλές και μέτριες τιμές του **BUI** χρησιμοποιείται μια πολυωνυμική σχέση. Για εξαιρετικά υψηλές τιμές, εφαρμόζεται εκθετική εξομάλυνση προκειμένου να περιοριστεί η ανεξέλεγκτη αύξηση της έντασης. Αυτός ο στη συνέχεια μετασχηματίζεται λογαριθμικά και εκθετικά σε αυτό που γίνεται ο τελικός Δείκτης **FWI**, έτσι ώστε να έχει καλύτερη κλίμακα. Μια κλίμακα την οποία όταν υπάρχει υψηλή τιμή, σημαίνει υψηλή ένταση άρα μεγαλύτερος κίνδυνος. Τέλος στο απόσπασμα του **Κώδικα 3.10** παρουσιάζεται ο υπολογισμός του **FWI**.

```
# Υπολογισμός FWI
if bui <= 80.0:
    bb = 0.1 * isi * (0.626 * bui**0.809 + 2.0)
else:
    bb = 0.1 * isi * (1000.0 / (25.0 + 108.64 * math.exp(-0.023 * bui)))
    if bb <= 1.0: fwi_val = bb
    else: fwi_val = math.exp(2.72 * (0.434 * math.log(bb))**0.647)
```

Κώδικας 3.10: Υπολογισμός Δείκτη FWI

Η διαδικασία ολοκληρώνεται κάνοντας μαζική εφαρμογή του αλγορίθμου σε ολόκληρο το αρχείο που περιέχει τα ιστορικά των μετεωρολογικών δεδομένων, παράγοντας τους δείκτες κινδύνου για κάθε ημέρα και περιοχή. Το κρίσιμο στάδιο σε αυτό το κομμάτι ήταν η ενοποίηση του αρχείου συμβάντων (fire / no fire) Με τους υπολογισμένους δείκτες του FWI, Χρησιμοποιώντας σαν κοινά πεδία την τοποθεσία και την ημερομηνία. Μέσω των δυο παραπάνω πεδίων κάθε εγγραφή εμπλουτίστηκε με τα απαραίτητα δεδομένα που προέκυψαν από τα προηγούμενα scripts. Τέλος τα δεδομένα αυτά αποθηκεύονται σε ένα τελικό αρχείο (**dataset_final.csv**), το οποίο αποτελείται πλέον από τις μεταβλητές εισόδου που θα χρησιμοποιήσουμε παρακάτω κατά την εκπαίδευση των μοντέλων.

Για μεγαλύτερη κατανόηση της ροής υπολογισμού των δεικτών του FWI, Στην **Εικόνα 3.3** παρουσιάζουμε σχηματικά τους παράγοντες που εξαρτώνται ο κάθε δείκτης ξεχωριστά.



Εικόνα 3.3: Ροή υπολογισμού δεικτών FWI

και η τελική μορφή του τελικού dataset παρουσιάζεται στον **Πίνακα 3.6**

Πίνακας 3.6: Τελικό dataset

lat	lon	acq_at	admin2	ndvi	target	temp_noon	temp_max	rh_noon	wind_noon	rain_24h	FFMC	DMC	DC	ISI	BUI	FWI
40.878	24.0478	2020-01-01 00:00:00	Grid_40.8_24.0	0.207	0	4.2	4.9	33	6.6	0.0	84.4	10.0	17.1	2.7	9.8	2.7
36.8466	25.8963	2020-01-01 00:00:00	Grid_36.8_25.8	0.175	0	10.9	11.2	47	28.4	0.0	84.7	10.1	20.4	8.5	10.0	8.8
38.9979	22.9227	2020-01-01 00:00:00	Grid_39.0_23.0	0.23	0	9.9	10.2	48	22.8	2.6	64.0	10.1	19.9	1.6	9.9	1.0
40.4923	23.1061	2020-01-01 00:00:00	Grid_40.4_23.2	0.25	0	7.2	9.2	55	17.8	0.0	82.8	10.0	18.6	3.9	9.9	4.1
40.5404	21.0547	2020-01-01 00:00:00	Grid_40.6_21.0	0.182	0	4.8	5.7	40	8.7	0.0	83.8	10.0	17.4	2.8	9.9	2.8

Κεφάλαιο 4ο: Φάση 2 - Ανάπτυξη & Αξιολόγηση Μοντέλων

4.1 Διαχωρισμός Προβλήματος

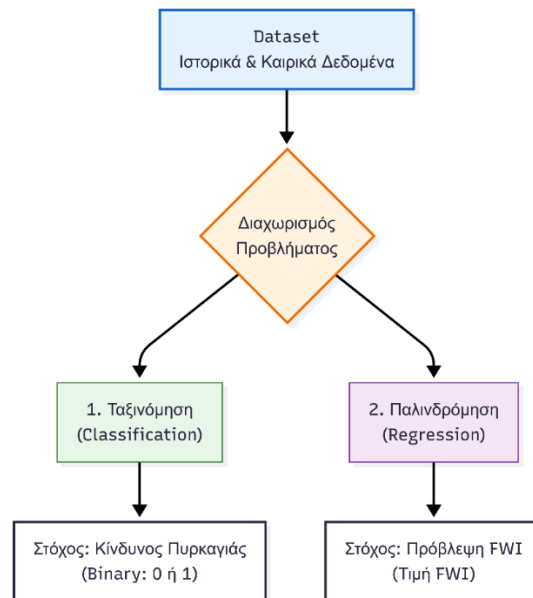
Το κεφάλαιο αυτό σηματοδοτεί την μετάβαση από το στάδιο της προετοιμασίας και οργάνωσης των δεδομένων, στην φάση της αξιοποίησης και ανάπτυξης προγνωστικών μοντέλων. Αφού έχουμε δημιουργήσει, μέσω των μεθοδολογιών που αναφέρθηκαν στο προηγούμενο κεφάλαιο, ένα καθαρό και δομημένο dataset, το οποίο συνδυάζει ιστορικά δεδομένα πυρκαγιών, καιρικές συνθήκες αυτών καθώς και τους υπολογισμένους δείκτες κινδύνου (FWI), το επόμενο στάδιο είναι η εκπαίδευση και αξιολόγηση μοντέλων μηχανικής μάθησης (ML) και βαθιάς μάθησης (DL).

Η καινοτομία της εργασίας αυτής είναι ο πειραματισμός σχεδιασμού σε 2 άξονες επίλυσης.

1. Την **Ταξινόμηση (Classification)**, έχοντας σαν τιμή στόχο την δυαδική πιθανότητα ενδείξεις πυρκαγιάς (1: Fire / 0: No Fire)
2. Την **Παλινδρόμηση (Regression)**, όπου εδώ η τιμή στόχος είναι η πρόβλεψη της τιμής επικινδυνότητας FWI.

Για κάθε μία από τις δυο προσεγγίσεις, συγκρίναμε διάφορες κατηγορίες αλγορίθμων. Κύριος στόχος δεν ήταν μόνο η επίτευξη υψηλών μετρικών απόδοσης για κάθε προσέγγιση, αλλά η επιλογή του καλύτερου μοντέλου προκειμένου να ενσωματωθεί στο τελικό σύστημα πρόβλεψης δασικών πυρκαγιών.

Τη σχηματική απεικόνιση των παραπάνω προσεγγίσεων την παρουσιάζουμε σχηματικά στην **Εικόνα 4.1**.

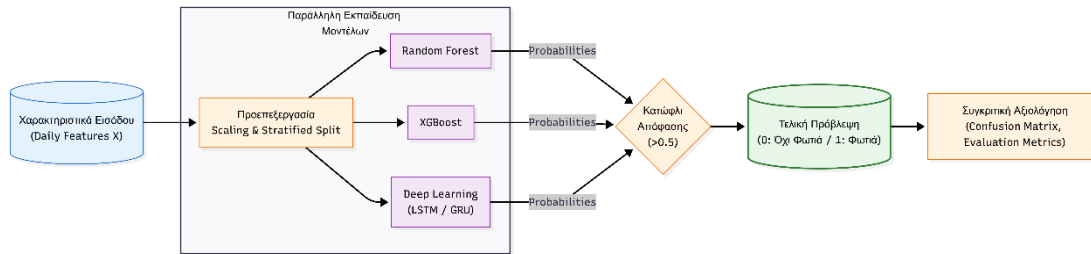


Εικόνα 4.1: Διαχωρισμός προβλήματος

4.1.1 Ταξινόμηση (Classification Pipeline)

Στην πρώτη περίπτωση εστιάζουμε στην επίλυση του προβλήματος της δυαδικής ταξινόμησης (1: Fire / 0: No Fire). Βασικός μας στόχος ήταν η επιλογή προγνωστικών μοντέλων τα οποία θα είναι ικανά να διακρίνουν ημέρες με υψηλή πιθανότητα εκδηλώσεις (target=1) πυρκαγιάς από τις ημέρες με μικρό κίνδυνο εκδηλώσεις πυρκαγιάς (target=0). Η διάκριση αυτή θα γίνει αποκλειστικά και μόνο με βάση

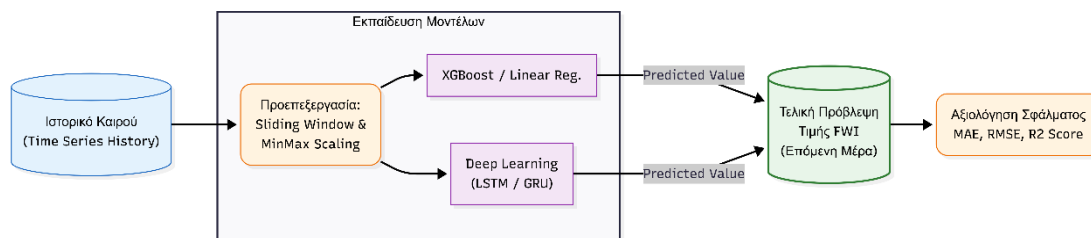
τους μετεωρολογικούς δείκτες καθώς και τους δείκτες του CFFDRS. Για να γίνει πιο κατανοητή η προσέγγιση αυτή, στη **Εικόνα 4.2** παρουσιάζουμε τη ροή εργασίας που περιλαμβάνει προεπεξεργασία, εκπαίδευση των τεσσάρων αλγορίθμων που επιλέχτηκαν και αξιολόγηση των αποτελεσμάτων τους με βάση τις **Μετρικές Αξιολόγησης (Evaluation Metrics) AUC ROC, Accuracy, F1-Score, Precision, Recall**.



Εικόνα 4.2: Αρχιτεκτονική πειράματος Ταξινόμησης

4.1.2 Παλινδρόμηση (Regression Pipeline)

Στην δεύτερη περίπτωση, έχουμε διαφορετική προσέγγιση. Εδώ θα ασχοληθούμε με το πρόβλημα της Παλινδρόμησης (Regression). Ο βασικός στόχος στην προσέγγιση αυτή, είναι η πρόβλεψη του δείκτη FWI της επόμενης ημέρας. Όπως αναφέραμε και παραπάνω, η πρόβλεψη αυτή βασίζεται στην χρονική αλληλουχία των καιρικών δεδομένων και των επιμέρους δεικτών του CFFDRS (ιστορικότητα 10 προηγούμενων ημερών). Για περισσότερη κατανόηση της διαδικασίας, στην **Εικόνα 4.3** παρουσιάζουμε την ροή εργασίας. Η διαδικασία περιλαμβάνει την δημιουργία χρονοσειρών (Sliding Window), την εκπαίδευση των μοντέλων και την εξαγωγή αποτελεσμάτων.



Εικόνα 4.3: Αρχιτεκτονική πειράματος Παλινδρόμησης

4.1.3 Επιλογή και Εκπαίδευση Αλγορίθμων

Μετά από εκτενή έρευνα στη βιβλιογραφία, επιλέξαμε να συγκρίνουμε τέσσερις (4) διαφορετικούς αλγόριθμους που εκπροσωπούν δυο διαφορετικές τεχνολογικές προσεγγίσεις.

Από τη μία πλευρά, δοκιμάσαμε σύγχρονα Νευρωνικά Δίκτυα, τα οποία έχουν την ικανότητα να εντοπίζουν σύνθετα μοτίβα. Από την άλλη, χρησιμοποιήσαμε κλασικές μεθόδους Μηχανικής Μάθησης που βασίζονται στα Δέντρα Αποφάσεων. Ουσιαστικά, κληθήκαμε να απαντήσουμε στο δίλημμα: Αξίζει το αυξημένο υπολογιστικό κόστος της Βαθιάς Μάθησης για να πετύχουμε μεγαλύτερη αξιοπιστία, ή μήπως οι αλγόριθμοι δέντρων αποφάσεων, που είναι σαφώς γρηγορότεροι, επαρκούν για το πρόβλημά μας;

Πιο συγκεκριμένα, οι αλγόριθμοι που εξετάστηκαν είναι οι εξής:

- **Random Forest** και **XGBoost** (Machine Learning).
- **LSTM** και **GRU** (Deep Learning).

4.1.4 Δεδομένα Εισόδου (Input Features)

Για την εκπαίδευση των μοντέλων, το τελικό σύνολο δεδομένων (dataset_final.csv) διαχωρίστηκε σε μεταβλητές εισόδου (X) και μεταβλητή στόχους (y). Ύστερα από πολλούς που πειραματισμούς επιλέχθηκε η προσέγγιση που εστιάζει αποκλειστικά στην φυσική συμπεριφορά της πυρκαγιάς ανεξαρτήτως γεωγραφικής θέσης. Ως εκ τούτου αφαιρέσαμε τις χωρικές πληροφορίες (lat / lon) Κι έτσι το διάνυσμα εισόδου αποτελείται από **10 χαρακτηριστικά** τα οποία χωρίζονται σε δυο κατηγορίες:

1. Μετεωρολογικά Δεδομένα:

- temp_noon: Θερμοκρασία στις 12:00 (°C).
- rh_noon: Σχετική Υγρασία (%).
- wind_noon: Ταχύτητα ανέμου (km/h).
- rain_24h: Βροχόπτωση τελευταίου 24ώρου (mm).

2. Δείκτες Επικινδυνότητας FWI: Περιλαμβάνουν τους 6 υπολογισμένους δείκτες του CFFDRS:

- FFMC, DMC, DC: Κώδικες υγρασίας καύσιμης ύλης.
- ISI, BUI: Δείκτες συμπεριφοράς πυρκαγιάς.
- FWI: Ο τελικός Δείκτης Καιρού Πυρκαγιάς.

Κάνοντας χρήση του δείκτη FWI ως είσοδο, Επιτρέπει στα μοντέλα να αξιολογήσει κατά πόσο ο μαθηματικός τύπος του συστήματος συσχετίζεται επιτυχώς με τα πραγματικά περιστατικά πυρκαγιών στην Ελλάδα.

Σημειώνεται ότι αφαιρέθηκαν βοηθητικές μεταβλητές (όπως το admin2 ή δείκτες βλάστησης τύπου pdnι), προκειμένου το μοντέλο να παραμείνει γενικεύσιμο και να βασίζεται στη φυσική συμπεριφορά του καιρού και των δεικτών FWI.

Έτσι λοιπόν ενδεικτικά στον **Πίνακα 4.1** εμφανίζονται τα χαρακτηριστικά εισόδου των προγνωστικών μοντέλων.

Πίνακας 4.1: Δείγματα Χαρακτηριστικών Εισόδου (Features)

Temp	RH (%)	Wind	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Target
4.2	33	6.6	0.0	84.4	10.0	17.1	2.7	9.8	2.7	0
10.9	47	28.4	0.0	84.7	10.1	20.4	8.5	10.0	8.8	0
9.9	48	22.8	2.6	64.0	10.1	19.9	1.6	9.9	1.0	0
7.2	55	17.8	0.0	82.8	10.0	18.6	3.9	9.9	4.1	0
...

```
# FEATURES: Οι μεταβλητές εισόδου (Input Variables).
FEATURES = ['temp_noon', 'rh_noon', 'wind_noon', 'rain_24h',
            'FFMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI']
```

Κώδικας 4.1: Μεταβλητές Εισόδου

4.1.5 Στρατηγικές Προεπεξεργασίας και Κανονικοποίηση (Scaling)

Για την σωστή εκπαίδευση των μοντέλων, η ποιότητα, καθώς και η μορφή των δεδομένων εισόδου αποτελούν τον σημαντικότερο παράγοντα για την επιτυχία. Ιδιαίτερη προσοχή δόθηκε στην προεπεξεργασία των δεδομένων για τα μοντέλα DL (Βαθιάς Μάθησης), τα οποία είναι ιδιαίτερα ευαίσθητα στην δομή την πληροφορίας εισόδου.

4.1.5.1 Μετασχηματισμός Χρονοσειρών (Sliding Window)

Σε αντίθεση με τους κλασικούς αλγορίθμους ML (XGBoost, Random Forest) που δέχονται δισδιάστατους πίνακες και εξετάζουν κάθε ημέρα μεμονωμένα ώστε να πάρουν μια απόφαση, τα Νευρωνικά Δίκτυα (LSTMs, GRUs) σε τρισδιάστατη δομή, ενσωματώνοντας και την χρονική πληροφορία (Time Series). Για το λόγο αυτό τα δεδομένα θα πρέπει να μετασχηματιστούν από 2D Array (N_samples, N_features) σε 3D Array (N_samples, N_timesteps, N_features) [39].

Πριν ξεκινήσουμε τον σχεδιασμό των μοντέλων, υλοποιήσαμε **δύο scripts** (ένα για κάθε προσέγγιση αλλά παρόμοια μεταξύ τους – Script 6, 12) προετοιμασίας δεδομένων με την τεχνική του **Sliding Window** που θα παράγει κατάλληλα δεδομένα [40]. Παρότι και οι δύο μοιράζονται κοινές τεχνικές βάσεις, παρουσιάζουν θεμελιώδεις διαφορές στον ορισμό του στόχου πρόβλεψης. Παρακάτω παρουσιάζουμε την κοινή μεθοδολογία, καθώς και τις διαφορές τους.

A. Κοινή Μεθοδολογία: Κανονικοποίηση (Min-Max Scaling)

Και στις δυο περιπτώσεις, εφαρμόστηκε η τεχνική **MinMaxScaler**. Η επιλογή αυτή δεν ήταν τυχαία, καθώς για τα μοντέλα των LSTM είναι επιβεβλημένη η ομαλή λειτουργία των εισόδων του Νευρωνικού Δικτιού. Συγκεκριμένα τα κρυφά επίπεδα των LSTM/GRU βασίζονται σε συναρτήσεις ενεργοποίησης όπως *sigmoid* και η *tanh*, οι οποίες λειτουργούν αποδοτικά μόνο όταν τον εύρος τιμών των εισόδων έχει συγκεκριμένο περιορισμένο διάστημα. Η χρήση λοιπόν του **MinMaxScaler**, συμπιέζει τις τιμές εισόδου στο (0,1), μειώνοντας τον κίνδυνο *vanishing gradient*, οδηγώντας σε σταθερότερη εκπαίδευση σε σχέση με άλλες τεχνικές [41].

Η διαδικασία που ακολουθήθηκε στον κώδικα είναι η εξής:

- **Fit:** Ο αλγόριθμος υπολογίζει το ελάχιστο και το μέγιστο όρο για κάθε χαρακτηριστικό (π.χ. για Θερμοκρασία: 0°C - 35°C) βασιζόμενος σε **όλο** το ιστορικό Μετεωρολογικών δεδομένων.
- **Transform:** Όλες οι τιμές μετασχηματίζονται γραμμικά σε τιμές από [0 έως 1].

Ακολουθεί το κοινό τμήμα του κώδικα που υλοποιεί την κανονικοποίηση:

```

from sklearn.preprocessing import MinMaxScaler
import joblib

# FEATURES: Οι μεταβλητές εισόδου (Input Variables).
FEATURES = ['temp_noon', 'rh_noon', 'wind_noon', 'rain_24h',
            'FFMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI']

# Δημιουργία και εφαρμογή του Scaler σε όλο το ιστορικό
scaler = MinMaxScaler()
df_history[FEATURES] = scaler.fit_transform(df_history[FEATURES])

# Αποθήκευση του Scaler για μελλοντική χρήση σε επόμενα scripts
joblib.dump(scaler, OUT_DIR / "scaler.pkl")

```

Κώδικας 4.2: Κανονικοποίηση Δεδομένων

B. Διαμόρφωση Χρονοσειρών (Sliding Window)

Για να εκμεταλλευτούμε τη μνήμη των δικτύων, υλοποιήσαμε ένα script προετοιμασίας δεδομένων όπου τα δεδομένα μετασχηματίζονται σε ακολουθίες σταθερού μήκους μέσω της τεχνικής του κυλιόμενου παραθύρου (**Sliding Window**) ώστε να παραχθούν δεδομένα κατάλληλα για LSTM/GRU.

Ορίστηκε η παράμετρος `LOOKBACK = 10`, που σημαίνει ότι το μοντέλο λαμβάνει ως είσοδο ένα «παράθυρο» των 10 προηγούμενων ημερών για να εξάγει την τελική πρόβλεψη (Y). Η δομή των δεδομένων εισόδου (X) είναι κοινή και στις δύο περιπτώσεις (`N_samples`, `N_timesteps`, `N_features`)[40].

C. Η Κρίσιμη Διαφορά: Στόχευση (Target Definition)

Στο χρονικό ορίζοντα πρόβλεψης, οι δυο προσεγγίσεις είναι ίδιες, στοχέυοντας την πρόβλεψη εκτίμησης της κατάστασης για την επόμενη μέρα. Η διαφορά στα δυο σενάρια παρουσιάζεται στον ορισμό της μεταβλητής εξόδου. Ποιο συγκεκριμένα στην περίπτωση της **Ταξινόμησης**, έχουμε Risk Warning, παίρνοντας σαν έξοδο την δυαδική τιμή (0: No Fire / Fire), που αντιπροσωπεύει, το αν οι αυριανές συνθήκες ευνοούν την δημιουργία πυρκαγιάς ή όχι.

Ενώ στην **Παλινδρόμηση** εξάγει την ακριβή πρόβλεψη της αυριανής τιμής του δείκτη επικινδυνότητας FWI.

4.1.5.2 Αποτελέσματα μετασχηματισμού

Κατά την εκτέλεση των σεναρίων (Script 5, 10), τα αρχικά δυσδιάστατα δεδομένα, μετασχηματίστηκαν σε τριδιάστατες δομές. Αναλύοντας το αρχείο καταγραφής (**Εικόνα 4.4 και 4.5**), καταλήξαμε στα εξής συμπεράσματα.

των μετεωρολογικών δεδομένων και των δεικτών FWI, μπορούν να προβλέψουν αν η επόμενη ημέρα κρύβει κίνδυνο εκδήλωσης πυρκαγιάς (Target = 1, Fire) ή όχι (Target = 0, No Fire).

Για να το πετύχουμε, επιλέξαμε να δοκιμάσουμε και να συγκρίνουμε δύο διαφορετικές «οικογένειες» μοντέλων. Από τη μία πλευρά χρησιμοποιούμε νευρωνικά δίκτυα τύπου RNN (LSTM και GRU), τα οποία είναι σχεδιασμένα για να εκμεταλλεύονται τη χρονική αλληλουχία των δεδομένων. Από την άλλη, χρησιμοποιούμε αλγορίθμους βασισμένους σε δέντρα αποφάσεων (Random Forest και XGBoost), οι οποίοι μαθαίνουν σύνολα κανόνων απόφασης πάνω στα χαρακτηριστικά εισόδου. Στις επόμενες υποενότητες παρουσιάζουμε αναλυτικά τη διαδικασία εκπαίδευσης και τα αποτελέσματα για κάθε κατηγορία μοντέλου ξεχωριστά.

4.2.1 Μοντέλα Βαθιάς Μάθησης (Deep Learning): Αρχιτεκτονική LSTM και GRU

Σε αυτή την υποενότητα παρουσιάζονται τα μοντέλα Βαθιάς Μάθησης που χρησιμοποιήθηκαν για το πρόβλημα της ταξινόμησης. Εστιάζουμε σε δύο τύπους RNNs, τα LSTM και GRU, τα οποία είναι κατάλληλα όταν τα δεδομένα έχουν τη μορφή χρονοσειράς και υπάρχει ανάγκη να «θυμούνται» πώς εξελίσσεται τα δεδομένα στο χρόνο. Με αυτόν τον τρόπο, τα μοντέλα μπορούν να αξιοποιήσουν όχι μόνο τις τρέχουσες μετεωρολογικές συνθήκες, αλλά και το πώς αυτές άλλαξαν τις προηγούμενες ημέρες.

A. Δομή και Αρχιτεκτονική Δικτύου

Για να γίνει η δίκαιη σύγκριση στα δίκτυα τύπου RNN, φτιάξαμε την ίδια βασική δομή και για τα δυο. Η υλοποίηση πραγματοποιήθηκε με τη χρήση της βιβλιοθήκης **Keras/TensorFlow**.

Όπως βλέπουμε στο απόσπασμα του **Κώδικα 4.3**, το δίκτυο έχει 2 στρώματα. Το πρώτο αποτελείται από **64 νευρώνες** και το δεύτερο από **32 νευρώνες**. Για την αποφυγή του Overfitting βάλουμε **Dropout=30% (0.3)**.

Στο τέλος του δικτύου, οι πληροφορίες από τα LSTM περνούν από ένα **πυκνό στρώμα 16 νευρώνων** με **ReLU**, που λειτουργεί σαν "φίλτρο" συμπυκνώνοντας τις πιο σημαντικές πληροφορίες, και καταλήγουν σε έναν **τελικό νευρώνα με Sigmoid**, ο οποίος βγάζει μια πιθανότητα από **0 έως 1** που απαντά στο ερώτημα "πόσο πιθανό είναι να πιάσει φωτιά αύριο;"

```

model = Sequential([
    # Input Layer
    Input(shape=(X_train.shape[1], X_train.shape[2])),

    # 1st LSTM Layer
    LSTM(64, return_sequences=True), # Το ίδιο και για GRU
    Dropout(0.3),

    # 2nd LSTM Layer
    LSTM(32, return_sequences=False), # Το ίδιο και για GRU
    Dropout(0.3),

    # Dense Layers
    Dense(16, activation='relu'),

    # Output Layer
    Dense(1, activation='sigmoid')
])

```

Κώδικας 4.3: Αρχιτεκτονική LSTM και GRU

B. Παραμετροποίηση και εκπαίδευση

Η διαδικασία εκπαίδευσης παραμετροποιήθηκε με τον αλγόριθμο βελτιστοποίησης **Adam** (Learning Rate: 0.001) και συνάρτηση κόστους την **Binary Crossentropy**, η οποία είναι η ενδεδειγμένη για προβλήματα δυαδικής ταξινόμησης (**Κώδικα 4.4**).

Ιδιαίτερη προσοχή δόθηκε στη διαχείριση της ανισορροπίας των κλάσεων, καθώς τα περιστατικά πυρκαγιάς είναι σπάνια. Έτσι λοιπόν υπολογίστηκαν τα βάρη των κλάσεων, ώστε το μοντέλο να τιμωρεί αυστηρότερα τα λάθη για τα θετικά δείγματα, εξισορροπώντας έτσι τη διαδικασία μάθησης.

```

# Υπολογισμός Βαρών
weights = class_weight.compute_class_weight('balanced', classes=np.unique(y_train), y=y_train)
class_weights = {0: weights[0], 1: weights[1]}

# Compile
model.compile(
    optimizer=Adam(learning_rate=0.001),
    loss='binary_crossentropy',
    metrics=['accuracy', 'AUC', 'Recall']
)

```

Κώδικας 4.4: Υπολογισμός Βαρών Κλάσεων και Ρύθμιση Μοντέλου

Γ. Έλεγχος Εκπαίδευσης (Callbacks)

Τέλος για τη βελτιστοποίηση της σύγκλισης, την αποφυγή άσκοπης χρήσης υπολογιστικής ισχύς και άσκοπων εποχών εκπαίδευσης, Ενσωματώσαμε τους μηχανισμούς **Callbacks**. Συγκεκριμένα χρησιμοποιήσαμε τον EarlyStopping, ώστε η εκπαίδευση να τερματιστεί Δεν παρατηρηθεί βελτίωση στο σφάλμα επικύρωσης (Validation Loss) για 10 διαδοχικές εποχές.

Και το ReduceLRonPlateau για δυναμική προσαρμογή του ρυθμού μάθησης. Πιο συγκεκριμένα μειώνει τον ρυθμό μάθησης στο μισό σε περίπτωση στασιμότητας (**Κώδικα 4.5**).

