



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΗΣ ΙΣΧΥΟΣ

**Υπερβραχυπρόθεσμη Πρόβλεψη Φωτοβολταϊκής
Παραγωγής με Χρήση Δεδομένων Εικόνων του Ουρανού**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αντόνιο Αντωνιάδης

Επίβλεψη : Παύλος Σ. Γεωργιάκης, Καθηγητής Ε.Μ.Π.

Επόπτευση: Μάρκος Κουσούναδης Κνούσεν, Υποψήφιος Διδάκτορας ΕΜΠ

Αθήνα, Ιούνιος 2024



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΗΣ ΙΣΧΥΟΣ

**Υπερβραχυπρόθεσμη Πρόβλεψη Φωτοβολταϊκής
Παραγωγής με Χρήση Δεδομένων Εικόνων του Ουρανού**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αντόνιο Αντωνιάδης

Επίβλεψη: Παύλος Σ. Γεωργιλιάκης, Καθηγητής Ε.Μ.Π.

Επόπτευση: Μάρκος Κουσούναδης Κνούσεν, Υποψήφιος Διδάκτορας ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 21^η Ιουνίου 2024

.....
Παύλος Γεωργιλιάκης
Καθηγητής Ε.Μ.Π.

.....
Αντόνιος Παπαβασιλείου
Επικ. Καθηγητής Ε.Μ.Π.

.....
Άρης-Ευάγγελος Δημέας
Επικ. Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2024

.....
Αντόνιο Αντωνιάδης
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών
Ε.Μ.Π.

Copyright © Αντόνιο Αντωνιάδης, 2024.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η υπερβραχυπρόθεσμη πρόβλεψη φωτοβολταϊκής παραγωγής, μέσω της ανάπτυξης ενός υβριδικού μοντέλου μηχανικής μάθησης. Το προτεινόμενο μοντέλο πρόβλεψης αποτελείται από την τεχνική μηχανικής μάθησης Random Forest και τον εξελικτικό αλγόριθμο Antlion Optimizer, με χρονικό ορίζοντα πρόβλεψης τα δεκαπέντε λεπτά.

Η εκπαίδευση και η αξιολόγηση του υβριδικού μοντέλου πραγματοποιείται με δύο σύνολα δεδομένων: ένα με αποκλειστικά αριθμητικά δεδομένα και ένα που περιλαμβάνει αριθμητικά δεδομένα και εικόνες του ουρανού. Στόχος του αναπτυσσόμενου μοντέλου είναι η πρόσεγγιση της πραγματικά παραγόμενης φωτοβολταϊκής παραγωγής, με τη χρήση μετρούμενων μετεωρολογικών μεγεθών.

Η απόδοση του προτεινόμενου υβριδικού μοντέλου συγκρίνεται με την απόδοση δύο άλλων μοντέλων μέσω υπολογισμού κατάλληλων μετρητικών σφάλματος. Η ανάπτυξη του κώδικα για τη δημιουργία των μοντέλων της παρούσας διπλωματικής εργασίας υλοποιήθηκε στη γλώσσα προγραμματισμού Python.

Η καινοτομία της παρούσας διπλωματικής εργασίας εγκείται στην ενδελεχή αξιολόγηση των δεδομένων ανά καιρική συνθήκη, η οποία αποσαφηνίζει τα ποιοτικά χαρακτηριστικά της διαδικασίας πρόβλεψης. Επιπλέον, το μοντέλο εκπαιδεύτηκε χρησιμοποιώντας επίγειες εικόνες του ουρανού, επιτρέποντας έτσι την ακριβή προσαρμογή του μοντέλου στις διαφορετικές συνθήκες νεφοκάλυψης.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Random Forest, Antlion Optimizer, υπερβραχυπρόθεσμη πρόβλεψη, φωτοβολταϊκή παραγωγή, επίγειες εικόνες ουρανού

ABSTRACT

The aim of this Diploma Thesis is the ultra-short-term forecasting of photovoltaic production through the development of a hybrid machine learning model. The proposed forecasting model consists of the machine learning technique Random Forest and the evolutionary algorithm Antlion Optimizer, with a forecasting horizon of fifteen minutes.

The training and evaluation of the hybrid model are conducted with two datasets: one with exclusively numerical data and another that includes numerical data and sky images. The aim of the developed model is to approach the actual produced photovoltaic output using measured meteorological parameters.

The performance of the proposed hybrid model is compared with the performance of two other models by calculating appropriate error metrics. The code for creating the models in this Thesis was implemented in the Python programming language.

The innovation of this Thesis lies in the thorough evaluation of the data under different weather conditions, which clarifies the qualitative characteristics of the forecasting process. Additionally, the model was trained using ground-based sky images, allowing for accurate adaptation of the model to different cloud cover conditions.

KEY WORDS

Random Forest, Antlion Optimizer, ultra-short-term forecasting, photovoltaic production, ground-based sky images

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2023–2024 υπό την επίβλεψη του κ. Παύλου Γεωργιάκη, Καθηγητή της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Ε.Μ.Π. στον οποίο οφείλω ιδιαίτερες ευχαριστίες για την ανάθεσή της, δίνοντάς μου την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον και δημιουργικό θέμα. Επίσης, θα ήθελα να ευχαριστήσω θερμά τον υποψήφιο διδάκτορα Μάρκο Κουσούνάδη - Κνούσεν για την υπομονή και την πολύτιμη βοήθεια και καθοδήγηση που μου παρείχε σε όλη τη διάρκεια εκπόνησης της εργασίας, καθώς και για τον πολύτιμο χρόνο που μου αφιέρωσε.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου Φώτο και Άστριτ για τη στήριξή τους σε όλη τη διάρκεια των σπουδών μου στο Ε.Μ.Π., καθώς και την Ήλια για τη συμπαράσταση και την αγάπη της σε όλες τις δυσκολίες που αντιμετώπισα.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	1
1.1 Αντικείμενο της διπλωματικής εργασίας	1
1.2 Βιβλιογραφική ανασκόπηση	3
1.3 Δομή διπλωματικής εργασίας	3
ΚΕΦΑΛΑΙΟ 2: ΠΡΟΒΛΕΨΗ ΦΩΤΟΒΟΛΤΑΪΚΗΣ ΠΑΡΑΓΩΓΗΣ	5
2.1 Γενικά	5
2.2 Τρόπος λειτουργίας φωτοβολταϊκών συστημάτων	6
2.3 Παράγοντες που επηρεάζουν τη φωτοβολταϊκή παραγωγή	7
2.4 Σημαντικότητα πρόβλεψης φωτοβολταϊκής παραγωγής	9
2.5 Μεθόδοι πρόβλεψης φωτοβολταϊκής παραγωγής	11
2.5.1 Φυσικές μεθόδοι	11
2.5.2 Στατιστικές μεθόδοι	11
2.5.3 Χρήση μηχανικής μάθησης	12
2.6 Χρονικός ορίζοντας μοντέλων πρόβλεψης	12
2.6.1 Βραχυπρόθεσμη πρόβλεψη	13
2.6.2 Μεσοπρόθεσμη πρόβλεψη	14
2.6.3 Μακροπρόθεσμη πρόβλεψη	14
2.6.4 Υπερβραχυπρόθεσμη πρόβλεψη	14
ΚΕΦΑΛΑΙΟ 3: ΜΟΝΤΕΛΟ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ RANDOM FOREST	17
3.1 Γενικά	17
3.2 Εισαγωγή στα Random Forests	19
3.2.1 Δέντρα Απόφασης	20
3.2.2 Συγκλιση των Random Forest	21
3.2.3 Τεχνική Bagging	23
3.2.4 Υπερπαράμετροι μοντέλου	24
3.3 Αντικειμενικές συναρτήσεις	26

3.4	Βελτιστοποίηση υπερπαραμέτρων Random Forest	27
3.5	Αλγόριθμος βελτιστοποίησης Ant-Lion Optimizer	27
ΚΕΦΑΛΑΙΟ 4: ΑΝΑΛΥΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΦΩΤΟΒΟΛΤΑΪΚΩΝ ΔΕΔΟΜΕΝΩΝ		33
4.1	Δεδομένα φωτοβολταϊκής παραγωγής	33
4.2	Αριθμητικά δεδομένα φωτοβολταϊκού πάρκου στην Ελλάδα	34
4.3	Επεξεργασία δεδομένων	35
4.3.1	Χρονική προετοιμασία δεδομένων	35
4.3.2	Κανονικοποίηση δεδομένων	36
4.4	Δεδομένα επίγειων εικόνων του ουρανού	36
4.5	Κλάσεις νεφώσεων	38
ΚΕΦΑΛΑΙΟ 5: ΜΟΝΤΕΛΟ ΥΠΕΡΒΡΑΧΥΠΡΟΘΕΣΜΗΣ ΠΡΟΒΛΕΨΗΣ ΦΩΤΟΒΟΛΤΑΪΚΗΣ ΠΑΡΑΓΩΓΗΣ		41
5.1	Μεθοδολογία	41
5.2	Αρχικοποίηση υπερπαραμέτρων μοντέλου	41
5.3	Υπολογιζόμενα σφάλματα και μετρητικά αξιολόγησης μοντέλου	43
5.4	Χωρισμός δεδομένων σε σύνολα εκπαίδευσης και ελέγχου	43
5.5	Επιλογή μεταβλητών εισόδου	44
5.5.1	Μονοδιάστατης είσοδος	44
5.5.2	Πολυδιάστατη είσοδος	45
5.6	Αλγόριθμος Ant-Lion Optimizer για βελτιστοποίηση υπερπαραμέτρων μοντέλου	47
5.6.1	Υλοποίηση του αλγορίθμου	48
ΚΕΦΑΛΑΙΟ 6: ΑΠΟΤΕΛΕΣΜΑΤΑ ΜΟΝΤΕΛΟΥ		51
6.1	Εισαγωγή	51
6.2	Αποτελέσματα πρώτης περίπτωσης προς μελέτη	51
6.2.1	Εύρεση αποδοτικότερων υπερπαραμέτρων	52
6.2.2	Συγκεντρωτικά αποτελέσματα	53

6.2.3	Ανάλυση ανά καιρική συνθήκη	55
6.3	Αποτελέσματα δεύτερης περίπτωσης προς μελέτη	61
6.3.1	Εύρεση αποδοτικότερων υπερπαραμέτρων	62
6.3.2	Συγκεντρωτικά αποτελέσματα	64
ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ		71
7.1	Σύνοψη εργασίας και συμπεράσματα	71
7.2	Προτάσεις μελλοντικής επέκτασης	72
ΚΕΦΑΛΑΙΟ 8: ΒΙΒΛΙΟΓΡΑΦΙΑ		75

ΕΙΣΑΓΩΓΗ

1.1 ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η πρόβλεψη της φωτοβολταϊκής (Φ/Β) παραγωγής μέσω της ανάπτυξης ενός υβριδικού μοντέλου μηχανικής μάθησης, το οποίο βασίζει τη λειτουργία του στο μοντέλο Random Forest (RF). Η άμεση απαίτηση για αντιμετώπιση του φαινομένου της κλιματικής αλλαγής, η ανάγκη για μείωση των εκπομπών διοξειδίου του άνθρακα, καθώς και η εξάντληση των αποθεμάτων ορυκτών καυσίμων, έχουν οδηγήσει στην αυξημένη ενσωμάτωση των Ανανεώσιμων Πηγών Ενέργειας (ΑΠΕ) στα Συστήματα Ηλεκτρικής Ενέργειας (ΣΗΕ) για την κάλυψη των ενεργειακών αναγκών[1]. Οι Φ/Β σταθμοί, συγκεκριμένα, αποτελούν βασική πηγή ανανεώσιμης ενέργειας τόσο στην Ελλάδα όσο και παγκοσμίως, χάρη στην υψηλή παραγωγική τους δυνατότητα, την εύκολη και γρήγορη εγκατάστασή τους, το χαμηλό κόστος συντήρησης, τη μακροχρόνια αξιοπιστία και την οικονομική τους αποδοτικότητα[2]. Ωστόσο, η ένταξη της Φ/Β ενέργειας στα ΣΗΕ συνοδεύεται από αυξημένη αβεβαιότητα και μεταβλητότητα, ιδιότητες που χαρακτηρίζουν τις περισσότερες ΑΠΕ. Η παραγωγή ηλεκτρικής ενέργειας από Φ/Β σταθμούς καθορίζεται από διάφορες μετεωρολογικές συνθήκες, με πιο σημαντικές την ηλιακή ακτινοβολία και την θερμοκρασία του περιβάλλοντος[3]. Οι απρόβλεπτες μεταβολές στις καιρικές συνθήκες, όπως η αυξημένη συννεφιά, οι ξαφνικές βροχοπτώσεις, και οι αλλαγές στην ένταση της ηλιακής ακτινοβολίας όσο και η απόκλιση από τις ιδανικές συνθήκες λειτουργίας των Φ/Β σταθμών καθιστούν την πρόβλεψη της Φ/Β παραγωγής μία δύσκολη αλλά κρίσιμη διαδικασία. Η ακριβής πρόβλεψη της παραγωγής είναι καθοριστικής σημασίας για πολλούς φορείς των ΣΗΕ, καθώς όσο μεγαλύτερο είναι το ποσοστό της Φ/Β παραγωγής, τόσο μεγαλύτερος είναι ο αντίκτυπος στη σταθερότητα των δικτύων μεταφοράς και διανομής[4]. Παράλληλα, είναι ζωτικής σημασίας για τον ενεργειακό προγραμματισμό, καθώς επιτρέπει την καλύτερη κατανομή και διαχείριση των ενεργειακών πόρων, μειώνοντας την εξάρτηση από τις συμβατικές μονάδες παραγωγής[5]. Οι σημαντικές αποκλίσεις μεταξύ προβλεπόμενης και πραγματικής παραγωγής μπορούν να οδηγήσουν σε απώλειες της παραγόμενης ηλεκτρικής ενέργειας λόγω μη αξιοποίησης της περίσσειας Φ/Β παραγωγής, να προκαλέσουν αστάθεια στο ηλεκτρικό δίκτυο λόγω διακυμάνσεων της τάσης και της συχνότητας και να αυξήσουν το συνολικό κόστος παραγωγής της ηλεκτρικής ενέργειας[6].

Στην προσπάθεια να περιοριστούν τα μειονεκτήματα που απορρέουν από τη τυχαιότητα της Φ/Β παραγωγής, έχουν αναπτυχθεί πολυάριθμες μέθοδοι πρόβλεψης που έχουν ως στόχο την προσέγγιση της πραγματικής Φ/Β παραγωγής. Οι μέθοδοι πρόβλεψης παραγωγής ταξινομούνται στις φυσικές μεθόδους, που αξιοποιούν μαθηματικές εξισώσεις και προσομοιώσεις του φυσικού περιβάλλοντος, στις στατιστικές μεθόδους, οι οποίες εξάγουν σχέσεις από ιστορικά δεδομένα με σκοπό την προσέγγιση της μελλοντικής μεταβολής των μεγεθών, και στα μοντέλα μηχανικής μάθησης, που αξιοποιούν αλγορίθμους για να βρουν

μοτίβα και συσχετίσεις σε μεγάλα σύνολα δεδομένων, ώστε να λαμβάνουν τις καλύτερες αποφάσεις και προβλέψεις[7]. Η παραγωγή προβλέψεων δύναται να προκύψει για διαφορετικούς χρονικούς ορίζοντες. Εκτός από τους μακροπρόθεσμους, μεσοπρόθεσμους και βραχυπρόθεσμους χρονικούς ορίζοντες, τα τελευταία χρόνια έχει δοθεί ιδιαίτερη έμφαση στην υπερβραχυπρόθεσμη πρόβλεψη[8]. Αυτού του είδους η πρόβλεψη αφορά χρονικούς ορίζοντες από μερικά δευτερόλεπτα έως μία ώρα, και είναι ιδιαίτερα σημαντική για τον βέλτιστο χειρισμό και έλεγχο των Φ/Β συστημάτων σε πραγματικό χρόνο. Η υπερβραχυπρόθεσμη πρόβλεψη συνδέεται άμεσα με τη χρήση μοντέλων μηχανικής μάθησης, τα οποία είναι ικανά να αποτυπώνουν πολύπλοκες σχέσεις μεταξύ εισόδου και εξόδου με υψηλή αποδοτικότητα και λεπτομέρεια, χωρίς να απαιτούν υπερβολικά μεγάλους υπολογιστικούς πόρους[9].

Στην παρούσα διπλωματική εργασία αναπτύσσεται ένα υβριδικό μοντέλο μηχανικής μάθησης RF, το οποίο εκπαιδεύεται με ιστορικά δεδομένα παραγωγής και μετεωρολογικών μεταβλητών, με σκοπό την υπερβραχυπρόθεσμη πρόβλεψη της Φ/Β παραγωγής σε χρονικό ορίζοντα 15 λεπτών. Για την βελτιστοποίηση των υπερπαραμέτρων του μοντέλου RF αξιοποιείται ο εξελικτικός αλγόριθμος Antlion Optimizer (ALO) [10], με στόχο την βελτίωση της απόδοσης του μοντέλου, διατηρώντας υψηλή τη δυνατότητα κλιμάκωσης (scalability) και τα επίπεδα ευελιξίας. Η ακρίβεια των παραγόμενων προβλέψεων αξιολογείται με προσεκτικά επιλεγμένα μετρητικά σφάλματος, με σκοπό τη διεξοδική εξέταση της αποτελεσματικότητάς του.

Για την εκπαίδευση και την αξιολόγηση του μοντέλου χρησιμοποιούνται δύο σετ δεδομένων. Το πρώτο σετ περιέχει αποκλειστικά αριθμητικά δεδομένα, ενώ το δεύτερο περιλαμβάνει αριθμητικά δεδομένα και εικόνες του ουρανού. Η προσθήκη των εικόνων του ουρανού προσφέρει στο προτεινόμενο μοντέλο περισσότερες δυνατότητες, επιτρέποντάς του να αναγνωρίζει εγκαίρως ταχείες μεταβολές στην προσπίπτουσα ηλιακή ακτινοβολία λόγω της ύπαρξης νεφώσεων[11]. Προκειμένου να αξιολογηθεί περαιτέρω η ακρίβεια του προτεινόμενου υβριδικού μοντέλου, το πρώτο σύνολο δεδομένων αναλύεται ανά καιρική συνθήκη. Αυτή η ανάλυση βοηθά στην εξαγωγή συμπερασμάτων σχετικά με την απόδοση του μοντέλου, την ύπαρξη ή απουσία περιοδικών φαινομένων, καθώς και τη διερεύνηση τυχόν συσχετίσεων μεταξύ των αποτελεσμάτων ανά καιρική συνθήκη. Στην περίπτωση του δεύτερου συνόλου δεδομένων, αξιοποιείται μία προκαθορισμένη ταξινόμηση των δεδομένων σε επτά κλάσεις νεφώσεων, και η εκπαίδευση του μοντέλου γίνεται ανά κλάση. Η ταξινόμηση αυτή γίνεται προκειμένου να αποκαλυφθούν τα πλεονεκτήματα και τα μειονεκτήματα του μοντέλου υπό διαφορετικές, λιγότερο ή περισσότερο μεταβαλλόμενες, καιρικές συνθήκες[12].

Οι βασικές καινοτομίες της παρούσας διπλωματικής εργασίας είναι οι εξής:

- Χρήση του αλγόριθμου ALO για τη βελτιστοποίηση όλων των βασικών υπερπαραμέτρων του μοντέλου RF.
- Χρήση δεδομένων εικόνων του ουρανού για τον εμπλουτισμό του μοντέλου με πρόσθετες πληροφορίες για τις καιρικές συνθήκες.
- Παραγωγή προβλέψεων με υψηλή χρονική ανάλυση, συγκεκριμένα με χρονική ανάλυση ενός λεπτού, ιδιαίτερα χρήσιμο για εφαρμογές που απαιτούν υπερβραχυπρόθεσμη πρόβλεψη.
- Τα δεδομένα αναλύονται ανά καιρική συνθήκη, εξάγοντας πολύτιμα συμπεράσματα για την απόδοση του μοντέλου κατά τις διάφορες καιρικές συνθήκες.

1.2 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

Η ακριβής πρόβλεψη της φωτοβολταϊκής παραγωγής, ιδιαίτερα σε υπερβραχυπρόθεσμους χρονικούς ορίζοντες, είναι ζωτικής σημασίας για την αποδοτική λειτουργία των ΣΗΕ και την απρόσκοπτη ένταξη των ΑΠΕ στο ενεργειακό μείγμα. Τα τελευταία χρόνια, τα μοντέλα μηχανικής μάθησης έχουν σημειώσει σημαντική πρόοδο, χάρη στην ικανότητά τους να ανακαλύπτουν περίπλοκα μοτίβα συσχετίσεων μεταξύ των δεδομένων. Σε αυτό το πλαίσιο, παρουσιάζονται ενδεικτικά ορισμένες εργασίες από τη διαθέσιμη βιβλιογραφία που εστιάζουν στην ανάπτυξη μοντέλων μηχανικής μάθησης για την υπερβραχυπρόθεσμη πρόβλεψη της Φ/Β παραγωγής.

Στην εργασία [12], οι D. Anagnostos, T. Schmidt, S. Cavadias, D. Soudris, J. Poortmans, και F. Catthoor παρουσιάζουν μια μέθοδο για λεπτομερή και αποτελεσματική πρόβλεψη της απόδοσης ενέργειας από Φ/Β συστήματα. Χρησιμοποιούνται εικόνες του ουρανού και με τη χρήση λογισμικού απεικόνισης, κάθε εικόνα κατατάσσεται σε μία από επτά κατηγορίες βάσει του κυρίαρχου τύπου νεφών. Για να αξιοποιηθούν τα χαρακτηριστικά κάθε κατηγορίας νεφών ξεχωριστά, αναπτύχθηκε ένα Νευρωνικό Δίκτυο με Εξωτερικές Ανατροφοδοτήσεις (NARX) για κάθε κατηγορία και εκπαιδεύτηκε με τα ταξινομημένα δεδομένα εισόδου της. Οι προβλέψεις παράχθηκαν με χρονικό ορίζοντα 15 λεπτών και χρονική ανάλυση ενός δευτερολέπτου.

Οι Luca Oneto, Federica Laureri, Michela Robba, Federico Delfino, Davide Anguita, στην εργασία [13] ασχολούνται με το πρόβλημα της υπερβραχυπρόθεσμης και της βραχυπρόθεσμης πρόβλεψης της Φ/Β παραγωγής με βάση πραγματικά δεδομένα, που προέρχονται από μικροδίκτυα. Στην εργασία αυτή γίνεται σύγκριση των αποτελεσμάτων με τρεις προηγμένες τεχνικές ανάλυσης δεδομένων, τα *kernelized regularized least squares*, τα *extreme learning machines*, και τα RF.

Οι Sonia Leva, Alfredo Nespoli, Marco Mussetta, και Emanuele Ogliari, στην εργασία [14] παρουσιάζουν μια ολοκληρωμένη μεθοδολογία για την ακριβή πρόβλεψη της παραγωγής ενέργειας από Φ/Β συστήματα σε διάφορους χρονικούς ορίζοντες: από μερικά λεπτά έως και 24 ώρες. Η πρόβλεψη ισχύος για την επόμενη ημέρα γίνεται με την χρήση της μεθόδου υβριδικού τεχνητού νευρωνικού δικτύου φυσικής προσέγγισης (PHANN), ενώ για την περαιτέρω βελτίωση των προβλέψεων για το άμεσο μέλλον, υιοθετείται μια στοχαστική μέθοδος τεχνητών νευρικών δικτύων.

Οι Ibrahim Anwar Ibrahim, M. J. Hossain, Benjamin C. Duck στη εργασία [15] αναπτύσσουν ένα νέο υβριδικό μοντέλο RF σε συνδυασμό με τον αλγόριθμο ALO, για την υπερβραχυπρόθεσμη πρόβλεψη των χαρακτηριστικών καμπυλών I–V των Φ/Β συστημάτων με χρονικό ορίζοντα 15 λεπτών. Για τον σκοπό αυτό, χρησιμοποιήθηκαν μετεωρολογικά δεδομένα με χρονική ανάλυση 15 λεπτών, η οποία ωστόσο είναι υπερβολικά χαμηλή για προβλέψεις σε υπερβραχυπρόθεσμους ορίζοντες. Επίσης, στην εργασία αυτή, δεν αξιοποιούνται εικόνες του ουρανού ως δεδομένα και δεν γίνεται ανάλυση των αποτελεσμάτων ανά καιρική συνθήκη. Επιπλέον, επιλέχθηκε να βελτιστοποιηθούν μόνο δύο υπεραπαράμετροι του μοντέλου RF, και όχι όλες οι βασικές του υπεραπαράμετροι.

1.3 ΔΟΜΗ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

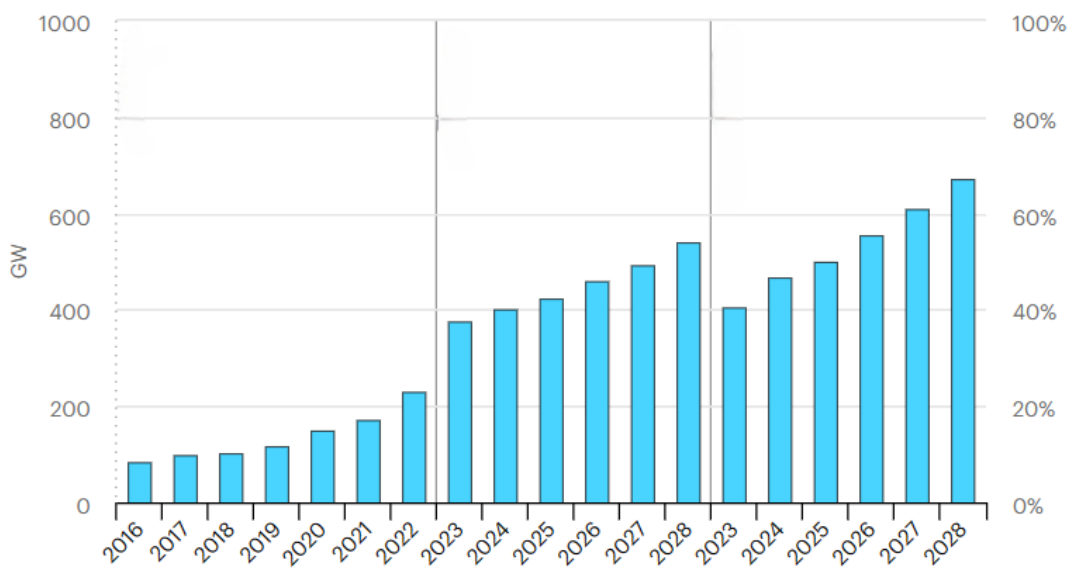
Η παρούσα διπλωματική εργασία οργανώνεται σε οκτώ κεφάλαια:

- Στο Κεφάλαιο 2 παρουσιάζεται ο βασικός τρόπος λειτουργίας ενός φωτοβολταϊκού σταθμού και επισημαίνεται η αξία της πρόβλεψης φωτοβολταϊκής παραγωγής. Επίσης, καταγράφονται οι κυριότεροι παράγοντες που επηρεάζουν τη φωτοβολταϊκή παραγωγή και υλοποιούνται διαχωρισμοί με βάση τις μεθόδους και τον χρονικό ορίζοντα των χρησιμοποιούμενων μοντέλων πρόβλεψης.
- Στο Κεφάλαιο 3 προσφέρεται το θεωρητικό υπόβαθρο του μοντέλου μηχανικής μάθησης RF και παρουσιάζονται οι αντικειμενικές συναρτήσεις που χρησιμοποιούνται. Επίσης παρουσιάζεται και αναλύεται ο αλγόριθμος ALO, καθώς και τα σημαντικότερα πλεονεκτήματα του.
- Στο Κεφάλαιο 4 γίνεται ανάλυση των δεδομένων που χρησιμοποιούνται στην παρούσα διπλωματική εργασία για την εκπαίδευση και αξιολόγηση του προτεινόμενου μοντέλου πρόβλεψης καθώς και περιγραφή των μεθόδων επεξεργασίας των δεδομένων.
- Στο Κεφάλαιο 5 παρουσιάζονται τα μετρητικά αξιολόγησης, η μεθοδολογία του προτεινόμενου υβριδικού μοντέλου πρόβλεψης, καθώς και η διαδικασία βελτιστοποίησης των παραμέτρων του.
- Στο Κεφάλαιο 6 αναλύονται τα αποτελέσματα της αξιολόγησης του προτεινόμενου υβριδικού μοντέλου πρόβλεψης.
- Στο Κεφάλαιο 7 πραγματοποιείται η σύνοψη της διπλωματικής εργασίας, παρουσιάζονται τα συνολικά συμπεράσματα και καταγράφονται προτάσεις μελλοντικής επέκτασης.
- Στο Κεφάλαιο 8 περιλαμβάνεται η χρησιμοποιούμενη βιβλιογραφία για την εκπόνηση της παρούσας διπλωματικής εργασίας.

ΠΡΟΒΛΕΨΗ Φ/Β ΠΑΡΑΓΩΓΗΣ

2.1 ΓΕΝΙΚΑ

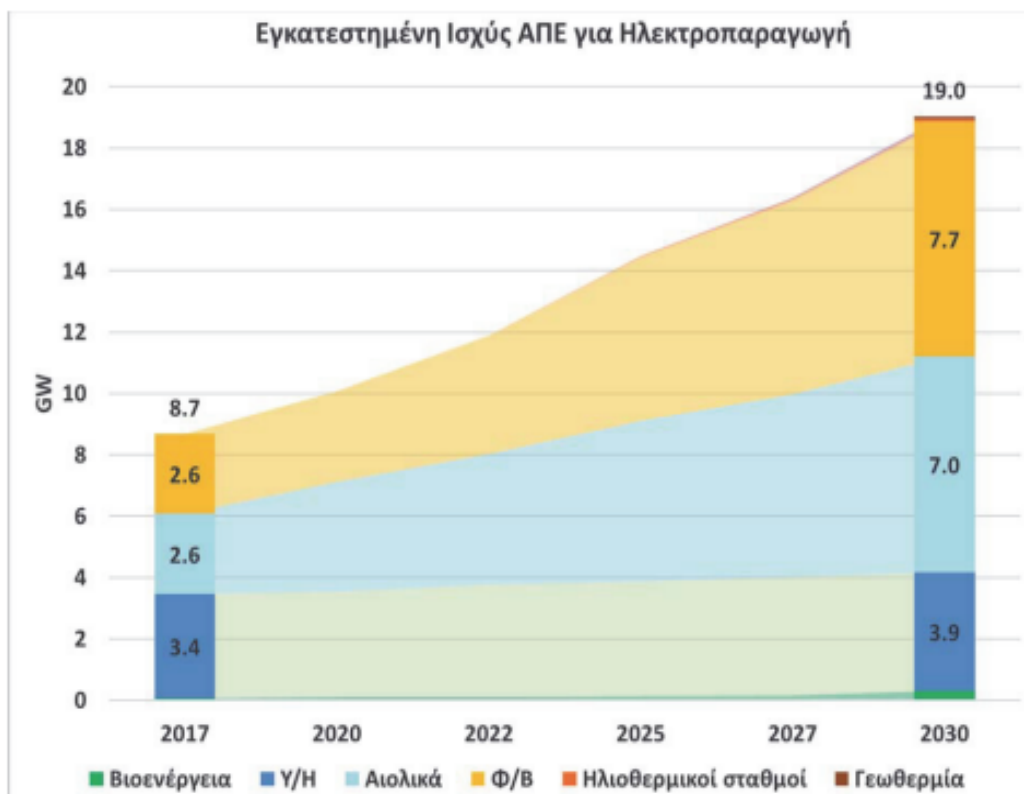
Με την πάροδο των χρόνων, το πρόβλημα της κλιματικής αλλαγής γίνεται όλο και πιο έντονο. Η συνεχής αύξηση των εκπομπών αερίων του θερμοκηπίου σε συνδυασμό με την απειλή της τρύπας του όζοντος απαιτούν επείγουσα δράση για τη μείωση των περιβαλλοντικών επιπτώσεων και την απεξάρτηση από τα ορυκτά καύσιμα. Εξαιτίας αυτής της κρίσιμης κατάστασης έχει παρατηρηθεί μία εμφανής αλλαγή προσανατολισμού η οποία εστιάζει στις ΑΠΕ ως βασική πηγή παραγωγής ενέργειας. Το 2023 οι ΑΠΕ συνεισέφεραν το 30,2% της συνολικής παγκόσμιας παραγωγής ηλεκτρικής ενέργειας [16]. Αυτό το ποσοστό συμπεριλαμβάνει τρεις κύριες πηγές: την αιολική ενέργεια, την ηλιακή ενέργεια, και την υδροηλεκτρική ενέργεια. Έτσι, μέχρι το 2028, η ηλιακή και αιολική παραγωγή θα αποτελούν το 95% της παγκόσμιας αύξησης παραγωγής ενέργειας που προέρχεται από ΑΠΕ [17]. Συγκεκριμένα, μεταξύ του 2022 και του 2023 υπήρξε μια συνολική αύξηση κατά 64% στις νέες εγκαταστάσεις Φ/Β, που αποτέλεσε παγκοσμίως το 75% περίπου της πρόσθετης εγκατεστημένης ισχύος στον τομέα των ΑΠΕ [18]. Στο Σχήμα 2.1 φαίνεται η παγκόσμια πρόσθετη εγκατεστημένη ισχύς των Φ/Β εγκαταστάσεων από το 2016 μέχρι το 2028[19].



Σχήμα 2.1: Η συνολική πρόσθετη εγκατεστημένη ισχύς από Φ/Β παγκοσμίως ανά έτος

Στην ελληνική επικράτεια, το 2021, το ποσοστό της κατανάλωσης ηλεκτρικής ενέργειας από ΑΠΕ ήταν 37,8%, ενώ το 2022 το ποσοστό έφθασε περίπου στο 43%. Επιπλέον το μερίδιο της ηλιακής ενέργειας στο συνολικό ενεργειακό μείγμα έφθασε στο 18,2% το 2022. Το 2023, παρατηρήθηκε σημαντική αύξηση της παραγωγής ηλεκτρικής ενέργειας μέσω Φ/Β στην Ελλάδα, φτάνοντας στο 19% της συνολικής εγχώριας παραγωγής [20].

Πέρα από αυτά, σύμφωνα με το αναθεωρημένο Εθνικό Σχέδιο για την Ενέργεια και το Κλίμα [21] προβλέπεται 7,7 GW εγκατεστημένης Φ/Β ισχύος μέχρι το 2030 και 40,3 GW εγκατεστημένης Φ/Β ισχύος μέχρι το 2050. Σύμφωνα με αυτά, η κατασκευή νέων Φ/Β σταθμών στην Ελλάδα προβλέπεται να ακολουθήσει αυξητική τάση και αναμένεται να εκτοξευθεί τα προσεχή έτη. Στο Σχήμα 2.2 φαίνεται η εξέλιξη της εγκατεστημένης ισχύος των Φ/Β σταθμών και γενικότερα των ΑΠΕ στην Ελλάδα.