Τέλος το μοντέλο εκπαιδεύεται μέχρι 50 epochs με batches των 32 δειγμάτων, δίνοντας μεγαλύτερη βαρύτητα στις πυρκαγιές (class_weights).

```
# Callbacks
stopper = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)
reducer = ReduceLRonPlateau(monitor='val_loss', factor=0.5, patience=3, min_lr=0.00001)

# Start Training
history = model.fit(
    X_train, y_train,
    validation_data=(X_test, y_test),
    epochs=50,
    batch_size=32,
    class_weight=class_weights,
    callbacks=[stopper, reducer],
    verbose=1
)
```

Κώδικας 4.5: Έλεγχος Εκπαίδευσης (Callbacks)

Παρ όλα αυτά όμως κατά την εξέλιξη του πειράματος στο **GRU**, παρατηρήσαμε ότι σε κάποιες παραμέτρους έπρεπε να επέμβουμε. Έτσι λοιπόν παρατηρήσαμε ότι η χρήση προεπιλεγμένου ρυθμού μάθησης (0.001), που λειτουργήσε άψογα στο LSTM, οδηγούσε το GRU σε αστάθεια (spikes) κατά την εκπαίδευση. Για αυτό το λόγο αλλάξαμε τον ρυθμό μάθησης του βελτιστοποιεί τη Adam σε **0.0001 (optimizer=Adam(learning_rate=0.0001))**.

Μείωση αυτή επέτρεψε την ομαλότερη σύγκλιση το μοντέλο, αποφεύγοντας τις έντονες διακυμάνσεις. Στον Πίνακα 4.2 συνοψίζονται οι υπερπαραμέτροι που χρησιμοποιήθηκαν.

Πίνακας 4.2: Υπερπαραμέτροι Εκπαίδευσης Μοντέλων LSTM/GRU

Παράμετρος	Τιμή / Ρύθμιση
Αρχιτεκτονική	Stacked (64 Units → 32 Units)
Συνάρτηση Ενεργοποίησης (Hidden)	Tanh (Default for RNNs) / ReLU (Dense)
Συνάρτηση Εξόδου	Sigmoid
Dropout Rate	0.3 (30%)

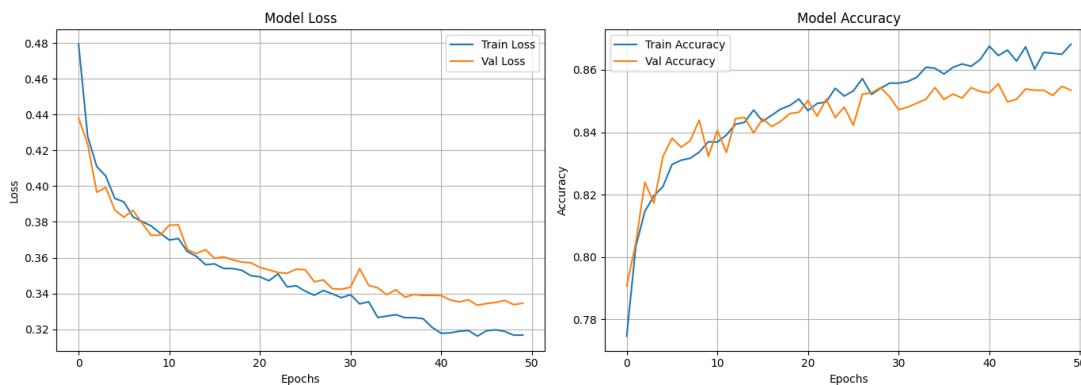
Optimizer	Adam (Learning Rate: 0.001)
Loss Function	Binary Crossentropy
Batch Size	32
Μέγιστες Εποχές (Epochs)	50 (με Early Stopping)

4.2.1.1 Αποτελέσματα Εκπαίδευσης Αρχιτεκτονικής LSTM

Συνεχίζουμε με την ανάδειξη των αποτελεσμάτων ανά αρχιτεκτονική. Σε αυτή την παράγραφο παρουσιαστούν τα αποτελέσματα της αρχιτεκτονικής LSTM. Εδώ θέλουμε να δούμε πόσο σταθερό είναι το μοντέλο μας και πώς συμπεριφέρεται σε πραγματικά δεδομένα ελέγχου, δηλαδή στα **6.017** δείγματα του test set (Περίπτωση 1:4).

A. Σταθερότητα Εκπαίδευσης (Learning Curves)

Η εξέλιξη της εκπαίδευσης αποτυπώνεται στην **Εικόνα 4.6**, όπου παρουσιάζονται οι καμπύλες της συνάρτησης κόστους (Loss) και της ακρίβειας (Accuracy) για 50 εποχές.



Εικόνα 4.6: Καμπύλες Loss και Accuracy κατά την εκπαίδευση του LSTM

Σχετικά με το **Loss**, στην παραπάνω εικόνα παρατηρούμε ότι η καμπύλη της εκπαίδευσης (μπλε) μειώνεται σταθερά, ενώ η καμπύλη επικύρωσης (πορτοκαλι) Ακολουθεί μια πτώση έως περίπου την τριακοστή εποχή, όπου από κει και πέρα παρουσιάζει μια σταθερότητα.

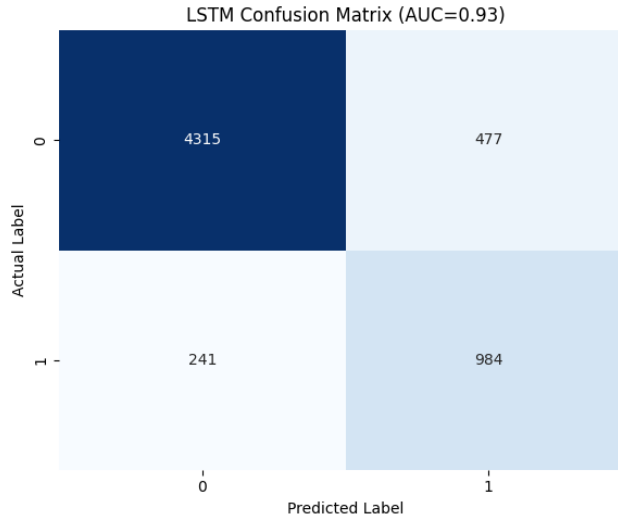
Όσον αφορά την ακρίβεια επικύρωσης (Accuracy) φτάνει περίπου στο 85-86% ακολουθώντας την ακρίβεια εκπαίδευσης.

B. Ανάλυση Σφαλμάτων (Confusion Matrix)

Η λεπτομερής συμπεριφορά του μοντέλου ως προς τα σφάλματα παρουσιάζεται στην **Εικόνα 4.7**.

- **Ανίχνευση Πυρκαγιών (True Positives):** Το μοντέλο κατάφερε να προβλέψει σωστά 984 συμβάντα πυρκαγιών.

- **Χαμένες Πυρκαγιές (False Negatives):** Υπήρξαν και 241 συμβάντα πυρκαγιών τις οποίες το μοντέλο δεν τις ανίχνευσε.
- **Ψευδείς Συναγερμοί (False Positives):** Βέβαια το μοντέλο ανίχνευσε και 477 συμβάντα με λανθασμένο κίνδυνο. Δηλαδή ανίχνευσε πυρκαγιά χωρίς να υπάρχει. Αυτό ήταν αναμενόμενο καθώς η χρήση των βαρών έδωσε προτεραιότητα στην ευαισθησία ακόμα και με κόστος ψευδών συναγερμών.



Εικόνα 4.7: Πίνακας Σύγκρισης του μοντέλου LSTM για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire)

Γ. Μετρικές Απόδοσης

Με βάση τα παραπάνω σχήματα οι τελικές επιδόσεις του μοντέλου διαμορφώνονται με βάση τις μετρικές απόδοσης που φαίνονται στον παρακάτω πίνακα (**Πίνακας 4.3**).

Πίνακας 4.3: Επιδόσεις Μοντέλου LSTM

ROC AUC	Accuracy	F1-Score	Precision (Fire)	Recall (Fire)
92.55%	88.1%	73.3%	67.4%	80.3%

Δ. Συμπέρασμα αποτελεσμάτων

Τα αποτελέσματα κρίνονται ικανοποιητικά για την περίπτωση του 1:4. Έχοντας μια υψηλή τιμή στο **ROC AUC (92.55%)** καταδεικνύει η στιβαρότητα του αλγόριθμου, ενώ το **Recall της τάξης του 80.3%** δείχνει ένα μοντέλο που προβλέπει σε μεγάλο βαθμό σωστά. Το Precision στο 67.4% είναι λίγο χαμηλότερο, αλλά αυτό είναι αποδεκτό, Γιατί σε ένα σύστημα πρόβλεψη δασικών πυρκαγιών είναι πολύ χειρότερο να χάσουμε μια πυρκαγιά (False Negative) παρά να έχουμε άλλον έναν συναγερμό (False Positive).

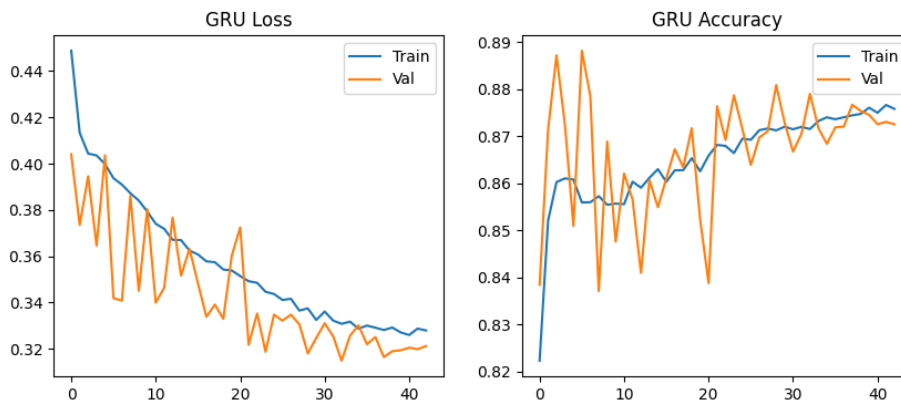
4.2.1.2 Αποτελέσματα Αρχιτεκτονικής GRU

Όπως αναφερθήκαμε και παραπάνω, στην περίπτωση του αλγορίθμου GRU χρειάστηκε να αλλάξουμε το Learning Rate από 0.001 σε 0.0001.

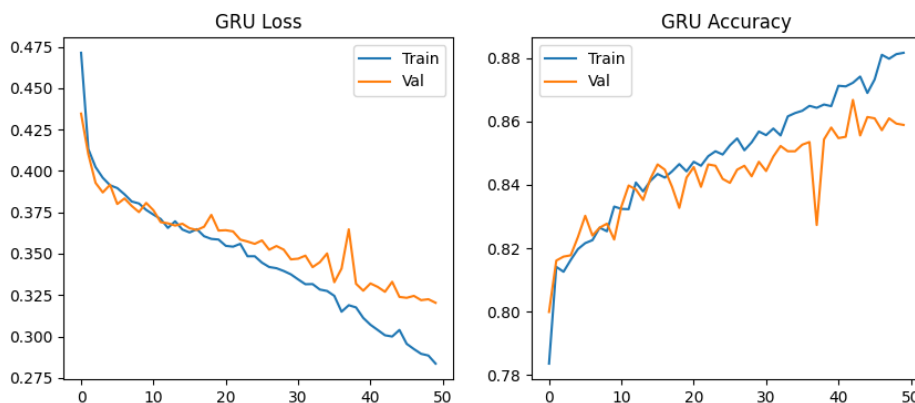
A. Σταθερότητα Εκπαίδευσης (Learning Curves)

Έτσι στο αρχικό μας πείραμα επιχειρήσαμε να εκπαιδέσουμε το δίκτυο με τις προεπιλεγμένες ρυθμίσεις ($lr=0.001$). Ωστόσο με τον συγκεκριμένο ρυθμό μάθησης η διαδικασία εμφάνιζε έντονη αστάθεια. Αυτό το φαινόμενο υποδηλώνει ότι το βήμα ενημέρωσης των βαρών ήταν πολύ μεγάλο, εμποδίζοντας το μοντέλο να "κατέβει" ομαλά στο ολικό ελάχιστο (**Εικόνα 4.8**).

Έτσι λοιπόν μειώσαμε το ρυθμό μάθησης σε ($lr=0.0001$). Το αποτέλεσμα όπως φαίνεται στην **Εικόνα 4.9** ήταν ότι με την αλλαγή αυτή εξομαλυνθεί και η καμπύλη με σταθερή σύγκλιση του μοντέλου χωρίς βίαιες αποκλίσεις.



Εικόνα 4.8: Καμπύλες Loss και Accuracy κατά την εκπαίδευση του GRU ($lr=0.001$)

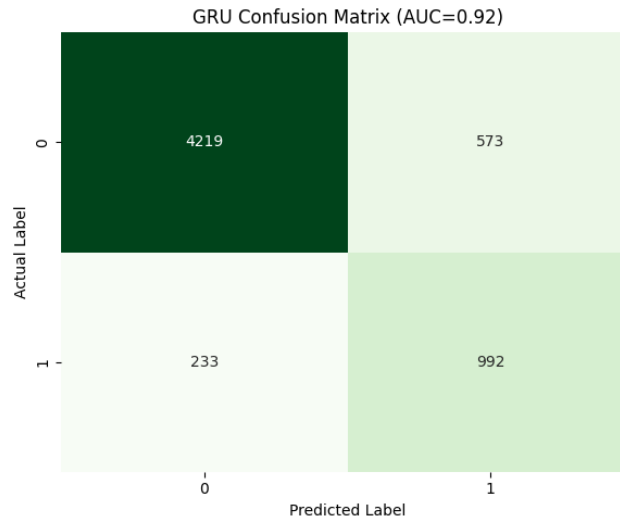


Εικόνα 4.9: Καμπύλες Loss και Accuracy κατά την εκπαίδευση του GRU ($lr=0.0001$)

B. Ανάλυση Σφαλμάτων (Confusion Matrix)

Η λεπτομερής συμπεριφορά του μοντέλου ως προς τα σφάλματα παρουσιάζεται στην **Εικόνα 4.10**.

- **Ανίχνευση Πυρκαγιών (True Positives):** Το μοντέλο κατάφερε να προβλέψει σωστά 992 συμβάντα πυρκαγιών.
- **Χαμένες Πυρκαγιές (False Negatives):** Υπήρξαν και 233 συμβάντα πυρκαγιών τις οποίες το μοντέλο δεν τις ανίχνευσε.
- **Ψευδείς Συναγερμοί (False Positives):** Ως αποτέλεσμα της αυξημένης ευαισθησίας, παρατηρήθηκε αύξηση των ψευδών συναγερμών (**573**), δεν επηρέασε αρνητικά τη συνολική ακρίβεια.



Εικόνα 4.10: Πίνακας Σύγχυσης του μοντέλου GRU για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire)

Γ. Μετρικές Απόδοσης

Με βάση τα παραπάνω σχήματα οι τελικές επιδόσεις του μοντέλου διαμορφώνονται με βάση τις μετρικές απόδοσης που φαίνονται στον παρακάτω πίνακα (**Πίνακας 4.4**).

Πίνακας 4.4: Επιδόσεις Μοντέλου GRU

ROC AUC	Accuracy	F1-Score	Precision (Fire)	Recall (Fire)
92.84%	87.9%	73.5%	66.3%	82.4%

Δ. Συμπεράσματα Αποτελεσμάτων

Τα αποτελέσματα κρίνονται ιδιαίτερα ενθαρρυντικά για την αξιοποίηση της αρχιτεκτονικής GRU. Η υψηλή τιμή του **Recall (82.4%)** αποδεικνύει ότι το μοντέλο είναι εξαιρετικά ευαίσθητο στην αναγνώριση των συνθηκών που οδηγούν σε πυρκαγιά, ελαχιστοποιώντας τον κίνδυνο μη ανίχνευσης (False Negatives). Παράλληλα, η διατήρηση του **ROC AUC σε επίπεδα άνω του 92%** πιστοποιεί τη σταθερότητα και τη στιβαρότητα του αλγορίθμου σε διάφορες μετεωρολογικές συνθήκες.

4.2.1.3 Ανάλυση Σημαντικότητας

Τα νευρωνικά δίκτυα LSTM/GRU, σε αντίθεση με τα μοντέλα με δέντρα αποφάσεων που δίνουν έτοιμες μετρήσεις σημαντικότητας χαρακτηριστικών, λειτουργούν σαν «μαύρα κουτιά». Για μπορέσουμε να εξάγουμε τα `features importance` για δίκτυα LSTM/GRU, εφαρμόσαμε την τεχνική **Permutation Importance** [42].

Η λογική είναι απλή. Μετράμε την αρχική απόδοση του μοντέλου. Για κάθε χαρακτηριστικό εισόδου ανακατεύουμε τυχαία τις τιμές του σπάζοντας τη σύνθεση του. Ξανά μετράμε την απόδοση και αν η πτώση στο σκορ είναι μεγάλη σημαίνει ότι το χαρακτηριστικό είναι σημαντικό για το μοντέλο αν η απόδοση παραμείνει η ίδια ή είναι μικρότερη τότε το χαρακτηριστικό είναι αδιάφορο για το μοντέλο [42].

Παρακάτω στο κομμάτι του **Κώδικα 4.6** παρουσιάζεται η μέθοδος αυτή.

```
def calculate_importance(model, X_test, y_test, model_name):
    # 1. Baseline Score
    y_pred_base = model.predict(X_test, verbose=0)
    baseline_auc = roc_auc_score(y_test, y_pred_base)

    # 2. Permutation Loop
    for i, feature_name in enumerate(FEATURE_NAMES):
        X_test_shuffled = X_test.copy()

        # Τυχαία αναδιάταξη (καταστροφή πληροφορίας) της στήλης i
        np.random.shuffle(X_test_shuffled[:, :, i])

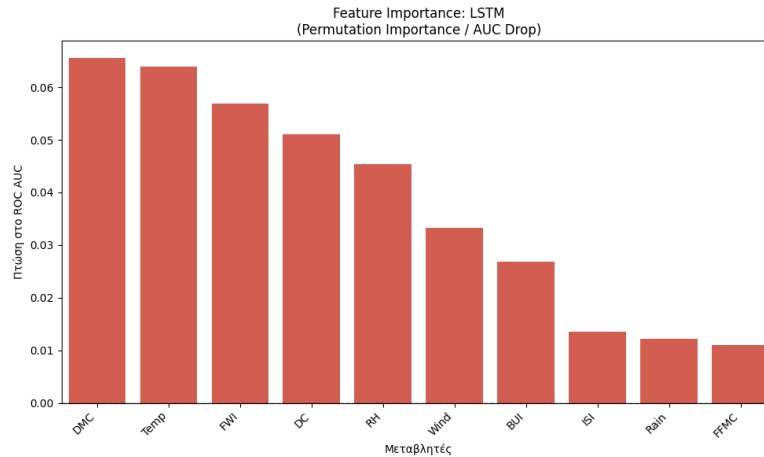
        # Πρόβλεψη
        y_pred_shuffled = model.predict(X_test_shuffled, verbose=0)
        shuffled_auc = roc_auc_score(y_test, y_pred_shuffled)

        # Υπολογισμός Πτώσης
        drop = baseline_auc - shuffled_auc
        if drop < 0: drop = 0

        importances[feature_name] = drop # Αποθήκευση
```

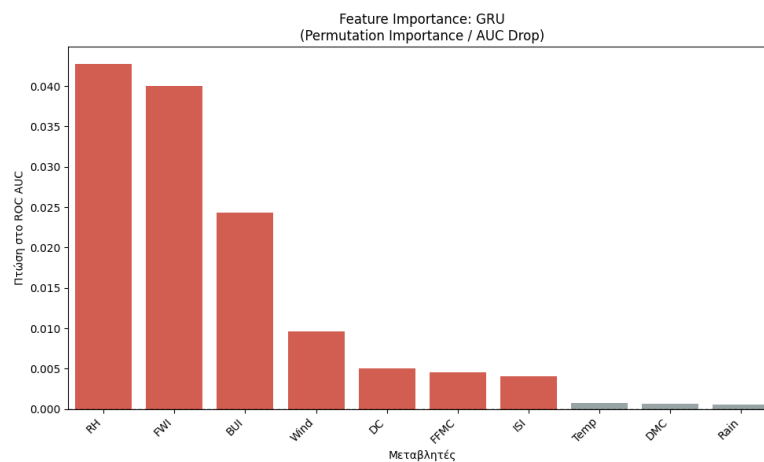
Κώδικας 4.6: Υλοποίηση Permutation Importance

Τα αποτελέσματα παρουσιάζονται σε ραβδογράμματα (**Εικόνα 4.11** και **Εικόνα 4.12**), επιτρέποντας **άμεση σύγκριση** μεταξύ μετεωρολογικών δεικτών και FWI στην τελική απόφαση του δικτύου [42].



Εικόνα 4.11: Features Importance του μοντέλου LSTM

Το παραπάνω διάγραμμα ανέδειξε για το μοντέλο LSTM, το δείκτη **DMC** (Duff Moisture Code) και τη **Θερμοκρασία** (Temp) ως τις πλέον καθοριστικές μεταβλητές για την πρόβλεψη. Το εύρημα αυτό υποδηλώνει ότι το δίκτυο «έμαθε» να δίνει προτεραιότητα στα επίπεδα υγρασίας της νεκρής οργανικής ύλης σε συνδυασμό με τις τοπικές θερμικές εκδήξεις, αξιολογώντας τες ως πιο κρίσιμες ακόμη και από τον σύνθετο δείκτη FWI.



Εικόνα 4.12: Features Importance του μοντέλου GRU

Στην άλλη περίπτωση, του μοντέλου GRU, η ιεράρχηση των μεταβλητών διαφοροποιείται αισθητά, αναδεικνύοντας τη **Σχετική Υγρασία (RH)** ως τον πλέον καθοριστικό παράγοντα, με τον δείκτη **FWI** να ακολουθεί. Η συμπεριφορά αυτή υποδηλώνει ότι ο συγκεκριμένος αλγόριθμος εστιάζει πρωτίστως στην ξηρασία της ατμόσφαιρας ως τον κύριο ρυθμιστή για την εκτίμηση της πιθανότητας ανάφλεξης.

4.2.2 Μοντέλα Μηχανικής Μάθησης (Machine Learning): Εκπαίδευση XGBoost και Random Forest.

Αφού ολοκληρώσαμε τα πειράματα με τα μοντέλα βαθιάς μάθησης, τώρα σειρά έχουν τα μοντέλα μηχανικής μάθησης και συγκεκριμένα αλγόριθμου συνόλου: το **Random Forest** και το **XGBoost**. Και στις δύο περιπτώσεις δώσαμε ιδιαίτερη προσοχή στη ρύθμιση των υπερπαραμέτρων, ώστε να αντιμετωπίσουμε αποτελεσματικά την ανισορροπία των κλάσεων και να δώσουμε προτεραιότητα στα σπάνια περιστατικά πυρκαγιάς. Στόχος είναι η συγκριτική αξιολόγηση της κλασικής προσέγγισης έναντι των Νευρωνικών Δικτύων.

4.2.2.1 Random Forest Classifier (Bagging)

A. Παραμετροποίηση και Εκπαίδευση

Ο αλγόριθμος **Random Forest** παραμετροποιήθηκε ως ένα σύνολο **300 εκτιμητών** (Δέντρα: `n_estimators=300`), σε σχέση με τα 100 που είναι η default τιμή. Αυτή η προσέγγιση δόθηκε για την εξασφάλιση την μεγαλύτερης σταθερότητας του μοντέλου στις προβλέψεις και την μείωση της διακύμανσης. Για τον έλεγχο της πολυπλοκότητας και την αποφυγή υπερπροσαρμογής, το μέγιστο βάθος κάθε δέντρου περιορίστηκε στα 10 επίπεδα (`max_depth=10`). Ο μηχανισμός αυτός, χρησιμοποιείτε σαν σταθεροποιητής, αποτρέποντας την απομνημόνευση θορύβου στα δεδομένα εκπαίδευσης (Overfitting)

Κρίσιμη για την απόδοση του μοντέλου ήταν η χρήση της παραμέτρου `class_weight='balanced'`. Η ρύθμιση αυτή προσαρμόζει αυτόματα τα βάρη κατά την εκπαίδευση, δίνοντας μεγαλύτερη βαρύτητα στα δείγματα της μειωνηφούσας κλάσης (**1: Φωτιά**), ώστε να αντισταθμιστεί η κυριαρχία των δειγμάτων «**0: Μη-Φωτιάς**».

Στο κομμάτι του κώδικα (**Κώδικας 4.7**) παρακάτω, φαίνονται οι τιμές των υπερπαραμέτρων για τον Random Forest.

```
# -----
# ΒΗΜΑ 2: ΕΚΠΑΙΔΕΥΣΗ RANDOM FOREST
# -----
# Προσθέτουμε class_weight='balanced' για να μην αγνοήσει τις φωτιές!
rf = RandomForestClassifier(
    n_estimators=300,      # 300 δέντρα
    max_depth=10,        # Περιορισμός βάθους 10
    class_weight='balanced', # Ζυγίζει αυτόματα τις κλάσεις
    random_state=42,
    n_jobs=-1            # Χρήση όλων των πυρήνων
)

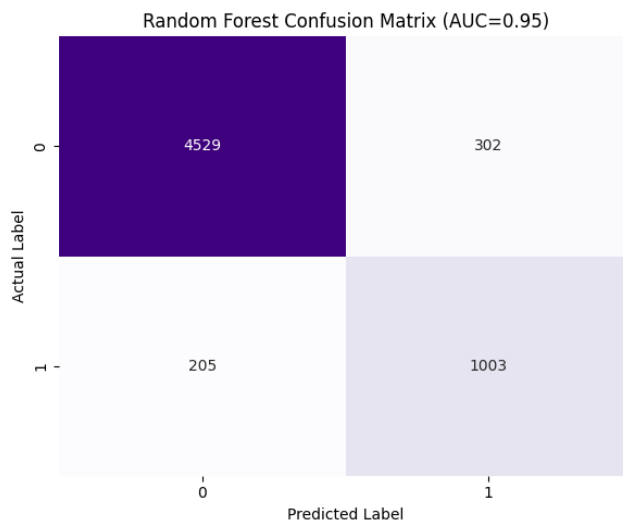
# Training Random Forest
rf.fit(X_train, y_train)
```

Κώδικας 4.7: Παράμετροι Random Forest

B. Ανάλυση Σφαλμάτων (Confusion Matrix)

Η ικανότητα ταξινόμησης του μοντέλου στο σύνολο ελέγχου (Test Set) αποτυπώνεται στον Πίνακα Σύγκρισης (Εικόνα 4.13). Παρατηρούμε ότι το Random Forest πέτυχε μια εξαιρετική ισορροπία μεταξύ ευαισθησίας και ακρίβειας, περιορίζοντας τα σφάλματα και στις δύο κατευθύνσεις. Συγκεκριμένα:

- Ανίχνευση Πυρκαγιών (True Positives): Το μοντέλο κατάφερε να προβλέψει σωστά 1003 συμβάντα πυρκαγιών.
- Χαμένες Πυρκαγιές (False Negatives): Υπήρξαν και 205 συμβάντα πυρκαγιών τις οποίες το μοντέλο δεν τις ανίχνευσε.
- Ψευδείς Συναγερμοί (False Positives): Ο αριθμός των ψευδών συναγερμών (False Positives) περιορίστηκε στους 302, γεγονός που βελτιώνει σημαντικά την τελική ακρίβεια (Precision) συγκριτικά με τις μεθόδους βαθιάς μάθησης.



Εικόνα 4.13: Πίνακας Σύγκρισης του αλγορίθμου Random Forest για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire)

Γ. Μετρικές Απόδοσης

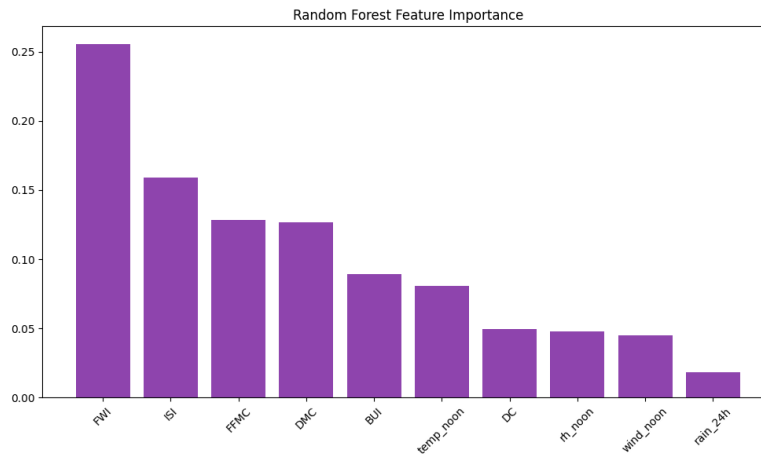
Στον Πίνακα 4.5 παρουσιάζονται συγκεντρωτικά οι επιδόσεις του Random Forest.

Πίνακας 4.5: Επιδόσεις Random Forest

ROC AUC	Accuracy	F1-Score	Precision (Fire)	Recall (Fire)
95.1%	91.6%	79.8%	76.9%	83%

Δ. Σημαντικότητα Χαρακτηριστικών (Feature Importance)

Ένα από τα σημαντικότερα πλεονεκτήματα του Random Forest είναι η ικανότητα ερμηνείας των αποτελεσμάτων. Στο διάγραμμα παρακάτω (Εικόνα 4.14) παρουσιάζεται η σημαντικότητα των μεταβλητών εισόδου για το Random Forest.



Εικόνα 4.14: Σημαντικότητα Χαρακτηριστικών στο Random Forest

Όπως φαίνεται στο διάγραμμα, οι δείκτες **FWI** (Fire Weather Index) και **ISI** (Initial Spread Index) αναδεικνύονται ως οι πλέον καθοριστικοί παράγοντες για την πρόβλεψη. Το γεγονός αυτό επιβεβαιώνει τη φυσική σημασία των δεικτών του συστήματος FWI, καθώς ο δείκτης **FWI** αποτελεί το γενικό δείκτη έντασης της πυρκαγιάς, ενώ ο **ISI** παρουσιάζει την αναμενόμενη ταχύτητα εξάπλωσης, ενσωματώνοντας την επίδραση του ανέμου.

4.2.2.2 XGBoost Classifier (Boosting)

Ο τελευταίος αλγόριθμος ML που επιλέχθηκε είναι ο **XGBoost** (Extreme Gradient Boosting) το οποίο κάνει χρήση την μέθοδο Boosting. Σε αντίθεση με το RF που κατασκευάζει δέντρα ανεξάρτητα, το XGBoost λειτουργεί σειριακά, όπου κάθε νέο δέντρο διορθώνει τα σφάλματα των προηγούμενων.

A. Παραμετροποίηση και Εκπαίδευση

Για την επίτευξη της μέγιστης γενίκευσης, το μοντέλο βασίζεται στην στρατηγική την αργής μάθησης. Οι βασικότερες ρυθμίσεις που εφαρμόστηκαν, όπως παρουσιάζονται στο παράρτημα του **Κώδικα 4.8** είναι:

```

# Υπολογισμός Βαρών (Scale Pos Weight)
neg = len(y_train) - sum(y_train)
pos = sum(y_train)
scale_weight = neg / pos # Δίνει έμφαση στη μειοψηφούσα κλάση

#####
# Ρύθμιση Μοντέλου
xgb = XGBClassifier(
    n_estimators=300,      # Αριθμός Δέντρων
    learning_rate=0.03,   # Ρυθμός μάθησης
    max_depth=6,         # Πόσο περίπλοκο μπορεί να είναι κάθε δέντρο
    scale_pos_weight=scale_weight, # Εφαρμογή της εξισορρόπησης
    random_state=42,
    eval_metric=['logloss', 'error'],
    use_label_encoder=False,
    early_stopping_rounds=10 # Early Stopping
)

#####
# Εκπαίδευση Μοντέλου
xgb.fit(
    X_train, y_train,
    eval_set=[(X_train, y_train), (X_test, y_test)],
    verbose=False
)

```

Κώδικας 4.8: Βασικές Ρυθμίσεις XGBoost

Ορίσαμε **χαμηλό ρυθμό μάθησης (learning_rate=0.03)** για να εξασφαλίσουμε σταδιακή σύγκλιση προς το ελάχιστο της συνάρτησης κόστους.

Επιλέξαμε **300 δέντρα (n_estimators=300)**, καθώς ο χαμηλός ρυθμός μάθησης απαιτεί περισσότερες επαναλήψεις για επιτεύγματα καλής απόδοσης.

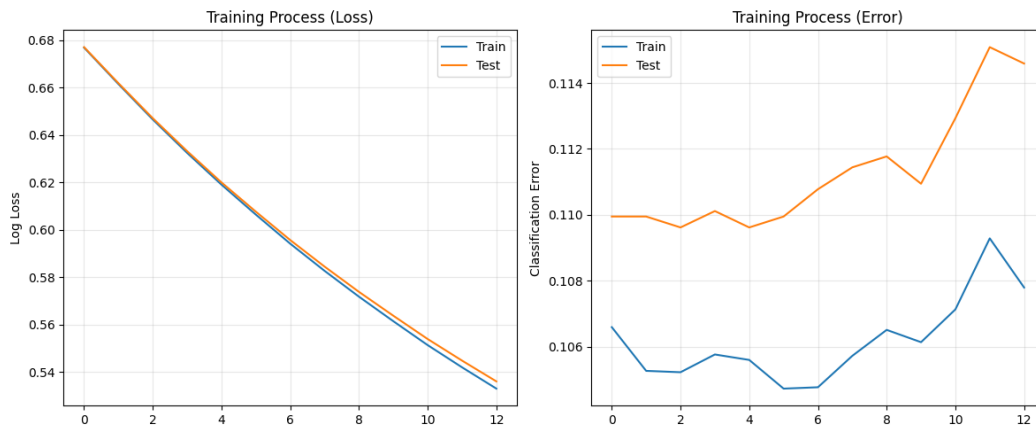
Για την **ανισορροπία των κλάσεων** ρυθμίσαμε δυναμικά την παράμετρο **scale_pos_weight** βάσει της αναλογίας αρνητικών/θετικών δειγμάτων, δίνοντας μεγαλύτερη βαρύτητα στα σπάνια περιστατικά πυρκαγιάς.

Ενεργοποιήσαμε **Early Stopping (early_stopping_rounds=10)**, ώστε η εκπαίδευση να διακόπτεται αυτόματα αν η απόδοση στο validation set δεν βελτιώνεται για 10 διαδοχικές επαναλήψεις, αποτρέποντας την υπερπροσαρμογή.

B. Καμπύλες Μάθησης (Learning Curves)

Η πορεία της εκπαίδευσης και η σύγκλιση του σφάλματος αποτυπώνονται στις καμπύλες μάθησης (Εικόνα 4.15).

Κεφάλαιο 4

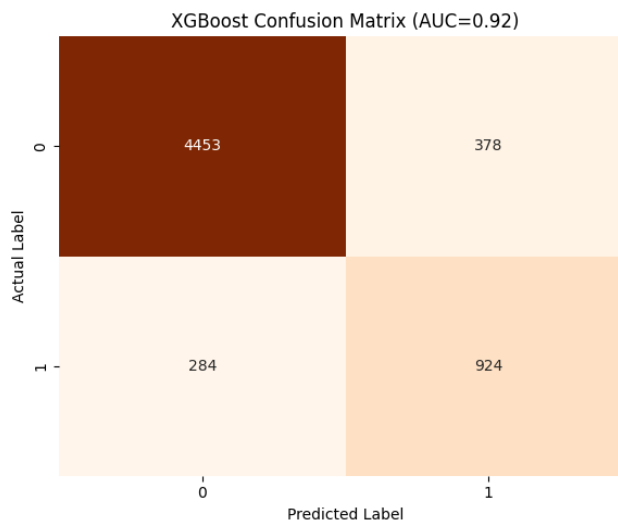


Εικόνα 4.15: Καμπύλες Μάθησης (Learning Curves)

Με βάση το παραπάνω διάγραμμα, επιβεβαιώνεται η σωστή παραμετροποίηση του μοντέλου, αλλά και η αναγκαιότητα του μηχανισμού Early Stopping. Πιο συγκεκριμένα, ενώ η συνάρτηση κόστους (**Log Loss**) παρουσιάζει ομαλή και ταυτόχρονη μείωση και για τα 2 σύνολα (train/test), το σφάλμα ταξινόμησης (**Classification Loss**), εμφανίζει μεγάλες αποκλίσεις (ένδειξη Overfitting). Αυτό έχει σαν αποτέλεσμα την απομνημόνευση θορύβου των δεδομένων εκπαίδευσης. Η ενεργοποίηση του Early Stopping μετά την 12^η εποχή λειτουργεί σαν ασφάλεια, διακόπτοντας την εκπαίδευση σε κρίσιμο σημείο, διασφαλίζοντας την ομαλή γενίκευση του ταξινομητή.

Γ. Ανάλυση Σφαλμάτων (Confusion Matrix)

Η απόδοση του XGBoost στην ταξινόμηση του συνόλου ελέγχου παρουσιάζεται στον Πίνακα Σύγκυσης (Εικόνα 4.16).



Εικόνα 4.16: Πίνακας Σύγκυσης του αλγορίθμου XGBoost για δυαδική ταξινόμηση πυρκαγιάς (Fire / No Fire)

- **Ανίχνευση Πυρκαγιών (True Positives):** Το μοντέλο κατάφερε να προβλέψει σωστά 924 συμβάντα πυρκαγιών.

- **Χαμένες Πυρκαγιές (False Negatives):** Υπήρξαν και 284 συμβάντα πυρκαγιών τις οποίες το μοντέλο δεν τις ανίχνευσε.
- **Ψευδείς Συναγερμοί (False Positives):** Ο αριθμός των ψευδών συναγερμών (False Positives) περιορίστηκε στους 378, γεγονός που βελτιώνει σημαντικά την τελική ακρίβεια (Precision) συγκριτικά με τις μεθόδους βαθιάς μάθησης.

Δ. Μετρικές Απόδοσης

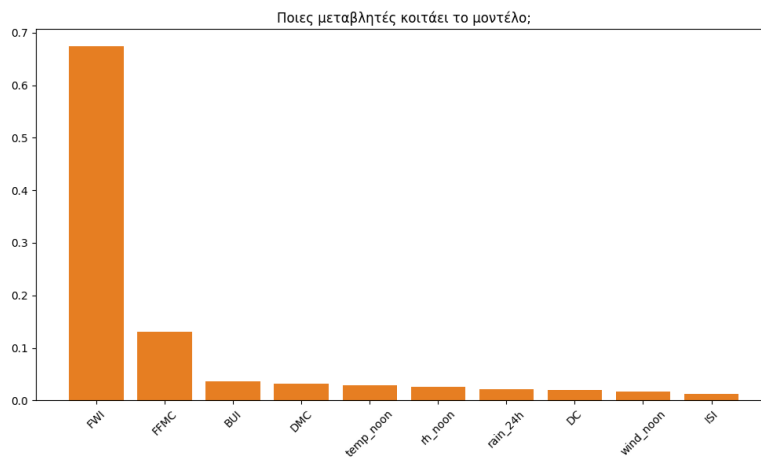
Στον Πίνακα 4.6 παρουσιάζονται συγκεντρωτικά οι επιδόσεις του Random Forest.

Πίνακας 4.6: Επιδόσεις XGBoost

ROC AUC	Accuracy	F1-Score	Precision (Fire)	Recall (Fire)
91.68%	89%	73.6%	71%	76.5%

Ε. Σημαντικότητα Χαρακτηριστικών (Feature Importance)

Παρόμοια με το RF, το XGBoost προσφέρει δυνατότητα ερμηνείας μέσω της αξιολόγησης της συνεισφοράς κάθε χαρακτηριστικού εισόδου στην κατασκευή των δέντρων. Στο διάγραμμα που ακολουθεί (Εικόνα 4.17), αποτυπώνεται η ιεράρχηση των μεταβλητών βάσει του Κέρδους (Gain), αναδεικνύοντας τους παράγοντες που συνέβαλαν καθοριστικά στη βελτιστοποίηση της ακρίβειας του μοντέλου.



Εικόνα 4.17: Σημαντικότητα Χαρακτηριστικών στο XGBoost

Στο παραπάνω διάγραμμα, αποτυπώνεται η κυριαρχία του δείκτη **FWI**, με τον **FFMC** να ακολουθεί σε πολύ χαμηλότερα επίπεδα. Αυτό μας δείχνει ότι ο αλγόριθμος «έμαθε» πως ο δείκτης FWI είναι ο καλύτερος δείκτης πρόβλεψης για την ταξινόμηση, κάνοντας χρήση τον XGBoost.

ΣΤ. Συμπέρασμα για XGBoost

Το XGBoost απέδειξε την ικανότητά του να διαχειρίζεται αποτελεσματικά την ανισορροπία των δεδομένων μέσω της παραμέτρου `scale_pos_weight`. Η χρήση χαμηλού ρυθμού μάθησης οδήγησε σε

ένα στιβαρό μοντέλο, το οποίο πέτυχε την **συνολική ακρίβεια (Accuracy ~89%)**. Ωστόσο, σε αντίθεση με τα Νευρωνικά Δίκτυα που εστίασαν στην ευαισθησία, το XGBoost λειτούργησε πιο συντηρητικά, ελαχιστοποιώντας τους ψευδείς συναγερούς (High Precision) με τίμημα όμως τη χαμηλότερη ικανότητα ανίχνευσης του συνόλου των πυρκαγιών (**Recall ~76.5%**).

4.2.3 Συζήτηση Αποτελεσμάτων και Επιλογή Βέλτιστου Μοντέλου

4.2.3.1 Ανάλυση Επιδόσεων

Στον Πίνακα 4.7 παρουσιάζεται η συγκεντρωτική αξιολόγηση των τεσσάρων αρχιτεκτονικών που αναπτύχθηκαν. Η σύγκριση βασίζεται στις μετρικές ROC AUC, Accuracy, F1-Score, Precision και Recall, οι οποίες αναδεικνύουν τη συμπεριφορά κάθε μοντέλου.

Πίνακας 4.7: Συγκεντρωτική αξιολόγηση Classification

Model	ROC AUC	Accuracy	F1-Score	Precision	Recall
Random Forest	0.9510	91.6%	79.8%	76.9%	83.0%
GRU	0.9284	87.9%	73.5%	66.3%	82.4%
LSTM	0.9255	88.1%	73.3%	67.4%	80.3%
XGBoost	0.9168	89.0%	73.6%	71.0%	76.5%

Από την επισκόπηση των αποτελεσμάτων προκύπτει η ξεκάθαρη υπεροχή του αλγορίθμου **Random Forest**, ο οποίος κατέγραψε τις υψηλότερες επιδόσεις σε όλους τους δείκτες.

Συγκεκριμένα:

- **Συνολική Αξιοπιστία:** Με Accuracy 91.6% και ROC AUC 0.951 (95,1%), το RF απέδειξε την μεγάλη ικανότητα σωστού διαχωρισμού των κλάσεων, με πολύ μεγάλο ποσοστό σωστών προβλέψεων
- **Ισορροπία Ευαισθησίας-Ακρίβειας:** Είναι το μοναδικό μοντέλο που διατήρησε υψηλή ευαισθησία (Recall 83.0%) χωρίς να θυσιάσει την ακρίβεια των συναγερούς (Precision 76.9%).
- **Σύγκριση DL vs ML:** Τα μοντέλα DL (GRU, LSTM), αν και πλησίασαν σε ευαισθησία (Recall ~80-82%), υστέρησαν σημαντικά σε ακρίβεια (Precision ~66-67%), παράγοντας περισσότερους ψευδείς συναγερούς. Το XGBoost, από την άλλη, λειτούργησε πιο συντηρητικά, χάνοντας όμως σημαντικό ποσοστό πυρκαγιών (Recall 76.5%).

4.2.3.2 Επιλογή βέλτιστου μοντέλου

Για την επιλογή του βέλτιστου μοντέλου μπορεί να εξεταστεί με βάση την περίοδο. Θα μπορούσαμε να αναφέρουμε μια σύγκριση μετρικών, χωρίζοντας την επιλογή σε 2 στρατηγικές [43] [45].

A. Αρχή αντιπυρικής περιόδου (Χαμηλό Ρίσκο)

Σε μια περίοδο χαμηλού κινδύνου, όπου η εκδήλωση πυρκαγιών είναι πολύ σπάνια (Μάιος, Οκτώβριος) και το κόστος ψευδούς συναγερούς είναι δυσανάλογο του κινδύνου, το RF (και δεύτερο το XGBoost)

παρουσιάζεται σαν βέλτιστη λύση. Η υψηλή ακρίβεια (76,9% και 71% αντίστοιχα) εγγυάται ότι αν εκπέμψει «πυρκαγιά», η πιθανότητα είναι πολύ μεγάλη. Άρα στην περίπτωση αυτή μπορούμε να λάβουμε υπόψη το **Precision** [43] [45].

B. Μέσα καλοκαιριού (Υψηλό Ρίσκο)

Σε περίοδο υψηλού κινδύνου, η απώλεια έστω και μιας πυρκαγιάς μπορεί να είναι καταστροφική. Άρα η ευαισθησία είναι πολύ σημαντική. Και εδώ το RF παραμένει σαν βέλτιστη λύση με **83%**. Βέβαια πολύ κοντά βρίσκετε και η προσέγγιση του μοντέλου **GRU (82.4%)** που ενδεχόμενος μπορεί να λειτουργήσει σαν δευτερεύουσα επιλογή μοντέλου, καθώς μπορεί να εντοπίσει συσσωρευμένη ξηρασία που οι στατικοί αλγόριθμοι δεν μπορούν να απομνημονεύσουν. Άρα στην περίπτωση αυτή μπορούμε να λάβουμε υπόψη το **Recall**.

Από την άποψη όλων των μετρήσεων, της υπολογιστικής πολυπλοκότητας και της ανάλυσης σπουδαιότητας χαρακτηριστικών, το **Random Forest** απέδειξε να έχει την καλύτερη απόδοση και ως εκ τούτου έχει επιλεγεί ως το βέλτιστο μοντέλο στην περίπτωση της ταξινόμησης. Προσέφερε τη μεγαλύτερη δυνατότητα ανίχνευσης πυρκαγιάς με τους ελάχιστους ψευδής συναγεμμους, αναδεικνύοντας έτσι την πιο αξιόπιστη λύση για σύστημα έγκαιρης προειδοποίησης πυρκαγιών [43] [45].

4.3 Πείραμα 2: Παλινδρόμηση (Regression Pipeline)

Σε αυτή την ενότητα μεταβαίνουμε από το πρόβλημα της διακριτής ταξινόμησης στο πεδίο της παλινδρόμησης (Regression). Στόχος μας πλέον δεν είναι απλώς να εντοπίσουμε αν υπάρχει κίνδυνος πυρκαγιάς, αλλά να προβλέψουμε την **ακριβή αριθμητική τιμή του δείκτη FWI** (Fire Weather Index) για την **επόμενη ημέρα**. Η προσέγγιση αυτή προσφέρει μεγαλύτερη λεπτομέρεια στη λήψη αποφάσεων, καθώς επιτρέπει την εκτίμηση όχι μόνο της **πιθανότητας**, αλλά και της **έντασης** του αναμενόμενου κινδύνου.

Για την υλοποίηση του πειράματος, αξιοποιούμε και προσαρμόζουμε τις αρχιτεκτονικές που αναλύθηκαν προηγουμένως, προσθέτοντας παράλληλα ένα κλασικό στατιστικό μοντέλο αναφοράς. Συγκεκριμένα, συγκρίνουμε τα Νευρωνικά Δίκτυα **LSTM** και **GRU**, τα οποία επιχειρούν να προβλέψουν τη μελλοντική τιμή βάσει της ιστορικής ακολουθίας, με τον αλγόριθμο **XGBoost Regressor**. Επιπλέον, εισάγουμε τη **Γραμμική Παλινδρόμηση (Linear Regression)** ως baseline μοντέλο, προκειμένου να διαπιστώσουμε εάν η πολυπλοκότητα των μη-γραμμικών μεθόδων προσφέρει ουσιαστική βελτίωση στην ακρίβεια της πρόβλεψης έναντι μιας απλούστερης προσέγγισης.

4.3.1 Γραμμική Παλινδρόμηση (Linear Regression)

Ως σημείο αναφοράς (Baseline) για το πείραμα της παλινδρόμησης επιλέχθηκε η **Απλή Γραμμική Παλινδρόμηση**. Σκοπός της χρήσης ενός γραμμικού μοντέλου είναι να διαπιστωθεί εάν η σχέση μεταξύ των ιστορικών μετεωρολογικών δεδομένων και του μελλοντικού δείκτη FWI μπορεί να προσεγγιστεί ικανοποιητικά από μια γραμμική εξίσωση, ή αν απαιτούνται πολυπλοκότερες μη-γραμμικές δομές (όπως τα Νευρωνικά Δίκτυα). Γι αυτό το λόγο επιλέξαμε να το αναλύσουμε πρώτο (Script 14).

A. Υλοποίηση και Μετασχηματισμός Δεδομένων

Η κύρια δυσκολία με τη Γραμμική Παλινδρόμηση (LR) στα δεδομένα χρονοσειρών είναι ότι δέχεται μόνο διδιάστατη είσοδο (2D Arrays), ενώ τα LSTM δεδομένα μας έχουν σχήμα (Samples, 10 ημέρες, 10 χαρακτηριστικά). Γι' αυτό προχωρήσαμε στην διαδικασία της **Επιπεδοποίησης (Flattening)** για κάθε ιστορικό 10 ημερών. Ουσιαστικά τα δεδομένα μετατρέπονται σε ένα διάνυσμα 100 χαρακτήρων στην είσοδο (10 ημέρες, 10 χαρακτηριστικά). Οπότε για κάθε τιμή εισόδου (π.χ. θερμοκρασία ημέρα - 3) γίνεται ξεχωριστή στήλη.

Εκπαιδύσαμε το μοντέλο στα **κανονικοποιημένα** δεδομένα, αλλά για την αξιολόγηση επαναφέραμε τις προβλέψεις FWI στην κανονική τους κλίμακα, ώστε να δούμε τα πραγματικά σφάλματα (**Κώδικα 4.9**).

```
# Μετατροπή από (Samples, 10, 10) σε (Samples, 100)
num_samples, num_timesteps, num_features = X_3d.shape
X_flat = X_3d.reshape(num_samples, num_timesteps * num_features)

# Διαχωρισμός και εκπαίδευση
X_train, X_test, y_train, y_test = train_test_split(
    X_flat, y_scaled, test_size=0.2, random_state=42)

lr = LinearRegression()
lr.fit(X_train, y_train)

# Μετατροπή και αξιολόγηση στις πραγματικές τιμές
y_pred_scaled = lr.predict(X_test)

# Επαναφορά στην πραγματική κλίμακα (Inverse Transform)
y_pred_real = scaler_y.inverse_transform(y_pred_scaled.reshape(-1, 1))
y_test_real = scaler_y.inverse_transform(y_test.reshape(-1, 1))
```

Κώδικας 4.9: Προετοιμασία και Εκπαίδευση Linear Regration

B. Αποτελέσματα LR

Για την αξιολόγηση της ακρίβειας των προβλέψεων, με βάση την βιβλιογραφία, επιλέχθηκαν οι παρακάτω μετρητές:

- Mean Absolute Error (MAE) για την εκτίμηση του μέσου σφάλματος
- Root Mean Squared Error (RMSE) που τιμωρεί αυστηρότερα τις μεγάλες αποκλίσεις
- Συντελεστής R^2 ως μέτρο της συνολικής ερμηνευτικής ικανότητας του μοντέλου.

Έτσι λοιπόν, η απόδοση του μοντέλου LR στο σύνολο ελέγχου (Test Set) καταγράφεται στον **Πίνακα 4.8**.

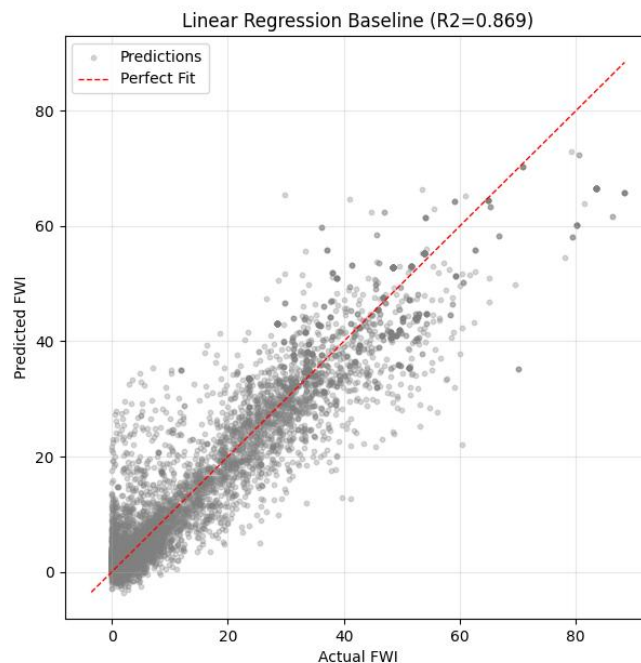
Πίνακας 4.8: Επιδόσεις Linear Regration

R2 - Score	MAE	RMSE
0.8693	4.1096	6.2475

Οι επιδόσεις κρίνονται αξιολογικές για μοντέλο αναφοράς, καθώς ο συντελεστής R^2 (**0.8693**) αποκαλύπτει ότι η σχέση των μετεωρολογικών δεδομένων με τον δείκτη FWI είναι πρωτίστως γραμμική. Το μέσο σφάλμα (**MAE = 4.1096**) κρίνεται αποδεκτό για μια αρχική εκτίμηση, επιβεβαιώνοντας ότι ακόμη και απλές αρχιτεκτονικές μπορούν να προσφέρουν μια βασική, αλλά αξιόπιστη, εικόνα του κινδύνου.

Γ. Διαγράμματα Πρόβλεψης και Σημαντικότητα Χαρακτηριστικών

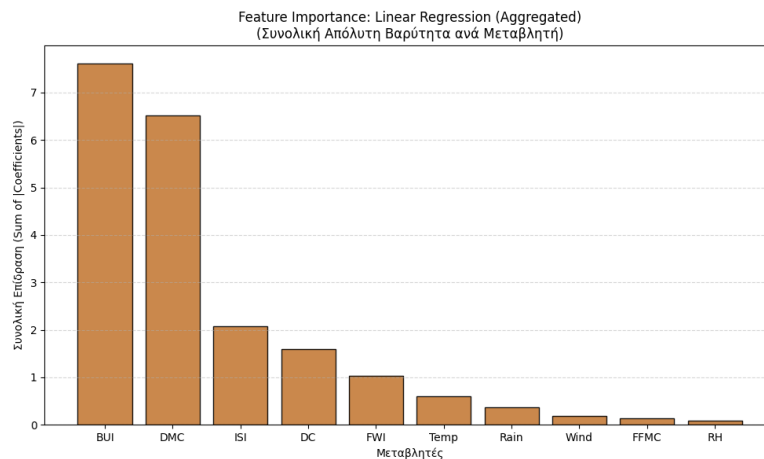
Για την οπτική αξιολόγηση της προσαρμογής του μοντέλου, δημιουργήθηκε το διάγραμμα διασποράς (**Scatter Plot**), όπου συγκρίνονται οι πραγματικές με τις προβλεπόμενες τιμές (**Εικόνα 4.18**).



Εικόνα 4.18: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI

Το διάγραμμα επιβεβαιώνει την ισχυρή γραμμική συσχέτιση, καθώς η πλειονηφία των σημείων ακολουθεί πολύ κοντά στη διαγώνιο της τέλειας πρόβλεψης (κόκκινη γραμμή), δικαιώνοντας το υψηλό $R^2 = 0.869$. Ωστόσο, παρατηρείται αισθητή αύξηση της διασποράς στις υψηλές τιμές του δείκτη FWI (>50), γεγονός που υποδεικνύει ότι το γραμμικό μοντέλο δυσκολεύεται να προβλέψει με την ίδια ακρίβεια τις ακραίες και πιο επικίνδυνες περιπτώσεις.

Για την ερμηνεία του μοντέλου, η συνολική βαρύτητα κάθε δείκτη υπολογίστηκε αθροίζοντας τις απόλυτες τιμές των συντελεστών του στο χρονικό παράθυρο των 10 ημερών. Η διαδικασία αυτή αποτυπώνει τη συνολική επίδραση κάθε μετεωρολογικού παράγοντα στη διαμόρφωση της τελικής τιμής του FWI, όπως φαίνεται στην **Εικόνα 4.19**.



Εικόνα 4.19: Σημαντικότητα Χαρακτηριστικών στο Linear Regression

Η ανάλυση ανέδειξε την απόλυτη κυριαρχία των δεικτών **BUI** και **DMC**, οι οποίοι επισκίασαν τα υπόλοιπα χαρακτηριστικά. Αυτό ερμηνεύεται από τη μαθηματική φύση του προβλήματος, καθώς η Γραμμική Παλινδρόμηση επιχειρεί ουσιαστικά να ανασυνθέσει τον τύπο υπολογισμού του δείκτη FWI, αποδίδοντας το μέγιστο βάρος στα βασικά συστατικά ξηρασίας από τα οποία προκύπτει.