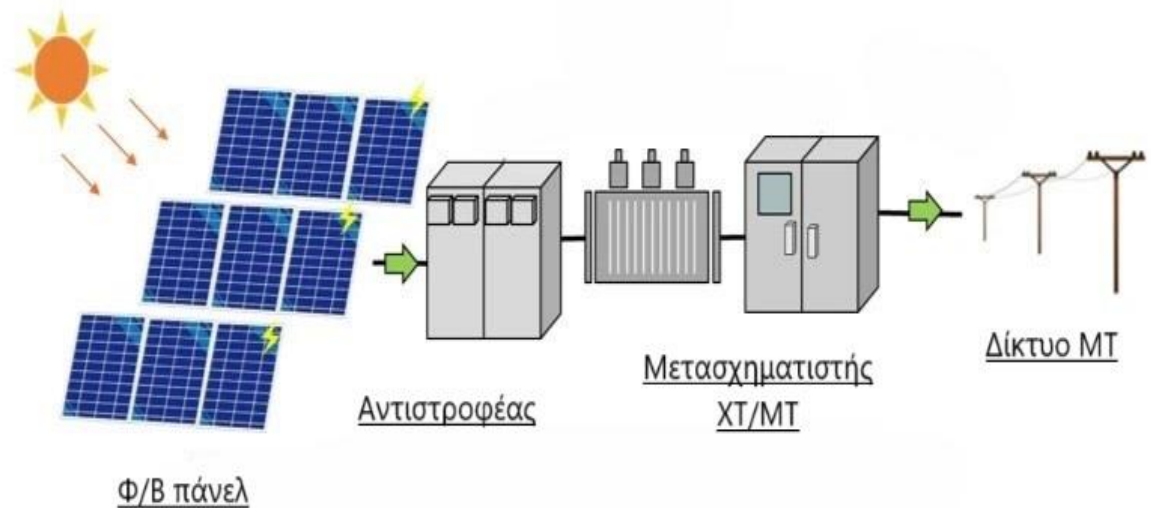
Σχήμα 2.2: Εξέλιξη της εγκατεστημένης ισχύος των ΑΠΕ από το 2017-2030 στην Ελλάδα[21]

2.2 ΤΡΟΠΟΣ ΛΕΙΤΟΥΡΓΙΑΣ ΦΩΤΟΒΟΛΤΑΪΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Τα Φ/Β συστήματα χρησιμοποιούνται για την παραγωγή ηλεκτρικού ρεύματος με τη μετατροπή μέρους του ορατού φάσματος της ηλιακής ακτινοβολίας. Αυτή η μετατροπή γίνεται συνήθως με την απορρόφηση του φωτός από κατάλληλες διατάξεις, όπως αυτή της διόδου p-n και σε ειδικές περιπτώσεις η απορρόφηση γίνεται με τη χρήση της διόδου Schottky με ημιδιαφανή επικάλυψη μετάλλου. Τα κύρια μέρη που απαρτίζουν ένα Φ/Β σύστημα είναι τα εξής:

- Φ/Β πάνελ (solar panel)

- Αντιστροφέας (inverter)
- Μετασχηματιστής Χαμηλής Τάσης / Μέσης Τάσης (transformer)
- Σύνδεση με το δίκτυο Μέσης Τάσης (electricity to the grid)



Σχήμα 2.3: Τυπική λειτουργία ενός Φ/Β σταθμού[22]

Όπως φαίνεται στο Σχήμα 2.3, τα Φ/Β πάνελ ή αλλιώς ηλιακοί συλλέκτες απορροφούν την ηλιακή ακτινοβολία και τη μετατρέπουν σε ηλεκτρική ενέργεια. Η ενέργεια που παράγεται είναι συνεχούς ρεύματος και το ρεύμα αυτό έπειτα μεταφέρεται στην είσοδο του αντιστροφέα (inverter). Ο αντιστροφέας είναι υπεύθυνος για τη μετατροπή της ηλεκτρικής ενέργειας από τη μορφή συνεχούς ρεύματος σε μορφή εναλλασσόμενου ρεύματος, το οποίο είναι αναγκαίο ώστε να είναι εφικτή η έγχυση της παραγόμενης ενέργειας στο δίκτυο μέσω του μετασχηματιστή ΧΤ/ΜΤ[23]. Ο Μ/Σ εκτός από το ότι είναι απαραίτητος για τη σύνδεση του συστήματος με το δίκτυο μεσαίας τάσης, βοηθά και στη μείωση των απωλειών κατά τη μεταφορά της ενέργειας. Τέλος, το Φ/Β σύστημα συνδέεται με το δίκτυο μέσης τάσης, όπου η παραγόμενη ηλεκτρική ενέργεια ενσωματώνεται στο ευρύτερο ηλεκτρικό δίκτυο[24].

2.3 ΠΑΡΑΓΟΝΤΕΣ ΠΟΥ ΕΠΗΡΕΑΖΟΥΝ ΤΗ ΦΩΤΟΒΟΛΤΑΪΚΗ ΜΕΤΑΤΡΟΠΗ

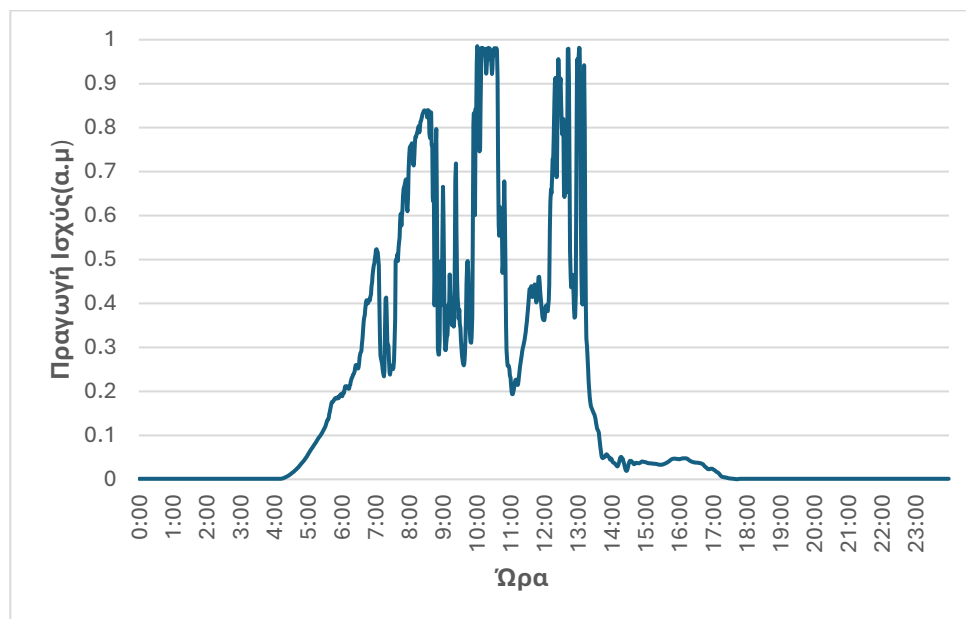
Η παραγωγή ηλεκτρικής ενέργειας από Φ/Β σταθμούς, όπως και η πλειοψηφία των ΑΠΕ, υπόκειται σε πολλαπλούς μη ελέγξιμους εξωτερικούς παράγοντες. Η παραγωγή ενός Φ/Β σταθμού εξαρτάται αναπόφευκτα από μετεωρολογικά και γεωγραφικά στοιχεία, καθώς και από την διαχείριση των δυναμικών περιβαλλοντικών παραμέτρων, όπως την διαθεσιμότητα του φωτός, τη θερμοκρασία, και τον άνεμο. Οι κυριότεροι από αυτούς τους παράγοντες είναι οι εξής [3]:

1. **Ηλιακή ακτινοβολία:** Αναμφισβήτητα, η κρισιμότερη παράμετρος για την παραγωγή ηλεκτρικής ενέργειας από Φ/Β συλλέκτες είναι η ένταση της ηλιακής ακτινοβολίας που φτάνει στην επιφάνειά τους. Αυτή η παράμετρος μπορεί να ποικίλει σημαντικά ανά έτος ή ακόμη και μέσα στην ίδια μέρα, παρά τη λεπτομερή μελέτη που έχει διεξαχθεί κατά την κατασκευή του Φ/Β πάρκου. Η έντονη παρουσία πυκνών νεφών μειώνει σημαντικά την παραγωγή ενέργειας, λόγω της μειωμένης προσπίπτουσας ηλιακής ακτινοβολίας. Για να απορροφήσει ένας Φ/Β συλλέκτης το μέγιστο ποσοστό της ηλιακής ακτινοβολίας, η επιφάνειά του πρέπει να είναι κάθετη στις ηλιακές ακτίνες. Αυτή η προϋπόθεση είναι δύσκολο να επιτευχθεί καθώς, η θέση του ήλιου στον ορίζοντα αλλάζει κατά τη διάρκεια της ημέρας, και κατά τις περιόδους όπου η γωνία είναι λιγότερο ευνοϊκή, όπως το πρωί και το απόγευμα, τα Φ/Β έχουν μειωμένη αποδοτικότητα. Ωστόσο, η πορεία του ήλιου έχει ημερήσια περιοδικότητα και ετήσιους κύκλους. Αυτό οδήγησε στην εισαγωγή νέων τεχνολογιών Φ/Β που μπορούν να περιστρέφονται. Συγκεκριμένα, οι Φ/Β ιχνηλάτες ακολουθούν έξυπνα την πορεία του ήλιου σε όλη τη διάρκεια της ημέρας, προσαρμόζοντας τους ηλιακούς συλλέκτες ώστε να βλέπουν απευθείας το φως του ήλιου, εξασφαλίζοντας βέλτιστη απορρόφηση ενέργειας.
2. **Θερμοκρασία:** Η ιδανική θερμοκρασία λειτουργίας των περισσότερων ειδών Φ/Β πάνελ, κατά την οποία παρέχεται η υψηλότερη δυνατή ισχύς εξόδου, βρίσκεται στους 25°C [25]. Με την αύξηση της θερμοκρασίας, το πάνελ παύει να μετατρέπει την ηλιακή ακτινοβολία σε ηλεκτρική ενέργεια με την ίδια απόδοση καθώς όσο αυξάνεται η θερμοκρασία, αυξάνεται και ο αριθμός των ηλεκτρονίων και των άκαμπτων θέσεων στον ημιαγωγό που μπορούν να διακινηθούν, εμποδίζοντας την αποτελεσματική μεταφορά τους. Με αποτέλεσμα, η απόδοση του Φ/Β πάνελ να μειωθεί. Κατά συνέπεια, υπάρχει μια μικρή αύξηση του ρεύματος και μια μεγαλύτερη πτώση τάσης άρα και μείωση της ισχύος εξόδου. Συγκεκριμένα, για κάθε βαθμό πάνω από τους 25°C, η απόδοση ενός Φ/Β πάνελ μπορεί να μειωθεί κατά περίπου 0,3% έως 0,5%. Τους καλοκαιρινούς μήνες τα πάνελ μπορεί να φτάσουν θερμοκρασίες περίπου 66°C ή ακόμη και υψηλότερες[26].
3. **Γήρανση:** Τα Φ/Β πάνελ συνήθως υφίστανται υποβάθμιση της απόδοσης ισχύος εξόδου κατά περίπου 0,5% ετησίως, που σημαίνει πως παράγουν τουλάχιστον 90% της ονομαστικής τους ικανότητας τα πρώτα 10 χρόνια και περίπου 80% στα επόμενα 10-15 χρόνια. Η υποβάθμιση αυτή οφείλεται στην σταδιακή φθορά του τεχνολογικού εξοπλισμού ενός Φ/Β σταθμού και στις συσσωρευμένες βλάβες που μπορεί αυτή να προκαλέσει[27].
4. **Άνεμος:** Ο άνεμος επηρεάζει τη θερμοκρασία και την ηλεκτρική απόδοση των Φ/Β πάνελ. Συγκεκριμένα, η αυξημένη ταχύτητα του ανέμου επηρεάζει την αεροδυναμική θέρμανση και τη συνεκτική ψύξη, μειώνοντας την θερμοκρασία των Φ/Β πάνελ και βελτιώνοντας την απόδοσή τους. Η ταχύτητα, η διεύθυνση, και οι συνθήκες ροής του ανέμου, συνεπώς, επηρεάζουν την απόδοση του Φ/Β συστήματος[28].

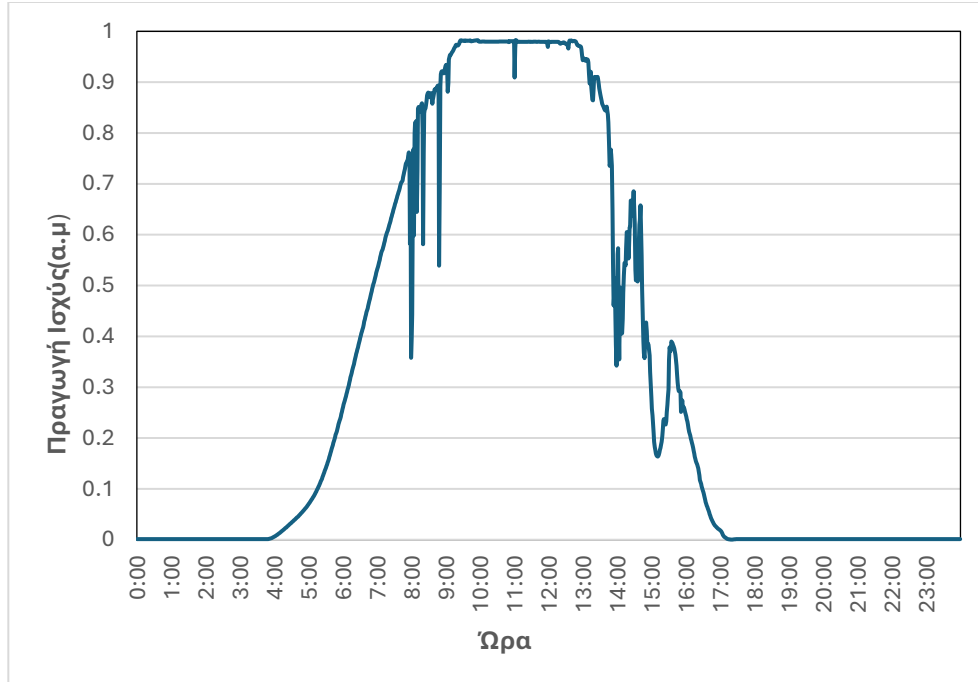
5. Σκόνη-υγρασία-φαινόμενο PID[29]: Η Δυνητική Επαγόμενη Υποβάθμιση (Potential Induced Degradation-PID) είναι ένα φαινόμενο που προκαλεί δυσλειτουργία στις κυψέλες των Φ/Β πλαισίων και οδηγεί στην πρόωρη γήρανσή τους. Το φαινόμενο PID προκαλείται από επιβλαβή ρεύματα διαρροής, τα οποία δημιουργούνται από τη διαφορά δυναμικού μεταξύ του πλαισίου και των Φ/Β κυψελών. Λόγω αυτής της ηλεκτρικής τάσης μπορεί να απελευθερωθούν ηλεκτρόνια από τα χρησιμοποιούμενα στη Φ/Β μονάδα υλικά τα οποία ρέουν μέσω του γειωμένου πλαισίου. Έτσι, παραμένει μία φόρτιση (πόλωση), η οποία επηρεάζει αρνητικά τη χαρακτηριστική καμπύλη των Φ/Β κυψελών. Η σκόνη και η ύπαρξη υγρασίας, ειδικά κατά τις πρωινές ώρες, αυξάνει την αγωγιμότητα του προστατευτικού γυαλιού των Φ/Β πάνελ, επιτρέποντας τη διαφυγή φορτίων προς τη γη όταν δημιουργείται τάση στο πλαίσιο από τις ακτίνες του ήλιου. Η υψηλή θερμοκρασία επιπλέον αυξάνει την κινητικότητα των φορέων και τη δημιουργία υδρατμών στην εσωτερική μεριά του γυαλιού, ενισχύοντας το φαινόμενο μέσω της αύξησης της αγωγιμότητας του ημιαγωγού.

2.4 ΣΗΜΑΝΤΙΚΟΤΗΤΑ ΠΡΟΒΛΕΨΗΣ ΦΩΤΟΒΟΛΤΑΪΚΗΣ ΠΑΡΑΓΩΓΗΣ

Οι ραγδαίες τεχνολογικές εξελίξεις έχουν συμβάλει στην επιταχυνόμενη μετάβαση από τα ορυκτά καύσιμα στις ΑΠΕ[30]. Παρόλο που η υψηλή διείσδυση των Φ/Β στα ηλεκτρικά συστήματα έχει φέρει πολλά οικονομικά και περιβαλλοντικά οφέλη, η συνολική διαδικασία παραγωγής ηλεκτρικής ενέργειας από Φ/Β απαιτεί αυξημένη απόδοση και αξιοπιστία. Από τη φύση τους, οι ΑΠΕ συνοδεύονται από σημαντικό βαθμό αβεβαιότητας λόγω της εξάρτησής τους από μετεωρολογικούς παράγοντες[31]. Στο Σχήμα 2.4 παρουσιάζεται η πραγματική ημερήσια καμπύλη της παραγόμενης ισχύος ενός Φ/Β σε μια ημέρα με έντονες νεφώσεις ενώ στο Σχήμα 2.5 παρουσιάζεται η αντίστοιχη καμπύλη κατά τη διάρκεια πλήρους ηλιοφάνειας.



Σχήμα 2.4: Ημερήσια καμπύλη παραγόμενης ισχύος Φ/Β σε μια ημέρα με έντονες νεφώσεις



Σχήμα 2.5: Ημερήσια καμπύλη παραγόμενης ισχύος Φ/Β σε μια ημέρα με ηλιοφάνεια

Η δυνατότητα ακριβούς πρόβλεψης της ενέργειας που παράγεται από τα Φ/Β συστήματα παίζει καθοριστικό ρόλο. Συγκεκριμένα, μερικές από τις μονάδες που απαρτίζουν τα Συστήματα Ηλεκτρικής Ενέργειας (ΣΗΕ) λειτουργούν ως εφεδρικές γεννήτριες στην περίπτωση απόκλισης μεταξύ της προβλεπόμενης και της παραγόμενης ενέργειας καλύπτονται από αυτές τις λοιπές τεχνολογίες. Έτσι, μια ακριβής πρόβλεψη της Φ/Β παραγωγής θα μπορούσε να μειώσει τον αριθμό των μονάδων σε εφεδρεία και, συνεπώς, να μειώσει τα λειτουργικά κόστη.

Επίσης, μια ακριβής πρόβλεψη είναι επωφελής όχι μόνο για τους λειτουργούς του συστήματος, αλλά και για τους διαχειριστές των Φ/Β εγκαταστάσεων, καθώς συμβάλλει στην αποφυγή πιθανών κυρώσεων που προκύπτουν λόγω αποκλίσεων μεταξύ της προβλεπόμενης και της παραγόμενης ηλεκτρικής ενέργειας. Πέρα από αυτά, η ακριβής πρόβλεψη της Φ/Β παραγωγής, συμβάλει στη διασφάλιση της ευστάθειας του ηλεκτρικού δικτύου μέσω σταθεροποίησης της τάσης και της συχνότητας. Συγκεκριμένα, μέσω της πρόβλεψης, οι διαχειριστές του δικτύου μπορούν να προβλέπουν τις διακυμάνσεις στην παραγωγή ηλεκτρικής ενέργειας και να λαμβάνουν μέτρα για να διατηρούν τη συχνότητα στα επιθυμητά επίπεδα για την ομαλή λειτουργία των ηλεκτρικών συσκευών. Παράλληλα, μπορούν να διατηρούν σταθερή την τάση, προβαίνοντας σε ρυθμίσεις για να διασφαλίσουν την ισορροπία μεταξύ της παραγωγής και της κατανάλωσης ηλεκτρικής ενέργειας.

Τέλος, η πρόβλεψη της παραγωγής ωφελεί τον σύγχρονο τρόπο λειτουργίας των αγορών ηλεκτρικής ενέργειας, ιδίως το χρηματιστήριο ενέργειας, με τους εξής τρόπους [32]:

- **Αύξηση της διαφάνειας:** Οι προβλέψεις της Φ/Β παραγωγής προσφέρουν ανοικτή πρόσβαση και ανταλλαγή πληροφοριών μεταξύ των συμμετεχόντων στην αγορά για τις αναμενόμενες ποσότητες ηλεκτρικής ενέργειας που θα εισαχθούν στο δίκτυο.

- Ελαχιστοποίηση κινδύνων: Μειώνονται οι κίνδυνοι που συνδέονται με τις απρόβλεπτες διακυμάνσεις της παραγωγής ηλιακής ενέργειας και έτσι βελτιώνεται η αποδοτικότητα της παραγωγής και η χρήση ενεργειακών πόρων.
- Υποστήριξη στην πλατφόρμα εμπορίου ενέργειας: Παρέχονται οι απαραίτητες πληροφορίες για τις συναλλαγές και τη διαχείριση των ενεργειακών συμβολαίων[33].

2.5 ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ ΦΩΤΟΒΟΛΤΑΪΚΗΣ ΠΑΡΑΓΩΓΗΣ

Στην προσπάθεια να περιοριστούν τα μειονεκτήματα που απορρέουν από την τυχαιότητα της Φ/Β παραγωγής, έχουν αναπτυχθεί πολυάριθμες μέθοδοι πρόβλεψης που έχουν ως στόχο την προσέγγιση της πραγματικής Φ/Β παραγωγής στον αντίστοιχο επιθυμητό χρονικό ορίζοντα. Οι μέθοδοι πρόβλεψης της Φ/Β παραγωγής ταξινομούνται σε τρεις κύριες κατηγορίες, κάθε μία από τις οποίες παρουσιάζει συγκεκριμένα πλεονεκτήματα και μειονεκτήματα[2].

2.5.1 Φυσικές μέθοδοι

Τα φυσικά μοντέλα πρόβλεψης Φ/Β παραγωγής αξιοποιούν μαθηματικές εξισώσεις και προσομοιώσεις του φυσικού περιβάλλοντος, τα οποία συσχετίζουν μεταβλητές όπως την ηλιακή ακτινοβολία, την ομοιομορφία της σκίασης, τη ρύπανση (βρωμιά), τη γήρανση, τη θερμοκρασία των κυψελών, τη γωνία πρόσπτωσης της ηλιακής ακτινοβολίας, την κατάσταση φόρτου, την κίνηση και μορφή των νεφών, που επηρεάζουν άμεσα τη λειτουργία ενός Φ/Β σταθμού και την παραγόμενη ισχύ του [34]. Συγκεκριμένα, στην περίπτωση υψηλότερων χωρικών (γεωγραφικών) αναλύσεων, μέσω προσομοιώσεων ατμοσφαιρικών φαινομένων προκύπτουν τα Αριθμητικά Μοντέλα Πρόβλεψης Καιρού (Numerical Weather Predictions - NWP's) [7], [35] τα οποία χρησιμοποιούν υπολογιστές για να επιλύσουν τις κυρίαρχες εξισώσεις της ατμόσφαιρας. Στην περίπτωση μικρότερων χωρικών αναλύσεων (π.χ. Φ/Β στέγης) αντί για αριθμητικά μοντέλα πρόβλεψης καιρού χρησιμοποιούνται δεδομένα που προκύπτουν από τη φυσική μοντελοποίηση της κίνησης των νεφώσεων και του ήλιου.

Το κύριο πλεονέκτημα αυτής της μεθόδου είναι πως δεν απαιτεί ιστορικά δεδομένα καθιστώντας δυνατή την εκτίμηση της ισχύος εξόδου ενός Φ/Β πάρκου ακόμα και πριν από την κατασκευή του γνωρίζοντας τις τεχνικές προδιαγραφές του [7]. Επίσης, βασικό πλεονέκτημα των φυσικών μοντέλων είναι πως έχουν τη δυνατότητα να παράγουν προβλέψεις ισχύος με υψηλή ακρίβεια σε μικρά χρονικά διαστήματα. Μπορούν να αναπαράγουν με ακρίβεια τις αλλαγές στην ισχύ και άλλες μετρήσεις σε πολύ σύντομα χρονικά διαστήματα, δίνοντας μια λεπτομερή εικόνα της καμπύλης ισχύος που είναι εξαιρετικά κοντά στις πραγματικές συνθήκες.

Κάποια από τα μειονέκτημα των φυσικών μεθόδων είναι πως συνήθως αναλώνουν μεγάλα χρηματικά ποσά για την υλοποίησή τους και την απόκτηση των απαραίτητων δεδομένων. Επιπλέον οι πληροφορίες που παρέχονται από τους κατασκευαστές είναι συνήθως περιορισμένες και αναφέρονται μόνο στις ονομαστικές λειτουργικές συνθήκες. Τέλος, το μεγαλύτερο τους μειονέκτημα είναι ότι κάθε φυσικό μοντέλο πρόβλεψης είναι εξατομικευμένο για ένα συγκεκριμένο Φ/Β σύστημα, λαμβάνοντας υπόψη την τοπογραφία, την κλιματολογία και τις τεχνικές προδιαγραφές του εκάστοτε τοποθετημένου εξοπλισμού .

2.5.2 Στατιστικές μέθοδοι

Σε αντίθεση με τα φυσικά μοντέλα, οι στατιστικές μέθοδοι δεν χρειάζονται τοπογραφικά χαρακτηριστικά του εδάφους, τεχνικές λεπτομέρειες του εξοπλισμού, δεδομένα από αισθητήρες και γενικά οποιαδήποτε εσωτερική πληροφορία από τον Φ/Β σταθμό για να τον μοντελοποιήσουν. Οι στατιστικές μέθοδοι εξάγουν σχέσεις από ιστορικά δεδομένα για να προβλέψουν τη μελλοντική παραγωγή του Φ/Β σταθμού. Συνεπώς, η ποιότητα των ιστορικών παρατηρήσεων είναι ουσιώδης για μια ακριβή πρόβλεψη [7]. Επίσης, απαιτείται συνήθως ένα μεγάλο ιστορικό σύνολο δεδομένων για το οποίο ο Φ/Β σταθμός πρέπει να λειτουργεί ήδη για κάποιο χρονικό διάστημα.

Κάποια από τα βασικά πλεονεκτήματα αυτών των μεθόδων είναι η απλότητά της στην υλοποίηση και τη χρήση, το ότι έχουν χαμηλότερο κόστος από τις φυσικές μεθόδους και πως δεν χρειάζονται κάποια ιδιαίτερη εξειδίκευση. Τα βασικά μειονεκτήματα των στατιστικών μεθόδων είναι η πλήρης εξάρτησή τους από την αξιοπιστία των μετρητικών οργάνων, η ευαισθησία τους σε ακραίες τιμές και ανωμαλίες, καθώς και η αδυναμία τους να εξάγουν σύνθετες μη γραμμικές σχέσεις

2.5.3 Μοντέλα μηχανικής μάθησης

Η πρόοδος στην επιστήμη των υπολογιστών σε συνδυασμό με την εξέλιξη της τεχνητής νοημοσύνης τα τελευταία χρόνια έχει επιτρέψει στις υπολογιστικές μηχανές να αναλύουν δεδομένα με μεγαλύτερη ταχύτητα και ακρίβεια. Η εφαρμογή των μοντέλων μηχανικής μάθησης έχει επεκταθεί σε ποικίλους τομείς, ένας από αυτούς είναι η πρόβλεψη παραγωγής ενέργειας από Φ/Β συστήματα. Τα μοντέλα μηχανικής μάθησης αξιοποιούν αλγορίθμους που εκπαιδεύονται για να βρουν μοτίβα και συσχετίσεις σε μεγάλα σύνολα δεδομένων, εισόδου και εξόδου, ώστε να λαμβάνουν τις καλύτερες αποφάσεις και προβλέψεις με βάση αυτή την ανάλυση [7]. Επιπλέον, μπορούν να ανιχνεύουν πολύπλοκες, μη γραμμικές και δύσκολα αναγνωρίσιμες συσχετίσεις μεταξύ μετεωρολογικών και τεχνικών παραμέτρων και της παραγόμενης ισχύος.

Η συγκεκριμένη προσέγγιση αποτελεί προέκταση των στατιστικών μεθόδων, απομακρύνοντας μερικά από τα μειονεκτήματά τους, καθώς επιτυγχάνουν σημαντικές βελτιώσεις στην δυνατότητα προσέγγισης σύνθετων μη γραμμικών σχέσεων.

Σε σχέση με τα φυσικά μοντέλα, ωστόσο, συνήθως είναι χειρότερα στο να κάνουν μακροπρόθεσμες προβλέψεις σε μεγάλες χωρικές αναλύσεις. Επίσης, δεν είναι σε θέση να αξιολογήσουν με την ίδια λεπτομέρεια την καμπύλη ισχύος, όπως ένα πιο περίπλοκο φυσικό μοντέλο και για να το επιτύχουν αυτό συνήθως απαιτείται αύξηση του υπολογιστικού φόρτου και των ιστορικών δεδομένων.

Για τη σωστή πρόβλεψη βάσει μοντέλων μηχανικής μάθησης απαιτείται η ύπαρξη αρκετών δεδομένων εισόδου υψηλής ποιότητας, η κατάλληλη διαδικασία εκπαίδευσης, και η αποφυγή υπερπροσαρμογής (overfitting) [36]. Οι αρχές λειτουργίας των μοντέλων μηχανικής μάθησης και μιας βασικής υποκατηγορίας τους που αφορά τα Random Forests, παρουσιάζονται αναλυτικότερα στο Κεφάλαιο 3.

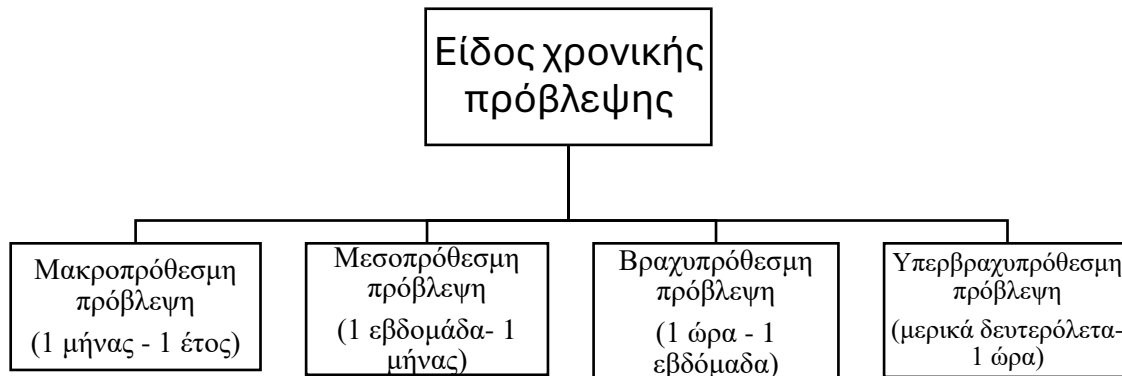
2.6 ΧΡΟΝΙΚΗ ΤΑΞΙΝΟΜΗΣΗ ΜΟΝΤΕΛΩΝ ΠΡΟΒΛΕΨΗΣ

Σχετικά με την χρονική πτυχή της παραγωγής προβλέψεων υπάρχουν τρεις σημαντικές έννοιες:

- Ο ορίζοντας πρόβλεψης (forecast horizon)
- Η ανάλυση πρόβλεψης (forecast resolution)
- Το διάστημα πρόβλεψης (forecast interval)

Ο ορίζοντας πρόβλεψης ορίζεται ως το χρονικό διάστημα μεταξύ χρονικής στιγμής παραγωγής της πρόβλεψης και της χρονικής στιγμής για την οποία παράγεται η πρόβλεψη. Η χρονική ανάλυση πρόβλεψης περιγράφει τη συχνότητα με την οποία εκδίδονται οι προβλέψεις και το διάστημα πρόβλεψης υποδεικνύει το χρονικό εύρος των προβλέψεων [34]. Ο σκοπός και η ακρίβεια ενός μοντέλου πρόβλεψης ισχύος Φ/Β εξαρτώνται από τον ορίζοντα πρόβλεψης. Ταυτόχρονα, όσο αυξάνεται το διάστημα της πρόβλεψης τόσο δυσχεραίνει η διαδικασία προσέγγισης της πραγματικής ισχύος, το οποίο ωστόσο δεν καθιστά μια μακροχρόνια πρόβλεψη πιο ασήμαντη από μια πιο άμεση, καθώς κάθε κατηγορία πρόβλεψης προσφέρει μοναδική πληροφορία, ενώ η συνδυαστική προσέγγισή τους είναι συχνά ορίζουσα για αποτελεσματικές ενεργειακές πολιτικές και στρατηγικές.

Οι τέσσερις βασικές χρονικές κλίμακες ταξινόμησης της Φ/Β πρόβλεψης είναι η υπερβραχυπρόθεσμη, η βραχυπρόθεσμη, η μεσοπρόθεσμη και η μακροπρόθεσμη [37], όπως φαίνεται και στο Σχήμα 2.6.



Σχήμα 2.6: Είδη χρονικών προβλέψεων ηλιακής μετατροπής

2.6.1 Μακροπρόθεσμη πρόβλεψη

Οι προβλέψεις αυτές αφορούν χρονικούς ορίζοντες που εκτείνονται από ένα μήνα μέχρι και πέραν του ενός έτους. Στις περιπτώσεις μακροπρόθεσμων προβλέψεων οι στατιστικές μέθοδοι και τα μοντέλα μηχανικής μάθησης έχουν μειωμένη απόδοση καθώς οι ιστορικές χρονοσειρές που αξιοποιούνται χάνουν σημαντική αξία σε τόσο μεγάλα χρονικά διαστήματα. Για τον λόγο αυτό, οι φυσικές μέθοδοι βρίσκουν σχεδόν καθολική εφαρμογή.

Η μακροπρόθεσμη πρόβλεψη είναι σημαντική για τον σχεδιασμό και την αποτελεσματική λειτουργία του συστήματος παραγωγής, μεταφοράς και διανομής ηλεκτρικής ενέργειας. Αυτό συμβαίνει επειδή οι προβλέψεις αυτές παρέχουν στους φορείς του ενεργειακού τομέα την αναγκαία προβλεπτική πληροφορία για να προγραμματίσουν και να διαχειριστούν σωστά τις εργασίες και τις απαιτήσεις τους με βάση την προβλεπόμενη παραγωγή ηλεκτρικής ενέργειας από τα Φ/Β συστήματα. Έτσι, μπορούν να ληφθούν αποφάσεις σχετικά με τον σχεδιασμό και την ανάπτυξη νέων εγκαταστάσεων παραγωγής ενέργειας, την αναβάθμιση του δικτύου μεταφοράς και διανομής, καθώς και τη διαχείριση των ενεργειακών πόρων. Επιπλέον, είναι απαραίτητη για την προσαρμογή της ενεργειακής πολιτικής και την ανάπτυξη μακροπρόθεσμων στρατηγικών παραγωγής ενέργειας από ΑΠΕ.