Δ. Συμπέρασμα για Linear Regration

Συνοψίζοντας, η Γραμμική Παλινδρόμηση λειτούργησε ως ένα ισχυρό μοντέλο αναφοράς ($R^2=0.87$), αναδεικνύοντας ότι ο δείκτης **FWI** καθορίζεται πρωτίστως από γραμμικούς συνδυασμούς των υποδεικτών ξηρασίας (**BUI**, **DMC**). Παρόλα αυτά, η σημαντική απόκλιση μεταξύ **RMSE** και **MAE**, σε συνδυασμό με την αυξημένη διασπορά στις υψηλές τιμές, καταδεικνύει την αδυναμία της να προβλέψει με ακρίβεια τις ακραίες συνθήκες κινδύνου, καθιστώντας αναγκαία τη χρήση μη-γραμμικών μοντέλων για τις κρίσιμες περιπτώσεις.

4.3.2 XGBoost Regressor

Μετά την εφαρμογή του γραμμικού μοντέλου αναφοράς, προχωρήσαμε στη χρήση του αλγορίθμου **XGBoost (Extreme Gradient Boosting)**. Σκοπός ήταν να διερευνηθεί εάν η ικανότητα του XGBoost να μοντελοποιεί μη-γραμμικές σχέσεις και πολύπλοκες αλληλεπιδράσεις μεταξύ των μεταβλητών μπορεί να μειώσει το σφάλμα πρόβλεψης, ειδικά στις υψηλές και πιο επικίνδυνες τιμές του δείκτη FWI, όπου η Γραμμική Παλινδρόμηση εμφάνισε αδυναμίες.

Α. Υλοποίηση και Μετασχηματισμός Δεδομένων

Όπως και στην προηγούμενη περίπτωση, τα δεδομένα εισόδου μετασχηματίστηκαν σε διδιάστατο πίνακα (Flattening), ώστε κάθε ημέρα του ιστορικού να γίνει ξεχωριστό χαρακτηριστικό. Το μοντέλο ρυθμίστηκε με έμφαση στην αποφυγή υπερ-προσαρμογής (overfitting), επιλέγοντας έναν σχετικά χαμηλό ρυθμό μάθησης (**learning_rate=0.05**) και περιορίζοντας το βάθος των δέντρων (**max_depth=6**).

Για την ενίσχυση της γενικευτικής ικανότητας, χρησιμοποιήθηκε η τεχνική της δειγματοληψίας δεδομένων (Subsampling), όπου κάθε δέντρο εκπαιδεύεται στο **80%** των δεδομένων και των χαρακτηριστικών (**Κώδικα 4.10**).

```
# FLATTEN ΔΕΔΟΜΕΝΩΝ
# Το XGBoost θέλει 2D πίνακα. Μετατρέπουμε το (Samples, 10, 10) σε (Samples, 100)
num_samples = X_3d.shape[0]
num_timesteps = X_3d.shape[1]
num_features = X_3d.shape[2]

X_flat = X_3d.reshape(num_samples, num_timesteps * num_features)

# -----
# SPLIT TRAIN/TEST (80-20)
X_train, X_test, y_train, y_test = train_test_split(
    X_flat, y_scaled, test_size=0.2, random_state=42)

# -----
# ΕΚΠΑΙΔΕΥΣΗ XGBOOST
xgb = XGBRegressor(
    n_estimators=300, # Πλήθος Δέντρων
    learning_rate=0.05, # Πόσο γρήγορα μαθαίνει
    max_depth=6, # Πόσο βαθιά είναι τα δέντρα
    subsample=0.8, # Χρήση μέρους των δεδομένων
    colsample_bytree=0.8, # Χρήση μέρους των features
    random_state=42,
    n_jobs=-1
)
xgb.fit(X_train, y_train)
```

Κώδικας 4.10: Ρύθμιση Υπερπαραμέτρων XGBoost Regressor

B. Αποτελέσματα XGBoost

Η απόδοση του μοντέλου παρουσιάζεται στον **Πίνακα 4.8**. Αναμένεται βελτίωση στις μετρικές σφάλματος (RMSE, MAE) σε σχέση με το baseline μοντέλο.

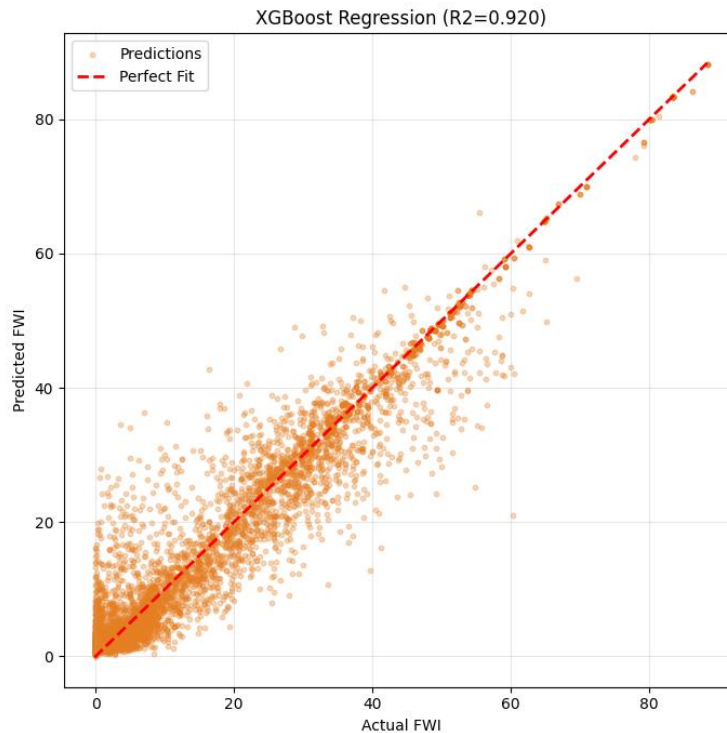
Πίνακας 4.9: Επιδόσεις XGBoost

R2 - Score	MAE	RMSE
0.9202	2.9727	4.8815

Ο XGBoost παρουσίασε μεγάλη βελτίωση, με τον συντελεστή R^2 (**0.9202**). Η αισθητή μείωση του μέσου σφάλματος (**MAE=2.97**) κατά 1.14 μονάδες σε σχέση με το Linear Regression, δείχνει την δυναμική των μη – γραμμικών αρχιτεκτονικών.

Γ. Διαγράμματα Πρόβλεψης και Σημαντικότητα Χαρακτηριστικών

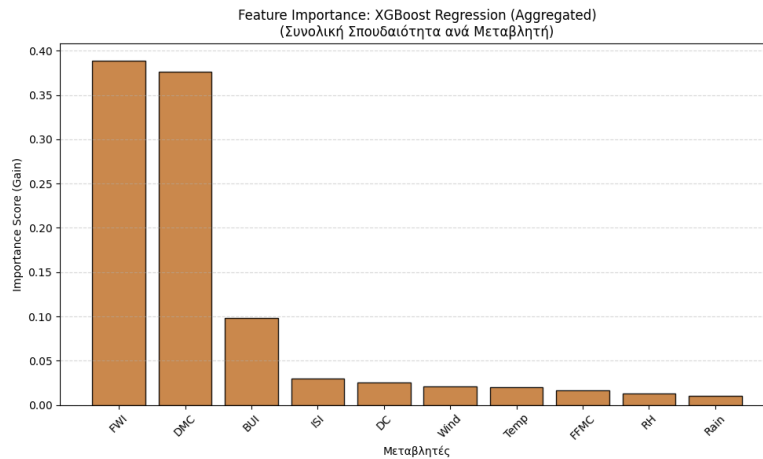
Για την οπτική αξιολόγηση της προσαρμογής του μοντέλου, δημιουργήθηκε **διάγραμμα διασποράς** (Scatter Plot) που συγκρίνει τις **πραγματικές με τις προβλεπόμενες τιμές FWI** (Εικόνα 4.20).



Εικόνα 4.20: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για XGBoost

Η γραφική απεικόνιση επιβεβαιώνει την ποιοτική αναβάθμιση της πρόβλεψης, καθώς το νέφος των σημείων εμφανίζει πολύ μεγαλύτερη συσπείρωση γύρω από τη διαγώνιο σε σχέση με το γραμμικό μοντέλο. Η μείωση των αποκλίσεων είναι ιδιαίτερα εμφανής στην κρίσιμη περιοχή των υψηλών τιμών (**FWI>40**), αποδεικνύοντας ότι το XGBoost μπορεί να προβλέψει με αξιοπιστία τις επικίνδυνες ημέρες, δικαιώνοντας απόλυτα τον υψηλό δείκτη R^2 (0.92).

Για την ερμηνεία του μοντέλου, η συνολική βαρύτητα κάθε δείκτη υπολογίστηκε αθροίζοντας τα Σημαντικά χαρακτηριστικά εισόδου που απέδωσε ο αλγόριθμος στο χρονικό παράθυρο των 10 ημερών. Η διαδικασία αυτή αποτυπώνει τη συνολική συνεισφορά κάθε παράγοντα στη βελτιστοποίηση της πρόβλεψης του FWI (**Εικόνα 4.21**).



Εικόνα 4.21: Σημαντικότητα Χαρακτηριστικών στο XGBoost

Δ. Συμπέρασμα για XGBoost

Το XGBoost αποδείχθηκε εξαιρετικό εργαλείο πρόβλεψης ($R^2 \approx 0.92$), ξεπερνώντας κατά πολύ τη Γραμμική Παλινδρόμηση. Κατάφερε να συνδυάσει έξυπνα το ιστορικό **FWI** με δείκτες ξηρασίας όπως το **DMC**, μειώνοντας τα σφάλματα ακόμα και στις ακραίες τιμές, κάτι που το κάνει ιδανικό για πραγματική χρήση.

4.3.3 Deep Learning Regression (LSTM & GRU)

Αφού δοκιμάσαμε τις "στατικές" μεθόδους (Linear Regression, XGBoost), τώρα περνάμε στα νευρωνικά δίκτυα τύπου RNN (**LSTM, GRU**), τα ίδια που τα πήγαν μια χαρά στην ταξινόμηση.

Η μόνη αλλαγή είναι στο τέλος: αντί για **Sigmoid** (που βγάζει πιθανότητες 0-1), βάλουμε γραμμικό νευρώνα (**Linear**), ώστε να μπορεί να προβλέψει οποιαδήποτε τιμή FWI.

Θέλουμε να δούμε αν η "μνήμη" τους μπορεί να πιάσει χρονικά μοτίβα που ξέφυγαν από το XGBoost και να βελτιώσει κι άλλο την πρόβλεψη.

4.3.3.1 Μοντέλο LSTM Regression

Το πρώτο δίκτυο DL που εξετάστηκε είναι το LSTM, το οποίο σχεδιάστηκε ειδικά για να λύνει το πρόβλημα της εξασθένισης της βαθμίδας (**vanishing gradient**) που ταλαιπωρούσε τα παλιά RNN. Στα LSTM, οι "πόρτες" (**forget, input, output gates**) επιτρέπουν στο δίκτυο να διατηρεί σημαντικές πληροφορίες ακόμα και μετά από πολλές ημέρες ιστορικού, κάτι απαραίτητο για Σύστημα πρόβλεψης πυρκαγιών.

A. Αρχιτεκτονική και Υλοποίηση

Η δομή του δικτύου αποτελείται από δύο επίπεδα (Stacked LSTM). Το πρώτο επίπεδο περιλαμβάνει **64 νευρώνες** και επιστρέφει **ολόκληρη την ακολουθία** (return_sequences=True), τροφοδοτώντας το δεύτερο επίπεδο των **32 νευρώνων**. Για την καταπολέμηση της υπερ-προσαρμογής, παρεμβάλλεται μηχανισμός **Dropout (20%)** μετά από κάθε επίπεδο LSTM.

Στο τέλος, η πληροφορία οδηγείται σε ένα πυκνό επίπεδο (**Dense**) 16 νευρώνων και καταλήγει σε έναν μοναδικό νευρώνα εξόδου με **γραμμική συνάρτηση ενεργοποίησης (Linear Activation)**, ώστε να παράγεται η συνεχής τιμή πρόβλεψης (**Κώδικα 4.11**).

```
# ΑΡΧΙΤΕΚΤΟΝΙΚΗ LSTM (REGRESSION MODE)
model = Sequential([
    Input(shape=(X_train.shape[1], X_train.shape[2])),
    # LSTM Layers
    LSTM(64, return_sequences=True),
    Dropout(0.2),
    LSTM(32, return_sequences=False),
    Dropout(0.2),
    # Dense Layer
    Dense(16, activation='relu'),
    # OUTPUT LAYER: Linear (για συνεχή τιμή FWI)
    Dense(1, activation='linear')
])
```

Κώδικας 4.11: Αρχιτεκτονική LSTM (Regression Mode)

Β. Αποτελέσματα LSTM

Οι επιδόσεις του μοντέλου παρουσιάζονται στον **Πίνακα 4.10**. Εδώ μας ενδιαφέρει να δούμε αν η "μήμη" του LSTM κατάφερε να μειώσει το σφάλμα (RMSE) σε σχέση με το XGBoost.

Πίνακας 4.10: Επιδόσεις LSTM Regression

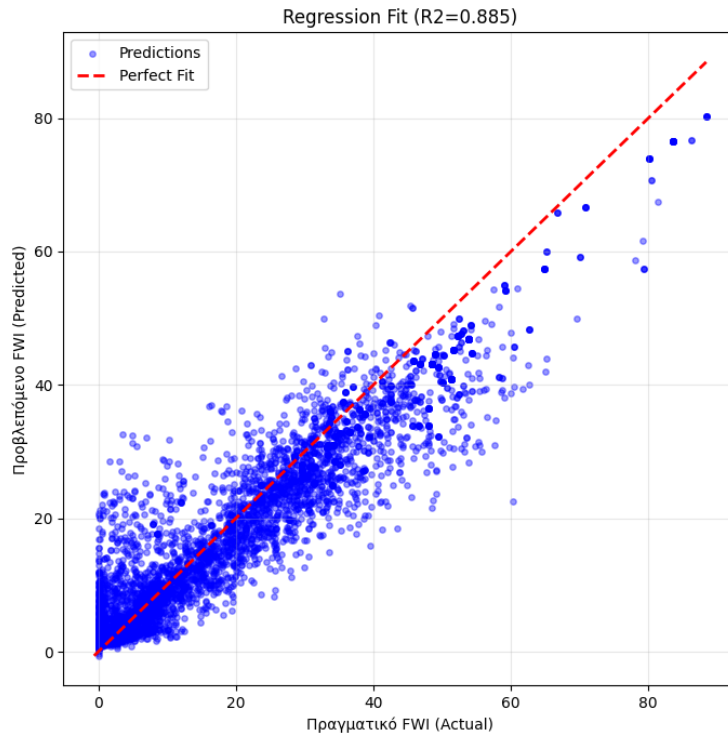
R2 - Score	MAE	RMSE
0.8853	3.9383	5.8533

Τα αποτελέσματα κρίνονται ικανοποιητικά, καθώς ο υψηλός συντελεστής προσδιορισμού (**R2≈0.89**) επιβεβαιώνει την ικανότητα του LSTM να μοντελοποιεί επιτυχώς τη χρονική ακολουθία του κινδύνου. Ωστόσο, οι τιμές των σφαλμάτων (**MAE≈3.94**, **RMSE≈5.85**) υποδηλώνουν ότι, παρόλο που η γενική τάση προβλέπεται σωστά, διατηρούνται ορισμένες αποκλίσεις στην ακρίβεια των σημειακών εκτιμήσεων.

του μέσου σφάλματος (**MAE=2.97**) κατά 1.14 μονάδες σε σχέση με το Linear Regression, δείχνει την δυναμική των μη – γραμμικών αρχιτεκτονικών.

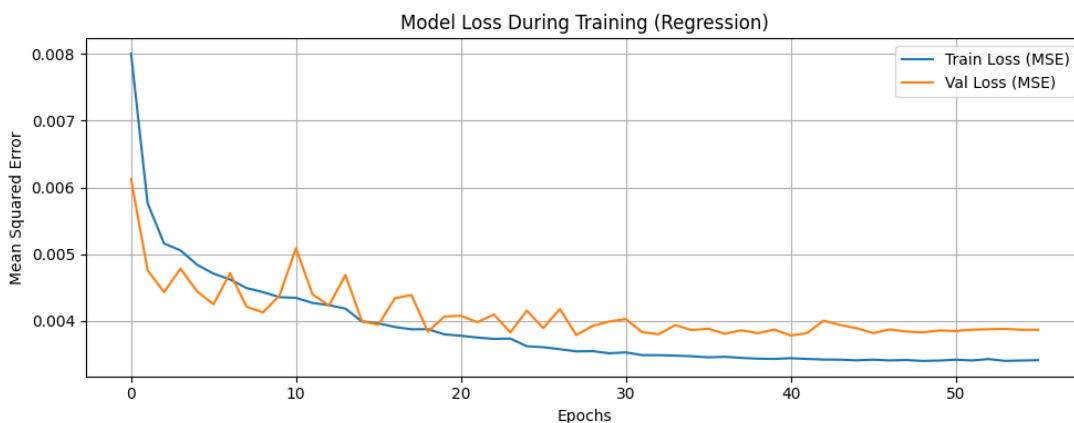
Γ. Διαγράμματα Πρόβλεψης και Σημαντικότητα Χαρακτηριστικών

Για την οπτική αξιολόγηση της προσαρμογής του μοντέλου, δημιουργήθηκε διάγραμμα διασποράς (Scatter Plot) που συγκρίνει τις πραγματικές με τις προβλεπόμενες τιμές FWI (**Εικόνα 4.22**), επιβεβαιώνοντας την υψηλή συσχέτιση. Στη συνέχεια παρουσιάζουμε τις καμπύλες απώλειας (**Εικόνα 4.23**), που επιβεβαιώνουν την ομαλή εκπαίδευση χωρίς υπερ-προσαρμογή, και η ανάλυση σημαντικότητας (**Εικόνα 4.24**) που ιεραρχεί τις μεταβλητές βάσει της επίδρασής τους στο σφάλμα.



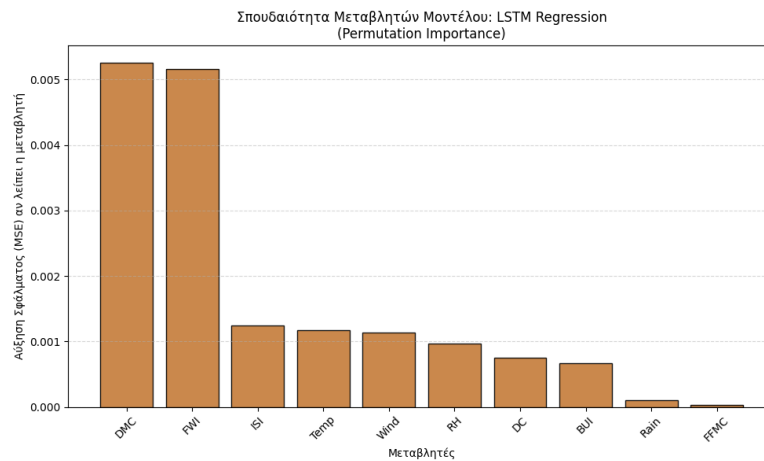
Εικόνα 4.22: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για LSTM

Η οπτική επιθεώρηση επιβεβαιώνει την ισχυρή συσχέτιση ($R^2 \approx 0.89$), καθώς το "νέφος" των προβλέψεων ακολουθεί τη διαγώνιο αναφοράς. Ωστόσο, η παρατηρούμενη διασπορά στις μεσαίες τιμές του δείκτη (**FWI 20-50**) ερμηνεύει το σφάλμα RMSE, υποδεικνύοντας ότι το μοντέλο, αν και συλλαμβάνει τη γενική τάση, παρουσιάζει αβεβαιότητα στις ενδιάμεσες συνθήκες κινδύνου.



Εικόνα 4.23: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για LSTM

Επίσης, ο τρόπος με τον οποίο αλλάζει η συνάρτηση απώλειας στην πάροδο των εποχών φαίνεται στο **Σχήμα 4.23**. Η σύγκλιση των καμπυλών εκπαίδευσης (Train) και επικύρωσης (Validation) σε χαμηλό επίπεδο σφάλματος αποδεικνύει ότι η διαδικασία είναι σταθερή και δεν υπάρχει πρόβλημα υπερπροσαρμογής.



Εικόνα 4.24: Σημαντικότητα Χαρακτηριστικών στο LSTM (Regression)

Η ανάλυση αποκαλύπτει μια ενδιαφέρουσα "στρατηγική" του LSTM. Βασίζεται σχεδόν ισοδύναμα στο **DMC** (πόσο ξηρή είναι η καύσιμη ύλη) και στο ιστορικό του **FWI**. Αυτό δείχνει ότι το δίκτυο έμαθε να συνδυάζει τη μακροπρόθεσμη ξηρασία με την πρόσφατη εικόνα κινδύνου, αντί να εστιάζει μονομερώς σε ένα από τα δύο όπως έκαναν τα προηγούμενα μοντέλα.

Δ. Συμπέρασμα για LSTM (Regression)

Συμπερασματικά, το LSTM λειτούργησε ως η «χρυσή τομή» ($R^2 \approx 0.89$), καθώς κατάφερε να συνδυάσει τη μνήμη του παρελθόντος (**FWI**) με την τρέχουσα ξηρασία (**DMC**). Αυτή η ικανότητα σύνθεσης του επιτρέπει να αντιλαμβάνεται τη δυναμική της πυρκαγιάς πιο ολοκληρωμένα, προσφέροντας αξιόπιστες προβλέψεις που δεν αγνοούν ούτε το ιστορικό ούτε τις περιβαλλοντικές συνθήκες.

4.3.3.2 Μοντέλο GRU Regression

Στη συνέχεια, δοκιμάστηκε η αρχιτεκτονική **GRU**, η οποία αποτελεί μια απλούστερη παραλλαγή του LSTM με λιγότερες παραμέτρους, προσφέροντας συχνά ταχύτερη εκπαίδευση με παρεμφερή απόδοση.

Α. Αρχιτεκτονική και Υλοποίηση

Προκειμένου να διασφαλιστεί η άμεση και δίκαιη σύγκριση με το δίκτυο LSTM, διατηρήθηκε αυστηρά η ίδια αρχιτεκτονική τοπολογία και οι ίδιες υπερπαραμέτροι εκπαίδευσης. Συγκεκριμένα, το μοντέλο δομήθηκε με δύο επάλληλα επίπεδα (Stacked GRU) των 64 και 32 νευρώνων αντίστοιχα, διατηρώντας το ποσοστό Dropout στο 20% για την αποφυγή υπερ-προσαρμογής.

Ομοίως, το επίπεδο εξόδου παρέμεινε αμετάβλητο (**Dense 16, Linear Output**), όπως και ο αλγόριθμος βελτιστοποίησης (**Adam, lr=0.001**). Με τον τρόπο αυτό, οποιαδήποτε διαφοροποίηση στην απόδοση μπορεί να αποδοθεί αποκλειστικά στη διαφορετική εσωτερική λειτουργία των κελιών GRU έναντι των LSTM (**Κώδικας 4.12**)

```

# ΑΡΧΙΤΕΚΤΟΝΙΚΗ GRU (STACKED)
print("\n Χτίσιμο Μοντέλου GRU...")
model = Sequential([
    Input(shape=(X_train.shape[1], X_train.shape[2])),

    # 1ο Layer: Διαβάζει την ακολουθία
    GRU(64, return_sequences=True),
    Dropout(0.2),

    # 2ο Layer: Συμπυκνώνει την πληροφορία
    GRU(32, return_sequences=False),
    Dropout(0.2),

    # Dense Layers
    Dense(16, activation='relu'),
    Dense(1, activation='linear') # Linear για Παλινδρόμηση
])

```

Κώδικας 4.12: Αρχιτεκτονική GRU (Regression Mode)

B. Αποτελέσματα GRU

Οι επιδόσεις του μοντέλου παρουσιάζονται στον **Πίνακα 4.11**. Εδώ εστιάζουμε στο κατά πόσο η απλούστερη εσωτερική δομή του GRU λειτούργησε πιο αποδοτικά από το LSTM στα συγκεκριμένα δεδομένα.

Πίνακας 4.11: Επιδόσεις GRU Regression

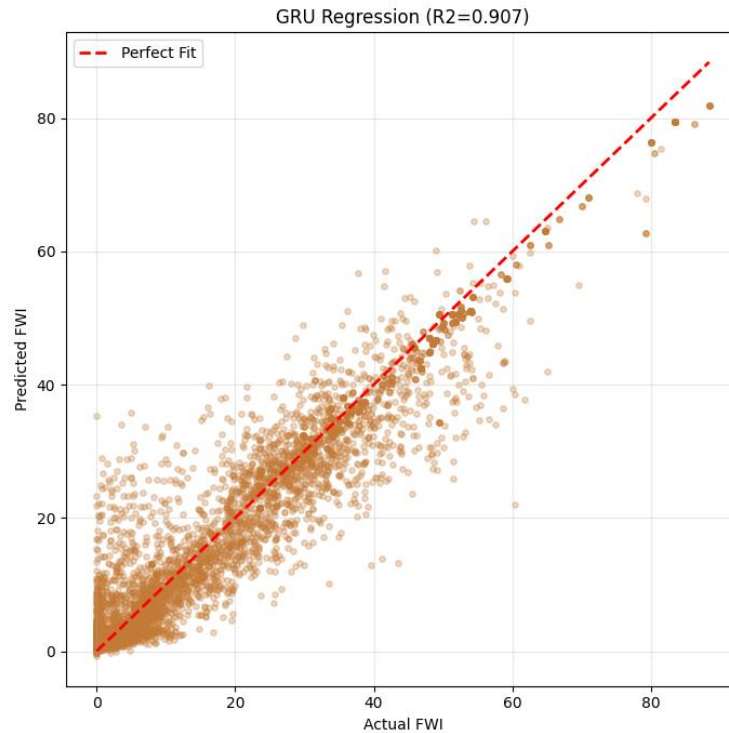
R2 - Score	MAE	RMSE
0.9067	3.3336	5.2773

Τα αποτελέσματα αποδεικνύονται ιδιαίτερα ενθαρρυντικά, καθώς το μοντέλο GRU επέτυχε υψηλότερη ακρίβεια σε σχέση με το LSTM, με τον συντελεστή προσδιορισμού να ανέρχεται στο **0.9067**. Η αξιοσημείωτη μείωση των σφαλμάτων (**MAE≈3.33**, **RMSE≈5.28**) υποδεικνύει ότι η αρχιτεκτονική του GRU επέτρεψε πιο αποδοτική προσαρμογή στα δεδομένα, επιβεβαιώνοντας ότι συχνά οι απλούστερες δομές RNN μπορούν να υπερτερήσουν σε προβλήματα πρόβλεψης χρονοσειρών μεσαίου μεγέθους.

Γ. Διαγράμματα Πρόβλεψης και Σημαντικότητα Χαρακτηριστικών

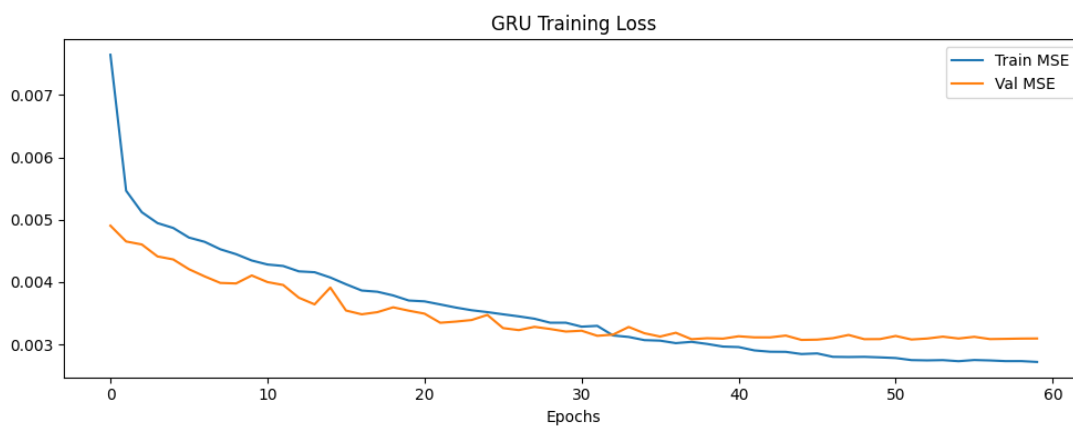
Για την οπτική αξιολόγηση της προσαρμογής του μοντέλου, παρατίθεται το διάγραμμα διασποράς (**Εικόνα 4.25**) που επιβεβαιώνει την υψηλή συσχέτιση ($R^2 = 0.9067$). Ακολουθούν οι καμπύλες απώλειας (Loss) (**Εικόνα 4.26**), που πιστοποιούν την ομαλή εκπαίδευση, και η ανάλυση σημαντικότητας (**Εικόνα 4.27**), η οποία ιεραρχεί τις μεταβλητές βάσει της συνεισφοράς τους στη μείωση του σφάλματος.