2.6.2 Μεσοπρόθεσμη πρόβλεψη

Στην κατηγορία αυτή εντάσσονται όλες οι προβλέψεις οι οποίες διενεργούνται για χρονικούς ορίζοντες μεταξύ της μίας εβδομάδας και του ενός μήνα. Στις περιπτώσεις αυτές, η διαδικασία παραγωγής των προβλέψεων μπορεί να γίνει είτε από πολύπλοκες φυσικές μεθόδους, ή με χρήση στατιστικών μεθόδων και μοντέλων μηχανικής μάθησης. Αυτός ο τύπος πρόβλεψης εξομαλύνει τον σχεδιασμό του ηλεκτρικού συστήματος, προβλέποντας τη διαθεσιμότητα της ηλεκτρικής ενέργειας στο μεσοπρόθεσμο μέλλον. Επίσης, παίζει σημαντικό ρόλο στον προγραμματισμό της συντήρησης των Φ/Β σταθμών παραγωγής με οικονομικά αποδοτικό τρόπο, ειδικότερα όταν αναμένεται χαμηλή Φ/Β παραγωγή.

2.6.3 Βραχυπρόθεσμη πρόβλεψη

Αφορά χρονικούς ορίζοντες από μία ώρα έως και επτά ημέρες, ενώ συχνά η κατηγορία αυτή συμπεριλαμβάνει την υπερ-βραχυπρόθεσμη πρόβλεψη, υποκατηγορία που παρουσιάζεται αναλυτικότερα στην Ενότητα 2.6.4. Στην μέθοδο αυτή αξιοποιούνται στατιστικές μέθοδοι, μοντέλα μηχανικής μάθησης και φυσικές μέθοδοι, χρησιμοποιώντας μεταξύ άλλων δεδομένα δορυφορικών εικόνων ως είσοδο σε μοντέλα για πρόβλεψη της νεφοκάλυψης.

Η βραχυπρόθεσμη πρόβλεψη της παραγωγής ηλεκτρικής ενέργειας από Φ/Β εξασφαλίζει τη δέσμευση μονάδων για την ικανοποίηση της αναμενόμενης ζήτησης, τον προγραμματισμό και σχεδιασμό του χρονοδιαγράμματος και των δραστηριοτήτων για παραγωγή ενέργειας, καθώς και την μεταφορά και παροχή ενέργειας στους τελικούς καταναλωτές. Πέρα από αυτά, έχει σημαντική συνεισφορά στη συμμετοχή στο χρηματιστήριο ενέργειας, αφού οι προβλέψεις αυτές χρησιμοποιούνται για να διαχειριστούν τους κινδύνους που συνδέονται με την αστάθεια της παραγωγής ενέργειας. Τέλος,

η βραχυπρόθεσμη πρόβλεψη είναι χρήσιμη για τη βέλτιστη διαχείριση ενός ΣΗΕ, καθώς και για την ασφαλή και εντός ορίων λειτουργία των δικτύων διανομής και μεταφοράς ηλεκτρικής ενέργειας.

2.6.4 Υπερβραχυπρόθεσμη πρόβλεψη

Οι προβλέψεις αυτές αφορούν χρονικούς ορίζοντες από μερικά δευτερόλεπτα έως μία ώρα, ενώ συχνά αναφέρεται και ως intra-hour forecasting ή nowcasting[13]. Αυτό το είδος πρόβλεψης αξιοποιείται για τον καλύτερο σχεδιασμό και την ασφαλή και αποδοτική λειτουργία των συστημάτων ενέργειας. Η ακριβής πρόβλεψη σε υπερβραχυπρόθεσμο χρονικό ορίζοντα είναι ζωτικής σημασίας για την πραγματική χρονική αποστολή μονάδων και τη διαχείριση της ισορροπίας και της ευστάθειας των ενεργειακών συστημάτων, καθώς και για τη διαχείριση της ενέργειας των διανομέων για να εξισορροπήσουν τη ζήτηση [14]. Επίσης, έχει σημαντική εφαρμογή σε μικρότερες εγκαταστάσεις, όπως Φ/Β στέγης, και γενικότερα αποτελεί την τωρινή τάση λόγω της βελτίωσης της ευελιξίας των Φ/Β που απαιτεί την λήψη αποφάσεων σε σχεδόν πραγματικό χρόνο.

Η υπερβραχυπρόθεσμη πρόβλεψη είναι απαραίτητη ώστε να δίνετε μια λεπτομερής εικόνα σχετικά με τις γρήγορες και απότομες διακυμάνσεις της παραγωγής ηλεκτρικής ενέργειας. Όμως, ο μικρός χρονικός ορίζοντάς της, την καθιστά εξαιρετικά απαιτητική, καθώς υπάρχουν περιορισμοί σε ότι αφορά την υπολογιστική πολυπλοκότητα. Σε τέτοιες υψηλές χρονικές αναλύσεις χρησιμοποιούνται δεδομένα εικόνων του ουρανού για να δώσουν μια καλή εικόνα για την πορεία των νεφών και αξιοποιούνται αισθητήρες οπτικής παρακολούθησης του ουρανού και ραντάρ [38].

Στις περιπτώσεις αυτές, η διαδικασία παραγωγής των προβλέψεων μπορεί να γίνει με χρήση μοντέλων μηχανικής μάθησης, τα οποία είναι πιο αποδοτικά και λεπτομερή, αλλά απαιτείται προσοχή ώστε να μην παραβιαστούν οι περιορισμοί σε ότι αφορά τους υπολογιστικούς πόρους. Τα φυσικά μοντέλα παρόλο που μπορούν να είναι εξαιρετικά αποδοτικά και να παράγουν λεπτομερέστερες προβλέψεις, χρειάζονται τεράστια εξειδίκευση για να αναπτυχθούν και για τον λόγο αυτό γενικά αποφεύγονται. Αντίστοιχα, τα στατιστικά μοντέλα γενικά δεν χρησιμοποιούνται επειδή αδυνατούν να παράγουν ακριβείς προβλέψεις σε πολύ υψηλή χρονική ανάλυση.

ΜΟΝΤΕΛΟ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ RANDOM FOREST

3.1 ΓΕΝΙΚΑ

Τεχνητή νοημοσύνη ορίζεται ως η ικανότητα μιας μηχανής να επιδεικνύει λειτουργίες που συνδέονται με την ανθρώπινη νοημοσύνη, όπως η δημιουργικότητα, η μάθηση και η επίλυση προβλημάτων. Τα συστήματα τεχνητής νοημοσύνης έχουν αδιαμφισβήτητα κατακλείσει την καθημερινότητα των ανθρώπων σε όλο και περισσότερους τομείς και σύμφωνα με προβλέψεις θα οδηγήσουν σε ακόμη μεγαλύτερες αλλαγές, ακόμη και στον πλήρη ψηφιακό μετασχηματισμό της κοινωνίας.

Κλάδος της επιστήμης των υπολογιστών και γενική υποκατηγορία της τεχνικής νοημοσύνης αποτελεί η μηχανική μάθηση, η οποία επικεντρώνεται στη χρήση αλγορίθμων και δεδομένων με στόχο να μιμηθεί τον τρόπο που μαθαίνουν οι άνθρωποι. Τα μοντέλα μηχανικής μάθησης συνεπικουρούμενα από την άνοδο της πληροφορικής, έχουν επεκταθεί και στον ενεργειακό τομέα, ιδιαίτερα στην περίπτωση της πρόβλεψης της φωτοβολταϊκής παραγωγής, καθώς αξιοποιούνται όλο και περισσότερο για προβλήματα ταξινόμησης και προβλέψεων. Η επέκταση αυτή οφείλεται στην ικανότητα των μοντέλων μηχανικής μάθησης να διαχειρίζονται σύνθετα και μη γραμμικά φαινόμενα, διακρίνοντας δυσνόητα μοτίβα μεταξύ δεδομένων εισόδου και εξόδου.

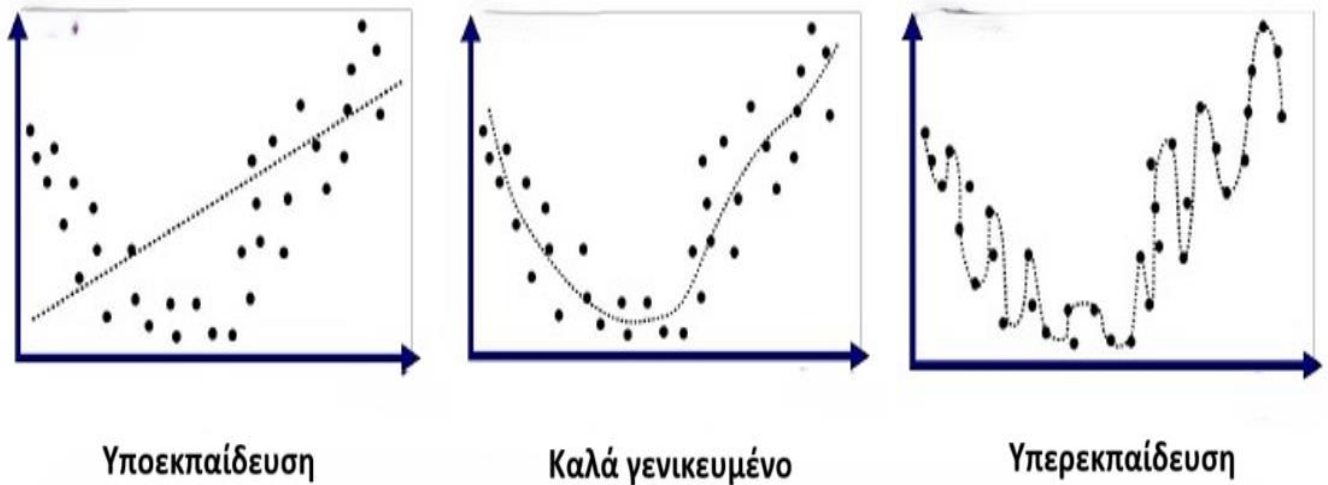
Ο τρόπος εκπαίδευσης των μοντέλων μηχανικής μάθησης συνήθως ταξινομείται σε δύο μεγάλες κατηγορίες, την επιτηρούμενη και μη επιτηρούμενη μάθηση [39].

- Η επιτηρούμενη μάθηση αναφέρεται στη διαδικασία εξαγωγής σχέσεων μεταξύ των δεδομένων εισόδου και εξόδου. Τα υποσύνολο δεδομένων που αξιοποιείται στην εκπαίδευση του μοντέλου περιλαμβάνει ετικέτες (labels) ώστε να μπορεί το μοντέλο μηχανικής μάθησης να εκπαιδευτεί σωστά και να αναπροσαρμόσει τις παραμέτρους του κατάλληλα, ώστε να προσομοιώσει την σχέση εισόδου-εξόδου. Στόχος της επιτηρούμενης μάθησης είναι η επιτυχής επέκταση της εκπαίδευσης σε άγνωστα δεδομένα χωρίς ετικέτες και τη δημιουργία αξιόπιστων προβλέψεων πάνω σε αυτά. Παραδείγματα μοντέλων μηχανικής μάθησης που αξιοποιούν επιτηρούμενη μάθηση είναι τα νευρωνικά δίκτυα και τα Random Forest (RF) καθώς και άλλοι αλγόριθμοι ταξινόμησης και παλινδρόμησης.
- Η μη επιτηρούμενη μάθηση δεν χρησιμοποιεί ετικέτες στα σύνολα δεδομένων. Αυτό σημαίνει ότι τα μοντέλα μηχανικής μάθησης λειτουργούν αυτόνομα για να ανακαλύψουν την ενσωματωμένη δομή και διάφορα μοτίβα μεταξύ των δεδομένων. Ως

επακόλουθο, απαιτείται η ανθρώπινη παρέμβαση για την επικύρωση των μεταβλητών εξόδου, η οποία εξαρτάται από την προοριζόμενη χρήση των δεδομένων. Αυτό το είδος εκπαίδευσης χρησιμοποιείται κυρίως για να αποκαλύψει ενδιαφέρουσες πληροφορίες μεγάλων όγκων δεδομένων και να ανιχνεύσει ανωμαλίες στα δεδομένα.

Τέλος, στα μοντέλα μηχανικής μάθησης, δύο συχνά υποδεικνύόμενα προβλήματα για τις κακές επιδόσεις και την έλλειψη ακρίβειας είναι η υποεκπαίδευση (underfitting) και η υπερεκπαίδευση (overfitting) [36]. Στο Σχήμα 3.1 φαίνονται παραδείγματα υποεκπαιδευμένων και υπερεκπαιδευμένων μοντέλων.

- **Υποεκπαίδευση:** Αντιπροσωπεύει την ανικανότητα του μοντέλου να αντιληφθεί και να μάθει τις πολύπλοκες σχέσεις μεταξύ των δεδομένων εκπαίδευσης, με αποτέλεσμα την κακή του απόδοση τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα ελέγχου. Στην περίπτωση αυτή, το μοντέλο είναι ανακριβές, ειδικά όταν εφαρμόζεται σε νέα δεδομένα. Η υποεκπαίδευση συμβαίνει κυρίως όταν χρησιμοποιούνται πολύ απλά μοντέλα στη βάση υπερβολικά απλοποιημένων υποθέσεων. Για να αντιμετωπιστεί το πρόβλημα της υποεκπαίδευσης, πρέπει να χρησιμοποιηθούν πιο πολύπλοκα μοντέλα, με μεγαλύτερες δυνατότητες αναπαράστασης χαρακτηριστικών. Επίσης, στη μείωση της υποεκπαίδευσης συμβάλει και η μείωση της έντασης της κανονικοποίησης των δεδομένων, προκειμένου να μην περιοριστεί υπερβολικά η ικανότητα του μοντέλου να μάθει την πολυπλοκότητα των δεδομένων.
- **Υπερεκπαίδευση:** Συμβαίνει όταν ένα μοντέλο εκπαιδεύεται υπερβολικά πολύ σε συγκεκριμένα δεδομένα και έτσι υπερπροσαρμόζεται σε αυτά. Ως εκ τούτου, πέρα από τις γενικές σχέσεις μεταξύ εισόδου και εξόδου, αρχίζει να προσομοιώνει θόρυβο και τις ανακριβείς καταχωρήσεις δεδομένων, με αποτέλεσμα να χάνει τη δυνατότητα γενίκευσης σε δεδομένα πέρα του συνόλου εκπαίδευσης. Οφείλεται κυρίως στην υπερβολική πολυπλοκότητα των μοντέλων, στο ότι αξιοποιεί πολλές μη παραμετρικές και μη γραμμικές μεθόδους και στη μη αντιπροσωπευτικότητα των δεδομένων εκπαίδευσης. Για να αντιμετωπιστεί το πρόβλημα της υπερεκπαίδευσης πρέπει να βελτιωθεί η ποιότητα των δεδομένων εστιάζοντας σε σημαντικά πρότυπα, αντιμετωπίζοντας τον κίνδυνο προσαρμογής στον θόρυβο και στα ανακριβή χαρακτηριστικά. Επίσης, στη μείωση της υπερεκπαίδευσης συμβάλουν οι τεχνικές κανονικοποίησης των δεδομένων και η επιλογή πιο απλών μοντέλων.

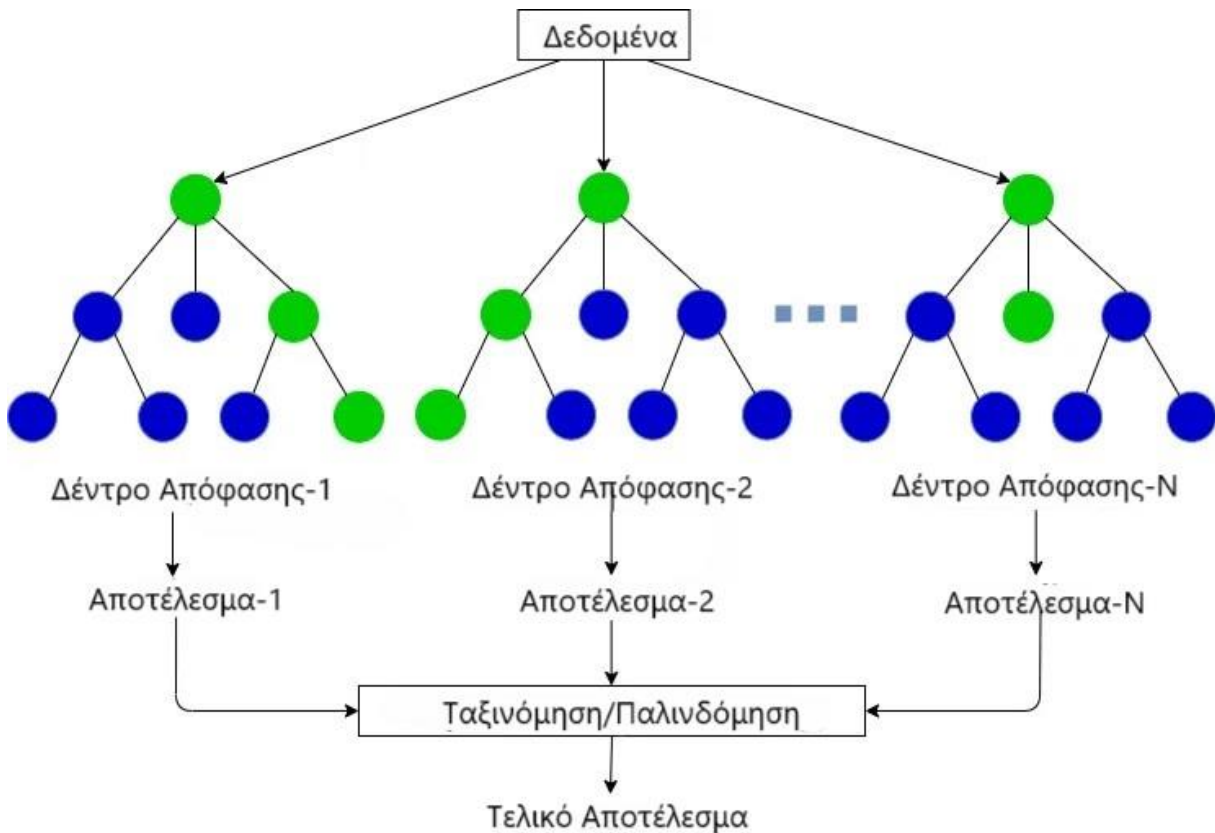


Σχήμα 3.1: Περιπτώσεις υποεκπαιδευμένου, υπερεκπαιδευμένου και καλά γενικευμένου μοντέλου

3.2 ΕΙΣΑΓΩΓΗ ΣΤΑ RANDOM FORESTS

Η τεχνική μηχανικής μάθησης RF προτάθηκε για πρώτη φορά το 1995 από τον Tin Kam Ho και αναπτύχθηκε από τους Leo Breiman και Adele Culter, οι οποίοι, επίσης καταχώρισαν το "Random Forests" ως εμπορικό σήμα το 2006 [40]. Η τεχνική αυτή, είναι μια μέθοδος συλλογιστικής μάθησης (ensemble learning) που αξιοποιείται κυρίως για προβλήματα παλινδρόμησης και ταξινόμησης, όπως φαίνεται και στο Σχήμα 3.2, κατασκευάζοντας πληθώρα δέντρων αποφάσεων κατά τη διάρκεια της εκπαίδευσης [41].

- Η παλινδρόμηση με το RF χρησιμοποιείται όταν το προβλεπόμενο αποτέλεσμα είναι μια συνεχής τιμή. Συγκεκριμένα, αντί να ψηφίζει για μια κλάση, το κάθε δέντρο απόφασης προβλέπει μια αριθμητική τιμή και το τελικό αποτέλεσμα (τελική πρόβλεψη) υπολογίζεται μέσω του μέσου όρου αυτών των τιμών.
- Στην περίπτωση της ταξινόμησης το RF χρησιμοποιείται όταν το αναμενόμενο αποτέλεσμα είναι μια κατηγορία. Στην περίπτωση αυτή, κάθε δέντρο απόφασης στο δάσος ψηφίζει για μια κατηγορία, και έτσι η τελική πρόβλεψη καθορίζεται από την κατηγορία με τις περισσότερες ψήφους.



Σχήμα 3.2: Μοντέλο RF για προβλήματα ταξινόμησης/παλινδρόμησης

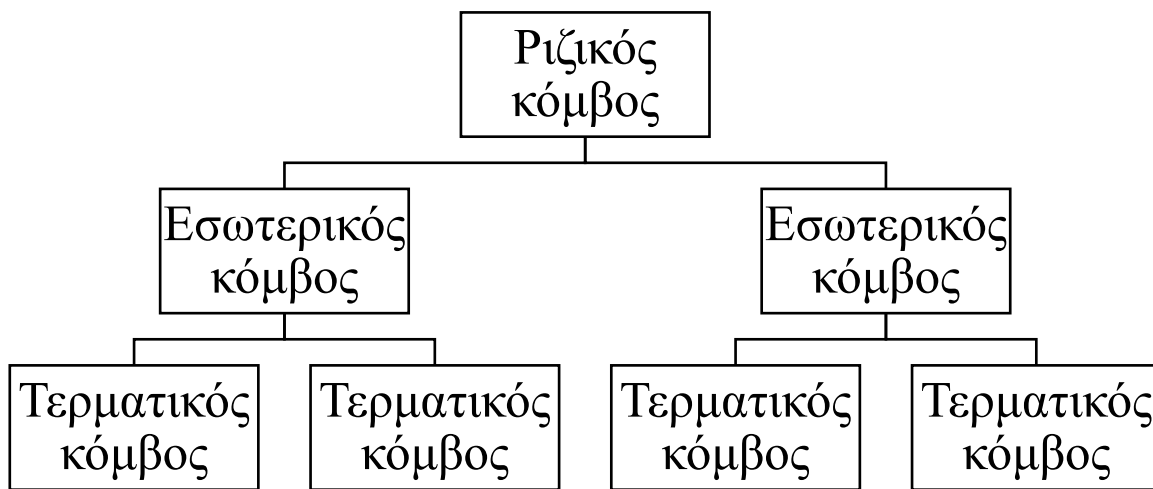
Οι σημαντικότεροι λόγοι για τους οποίους χρησιμοποιούνται τα μοντέλα RF είναι η δυνατότητά τους να χειρίζονται μεγάλα σύνολα δεδομένων χωρίς προηγούμενες υποθέσεις, η γρήγορη λειτουργία τους για τη διανομή των δεδομένων, η αποφυγή υπερεκπαίδευσης (overfitting) και η δυνατότητά τους να αντιμετωπίζουν τόσο συνεχείς όσο και διακριτές μεταβλητές. Επίσης, έχει αποδειχθεί ότι είναι ιδιαίτερα αποτελεσματικά στην πρόβλεψη της ηλιακής ακτινοβολίας και την μοντελοποίηση της απόδοσης των Φ/Β πάνελ [15].

3.2.1 Δέντρα Απόφασης

Το βασικό στοιχείο ενός μοντέλου RF είναι τα δέντρα αποφάσεων. Παράδειγμα ενός δέντρου απόφασης φαίνεται στο Σχήμα 3.3. Τα γενικά χαρακτηριστικά ενός δέντρου απόφασης (decision tree) είναι τα εξής [41]:

- Ξεκινά με έναν ριζικό κόμβο, ο οποίος δεν έχει κανένα εισερχόμενο κλάδο.
- Ένα δέντρο απόφασης θεωρείται σημαντικό όταν περιέχει τουλάχιστον τρεις κόμβους.
- Ένα δέντρο απόφασης θεωρείται ασήμαντο όταν έχει μόνο έναν κόμβο, δηλαδή τον ριζικό κόμβο.

- Κάθε άλλος κόμβος ενός δέντρου απόφασης έχει ακριβώς έναν γονικό κόμβο, δηλαδή είναι διάδοχος αποκλειστικά ενός άλλου κόμβου.
- Οι εσωτερικοί κόμβοι (internal nodes), γνωστοί και ως κόμβοι απόφασης/ελέγχου περιέχουν έναν κατάλληλο έλεγχο, που ονομάζεται έλεγχος διαίρεσης κόμβου, ο οποίος παράγει ακριβώς δύο διάδοχους κόμβους.
- Τερματικοί κόμβοι (leaf nodes) ονομάζονται αυτοί που δεν έχουν διάδοχους κόμβους και είναι αυτοί που ταξινομούν την περίπτωση που αναλύεται σε μία από τις δύο ή περισσότερες κλάσεις, ανάλογα με τον αριθμό των κλάσεων που υπάρχουν στα δεδομένα εκπαίδευσης.



Σχήμα 3.3: Δέντρο Απόφασης

3.2.2 Σύγκλιση των Random Forest

Στην ενότητα αυτή αναλύεται το θεωρητικό υπόβαθρο που αποδεικνύει την ικανότητα των RF να συγκλίνουν σε μία λύση και να μην υπερεκπαιδεύονται [42].

Δεδομένου ενός συνόλου ταξινομητών $\{h_1(x), h_2(x), \dots, h_K(x)\}$ και με το σύνολο εκπαίδευσης να επιλέγεται τυχαία από την κατανομή του τυχαίου διανύσματος (X, Y) , όπου X το διάνυσμα χαρακτηριστικών και Y η ετικέτα (label) της πρόβλεψης, ορίζεται η συνάρτηση περιθωρίου (margin function) ως εξής:

$$mg(X, Y) = av_K(I(h_K(X) = Y)) - \max_{j \neq Y} av_K(I(h_K(X) = j)) \quad (3.1)$$

όπου:

- $mg(X, Y)$: Η συνάρτηση περιθωρίου που μετρά τον βαθμό με τον οποίο ο μέσος όρος των ψήφων στο X και Y για τη σωστή κλάση υπερβαίνει τη μέση ψήφο για οποιαδήποτε άλλη κλάση. Όσο πιο μεγάλο είναι το περιθώριο, τόσο μεγαλύτερη είναι η εμπιστοσύνη στην κατηγοριοποίηση/ταξινόμηση (classification).
- $I(\dots)$: Αποτελεί την συνάρτηση δείκτη (indicator function), που επιστρέφει 1 αν η συνθήκη είναι αληθής αλλιώς 0.
- av_K : Ο μέσος όρος κατά μήκος όλων των ταξινομητών $\{h_1(x), h_2(x), \dots, h_K(x)\}$.
- $max_{j \neq Y}$: Ο μέγιστος όρος κατά μήκος όλων των δυνατών κλάσεων j , εξαιρώντας την σωστή κλάση.

Το σφάλμα γενίκευσης δίνεται από την:

$$PE^* = P_{X,Y}(mg(X, Y) < 0) \quad (3.2)$$

όπου:

- PE^* : το σφάλμα γενίκευσης.
- $P_{X,Y}$: Η κατανομή πιθανότητας του ζεύγους X, Y . Οι δείκτες X, Y δείχνουν ότι βρίσκεται πάνω στον χώρο X, Y .
- $mg(X, Y)$: Η συνάρτηση περιθωρίου.

Στα RF, επίσης ισχύει ότι :

$$h_K(X) = h(X, \theta_K) \quad (3.3)$$

όπου:

- $h_K(X)$: Η πρόβλεψη του ταξινομητή K για την μεταβλητή X .
- $h(X, \theta_K)$: Η πρόβλεψη ενός ταξινομητή για τη μεταβλητή X και τη παράμετρο θ_K .

Παράλληλα, για έναν μεγάλο αριθμό δέντρων, προκύπτει από τον Ισχυρό Νόμο των Μεγάλων Αριθμών και τη δομή του δέντρου το θεώρημα ότι, καθώς αυξάνεται ο αριθμός των δέντρων του τυχαίου δάσους, σχεδόν για όλες τις ακολουθίες $\theta_1, \theta_2, \dots$, η PE^* συγκλίνει στην:

$$PE^* = P_{X,Y} \left(P_\theta(h(X, \theta) = Y) - \max_{j \neq Y} P_\theta(h(X, \theta) = j) < 0 \right) \quad (3.4)$$

όπου:

- $P_\theta(h(X, \theta) = Y)$: Η πιθανότητα ότι η πρόβλεψη του ταξινομητή h για το δείγμα X , με την παράμετρο θ , είναι ίση με την πραγματική ετικέτα Y .
- $P_\theta(h(X, \theta) = j)$: Η πιθανότητα ότι η πρόβλεψη του ταξινομητή h για το δείγμα X , με την παράμετρο θ , είναι ίση με μια άλλη ετικέτα j .

Βάσει αυτού του αποτελέσματος εξηγείται γιατί τα τυχαία δάση δεν υπερπροσαρμόζονται καθώς αυξάνεται ο αριθμός των δέντρων και αντί αυτού, παράγουν μια περιοριστική τιμή του σφάλματος γενίκευσης. Έτσι, αντιμετωπίζεται το βασικό πρόβλημα των απλών δέντρων απόφασης που υπόκειται στον κίνδυνο υπερεκπαίδευσης καθώς τείνουν να προσαρμόζονται στα δείγματα εντός των δεδομένων εκπαίδευσης.

3.2.3 Τεχνική Bagging

Σημαντικό στοιχείο για την επίτευξη υψηλής απόδοσης στην τεχνική μηχανικής μάθησης RF είναι το bootstrap aggregating (bagging). Ο τρόπος με τον οποίο λειτουργεί αυτή η τεχνική είναι ο εξής [43]:

Με δεδομένο ένα σύνολο εκπαίδευσης $X = \{x_1, x_2, \dots, x_n\}$ που έχει τις αντιστοιχίσεις $Y = \{y_1, y_2, \dots, y_n\}$, η τεχνική bagging επιλέγει επαναλαμβανόμενα (B φορές) ένα διαφορετικό υποσύνολο των δεδομένων εκπαίδευσης με αντικατάσταση, και έπειτα εφαρμόζει δέντρα απόφασης για αυτά τα δείγματα. Συγκεκριμένα, για $b=1, 2, \dots, B$:

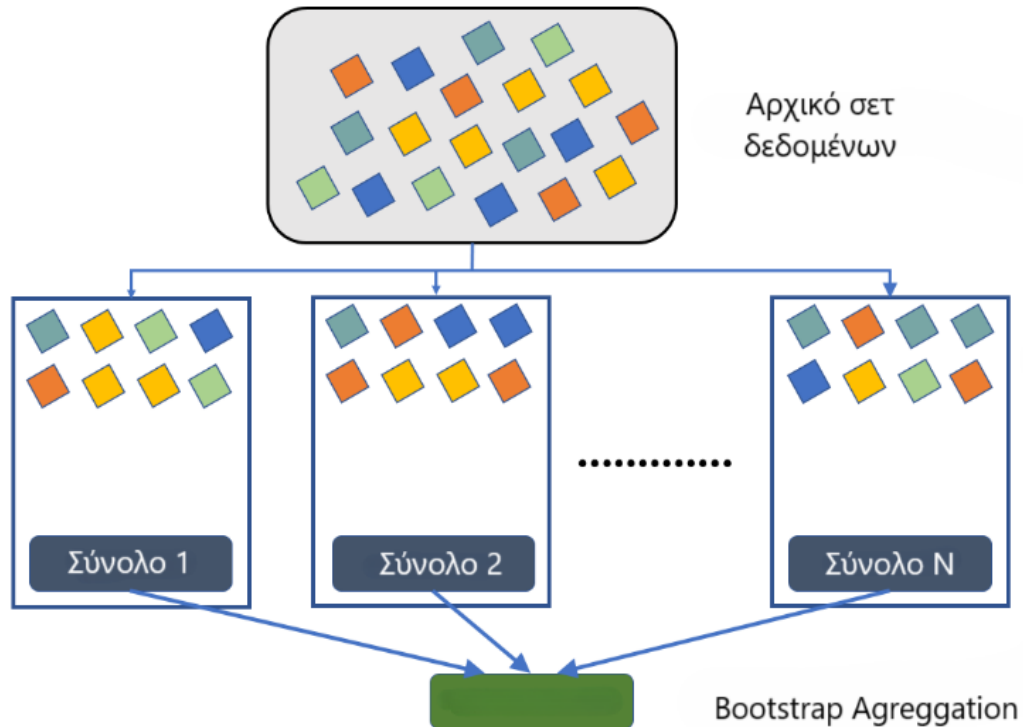
1. Δημιουργεί ένα υποσύνολο δεδομένων $\{X_b, Y_b\}$ από το $\{X, Y\}$ με αντικατάσταση, και το ίδιο υποσύνολο μπορεί να επαναληφθεί στην εκπαίδευση n φορές.
2. Εκπαιδεύει ένα δέντρο ταξινόμησης ή παλινδρόμησης f_b βάσει των $\{X_b, Y_b\}$.

Αφού ολοκληρωθεί η εκπαίδευση, στην περίπτωση προβλημάτων παλινδρόμησης, οι προβλέψεις για μη εισαγόμενα δείγματα x' , τα οποία δεν αξιοποιήθηκαν κατά την διάρκεια της εκπαίδευσης, γίνονται υπολογίζοντας τον μέσο όρο των προβλέψεων από όλα τα μεμονωμένα δέντρα παλινδρόμησης για το x' , με την χρήση της εξίσωσης []:

$$\hat{f} = \frac{1}{B} \cdot \sum_{b=1}^B f_b(x') \quad (3.5)$$

Στην περίπτωση προβλημάτων ταξινόμησης, οι προβλέψεις παράγονται παίρνοντας την ψήφο με την μεγαλύτερη πλειοψηφία. Παράδειγμα Bagging σε πρόβλημα ταξινόμησης φαίνεται στο Σχήμα 3.4.

Παράλληλα, τα μοντέλα RF περιλαμβάνουν μία ακόμη τεχνική bagging, την επονομαζόμενη feature bagging, η οποία επιλέγει σε κάθε υποψήφια διαίρεση κατά της διαδικασίας της μάθησης ένα τυχαίο υποσύνολο των χαρακτηριστικών εισόδου, αντί του συνόλου των χαρακτηριστικών. Ο λόγος που γίνεται αυτό είναι για να αποφευχθεί η ισχυρή συσχέτιση μεταξύ των δέντρων, καθώς εάν ένα ή περισσότερα χαρακτηριστικά εισόδου είναι έντονα συσχετισμένα με τη μεταβλητή εξόδου, αυτά τα χαρακτηριστικά θα επιλεγούν σε πολλά από τα B δέντρα.



Σχήμα 3.4: Παράδειγμα Bootstrap Aggregation (bagging)

Η διαδικασία αυτή, στο σύνολό της, οδηγεί σε καλύτερη απόδοση του μοντέλου καθώς μειώνει τη διακύμανση του μοντέλου χωρίς να αυξάνει την μεροληψία του. Αυτό σημαίνει ότι σε αντίθεση με τις προβλέψεις ενός μόνο δέντρου, ο μέσος όρος των προβλέψεων πολλών δέντρων δεν είναι ευαίσθητος στον θόρυβο του συνόλου εκπαίδευσης, αρκεί τα δέντρα να μην είναι συσχετισμένα. Επιπλέον, μέσα από τη χρήση των RF, μπορεί να υπολογιστεί μια εκτίμηση της αβεβαιότητας της πρόβλεψης, ως η τυπική απόκλιση των προβλέψεων από όλα τα μεμονωμένα δέντρα παλινδρόμησης στο x' :

$$\sigma = \sqrt{\frac{\sum_{b=1}^B (f_b(x') - \hat{f})^2}{B - 1}} \quad (3.6)$$

3.2.3 Υπερπαράμετροι μοντέλου

Σε ένα μοντέλο RF, η επιλογή των κατάλληλων υπερπαραμέτρων επηρεάζει σημαντικά την απόδοση του μοντέλου και την ικανότητά του να γενικεύει καλά σε νέα δεδομένα []. Οι υπερπαραμέτροι, επίσης γνωστοί ως παράμετροι ρύθμισης, πρέπει να βελτιστοποιηθούν προσεκτικά καθώς οι βέλτιστες τιμές τους εξαρτώνται από το σύνολο δεδομένων που διατίθεται και δεν μπορούν να εκτιμηθούν απευθείας από αυτό κατά τη διαδικασία εκπαίδευσης του μοντέλου.

Σύμφωνα με την βιβλιογραφία, οι υπερπαραμέτροι που επηρεάζουν σημαντικότερα την απόδοση ενός μοντέλου RF είναι οι εξής [44]:

- Πλήθος δέντρων: Η παράμετρος αυτή ελέγχει τον αριθμό των δέντρων μέσα στο δάσος. Ο αριθμός των δέντρων που απαιτούνται για να επιτευχθεί η βέλτιστη απόδοση, εξαρτάται από τις ιδιότητες του συνόλου δεδομένων και η χρήση πολλών δέντρων για την εκπαίδευση δεν βοηθάει πάντα στην παραγωγή μιας αποδοτικότερης λύσης. Παράλληλα, ενώ δεν προκαλείται υπερεκπαίδευση (overfitting), ο χρόνος υπολογισμού αυξάνεται γραμμικά με τον αριθμό των δέντρων. Έχει αποδειχθεί ότι για ένα μεγάλο σύνολο δεδομένων η μεγαλύτερη κερδοφορία στην απόδοση μπορεί συχνά να επιτυγχάνεται με την ανάπτυξη των πρώτων εκατό δέντρων [44].
- Ελάχιστος αριθμός δειγμάτων για διαχωρισμό (min samples split): Καθορίζει τον ελάχιστο αριθμό δειγμάτων που πρέπει να έχει ένας εσωτερικός κόμβος του δέντρου για να μπορεί να διαχωριστεί σε περισσότερους κόμβους. Η παράμετρος αυτή δεν πρέπει να έχει χαμηλή τιμή, καθώς έτσι το δέντρο θα αυξάνεται ραγδαία και ενδέχεται να υπερεκπαιδευτεί. Αυξάνοντας την τιμή του ελάχιστου αριθμού δειγμάτων για διαχωρισμό, μειώνεται ο συνολικός αριθμός των διαιρέσεων, περιορίζοντας έτσι τον αριθμό των συνολικών παραμέτρων, βοηθώντας έτσι στην μείωση της πιθανότητας υπερεκπαίδευσης. Παράλληλα, η τιμή δεν πρέπει να είναι πολύ μεγάλη, καθώς αυτό μπορεί να οδηγήσει σε παραγωγή ελάχιστων παραμέτρων και υποεκπαίδευση του μοντέλου.
- Μέγιστο βάθος: Ρυθμίζει το μέγιστο βάθος που μπορούν να φτάσουν τα δέντρα μέσα στο δάσος. Παίζει καθοριστικό ρόλο στην ακρίβεια του μοντέλου. Καθώς αυξάνεται το βάθος του δέντρου, η ακρίβεια του μοντέλου αυξάνεται, ωστόσο αφού το βάθος ξεπεράσει ένα συγκεκριμένο όριο, η ακρίβεια μειώνεται σταδιακά λόγω της υπερεκπαίδευσης του μοντέλου. Όταν δεν καθοριστεί τιμή σε αυτή την παράμετρο, οι κόμβοι μέσα στο δέντρο συνεχίζουν να αυξάνονται μέχρι όλα τα φύλλα να γίνουν τερματικά, δηλαδή να μην υπάρχουν άλλες διακλαδώσεις ή όλα τα φύλλα να περιέχουν λιγότερα δείγματα από τον ελάχιστο αριθμό δειγμάτων για διαχωρισμό.
- Ελάχιστος αριθμός δειγμάτων ανά φύλλο: Η παράμετρος αυτή καθορίζει τον ελάχιστο αριθμό δειγμάτων που πρέπει να υπάρχουν σε κάθε φύλλο του δέντρου. Το δέντρο σταματάει να διαιρείται όταν ο αριθμός των δειγμάτων σε ένα φύλλο είναι μικρότερος από την τιμή της παραμέτρου αυτής, ακόμα και αν θα βελτιωνόταν η απόδοση του μοντέλου με επιπλέον διαιρέσεις. Βοηθάει στην αποφυγή της υπερεκπαίδευσης και στη δημιουργία πιο γενικευμένων μοντέλων.
- Μέγιστος αριθμός δειγμάτων: Βοηθάει στη δημιουργία διαφορετικών δέντρων σε κάθε επανάληψη του δάσους, εφαρμόζοντας μια μορφή τυχαιότητας στην επιλογή των δειγμάτων που θα χρησιμοποιηθούν για εκπαίδευση σε κάθε δέντρο.
- Μέγιστος αριθμός χαρακτηριστικών: Για κάθε δέντρο, καθορίζει το μέγιστο αριθμό χαρακτηριστικών που θα ληφθούν υπόψη, δημιουργώντας έτσι τυχαία υποσύνολα των χαρακτηριστικών εισόδου για την εύρεση του βέλτιστου διαχωρισμού.

3.3 Αντικειμενικές συναρτήσεις

Οι αντικειμενικές συναρτήσεις είναι καθοριστικής σημασίας στην αξιολόγηση της απόδοσης ενός μοντέλου RF. Ο σκοπός των αντικειμενικών συναρτήσεων είναι να μετρήσουν τη διαφορά μεταξύ των παραγόμενων και των επιθυμητών εξόδων συνολικά, καθώς και να ελαχιστοποιήσουν αυτήν την απόκλιση. Αυτή η διαδικασία επιτρέπει τη βελτίωση της ακρίβειας και της γενίκευσης του μοντέλου, προσφέροντας έτσι ένα πιο αξιόπιστο και αποτελεσματικό σύστημα πρόβλεψης. Ορισμένες πολύ συχνά χρησιμοποιούμενες αντικειμενικές συναρτήσεις είναι οι εξής:

- Ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (Mean Squared Error – MSE):

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.7)$$

- Ελαχιστοποίηση της ρίζας του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error – RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.8)$$

- Ελαχιστοποίηση του μέσου απόλυτου σφάλματος (Mean Absolute Error – MAE):

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.9)$$

- Ελαχιστοποίηση του μέσου ποσοστιαίου απόλυτου σφάλματος (Mean Absolute Percentage Error – MAPE):

$$MAPE = \frac{100}{n} \cdot \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3.10)$$

όπου:

1. n : το πλήθος των δεδομένων εκπαίδευσης
2. y_i : η παραγόμενη έξοδος για το δείγμα i
3. \hat{y}_i : η επιθυμητή έξοδος για το δείγμα i .

3.4 ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΥΠΕΡΠΑΡΑΜΕΤΡΩΝ RANDOM FOREST

Η διαδικασία βελτιστοποίησης των υπερπαραμέτρων ενός μοντέλου RF είναι μια εμπειρική διαδικασία με στόχο την καλύτερη προσαρμογή στο τοπίο των δεδομένων που παρέχονται και την δημιουργία πιο αποδοτικών και αξιόπιστων μοντέλων. Για την αποφυγή επιλογής μη βέλτιστων τιμών παραμέτρων χρησιμοποιούνται διάφορες μέθοδοι, όπως:

- Διασταυρούμενη επικύρωση (cross validation)
- Αναζήτηση πλέγματος (grid search)
- Τυχαία αναζήτηση (random search)
- Χρήση εξελικτικών / μεταερευτικών αλγορίθμων

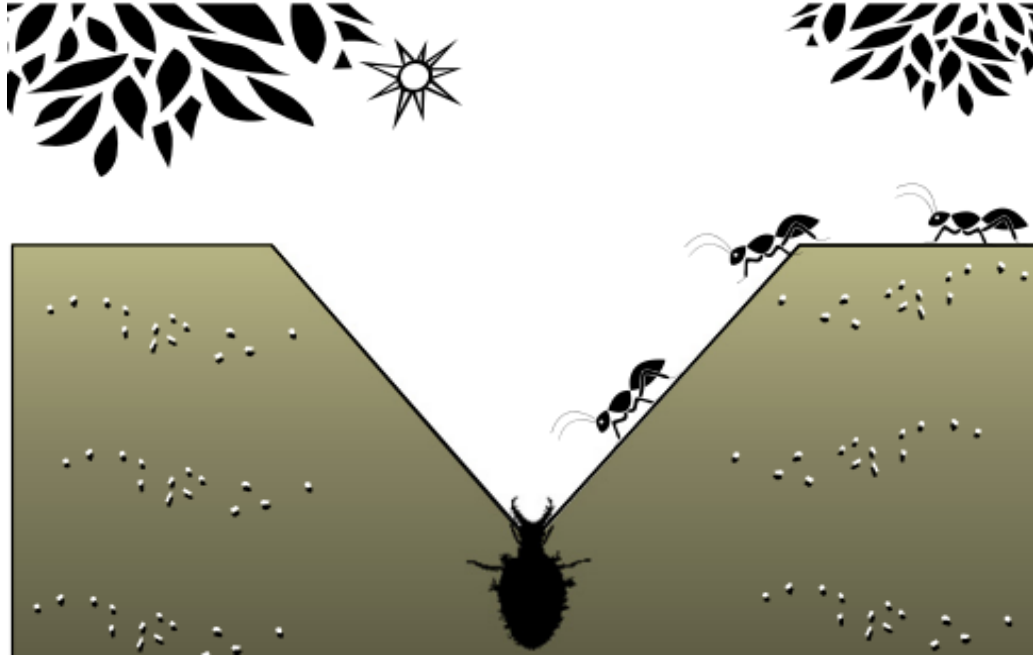
Από αυτό το σύνολο μεθόδων, οι διαδικασίες με τη μεγαλύτερη αξιοπιστία είναι η διασταυρούμενη επικύρωση και η αναζήτηση πλέγματος, γιατί δοκιμάζουν σχεδόν όλους τους δυνατούς συνδυασμούς, και στην περίπτωση της διασταυρούμενης επικύρωσης, με διαφορετικά υποσύνολα εκπαίδευσης. Η απλούστερη μέθοδος είναι η τυχαία αναζήτηση, ενώ οι μεταερευτικοί αλγόριθμοι διαθέτουν τη μεγαλύτερη ευελιξία, ωστόσο δεν μπορούν πάντα να εγγυηθούν πως θα βρουν τον βέλτιστο συνδυασμό καθώς είναι στοχαστικοί.

Στην περίπτωση όπου δεν χρειάζονται πολλές επανεκπαιδεύσεις ή είναι απαραίτητη η εύρεση του ολικού βέλτιστου, αξιοποιούνται οι εξαντλητικές διαδικασίες αναζήτησης, δηλαδή η διασταυρούμενη επικύρωση και η αναζήτηση πλέγματος. Σε περιπτώσεις όπου είναι απαραίτητη η συχνή επανεκπαίδευση, οι μεταερευτικοί αλγόριθμοι αποτελούν την προτιμότερη διαδικασία. Επίσης, οι μεταερευτικοί αλγόριθμοι είναι ιδιαίτερα καλοί στην αντιμετώπιση πολυδιάστατων χώρων υπερπαραμέτρων, αφού είναι σχεδιασμένοι να ερευνούν αυτούς τους χώρους αποτελεσματικά, χρησιμοποιώντας ευέλικτες τεχνικές, όπως την επιλογή (selection), τη διασταύρωση (crossover) και τη μετάλλαξη (mutation) [9]. Σε σχέση με τους αλγορίθμους που βασίζονται στη μείωση της κλίσης (gradient-descent algorithms), οι εξελικτικοί/μεταερευτικοί αλγόριθμοι είναι αποτελεσματικότεροι για την εύρεση του ολικού βέλτιστου, καθώς αποφεύγουν τα τοπικά βέλτιστα και επιτυγχάνουν την εξερεύνηση του χώρου υπερπαραμέτρων πιο ολιστικά. Τέλος, οι εξελικτικοί/μεταερευτικοί αλγόριθμοι προσφέρουν αυτοματοποιημένη διαδικασία βελτιστοποίησης, σε αντίθεση με τη διασταυρούμενη επικύρωση, την τυχαία αναζήτηση και την αναζήτηση πλέγματος, ελαχιστοποιώντας την ανάγκη για χειροκίνητη επίβλεψη, κάτι που τους καθιστά πιο αποδοτικούς.

3.5 ΑΛΓΟΡΙΘΜΟΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ALO

Ο αλγόριθμος βελτιστοποίησης ALO είναι ένας σύγχρονος μεταερευτικός αλγόριθμος που στηρίζεται σε πληθυσμούς και δημιουργήθηκε το 2015 [10]. Πρόκειται για αλγόριθμο που ανήκει στην κατηγορία αυτών που αντλούν την έμπνευσή τους από κάποιο φυσικό φαινόμενο. Το φυσικό φαινόμενο

που προσομοιάζει, είναι η συμπεριφορά των μυρμηγκιών και των λιονταροσκορπιών (αντίλια) στην φύση, συγκεκριμένα μοντελοποιεί τα μυρμήγκια που δημιουργούν μονοπάτια καθώς αναζητούν τροφή, και τα αντίλια που τα παγιδεύουν αυτά στις λάκκους άμμου, όπως φαίνεται στο Σχήμα 3.5.



Σχήμα 3.5: Αντίλιο που παγιδεύει μυρμήγκια στο λάκκο άμμου

Στον αλγόριθμο ALO τα μυρμήγκια αποτελούν τους πράκτορες αναζήτησης και είναι ελεύθερα να κινηθούν στον χώρο αναζήτησης. Αντίστοιχα, στον αλγόριθμο, τα αντίλια αντιπροσωπεύουν τις λύσεις στο πρόβλημα βελτιστοποίησης. Σε ένα πρόβλημα βελτιστοποίησης με N μεταβλητές απόφασης, οι θέσεις των μυρμηγκιών/αντιλιών σε έναν χώρο N διαστάσεων αποτελούν τις μεταβλητές απόφασης []. Καθώς ο αλγόριθμος εξελίσσεται, τα μυρμήγκια προσεγγίζουν σταδιακά τις βέλτιστες λύσεις γύρω από τις θέσεις των αντιλιών, ενώ αντίστοιχα τα αντίλια προσαρμόζουν τις θέσεις τους ανάλογα με το πως εξελίσσεται η αναζήτηση των μυρμηγκιών. Η συνεργατική αυτή αναζήτηση μεταξύ μυρμηγκιών και αντιλιών εξασφαλίζει την αποτελεσματική εξερεύνηση του χώρου και τη σύγκλιση του αλγορίθμου.

Η κίνηση των μυρμηγκιών έχει στοχαστικό χαρακτήρα και μοντελοποιείται με την χρήση ενός τυχαίου περιπάτου (random walk). Σκοπός των μυρμηγκιών είναι να αποφύγουν τις παγίδες των αντιλιών και να αξιολογηθεί η θέση τους χρησιμοποιώντας μια αντικειμενική συνάρτηση. Ο τυχαίος περίπατος ενός μυρμηγκιού υπολογίζεται χρησιμοποιώντας την ακόλουθη εξίσωση:

$$X(t) = [0, \text{cumsum}(2r(t_1) - 1), \text{cumsum}(2r(t_2) - 1), \dots, \text{cumsum}(2r(t_n) - 1)] \quad (3.11)$$

όπου:

- $\text{cumsum}(x)$: συνάρτηση σωρευτικού αθροίσματος
- n : ο μέγιστος αριθμός επαναλήψεων του αλγορίθμου ALO
- $r(t)$: τυχαία συνάρτηση δειγματοληψίας στο εύρος $[0,1]$ που περιγράφεται από την εξίσωση (3.12):

$$r(t) = \begin{cases} 0, & t < 0.5 \\ 1, & t \geq 0.5 \end{cases} \quad (3.12)$$

Οι θέσεις που παίρνει το μυρμήγκι κατά την διάρκεια του τυχαίου περιπάτου αποθηκεύονται σε ένα πίνακα για να μπορέσουν να αξιοποιηθούν στη συνέχεια:

$$M_{ant,i} = \begin{pmatrix} A_{1,1} & \cdots & A_{1,d} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \cdots & A_{n,d} \end{pmatrix}, i \leq N_{ants} \quad (3.13)$$

όπου:

- $M_{ant,i}$: Ο πίνακας που χρησιμοποιείται για την αποθήκευση της θέσης του μυρμηγκιού i σε κάθε επανάληψη
- n : ο μέγιστος αριθμός επαναλήψεων του αλγόριθμου ALO
- d : ο αριθμός των μεταβλητών απόφασης
- N_{ants} : ο αριθμός των μυρμηγκιών

Στόχος των αντίλων είναι να αιχμαλωτίσουν τα μυρμήγκια. Για να το επιτύχουν, δημιουργούν λάκκους, οι οποίοι λειτουργούν ως παγίδες, στις οποίες τα μυρμήγκια μπορούν να πέσουν. Με αυτόν τον τρόπο, τα αντίλια επηρεάζουν την τυχαία περιπλάνηση των μυρμηγκιών. Συγκεκριμένα, κάθε μυρμήγκι κινείται σε μια υπερσφαίρα γύρω από την παγίδα ενός αντίλιου. Το αντίλιο επιλέγεται με την χρήση τελεστή ρουλέτας (roulette wheel - fitness proportionate selection), μια τυχαία διαδικασία όπου το αντίλιο που ικανοποιεί περισσότερο την αντικειμενική συνάρτηση είναι πιο πιθανό να επιλεγεί.

Παράλληλα, οι τυχαίοι περίπατοι κανονικοποιούνται με τη μέθοδο μεγίστου-ελαχίστου (min-max normalization [45]), ώστε να διατηρηθούν εντός του χώρου αναζήτησης. Η τυχαία περιπλάνηση των μυρμηγκιών, μοντελοποιημένη για να συμπεριλαμβάνει την επίδραση των παγίδων των αντίλιων, περιγράφεται από την εξίσωση (3.14):

$$Z_n(It) = \frac{(X_n(It) - a_n) \cdot (d_n(It) - c_n(It))}{b_n - a_n} \quad (3.14)$$

όπου:

- $Z_n(It)$: Η κανονικοποιημένη θέση του τυχαίου περιπάτου της $n^{\text{ης}}$ μεταβλητής απόφασης στην It επανάληψη του αλγορίθμου ALO
- $X_n(It)$: Η θέση του τυχαίου περιπάτου της μεταβλητής απόφασης n στην It επανάληψη πριν την κανονικοποίηση
- a_n : Η ελάχιστη τιμή σε ολόκληρο τον τυχαίο περίπατο για την μεταβλητή απόφασης n .
- b_n : Η μέγιστη τιμή σε ολόκληρο τον τυχαίο περίπατο για την μεταβλητή απόφασης n .
- $d_n(It)$: Η μέγιστη τιμή της μεταβλητής απόφασης n στην It επανάληψη.

- $c_n(It)$: Η ελάχιστη τιμή της μεταβλητής απόφασης n στην It επανάληψη.

Τα $d_n(It)$, $c_n(It)$ δίνονται από τις εξισώσεις (3.15) και (3.16):

$$d_n(It) = SelectedAntlion(It) + d(It) \quad (3.15)$$

$$c_n(It) = SelectedAntlion(It) + c(It) \quad (3.16)$$

όπου $d(It)$, $c(It)$, η μέγιστη και ελάχιστη τιμή του εύρους αναζήτησης στην It επανάληψη.

Όταν ένα μυρμήγκι πέσει στην παγίδα, το αντίλιο αρχίζει να ρίχνει άμμο προς το κέντρο του λάκκου για να αποτρέψει το μυρμήγκι από το να ξεφύγει. Αυτή η συμπεριφορά προσομοιώνεται μαθηματικά με τη συρρίκνωση της ακτίνας του τυχαίου περιπάτου του μυρμηγκιού. Έτσι, το εύρος των ορίων για όλες τις μεταβλητές απόφασης μειώνεται, όπως φαίνεται στις εξισώσεις (3.17) και (3.18).

$$c(It) = \frac{c(It)}{R} \quad (3.17)$$

$$d(It) = \frac{d(It)}{R} \quad (3.18)$$

όπου η R καθορίζεται βάσει της εξίσωσης:

$$R = 10^w \cdot \frac{It}{IT} \quad (3.19)$$

Στην εξίσωση (3.19), IT είναι ο μέγιστος αριθμός επαναλήψεων του αλγόριθμου ALO και w μια σταθερά που υπολογίζεται με βάση την εξίσωση (3.20):

$$w = \begin{cases} 2, It > 0.1IT \\ 3, It > 0.5IT \\ 4, It > 0.75IT \\ 5, It > 0.9IT \\ 6, It > 0.95IT \end{cases} \quad (3.20)$$

Σημαντικό χαρακτηριστικό των εξελικτικών αλγορίθμων είναι η προνομιούχα/ελιτιστική αντιμετώπιση (elitism), δηλαδή η διατήρηση της καλύτερης εξερευνημένης λύσης σε όλη την διαδικασία βελτιστοποίησης. Στην περίπτωση του ALO, ως καλύτερη λύση θεωρείται το αντίλιο που ικανοποιεί περισσότερο την αντικειμενική συνάρτηση κατά την τρέχουσα επανάληψη. Η τυχαία περιπλάνηση κάθε μυρμηγκιού επηρεάζεται ταυτόχρονα από το επιλεγμένο αντίλιο που αντιστοιχεί στο μυρμήγκι και από το αντίλιο που θεωρείται η τρέχουσα καλύτερη λύση. Ως εκ τούτου, οι τυχαίοι περίπατοι των μυρμηγκιών μοντελοποιούνται έτσι ώστε να συμπεριλαμβάνουν την επίδραση και των δύο παγίδων αυτών των αντίλιων.

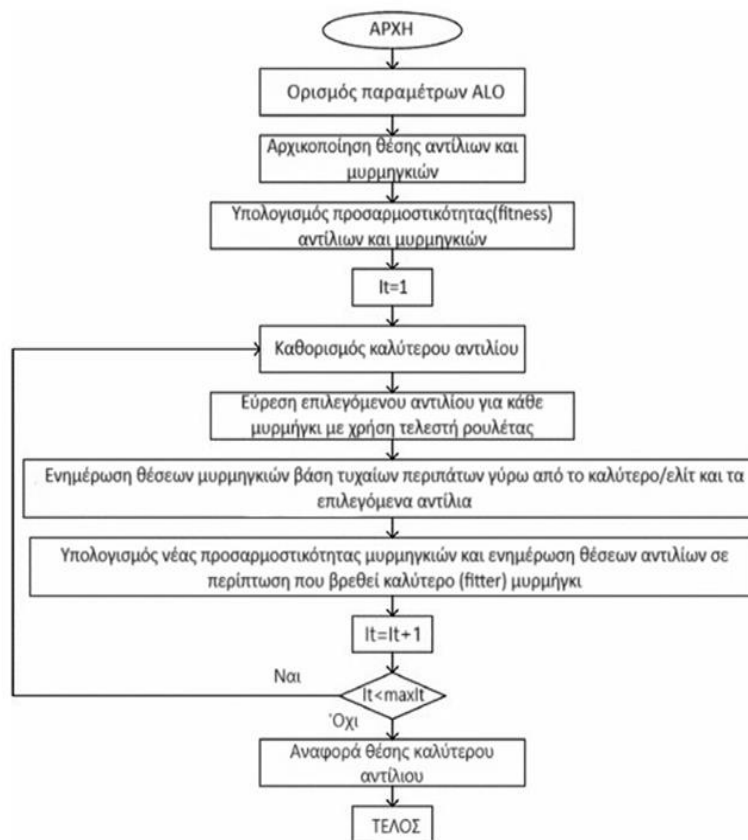
Οι νέες θέσεις που παίρνουν τα μυρμήγκια σε κάθε επανάληψη του αλγορίθμου, υπολογίζονται ως εξής:

$$Ant_i(It) = \frac{Z_e(It) + Z_m(It)}{2} \quad (3.20)$$

όπου:

- $Ant_i(It)$: Η θέση του i μυρμηγκιού κατά τη It επανάληψη
- $Z_e(It)$: Κανονικοποιημένος τυχαίος περίπατος γύρω από το επιλεγόμενο αντίλιο στην It επανάληψη
- $Z_m(It)$: Κανονικοποιημένος τυχαίος περίπατος γύρω από το καλύτερο/ελίτ αντίλιο έως την It επανάληψη

Τελευταίο στάδιο του αλγορίθμου ALO είναι η στιγμή που το μυρμήγκι φτάνει στον πάτο του λάκκου και καταναλώνεται από το αντίλιο. Αυτό συμβαίνει όταν το μυρμήγκι ικανοποιήσει περισσότερο την αντικειμενική συνάρτηση σε σχέση με το αντίστοιχο αντίλιο του. Σε αυτήν την κατάσταση, το αντίλιο αλλάζει τη θέση του με αυτή του παγιδευμένου μυρμηγκιού.



Σχήμα 3.6: Διάγραμμα ροής αλγορίθμου ALO

Η διαδικασία του αλγορίθμου ALO επαναλαμβάνεται έως ότου ικανοποιηθεί το κριτήριο τερματισμού του αλγορίθμου, το οποίο είναι η συμπλήρωση του μέγιστου αριθμού επαναλήψεων. Η τελική θέση και τιμή προσαρμοστικότητας του καλύτερου/ελίτ αντιλίου αποτελεί και τη λύση του προβλήματος. Το διάγραμμα ροής του αλγορίθμου ALO απεικονίζεται στο Σχήμα 3.6.

Ο εξελικτικός αλγόριθμος βελτιστοποίησης ALO χρησιμοποιείται όλο και συχνότερα τα τελευταία χρόνια, κυρίως σε προβλήματα βελτιστοποίησης της διαχείρισης ενέργειας σε συστήματα παραγωγής, λόγω των πολλών πλεονεκτημάτων του [10]. Τα σημαντικότερα πλεονεκτήματα του ALO είναι πως μπορεί πολύ εύκολα να συνδυαστεί με άλλους αλγορίθμους βελτιστοποίησης, έχει γρήγορη ταχύτητα σύγκλισης και προσφέρει τη δυνατότητα εύρεσης εξαιρετικών λύσεων σε πολύ λίγο χρόνο. Επίσης, είναι κατάλληλος για πολλούς τύπους προβλημάτων βελτιστοποίησης και είναι ικανός να ανιχνεύσει λύσεις σε ανισότροπους χώρους, είτε διακριτών ή συνεχών μεταβλητών. Πέρα από αυτά, είναι απλός στην υλοποίηση, έχει λίγες παραμέτρους προσαρμογής και δεν απαιτεί αρχικές λύσεις στα προβλήματα που προσπαθεί να επιλύσει. Τέλος, ο αλγόριθμος ALO είναι εξαιρετικά αποδοτικός στην εύρεση του ολικού μέγιστου, καθώς είναι χαμηλή η πιθανότητα παγίδευσής του σε κάποιο τοπικό βέλτιστο.

ΑΝΑΛΥΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΦΩΤΟΒΟΛΤΑΪΚΩΝ ΔΕΔΟΜΕΝΩΝ

4.1 ΔΕΔΟΜΕΝΑ ΦΩΤΟΒΟΛΤΑΪΚΗΣ ΠΑΡΑΓΩΓΗΣ

Η καταγραφή δεδομένων σε πραγματικό χρόνο κοντά σε Φ/Β συστήματα, δηλαδή σε απόσταση που επιτρέπει τη συσχέτιση των μετρήσεων με την απόδοση των Φ/Β πάνελ, είναι ζωτικής σημασίας για τη βελτιστοποίηση και την μακροχρόνια αξιοπιστία τους. Αυτή η διαδικασία επιτρέπει την άμεση ανάλυση και αξιολόγηση των παραμέτρων που επηρεάζουν την απόδοση των Φ/Β συστημάτων και έτσι παίζει τεράστιο ρόλο στην εξομάλυνση των μεταβολών της παραγωγής, στις βραχυπρόθεσμες συναλλαγές, στη διαχείριση της ζήτησης, στην εφεδρική παραγωγή και στη διαχείριση της αποθήκευσης [2]. Για να επιτευχθούν αυτά, τα δεδομένα πρέπει να συλλέγονται σε υψηλές χρονικές αναλύσεις ώστε να διευκολύνεται η λεπτομερής μοντελοποίηση της Φ/Β παραγωγής, γεγονός που αυξάνει την υπολογιστική πολυπλοκότητα.

Τα όργανα που χρησιμοποιούνται για την καταγραφή αυτών των δεδομένων περιλαμβάνουν τα πυρανόμετρα για τη μέτρηση της ηλιακής ακτινοβολίας, αισθητήρες θερμοκρασίας για τη μέτρηση της θερμοκρασίας των πάνελ και του περιβάλλοντος και ανεμόμετρα για την καταγραφή της ταχύτητας και της διεύθυνσης του ανέμου. Πέρα από αυτά, αξιοποιούνται κάμερες για τη λήψη φωτογραφιών ευρείας κλίμακας και την παρακολούθηση των πάνελ και του περιβάλλοντος χώρου. Η εγγύτητα αυτών των οργάνων (Σχήμα 4.1), εξασφαλίζει ότι οι μετρήσεις είναι αντιπροσωπευτικές των πραγματικών συνθηκών λειτουργίας των Φ/Β πάνελ.



Σχήμα 4.1: Παράδειγμα οργάνων καταγραφής δεδομένων σε Φ/Β πάρκο

Από αυτούς τους παράγοντες, κάποιοι είναι σταθεροί στη διάρκεια του χρόνου ή διαφοροποιούνται με πολύ αργό ρυθμό (κατασκευαστικοί παράγοντες), ενώ άλλοι παράγοντες μεταβάλλονται στη διάρκεια του χρόνου (περιβαλλοντικοί και μετεωρολογικοί παράγοντες). Αν και η Φ/Β παραγωγή εξαρτάται από όλους τους παραπάνω παράγοντες, η μεταβολή της Φ/Β παραγωγής εξαρτάται μόνο από τους παράγοντες που μεταβάλλονται με τον χρόνο, καθώς η συνεισφορά των υπόλοιπων παραγόντων είναι σταθερή. Επομένως, για την δημιουργία μίας ιστορικής χρονοσειράς και την αποτύπωση μοτίβων με σκοπό την πρόβλεψη της Φ/Β παραγωγής, γίνεται χρήση μόνο των παραγόντων που μεταβάλλονται με τον χρόνο.

Στην περίπτωση της πρόβλεψης Φ/Β παραγωγής με χρήση μηχανικής μάθησης, τα δεδομένα της Φ/Β παραγωγής χρησιμοποιούνται για την εκπαίδευση και την αξιολόγηση του ίδιου του μοντέλου. Αυτό σημαίνει ότι η ποσότητα και η ποιότητα αυτών των δεδομένων είναι καθοριστικής σημασίας για την ακρίβεια και την αξιοπιστία των παραγόμενων αποτελεσμάτων.

4.2 ΑΡΙΘΜΗΤΙΚΑ ΔΕΔΟΜΕΝΑ ΦΩΤΟΒΟΛΤΑΪΚΟΥ ΠΑΡΚΟΥ ΣΤΗΝ ΕΛΛΑΔΑ

Τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία για την εκπαίδευση και τον έλεγχο του μοντέλου RF σε συνδυασμό με τον αλγόριθμο ALO αφορούν ένα φωτοβολταϊκό πάρκο στην Ελλάδα με ονομαστική ισχύ 13 MW. Σε ότι αφορά την πρόβλεψη της Φ/Β παραγωγής, η χρονοσειρά δεδομένων αποτελείται από τις εξής μεταβλητές:

- Χρονική σήμανση (Timestamp)
- Θερμοκρασία περιβάλλοντος (°C)
- Θερμοκρασία πάνελ (°C)
- Οριζόντια ακτινοβολία (W/m²)
- Ακτινοβολία με κλίση (W/m²)
- Σχετική υγρασία (%)
- Διεύθυνση ανέμου (°)
- Ταχύτητα ανέμου (m/s)
- Παραγόμενη ισχύς (kW)

Η χρονική ανάλυση των χρονοσειρών δεδομένων είναι ένα (1) λεπτό, επομένως για κάθε κατηγορία δεδομένων αντιστοιχούν 1.440 μετρήσεις ανά ημέρα. Τα δεδομένα αφορούν περίοδο 45 ημερών, από την 20^η Οκτωβρίου του 2022, έως και την 2^η Νοεμβρίου του 2022.

4.3 ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Προκειμένου το μοντέλο πρόβλεψης Φ/Β παραγωγής να εκπαιδευθεί σωστά και αποδοτικά για την επίτευξη των καλύτερων δυνατών προβλέψεων, απαιτούνται κατάλληλα δεδομένα. Αυτό περιλαμβάνει την επιλογή δεδομένων που να είναι αντιπροσωπευτικά του προβλήματος, να μην περιέχουν ακραίες ή κενές τιμές και να υπάρχει ένα επαρκές δείγμα. Για τον λόγο αυτό, για να εξασφαλιστούν τα καλύτερα δυνατά αποτελέσματα, τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία υπέστησαν μία σειρά απαραίτητων διαδικασιών εκκαθάρισης. Συγκεκριμένα, υλοποιήθηκαν τα ακόλουθα βήματα:

- Αφαίρεση χρονικών στιγμών όπου, ενώ υπήρχε επαρκής ηλιακή ακτινοβολία, η παραγόμενη ισχύς του Φ/Β σταθμού ήταν σχεδόν μηδενική.
- Αφαίρεση των χρονικών στιγμών από το σύνολο δεδομένων όταν κάποιες ή όλες από τις μεταβλητές εισόδου της χρονικής στιγμής ήταν μηδενικές.

Μετά τη διαδικασία εκκαθάρισης των δεδομένων, έγινε αναπλήρωση των διαγραμμένων τιμών, με τον εξής τρόπο:

- Οπουδήποτε υπήρχε κενό κελί, αυτό συμπληρώθηκε με την αμέσως προηγούμενη μετρούμενη τιμή του αντίστοιχου μεγέθους, έτσι ώστε να υπάρχει πληρότητα και ομοιομορφία στα δεδομένα εισόδου.