Κεφάλαιο 4



Εικόνα 4.25: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για GRU

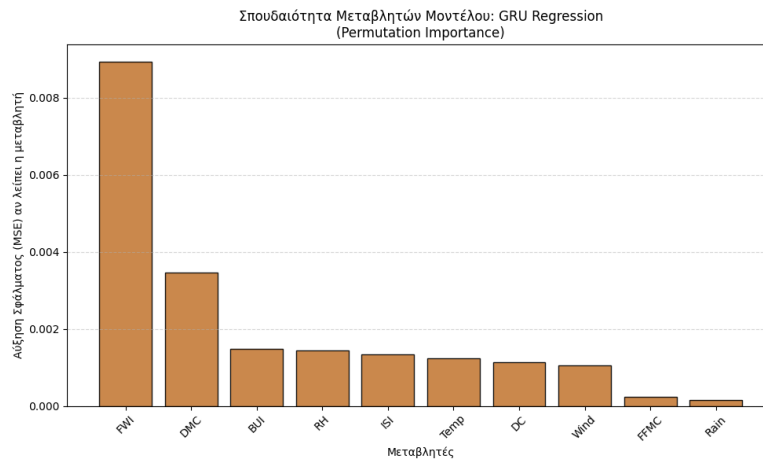
Η γραφική απεικόνιση επαληθεύει την εξαιρετική προσαρμογή του μοντέλου, καθώς το νέφος των σημείων εμφανίζει έντονη και πυκνή συσπείρωση γύρω από τη διαγώνιο αναφοράς. Η υψηλή ακρίβεια διατηρείται σταθερή ακόμα και στις ενδιάμεσες τιμές του δείκτη, γεγονός που υποδηλώνει ότι το δίκτυο διαχειρίστηκε αποτελεσματικά τον θόρυβο των δεδομένων, επιτυγχάνοντας αξιόπιστες εκτιμήσεις σε όλο το εύρος του δείγματος.



Εικόνα 4.26: Διάγραμμα Διασποράς Προβλεπόμενων vs Πραγματικών τιμών FWI για GRU

Από το παραπάνω διάγραμμα παρατηρείται ταχεία και σταθερή πτώση του σφάλματος (MSE), με τις καμπύλες εκπαίδευσης (Train) και επικύρωσης (Validation) να συγκλίνουν αρμονικά. Η απουσία

διακύμανσης μεταξύ τους, επιβεβαιώνει ότι η απλούστερη δομή του GRU λειτούργησε ευεργετικά, αποτρέποντας φαινόμενα υπερ-προσαρμογής.



Εικόνα 4.27: Σημαντικότητα Χαρακτηριστικών στο GRU

Η ανάλυση σημαντικότητας αναδεικνύει το ιστορικό του δείκτη **FWI** ως τον αδιαμφισβήτητο κυρίαρχο παράγοντα πρόβλεψης. Ακολουθεί ο δείκτης ξηρασίας **DMC** με σαφώς μικρότερη, αλλά διακριτή βαρύτητα. Το μοντέλο φαίνεται να δίνει προτεραιότητα στην «αδράνεια» του φαινομένου (δηλαδή, η προηγούμενη κατάσταση κινδύνου καθορίζει σε μεγάλο βαθμό την τρέχουσα), αξιοποιώντας τους δείκτες ξηρασίας συμπληρωματικά για τη βελτιστοποίηση της εκτίμησης.

Δ. Συμπέρασμα για GRU Regression

Συμπερασματικά, το μοντέλο GRU επέδειξε υψηλή προσαρμοστικότητα και ακρίβεια ($R^2 \approx 0.91$). Η ικανότητά του να εστιάζει στο πρόσφατο ιστορικό κινδύνου (FWI), αξιοποιώντας παράλληλα την απλή αλλά αποδοτική δομή του, του επέτρεψε να επιτύχει υψηλά επίπεδα γενίκευσης και να ελαχιστοποιήσει το σφάλμα πρόβλεψης.

4.3.4 Συζήτηση Αποτελεσμάτων και Επιλογή Βέλτιστου Μοντέλου

4.3.4.1 Συγκριτική Ανάλυση Επιδόσεων

Στον Πίνακα 4.12 παρουσιάζεται η συγκεντρωτική αξιολόγηση των τεσσάρων μεθόδων παλινδρόμησης. Η σύγκριση βασίζεται στον Συντελεστή Προσδιορισμού (R^2), το Μέσο Απόλυτο Σφάλμα (MAE) και το Τυπικό Σφάλμα (RMSE), δείκτες που αποτυπώνουν την ακρίβεια πρόβλεψης του δείκτη FWI.

Πίνακας 4.12: Συγκεντρωτική αξιολόγηση Regression

Model	R2 Score	MAE	RMSE
XGBoost	0.9202	2.9727	4.8815
GRU	0.9067	3.3336	5.2773
LSTM	0.8853	3.9383	5.8533
Linear Reg.	0.8693	4.1096	6.2475

Από την επισκόπηση των αποτελεσμάτων για την περίπτωση του Regression (δεδομένα 1:4), προκύπτει η σαφής υπεροχή του αλγορίθμου **XGBoost**, ο οποίος κατέγραψε τις καλύτερες επιδόσεις σε όλες τις μετρικές. Δεύτερο ακολουθεί το GRU.

4.3.4.2 Επιλογή Βέλτιστου Μοντέλου

Η επιλογή του καταλληλότερου μοντέλου παλινδρόμησης για την πρόβλεψη του δείκτη FWI, δεν εξαρτάται αποκλειστικά και μόνο από τον δείκτη αυτό. Η κάθε περίοδος του χρόνου χρειάζεται ειδική μεταχείριση από το σύστημα με βάση την εποχικότητα του κινδύνου. Σύγχρονες μελέτες τονίζουν την ανάγκη για προσαρμοσμένους δείκτες (tailored indices) που λαμβάνουν υπόψη τα τοπικά πρότυπα και τις μεταβαλλόμενες συνθήκες, αντί για τη χρήση καθολικών μοντέλων [43-45].

Έτσι λοιπόν μπορούμε να χωρίσουμε την επιλογή μας σε δυο (2) στρατηγικές αξιολόγησης μοντέλων:

A. Αρχή αντιπυρικής περιόδου (Χαμηλό Ρίσκο - Σταθερή Παρακολούθηση - MAE)

Στην αρχή της σεζόν, με χαμηλό-μέτριο κίνδυνο, θέλουμε απλά το μοντέλο να "πιάνει" σωστά τις καθημερινές αλλαγές του FWI. Εδώ μετράει το MAE, δηλαδή πόσο κοντά είναι κατά μέσο όρο οι προβλέψεις στην πραγματικότητα. Το **XGBoost** κερδίζει εύκολα (**MAE≈2.97**) για τον καθημερινό σχεδιασμό. Το **GRU** ακολουθεί κοντά (**MAE≈3.33**), πολύ καλό για εναλλακτική λύση [43-45].

B. Μέσα καλοκαιριού (Υψηλό Ρίσκο - Έλεγχος Ακραίων - RMSE)

Μέσα Ιουλίου-Αυγούστου, ένα μεγάλο λάθος (π.χ. **FWI 30 αντί για 70**) μπορεί να κοστίζει ακριβά. Το **RMSE** τιμωρεί τις μεγάλες αποκλίσεις, και πάλι το **XGBoost** υπερτερεί (**RMSE≈4.88**) να είναι σταθερό ακόμα και στα άκρα. Το **GRU** (**RMSE≈5.27**) αντέχει αξιοπρεπώς, ενώ η γραμμική παλινδρόμηση αποτυγχάνει (**RMSE≈6.25**) με επικίνδυνα μεγάλα λάθη [43-45].

Ως ένα **Τελικό Συμπέρασμα**, η συνολική Ακρίβεια των μοντέλων φαίνεται ξεκάθαρα από τον συγκεντρωτικό πίνακα παραπάνω. Το XGBoost πετυχαίνει την καλύτερη επίδοση όλων με **R²=0.92**, που σημαίνει ότι το 92% των προβλέψεων προβλέπουν σωστά το FWI. Με **MAE=2.97** (δηλαδή το μέσο σφάλμα πρόβλεψης κατά μέσο όρο <3) και **RMSE=4.88** (που τιμωρεί τις μεγάλες αποκλίσεις), είναι το πιο αξιόπιστο για καθημερινή χρήση σε σύστημα πρόγνωσης Πυρκαγιών. Τα μοντέλα DL (**LSTM/GRU**) ακολουθούν αξιοπρεπώς (**R²~0.89-0.91**), αλλά η γραμμική παλινδρόμηση μένει αρκετά πίσω (**R²=0.87**, μεγαλύτερα σφάλματα), επιβεβαιώνοντας ότι τα μη γραμμικά μοντέλα ταιριάζουν καλύτερα στα σύνθετα καιρικά δεδομένα της συγκεκριμένης εργασίας.

Κεφάλαιο 5ο: Φάση 3 – Εφαρμογές & Προσομοίωση

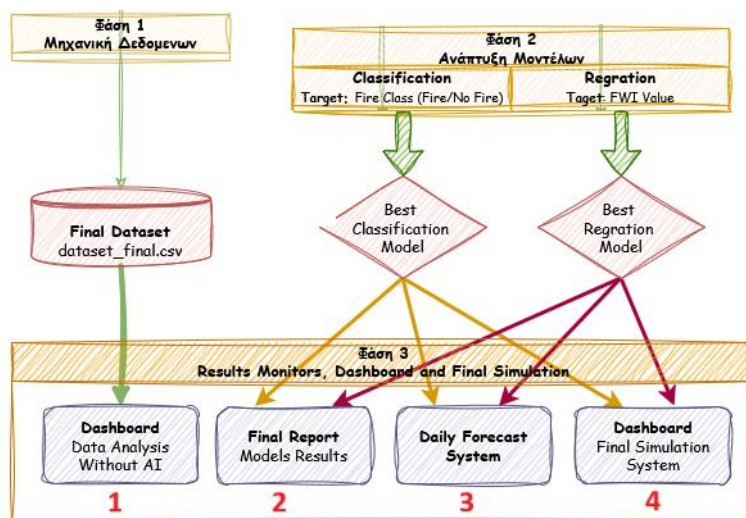
5.1 Εισαγωγή: Αρχιτεκτονική Συστήματος (Systems Architecture)

Η μετάβαση από τα μαθηματικά μοντέλα και τα *ξερά* αποτελέσματα δεικτών, σε ένα σύνολο εργαλείων οπτικοποίησης και κατανόησης, απαιτεί την κατασκευή ενός ολοκληρωμένου συστήματος λογισμικού το οποίο θα είναι εύχρηστο, γρήγορο και αξιόπιστο. Στο πλαίσιο της παρούσας διπλωματικής, σχεδιάστηκε και υλοποιήθηκε μια ευέλικτη αρχιτεκτονική με βάση τη γλώσσα **Python**, αξιοποιώντας πολλές βιβλιοθήκες όπως **Streamlit**, **Folium**, **geopy** και άλλες, για την δημιουργία των βοηθητικών dashboards, καθώς και του κεντρικού framework.

Η επιλογή του **Streamlit** ως κεντρικού framework βασίστηκε στην ικανότητά του να μετατρέπει **scripts** δεδομένων σε διαδραστικές web εφαρμογές σε πραγματικό χρόνο, επιτρέποντας την άμεση οπτικοποίηση των αποτελεσμάτων των μοντέλων ML και DL χωρίς την ανάγκη περίπλοκων διαδικασιών **Front-End** ανάπτυξης [46-47].

Το σύστημα αποτελείται (**Εικόνα 4.1**) από δύο υπο-εφαρμογές (Dashboards) και δυο κύριες εφαρμογές πρόβλεψης κινδύνου, κάθε μία από τις οποίες εξυπηρετεί διαφορετικό σκοπό:

1. Ανάλυση Κινδύνου Δασικών Πυρκαγιών στην Ελλάδα (5_17_dashboard_analytics_without_ai.py): Μια συγκεντρωτική παρουσίαση δεδομένων, των ιστορικών δεδομένων χωρίς την χρήση ΑΙ.
2. Τελική Έκθεση Αξιολόγησης Μοντέλων (4_16_generate_final_report.py): Μια συγκεντρωτική ανάλυση και οπτικοποίηση όλων των αποτελεσμάτων της φάσης 2.
3. AI Fire Risk Forecast – Σύστημα Διαχείρισης Κινδύνου Πυρκαγιάς (5_18_dashboard_daily_forecast_system.py): Κύριο εργαλείο πρόβλεψης κινδύνου επόμενης ημέρας.
4. AI Wildfire Risk Simulator - Live Πρόγνωση Κινδύνου (5_19_dashboard_final_simulation_app.py): Live πρόβλεψη κινδύνου με βάση τα καιρικά φαινόμενα, κάνοντας χρήση την υβριδική λογική.



Εικόνα 5.1: Τελικά συστήματα του Project

5.1.1 Ανάλυση Κινδύνου Δασικών Πυρκαγιών στην Ελλάδα (Fire Data Analysis without AI)

5.1.1.1 Σκοπός και Λειτουργία

Ξεκινώντας τον σχεδιασμό των εφαρμογών, ήταν αναγκαίο να δημιουργηθεί μια, η οποία θα εστίαζε στην οπτικοποίηση του τελικού dataset. Έτσι με την βοήθεια των δυνατοτήτων των βιβλιοθηκών όπως την **Folium** και την **Plotly** μετατρέψαμε τα στατικά δεδομένα σε ένα διαδραστικό αρχείο html.

Το Script ανοίγει αυτόματα σε φυλλομετρητή (με την βοήθεια την βιβλιοθήκης webbrowser) και παρουσιάζει όλα τα ιστορικά δεδομένα πυρκαγιών πάνω στο χάρτη. Ο χρήστης έχει την δυνατότητα να επιλέξει τα σημεία, με βάση την χρονιά.

Πολύ χρήσιμες είναι ο πίνακας των «**Top 50**» συμβάντων με βάση το FWI καθώς και τα σχήματα κατανόησης των συμβάντων ανά έτος, μήνα κλπ. Να σημειωθεί εδώ ότι δεν γίνεται καμία παρεμβολή αλγορίθμων μηχανικής και βαθιάς μάθησης.

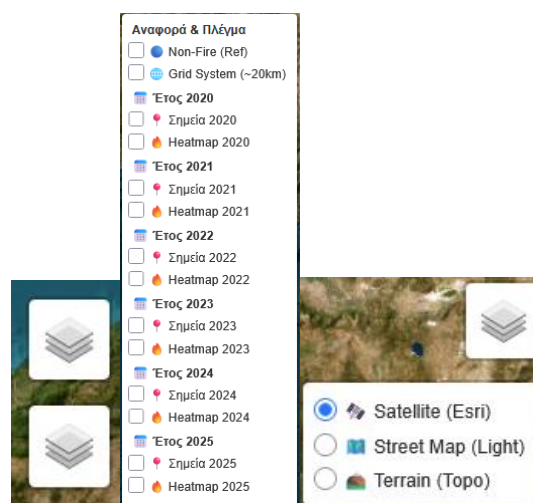
Στόχος είναι να κατανοήσει ο χρήστης την συμπεριφορά των μετεωρολογικών δεικτών, καθώς και των δεικτών του FWI σε σχέση με τις εκδηλώσεις πυρκαγιών, παρουσιάζοντας την εποχικότητα και την γεωγραφική κατανομή των κινδύνων για τα τελευταία 5 έτη.

5.1.1.2 Τεχνική Υλοποίηση & Εργαλεία

Όπως αναφέρθηκε και παραπάνω, αξιοποιήσαμε τις εξειδικευμένες βιβλιοθήκες της Python έτσι ώστε να υλοποιήσουμε το κάθε ένα feature της ενότητας αυτής. Ποιο συγκεκριμένα:

1. Γεωχωρική Ανάλυση (folium):

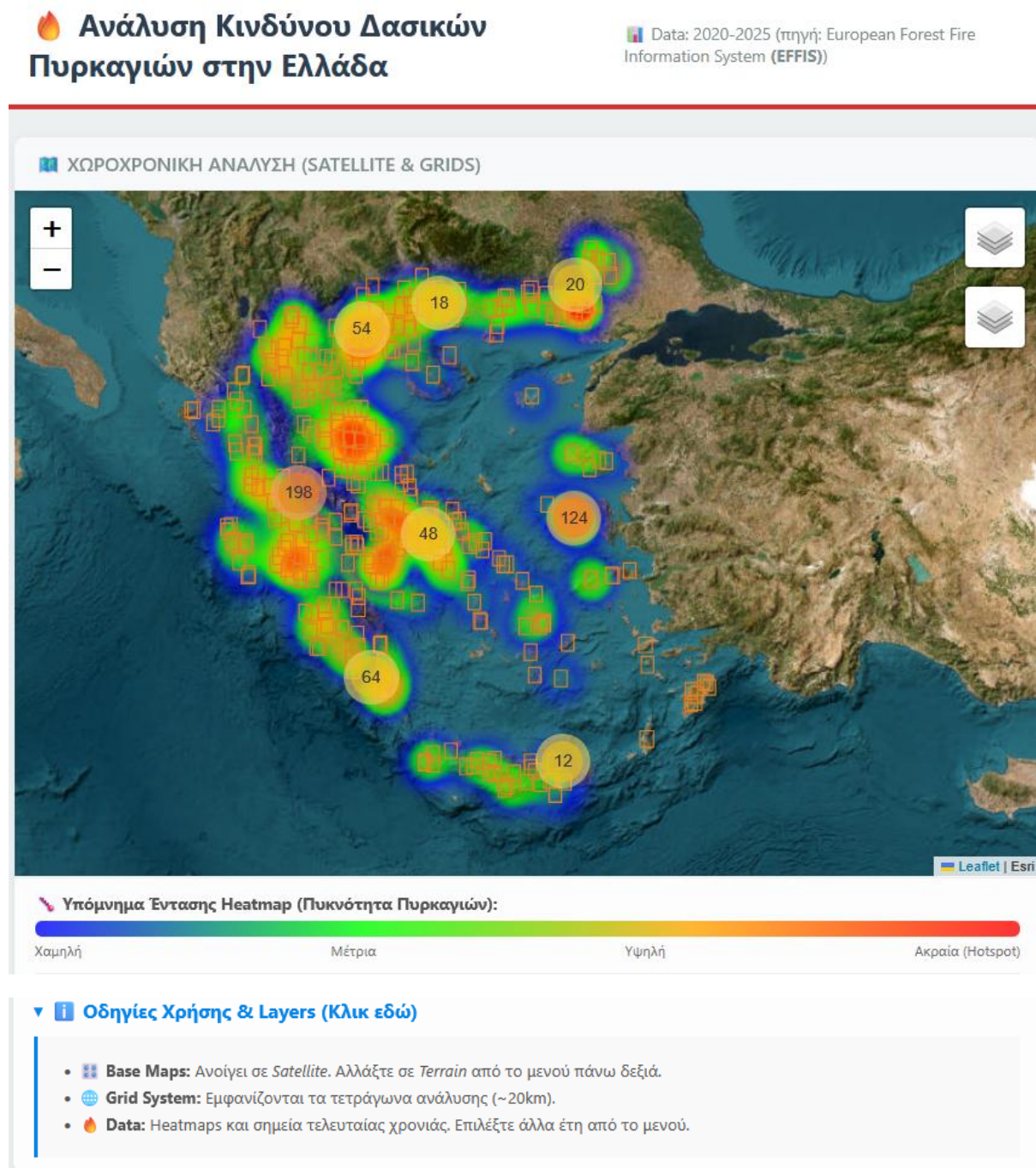
Ένα από σημαντικότερα πράγματα που προσφέρει το συγκεκριμένο υποσύστημα, είναι η οπτικοποίηση στον χάρτη των σημείων (**MarkerCluster**), καθώς και την πυκνότητα των πυρκαγιών ανά περιοχή μέσω θερμικών χαρτών (**HeatMap**) (**Εικόνα 5.2**). Στο τέλος έχουμε κάποιες οδηγίες για το τι είναι το καθένα.



Εικόνα 5.2: Πλαϊνή μπάρα επιλογών Χάρτη

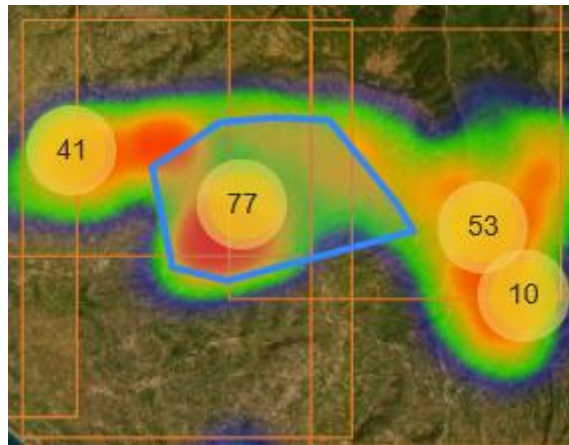
Στο κάτω μέρος του χάρτη, υπάρχει το **Υπόμνημα Έντασης Heatmap**. Στην ουσία υποδηλώνει την πυκνότητα πυρκαγιών σε συγκεκριμένη περιοχή πάνω στον χάρτη.

Στα δεξιά της Εικόνας 5.3 μπορεί ο χρήστης να επιλέξει, το layout του χάρτη που επιθυμεί (**Satellite, Street Map ή Terrain**), καθώς και να επιλέξει τι ακριβώς θα βλέπει στον χάρτη (Σημεία πυρκαγιάς, Heatmaps, Grids ή ακόμα και τα σημεία Non – Fire που έχει δημιουργήσει το σύστημα).



Εικόνα 5.3: Χάρτης Συμβάντων

Κάνοντας μεγέθυνση (Zoom) σε κάποιο σημείο στον χάρτη, μας εμφανίζει πόσα συμβάντα πυρκαγιών υπάρχει σε κάθε Grid. Και αν συνεχίσουμε να κάνουμε μεγέθυνση στον χάρτη η πυκνότητα, γίνεται «κόκκινη» κουκίδα, όπου ουσιαστικά απεικονίζει στο συγκεκριμένο συμβάν πυρκαγιάς (Εικόνα 5.4 και Εικόνα 5.5).



Εικόνα 5.4: Σύνολο Συμβάντων ανά Grid



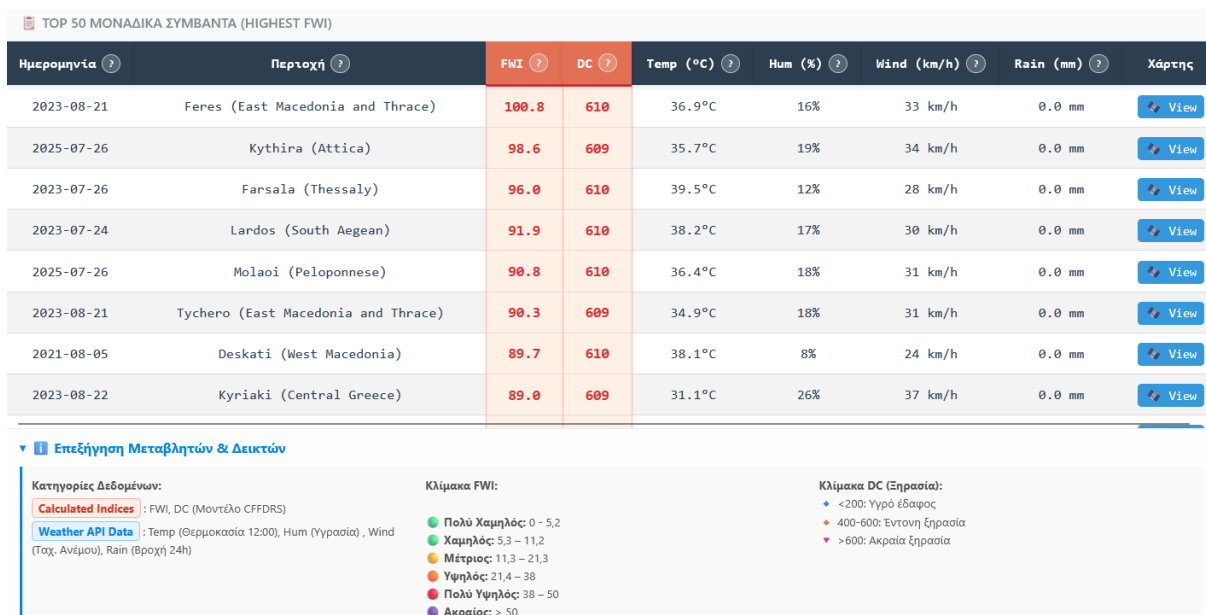
Εικόνα 5.5: Σημείο πυρκαγιάς

Επιπρόσθετα στην Εικόνα 5.5 παρουσιάζονται και άλλες πληροφορίες για το σημείο. Αυτές είναι:

- Η τιμή του FWI (Δείκτης Κινδύνου)
- Η τιμή του DC (Δείκτης Ξηρασίας)
- Καιρικές μεταβλητές (Θερμοκρασία, Αέρας)
- Και τέλος την Περιοχή και την Ημερομηνία συμβάντος

Το επόμενο Feature με σειρά προβολής είναι ο πίνακας των «**TOP 50 Μοναδικών Συμβάντων**» με βάση το υψηλότερο FWI (Εικόνα 5.6). Είναι ένας πίνακας που δείχνει συγκεντρωμένα όλη την πληροφορία του dataset που θα δοθεί σαν στοιχεία εισόδου στα μοντέλα. Πιο συγκεκριμένα και όπως φαίνεται στην εικόνα παρακάτω, η πληροφορία που παίρνουμε από τον συγκεκριμένο πίνακα είναι ημερομηνία του συμβάντος, η περιοχή, οι δείκτες FWI και DC (μαθηματικός υπολογισμός), οι μετεωρολογικοί δείκτες που επικρατούσαν στην περιοχή την ημέρα του συμβάντος (θερμοκρασία,

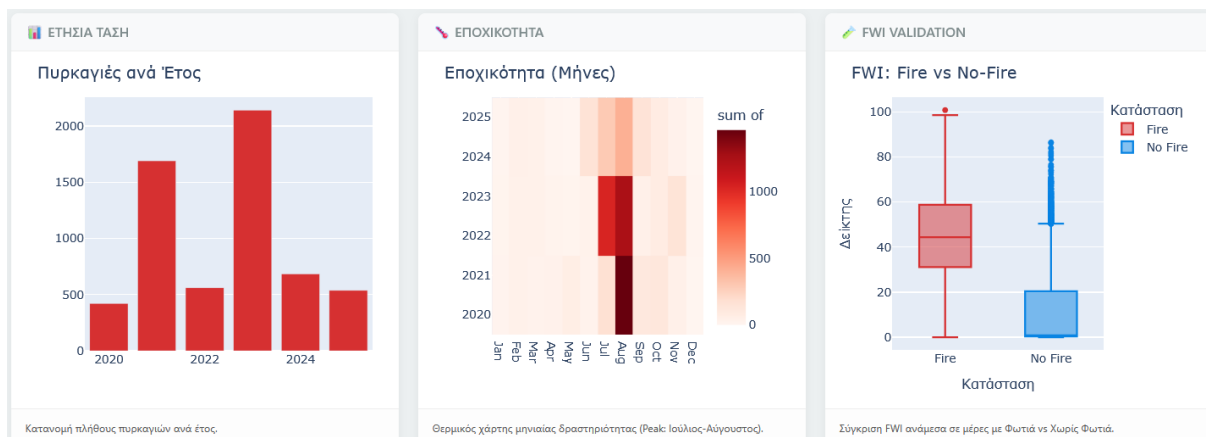
υγρασία, αέρας, βροχή) και ένα κουμπί μετάβασης στον χάρτη, που μας αποτυπώνει το σημείο του συμβάντος.



Εικόνα 5.6: TOP 50 Μοναδικά Συμβάντα (Highest FWI)

Στην ονομασία κάθε στήλης υπάρχει ένα Αγγλικό ερωτηματικό (?), που αν ο χρήστης βάλει τον κέρσορα του ποντικιού εκεί, εμφανίζεται μήνυμα με την εξήγηση του κάθε δείκτη. Και εδώ έχουμε ένα υπόμνημα με κάποια χρήσιμα στοιχεία του πίνακα, στο κάτω μέρος αυτού.