Αυτή η μέθοδος είναι αξιόπιστη λόγω της υψηλής χρονικής ανάλυσης των δεδομένων. Σε άλλες περιπτώσεις, π.χ. αν τα δεδομένα μετρούνταν κάθε 1 ώρα, τότε αυτή η μέθοδος θα ήταν αναξιόπιστη.

Προκειμένου να διασφαλιστεί η εγκυρότητα και η ακρίβεια των αποτελεσμάτων, από το συνολικό σύνολο δεδομένων των 45 ημερών αποφασίστηκε να χρησιμοποιηθούν μόνο 43, καθώς δύο από αυτές ήταν ελλιπείς σε πολλά σημεία.

4.3.1 Χρονική προετοιμασία δεδομένων

Για να μπορέσουν να χρησιμοποιηθούν τα δεδομένα που θα τροφοδοτηθούν στο μοντέλο RF για την πρόβλεψη της Φ/Β παραγωγής, πρέπει οι χρονοσειρές δεδομένων να διατυπωθούν ως προβλήματα επιτηρούμενης μάθησης [39]. Η επιτηρούμενη μάθηση απαιτεί ετικέτες για τα δεδομένα εισόδου. Στην πρόβλεψη, αυτές οι ετικέτες αντιστοιχούν στις τιμές που είναι επιθυμητό να προβλεφθούν για κάθε σύνολο δεδομένων εισόδου. Αφού πρόκειται για χρονοσειρές, μπορούν να προσδοθούν αυτές οι ετικέτες κάνοντας μια μετατόπιση της χρονοσειράς στόχου (Φ/Β παραγωγή) στον άξονα του χρόνου. Στην παρούσα διπλωματική εργασία, οι προβλέψεις της Φ/Β παραγωγής έχουν παραχθεί με χρονικό ορίζοντα πρόβλεψης 15 λεπτών ενώ οι μετρήσεις παρέχονται σε ανάλυση 1 λεπτού, έτσι η μετατόπιση που έγινε στη χρονοσειρά Φ/Β παραγωγής είναι 15 θέσεων. Ένα απλό παράδειγμα αυτής της μετατροπής για πρόβλεψη ενός βήματος μπροστά (1 step ahead forecasting) φαίνεται στο Σχήμα 4.2. Όπως φαίνεται και στο Σχήμα 4.2, η πρώτη και τελευταία σειρά δεδομένων δεν μπορεί να αξιοποιηθεί για την εκπαίδευση του μοντέλου. Με αυτόν τον τρόπο, το μοντέλο RF επιλύει ένα πρόβλημα παλινδρόμησης, όπου ο στόχος είναι να προβλεφθεί η επόμενη τιμή της μεταβλητής εξόδου με βάση τις προηγούμενες τιμές.

time, measure	Input, Output
1, 100	?, 100
2, 110	100, 110
3, 108	110, 108
4, 115	108, 115
5, 120	115, 120
	120, ?

Σχήμα 4.2: Μετατροπή δεδομένων χρονοσειράς σε δεδομένα επιβλεπόμενης μάθησης για πρόβλεψη ενός βήματος μπροστά (1 step ahead forecast)

4.3.2 Κανονικοποίηση δεδομένων

Ακόμη μία σημαντική και συνήθης διαδικασία προ-επεξεργασίας των δεδομένων είναι η κανονικοποίηση των δεδομένων εισόδου και εξόδου. Η κανονικοποίηση στον τομέα της μηχανικής μάθησης αναφέρεται στη διαδικασία επεξεργασίας των δεδομένων με στόχο την προσαρμογή τους σε ένα νέο εύρος τιμών, διατηρώντας παράλληλα τη μορφή της κατανομής τους. Με αυτή τη διαδικασία εξασφαλίζεται ότι δεν θα υπάρχουν μεγάλες αποκλίσεις στην τάξη μεγέθους των δεδομένων, ενώ επίσης μειώνεται η επιρροή των ακραίων τιμών. Συνήθως επιλέγεται η κανονικοποίηση στο διάστημα $[-1, 1]$ ή στο διάστημα $[0, 1]$.

Στην παρούσα διπλωματική εργασία, τα δεδομένα κανονικοποιήθηκαν στο διάστημα $[0, 1]$ με τη μέθοδο κανονικοποίησης μεγίστου – ελαχίστου, η οποία βασίζεται στην εξίσωση (4.1) [45]:

$$x_{i,scaled} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (4.1)$$

όπου:

- $x_{i,scaled}$: Η κανονικοποιημένη τιμή του δεδομένου i
- x_i : Η τιμή του δεδομένου i πριν την κανονικοποίηση
- x_{min} : Η ελάχιστη τιμή της κατηγορίας δεδομένων στην οποία ανήκει το δεδομένο i
- x_{max} : Η μέγιστη τιμή της κατηγορίας δεδομένων στην οποία ανήκει το δεδομένο i

4.4 ΔΕΔΟΜΕΝΑ ΕΠΙΓΕΙΩΝ ΕΙΚΟΝΩΝ ΤΟΥ ΟΥΡΑΝΟΥ

Τα σύννεφα έχουν σημαντική επίδραση στην Φ/Β παραγωγή ηλεκτρικής ενέργειας. Κατά τη διάρκεια της ημέρας, η μεταβαλλόμενη κατάσταση του ουρανού μπορεί να προκαλέσει ραγδαίες μεταβολές στην ένταση της ηλιακής ακτινοβολίας που φτάνει στα Φ/Β συστήματα, με αποτέλεσμα τη μείωση της παραγωγής ηλεκτρικής ενέργειας [12]. Οι απρόβλεπτες διακυμάνσεις επηρεάζουν την αποδοτικότητα και

τη σταθερότητα του συστήματος, καθιστώντας απαραίτητη την ανάπτυξη τεχνικών πρόβλεψης της κίνησης των νεφών για την ορθή υπερβραχυπρόθεσμη πρόβλεψη της Φ/Β παραγωγής ενέργειας.

Για αυτούς τους υπερβραχυπρόθεσμους χρονικούς ορίζοντες, όπου απαιτείται η γνώση της κατάστασης του ουρανού σε πραγματικό χρόνο, η χρήση αριθμητικών προβλέψεων καιρού και η εικονογράφηση με βάση δορυφορικές παρατηρήσεις είναι ακατάλληλη. Λύση σε αυτό το ζήτημα δίνουν οι επίγειες εικόνες του ουρανού με χρήση ψηφιακών καμερών, εξοπλισμένων με ευρυγώνιους φακούς προβολής 180° (fish-eyed lens) για να αποτυπώνουν ολόκληρο τον ουράνιο θόλο. Με βάση την ανάλυση αυτών των εικόνων, μια εικόνα της μελλοντικής κατάστασης του ουρανού μπορεί να αποτυπωθεί με μεγάλη λεπτομέρεια και υπο-χιλιομετρική ακρίβεια, δηλαδή ανάλυση που αφορά αποστάσεις μικρότερες του ενός χιλιομέτρου. Οι επίγειες εικόνες του ουρανού, συνδυαζόμενες με πρόσθετες πληροφορίες, όπως το ύψος των νεφών πάνω από το έδαφος, μπορούν να οδηγήσουν στην ορθή εκτίμηση της μελλοντικής κατάστασης του ουρανού και κατά επέκταση της Φ/Β παραγωγής, κάνοντας χρήση των κατάλληλων φυσικών εξισώσεων και μεθόδων μηχανικής μάθησης.

Το σύνολο δεδομένων που παρουσιάστηκε στην Ενότητα 4.2, αν και επαρκές για βραχυπρόθεσμες προβλέψεις, δεν επαρκεί για υπερβραχυπρόθεσμους ορίζοντες, καθώς δεν αξιοποιεί δεδομένα από εικόνες και στοιχεία του ουρανού. Για την αντιμετώπιση αυτού του ζητήματος, εισάχθηκε ένα δεύτερο σύνολο δεδομένων, το οποίο περιλαμβάνει επίγειες εικόνες του ουρανού. Δοκιμάζοντας το μοντέλο RF σε αυτό το δεύτερο σύνολο δεδομένων, η διαδικασία επικύρωσης γίνεται πιο αξιόπιστη και εξασφαλίζεται καλύτερη προσαρμοστικότητα στις μεταβαλλόμενες καιρικές συνθήκες σε πραγματικό χρόνο.

Το δεύτερο σύνολο δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση και τον έλεγχο του μοντέλου πρόβλεψης Φ/Β παραγωγής προέρχονται από το Φ/Β σύστημα στέγης τοποθετημένο σε πανεπιστήμιο της Γερμανίας. Το χρονικό εύρος των μετρήσεων εκτείνεται από την 19^η Ιουλίου 2015 μέχρι την 31^η Αυγούστου 2015 [38].

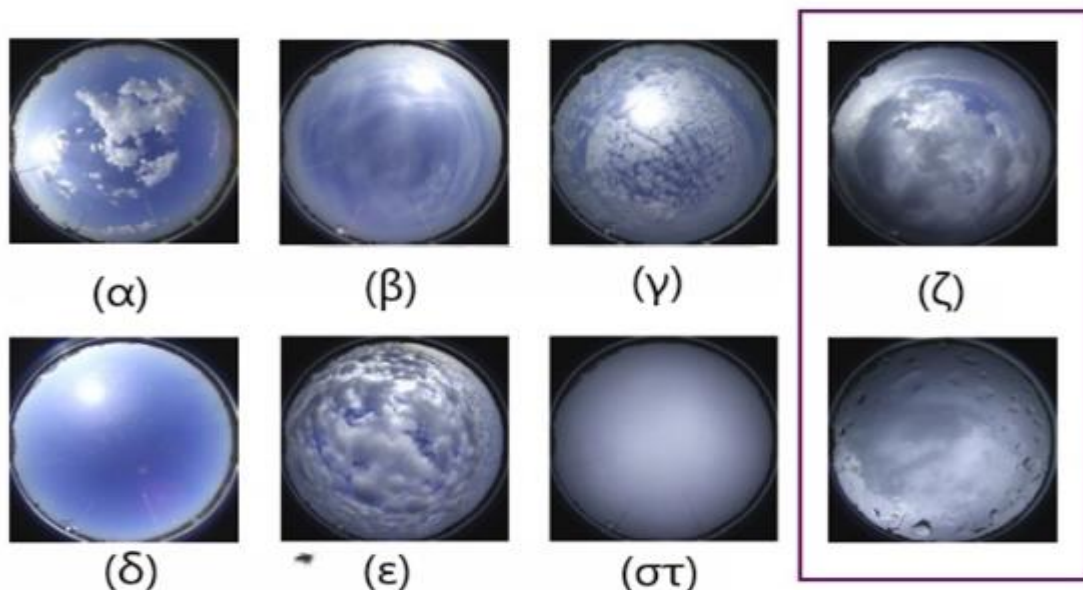
Τα δεδομένα που αξιοποιούνται είναι χρονικής ανάλυσης ενός (1) δευτερολέπτου και ο χρονικός ορίζοντας της πρόβλεψης είναι τα 15 λεπτά. Στην παρούσα διπλωματική, οι χρονοσειρές εισόδου και εξόδου ελήφθησαν έτοιμες, επομένως δεν χρειάστηκε η περεταίρω ανάλυση και επεξεργασία τους.

Σε ότι αφορά την πρόβλεψη της Φ/Β παραγωγής, το σύνολο δεδομένων χρονοσειράς περιλαμβάνει τα εξής [38]: Πρώτον, τις προβλέψεις του μέσου κόκκινου, πράσινου και μπλε χρώματος (Red, Green, Blue – RGB) των εικονοστοιχείων (pixels) της εικόνας του ουράνιου θόλου που αντιστοιχεί στον χρονικό ορίζοντα πρόβλεψης. Οι προβλέψεις αυτές αντιστοιχούν ουσιαστικά σε προβλέψεις νεφοκάλυψης και προσπίπτουσας ηλιακής ακτινοβολίας, και προέρχονται από σειρά βημάτων επεξεργασίας των εικόνων []. Επίσης, στο σύνολο δεδομένων περιλαμβάνονται η ηλιακή αζιμούθια γωνία (solar azimuth angle) και η ηλιακή ζενιθιακή γωνία (solar zenith angle), για την καταγραφή της θέσης του ηλίου, καθώς και οι μετρήσεις της άμεσης κανονικής ακτινοβολίας (Direct Normal Irradiance – DNI), της διάχυτης οριζόντιας ακτινοβολίας (Diffuse Horizontal Irradiance – DHI), της θερμοκρασίας του Φ/Β πάνελ, και ο λόγος του μέσου κόκκινου χρώματος προς το μέσο μπλε χρώμα. Τέλος, το σύνολο δεδομένων περιλαμβάνει τη μέση διαφορά μεταξύ μπλε και κόκκινου χρώματος, καθώς και την τυπική απόκλιση αυτής. Η σχέση μεταξύ κόκκινου και μπλε χρώματος δίνουν μια καλή αίσθηση της ύπαρξης ή μη των νεφών στον ουρανό, καθώς όταν ο ουρανός είναι καθαρός κυριαρχεί το μπλε χρώμα, ενώ στην περίπτωση ύπαρξης νεφών το κόκκινο και το μπλε παίρνουν παραπλήσιες τιμές.

4.5 ΚΛΑΣΕΙΣ ΝΕΦΩΣΕΩΝ

Με σκοπό την καλύτερη πρόβλεψη της Φ/Β παραγωγής, οι επίγειες εικόνες του ουρανού κατατάχθηκαν σε 7 κλάσεις. Η κατηγοριοποίηση της εισόδου συμβάλλει στην βελτίωση της πρόβλεψης Φ/Β παραγωγής με διάφορους τρόπους. Πρώτον, μέσω της κατηγοριοποίησης, το μοντέλο RF μπορεί να μάθει πιο συγκεκριμένα μοτίβα που συνδέονται με κάθε κλάση νεφώσεων. Επιπλέον, η χρήση κλάσεων μειώνει την πολυπλοκότητα των δεδομένων εισόδου, καθιστώντας τη διαδικασία εκπαίδευσης πιο γρήγορη και αποδοτική. Τέλος, η κατηγοριοποίηση της εισόδου βοηθάει το μοντέλο RF να γίνει πιο ευέλικτο και ικανό να προβλέπει την παραγωγή ενέργειας σε ποικίλες καιρικές συνθήκες. Οι επτά κλάσεις [12] των νεφώσεων (Σχήμα 4.3) που χρησιμοποιήθηκαν είναι οι ακόλουθες :

- Cirrus : χαμηλά, αφράτα σύννεφα με χρώμα λευκό ή ανοιχτό γκρι.
- Cumulus: ψηλά, λεπτά σύννεφα που καλύπτουν τον ουρανό, με λευκό χρώμα
- Cirrocumulus: ψηλά διάσπαρτα μικρά σύννεφα, με λευκό χρώμα
- Clear Sky: χωρίς σύννεφα ή με νέφωση κάτω του 10%
- Stratocumulus: χαμηλού ή μεσαίου επιπέδου, σύννεφα σε μορφή σβώλων, σχεδόν καλύπτουν τον ουρανό, με χρώμα λευκό ή γκρι
- Stratus: χαμηλού ή μεσαίου επιπέδου, ομοιόμορφα καλύπτουν τον ουρανό, με γκρι χρώμα
- Nimbostratus: σκοτεινά, πυκνά σύννεφα, με γκρι χρώμα



Σχήμα 4.3: Κατηγορίες νεφών (α): Cirrus (β): Cumulus (γ): Cirrocumulus (δ): Clear Sky (ε): Stratocumulus (στ): Stratus (ζ): Nimbostratus

ΜΟΝΤΕΛΟ ΥΠΕΡΒΡΑΧΥΠΡΟΘΕΣΜΗΣ ΠΡΟΒΛΕΨΗΣ ΦΩΤΟΒΟΛΤΑΙΚΗΣ ΠΑΡΑΓΩΓΗΣ

5.1 ΜΕΘΟΔΟΛΟΓΙΑ

Για την υπερβραχυπρόθεσμη πρόβλεψη της Φ/Β παραγωγής θα αξιοποιηθεί ένα μοντέλο μηχανικής μάθησης RF σε συνδυασμό με τον εξελικτικό αλγόριθμο ALO. Το μοντέλο RF αξιοποιείται για την πρόβλεψη της Φ/Β ισχύος με χρονικό ορίζοντα 15 λεπτών, ενώ ο αλγόριθμος ALO χρησιμοποιείται για τη βελτιστοποίηση των υπερ-παραμέτρων του μοντέλου RF, βελτιώνοντας την ακρίβεια της πρόβλεψης του μοντέλου.

Στην πρώτη εκτέλεση του αλγορίθμου RF γίνεται ο καθορισμός των δειγμάτων και των μεταβλητών που θα χρησιμοποιηθούν στην είσοδο του μοντέλου. Με αυτόν τον τρόπο, μπορεί να αφαιρεθούν οι λιγότερο σημαντικές μεταβλητές, βελτιώνοντας την απόδοση και την ακρίβεια του μοντέλου. Στη συνέχεια, αφού οι μεταβλητές και τα δείγματα τροποποιηθούν κατάλληλα, ο αλγόριθμος ALO εκτελείται για έναν καθορισμένο αριθμό επαναλήψεων. Σε κάθε επανάληψη, ο αλγόριθμος αξιολογεί τις νέες τιμές των υπερ-παραμέτρων του μοντέλου RF με σκοπό να ελαχιστοποιήσει την αντικειμενική συνάρτηση (ελαχιστοποίηση RMSE). Αυτό το καταφέρνει με το να κανονικοποιήσει τις υπερπαραμέτρους σε ένα συγκεκριμένο εύρος τιμών, ώστε να προσαρμοστούν στις ανάγκες του μοντέλου.

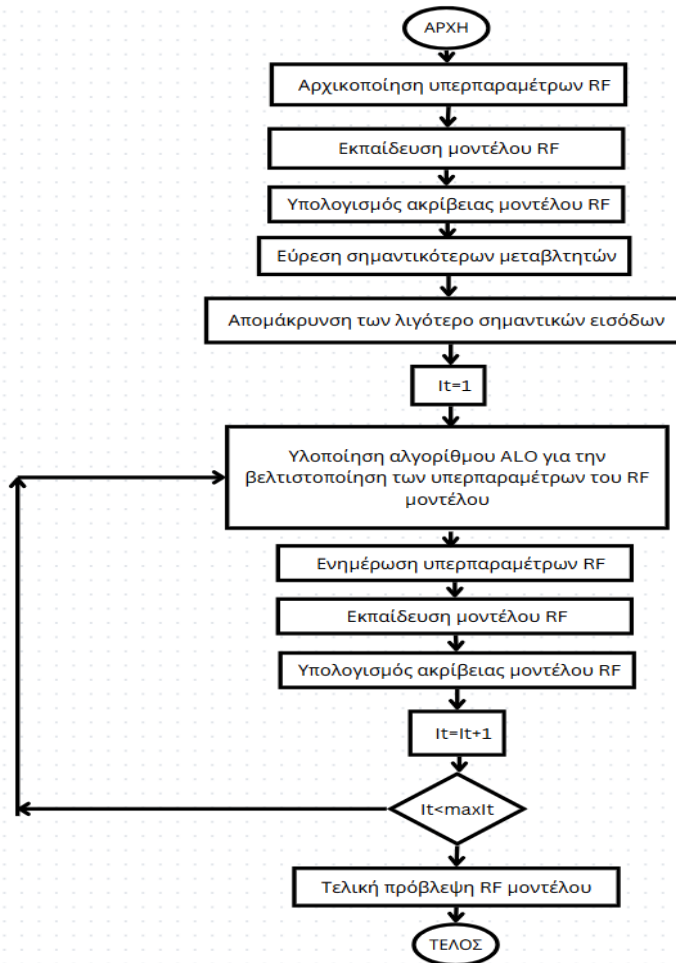
Αρχικά, ο αλγόριθμος ALO δημιουργεί έναν πληθυσμό από μυρμήγκια και αντίλια και εκτιμά την καταλληλότητα κάθε συνδυασμού υπερ-παραμέτρων. Έπειτα, το κάθε μυρμήγκι εκτελεί έναν τυχαίο περίπατο γύρω από ένα επιλεγμένο αντίλιο και η καταλληλότητα της κάθε νέας θέσης αξιολογείται σε κάθε επανάληψη. Η διαδικασία συνεχίζεται επαναλαμβανόμενα, με τα μυρμήγκια να μετακινούνται στον χώρο των υπερ-παραμέτρων και να αξιολογούνται. Σε κάθε επανάληψη, οι υπερ-παραμέτροι που οδηγούν σε χαμηλότερη τιμή RMSE διατηρούνται ως βέλτιστες. Εάν κάποια νέα θέση έχει μικρότερη τιμή RMSE από την τρέχουσα καλύτερη, τότε αυτή υιοθετείται ως η νέα καλύτερη θέση.

Αυτή η διαδικασία επαναλαμβάνεται μέχρι να ολοκληρωθούν όλες οι επαναλήψεις. Τελικά, οι βέλτιστες υπερ-παραμέτροι που προκύπτουν από αυτή τη διαδικασία χρησιμοποιούνται για την τελική εκπαίδευση του μοντέλου RF, εξασφαλίζοντας έτσι ότι το μοντέλο κάνει την ακριβέστερη πρόβλεψη της Φ/Β παραγωγής.

Το διάγραμμα ροής του μοντέλου RF σε συνδυασμό με το ALO απεικονίζεται στο Σχήμα 5.1.

5.2 ΑΡΧΙΚΟΠΟΙΗΣΗ ΥΠΕΡΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

Για την υλοποίηση του μοντέλου RF χρησιμοποιήθηκε η βιβλιοθήκη scikit-learn [46]. Πριν την πρώτη εκτέλεση του μοντέλου, οι υπερπαραμέτροι του μοντέλου RF, οι οποίες παρουσιάστηκαν αναλυτικά στην Ενότητα 3.2.3, πρέπει να οριστούν για να γίνει σωστά η διαδικασία αφαίρεσης των



Σχήμα 5.1: Διάγραμμα ροής του μοντέλου υπερβραχυπροθεσμής πρόβλεψης Φ/Β παραγωγής

λιγότερων σημαντικών δειγμάτων και μεταβλητών. Οι αρχικές τιμές των υπερπαραμέτρων είναι οι εξής [15]:

- Πλήθος δέντρων: 500
- Μέγιστο βάθος δέντρων: 5
- Αντικειμενική συνάρτηση: ελαχιστοποίηση RMSE
- Ελάχιστος αριθμός δειγμάτων για διάσπαση: 2
- Ελάχιστος αριθμός δειγμάτων ανά φύλλο: 1
- Μέγιστος αριθμός χαρακτηριστικών: όσο το σύνολο των χαρακτηριστικών εισόδου

Οι τιμές για το πλήθος δέντρων και το μέγιστο βάθος δέντρων επιλέχθηκαν λόγω των ευρημάτων της εργασίας [15], όπου προτείνεται αυτός ο αριθμός για τη σταθερότητα των αποτελεσμάτων, ενώ για τις υπόλοιπες υπερπαραμέτρους, χρησιμοποιήθηκαν οι προεπιλεγμένες τιμές της βιβλιοθήκης scikit-learn, καθώς θεωρούνται επαρκείς και αξιόπιστες για ένα ευρύ φάσμα εφαρμογών.

5.3 ΥΠΟΛΟΓΙΖΟΜΕΝΑ ΣΦΑΛΜΑΤΑ ΚΑΙ ΜΕΤΡΗΤΙΚΑ ΑΞΙΟΛΟΓΗΣΗΣ ΜΟΝΤΕΛΟΥ

Κατά την εκτέλεση του μοντέλου RF για την υπερβραχυπρόθεσμη πρόβλεψη της Φ/Β παραγωγής, υπολογίζεται το σφάλμα RMSE μέσω της εξίσωσης (3.7), με σκοπό την αξιολόγηση της απόδοσης του μοντέλου. Η μέτρηση αυτή υπολογίζει τη διαφορά μεταξύ της πραγματικής τιμής της Φ/Β παραγωγής και αυτής που προβλέπεται μετά την εκπαίδευση του μοντέλου. Η χρήση αυτού του μετρητικού σφάλματος επιλέχθηκε καθώς το RMSE παρέχει μια σαφή ένδειξη του μέσου σφάλματος ενώ ταυτόχρονα τιμωρεί μεγαλύτερες αποκλίσεις αναλογικά περισσότερο σε σχέση με τις μικρότερες.

Για να αξιολογηθεί αν το μοντέλο RF λειτουργεί αποτελεσματικά και αποδοτικά, χρησιμοποιήθηκε ως μέτρο σύγκρισης το μοντέλο persistence. Το μοντέλο persistence είναι ένα μοντέλο πρόβλεψης που δεν απαιτεί κάποια εκπαίδευση ή περίπλοκους υπολογισμούς, και παρέχει ένα βασικό σημείο αναφοράς για την αξιολόγηση πιο σύνθετων μοντέλων. Συγκεκριμένα, το μοντέλο persistence βασίζεται στην υπόθεση ότι η μελλοντική τιμή μιας μεταβλητής θα είναι ίδια με την πιο πρόσφατη παρατηρούμενη τιμή της, και ορίζεται ως:

$$\hat{P}(t + \Delta t) = P(t) \quad (5.1)$$

Όπου:

- $\hat{P}(t + \Delta t)$: Η προβλεπόμενη τιμή της μεταβλητής για τη χρονική στιγμή $t + \Delta t$
- $P(t)$: Η πραγματική τιμή της μεταβλητής τη χρονική στιγμή t

Αυτό το καθιστά ιδιαίτερα χρήσιμο για υπερβραχυπρόθεσμες προβλέψεις, όπου ο χρονικός ορίζοντας της πρόβλεψης είναι ιδιαίτερα μικρός, με αποτέλεσμα η προβλεπόμενη και η τελευταία παρατηρούμενη τιμή να έχουν μεγάλη συσχέτιση. Επομένως, είναι δύσκολο για ένα άλλο μοντέλο να κάνει προβλέψεις καλύτερες από αυτές που παρέχει το μοντέλο persistence. Εάν ένα μοντέλο δεν μπορεί να ξεπεράσει την απόδοση του μοντέλου persistence, τότε δεν αξίζει να χρησιμοποιηθεί. Στην περίπτωση της πρόβλεψης της Φ/Β παραγωγής, το μοντέλο persistence προβλέπει ότι η παραγωγή σε 15 λεπτά θα είναι ίση με την τρέχουσα παραγωγή. Έτσι, για να θεωρηθεί ότι το προτεινόμενο μοντέλο (RF+ALO) λειτουργεί αποδοτικά, πρέπει να επιτυγχάνει χαμηλότερη τιμή σφάλματος RMSE από αυτήν του μοντέλου persistence.

5.4 ΧΩΡΙΣΜΟΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΣΥΝΟΛΑ ΕΛΕΓΧΟΥ ΚΑΙ ΕΚΠΑΙΔΕΥΣΗΣ

Για την ανάπτυξη του μοντέλου RF σε συνδυασμό με τον αλγόριθμο ALO με σκοπό την υπερβραχυπρόθεσμη πρόβλεψη της Φ/Β παραγωγής, τα δύο σύνολα δεδομένων χωρίστηκαν σε σετ

εκπαίδευσης και σετ ελέγχου με ποσοστά 70% και 30%, αντίστοιχα. Στην περίπτωση του πρώτου σετ δεδομένων, που περιέχει αποκλειστικά αριθμητικά δεδομένα και παρουσιάστηκε στην Ενότητα 4.2, το σετ εκπαίδευσης περιλαμβάνει δεδομένα από τις 20/07/2022 μέχρι τις 19/08/2022, ενώ το σετ ελέγχου περιλαμβάνει δεδομένα από τις 20/08/2022 μέχρι τις 02/09/2022. Στην περίπτωση του δεύτερου σετ δεδομένων, το οποίο παρουσιάστηκε στην Ενότητα 4.4 και περιλαμβάνει εικόνες του ουρανού και αριθμητικά δεδομένα, το σετ εκπαίδευσης περιλαμβάνει δεδομένα από τις 19/07/2015 μέχρι τις 18/08/2015, ενώ το σετ ελέγχου περιλαμβάνει δεδομένα από τις 19/08/2015 μέχρι τις 31/08/2015.

Η επιλογή αυτή έγινε επειδή η αξιοποίηση του 70% του συνόλου των δεδομένων για την εκπαίδευση συνήθως παρέχει αρκετές παρατηρήσεις για να εκπαιδευτεί το μοντέλο αποτελεσματικά, ενώ το υπόλοιπο 30% αποτελεί ένα ικανό δείγμα για να ελεγχθεί η επίδοσή του σε νέα δεδομένα. Η χρήση αυτής της αναλογίας είναι συχνή στην πράξη και έχει αποδειχθεί αποτελεσματική σε πολλές εφαρμογές μηχανικής μάθησης [47], καθώς προσφέρει μια ισορροπία μεταξύ της αποδοτικότητας και της αξιοπιστίας της αξιολόγησης του μοντέλου.

5.5 ΕΠΙΛΟΓΗ ΜΕΤΑΒΛΗΤΩΝ ΕΙΣΟΔΟΥ

Όπως προαναφέρθηκε στην Ενότητα 5.1, οι μεταβλητές που θα χρησιμοποιηθούν στην είσοδο του μοντέλου πρόβλεψης παίζουν σημαντικό ρόλο στην απόδοσή του. Η σωστή επιλογή αυτών των μεταβλητών εισόδου παίζει κρίσιμο ρόλο στην επίτευξη της υψηλής ακρίβειας των προβλέψεων, καθώς επηρεάζει άμεσα την ικανότητα του μοντέλου να κατανοεί και να προβλέπει τις τάσεις και τις διακυμάνσεις της Φ/Β παραγωγής. Επομένως, η διαδικασία επιλογής των καταλληλότερων μεταβλητών είναι απαραίτητη για τη βελτιστοποίηση της απόδοσης του μοντέλου και τη μείωση του σφάλματος RMSE της πρόβλεψης. Είναι σημαντικό να σημειωθεί ότι αυτή η διαδικασία είναι απαραίτητη μόνο για το πρώτο σετ δεδομένων (Ενότητα 4.2), καθώς για το δεύτερο σετ δεδομένων (Ενότητα 4.4) η διαδικασία επιλογής των καταλληλότερων μεταβλητών έχει ήδη ολοκληρωθεί, δεδομένου ότι το σετ αυτό έχει ήδη υποστεί προεπεξεργασία.

5.5.1 Μονοδιάστατης είσοδος

Πριν εκτελεστεί ο αλγόριθμος RF με όλες τις διαθέσιμες μεταβλητές εισόδου, επιλέχθηκε να γίνει η υλοποίησή του με μόνη είσοδο την τρέχουσα μέτρηση της τιμής της Φ/Β παραγωγής σε κάθε χρονική στιγμή. Αυτή η επιλογή έγινε για να εξεταστεί εάν το μοντέλο μπορεί να προβλέπει αποτελεσματικά βασισμένο μόνο σε αυτήν την μεταβλητή. Υπολογίστηκε το RMSE μοντέλου RF και παρόλο που εμφάνισε μικρή τιμή, η οποία φαίνεται στον Πίνακα 5.1, δεν παρουσίασε καλύτερη απόδοση σε σχέση με το μοντέλο persistence. Η χαμηλή τιμή του RMSE οφείλεται στον μικρό χρονικό ορίζοντα πρόβλεψης, αλλά και στο γεγονός ότι το μοντέλο RF είναι αποδοτικό στην εύρεση προβλέψεων ακόμη και με περιορισμένο αριθμό δεδομένων. Ωστόσο, το μοντέλο RF με μονοδιάστατη είσοδο δεν πέτυχε χαμηλότερη τιμή RMSE από το μοντέλο persistence, καθώς η πρόβλεψη της Φ/Β παραγωγής με βάση μόνο την τρέχουσα τιμή της είναι δύσκολη, ενώ το μοντέλο persistence αποτελεί ισχυρό εργαλείο όταν δεν υπάρχουν σημαντικές αλλαγές στις εξωγενείς συνθήκες (υπερβραχυπρόθεσμη πρόβλεψη). Με βάση αυτήν την παρατήρηση, διαπιστώθηκε ότι απαιτούνται περισσότερες μεταβλητές στην είσοδο του μοντέλου προκειμένου να βελτιωθεί η απόδοσή του.

5.5.2 Πολυδιάστατη είσοδος

Η αύξηση των μεταβλητών εισόδου είναι αναμενόμενο ότι θα αυξήσει την απόδοση του μοντέλου. Ωστόσο, δεν συμβάλλουν όλες οι μεταβλητές στην αύξηση της απόδοσης και της ακρίβειας του μοντέλου. Κάθε μεταβλητή έχει διαφορετική συσχέτιση με την Φ/Β παραγωγή, και οι μεταβλητές που έχουν πολύ μικρή συσχέτιση ενδεχομένως να λειτουργούν ως θόρυβος. Ως αποτέλεσμα, δεν έχει νόημα να χρησιμοποιηθούν αυτές οι μεταβλητές, καθώς η παράλειψή τους όχι μόνο βελτιώνει την απόδοση του μοντέλου, αλλά μειώνει και την συνολική υπολογιστική πολυπλοκότητα. Για να επιλεγούν οι καταλληλότερες μεταβλητές, αξιοποιήθηκε ένας πίνακας συσχετίσεων (correlation matrix). Ο πίνακας συσχετίσεων είναι μια στατιστική τεχνική που χρησιμοποιείται για να αξιολογήσει τη σχέση μεταξύ μεταβλητών σε ένα σύνολο δεδομένων. Κάθε κελί του πίνακα περιέχει έναν συντελεστή συσχέτισης, με τις τιμές να κυμαίνονται από -1 έως 1. Τιμή 1 θεωρείται ισχυρή θετική σχέση μεταξύ των μεταβλητών, τιμή 0 ουδέτερη σχέση, και τιμή -1 ισχυρή αρνητική σχέση. Ο συντελεστής συσχέτισης που αξιοποιήθηκε στον πίνακα συσχέτισης είναι ο συντελεστής συσχέτισης του Pearson, ο οποίος είναι ένας γραμμικός συντελεστής και υπολογίζεται με βάση την εξίσωση (5.2):

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sqrt{(\sum_{i=1}^n (X_i - \bar{X})^2)(\sum_{i=1}^n (Y_i - \bar{Y})^2)}} \quad (5.2)$$

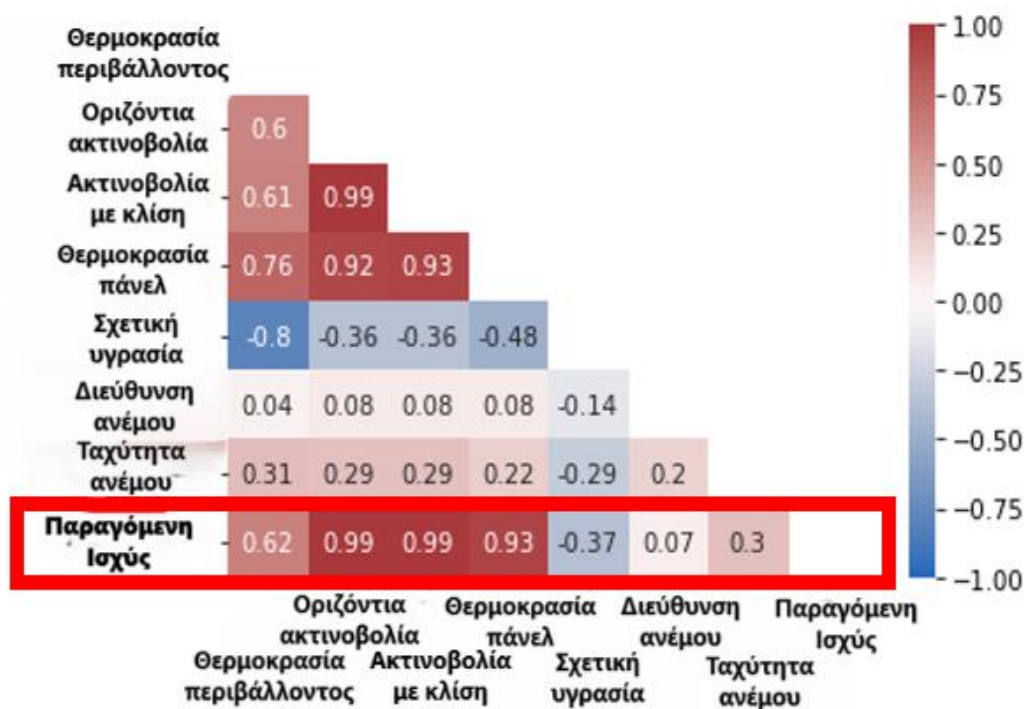
Όπου:

- X_i, Y_i : Οι τιμές του δείγματος i των μεταβλητών X και Y που μελετώνται.
- \bar{X}, \bar{Y} : Οι μέσοι όροι των δύο μεταβλητών που X και Y που μελετώνται.
- n : Το σύνολο των δειγμάτων των μεταβλητών X και Y .

Οι μεταβλητές που εξετάστηκαν ήταν: η θερμοκρασία περιβάλλοντος, η οριζόντια ακτινοβολία, η ακτινοβολία με κλίση, η θερμοκρασία του Φ/Β πάνελ, η σχετική υγρασία, η διεύθυνση ανέμου και η ταχύτητα ανέμου. Μέσω του πίνακα συσχετίσεων εξετάστηκε ποιες από αυτές τις μεταβλητές έχουν τη μεγαλύτερη συσχέτιση με την παραγόμενη Φ/Β ισχύ. Τα αποτελέσματα του πίνακα συσχέτισης φαίνονται στο Σχήμα 5.2.

Αφού αναλύθηκαν τα δεδομένα, ο πίνακας συσχετίσεων ανέδειξε ποιες μεταβλητές έχουν τη μεγαλύτερη επίδραση στην παραγόμενη Φ/Β ισχύ. Στην συνέχεια, οι μεταβλητές αυτές επιλέχθηκαν για την περαιτέρω ανάπτυξη του μοντέλου, με στόχο την αύξηση της απόδοσης και της ακρίβειας των προβλέψεών του.

Για το προτεινόμενο μοντέλο RF, ως κατώφλι της απόλυτης τιμής του συντελεστή συσχέτισης για την επιλογή των μεταβλητών εισόδου ορίστηκε η τιμή 0,6. Η τιμή αυτή επιλέχθηκε για να επιτευχθεί μια μέση λύση μεταξύ καλής προσαρμογής του μοντέλου και χαμηλής υπολογιστικής πολυπλοκότητας. Με βάση αυτό το κριτήριο, επιλέχθηκαν οι εξής μεταβλητές:



Σχήμα 5.2: Πίνακας συσχετίσεων των μεταβλητών που θα αξιοποιηθούν στο μοντέλο RF

- Θερμοκρασία περιβάλλοντος
- Οριζόντια ακτινοβολία
- Ακτινοβολία με κλίση
- Θερμοκρασία πάνελ

Οι υπόλοιπες μεταβλητές δεν θα χρησιμοποιηθούν ως είσοδοι στο μοντέλο RF για τους εξής λόγους: η υγρασία έχει αρνητική συσχέτιση με τη Φ/Β παραγωγή, αλλά όχι αρκετά ισχυρή. Παρομοίως, η ταχύτητα ανέμου παρουσιάζει μικρή θετική συσχέτιση, ενώ η διεύθυνση του ανέμου έχει σχεδόν μηδενική συσχέτιση.

Με τη χρήση αυτών των μεταβλητών εισόδου, το μοντέλο έδειξε καλύτερη απόδοση από την περίπτωση χρήσης μονοδιάστατης εισόδου, ενώ επίσης ξεπέρασε το μοντέλο persistence. Για την αξιολόγηση και σύγκριση των μοντέλων χρησιμοποιήθηκαν τρία μετρητικά σφάλματος: το RMSE, το MAE, και το MSE, που περιγράφονται από τις εξισώσεις (3.7), (3.8) και (3.9), αντίστοιχα. Τα κανονικοποιημένα αποτελέσματα παρουσιάζονται στον Πίνακα 5.1. Σε σύγκριση με το μοντέλο persistence, το μοντέλο RF με πολυδιάστατη είσοδο βελτιώνει κατά 48,45%, 49,83% και 24,97% τα MAE, MSE και RMSE, αντίστοιχα. Επιπλέον, σε σχέση με το μοντέλο RF με μονοδιάστατη είσοδο, το μοντέλο RF με πολυδιάστατη είσοδο βελτιώνει κατά 51,61%, 50% και 25,16% τα MAE, MSE και RMSE,

Πίνακας 5.1: Αποτελέσματα μονοδιάστατης και πολυδιάστατης εισόδου του απλού μοντέλου RF

Σφάλματα μέτρησης			
	MAE(α.μ)	MSE(α.μ.)	RMSE(α.μ.)
Random Forest model (μονοδιάστατη είσοδος)	0,031	0,006	0,0775
Random Forest model (πολυδιάστατη είσοδος)	0,015	0,003	0,058
Persistence model	0,0291	0,00598	0,0773
Ποσοστιαία βελτίωση μεταξύ RF με πολυδιάσταση είσοδο και RF με μονοδιάστατη είσοδο	51,61%	50%	25,16%
Ποσοστιαία βελτίωση μεταξύ RF με πολυδιάσταση είσοδο και persistence model	48,45%	49,83%	24,97%

αντίστοιχα. Αυτά τα αποτελέσματα υποδεικνύουν ότι η χρήση πολυδιάστατης εισόδου βελτιώνει σημαντικά την απόδοση του μοντέλου RF, τόσο σε σύγκριση με τη μονοδιάστατη είσοδο όσο και σε σχέση με το μοντέλο persistence, επιτρέποντας στο μοντέλο να παράγει πιο ακριβείς προβλέψεις για τη Φ/Β παραγωγή.

5.6 ΑΛΓΟΡΙΘΜΟΣ ANTLION-LION OPTIMIZER ΓΙΑ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΥΠΕΡΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

Για να έχει το μοντέλο RF τη μεγαλύτερη δυνατή ακρίβεια και απόδοση, πρέπει να βελτιστοποιηθούν οι υπερπαράμετροί του. Ο αλγόριθμος που επιλέχθηκε για αυτόν τον σκοπό είναι ο ALO, καθώς προσφέρει έναν καινοτόμο και αποδοτικό τρόπο αναζήτησης των βέλτιστων υπερπαραμέτρων. Όπως έχει αναφερθεί στην Ενότητα 3.5, ο αλγόριθμος ALO είναι εμπνευσμένος από τη συμπεριφορά των αντιλίων και των μυρμηγκιών στη φύση και η μεταφορά της φυσικής διαδικασίας θήρευσης σε αλγοριθμική μορφή επιτρέπει την αποτελεσματική εξερεύνηση και εκμετάλλευση του χώρου αναζήτησης, αποφεύγοντας τις τοπικές λύσεις και εντοπίζοντας την ολικά βέλτιστη. Χάρη σε αυτές τις ιδιότητες, ο ALO μπορεί να εντοπίσει τις ιδανικές ρυθμίσεις για το μοντέλο RF. Οι υπερπαράμετροι που πρέπει ο αλγόριθμος ALO να ρυθμίσει για να βελτιστοποιηθεί το μοντέλο Random Forest είναι οι εξής:

- Ο αριθμός των δέντρων
- Το μέγιστο βάθος των δέντρων
- Ο ελάχιστος αριθμός δειγμάτων για διάσπαση
- Ο ελάχιστος αριθμός δειγμάτων ανά φύλλο

- Ο μέγιστος αριθμός χαρακτηριστικών

5.6.1 Υλοποίηση του αλγορίθμου

Πρώτο στάδιο στην υλοποίηση του αλγορίθμου ALO για την βελτιστοποίηση των υπερπαραμέτρων του μοντέλου RF είναι η αρχικοποίηση των παραμέτρων του. Οι παράμετροι που περιλαμβάνει ο αλγόριθμος ALO είναι ο μέγιστος αριθμός επαναλήψεων του αλγορίθμου, ο αριθμός μυρμηγκιών στο πλυθησμό, ο αριθμός αντίλιων στο πλυθησμό, τα κατώτατα και ανώτατα όρια των διαστάσεων του χώρου αναζήτησης και ο αριθμός των διαστάσεων της λύσης. Συγκεκριμένα οι διαστάσεις στον αλγόριθμο ALO αντιστοιχούν με τις υπερπαραμέτρους προς βελτιστοποίηση του μοντέλου RF.

Οι τιμές που παίρνουν οι παράμετροι του αλγορίθμου ALO για το πρώτο σετ δεδομένων που αφορά μόνο αριθμητικά δεδομένα (Ενότητα 4.2), παρουσιάζονται στον Πίνακα 5.2. Όπως φαίνεται στον Πίνακα 5.2, χρησιμοποιήθηκε μικρός αριθμός επαναλήψεων, μυρμηγκιών και αντίλιων (50 επαναλήψεις, 60 μυρμηγκία και 30 αντίλια). Αυτό έγινε επειδή το σετ δεδομένων που χρησιμοποιήθηκε είναι πολύ μεγάλο καθώς τα δεδομένα έχουν πολύ μεγάλη χρονική ανάλυση (1 λεπτό), και ως χρονικό όριο για το τρέξιμο του αλγορίθμου τέθηκαν οι 24 ώρες. Ο περιορισμός αυτός επιβλήθηκε για να διασφαλιστεί ότι η διαδικασία βελτιστοποίησης θα ολοκληρωθεί εντός λογικού χρονικού πλαισίου.

Στην περίπτωση του δεύτερου σετ δεδομένων (Ενότητα 4.4), όπου υπάρχουν και εικόνες του ουρανού, χρησιμοποιήθηκε μεγάλος αριθμός επαναλήψεων, μυρμηγκιών και αντίλιων καθώς τα δεδομένα σε αυτή την περίπτωση ήταν προεπεξεργασμένα και πολύ λιγότερα ανά κλάση σε σχέση με το πρώτο σετ δεδομένων, λόγω της κατηγοριοποίησης ανά κλάση νεφώσεων. Οι τιμές που παίρνουν οι παράμετροι του αλγορίθμου ALO για το δεύτερο σετ δεδομένων, φαίνονται στον Πίνακα 5.3.

Η υλοποίηση του αλγορίθμου ALO στα πλαίσια της βελτιστοποίησης των υπερπαραμέτρων του RF περιλαμβάνει τα εξής στάδια:

1. Δημιουργία των αρχικών πληθυσμών των μυρμηγκιών και των αντίλιων μέσα στο χώρο αναζήτησης. Κάθε αντίλιο αντιπροσωπεύει μια πιθανή λύση στο πρόβλημα βελτιστοποίησης των υπερπαραμέτρων του μοντέλου RF, ενώ τα μυρμηγκία εκτελούν τυχαίους περιπάτους γύρω από τα αντίλια για να βρουν καλύτερους συνδυασμούς υπερπαραμέτρων.

Πίνακας 5.2: Τιμές παραμέτρων αλγορίθμου ALO για το πρώτο σετ δεδομένων

Τιμές παραμέτρων	
Παράμετρος	Τιμή
Μέγιστος αριθμός επαναλήψεων	50
Αριθμός μυρμηγκιών	60
Αριθμός αντίλιων	30
Ανώτατο όριο	5
Κατώτατο όριο	-5
Αριθμός διαστάσεων	5

Πίνακας 5.3: Τιμές παραμέτρων αλγόριθμου ALO για το δεύτερο σετ δεδομένων

Τιμές παραμέτρων	
Παράμετρος	Τιμή
Μέγιστος αριθμός επαναλήψεων	100
Αριθμός μυρμηγκιών	250
Αριθμός αντίλιων	60
Ανώτατο όριο	5
Κατώτατο όριο	-5
Αριθμός διαστάσεων	5

2. Με τη χρήση μια αντικειμενικής συνάρτησης, υπολογίζεται η καταλληλότητα (fitness) της θέσης κάθε αντίλιου μετρώντας έτσι την ποιότητα της λύσης. Στην περίπτωση αυτή, η αντικειμενική συνάρτηση υπολογίζει το σφάλμα RMSE μεταξύ των προβλεπόμενων και των πραγματικών τιμών της Φ/Β παραγωγής, για κάθε συνδυασμό παραμέτρων του RF.
3. Τα μυρμηγκία επιλέγουν ένα αντίλιο για να κάνουν τον τυχαίο περίπατο τους γύρω από αυτό, χρησιμοποιώντας έναν μηχανισμό ρουλέτας (roulette wheel) που δίνει μεγαλύτερες πιθανότητες επιλογής σε αντίλια με καλύτερη καταλληλότητα.
4. Ο τυχαίος περίπατος των μυρμηγκιών προσαρμόζεται ανάλογα με τον αριθμό των εκτελεσμένων επαναλήψεων. Καθώς προχωρούν οι επαναλήψεις, το εύρος του περιπάτου μειώνεται, προσομοιώνοντας την σταδιακή σύγκλιση προς την καλύτερη λύση. Συγκεκριμένα, τα όρια του τυχαίου περιπάτου των μυρμηγκιών προσαρμόζονται ανάλογα με την πρόοδο των επαναλήψεων, επιτρέποντας έτσι την καλύτερη εξερεύνηση του χώρου αναζήτησης κατά τις αρχικές επαναλήψεις και την σύγκλιση σε μια λύση κατά τις τελευταίες επαναλήψεις.
5. Η καλύτερη λύση (elite antlion) διατηρείται σε τη διάρκεια της διαδικασίας και τα μυρμηγκία εκτελούν επίσης περιπάτους γύρω από αυτή τη λύση. Αν κάποιο μυρμηγκί βρει μια καλύτερη λύση, το elite antlion ενημερώνεται με αυτή τη νέα λύση.
6. Ο αλγόριθμος ALO ολοκληρώνεται όταν ο προκαθορισμένος αριθμός επαναλήψεων έχει ολοκληρωθεί. Τότε, η καλύτερη λύση που έχει βρεθεί κατά τη διάρκεια όλων των επαναλήψεων θεωρείται ως η τελική βέλτιστη λύση. Για κάθε διάσταση της βέλτιστης λύσης, πραγματοποιείται μετατροπή στις αντίστοιχες τιμές των υπερπαραμέτρων του μοντέλου RF και αυτή η μετατροπή γίνεται χρησιμοποιώντας ένα συγκεκριμένο εύρος για κάθε υπερπάρμετρο. Το εύρος των τιμών που μπορούν να πάρουν οι υπερπάρμετροι του μοντέλου RF φαίνεται στον Πίνακα 5.4.

Πίνακας 5.4: Εύρος υπερπαραμέτρων του μοντέλου RF

Εύρος υπερπαραμέτρων		
Παράμετρος	Ελάχιστη τιμή	Μέγιστη τιμή
Αριθμός δέντρων	200	500
Μέγιστο βάθος	3	8
Ελάχιστος αριθμός δειγμάτων για διάσπαση	2	300
Ελάχιστος αριθμός δειγμάτων άνα φύλλο	1	30
Μέγιστος αριθμός χαρακτηριστικών	1	6

Τα εύρη για τις υπερπαραμέτρους του μοντέλου RF επιλέχθηκαν έτσι ώστε να αποφευχθεί η υπερεκπαίδευση και ταυτόχρονα για να μπορεί το μοντέλο να αποδίδει καλά σε νέα, ανεξάρτητα δεδομένα. Συγκεκριμένα, επιλέχθηκε ο αριθμός των δέντρων να είναι άνω των 200 για να έχει επαρκή ποικιλομορφία και όχι πάνω των 500 για να αποφευχθεί το πρόσθετο υπολογιστικό κόστος. Επίσης, το μέγιστο βάθος των δέντρων κυμαίνεται στο διάστημα [3, 8], ώστε να αναγνωρίζουν τις σημαντικές σχέσεις μεταξύ των δεδομένων χωρίς να οδηγούν σε υπερεκπαίδευση.

ΑΠΟΤΕΛΕΣΜΑΤΑ ΜΟΝΤΕΛΟΥ

6.1 ΕΙΣΑΓΩΓΗ

Για την εκπαίδευση και την αξιολόγηση του μοντέλου RF σε συνδυασμό με τον αλγόριθμο ALO, η γλώσσα προγραμματισμού που αξιοποιήθηκε είναι η Python με χρήση του εύχρηστου περιβάλλοντος προγραμματισμού Spyder για την περαιτέρω ανάλυση των αποτελεσμάτων. Το υπολογιστικό σύστημα που αξιοποιήθηκε για την εκπαίδευση και υλοποίηση του μοντέλου είναι ένας φορητός υπολογιστής με τα εξής χαρακτηριστικά:

- Κεντρική μονάδα επεξεργασίας(CPU): Intel Core i7-7820HQ CPU @ 2.90GHz
- Μνήμη RAM: 32GB
- Τύπος συστήματος: 64-bit λειτουργικό σύστημα, επεξεργαστής βασισμένος σε x64

Οι σημαντικότερες βιβλιοθήκες που αξιοποιήθηκαν στη γλώσσα Python για την σωστή και αποδοτική εκπαίδευση του μοντέλου RF είναι :

- Scikit-learn [46]: Παρέχει ένα ευρύ φάσμα αλγορίθμων για την εκπαίδευση μοντέλων μηχανικής μάθησης.
- Pandas [48]: Προσφέρει δομές δεδομένων και λειτουργίες για την ευέλικτη και αποτελεσματική επεξεργασία και ανάλυση δεδομένων.
- Numpy [49]: Παρέχει υψηλής απόδοσης πίνακες και λειτουργίες πάνω σε αυτούς για την εκτέλεση αριθμητικών υπολογισμών.

6.2 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΩΤΗΣ ΠΕΡΙΠΤΩΣΗΣ ΠΡΟΣ ΜΕΛΕΤΗ

Το πρώτο σετ δεδομένων περιέχει μόνο αριθμητικά δεδομένα για την πρόβλεψη της Φ/B παραγωγής. Η ανάλυση ξεκίνησε με την εφαρμογή του αλγορίθμου ALO, ο οποίος χρησιμοποιήθηκε για να εντοπιστούν οι πιο αποδοτικές υπερπαραμέτροι για το συγκεκριμένο σετ δεδομένων. Οι υπερπαραμέτροι αυτοί βελτιστοποιήθηκαν με στόχο τη μεγιστοποίηση της ακρίβειας της πρόβλεψης. Μετά την εφαρμογή του αλγορίθμου και τον καθορισμό των υπερπαραμέτρων, υπολογίστηκαν τρία μετρητικά σφάλματος για την αξιολόγηση της απόδοσης του μοντέλου: το RMSE, το MAE και το MSE, τα οποία

Πίνακας 6.1: Βέλτιστες τιμές των υπερπαραμέτρων του μοντέλου RF

Βέλτιστες τιμές των υπερπαραμέτρων	
Παράμετρος	Βέλτιστη τιμή
Αριθμός δέντρων	368
Μέγιστο βάθος	5
Ελάχιστος αριθμός δειγμάτων για διάσπαση	92
Ελάχιστος αριθμός δειγμάτων άνα φύλλο	24
Μέγιστος αριθμός χαρακτηριστικών	5

περιγράφονται από τις εξισώσεις (3.7), (3.8) και (3.9), αντίστοιχα. Τα μετρητικά αυτά παίζουν κρίσιμο ρόλο στην εκτίμηση της ακρίβειας και της αξιοπιστίας των προβλέψεων του μοντέλου. Για περαιτέρω ανάλυση των αποτελεσμάτων του πρώτου σετ δεδομένων, πραγματοποιήθηκε εποχική ανάλυση προκειμένου να φανερωθούν τυχόν μοτίβα που ενδέχεται να επηρεάζουν τη Φ/Β παραγωγή.

6.2.1 Εύρεση αποδοτικότερων υπερπαραμέτρων

Οι βέλτιστες τιμές των υπερπαραμέτρων που προέκυψαν από τον αλγόριθμο ALO και θα χρησιμοποιηθούν για τη ρύθμιση του μοντέλου RF ώστε να αποδώσει βέλτιστα στο πρώτο σετ δεδομένων, φαίνονται στον Πίνακα 6.1. Όπως φαίνεται από τον Πίνακα 6.1, τα εύρη που τέθηκαν για τις υπερπαραμέτρους είναι κατάλληλα, καθώς καμία υπερπαραμέτρος δεν πήρε κάποια ακραία τιμή. Αυτό υποδηλώνει ότι τα επιλεγμένα όρια κάλυψαν αποτελεσματικά τις ανάγκες του μοντέλου και επέτρεψαν την εύρεση βέλτιστων τιμών. Συγκεκριμένα, το μέγιστο βάθος των δέντρων πήρε τιμή 5, επιλογή που έγινε για την αποφυγή υπερπροσαρμογής. Το βάθος αυτό είναι κατάλληλο για να μπορέσει το μοντέλο να μάθε από τα δεδομένα, χωρίς να υπάρχει κίνδυνος να χάσει από τη δυνατότητα γενίκευσής του. Αντίστοιχα, ο αριθμός των δέντρων πήρε τιμή 368, έτσι ώστε να εξασφαλίζεται η κατάλληλη ακρίβεια χωρίς τον κίνδυνο της υπερπροσαρμογής. Με περισσότερα δέντρα θα μπορούσε να υπάρχει περαιτέρω βελτίωση στην ακρίβεια αλλά και μεγαλύτερη πιθανότητα να είναι επιρρεπές το μοντέλο σε υπερπροσαρμογή. Επιπλέον, ο μέγιστος αριθμός χαρακτηριστικών που εξετάζονται σε κάθε δέντρο είναι 5, άρα σε πολλές περιπτώσεις αξιοποιούνται ταυτόχρονα 5 από τα 6 χαρακτηριστικά εισόδου στην δημιουργία των δέντρων. Με την επιλογή αυτή, το μοντέλο μπορεί να κάνει καλή διάκριση των δεδομένων, χωρίς να υπερφορτώνεται με πολλές επιλογές.

Πίνακας 6.2: Αποτελέσματα μοντέλων για το πρώτο σετ δεδομένων.

Μέθοδος	Σφάλματα μέτρησης		
	RMSE (α.μ)	MSE (α.μ)	MAE (α.μ)
Random Forest χωρίς ALO:	0,096	0,0092	0,037
Persistence model:	0,093	0,0086	0,0376
Random Forest με ALO (προτεινόμενο):	0,0894	0,0079	0,036
Ποσοστιαία βελτίωση μεταξύ προτεινόμενου μοντέλου και μοντέλου persistence	3,9%	8,1%	4,22%
Ποσοστιαία βελτίωση μεταξύ προτεινόμενου μοντέλου και RF χωρίς ALO	6,92%	14,13%	2,7%

6.2.2 Συγκεντρωτικά αποτελέσματα

Με τη χρήση των υπερπαραμέτρων που βρεθήκαν μέσω του αλγορίθμου ALO πραγματοποιήθηκε η εκτέλεση του μοντέλου RF για την πρόβλεψη της Φ/Β παραγωγής. Τα αποτελέσματα του μοντέλου RF σε συνδυασμό με τον αλγόριθμο ALO παρουσιάζονται στον Πίνακα 6.2, όπου φαίνεται η ακρίβεια και η αποτελεσματικότητα της πρόβλεψης. Για να αξιολογηθεί η απόδοση του προτεινόμενου μοντέλου, πραγματοποιήθηκε σύγκριση με δύο άλλες περιπτώσεις, το μοντέλο persistence και το μοντέλο RF χωρίς τη χρήση του ALO.

Ο Πίνακας 6.2 δείχνει ότι το προτεινόμενο μοντέλο παρουσιάζει καλύτερη απόδοση σε σύγκριση με το μοντέλο persistence και το RF χωρίς ALO, επιτυγχάνοντας χαμηλότερα σφάλματα RMSE, MSE και MAE. Η σύγκριση με το μοντέλο persistence δείχνει ότι το προτεινόμενο μοντέλο επιτυγχάνει ποσοστιαίες βελτιώσεις 3,9% στο RMSE, 8,1% στο MSE, και 4,22% στο MAE. Η χρήση του αλγορίθμου ALO για τη βελτιστοποίηση των υπερπαραμέτρων του RF επιφέρει ακόμα μεγαλύτερες βελτιώσεις σε σχέση με το μη βελτιστοποιημένο RF μοντέλο, καθώς υπάρχει ποσοστιαία βελτίωση 6,92% στο RMSE, 14,13% στο MSE, και 2,7% στο MAE. Στην περίπτωση αυτή, η μικρότερη ποσοστιαία βελτίωση του MAE συγκριτικά με το RMSE και το MSE οφείλεται στο ότι το MAE δεν επηρεάζεται από τις ακραίες τιμές όσο τα άλλα δύο μετρητικά σφάλματος. Συγκεκριμένα, το ALO βοηθά κυρίως στη μείωση των μεγάλων αποκλίσεων, κάτι που αντανακλάται περισσότερο στα RMSE και MSE. Συμπερασματικά, η βελτιστοποίηση των υπερπαραμέτρων μέσω του ALO είναι καθοριστική για το μοντέλο RF. Χωρίς το ALO, το RF έχει χειρότερη επίδοση από το persistence, γεγονός που υποδουλώνει ότι το μοντέλο αδυνατεί να επιτελέσει το

Πίνακας 6.3: Υπολογιστική διάρκεια για την βελτιστοποίηση υπερπαραμέτρων, την εκπαίδευση και την εκτέλεση του μοντέλου RF.

Διαδικασία	Χρόνος (s)
Εύρεση βέλτιστων υπερπαραμέτρων με την χρήση του αλγόριθμου ALO	3.793,4
Εκπαίδευση μοντέλου RF	14,81
Εκτέλεση μοντέλου RF	$1,15 \times 10^{-5}$

ρόλο του αποτελεσματικά. Με το ALO, το μοντέλο αποδίδει καλύτερα από το persistence με μία ποσοστιαία βελτίωση περίπου 4% στο μέσο απόλυτο σφάλμα, σημαντική βελτίωση στα πλαίσια της υπερβραχυπρόθεσμης πρόβλεψης.

Ένας από τους λόγους που επιλέχθηκε να γίνει η πρόβλεψη της Φ/Β παραγωγής βάσει του προτεινόμενου υβριδικού μοντέλου, είναι η ταχύτητα και η χαμηλή υπολογιστική πολυπλοκότητα που χαρακτηρίζει τα RF και τον αλγόριθμο ALO. Ο αλγόριθμος ALO είναι ικανός να βρει τις βέλτιστες υπερπαραμέτρους του μοντέλου πολύ πιο γρήγορα σε σύγκριση με τις εξαντλητικές μεθόδους, όπως την αναζήτηση πλέγματος και την διασταυρούμενη επικύρωση. Η ταχύτητα αυτή είναι κρίσιμη καθώς επιτρέπει ταχύτερες διαδικασίες ανάπτυξης μοντέλων διατηρώντας υψηλά επίπεδα αξιοπιστίας. Η ευελιξία αυτή είναι χρήσιμη σε περιπτώσεις όπου απαιτείται η επαναλαμβανόμενη εκπαίδευση του μοντέλου πρόβλεψης σε καινούρια σετ δεδομένων. Επιπλέον, το μοντέλο RF είναι γνωστό για την ικανότητά του να εκπαιδεύεται γρήγορα πάνω στα σετ δεδομένων και να κάνει προβλέψεις με υψηλή ταχύτητα. Η αποδοτικότητα αυτή καθιστά το RF ιδιαίτερα κατάλληλο για εφαρμογές υπερβραχυπρόθεσμης πρόβλεψης με μεγάλα σετ δεδομένων, ενώ τα μικρά χρονικά περιθώρια για την εξαγωγή των προβλέψεων καθιστούν την ταχύτητα επεξεργασίας ως ζωτικό παράγοντα. Ο Πίνακας 6.3 παρουσιάζει τον χρόνο που απαιτείται για την εύρεση των βέλτιστων υπερπαραμέτρων με τον αλγόριθμο ALO, καθώς και τους χρόνους εκπαίδευσης και εκτέλεσης του μοντέλου RF για το συγκεκριμένο σετ δεδομένων.

Όπως φαίνεται και από τον Πίνακα 6.3 η διαδικασία εύρεσης των βέλτιστων υπερπαραμέτρων μέσω του αλγόριθμου ALO απαιτεί 3.793,4 δευτερόλεπτα (περίπου 1,05 ώρες). Αν η διαδικασία εύρεσης των υπερπαραμέτρων γίνονταν με αναζήτηση πλέγματος (grid search), τότε χονδρικά θα χρειαζόνταν 2,5 μήνες για την εύρεση των βέλτιστων υπερπαραμέτρων. Η αναζήτηση πλέγματος είναι μια εξαντλητική μέθοδος που, θεωρητικά, μπορεί να βρει την καλύτερη δυνατή λύση για τις υπερπαραμέτρους, καθώς εξετάζει συστηματικά όλες τις δυνατές συνδυαστικές τιμές. Ωστόσο, η διαδικασία αυτή απαιτεί σημαντικούς πόρους και πολύ χρόνο, καθιστώντας την μη πρακτική. Για τον λόγο αυτό αξιοποιήθηκε ο αλγόριθμος ALO, ο οποίος παρέχει μια πολύ καλή λύση σε ένα κλάσμα του χρόνου και των πόρων που απαιτούν οι εξαντλητικές μέθοδοι αναζήτησης.

Είναι σημαντικό να αναφερθεί το γεγονός ότι δοκιμάστηκαν και συνδυασμοί παραμέτρων του αλγορίθμου ALO που αντιστοιχούν σε μεγαλύτερους υπολογιστικούς χρόνους, ωστόσο βρέθηκαν παρόμοια αποτελέσματα, υποδηλώνοντας ότι η μία ώρα είναι επαρκής για τη διαδικασία βελτιστοποίησης

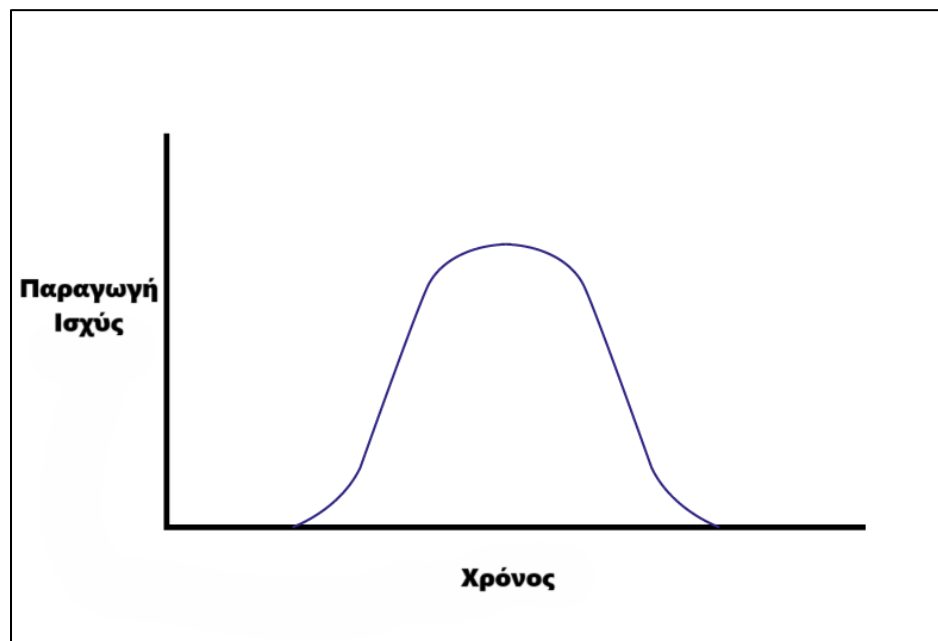
των υπερπαραμέτρων του μοντέλου RF. Επίσης, η εκπαίδευση του μοντέλου RF απαιτεί μόλις 14,81 δευτερόλεπτα, άρα η διαδικασία εκπαίδευσης του μοντέλου είναι εξαιρετικά γρήγορη και μπορεί να επαναληφθεί πολλές φορές χωρίς ιδιαίτερο υπολογιστικό κόστος. Επιπλέον, ο χρόνος που απαιτείται για την εκτέλεση του μοντέλου RF πάνω σε ένα δείγμα είναι περίπου 11,15μs, αποδεικνύοντας ότι το μοντέλο μπορεί να κάνει προβλέψεις σχεδόν σε πραγματικό χρόνο.

6.2.3 Ανάλυση ανά καιρική συνθήκη

Για την περαιτέρω ανάλυση των αποτελεσμάτων του πρώτου σετ δεδομένων, πραγματοποιήθηκε ανάλυση με βάση τις ημερήσιες επικρατούσες καιρικές συνθήκες. Η κατηγοριοποίηση της εισόδου με βάση μια συγκεκριμένη συνθήκη και η ανάλυση των αποτελεσμάτων ανά κατηγορία είναι μια στατιστική διαδικασία που αξιοποιείται για την καλύτερη και λεπτομερέστερη κατανόηση και ερμηνεία των μοτίβων που εμφανίζονται σε χρονοσειρές δεδομένων. Στην περίπτωση της Φ/Β παραγωγής, η ανάλυση ανά καιρική συνθήκη είναι ιδιαίτερα σημαντική καθώς οι επικρατούσες καιρικές συνθήκες, όπως το ποσοστό νεφοκάλυψης και η ηλιοφάνεια, παρουσιάζουν επηρεάζουν άμεσα την απόδοση των Φ/Β συστημάτων, όπως αναφέρθηκε στο Κεφάλαιο 2.