Η τελευταία απεικόνιση της εφαρμογής αυτής, είναι τρία (γραφήματα) που δείχνουν την κατανομή των πλήθων πυρκαγιών ανά έτος, την δραστηριότητα ανά μήνα και ένα box plot με τον διαχωρισμό των δυο κλάσεων (με την χρήση της βιβλιοθήκης plotly.express).



Εικόνα 5.7: Χρονική εξέλιξη περιστατικών πυρκαγιάς (Ετη, Μήνες) και συσχέτιση με τον δείκτη κινδύνου FWI

5.1.2 Τελική Έκθεση Αξιολόγησης Μοντέλων

Το επόμενο εργαλείο που υλοποιήσαμε, ήταν αυτό που απεικονίσαμε όλα τα αποτελέσματα σε μια σελίδα html. Ποιο συγκεκριμένα το συγκεκριμένο script λειτουργεί ως μηχανισμός δημιουργίας μια

Κεφάλαιο 5

αυτοματοποιημένης αναφοράς και τεκμηρίωσης, συγκεντρώνοντας όλα τα αποτελέσματα όλων των πειραμάτων σε ένα ενιαίο αρχείο HTML.

Αρχικά φορτώνει όλα τα εκπαιδευμένα μοντέλα (Classification & Regression) και υπολογίζει απ την αρχή της όλες τις μετρικές αξιολόγησης. Στην συνέχεια δημιουργεί από ένα πίνακα κατάταξης για κάθε πείραμα (Εικόνα 5.8). Παράλληλα εντοπίζει όλα τα γραφήματα που έχουν παραχθεί για την κάθε περίπτωση (Πίνακες Σύγκυσης, Καμπύλες Εκμάθησης, Γράφημα Παλινδρόμησης, Feature Importance). Με βάση όλα τα παραπάνω έχουμε μια ολική εικόνα της απόδοσης του συστήματος.

1. Classification Results (Metrics)						
Model	ROC AUC	Accuracy	F1-Score	Precision	Recall	
Random Forest	0.9510	91.6%	79.8%	76.9%	83.0%	
GRU	0.9284	87.9%	73.5%	66.3%	82.4%	
LSTM	0.9255	88.1%	73.3%	67.4%	80.3%	
XGBoost	0.9168	89.0%	73.6%	71.0%	76.5%	

Επεξήγηση Δεικτών:

- ROC AUC:** Η ικανότητα διαχωρισμού των κλάσεων (0.5=Τυχαίο, 1.0=Άριστο).
- Accuracy (Ακρίβεια):** Το ποσοστό των συνολικών προβλέψεων που ήταν σωστές.
- F1-Score:** Ο αρμονικός μέσος Precision & Recall. Κρίσιμος δείκτης για imbalanced δεδομένα.
- Precision:** Πόσες από τις προβλέψεις "θιγατά" ήταν πραγματικά φωτιά (Αποφυγή ψευδών συναγερμών).
- Recall:** Πόσες από τις πραγματικές φωτιές κατάφερε να εντοπίσει το μοντέλο.

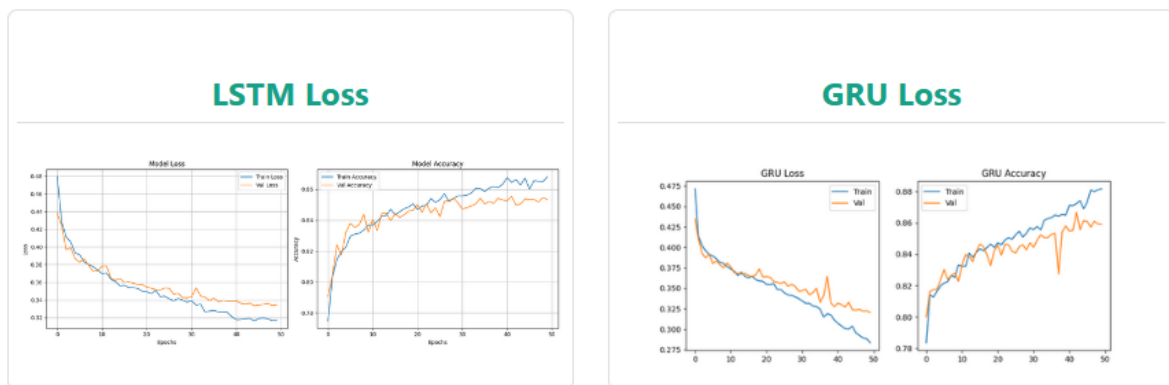
2. Regression Results (Metrics)				
Model	R2 Score	MAE	RMSE	
XGBoost	0.9202	2.9727	4.8815	
GRU	0.9067	3.3336	5.2773	
LSTM	0.8853	3.9383	5.8533	
Linear Reg	0.8693	4.1096	6.2475	

Επεξήγηση Δεικτών:

- R2 Score:** Το ποσοστό της μεταβλητότητας του FMI που ελγίνεται από το μοντέλο (Max=1.0).
- MAE (Mean Absolute Error):** Το μέσο σφάλμα πρόβλεψης σε μονάδες FMI (χαμηλότερο = < 5 μονάδες FMI) = καλύτερο).
- RMSE (Root Mean Squared Error):** Τυπικό σφάλμα που "τιμωρεί" περισσότερο τις μεγάλες αποκλίσεις.

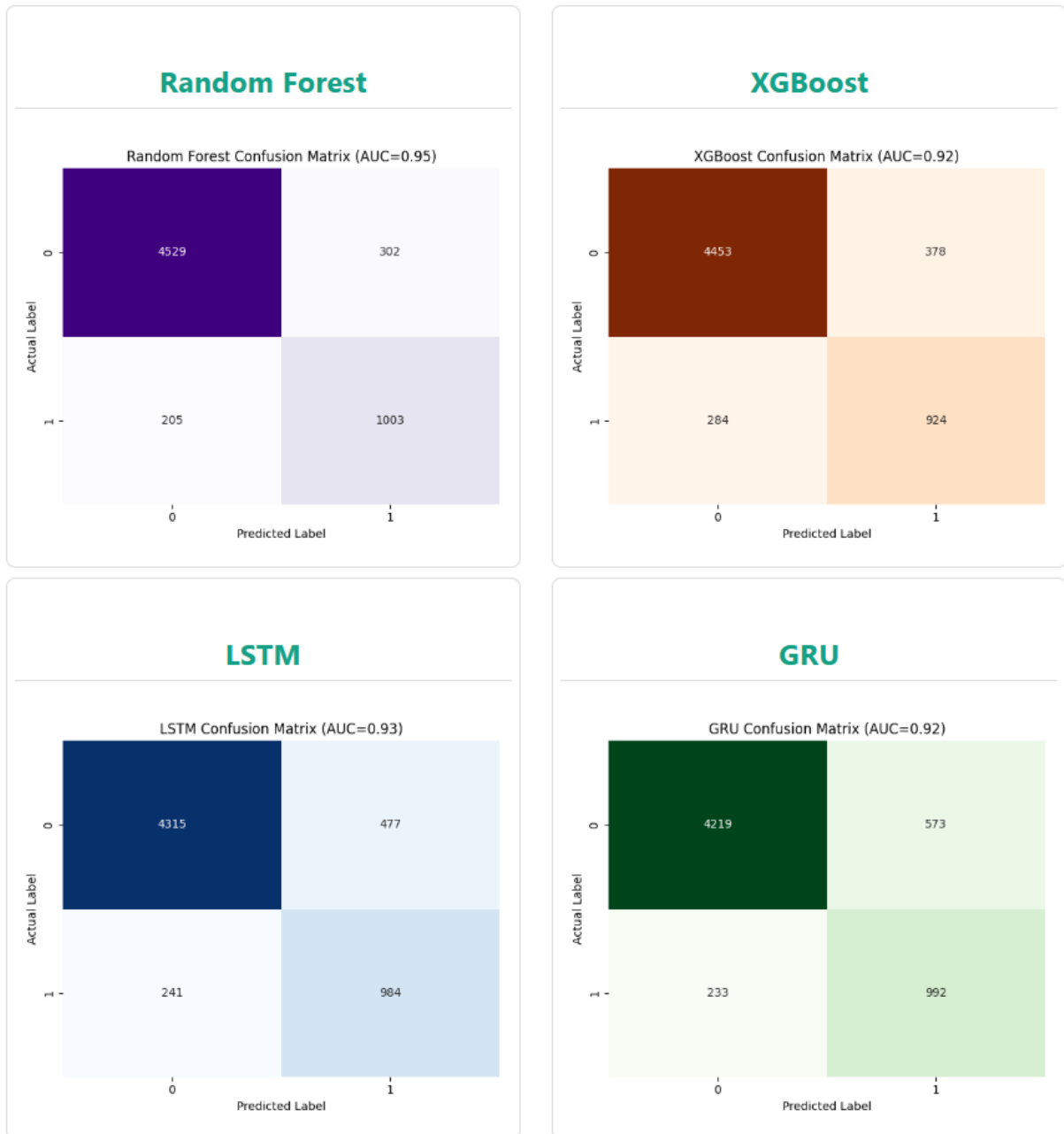
Εικόνα 5.8: Αποτελέσματα Classification / Regression και κατάταξη μοντέλων

3. Learning Curves (Deep Learning Training)



Εικόνα 5.9: Καμπύλες εκμάθησης (Μοντέλα DL)

4. Confusion Matrices (Classification Performance)

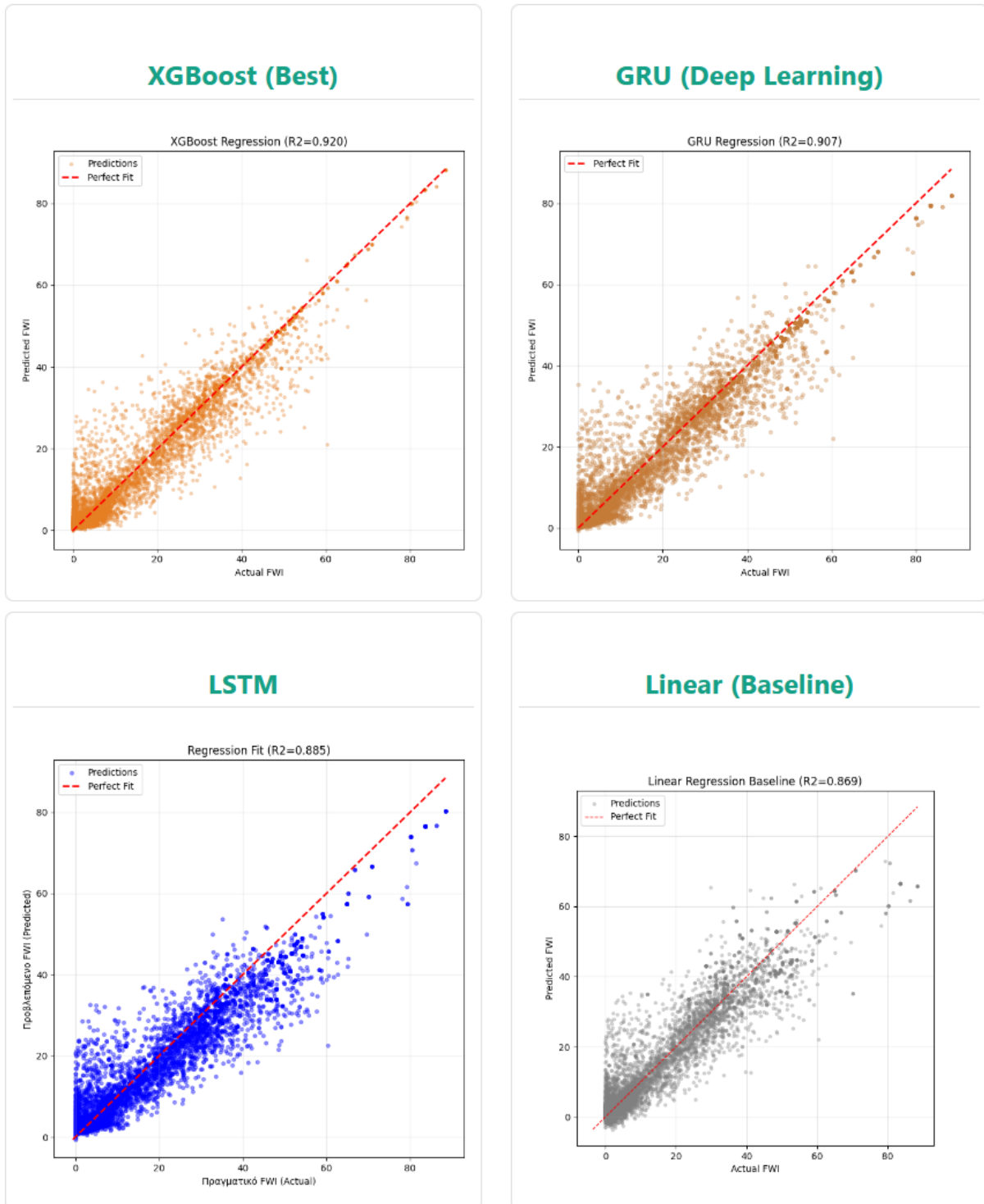


💡 Ερμηνεία Πίνακα Σύγκρισης:

- **True Negative (Πάνω Αριστερά):** Σωστή πρόβλεψη "Όχι Φωτιά".
- **False Positive (Πάνω Δεξιά):** Ψευδής συναγερμός (Προβλέψαμε Φωτιά, αλλά δεν είχαμε).
- **False Negative (Κάτω Αριστερά):** Χαμένη Φωτιά (Προβλέψαμε Όχι, αλλά είχαμε Φωτιά).
- **True Positive (Κάτω Δεξιά):** Σωστή πρόβλεψη "Φωτιά".

Εικόνα 5.10: Πίνακες Σύγκρισης

5. Regression Scatter Plots (Predicted vs Actual)



Εικόνα 5.11: Διάγραμμα Διασποράς Παλινδρόμησης

6. Feature Importance Analysis

6.1 Classification Models



Εικόνα 5.12: Ανάλυση Σημαντικότητας Χαρακτηριστικών (Classification Models)

6. Feature Importance Analysis

6.2 Regression Models



Εικόνα 5.13: Ανάλυση Σημαντικότητας Χαρακτηριστικών (Regression Models)

Συνοψίζοντας, ο παραπάνω μηχανισμός είναι ουσιαστικά το τελικό στάδιο επαλήθευσης της διαδικασίας. Παρέχει μια ολική εικόνα της απόδοσης τεκμηριώνοντας την επιλογή του βέλτιστου μοντέλου ανά κατηγορία που θα τροφοδοτηθεί τα δυο επόμενα συστήματα πρόγνωσης κινδύνου.

5.1.3 AI Fire Risk Forecast (Σύστημα Διαχείρισης Κινδύνου Πυρκαγιάς)

Το συγκεκριμένο script αποτελεί τον πυρήνα της επιχειρησιακής λειτουργίας του συστήματος, καθώς υλοποιεί την αυτοματοποιημένη ημερήσια πρόγνωση κινδύνου αξιοποιώντας το εκπαιδευμένο μοντέλο.

Αρχικά, το πρόγραμμα φορτώνει τα ιστορικά μετεωρολογικά δεδομένα και, ελέγχοντας την τρέχουσα ημερομηνία συστήματος, προσαρμόζει τη λειτουργία του είτε σε πραγματικό χρόνο (Live Mode - Αντιπυρική περίοδος) είτε σε προσομοίωση (Demo Mode), ζητώντας από τον χρήστη την ημερομηνία ενδιαφέροντος. *Να τονίσουμε πως η εκδοχή που βγάζει προσομοίωση, είναι για καθαρά λόγους παρουσίασης. Το σύστημα θα δουλεύει για όλες τις περιόδους χωρίς demo mode.* Στη συνέχεια, προετοιμάζει τα δεδομένα ομαδοποιώντας τα ανά γεωγραφική περιοχή (Grid) και δημιουργώντας χρονικές ακολουθίες 10 ημερών (Lookback), τις οποίες τροφοδοτεί στο νευρωνικό δίκτυο για να υπολογίσει την πιθανότητα εκδήλωσης πυρκαγιάς. Τα αποτελέσματα φιλτράρονται δυναμικά (διατηρώντας μόνο περιοχές με δείκτη FWI ≥ 10 , για λόγους υπολογιστικής ισχύος), ταξινομούνται βάσει επικινδυνότητας και ενσωματώνονται αυτόματα σε μια αναφορά HTML, η οποία περιλαμβάνει χρωματική κωδικοποίηση επιπέδων κινδύνου και συνδέσμους χαρτών, ανοίγοντας τελικά το δελτίο απευθείας στον φυλλομετρητή του χρήστη (Εικόνα 5.14).

<p style="text-align: center;">▲ ΛΕΙΤΟΥΡΓΙΑ ΠΡΟΣΟΜΟΙΩΣΗΣ (DEMO MODE): Η σημερινή ημερομηνία (01/02/2026) είναι εκτός αντιπυρικής περιόδου. Το σύστημα εκτελείται σε λειτουργία επίδειξης.</p> <p style="text-align: center;">Προβάλλονται ιστορικά δεδομένα για την ημερομηνία: 06/07/2022</p>							
#	ΠΕΡΙΟΧΗ	ΣΥΝΘΗΚΕΣ	ΠΙΘΑΝΟΤΗΤΑ	FWI (True)	DC	ΚΙΝΔΥΝΟΣ	ΧΑΡΤΗΣ
1	Meligalas (Peloponnese) Lat: 37.33, Lons: 22.01	34.1°C 12.2 km/h	92.0%	40.3	517.0	ΑΚΡΑΙΟΣ	View
2	Amaliada (West Greece) Lat: 37.77, Lons: 21.42	34.9°C 14.6 km/h	74.1%	38.9	517.5	ΠΟΛΥ ΥΨΗΛΟΣ	View
3	Nemea (Peloponnese) Lat: 37.88, Lons: 22.68	32.2°C 11.7 km/h	63.1%	39.2	516.1	ΠΟΛΥ ΥΨΗΛΟΣ	View
4	Karystos (Central Greece) Lat: 36.13, Lons: 24.37	24.8°C 18.8 km/h	56.7%	31.8	512.4	ΥΨΗΛΟΣ	View
5	Thasos (East Macedonia and Thrace) Lat: 40.80, Lons: 24.79	25.7°C 4.7 km/h	27.4%	19.7	512.9	ΧΑΜΗΛΟΣ	View

Εικόνα 5.14: Σύστημα Διαχείρισης Κινδύνου Πυρκαγιάς

Στην συγκεκριμένη περίπτωση (Demo Mode) και για την ημερομηνία 06/07/2022 το σύστημα «Ανακάλυψε» 12 περιοχές (εδώ παρουσιάζουμε τις πρώτες 5) για τις οποίες το σύστημα έβγαλε alarm. Αυτό που βλέπουμε στην εικόνα, είναι ένας πίνακας 12 περιοχών που με βάση τα καιρικά τους φαινόμενα για την συγκεκριμένη ημερομηνία που δώσαμε (Demo Mode) έβγαλε τα alarm. Οι μετρικές του FWI και DC έρχονται από τα ιστορικά δεδομένα του dataset_final.csv. Στην Live λειτουργεί όμως, το σύστημα υπολογίζει από την αρχή το παράθυρο των 10 ημερών με βάση τα καιρικά φαινόμενα της εκάστοτε περιοχής ώστε τα αποτελέσματα να αντικατοπτρίζουν την πραγματικότητα. Έτσι ο χειριστής του συστήματος έχει μια εικόνα όλης της Ελλάδας με το alarm αν περιοχή (grid).

5.1.4 AI Wildfire Risk Simulation (Live Πρόγνωση Κινδύνου Πυρκαγιών)

Κλείνοντας το κομμάτι της παραγωγής των συστημάτων, τελευταία εφαρμογή είναι η **AI Wildfire Risk Simulation**. Η εφαρμογή αυτή αποτελεί τον ακρογωνιαίο λίθο της παρούσας διπλωματικής. Είναι ένα ολοκληρωμένο περιβάλλον προσομοίωσης και πρόγνωσης που δεν χρησιμοποιεί απλά την χρήση και την απεικόνιση των αλγορίθμων ML και DL, αλλά εισάγει και μια δικλείδα ασφαλείας υβριδικής λογικής για τη λήψη αποφάσεων.

Είναι πολύ λογικό τα μοντέλα AI, ενδέχεται να φέρουν και ψευδείς συναγερμούς (false Negatives), όταν οι συνθήκες είναι σπάνιες ή ακραίες. Έτσι το σύστημα ενσωματώνει ένα μηχανισμό ασφαλείας. Η λογική αυτή δεν μπαίνει σε πρώτο πλάνο, αλλά λειτουργεί σαν επιδιόρθωση των μοντέλων τεχνικής

Κεφάλαιο 5

νοημοσύνης, ελέγχοντας διαρκώς αν η πρόβλεψη συμβαδίζει με τους φυσικούς δείκτες κινδύνου (FWI). Σε περίπτωση απόκλισης, όπου ο δείκτης FWI (μαθηματικές εξισώσεις) υποδεικνύει ακραίο κίνδυνο αλλά το μοντέλο υπό-εκτιμά αυτόν, το σύστημα ενεργοποιεί αυτόματα συναγερμό βάσει των φυσικών κανόνων, αγνοώντας το μοντέλο πρόβλεψης. Με αυτό τον τρόπο επιτυγχάνεται η βέλτιστη ισορροπία της ακρίβειας των μοντέλων AI Και της αξιοπιστίας των φυσικών κανόνων.

Το σκεπτικό για την υλοποίηση του συγκεκριμένου συστήματος Βασίστηκε σε 2 πυλώνες: αφενός η επιχειρησιακή πρόγνωση σε πραγματικό χρόνο και αφετέρου η επιστημονική επαλήθευση των αποξεσμάτων που παρουσιάστηκε στα προηγούμενα κεφάλαια.

Στην **Εικόνα 5.15** παρουσιάζουμε την αρχική οθόνη του συστήματος.



Εικόνα 5.15: Αρχική οθόνη AI Wildfire Risk Simulator

Στην παραπάνω εικόνα φαίνονται οι βασικές επιλογές / λειτουργίες του συστήματος. Περιληπτικά αυτές είναι:

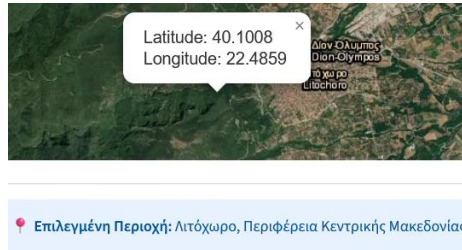
- Διαδραστικός Χάρτης: Ο χώρος όπου ο χρήστης επιλεγεί το σημείο ενδιαφέροντος ή οπτικοποιεί την τοποθεσία του συμβάντος (Ιστορική Επαλήθευση).
- Επιλογή Λειτουργίας: Δυνατότητα λειτουργίας είτε σε «Ζωντανή Πρόβλεψη», είτε σε «Ιστορικής Επαλήθευσης».
- Παράμετροι Μοντέλου: Πίνακας προβολής μετεωρολογικών δεδομένων (θερμοκρασία, υγρασία, άνεμος, βροχή), είτε σε πραγματικό χρόνο, είτε ιστορικά συμβάντα.
- Προβολή Αναφορών: Επιλογή προβολής αποτελεσμάτων, ενεργοποιώντας τα τρία υποσυστήματα της παρούσας διπλωματικής.

Η διεπαφή της εφαρμογής σχεδιάστηκε με γνώμονα την ευχρηστία και την εύκολη κατανόηση του συστήματος από τον τελικό χρήστη. Παρακάτω ακολουθεί αναλυτική περιγραφή των τεσσάρων βασικών λειτουργιών του συστήματος:

5.1.4.1 Διαδραστικός Χάρτης

Όπως είδαμε και στην Εικόνα 5.15, στο κέντρο της υπάρχει ένας μηχανισμός γεωχωρικής επιλογής, οποίος υλοποιήθηκε με την χρήση της βιβλιοθήκης **Folium**. Ο χρήστης μπορεί να επιλέξει το σημείο ενδιαφέροντος.

Το σύστημα αυτόματα κάνει μετατροπή το στίγμα σε συντεταγμένες (Latitude/Longitude). Στην συνέχεια κάνει έλεγχο εγκυρότητας σημείου. Συγκεκριμένα με την χρήση της βιβλιοθήκης **geopy**, διασφαλίζεται ότι το επιλεγμένο σημείο είναι εντός της Ελληνικής επικράτειας και δεν είναι θαλάσσια περιοχή. Αυτό έγινε για την αποφυγή λανθασμένων προβλέψεων (θάλασσα ή εκτός Ελλάδας). Και τέλος το σύστημα μας εμφανίζει την ονομασία της περιοχής (ή όνομα τοποθεσίας) (**Εικόνα 5.16**).



Εικόνα 5.16: Επιλογή σημείου στο Χάρτη

5.1.4.2 Επιλογή Λειτουργίας Συστήματος

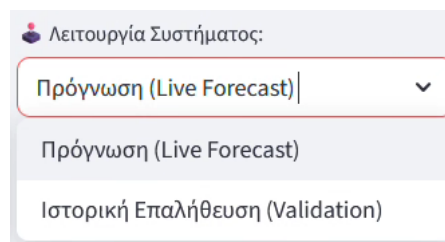
Με την επιλογή αυτή, ο χρήστης καθορίζει την ροή δεδομένων που θα ακολουθήσει η εφαρμογή για να βγάλει τα επιθυμητά αποτελέσματα επιλέγοντας μια από τις δυο καταστάσεις.

- Πρόγνωση (Live Forecast): Με την επιλογή αυτή, το σύστημα συνδέεται σε πραγματικό χρόνο με το API του Open-Meteo, αντλώντας τα πραγματικά μετεωρολογικά δεδομένα που επικρατούν στην περιοχή. Με βάση αυτά, το σύστημα κάνει χρήση και του Looorback δέκα ημερών, ώστε να βγάλει αποτελέσματα πρόβλεψης κινδύνου.
- Ιστορική Επαλήθευση (Validation): Για μεγαλύτερη εμπειρία χρήστη, υλοποιήσαμε μια επιλογή που με βάση ιστορικές μεγάλες πυρκαγιές στην Ελλάδα (π.χ. Μάτι 2018, Βαρνάβας 2024, Δαδιά 2023), το σύστημα αντλεί τις ακριβείς συνθήκες που επικρατούσαν την ημέρα εκείνη και συγκρίνει με τις προβλέψεις του συστήματος. Τα μηνύματα που εμφανίζει είναι τα παρακάτω:

→Επιλέξατε περιοχή εκτός Ελλάδας (Χώρα: TR). Προσπαθήστε ξανά.

→Η επιλεγμένη τοποθεσία βρίσκεται στη θάλασσα. Παρακαλώ επιλέξτε χερσαία περιοχή.

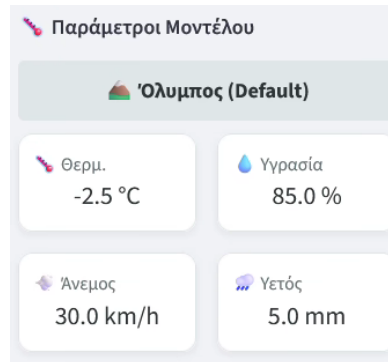
Ανάλυση λειτουργίας του συστήματος θα γίνει σε επόμενη παράγραφο.



Εικόνα 5.17: Επιλογές λειτουργίας συστήματος

5.1.4.3 Παράμετροι Μοντέλου

Αριστερά στην πλευρική στήλη και κάτω από τις επιλογές του συστήματος υπάρχουν τέσσερες καιρικές μεταβλητές: **Θερμοκρασία (°C)**, **Υγρασία (%)**, **Άνεμος (km/h)** και **Υετός (mm)**. Οι τιμές που θα αντικρίσει ο χρήστης κατά την εισαγωγή του στο σύστημα, είναι με βάση το default σημείο αναφοράς που έχουμε ορίσει αρχικά και αυτό είναι ο Όλυμπος (Εικόνα 5.18). Οι τιμές που βλέπουμε δεν είναι τυχαίες, αλλά οι πραγματικές τιμές που επικρατούν την συγκεκριμένη στιγμή.



Εικόνα 5.18: Εμφάνιση καιρικών δεδομένων για το όρος Όλυμπος (default σημείο)

Σε περίπτωση **Ιστορικής Επαλήθευσης** και επιλογής ιστορικού συμβάντος, οι καιρικοί παράμετροι που εμφανίζει είναι οι καιρικές συνθήκες που επικρατούσαν την ημέρα της πυρκαγιάς.