Το πρώτο σετ δεδομένων, μετά από διαδικασία προεπεξεργασίας και καθαρισμού, περιέχει δεδομένα για 43 ημέρες. Από αυτές, το 30%, δηλαδή 13 ημέρες, χρησιμοποιήθηκαν για έλεγχο. Χρησιμοποιώντας ως βάση την ιδεατή χαρακτηριστική ημερήσια καμπύλη της Φ/Β ισχύος, η οποία φαίνεται στο Σχήμα 6.1, οι 13 ημέρες του σετ ελέγχου χωρίστηκαν σε 3 κατηγορίες ημερών:

- Ημέρες με ηλιοφάνεια ή ελάχιστη συννεφιά
- Ημέρες με μερική συννεφιά



Σχήμα 6.1: Ιδεατή καμπύλη ισχύος

- Ημέρες με έντονη συννεφιά

Οι ημέρες που αντιστοιχούν στην πρώτη κατηγορία, δηλαδή οι ημέρες με ηλιοφάνεια ή ελάχιστη συννεφιά, είναι πέντε, ενώ αμφοτέρως οι υπόλοιπες δύο κατηγορίες περιλαμβάνουν τέσσερις μέρες.

Για κάθε καιρική συνθήκη, υπολογίστηκε το μετρητικό σφάλμα RMSE, το οποίο περιγράφεται από την εξίσωση (3.7), για τις αντίστοιχες προβλέψεις που παρήγαγε το προτεινόμενο υβριδικό μοντέλο. Για λόγους πληρότητας, υπολογίστηκε το ίδιο σφάλμα για την πρόβλεψη με το μοντέλο RF χωρίς βελτιστοποίηση, καθώς και για την πρόβλεψη με το μοντέλο persistence. Τα αποτελέσματα αυτών των υπολογισμών, καθώς και οι ποσοστιαίες βελτιώσεις που επιφέρει το προτεινόμενο μοντέλο σε σχέση με τα μοντέλα σύγκρισης, παρουσιάζονται στον Πίνακα 6.4.

Τα αποτελέσματα που παρουσιάζονται στον Πίνακα 6.4 δείχνουν ότι το προτεινόμενο μοντέλο παρουσιάζει καλύτερη απόδοση από το μοντέλο persistence και από το μοντέλο RF χωρίς ALO για όλες τις καιρικές συνθήκες. Ειδικότερα, τα αποτελέσματα στην περίπτωση ηλιοφάνειας είναι καλύτερα από αυτά σε συνθήκες μερικής ή έντονης συννεφιάς. Αυτό είναι αναμενόμενο καθώς σε συνθήκες ηλιοφάνειας υπάρχει λιγότερη μεταβλητότητα κάτι που καθιστά την πρόβλεψη πιο εύκολη. Αντίθετα, σε συνθήκες μερικής ή έντονης συννεφιάς, υπάρχει περισσότερη αστάθεια στην ένταση της προσπίπτουσας ηλιακής ακτινοβολίας καθώς οι καιρικές συνθήκες αλλάζουν απρόβλεπτα, καθιστώντας την πρόβλεψη πιο δύσκολη.

Η σύγκριση με το μοντέλο persistence δείχνει ότι το προτεινόμενο μοντέλο επιτυγχάνει ποσοστιαίες βελτιώσεις 15,48% στην περίπτωση ηλιοφάνειας/ελάχιστης συννεφιάς, 9,68% στην περίπτωση μερικής συννεφιάς, και 8,46% στην περίπτωση έντονης συννεφιάς. Αυτό υποδηλώνει ότι η μεγαλύτερη βελτίωση παρουσιάζεται στις ημέρες με ηλιοφάνεια, κάτι που εξηγείται δεδομένου ότι η εκπαίδευση πάνω σε αυτές τις μέρες είναι ευκολότερη και το σετ δεδομένων σε αυτή την περίπτωση είναι μεγαλύτερο. Έτσι, είναι πιο εύκολο για το προτεινόμενο μοντέλο να προσαρμοστεί στις ημέρες αυτές ώστε να μειώσει το μέσο σφάλμα.

Η σύγκριση με το μοντέλο RF χωρίς ALO δείχνει ότι το προτεινόμενο μοντέλο επιτυγχάνει ποσοστιαίες βελτιώσεις 1,08% στην περίπτωση ηλιοφάνειας/ελάχιστης συννεφιάς, 5,6% στην περίπτωση μερικής συννεφιάς, και 14,8% στην περίπτωση έντονης συννεφιάς. Τα αποτελέσματα αυτά οδηγούν στο συμπέρασμα ότι όσο μεγαλώνει η δυσκολία πρόβλεψης, τόσο μεγαλώνει και η βελτίωση που επιφέρει το ALO. Αυτό δείχνει ότι χωρίς τη βελτιστοποίηση ALO, το RF λειτουργεί ικανοποιητικά για ηλιοφάνεια, αλλά αδυνατεί να μοντελοποιήσει την παραγωγή σε συννεφιασμένες μέρες.

Συνολικά, τα αποτελέσματα υποδεικνύουν πως είναι χρήσιμο να γίνει ξεχωριστή εκπαίδευση ανά καιρική συνθήκη, ώστε να υπάρχει βέλτιστη προσαρμογή του μοντέλου σε κάθε συνθήκη ανεξαρτήτως δυσκολίας και συχνότητας εμφάνισης αυτής της συνθήκης. Με αυτόν τον τρόπο, το μοντέλο θα μπορεί να βελτιστοποιεί τις υπερπαραμέτρους για να προσαρμόζεται στις μεταβαλλόμενες συνθήκες και να αποδίδει καλύτερα σε κάθε συγκεκριμένο σενάριο.

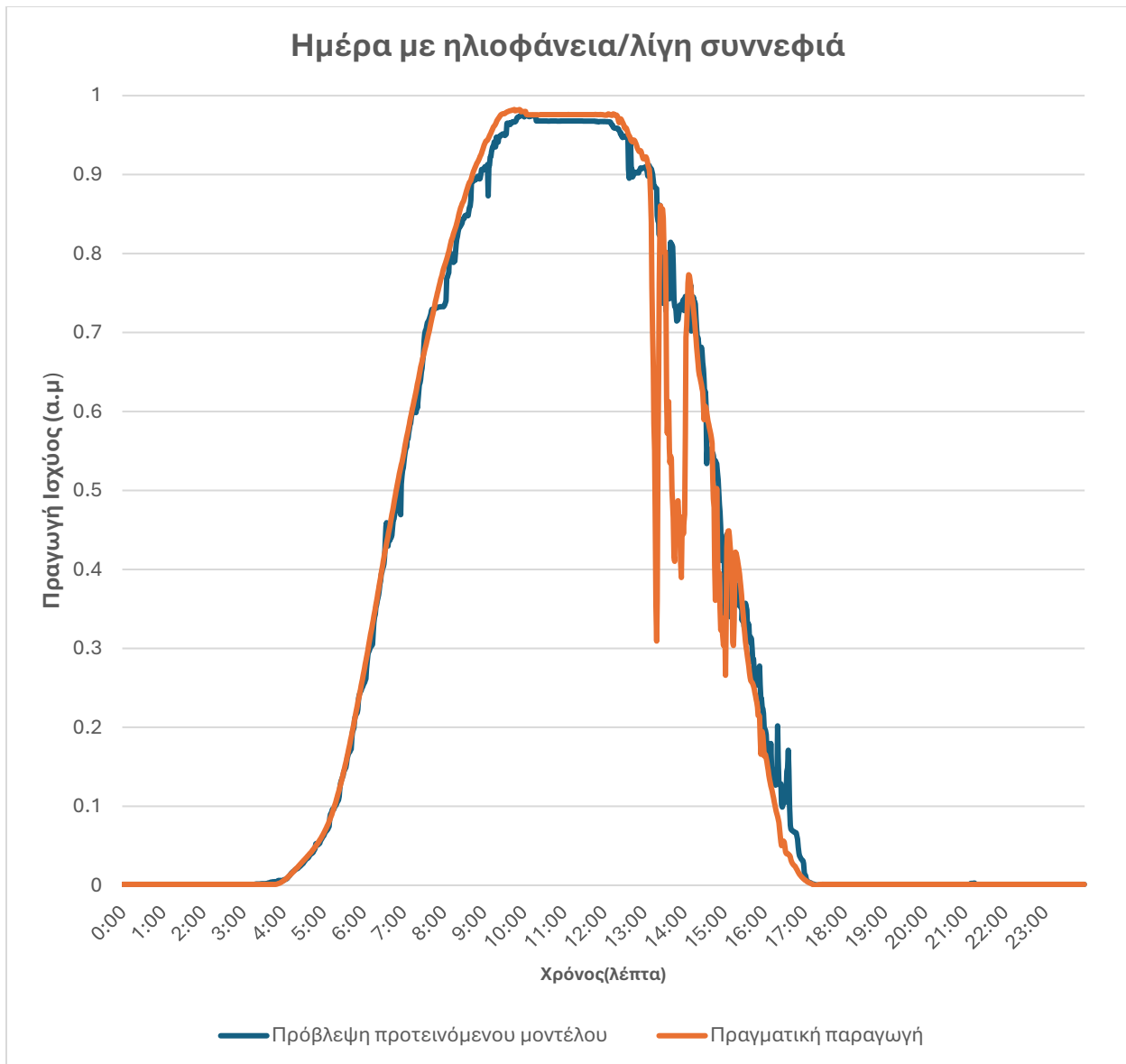
Για την πληρέστερη παρουσίαση των αποτελεσμάτων της ανάλυσης ανά καιρική συνθήκη του πρώτου σετ δεδομένων, παρουσιάζονται ενδεικτικές ημερήσιες καμπύλες Φ/Β ισχύος που προκύπτουν μέσω του προτεινόμενου μοντέλου πρόβλεψης για κάθε διαφορετική καιρική συνθήκη. Οι ημερήσιες καμπύλες πρόβλεψης Φ/Β ισχύος παρέχουν μια λεπτομερή απεικόνιση της απόδοσης του μοντέλου υπό

Πίνακας 6.4: Σφάλμα RMSE για κάθε καιρική συνθήκη

Μέθοδος	Ημερήσιες επικρατούσες καιρικές συνθήκες		
	Ηλιοφάνεια/ελάχιστη συννεφιά	Μερική συννεφιά	Έντονη συννεφιά
	RMSE (α.μ.)		
Random Forest χωρίς ALO	0,06609	0,1180	0,1472
Persistence	0,07734	0,1233	0,1370
Random Forest με ALO (προτεινόμενο)	0,06537	0,1113	0,1254
Ποσοστιαία βελτίωση μεταξύ προτεινόμενου μοντέλου και μοντέλου persistence	15,48%	9,68%	8,46%
Ποσοστιαία βελτίωση μεταξύ προτεινόμενου μοντέλου και RF χωρίς ALO	1,08%	5,6%	14,8%

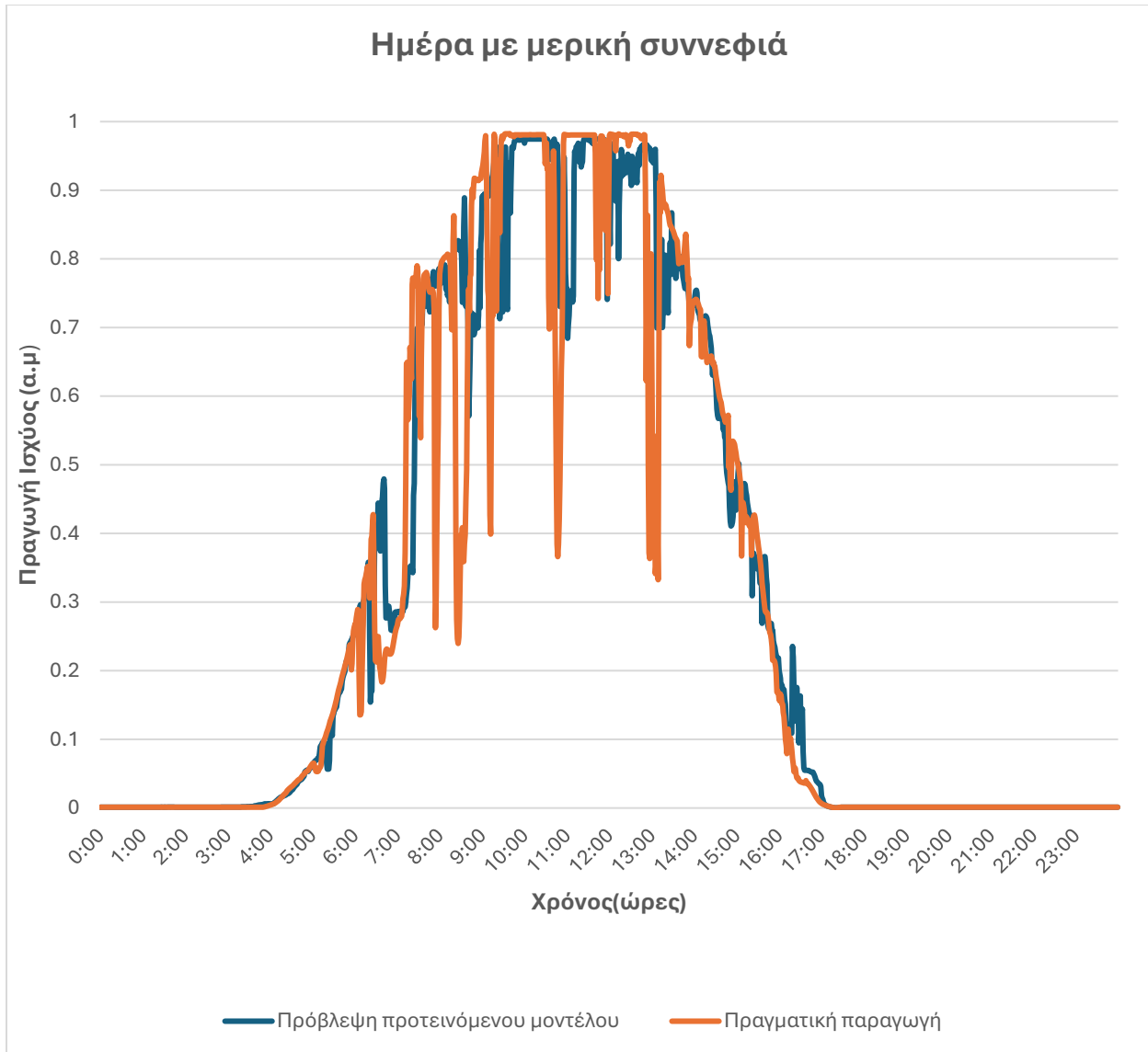
διαφορετικές καιρικές συνθήκες, επιτρέποντας μια σαφή σύγκριση μεταξύ της προβλεπόμενης και της πραγματικής παραγωγής.

Αρχικά, στο Σχήμα 6.2, παρουσιάζονται η πραγματική καμπύλη ισχύος για την περίπτωση ημέρας με ηλιοφάνεια ή λίγη συννεφιά και η καμπύλη που προβλέπει το προτεινόμενο υβριδικό μοντέλο. Η καμπύλη με το μπλε χρώμα δείχνει την προβλεπόμενη παραγωγή ισχύος από το προτεινόμενο μοντέλο ενώ



Σχήμα 6.2: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για ημέρα με ηλιοφάνεια/λίγη συννεφιά

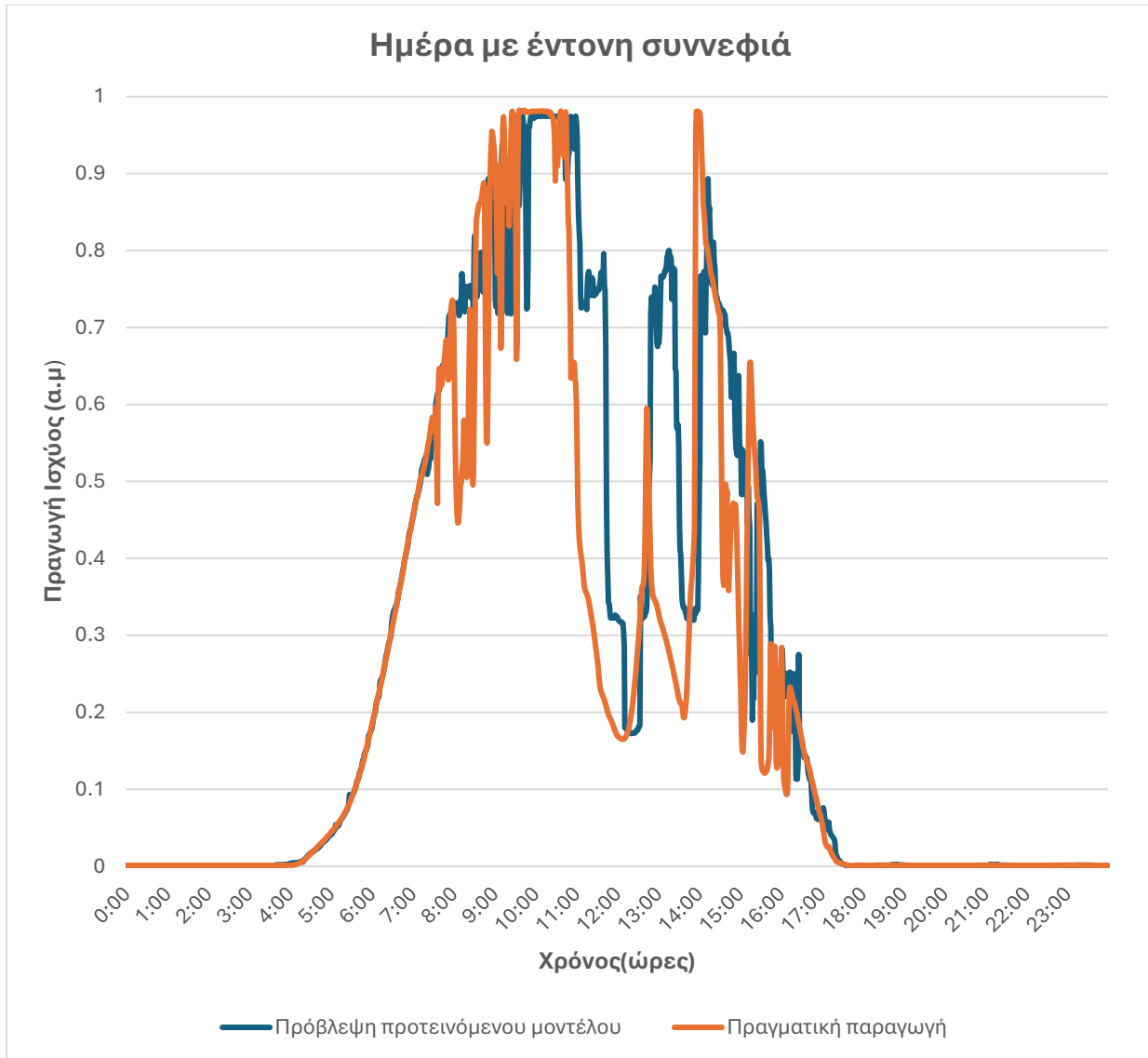
η καμπύλη με το πορτοκαλί χρώμα αντιστοιχεί στην πραγματική παραγωγή. Όπως φαίνεται από στο Σχήμα 6.2, οι δύο καμπύλες σχεδόν συμπίπτουν, υποδεικνύοντας ότι το μοντέλο RF σε συνδυασμό με τον αλγόριθμο ALO αποδίδει με υψηλή ακρίβεια τις προβλέψεις της Φ/Β παραγωγής υπό συνθήκες ηλιοφάνειας. Στην συγκεκριμένη περίπτωση, το μετρητικό σφάλματος RMSE του προτεινόμενου μοντέλου πρόβλεψης έχει τιμή 0,0543 α.μ. Από το Σχήμα 6.2 φαίνεται επίσης πως μέχρι το μεσημέρι η παραγωγή ήταν σχετικά ομαλή, ενώ το απόγευμα εμφανίστηκαν κάποιες διακυμάνσεις, οι οποίες πιθανότατα οφείλονται στη διέλευση νεφώσεων. Επίσης, η καμπύλη προβλεπόμενης παραγωγής ισχύος δείχνει ότι κατά το απόγευμα, το μοντέλο αδυνατεί να προβλέψει με ακρίβεια τις απότομες διακυμάνσεις, κάτι αναμενόμενο καθώς δεν του δίνεται κάποια ένδειξη ότι θα υπάρξει διέλευση νεφώσεων. Αυτό υποδεικνύει τη σημασία χρήσης εικόνων του ουρανού ως είσοδο στο μοντέλο, ώστε να μπορεί να προβλέψει καλύτερα



Σχήμα 6.3: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για ημέρα με μερική συννεφιά

τέτοιες διακυμάνσεις. Επίσης, αναδεικνύει τη σημασία της υψηλής χρονικής ανάλυσης των δεδομένων καθώς αν η ανάλυση των δεδομένων ήταν 15 λεπτά και όχι 1 λεπτό, οι διακυμάνσεις αυτές δεν θα φαινότουσαν.

Στο Σχήμα 6.3 παρουσιάζεται η πραγματική και η προβλεπόμενη καμπύλη ισχύος για την περίπτωση ημέρας με μερική συννεφιά. Η καμπύλη με το μπλε χρώμα δείχνει την προβλεπόμενη παραγωγή ισχύος από το προτεινόμενο μοντέλο ενώ η καμπύλη με το πορτοκαλί χρώμα αντιστοιχεί στην πραγματική παραγωγή. Όπως φαίνεται στο Σχήμα 6.3, οι καμπύλες της προβλεπόμενης παραγωγής και της πραγματικής παραγωγής τείνουν να συμπίπτουν, αν και σε μικρότερο βαθμό σε σχέση με την περίπτωση της ηλιοφάνειας. Αυτό δείχνει ότι το μοντέλο RF με χρήση του αλγορίθμου ALO μπορεί να ακολουθεί τις μεταβολές της ισχύος σε γενικές γραμμές, αλλά η ακρίβεια της πρόβλεψης μειώνεται σε συνθήκες με



Σχήμα 6.4: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για ημέρα με έντονη συννεφιά

μερική συννεφιά, όπου η μεταβλητότητα είναι μεγαλύτερη. Τέλος, στην συγκεκριμένη περίπτωση, το μετρητικό σφάλματος RMSE του προτεινόμενου μοντέλου πρόβλεψης έχει τιμή 0,1055 α.μ.

Στο Σχήμα 6.4 παρουσιάζεται η πραγματική και η προβλεπόμενη καμπύλη ισχύος για την περίπτωση ημέρας με έντονη συννεφιά. Η καμπύλη με το μπλε χρώμα δείχνει την προβλεπόμενη παραγωγή ισχύος από το προτεινόμενο μοντέλο ενώ η καμπύλη με το πορτοκαλί χρώμα αντιστοιχεί στην πραγματική παραγωγή. Όπως φαίνεται στο Σχήμα 6.4, οι καμπύλες της προβλεπόμενης παραγωγής ALO και της πραγματικής παραγωγής συμπίπτουν ακόμη λιγότερο σε σύγκριση με τις περιπτώσεις ηλιοφάνειας και μερικής συννεφιάς. Αυτό δείχνει ότι η προβλεψιμότητα του μοντέλου μειώνεται σημαντικά σε συνθήκες έντονης συννεφιάς, καθώς η μεταβλητότητα και οι απότομες αλλαγές στην προσπίπτουσα ηλιακή ακτινοβολία στα Φ/Β πάνελ καθιστούν την πρόβλεψη πιο δύσκολη. Σε αυτή την περίπτωση, το μετρητικό σφάλματος RMSE του προτεινόμενου μοντέλου πρόβλεψης έχει τιμή 0,1437 α.μ. Πέρα από αυτά,

συγκρίνοντας τις δύο καμπύλες, φαίνεται να υπάρχει μια μετατόπιση στον άξονα x' , η οποία ισοδυναμεί περίπου με τον χρονικό ορίζοντα πρόβλεψης. Επομένως, είναι ακόμα πιο φανερή η ανάγκη για εμπλουτισμό της εισόδου με εικόνες, ώστε να προβλέπονται έγκαιρα οι διακυμάνσεις λόγω διέλευσης νεφώσεων και να εξαλειφθεί η χρονική υστέρηση στον οριζόντιο άξονα. Ωστόσο, πέρα από τη χρονική καθυστέρηση, η προσομοίωση της πραγματικής καμπύλης είναι αρκετά καλή και υποδεικνύει τις δυνατότητες του RF, ειδικά αν υποβοηθηθεί με καλύτερη είσοδο (εικόνες του ουρανού) και αν εκπαιδευτεί ανά καιρική συνθήκη.

Συμπερασματικά, από τα Σχήματα 6.2, 6.3, και 6.4, διαφαίνεται ότι το μοντέλο RF σε συνδυασμό με τον αλγόριθμο ALO λειτουργεί εξαιρετικά αποτελεσματικά σε συνθήκες ηλιοφάνειας ή ελάχιστης συννεφιάς, με τις καμπύλες της προβλεπόμενης και πραγματικής παραγωγής να συμπίπτουν σχεδόν απόλυτα. Ωστόσο, στην περίπτωση μερικής ή έντονης συννεφιάς όπου οι καιρικές συνθήκες γίνονται πιο μεταβλητές, η ακρίβεια της πρόβλεψης μειώνεται, παρουσιάζοντας μεγαλύτερες αποκλίσεις μεταξύ των καμπυλών πραγματικής και προβλεπόμενης παραγωγής. Αυτό υποδεικνύει τη δυσκολία που θέτουν οι μεταβλητές καιρικές συνθήκες στην πρόβλεψη της Φ/Β παραγωγής και την ανάγκη για τροφοδότηση επιπλέον πληροφοριών στο μοντέλο πρόβλεψης, όπως εικόνες του ουρανού, προκειμένου να βελτιωθεί η ακρίβεια των προβλέψεων.

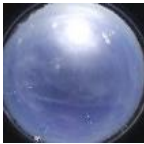
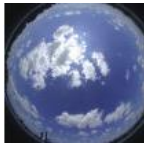
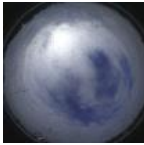
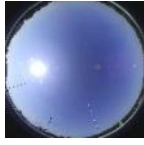
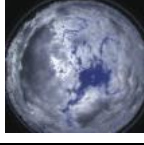

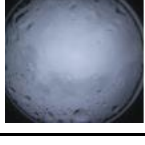
6.3 ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΥΤΕΡΗΣ ΠΕΡΙΠΤΩΣΗΣ ΠΡΟΣ ΜΕΛΕΤΗ

Το δεύτερο σύνολο δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση και την αξιολόγηση του μοντέλου πρόβλεψης της Φ/Β παραγωγής περιλαμβάνει επίγειες εικόνες του ουρανού. Για την καλύτερη πρόβλεψη της Φ/Β παραγωγής, οι εικόνες κατηγοριοποιήθηκαν σε επτά κατηγορίες, ανάλογα με το είδος της νέφωσης που επικρατούσε, χρησιμοποιώντας μια προκαθορισμένη κατηγοριοποίηση που προήλθε από εξωτερική πηγή[11]. Για κάθε κατηγορία έγινε ξεχωριστή εκπαίδευση επιτρέποντας την ακριβή προσαρμογή στις διαφορετικές συνθήκες νεφοκάλυψης. Η κατηγοριοποίηση των εισερχόμενων δεδομένων συμβάλλει στη βελτίωση της πρόβλεψης με τους εξής τρόπους:

- Μειώνει την πολυπλοκότητα των δεδομένων εισόδου, αφού αντί το μοντέλο να επεξεργάζεται κάθε εικόνα ως μοναδική πληροφορία, αναγνωρίζει συγκεκριμένες κατηγορίες, καθιστώντας την εκπαίδευση πιο αποδοτική.
- Επιτρέπει στο μοντέλο RF να γίνει πιο ευέλικτο και ικανό να προβλέπει την παραγωγή ενέργειας σε διάφορες καιρικές συνθήκες.

Οι κλάσεις των νεφώσεων που χρησιμοποιήθηκαν παρουσιάστηκαν στην Ενότητα 4.5. Το πλήθος των δεδομένων εικόνων που αντιστοιχούν σε κάθε κλάση, καθώς και μια ενδεικτική εικόνα για κάθε κλάση, παρουσιάζονται στον Πίνακα 6.5.

Πίνακας 6.5: Κλάσεις νεφώσεων του δεύτερου σετ δεδομένων

Όνομα κλάσης	Πλήθος εικόνων	Ενδεικτική εικόνα
Cirrus	2.726	
Cumulus	3.131	
Cirrocumulus	856	
Clear sky	1.271	
Stratocumulus	7.400	
Stratus	471	
Nimbostratus	2.511	

Από τις κατηγορίες αυτές, η κλάση Stratocumulus έχει τα περισσότερα δεδομένα με 7.400 εικόνες, γεγονός που την καθιστά την πιο αντιπροσωπευτική κατηγορία στο σύνολο δεδομένων. Αντίθετα, η κλάση Stratus έχει τα λιγότερα δεδομένα, με μόλις 471 εικόνες, καθιστώντας την ως την πιο σπάνια κατηγορία. Οι υπόλοιπες κατηγορίες νεφώσεων κυμαίνονται στο διάστημα των 800-3.200 εικόνων. Αυτή η διακύμανση στο μέγεθος της κάθε κλάσης θα πρέπει να ληφθεί υπόψη κατά την εκπαίδευση και την αξιολόγηση του προτεινόμενου μοντέλου καθώς και στην αποδοτικότητα του αλγορίθμου ALO.

6.3.1 Εύρεση αποδοτικότερων υπερπαραμέτρων

Οι βέλτιστες τιμές των υπερπαραμέτρων που προέκυψαν από τον αλγόριθμο ALO και θα χρησιμοποιηθούν για τη ρύθμιση του μοντέλου πρόβλεψης RF για κάθε κλάση νεφώσεων στο δεύτερο σετ

Πίνακας 6.6: Βέλτιστες τιμές των υπερπαραμέτρων του προτεινόμενου μοντέλου πρόβλεψης για κάθε κλάση

Όνομα κλάσης	Αριθμός δέντρων	Μέγιστο βάθος	Ελάχιστος αριθμός δειγμάτων για διάσπαση	Ελάχιστος αριθμός δειγμάτων ανά φύλλο	Μέγιστος αριθμός χαρακτηριστικών
Cirrus	261	6	26	16	4
Cumulus	274	6	23	8	5
Cirrocumulus	360	6	24	10	5
Clear sky	316	6	26	9	5
Stratocumulus	400	8	30	12	6
Stratus	370	7	22	6	6
Nimbostratus	419	6	36	4	6

δεδομένων παρουσιάζονται στον Πίνακα 6.6. Οι υπερπαραμέτροι αυτές βρέθηκαν με σκοπό να διασφαλιστεί η καλύτερη δυνατή απόδοση του μοντέλου πρόβλεψης για κάθε κατηγορία νεφώσεων.

Όπως φαίνεται από τον Πίνακα 6.6, υπάρχει μεγάλη ποικιλία στις τιμές των υπερπαραμέτρων για κάθε κλάση. Ο αριθμός των δέντρων κυμαίνεται μεταξύ των 250 και των 420 για κάθε κλάση, υποδεικνύοντας ότι η επιλογή του εύρους αναζήτησης για αυτήν την παράμετρο είναι κατάλληλη. Είναι επίσης σημαντικό να σημειωθεί ότι η κλάση Nimbostratus έχει το μεγαλύτερο πλήθος δέντρων, με τιμή 419, πιθανώς λόγω της μεγαλύτερης πολυπλοκότητας των προβλέψεων σε αυτήν την κατηγορία. Επιπρόσθετα, το μέγιστο βάθος των δέντρων σε όλες τις κλάσεις είναι μεγαλύτερο από 6, υποδεικνύοντας ότι τα 6 επίπεδα σε κάθε δέντρο είναι απαραίτητα για να αντιμετωπιστούν ορισμένες πολύπλοκες συνθήκες. Επίπλέον, οι τιμές για τον αριθμό δειγμάτων για διάσπαση και ο αριθμός δειγμάτων ανά φύλλο παρουσιάζει μεγάλη μεταβλητότητα. Συνοπτικά, οι επιλεγμένες τιμές υπερπαραμέτρων φαίνεται να καλύπτουν αποτελεσματικά τις ανάγκες του μοντέλου, επιτρέποντας την εύρεση βέλτιστων προσαρμογών για κάθε κατηγορία νεφών.

Πίνακας 6.7: Σφάλμα RMSE (α.μ.) για κάθε κλάση νέφωσης

Όνομα κλάσης	RF χωρίς ALO	RF με ALO (προτεινόμενο)	Persistence μοντέλο	Ποσοστιαία βελτίωση μεταξύ προτεινόμενου μοντέλου και persistence	Ποσοστιαία βελτίωση μεταξύ προτεινόμενου μοντέλου και RF χωρίς ALO
	RMSE (α.μ.)				
Cirrus	0,195	0,1911	0,2677	29%	2%
Cumulus	0,121	0,116	0,1269	9%	4,1%
Cirrocumulus	0,182	0,167	0,2136	22%	8,2%
Clear sky	0,04	0,037	0,06149	40%	7.5%
Stratocumulus	0,155	0,1518	0,1967	23%	2,1%
Stratus	0,096	0,0959	0,0998	4%	0,1%
Nimbostratus	0,108	0,106	0,1157	8%	1,8%

6.3.2 Συγκεντρωτικά αποτελέσματα

Με τη χρήση των υπερπαραμέτρων που βρέθηκαν μέσω του αλγορίθμου ALO πραγματοποιήθηκε η εκτέλεση του μοντέλου RF για την πρόβλεψη της Φ/Β παραγωγής. Τα αποτελέσματα του μοντέλου RF σε συνδυασμό με τον αλγόριθμο ALO παρουσιάζονται στον Πίνακα 6.7, όπου φαίνεται η ακρίβεια και η αποτελεσματικότητα της πρόβλεψης για τις επτά κλάσεις νεφώσεων. Για να αξιολογηθεί η απόδοση του προτεινόμενου μοντέλου, πραγματοποιήθηκε σύγκριση με άλλα δύο μοντέλα, το μοντέλο persistence και το μοντέλο RF χωρίς τη χρήση του ALO.

Όπως φαίνεται και από τον Πίνακα 6.7, το προτεινόμενο μοντέλο παρουσιάζει το μικρότερο σφάλμα RMSE για την κλάση Clear sky (0,037 α.μ.). Η κλάση Clear sky παρουσιάζει το μικρότερο σφάλμα επειδή στην περίπτωση αυτή ο ουρανός είναι καθαρός, χωρίς σύννεφα ή άλλες δυναμικά μεταβαλλόμενες

συνθήκες, καθιστώντας την εκπαίδευση του μοντέλου πιο εύκολη και ακριβή. Από την άλλη πλευρά, το μεγαλύτερο σφάλμα παρουσιάζει η κλάση Cirrus με RMSE ίσο με 0,1911 α.μ. Αυτό οφείλεται στην πολυπλοκότητα και τη μεταβλητότητα που χαρακτηρίζει τα σύννεφα τύπου Cirrus, γεγονός που καθιστά την πρόβλεψη πιο δύσκολη. Πέρα από αυτά, ο Πίνακας 6.7 δείχνει ότι το προτεινόμενο μοντέλο παρουσιάζει και στις επτά κλάσεις νεφώσεων καλύτερη απόδοση σε σύγκριση με το μοντέλο persistence και το RF χωρίς ALO, επιτυγχάνοντας χαμηλότερα σφάλματα RMSE.