Τέλος στην Live λειτουργία, ο χρήστης μπορεί επιπλέον να τροποποιήσει τους καιρικούς δείκτες (με slide bar) για εξέταση σεναρίων. Δηλαδή τι θα γίνει αν...

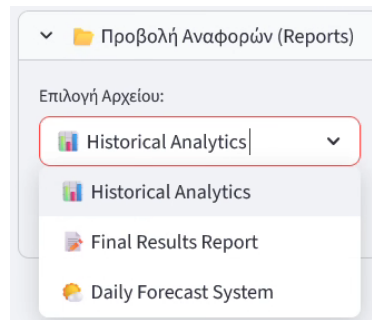
- **Για παράδειγμα:** Τι θα γίνει αν (μεταβολή κινδύνου) ο άνεμος αυξηθεί από 20km/h σε 60km/h; Αυτή η λειτουργία κάνει την εφαρμογή ένα ισχυρό εργαλείο λήψης αποφάσεων.



Εικόνα 5.19: Τροποποίηση τιμών μετεωρολογικών δεικτών

5.1.4.4 Προβολή Αναφορών (Reports)

Στο κάτω μέρος της πλαϊνής μπάρας ο χρήστης έχει πρόσβαση στα τρία υποσυστήματα που αναφέρθηκαν στις παραπάνω παραγράφους. Ουσιαστικά είναι ένα dropdown menu με τα link των υποσυστημάτων.



Εικόνα 5.20: Dropdown Menu (Προβολή Αναφορών)

5.1.4.5 Επίδειξη λειτουργίας Συστήματος AI Wildfire Risk Simulation

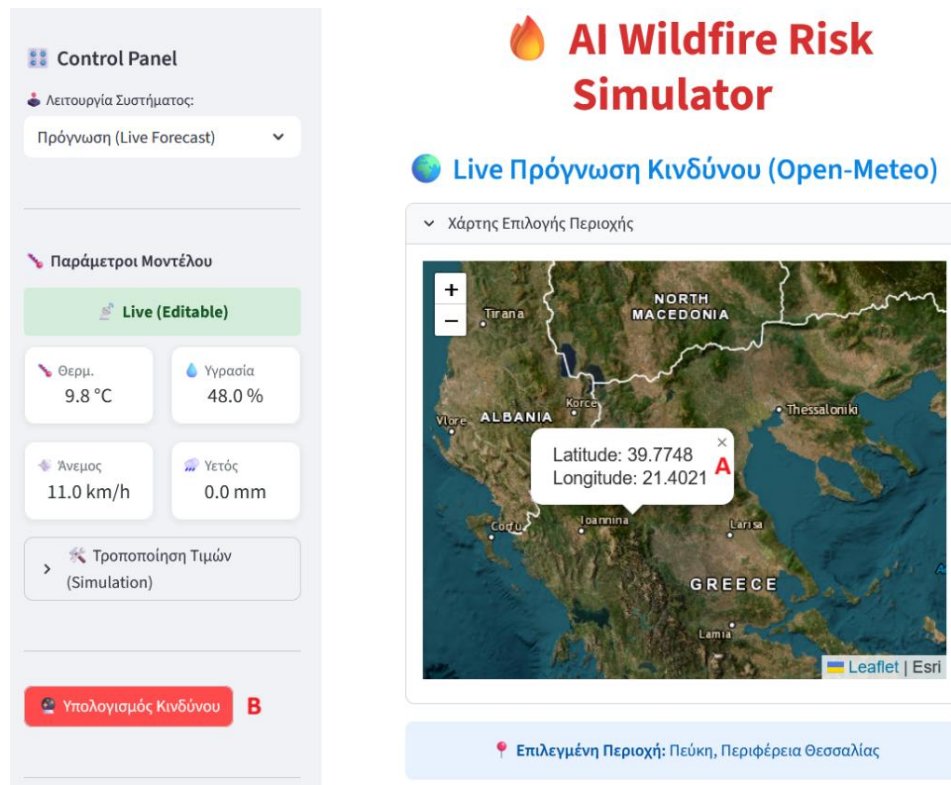
Όπως αναφέρθηκε και παραπάνω, οι δυο κύριες λειτουργίες του συστήματος είναι οι παρακάτω:

- Λειτουργία «Πρόγνωσης (Live Forecast)»:

Έχοντας αναφέρει παραπάνω τις βασικές ιδιότητες της λειτουργίας αυτής, θα προχωρήσουμε και στην επίδειξη. Η λειτουργία του είναι πολύ εύκολη και απολύτως κατανοητή στον απλό χρήστη. Η λειτουργία αυτή ακολουθεί της εξής ροή:

α. Επιλογή Σημείου (Χάρτη) → **β.** Υπολογισμός Κινδύνου (Κουμπί) → **γ.** Προβολή Αποτελεσμάτων

α & β) Στα πλαίσια της προσομοίωσης επιλέχτηκε ένα τυχαίο μέρος στον Χάρτη, που στην προκυμμένη περίπτωση αφορά το χωριό **Πεύκη** στην **Περιφέρεια Θεσσαλίας** [48], ένα ορεινό χωριό στον Δήμο Μαλακασίου που βρίσκεται σε υψόμετρο 850μ που βρίσκετε μέσα στο δάσος [46]. Είναι φανερό ότι μιλάμε για μια περιοχή με αρκετή **υγρασία (48%)** τη στιγμή που έγινε επιλογή του σημείου, με ταχύτητα ανέμου στα **11km/h**, καθόλου βροχή (**0.0 mm**) και θερμοκρασία στους **9.8 °C** (**Εικόνα 5.21**) Θα λέγαμε ότι οι συνθήκες δεν εμπνέουν για ένδειξη εκδήλωσης πυρκαγιάς. Πατώντας το κουμπί «**Υπολογισμός Κινδύνου**» (**β**), σύστημα θα κάνει όλους τους υπολογισμούς και θα μας φέρει όλες τις τιμές των δεικτών για την πρόβλεψη κινδύνου (**γ**).



Εικόνα 5.21: Πρόβλεψη για Χωρίο Πεύκη (περιφέρεια Θεσσαλίας)

Τα αποτελέσματα που εμφανίζει το σύστημα στον χρήστη, χωρίζονται σε τέσσερις κατηγορίες. Την γενική εκτίμηση κινδύνου έως την εξειδικευμένη ανάλυση δεικτών.

Κύριοι Δείκτες Κινδύνου και Μπάρα Συναγερμού: Κάτω από τον Χάρτη, αρχικά το σύστημα μας υπενθυμίζει ποια είναι τα βέλτιστα μοντέλα, τα οποία χρησιμοποιεί το σύστημα για τις προβλέψεις του. Ακριβώς από κάτω παρουσιάζονται τρία μεγέθη που προκύπτουν από την επεξεργασία (Εικόνα 5.22).

- FWI (Classic Calc): Η τιμή του δείκτη, η οποία υπολογίστηκε από τις μαθηματικές εξισώσεις του συστήματος Van Wagner. Το χρησιμοποιούμε σαν μέτρο σύγκρισης.
- Πιθανότητα (AI): Είναι η τελική εκτίμηση κινδύνου πυρκαγιάς (0-100%). Η τιμή αυτή εξάγεται από την Υβριδική Λογική, συνδυάζοντας την πρόβλεψη του μοντέλου με τους κανόνες ασφαλείας που έχουμε θέσει.
- Πρόβλεψη FWI (AI): Η εκτίμηση του μοντέλου Παλινδρόμησης (Regression).



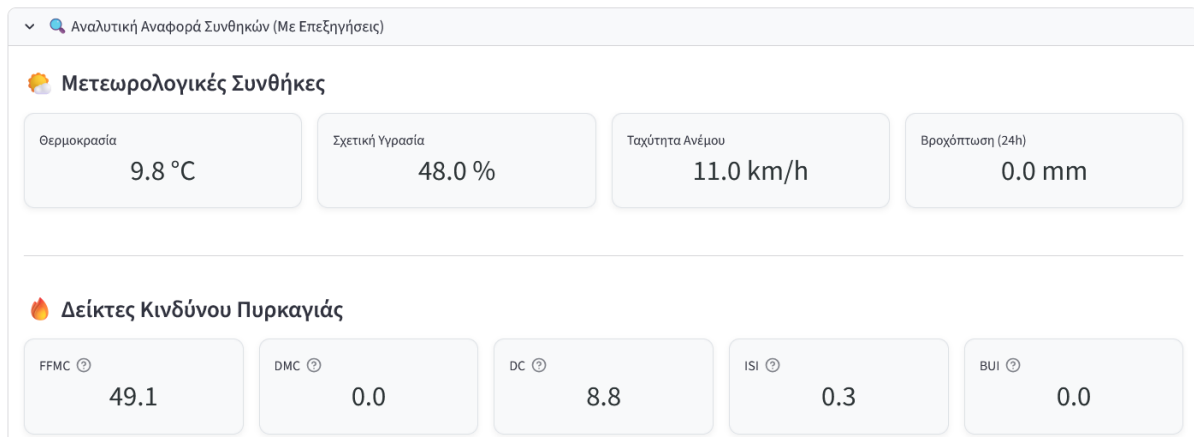
Εικόνα 5.22: Κύριοι Δείκτες Κινδύνου και Μπάρα Συναγερμού

Η **Μπάρα Συναγερμού** είναι ουσιαστικά μια μπάρα κατάστασης που ταξινομείται με βάση συγκεκριμένα νούμερα που έχουν οριστεί από το σύστημα CFFDRS. Ποιο συγκεκριμένα έχουμε τις εξής χρωματικές μεταβολές:

- **ΑΣΦΑΛΕΙΑ:** Χαμηλός Κίνδυνος.
- **Μέτριος Κίνδυνος**
- **ΠΡΟΣΟΧΗ:** Υψηλός Κίνδυνος
- **ΣΥΝΑΓΕΡΜΟΣ:** Πολύ Υψηλός Κίνδυνος!
- **ΕΞΑΙΡΕΤΙΚΑ ΚΡΙΣΙΜΟ:** Ακραίος Κίνδυνος!
- **ΚΙΝΔΥΝΟΣ ΘΑΝΑΤΟΣ:** Κατάσταση Συναγερμού!

Αναφορά Συνθηκών: Για να κατανοήσει ο χρήστης για το πως προέκυψαν τα αποτελέσματα το σύστημα εμφανίζει (**Εικόνα 5.23**):

- **Μετεωρολογικές Συνθήκες:** Τους Μετεωρολογικούς Δείκτες που έτσι και αλλιώς είχαμε δει στην πλαϊνή μπάρα.
- **Υπο-δείκτες FWI:** Όλα τα αποτελέσματα των δεικτών **FFMC, DMC, DC, ISI, BUI**. Οι τιμές των δεικτών αυτών είναι πολύ σημαντικές, για την εξαγωγή σημαντικών συμπερασμάτων. Βλέποντας τα νούμερα αυτά, μπορούμε να διακρίνουμε αν ο κίνδυνος οφείλεται για παράδειγμα στην επιφανειακή ξηρασία (FFMC) ή στην ξηρασία βάθους (DC). Περνώντας τον δείκτη του ποντικιού στα ερωτηματικά (?) πάνω από κάθε δείκτη, μας εμφανίζεται η πλήρη ονομασία του καθενός.



Εικόνα 5.23: Αναφορά Συνθηκών

Αναφοράς Δεικτών & Ιστορικότητα Καιρού: Τέλος, για την ερμηνεία των αποτελεσμάτων παρέχονται:

- **Πίνακας Κλίμακας Επικινδυνότητας (Εικόνα 5.24):** Παρουσιάζονται όλοι οι δείκτες του συστήματος CFFDRS, όπου τους χωρίσαμε σε αριθμητικές κλάσεις με βάση συγκεκριμένα πρότυπα [47]. Με βάση τα αποτελέσματα των υπολογισμών, χρωματίζει αυτόματα την κατηγορία τιμών που προκύπτει από τις τιμές υπολογισμού.
- **Ιστορικά Δεδομένα Καιρού - Δεδομένα Εισόδου (Εικόνα 5.25):** Είναι ο πίνακας των καιρικών συνθηκών των 10 προηγούμενων ημερών (LOOPBACK=10). Αυτό μας δείχνει τις τιμές που έλαβε υπόψη το μοντέλο, ώστε να βγάλει την τελική πρόβλεψη. Επίσης είναι

Κεφάλαιο 5

και ένα καλό συμπέρασμα για να αντιληφθεί ο χρήστης πως οι καιρικές συνθήκες επηρεάζουν την πρόβλεψη κινδύνου δασικών πυρκαγιών.

Κλίμακα Επικινδυνότητας (Fire Danger Classes)

Class	FWI (Δείκτης Καιρού)	FFMC (Λεπτή Ύλη)	DMC (Μέσο Στρώμα)	DC (Δείκτης Ξηρασίας)	ISI (Ταχ. Εξάπλωσης)	BUI (Καύσιμη Ύλη)
Low	< 11.2	< 82.7	< 15.7	< 256.1	< 3.2	< 24.2
Moderate	11.2 - 21.3	82.7 - 86.1	15.7 - 27.9	256.1 - 334.1	3.2 - 5.0	24.2 - 40.7
High	21.3 - 38.0	86.1 - 89.2	27.9 - 53.1	334.1 - 450.6	5.0 - 7.5	40.7 - 73.3
Very High	38.0 - 50.0	89.2 - 93.0	53.1 - 83.6	450.6 - 600.0	7.5 - 13.4	73.3 - 133.1
Extreme	50.0 - 70.0	93.0 - 96.0	83.6 - 160.7	600.0 - 749.4	13.4 - 26.8	133.1 - 193.1
Very Ext.	> 70.0	> 96.0	> 160.7	> 749.4	> 26.8	> 193.1

Πηγή: Copernicus EFFIS - Fire Danger Forecast

Εικόνα 5.24: Πίνακας Κλίμακα Επικινδυνότητας των δεικτών του CFFDRS [49]

Ημερομηνία	Θερμοκρασία (°C)	Υγρασία (%)	Άνεμος (km/h)	Βροχή (mm)	
24/01/2026		5.5	87	5.8	0.1
25/01/2026		8	98	7.2	0
26/01/2026		8.7	67	10.9	8.4
27/01/2026		11	52	10.7	0.3
28/01/2026		7.8	53	8.6	2.6
29/01/2026		8	83	5.7	14.6
30/01/2026		11	46	15.8	0.2
31/01/2026		11.3	44	8	8.4
01/02/2026		5.7	84	10	60.3
02/02/2026 (Στόχος)		9.8	48	11	0

Εικόνα 5.25: Ιστορικά Δεδομένα Καιρού - Δεδομένα Εισόδου

Συμπερασματικά για το συγκεκριμένο σενάριο της Πεύκη Τρικάλων, το σύστημα μας απάντησε σωστά (True Negative). Παρόλο που την μέρα την πρόγνωσης δεν υπήρχε βροχή, δεδομένου της εποχής (Χειμώνας) και του παραθύρου των 10 ημερών, όπου αναγνώρισε έντονη βροχόπτωση την προηγούμενη ημέρα, η πιθανότητα πρόβλεψης κρατήθηκε σε φυσιολογικά για την εποχή επίπεδα με μόλις 2.1% πιθανότητα.

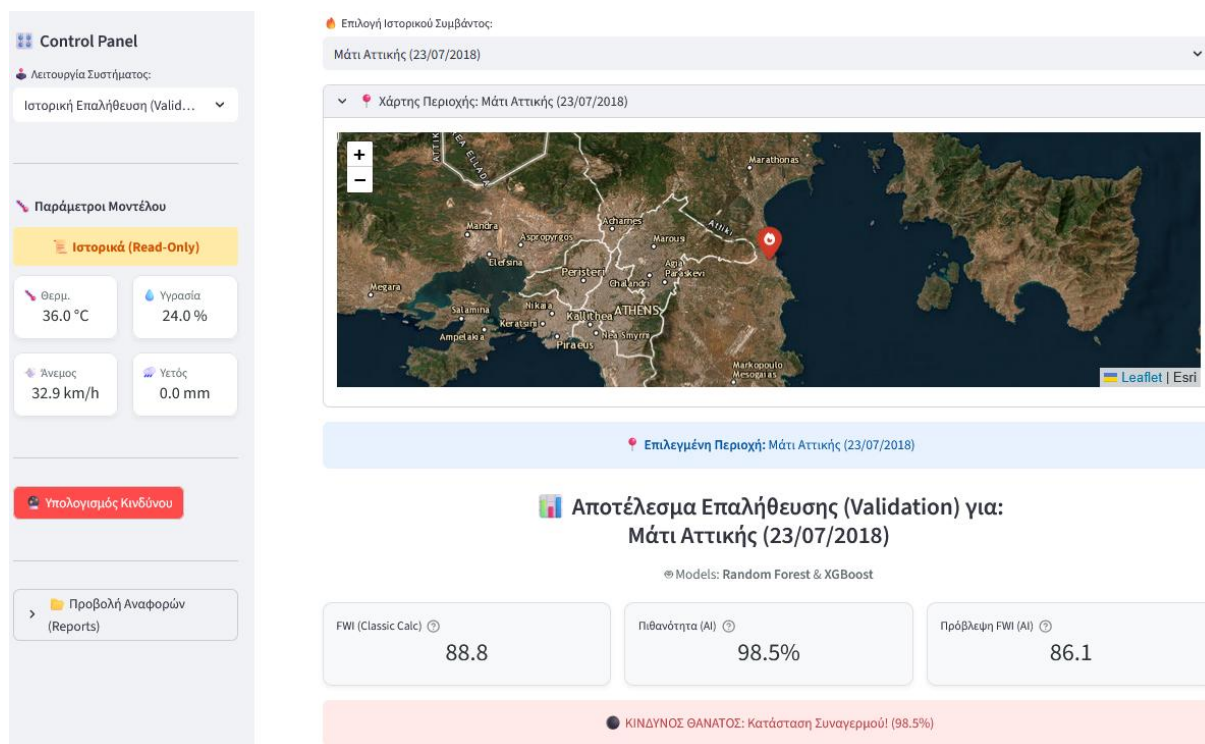
- Λειτουργία «Ιστορικής Επαλήθευσης (Validation)»:

Ένα από τα πιο δυνατά σημεία του συστήματος είναι η **Ιστορική Επαλήθευση**. Μας επιτρέπει να αναπαραστήσουμε παρελθοντικά γεγονότα και να ελέγξουμε αν το μοντέλο θα τα είχε προβλέψει σωστά, με βάση τις συνθήκες που ίσχυαν τότε. Η λειτουργία αυτή ακολουθεί την εξής ροή:

α. Επιλογή Ιστορικού Συμβάν → **β.** Υπολογισμός Κινδύνου (Κουμπί) → **γ.** Προβολή Αποτελεσμάτων

Αξίζει να σημειωθεί ότι το σύστημα εμφανίζει με «**πινέζα**» το σημείο εστίας της πυρκαγιάς. Κάνοντας κλήση στο API του Open-Meteo για παρελθοντικά δεδομένα, μας παρουσιάζει τις μετεωρολογικές συνθήκες που επικρατούσαν εκείνη την ημέρα (και τα καιρικά δεδομένα 10 μέρες πριν την πυρκαγιά).

Τρέξαμε το σενάριο για την φονική πυρκαγιά στο **Μάτι Αττικής στις 23/7/2018**. Όπως βλέπουμε στην **Εικόνα 5.26** το σύστημα προέβλεψε σωστά με πιθανότητα **98.5%**.



Εικόνα 5.26: Μάτι Αττικής (23/7/2018)

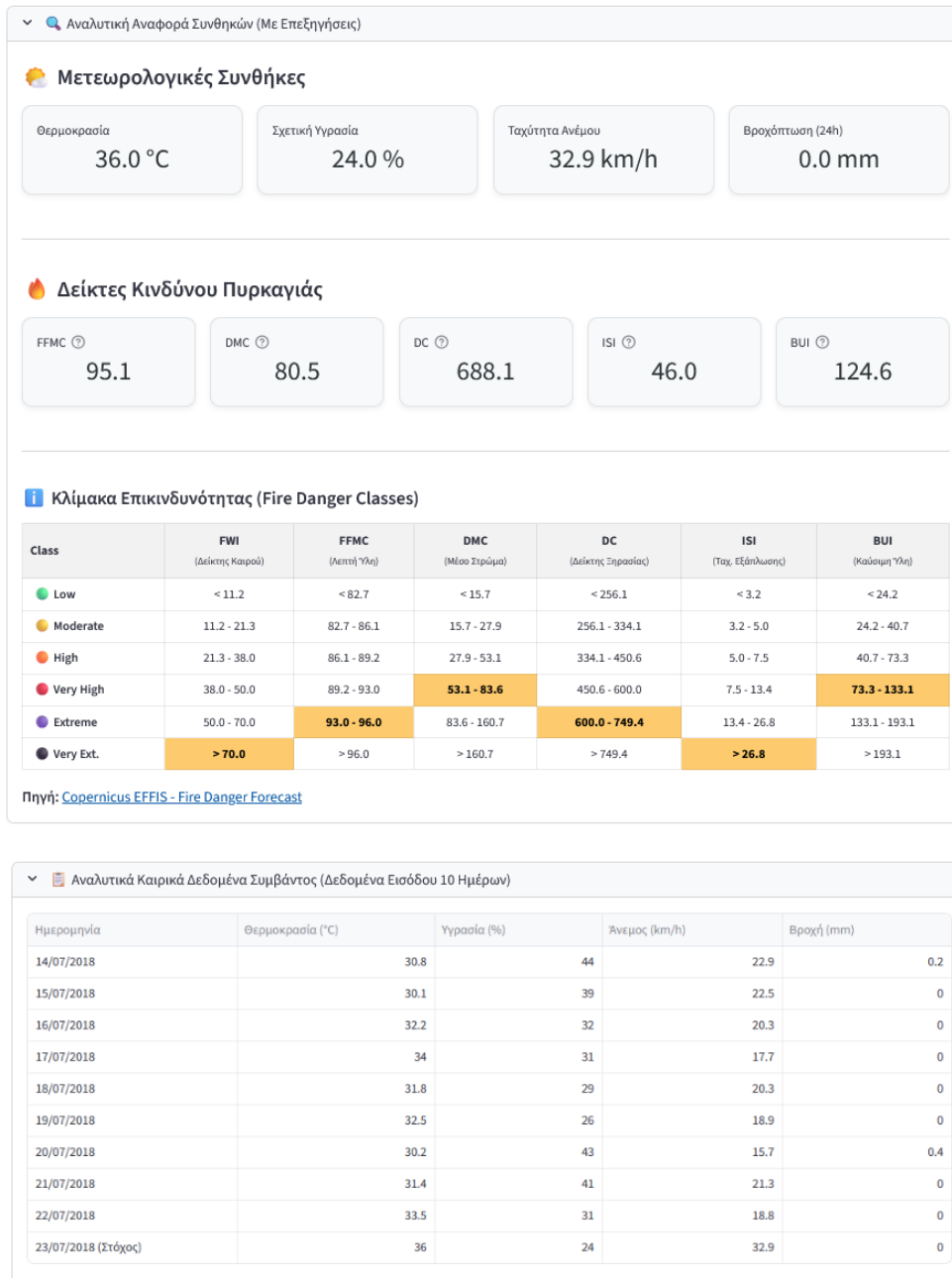
Όλοι οι δείκτες εκείνης της ημέρας κυμαίνονταν σε **ακραία υψηλά επίπεδα**, συνδυάζοντας **ισχυρούς ανέμους, παντελή απουσία βροχοπτώσεων** το προηγούμενο διάστημα και ιδιαίτερα **υψηλές θερμοκρασίες**.

Τέλος στον **Κώδικα 5.1** φαίνονται τα σενάρια ιστορικών συμβάντων που εισάγαγέ στο σύστημα προς επαλήθευση.

```
# =====
# ΔΕΔΟΜΕΝΑ ΙΣΤΟΡΙΚΩΝ ΠΥΡΚΑΓΙΩΝ
# =====
FAMOUS_FIRES = {
  "Επιλέξτε Ιστορικό Συμβάν...": None,
  "Βά Αττική - Βαρνάβας (11/08/2024)": {"lat": 38.22, "lon": 23.92, "date": "2024-08-11"},
  "Ξυλόκαστρο Κορινθίας (29/09/2024)": {"lat": 38.08, "lon": 22.58, "date": "2024-09-29"},
  "Δαδιά Εβρου (21/08/2023)": {"lat": 41.12, "lon": 26.22, "date": "2023-08-21"},
  "Πάρνηθα - Φωλή (22/08/2023)": {"lat": 38.12, "lon": 23.68, "date": "2023-08-22"},
  "Ρόδος (18/07/2023)": {"lat": 36.17, "lon": 27.92, "date": "2023-07-18"},
  "Πεντέλη (19/07/2022)": {"lat": 38.06, "lon": 23.90, "date": "2022-07-19"},
  "Βαρυμύμη (03/08/2021)": {"lat": 38.13, "lon": 23.79, "date": "2021-08-03"},
  "Λίμνη Ευβοίας (03/08/2021)": {"lat": 38.76, "lon": 23.32, "date": "2021-08-03"},
  "Βίλια Αττικής (16/08/2021)": {"lat": 38.16, "lon": 23.33, "date": "2021-08-16"},
  "Σχίνος Κορινθίας (19/05/2021)": {"lat": 38.03, "lon": 23.06, "date": "2021-05-19"},
  "Μάτι Αττικής (23/07/2018)": {"lat": 38.04, "lon": 23.99, "date": "2018-07-23"},
}
```

Κώδικας 5.1: Ιστορικές Πυρκαγιές στην Ελλάδα

Κεφάλαιο 5



Εικόνα 5.27: Δείκτες Σεναρίου - Μάτι Αττικής

Κεφάλαιο 6ο: Συμπεράσματα, Περιορισμοί και Μελλοντικές Επεκτάσεις

6.1 Διερεύνηση της Επίδρασης Ανισορροπίας Κλάσεων

Στην πειραματική διαδικασία, δεν ασχοληθήκαμε αποκλειστικά και μόνο με ένα σύνολο δεδομένων για τις αρχιτεκτονικές που αναφέρθηκαν παραπάνω. Κάναμε τέσσερα διαφορετικά σενάρια ανισορροπίας δεδομένων. (1:1, 1:4, 1:10 και 1:200). Τα αποτελέσματα αυτών οδήγησαν στα παρακάτω συμπεράσματα:

1. Περίπτωση 1:1 (Απόλυτη Ισορροπία - Εργαστηριακό Σενάριο) (Πίνακας 6.1)

- **Classification:** Εδώ σημειώθηκαν οι **υψηλότερες επιδόσεις** όλων των πειραμάτων. Πολύ λογικό γιατί υπήρχε η απόλυτη ισορροπία. Πιο συγκεκριμένα ο Random Forest με **Precision 91.2%** και **Recall 87.9%** αναδείχθηκε πρώτος. Αν και οι επιδόσεις ήταν εξαιρετικές, η περίπτωση αυτή θεωρήθηκε ως μη ρεαλιστική καθώς είναι αδύνατο να υπάρχει αναλογία 1:1.
- **Regression:** Στην περίπτωση αυτή το **XGBoost** είχε σχεδόν το απόλυτο στην προσαρμογή (**R² 0.9997**, **MAE 0.0377**), δείχνοντας την φοβερή Του μοντέλου σε τέτοιου είδους δεδομένα.

Πίνακας 6.1: Αποτελέσματα Περίπτωσης 1:1

1:1				
Classification Results				
Model	ROC AUC	Accuracy	Precision	Recall
Random Forest	0,9618	89.7%	91.2%	87.9%
XGBoost	0,9543	88.3%	89.6%	86.8%
GRU	0,9355	85.9%	88.9%	82.4%
LSTM	0,9303	85.4%	87.8%	82.6%

Regression Results			
Model	R2 Score	MAE	RMSE
XGBoost	0,9997	0,0377	0,0999
Linear Reg	0,9890	0,4277	0,6603
LSTM	0,9404	0,9327	1,5484
GRU	0,9368	0,9967	1,5939

2. Περίπτωση 1:4 (Πίνακας 6.2)

- **Classification:** Η «χρυσή τομή». Και εδώ ο **Random Forest** είχε το υψηλότερο **F1-Score (79.8%)**, συνδυάζοντας πολύ καλή ικανότητα εντοπισμού (**Recall 83.0%**) και πολύ καλή αξιοπιστία (**Precision 76.9%**). Καταλήξαμε σε αυτή την προσέγγιση, λόγω την μεγάλης ασφάλειας, χωρίς υπερβολή στους ψευδής συναγεμμούς.
- **Regression:** Και εδώ το **XGBoost** παραμένει πολύ στιβαρό (**R²= 0.9202**) με αρκετά χαμηλό σφάλμα (**MAE = 2.97**), αποδεικνύοντας ότι η αναλογία 1:4 επιτρέπει ακριβή πρόβλεψη της έντασης του δείκτη.

Πίνακας 6.2: Αποτελέσματα Περίπτωσης 1:4

Classification Results					
Model	ROC AUC	Accuracy	F1-Score	Precision	Recall
Random Forest	0,9510	91.6%	79.8%	76.9%	83.0%
GRU	0,9284	87.9%	73.5%	66.3%	82.4%
LSTM	0,9255	88.1%	73.3%	67.4%	80.3%
XGBoost	0,9168	89.0%	73.6%	71.0%	76.5%

Regression Results			
Model	R2 Score	MAE	RMSE
XGBoost	0,9202	2,9727	4,8815
GRU	0,9067	3,3336	5,2773
LSTM	0,8853	3,9383	5,8533
Linear Reg	0,8693	4,1096	6,2475

3. Περίπτωση 1:10 (Πίνακας 6.3)

- **Classification:** Ενώ η Ακρίβεια (Accuracy) ανέβηκε φαινομενικά στο **93.1%**, Για τα υπόλοιπα παρατηρείται σημαντική πτώση. Το το**F1-Score έπεσε στο 67.9%** και το **Precision στο 59.1%**. Το μοντέλο αρχίζει να χάνει την ισορροπία του υπέρ της «μεγάλης» κλάσης (No-Fire).
- **Regression:** Το **XGBoost** εμφάνισε πτώση στο $R^2 = 0.8926$ και αύξηση του σφάλματος **MAE = 3.16**, δείχνοντας ότι η αύξηση των αρνητικών δειγμάτων εισάγει "**θόρυβο**" στην πρόβλεψη τιμών.

Πίνακας 6.3: Αποτελέσματα Περίπτωσης 1:10

Classification Results					
Model	ROC AUC	Accuracy	F1-Score	Precision	Recall
🔪 Random Forest	0,9474	93.1%	67.9%	59.1%	79.8%
🔪 GRU	0,9345	89.0%	57.5%	43.8%	83.5%
🔪 XGBoost	0,9100	90.9%	59.2%	50.1%	72.3%
LSTM	0,9040	89.7%	56.7%	45.5%	75.3%




Regression Results			
Model	R2 Score	MAE	RMSE
🔪 XGBoost	0,8926	3,1596	5,0674
🔪 GRU	0,8782	3,538	5,3949
🔪 LSTM	0,8732	3,5394	5,5053
Linear Reg	0,8554	3,7555	5,8780




4. Περίπτωση 1:200 (Φυσική Κατανομή - Το Πρόβλημα) (Πίνακας 6.4)

- **Classification:** Το φαινόμενο του "Accuracy Paradox". Το **LSTM** έδειξε τεράστια Ακρίβεια (**96.7%**), αλλά στην πραγματικότητα απέτυχε πλήρως, με **F1-Score μόλις 18.9%**. Το σύστημα έμαθε να προβλέπει συνεχώς "Όχι Φωτιά", αγνοώντας τα πραγματικά συμβάντα.

- **Regression:** Η πιο δύσκολη περίπτωση. Οι επιδόσεις έπεσαν στο χαμηλότερο επίπεδο ($R^2 \sim 0.85-0.86$) για XGBoost/GRU, καθώς το πλήθος των μηδενικών τιμών δυσκόλεψε τα μοντέλα να εκτιμήσουν σωστά τις εξάρσεις του δείκτη κινδύνου.

Πίνακας 6.4: Αποτελέσματα Περίπτωσης 1:200

Classification Results					
Model	ROC AUC	Accuracy	F1-Score	Precision	Recall
 LSTM	0,9470	96.7%	18.9%	10.7%	80.4%
 Random Forest	0,9450	94.1%	12.1%	6.5%	81.0%
 GRU	0,9365	89.2%	6.9%	3.6%	83.9%
XGBoost	0,9164	91.6%	8.2%	4.3%	74.9%

Regression Results			
Model	R2 Score	MAE	RMSE
 GRU	0.8685	3,3824	4,7615
 XGBoost	0.8565	3,1692	4,9742
 LSTM	0.8538	3,8598	5,0203
Linear Reg	0.8256	3,4517	5,4831

Γενικό Συμπέρασμα: Η Επιλογή της Βέλτιστης Αρχιτεκτονικής

Η διεξοδική αξιολόγηση των τεσσάρων σεναρίων ανισορροπίας δεδομένων (1:1, 1:4, 1:10, 1:200) κατέδειξε ότι η επιλογή της κατάλληλης αναλογίας εκπαίδευσης είναι εξίσου κρίσιμη με την επιλογή του αλγορίθμου.

Ενώ η φυσική κατανομή (1:200) ανέδειξε τις αδυναμίες των μοντέλων να εντοπίσουν σπάνια συμβάντα (κατάρρευση F1-Score < 20%), και η απόλυτη εξισορρόπηση (1:1) κρίθηκε μη ρεαλιστική για χρήση πρόβλεψης κινδύνου, η αναλογία 1:4 αναδείχθηκε ως η «χρυσή τομή» του συστήματος.

Η επιλογή της αναλογίας 1:4 στηρίζεται σε τρεις βασικούς πυλώνες:

- **Μέγιστη Επιχειρησιακή Αξιοπιστία:** Πέτυχε τη βέλτιστη ισορροπία μεταξύ ευαισθησίας (Recall 83.0%) και ακρίβειας (Precision 76.9%) με τον αλγόριθμο **Random Forest**, εξασφαλίζοντας ότι το σύστημα εντοπίζει τις πυρκαγιές χωρίς να γεμίζει με ψευδείς συναγερμούς.
- **Ακρίβεια στην Εκτίμηση Κινδύνου:** Στην παλινδρόμηση, ο αλγόριθμος **XGBoost** κατάφερε να ελαχιστοποιήσει το σφάλμα πρόβλεψης ($R^2 > 0.92$), προσφέροντας μια ρεαλιστική και ακριβή εικόνα της έντασης του δείκτη FWI.
- **Ανθεκτικότητα:** Σε αντίθεση με την αναλογία 1:10, όπου η απόδοση άρχισε να πέφτει, η περίπτωση του 1:4 απέδειξε ότι διαθέτει την απαραίτητη πυκνότητα θετικών δειγμάτων για να εκπαιδεύσει στιβαρά μοντέλα, ικανά να γενικεύσουν σωστά σε άγνωστα δεδομένα.

Συνεπώς, το Σύστημα κάνει χρήση του Random Forest για Ταξινόμηση και XGBoost για Παλινδρόμηση υπό την αρχιτεκτονική δειγματοληψίας 1:4, αποτελώντας την πλέον ολοκληρωμένη και εφαρμόσιμη λύση για την έγκαιρη και έγκυρη πρόβλεψη δασικών πυρκαγιών.

6.2 Συμπεράσματα

1. Η Υπεροχή της Υβριδικής Προσέγγισης (The Hybrid Advantage): Ένα από τα σημαντικότερα πορίσματα της παρούσας διπλωματικής εργασίας, είναι ότι η βέλτιστη στρατηγική πρόγνωσης δεν προκύπτει από την αποκλειστική χρήση μίας μεθόδου, αλλά από τη αλληλουχία διαφορετικών μοντέλων. Η ανάπτυξη της **Υβριδικής Λογικής (Hybrid Logic)**, η οποία συνδυάζει τη προβλεπτική ικανότητα του DL και ML με τη βεβαιότητα των φυσικών εξισώσεων του συστήματος FWI, αποδείχθηκε καθοριστική. Μέσω του μηχανισμού ασφαλείας "**Override Protocol**", το σύστημα πέτυχε να εξαλείψει τα ψευδώς αρνητικά αποτελέσματα (False Negatives) σε συνθήκες ακραίου κινδύνου, γεφυρώνοντας το χάσμα μεταξύ της στατιστικής ακρίβειας και της επιχειρησιακής ασφάλειας.

2. Η Σημασία της Χρονικής Συνέχειας (Temporal Dependency): Η πειραματική διαδικασία ανέδειξε ένα ενδιαφέρον εύρημα: αν και το **Random Forest** σημείωσε τις υψηλότερες συνολικές επιδόσεις (Accuracy 91.6%, F1 79.8%), επιβεβαιώνοντας την υπεροχή του σε δομημένα δεδομένα, η ανάλυση μέσω μοντέλων Βαθιάς Μάθησης (LSTM) προσέφερε κρίσιμα ποιοτικά συμπεράσματα για τη φύση του προβλήματος. Συγκεκριμένα, η χρήση χρονικού παραθύρου εισόδου 10 ημερών (Lookback Window) κατέδειξε ότι η πρόβλεψη πυρκαγιάς δεν είναι μια απλή στατική ταξινόμηση, αλλά ένα δυναμικό πρόβλημα χρονοσειρών.

Το παράδειγμα στο **Μάτι Αττικής**, όπου αναλύθηκε η συσσώρευση ξηρασίας σε βάθος χρόνου, επιβεβαιώνει ότι η ενσωμάτωση ιστορικών δεδομένων (είτε μέσω της αρχιτεκτονικής μνήμης των LSTM είτε μέσω των χρονικών χαρακτηριστικών που τροφοδότησαν το Random Forest) είναι απαραίτητη για την ορθή εκτίμηση κινδύνου, αποτρέποντας ψευδείς συναγερμούς μετά από πρόσφατες βροχοπτώσεις.

3. Επιχειρησιακή Ετοιμότητα (Operational Feasibility) Η υλοποίηση της διαδικτυακής εφαρμογής προσομοίωσης (**AI Wildfire Simulator**) κατέδειξε ότι η μετάβαση από το εργαστηριακό περιβάλλον στην πραγματική χρήση είναι εφικτή. Η δυνατότητα του συστήματος να λειτουργεί σε πραγματικό χρόνο (Live Forecast) αντλώντας δεδομένα από API, αλλά και να εκτελεί σενάρια ιστορικής επαλήθευσης (Validation Mode), το καθιστά ένα ισχυρό εργαλείο υποστήριξης λήψης αποφάσεων. Η ενσωμάτωση δυναμικών δεικτών (sliders) για ανάλυση σεναρίων "What-If" προσφέρει στους χειριστές τη δυνατότητα προληπτικού σχεδιασμού έναντι ακραίων καιρικών φαινομένων.

6.3 Περιορισμοί της Έρευνας

Παρά τα ενθαρρυντικά αποτελέσματα και την καινοτομία της Υβριδικής Προσέγγισης, η παρούσα διπλωματική εργασία υπόκειται σε συγκεκριμένους περιορισμούς που πρέπει να ληφθούν υπόψη κατά την ερμηνεία των αποτελεσμάτων:

1. Χωρική Ανάλυση και Μικροκλίμα: Ένας βασικός περιορισμός αφορά τη χωρική ανάλυση των μετεωρολογικών δεδομένων. Τα δεδομένα που αντλούνται από το API (Open-Meteo) βασίζονται σε μοντέλα πλέγματος (grid models) που καλύπτουν ευρύτερες περιοχές (ανά 20km). Ως εκ τούτου, το σύστημα ενδέχεται να μην μπορεί να αποτυπώσει με ακρίβεια το **τοπικό μικροκλίμα**, το οποίο παίζει καθοριστικό ρόλο στην εξέλιξη μιας πυρκαγιάς. Για παράδειγμα, οι συνθήκες υγρασίας εντός μιας βαθιάς χαράδρας μπορεί να διαφέρουν σημαντικά από αυτές σε μια εκτεθειμένη σε ένα μεγάλο οροπέδιο, παρόλο που και τα δύο σημεία ανήκουν στο ίδιο κελί πρόβλεψης.

2. Εξάρτηση από την Ακρίβεια της Πρόγνωσης: Η επιχειρησιακή λειτουργία του συστήματος (Live Mode) βασίζεται σε μετεωρολογικές **προβλέψεις** και όχι σε πραγματικές μετρήσεις εκείνης της στιγμής.

Οποιαδήποτε απόκλιση ή σφάλμα της μετεωρολογικής πρόγνωσης (π.χ. υποτίμηση της ταχύτητας του ανέμου από το μετεωρολογικό μοντέλο) μεταφέρεται αναπόφευκτα και στο μοντέλο πρόβλεψης κινδύνου πυρκαγιάς.

3. Ο Ανθρώπινος Παράγοντας: Το σύστημα έχει σχεδιαστεί για να προβλέπει την **επικινδυνότητα** (Risk/Danger), δηλαδή την προδιάθεση του **περιβάλλοντος** για εκδήλωση ή την ανάφλεξη και εξάπλωση. Το μοντέλο δεν μπορεί να προβλέψει τον ανθρώπινο παράγοντα, όπως τυχαία ατυχήματα (π.χ. σπινθήρες από καλώδια ΔΕΗ) ή κακόβουλες ενέργειες (εμπρησμός). Επομένως, μια περιοχή μπορεί να χαρακτηριστεί ως "**Extreme Risk**", αλλά να μην εκδηλωθεί ποτέ πυρκαγιά αν δεν υπάρξει πηγή ανάφλεξης.

4. Απουσία Στατικών Γεωχωρικών Χαρακτηριστικών: Στην παρούσα φάση, το μοντέλο εκπαιδεύτηκε αποκλειστικά με δυναμικά μετεωρολογικά δεδομένα και δείκτες FWI. Δεν έχουν συμπεριληφθεί στατικά χαρακτηριστικά, όπως η **μορφολογία του εδάφους** (κλίση/slope, προσανατολισμός/aspect) και ο **τύπος βλάστησης**. Η έλλειψη αυτών των παραμέτρων σημαίνει ότι το μοντέλο αντιμετωπίζει με τον ίδιο τρόπο μια επίπεδη αγροτική έκταση και μια απότομη δασική πλαγιά, εφόσον έχουν τις ίδιες καιρικές συνθήκες, παρόλο που η συμπεριφορά της φωτιάς θα ήταν διαφορετική.

6.4 Μελλοντικές Επεκτάσεις

1. Ενσωμάτωση Χωρικών Δεδομένων & Δορυφορικής Εικόνας Μια κρίσιμη επέκταση του συστήματος αφορά την εισαγωγή στατικών χαρακτηριστικών που σχετίζονται με τη μορφολογία του εδάφους και τη βλάστηση. Η ενσωμάτωση δεικτών όπως ο **NDVI (Normalized Difference Vegetation Index)** από δορυφορικά δεδομένα (Sentinel-2) και ψηφιακών μοντέλων εδάφους (DEM) θα επέτρεπε στο μοντέλο να διαφοροποιεί τον κίνδυνο όχι μόνο βάσει καιρού, αλλά και βάσει της καύσιμης ύλης. **Για παράδειγμα**, η ίδια ξηρασία σε πευκοδάσος εγκυμονεί μεγαλύτερο κίνδυνο απ' ό,τι σε περιοχή με χαμηλή βλάστηση.

2. Χωρική Χαρτογράφηση Υψηλής Ευκρίνειας: Στην παρούσα υλοποίηση το σύστημα λειτουργεί με ανάλυση 20x20 km. Η επιλογή αυτή δεν έγινε τυχαία, αλλά βασίστηκε στην πραγματική ανάλυση των μετεωρολογικών δεδομένων που μας έδινε το API (δεδομένα ανά 20-25km). Έτσι το επόμενο βήμα είναι η αύξηση της ανάλυσης σε 1x1 km. Στόχος αυτού ο εντοπισμός κινδύνου ακόμα και σε μικρές δασικές εκτάσεις, που τώρα λόγω της γενίκευσης του υπάρχοντος πλέγματος ενδεχομένως να χάνονται. Μια τέτοια αναβάθμιση θα είναι απαραίτητη για το δύσκολο ανάγλυφο της Ελλάδας. Το μειονέκτημα στην προσέγγιση αυτή είναι η μεγάλη υπολογιστική ισχύ.

3. Δυναμική Παρακολούθηση Εξέλιξης Κινδύνου: Στην παρούσα φάση το σύστημα στοχεύει στην ημερήσια πρόβλεψη. Μια εξέλιξη του συστήματος αυτού θα ήταν η παρακολούθηση της μεταβολής του κινδύνου. Ποιο συγκεκριμένα, η παραγωγή χαρτών επικινδυνότητας ανά ώρα ή ανά εξάωρο, με στόχο την κατανόηση της μετατόπισης κινδύνου ενός μετώπου κατά την διάρκεια της ημέρας, ανάλογα με την μεταβολή των καιρικών δεικτών.

4. Διασύνδεση με IoT Αισθητήρες Πραγματικού Χρόνου: Τέλος, η ακρίβεια του συστήματος θα μπορούσε να μεγιστοποιηθεί μέσω της απευθείας διασύνδεσης με δίκτυα επίγειων αισθητήρων IoT εντός των δασικών εκτάσεων. Η αντικατάσταση των μετεωρολογικών προβλέψεων (API Forecasts) με πραγματικές μετρήσεις μικρο-κλίματος (υγρασία φυλλάδας, θερμοκρασία εδάφους) θα μείωνε δραστικά το σφάλμα εισόδου και θα επέτρεπε τον εντοπισμό πυρκαγιών στα πρώτα λεπτά της εκδήλωσής τους (Early Warning System).

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] tovima.com, “Greece Sees Surge in Once-Rare Megafires,” *tovima.com*, Sep. 11, 2025. [Online]. Available: <https://www.tovima.com/climate/greece-sees-surge-in-one-rare-megafires/> (accessed Feb. 03, 2026).
- [2] “Weather conditions leading to deadly wildfires in Türkiye, Cyprus and Greece made 10 times more likely due to climate change – World Weather Attribution,” *Worldweatherattribution.org*, 2025. [Online]. Available: <https://www.worldweatherattribution.org/weather-conditions-leading-to-deadly-wildfires-in-turkiye-cyprus-and-greece-made-10-times-more-likely-due-to-climate-change/>
- [3] OECD, “Taming wildfires in the context of climate change: The case of Greece,” *OECD environment policy papers*, Jun. 2024.
- [4] “Guidelines on the use of tools, science and best practices for fire analysis.” Accessed: Feb. 04, 2026. [Online]. Available: https://civil-protection-knowledge-network.europa.eu/system/files/2023-07/AFAN%20D2.5%20GUIDELINES%20ON%20THE%20USE%20OF%20TOOLS_0.pdf
- [5] E. Kostopoulou and G. Stavridis, “Wildfire Risk Assessment Using the Fire Weather Index (FWI) in Greece,” *Climate*, vol. 13, no. 6, pp. 109–109, May 2025.
- [6] N. Omar, A. Al-zebari, and A. Sengur, “Deep Learning Approach to Predict Forest Fires Using Meteorological Measurements,” *IEEE Xplore*, Dec. 01, 2021.
- [7] M. Hojati and C. Robertson, “Integrating cellular automata and discrete global grid systems: a case study into wildfire modelling,” *AGILE: GIScience Series*, vol. 1, pp. 1–23, Jul. 2020.
- [8] V. Wagner, T. L. Pickett, and C. F. Service, *Equations and FORTRAN Program for the Canadian Forest Fire Weather Index System*. 1985.
- [9] “Fire Weather Indices Wiki | Fine fuel moisture code,” *Wikifire.wsl.ch*, 2025. [Online]. Available: <https://wikifire.wsl.ch/tiki-index91f7.html?page=Fine+fuel+moisture+code>
- [10] “Fire Weather Indices Wiki | Duff moisture code,” *Wikifire.wsl.ch*, 2025. [Online]. Available: <https://wikifire.wsl.ch/tiki-index9436.html?page=Duff+moisture+code>
- [11] “Fire Weather Indices Wiki | Drought code,” *Wikifire.wsl.ch*, 2026. [Online]. Available: <https://wikifire.wsl.ch/tiki-indexd5c6.html?page=Drought+code> (accessed Feb. 03, 2026).
- [12] “Fire Weather Indices Wiki | Initial spread index,” *Wikifire.wsl.ch*, 2026. [Online]. Available: <https://wikifire.wsl.ch/tiki-index4de6.html?page=Initial+spread+index> (accessed Feb. 03, 2026).
- [13] “Fire Weather Indices Wiki | Buildup index,” *Wikifire.wsl.ch*, 2026. [Online]. Available: <https://wikifire.wsl.ch/tiki-index8720.html?page=Buildup+index> (accessed Feb. 03, 2026).
- [14] “Fire Weather Indices Wiki | Fire weather index,” *Wikifire.wsl.ch*, 2026. [Online]. Available: <https://wikifire.wsl.ch/tiki-index259b.html?page=Fire+weather+index>
- [15] L. Van Vliet, J. Fyke, S. Nakoneczny, T. Q. Murdock, and P. Jafarpur, “Developing user-informed fire weather projections for Canada,” *Climate Services*, vol. 35, p. 100505, Aug. 2024.
- [16] Arunkumar Gopu, A. Ramakrishnan, G. Balasubramanian, and Kuna Srinidhi, “A Comparative Study on Forest Fire Prediction using ARIMA, SARIMA, LSTM, and GRU Methods,” pp. 1–5, Apr. 2023.

- [17] R. Patil, J. Pawar, K. Shah, D. Shetty, Aparna Ajith, and S. Jadhav, "Machine Learning based Forest Fire Prediction: A Comparative Approach," *International Research Journal of Multidisciplinary Technovation*, pp. 32–39, Jan. 2024.
- [18] P. Rakshit *et al.*, "Prediction of Forest Fire Using Machine Learning Algorithms: The Search for the Better Algorithm," *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, Nov. 2021.
- [19] S. Kondylatos *et al.*, "Wildfire Danger Prediction and Understanding With Deep Learning," *Geophysical Research Letters*, vol. 49, no. 17, Sep. 2022.
- [20] P. Moral *et al.*, "Forest Fire Forecasting Leveraging Modis Satellite Fire Data Using Machine Learning For Jharkhand State, India," *2024 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*, pp. 1–4, Dec. 2024.
- [21] P. K. Srivastava, S. Mukherjee, M. Gupta, and T. Islam, *Remote Sensing Applications in Environmental Research*. 2014. doi: <https://doi.org/10.1007/978-3-319-05906-8>.
- [22] V. Bharathi and C. PeddaReddy, "Comparison of Forest Fire Prediction System using Machine Learning Algorithms," *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 1539–1542, May 2023.
- [23] R. Shanmuga Priya and K. Vani, "Climate Change Forecast for Forest Fire Risk Prediction using Deep Learning," Mar. 2023, Accessed: Nov. 14, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10112983>
- [24] X. Lin, Z. Li, W. Chen, X. Sun, and D. Gao, "Forest Fire Prediction Based on Long- and Short-Term Time-Series Network," *Forests*, vol. 14, no. 4, p. 778, Apr. 2023, Accessed: May 31, 2023. [Online]. Available: <https://www.mdpi.com/1999-4907/14/4/778>
- [25] Evizal Abdul Kadir, Hsiang Tsung Kung, A. Almansour, H. Irie, Sri Listia Rosa, and S. Mohd, "Wildfire Hotspots Forecasting and Mapping for Environmental Monitoring Based on the Long Short-Term Memory Networks Deep Learning Algorithm," *Environments*, vol. 10, no. 7, pp. 124–124, Jul. 2023.
- [26] E. A. Kadir, H. Irie, S. L. Rosa, W. A. W. A. Bakar, M. F. Evizal, and J. Hassan, "Data Analysis and Forecasting of Forest Fires in the Indonesia and Australia Region," *2024 17th International Conference on Sensing Technology (ICST)*, pp. 1–6, Dec. 2024.
- [27] I. Karasante, L. Alonso, I. Prapas, A. Ahuja, N. Carvalhais, and I. Papoutsis, "SeasFire cube - a multivariate dataset for global wildfire modeling," *Scientific Data*, vol. 12, no. 1, Mar. 2025.
- [28] V. Kotroni *et al.*, "DISARM Early Warning System for Wildfires in the Eastern Mediterranean," *Sustainability*, vol. 12, no. 16, p. 6670, Aug. 2020.
- [29] R. Mekala, S. Srinath, S. Gokul, E. Balavigneshwar, and R. Muralidharan, "Forest Fire Probability Prediction based on Humidity and Temperature," pp. 1–5, Mar. 2023.
- [30] European Commission. (2026). *EFFIS/WILDFIRE Database: MODIS Burnt Areas Metadata*. European Forest Fire Information System. Available at: https://ec.europa.eu/info/legal-notice_en
- [31] J. San-Miguel-Ayanz *et al.*, "Comprehensive Monitoring of Wildfires in Europe: The European Forest Fire Information System (EFFIS)," *Approaches to Managing Disaster - Assessing Hazards, Emergencies and Disaster Impacts*, Mar. 2012.

- [32] NASA, “MODIS Web,” *modis.gsfc.nasa.gov*, [Online]. Available: <https://modis.gsfc.nasa.gov/>
- [33] J. D. Blower and A. Clegg, “Fast regridding of large, complex geospatial datasets,” Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications - COM.Geo '11, pp. 1–6, Mar. 2011.
- [34] H. Veregin, “How Big is a Degree?,” State Cartographer’s Office, Jan. 21, 2022. [Online]. Available: <https://www.sco.wisc.edu/2022/01/21/how-big-is-a-degree/>
- [35] *Nasa.gov*, 2026. [Online]. Available: https://firms.modaps.eosdis.nasa.gov/data/active_fire/modis-c6.1/csv/MODIS_C6_1_Europe_7d.csv (accessed Jan. 16, 2026).
- [36] Open-Meteo, “Historical Weather API | Open-Meteo.com,” *open-meteo.com*, 2025. [Online]. Available: <https://open-meteo.com/en/docs/historical-weather-api>
- [37] apipark, “How to Circumvent API Rate Limiting: Best Practices,” *APIPark*, Dec. 27, 2025. [Online]. Available: <https://apipark.com/techblog/en/how-to-circumvent-api-rate-limiting-best-practices-5/> (accessed Jan. 18, 2026).
- [38] M. McElhinny, J. F. Beckers, C. Hanes, M. Flannigan, and P. Jain, “A high-resolution reanalysis of global fire weather from 1979 to 2018 – overwintering the Drought Code,” *Earth System Science Data*, vol. 12, no. 3, pp. 1823–1833, Aug. 2020.
- [39] J. Brownlee, “Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras,” *Machine Learning Mastery*, Apr. 26, 2019. <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/> (accessed Nov. 24, 2025).
- [40] J. Shen, W. Wu, and Q. Xu, “Accurate Prediction of Temperature Indicators in Eastern China Using a Multi-Scale CNN-LSTM-Attention model,” *Applied and Computational Engineering*, vol. 120, no. 1, pp. 164–170, Jan. 2025.
- [41] I Nyoman Switrayana, R. Hammad, P. Irfan, T. T. Sujaka, and M. H. Nasri, “Comparative Analysis of Stock Price Prediction Using Deep Learning with Data Scaling Method,” *JTIM Jurnal Teknologi Informasi dan Multimedia*, vol. 7, no. 1, pp. 78–90, Jan. 2025.
- [42] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, “Permutation importance: a corrected feature importance measure,” *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, Apr. 2010.
- [43] Assaf Shmuel, T. Lazebnik, E. Heifetz, O. Glickman, and C. Price, “Fire weather indices tailored to regional patterns outperform global models,” *npj natural hazards*, vol. 2, no. 1, Jul. 2025, Accessed: Oct. 17, 2025. [Online]. Available: <https://www.nature.com/articles/s44304-025-00126-y>
- [44] M. Ramesh *et al.*, “Assessing WildfireGPT: a comparative analysis of AI models for quantitative wildfire spread prediction,” *Natural Hazards*, vol. 121, no. 11, pp. 13117–13130, May 2025.
- [45] D. A. Wood, “Prediction and data mining of burned areas of forest fires: Optimized data matching and mining algorithm provides valuable insight,” *Artificial Intelligence in Agriculture*, vol. 5, pp. 24–42, 2021.
- [46] Streamlit, “Streamlit Docs,” *docs.streamlit.io*, 2024. [Online]. Available: <https://docs.streamlit.io/>
- [47] A. Merced, “Deep Dive into Data Apps with Streamlit,” *DEV Community*, Sep. 22, 2024. [Online]. Available: <https://dev.to/alexmercedcoder/deep-dive-into-data-apps-with-streamlit-3e40>

[48] Συνεισφέροντας στα εγχειρίδια Wikimedia, “οικισμός της Ελλάδας,” *Wikipedia.org*, Apr. 21, 2023. [Online]. Available: https://el.wikipedia.org/wiki/Πεύκη_Τρικάλων (accessed Feb. 02, 2026).

[49] “EFFIS - Fire Danger Forecast,” *forest-fire.emergency.copernicus.eu*. [Online]. Available: <https://forest-fire.emergency.copernicus.eu/about-effis/technical-background/fire-danger-forecast> (accessed Feb. 02, 2026).