Η μεγαλύτερη ποσοστιαία βελτίωση μεταξύ του προτεινόμενου μοντέλου και του μοντέλου persistence παρατηρείται για την κλάση Clear sky με βελτίωση 40%, ενώ η ποσοστιαία βελτίωση μεταξύ του προτεινόμενου μοντέλου και του μοντέλου RF χωρίς ALO είναι 7,5%. Η μικρότερη ποσοστιαία βελτίωση μεταξύ του προτεινόμενου μοντέλου και του μοντέλου persistence παρατηρείται για την κλάση Stratus με βελτίωση 4%. Αντίστοιχα, η βελτίωση του προτεινόμενου μοντέλου σε σχέση με το μοντέλο RF χωρίς ALO για αυτήν την κλάση είναι σχεδόν μηδαμινή (0,1%). Η σχεδόν μηδαμινή βελτίωση υποδηλώνει ότι οι βελτιστοποιημένες υπερπαραμέτροι δεν προσφέρουν σημαντική βελτίωση σε αυτή την κλάση, πιθανόν επειδή το βασικό μοντέλο RF χωρίς ALO ήδη αποδίδει πολύ καλά σε συνθήκες πλήρους νεφοκάλυψης όπου οι μεταβολές στην κατάσταση του ουρανού είναι λίγες και η παραγωγή σχεδόν μηδαμινή.

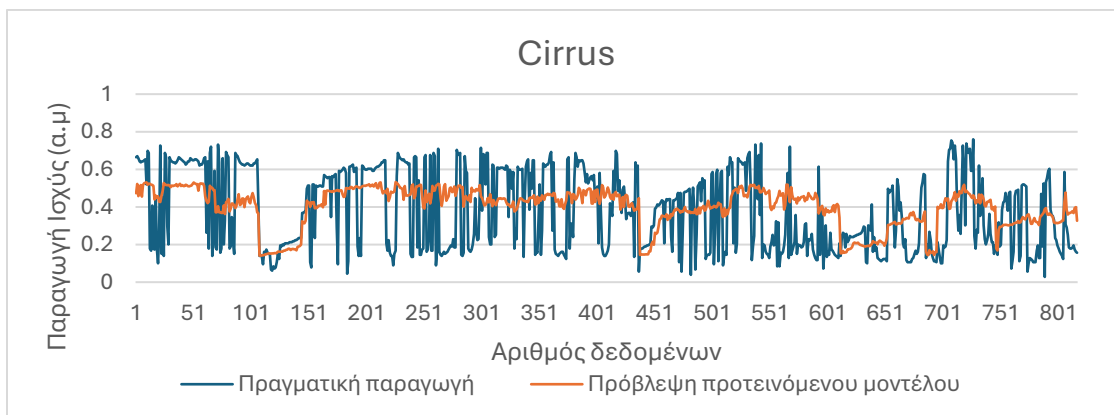
Στις υπόλοιπες κλάσεις υπάρχει μια βελτίωση μεταξύ του προτεινόμενου μοντέλου και του μοντέλου persistence εύρους από 8% έως 29%. Η αντίστοιχη βελτίωση μεταξύ του προτεινόμενου μοντέλου και του μοντέλου RF χωρίς ALO κυμαίνεται από 1,8% έως 8,2%. Αυτές οι βελτιώσεις δείχνουν ότι μέσω του ALO το μοντέλο προσαρμόζεται καλύτερα στις πιο μεταβλητές συνθήκες αυτών των κλάσεων, βελτιώνοντας την πρόβλεψη σε σχέση με το μοντέλο persistence. Η μικρότερη βελτίωση σε σχέση με το βασικό μοντέλο RF δείχνει ότι η βελτιστοποίηση των υπερπαραμέτρων μέσω του ALO προσφέρει επιπλέον πλεονεκτήματα, βελτιώνοντας την ακρίβεια της πρόβλεψης και μειώνοντας το σφάλμα πρόβλεψης.

Όπως φαίνεται και από τον Πίνακα 6.8, η διαδικασία εύρεσης των βέλτιστων υπερπαραμέτρων μέσω του αλγόριθμου ALO κυμαίνεται από 3,5 έως 23 ώρες, με την κλάση Stratocumulus να έχει τον μεγαλύτερο χρόνο και την κλάση Stratus τον μικρότερο. Η διαδικασία εύρεσης των βέλτιστων παραμέτρων για κάθε κλάση ποικίλει λόγω της διαφοράς στο μέγεθος των δεδομένων της κάθε κλάσης. Ωστόσο, πρέπει να τονιστεί πως η διαδικασία υλοποιείται εξαιρετικά γρήγορα σε σχέση με άλλες μεθόδους σαν αυτή της αναζήτησης πλέγματος. Επίσης, η εκπαίδευση των μοντέλων RF απαιτούν πολύ λίγο χρόνο, με τους χρόνους να κυμαίνονται από 1 έως 9,5 δευτερόλεπτα, άρα η διαδικασία εκπαίδευσης του μοντέλου είναι εξαιρετικά γρήγορη και μπορεί να επαναληφθεί πολλές φορές χωρίς ιδιαίτερο υπολογιστικό κόστος. Επιπλέον, ο χρόνος που απαιτείται για την εκτέλεση των μοντέλων RF πάνω σε ένα δείγμα είναι σε κάθε περίπτωση στην τάξη των ms, αποδεικνύοντας ότι το μοντέλο μπορεί να κάνει προβλέψεις σχεδόν σε πραγματικό χρόνο για κάθε κλάση νέφωσης.

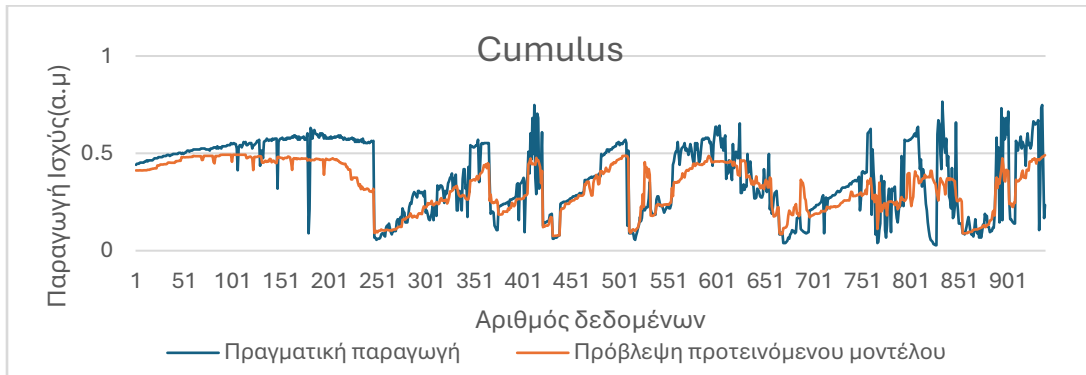
Για την πληρέστερη παρουσίαση των αποτελεσμάτων του δεύτερου σετ δεδομένων, παρουσιάζεται για κάθε μία από τις επτά κλάσεις η καμπύλη της προβλεπόμενης παραγωγής ισχύος μέσω του προτεινόμενου μοντέλου πρόβλεψης σε σύγκριση με την πραγματική παραγωγή. Οι καμπύλες αυτές φαίνονται στα Σχήματα 6.5 έως 6.11.

Πίνακας 6.8: Χρόνοι που απαιτούνται για το μοντέλο RF σε συνδυασμό με τον αλγόριθμο ALO

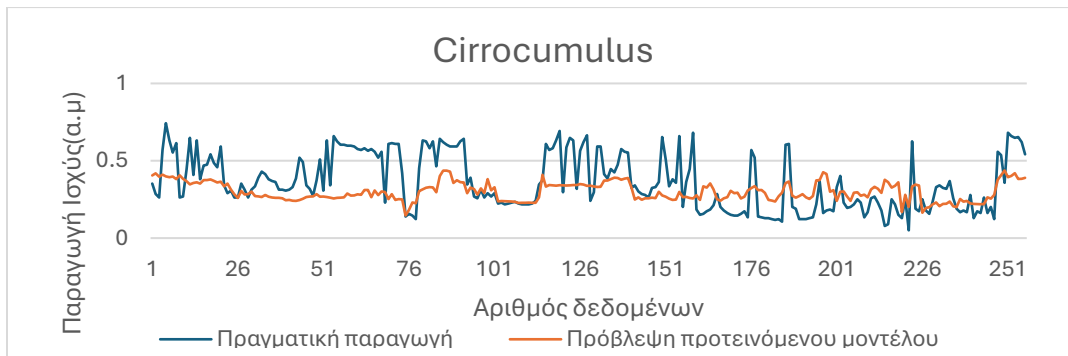
Χρόνος εκτέλεσης διαδικασίας (s)			
Όνομα κλάσης	Βελτιστοποίηση υπερπαραμέτρων με χρήση ALO	Εκπαίδευση μοντέλου RF	Εκτέλεση μοντέλου RF
Cirrus	30945	4,502	$1,98 \times 10^{-5}$
Cumulus	39381	4,852	$2,14 \times 10^{-5}$
Cirrocumulus	14222	1,936	$5,49 \times 10^{-5}$
Clear sky	19281	2,046	$3,7 \times 10^{-5}$
Stratocumulus	85547	9,463	$2,55 \times 10^{-5}$
Stratus	12537	1,277	$8,49 \times 10^{-5}$
Nimbostratus	28477	3,934	$1,99 \times 10^{-5}$



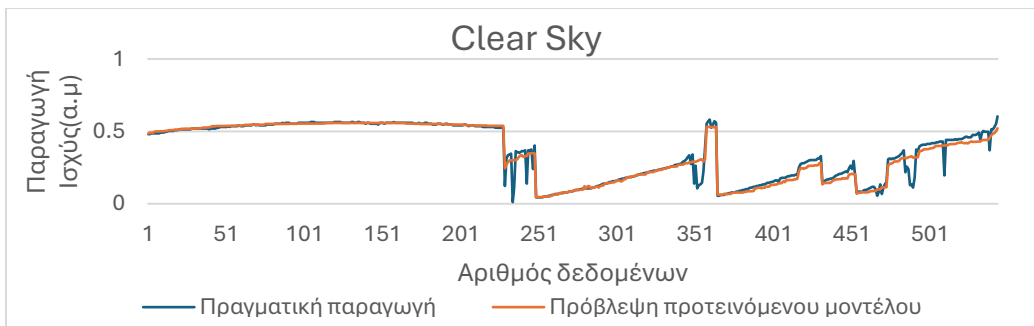
Σχήμα 6.5: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για την κλάση Cirrus.



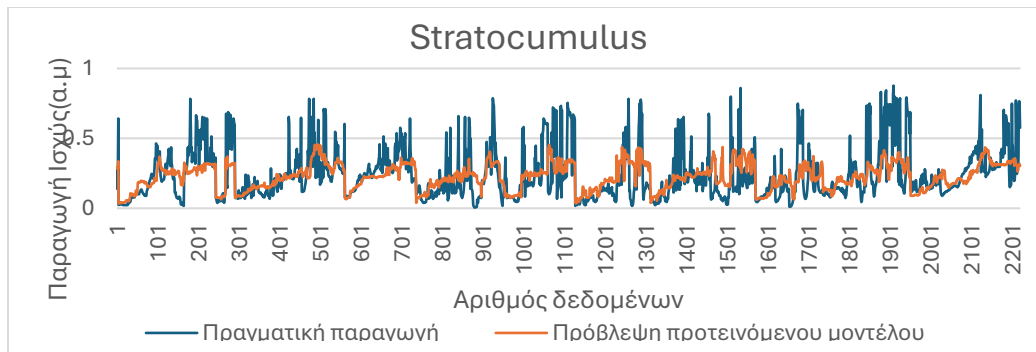
Σχήμα 6.6: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για την κλάση Cumulus.



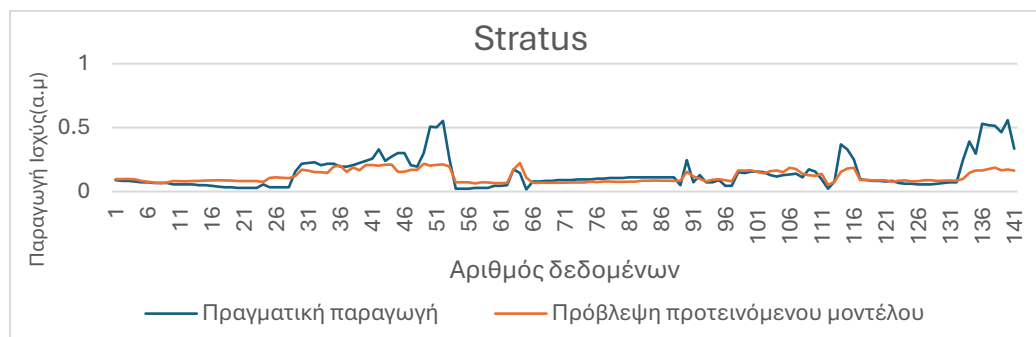
Σχήμα 6.7: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για την κλάση Cirrocumulus.



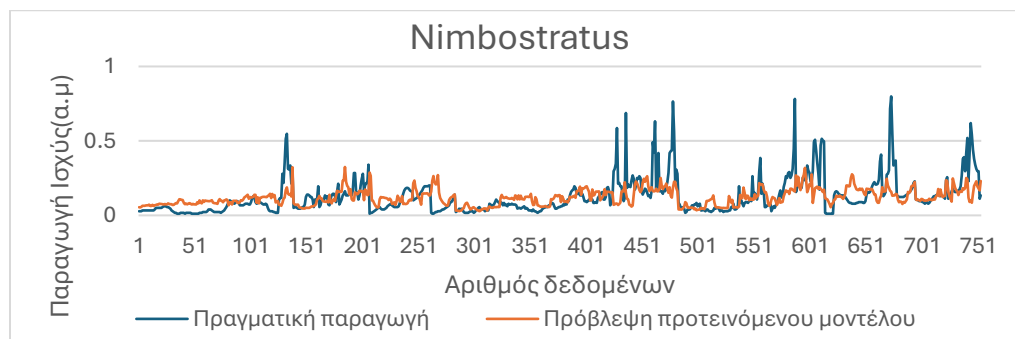
Σχήμα 6.8: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για την κλάση Clear sky.



Σχήμα 6.9: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για την κλάση Stratocumulus.



Σχήμα 6.10: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για την κλάση Stratus.



Σχήμα 6.11: Καμπύλη προβλεπόμενης παραγωγής ισχύος σε σύγκριση με την πραγματική παραγωγή για την κλάση Nimbostratus.

Στα Σχήματα 6.5 έως 6.11, η καμπύλη με το πορτοκαλί χρώμα δείχνει την προβλεπόμενη παραγωγή ισχύος από το προτεινόμενο μοντέλο, ενώ η καμπύλη με το μπλε χρώμα αντιστοιχεί στην πραγματική παραγωγή. Τα σχήματα 6.5, 6.7 και 6.9, που αντιστοιχούν στις κλάσεις Cirrus, Cirrocumulus και Stratocumulus αντίστοιχα, δείχνουν ότι οι καμπύλες της προβλεπόμενης παραγωγής και της πραγματικής παραγωγής συμπίπτουν λιγότερο σε σύγκριση με τις υπόλοιπες κλάσεις. Αυτό είναι αναμενόμενο, καθώς αυτές οι κλάσεις χαρακτηρίζονται από μεγαλύτερη πολυπλοκότητα και μεταβλητότητα λόγω των νεφώσεων που υπάρχουν σε αυτές τις κατηγορίες. Αντίθετα, στις κλάσεις Cumulus (Σχήμα 6.6), Stratus (Σχήμα 6.10) και ιδιαίτερα στην κλάση Clear Sky (Σχήμα 6.8), οι δύο

καμπύλες σχεδόν συμπίπτουν. Αυτό υποδεικνύει ότι το μοντέλο RF σε συνδυασμό με τον αλγόριθμο ALO αποδίδει με υψηλή ακρίβεια τις προβλέψεις της Φ/Β παραγωγής υπό αυτές τις συνθήκες.

Συμπερασματικά, η ανάλυση των καμπύλων της προβλεπόμενης και της πραγματικής παραγωγής ισχύος για τις επτά κλάσεις νεφώσεων δείχνει ότι το προτεινόμενο μοντέλο πρόβλεψης αποδίδει με ικανοποιητική ακρίβεια στις περισσότερες κλάσεις. Οι κλάσεις με υψηλότερη πολυπλοκότητα και μεταβλητότητα, όπως οι κλάσεις Cirrus, Cirrocumulus, και Stratocumulus, υπογραμμίζουν την ανάγκη για περαιτέρω βελτιώσεις στο μοντέλο για την αντιμετώπιση αυτών των πιο σύνθετων συνθηκών.

ΣΥΜΠΕΡΑΣΜΑΤΑ

7.1 ΣΥΝΟΨΗ ΕΡΓΑΣΙΑΣ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα διπλωματική εργασία αναπτύχθηκε ένα μοντέλο μηχανικής μάθησης RF που αξιοποιήθηκε για παλινδρόμηση με σκοπό την πρόβλεψη Φ/Β παραγωγής. Συγκεκριμένα, το μοντέλο RF επικεντρώθηκε στην υπερβραχυπρόθεσμη πρόβλεψη, με χρονικό ορίζοντα πρόβλεψης 15 λεπτών, μια κατηγορία πρόβλεψης που αποκτά όλο και μεγαλύτερη σημασία για τους διαχειριστές των δικτύων και τους παραγωγούς, για τη διαχείριση της ενεργειακής παραγωγής και κατανάλωσης. Για τη βελτιστοποίηση του μοντέλου RF χρησιμοποιήθηκε ο αλγόριθμος ALO, ένας μετεωρετικός αλγόριθμος που είναι γρήγορος και αποτελεσματικός στην εύρεση ολικά βέλτιστων λύσεων. Συγκεκριμένα, ο αλγόριθμος αυτός αξιοποιήθηκε για τη βελτιστοποίηση των πέντε υπερπαραμέτρων του μοντέλου RF. Με τη χρήση αυτού του αλγορίθμου το μοντέλο RF προσαρμόστηκε καλύτερα στις μεταβαλλόμενες καιρικές συνθήκες και εξασφαλίστηκε ότι έκανε την ακριβέστερη δυνατή πρόβλεψη της Φ/Β παραγωγής.

Για την ολοκληρωμένη και αμερόληπτη αξιολόγηση των προβλέψεων του προτεινόμενου μοντέλου RF σε συνδυασμό με τον αλγόριθμο ALO, πραγματοποιήθηκε σύγκριση των αποτελεσμάτων του με τις προβλέψεις δύο άλλων μοντέλων: το μοντέλο RF χωρίς τον αλγόριθμο ALO και το μοντέλο persistence. Η σύγκριση αυτή ήταν απαραίτητη για να διαπιστωθεί η πραγματική αποδοτικότητα και αξιοπιστία του προτεινόμενου μοντέλου σε σχέση με άλλες μεθόδους πρόβλεψης. Ως μέτρα αξιολόγησης για την αποδοτικότητα του προτεινόμενου μοντέλου αξιοποιήθηκαν τρία μετρητικά σφάλματος, το RMSE, το MAE, και το MSE, ώστε οι προβλέψεις να αξιολογηθούν από πολλαπλές οπτικές γωνίες.

Για την εκπαίδευση και αξιολόγηση των μοντέλων χρησιμοποιήθηκαν δύο σετ δεδομένων. Το πρώτο σετ δεδομένων περιείχε αποκλειστικά αριθμητικά δεδομένα για 43 ημέρες Φ/Β παραγωγής, περιλαμβάνοντας χαρακτηριστικά όπως η ηλιακή ακτινοβολία, η θερμοκρασία των Φ/Β πάνελ και άλλα συναφή δεδομένα. Το δεύτερο σετ δεδομένων περιείχε, πέρα από αριθμητικά δεδομένα, και εικόνες του ουρανού. Για το πρώτο σετ δεδομένων, πέρα από τη συνολική ανάλυση των αποτελεσμάτων, πραγματοποιήθηκε ανάλυση ανά καιρική συνθήκη. Συγκεκριμένα, τα δεδομένα ελέγχου χωρίστηκαν σε τρεις κατηγορίες: ημέρες με ηλιοφάνεια, ημέρες με μερική συννεφιά και ημέρες με ολική συννεφιά. Αυτό έγινε για την εξέταση σε μεγαλύτερο βάθος της αποδοτικότητας του προτεινόμενου μοντέλου, αλλά και για την επισήμανση των πλεονεκτημάτων και των αδυναμιών του προτεινόμενου μοντέλου υπό τις διάφορες καιρικές συνθήκες. Στην περίπτωση του δεύτερου σετ δεδομένων, οι εικόνες κατηγοριοποιήθηκαν σε επτά κατηγορίες, ανάλογα με το είδος της νέφωσης που επικρατούσε. Για κάθε κατηγορία έγινε ξεχωριστή εκπαίδευση, επιτρέποντας την ακριβή προσαρμογή στις διαφορετικές συνθήκες νέφωσης. Το προτεινόμενο μοντέλο εκτελέστηκε και στις επτά κλάσεις νεφώσεων, παρέχοντας έτσι μια λεπτομερή αξιολόγηση της αποδοτικότητάς του όταν λαμβάνονται υπόψη και δεδομένα εικόνων. Και στις δύο υπό εξέταση περιπτώσεις, το μοντέλο RF εκπαιδεύτηκε με πραγματικά ιστορικά δεδομένα και υπό ρεαλιστικές συνθήκες. Αυτή η προσέγγιση εξασφάλισε ότι το μοντέλο ήταν ικανό να παρέχει αξιόπιστες

προβλέψεις σε πραγματικές συνθήκες, ενισχύοντας έτσι τη σημασία και την πρακτική εφαρμογή του προτεινόμενου μοντέλου μηχανικής μάθησης.

Ο αλγόριθμος ALO που χρησιμοποιήθηκε για τη βελτιστοποίηση των υπερπαραμέτρων του μοντέλου RF πέτυχε τον στόχο του σε πολύ μικρότερους χρόνους σε σύγκριση με τις εξαντλητικές μεθόδους. Με τις κατάλληλες υπερπαραμέτρους, το προτεινόμενο υβριδικό μοντέλο παρουσίασε υψηλότερη ακρίβεια πρόβλεψης σε σύγκριση με το απλό μοντέλο RF χωρίς ALO και το μοντέλο persistence, βάσει όλων των υπολογιζόμενων μετρητικών σφαλμάτων (RMSE, MAE, MSE) σε όλες τις περιπτώσεις. Συμπερασματικά, υπήρχε βελτιωμένη απόδοση με μικρό κόστος σε χρόνο για τη βελτιστοποίηση των υπερπαραμέτρων. Επιπλέον, τα καλά αποτελέσματα του προτεινόμενου μοντέλου δείχνουν ότι η υψηλή χρονική ανάλυση των δεδομένων επιτρέπει λεπτομερέστερη απεικόνιση της πραγματικής παραγωγής. Μέσω της ανάλυσης ανά καιρική συνθήκη του πρώτου σετ δεδομένων, αναδείχθηκε ότι το μοντέλο δυσκολεύεται να μοντελοποιήσει την παραγωγή σε συνεφασμένες μέρες, αναδεικνύοντας τη σημασία χρήσης εικόνων ως είσοδο και την εκπαίδευση ανά καιρική συνθήκη. Στην περίπτωση του δεύτερου σετ δεδομένων, το προτεινόμενο μοντέλο παρουσίασε συνολικά καλύτερη απόδοση σε σχέση με τα δύο προαναφερθέντα μοντέλα για τις επτά κλάσεις νεφώσεων, ωστόσο υπάρχουν περιθώρια βελτίωσης.

7.2 ΠΡΟΤΑΣΕΙΣ ΜΕΛΛΟΝΤΙΚΗΣ ΕΠΕΚΤΑΣΗΣ

Η περαιτέρω βελτίωση της αποδοτικότητας του προτεινόμενου μοντέλου είναι εφικτή μέσω της υποβολής του σε ορισμένες τροποποιήσεις. Οι πιο σημαντικές πιθανές τροποποιήσεις, οι οποίες είτε δεν ήταν άμεσα εφαρμόσιμες, είτε δεν συνάδουν με τις κύριες επιδιώξεις της παρούσας εργασίας, παρουσιάζονται παρακάτω:

- Αντικατάσταση του μοντέλου RF με κάποιο μοντέλο βαθιάς μάθησης (deep learning): Τα μοντέλα βαθιάς μάθησης, όπως τα δίκτυα Long Short-Term Memory (LSTM) και τα συνελκτικά (convolutional) νευρωνικά δίκτυα, έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά στην επεξεργασία μεγάλων όγκων δεδομένων και στην προσαρμογή σε πολύπλοκες και μεταβαλλόμενες συνθήκες. Ως επακόλουθο, η χρήση ενός τέτοιου μοντέλου θα μπορούσε να επιτρέψει την καλύτερη εκμετάλλευση των πολυδιάστατων δεδομένων και των σχέσεων μεταξύ τους, βελτιώνοντας έτσι την ακρίβεια των προβλέψεων της Φ/Β παραγωγής.
- Λεπτομερέστερη ανάλυση ανά καιρική συνθήκη: Η λεπτομερέστερη ανάλυση και η εκπαίδευση ανά συγκεκριμένες καιρικές συνθήκες θα μπορούσε να βελτιώσει σημαντικά την αποδοτικότητα του μοντέλου στην πρόβλεψη της Φ/Β παραγωγής. Με τη διάκριση σε περισσότερες κατηγορίες καιρικών συνθηκών, όπως διαφορετικά είδη νεφώσεων, διαφορετικές τιμές θερμοκρασίας ή ανέμου, το μοντέλο θα μπορούσε να μάθει να προσαρμόζεται καλύτερα στις ιδιαιτερότητες κάθε συνθήκης. Αυτό θα οδηγούσε σε πιο αξιόπιστες προβλέψεις, καθώς θα λαμβάνονταν υπόψη οι πιο εξειδικευμένες επιδράσεις του καιρού στην Φ/Β παραγωγή.
- Επέκταση των δεδομένων των εικόνων: Η περαιτέρω αύξηση του πλήθους των εικόνων που περιείχε το δεύτερο σετ δεδομένων για κάθε κλάση νεφώσεων θα βοηθούσε σημαντικά στη δυνατότητα του μοντέλου να εκπαιδευτεί πιο σωστά. Με περισσότερες εικόνες, το μοντέλο θα μπορούσε να μάθει και να προσαρμοστεί καλύτερα στις πιο πολύπλοκες και μεταβλητές καιρικές

συνθήκες, βελτιώνοντας την ακρίβεια και την αξιοπιστία των προβλέψεων της Φ/Β παραγωγής. Επιπρόσθετα, η μεγαλύτερη ποικιλία στις εικόνες θα αύξανε την δυνατότητα του μοντέλου να γενικεύει, αντιμετωπίζοντας αποτελεσματικότερα απρόβλεπτες αλλαγές στον καιρό.

- Αξιολόγηση του μοντέλου σε κάποιο πραγματικό πρόβλημα διαχείρισης ηλεκτρικής ενέργειας: Η εφαρμογή του προτεινόμενου μοντέλου σε ένα πραγματικό πρόβλημα διαχείρισης ηλεκτρικής ενέργειας και η μετέπειτα αξιολόγηση του θα επέκτειναν σημαντικά την παρούσα διπλωματική εργασία. Μέσω αυτής της δοκιμής σε πραγματικές συνθήκες, θα μπορούσε να αναδειχθεί η χρησιμότητα και η αποτελεσματικότητα του, ενώ παράλληλα θα προέκυπταν σημαντικά δεδομένα σχετικά με τις επιδόσεις του μοντέλου, τα οποία θα μπορούσαν να αξιοποιηθούν για την περαιτέρω βελτίωση του.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] “ΕΘΝΙΚΟ ΣΧΕΔΙΟ ΕΝΕΡΓΕΙΑΣ ΚΑΙ ΚΛΙΜΑΤΟΣ-ΠΡΟΣΧΕΔΙΟ ΑΝΑΘΕΩΡΗΜΕΝΗΣ ΕΚΔΟΣΗΣ ΟΚΤΩΒΡΙΟΣ 2023.”
- [2] U. K. Das *et al.*, “Forecasting of photovoltaic power generation and model optimization: A review,” *Renewable and Sustainable Energy Reviews*, vol. 81. Elsevier Ltd, pp. 912–928, 2018. doi: 10.1016/j.rser.2017.08.017.
- [3] K. V Vidyanandan, “An Overview of Factors Affecting the Performance of Solar PV Systems,” 2017. [Online]. Available: <https://www.researchgate.net/publication/319165448>
- [4] Π. Σ. Γεωργιλάκης, *Σύγχρονα Συστήματα Μεταφοράς και Διανομής Ηλεκτρικής Ενέργειας*. Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών (ΣΕΑΒ), 2015.
- [5] Sebastian. Busch, Ruben. Kasdorp, Derck. Koolen, Arnaud. Mercier, Magdalena. Spooner, and European Commission. Directorate-General for Economic and Financial Affairs., *The development of renewable energy in the electricity market*.
- [6] Σ. Α. Παπαθανασίου and Κ. Εμπ, “Σύνδεση στο Δίκτυο διεσπαρμένων ενεργειακών πόρων Μάθημα: ΑΝΑΝΕΩΣΙΜΕΣ ΠΗΓΕΣ ΕΝΕΡΓΕΙΑΣ.”
- [7] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” *Solar Energy*, vol. 136. Elsevier Ltd, pp. 78–111, Oct. 15, 2016. doi: 10.1016/j.solener.2016.06.069.
- [8] K. Başaran, F. Bozyiğit, P. Siano, P. Y. Taşer, and D. Kılınc, “Systematic literature review of photovoltaic output power forecasting,” *IET Renewable Power Generation*, vol. 14, no. 19, pp. 3961–3973, Dec. 2020, doi: 10.1049/iet-rpg.2020.0351.
- [9] S. Mirjalili, H. Faris, I. Aljarah, and A. Algorithms, “Evolutionary Machine Learning Techniques.” [Online]. Available: <http://www.springer.com/series/16171>
- [10] S. Mirjalili, “The ant lion optimizer,” *Advances in Engineering Software*, vol. 83, pp. 80–98, 2015, doi: 10.1016/j.advengsoft.2015.01.010.
- [11] T. Schmidt, “High resolution solar irradiance forecasts based on sky images in Bad Berleburg,” 1986.
- [12] D. Anagnostos, T. Schmidt, S. Cavadias, D. Soudris, J. Poortmans, and F. Catthoor, “A method for detailed, short-term energy yield forecasting of photovoltaic installations,” *Renew Energy*, vol. 130, pp. 122–129, Jan. 2019, doi: 10.1016/j.renene.2018.06.058.
- [13] L. Oneto, F. Laureri, M. Robba, F. Delfino, and D. Anguita, “Data-driven photovoltaic power production nowcasting and forecasting for polygeneration microgrids,” *IEEE Syst J*, vol. 12, no. 3, pp. 2842–2853, Sep. 2018, doi: 10.1109/JSYST.2017.2688359.
- [14] S. Leva, A. Nespoli, S. Pretto, M. Mussetta, and E. G. C. Ogliari, “PV plant power nowcasting: A real case comparative study with an open access dataset,” *IEEE Access*, vol. 8, pp. 194428–194440, 2020, doi: 10.1109/ACCESS.2020.3031439.

- [15] I. A. Ibrahim, M. J. Hossain, and B. C. Duck, “An optimized offline random forests-based model for ultra-short-term prediction of PV characteristics,” *IEEE Trans Industr Inform*, vol. 16, no. 1, pp. 202–214, Jan. 2020, doi: 10.1109/TII.2019.2916566.
- [16] “<https://www.iea.org/reports/renewables-2023/executive-summary>”.
- [17] “<https://www.iea.org/reports/renewables-2023/electricity>”.
- [18] “<https://www.abovesurveying.com/growth-of-solar-energy-in-2023/>”.
- [19] “<https://www.iea.org/energy-system/renewables/solar-pv>”.
- [20] “<https://www.ot.gr/2024/02/07/green/ape/ember-igetis-tis-iliakis-energeias-i-ellada-to-alma-tis-4etias/>”.
- [21] “World Energy Outlook 2023.”
- [22] “<https://daelim-electric.com/solar-transformer/>”.
- [23] P. S. Georgilakis, “Spotlight on Modern Transformer Design (Power Systems).”
- [24] Κ. Σ. Παπαθανασίου, “Φωτοβολταϊκά Συστήματα Σταύρος Παπαθανασίου Καθηγητής ΕΜΠ.”
- [25] “<https://blog.ecoflow.com/us/effects-of-temperature-on-solar-panel-efficiency/>”.
- [26] S. Dubey, J. N. Sarvaiya, and B. Seshadri, “Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world - A review,” in *Energy Procedia*, Elsevier Ltd, 2013, pp. 311–321. doi: 10.1016/j.egypro.2013.05.072.
- [27] S. A. Alves dos Santos, J. P. João, C. A. Carlos, and R. A. Marques Lameirinhas, “The impact of aging of solar cells on the performance of photovoltaic panels,” *Energy Conversion and Management: X*, vol. 10, Jun. 2021, doi: 10.1016/j.ecmx.2021.100082.
- [28] H. Goverde *et al.*, “Spatial and temporal analysis of wind effects on PV module temperature and performance,” *Sustainable Energy Technologies and Assessments*, vol. 11, pp. 36–41, Sep. 2015, doi: 10.1016/j.seta.2015.05.003.
- [29] S. Pingel *et al.*, “Potential Induced Degradation of solar cells and panels,” in *2010 35th IEEE Photovoltaic Specialists Conference*, IEEE, Jun. 2010, pp. 002817–002822. doi: 10.1109/PVSC.2010.5616823.
- [30] Σ. Αθ Παπαθανασίου, “Ανάπτυξη των ανανεώσιμων πηγών ενέργειας διεθνώς και στην Ελλάδα.”
- [31] K. Van Der Wiel *et al.*, “The influence of weather regimes on European renewable energy production and demand,” *Environmental Research Letters*, vol. 14, no. 9, Sep. 2019, doi: 10.1088/1748-9326/ab38d3.
- [32] Π. Γεωργιλάκης, *Οικονομική και Αξιοπίστη Λειτουργία Σύγχρονων Συστημάτων Ηλεκτρικής Ενέργειας*. Αθήνα: Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις, 2023.
- [33] “https://el.wikipedia.org/wiki/%CE%95%CE%BB%CE%BB%CE%B7%CE%BD%CE%B9%CE%BA%CF%8C_%CE%A7%CF%81%CE%B7%CE%BC%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%AE%CF%81%CE%B9%CE%BF_%CE%95%CE%BD%CE%AD%CF%81%CE%B3%CE%B5%CE%B9%CE%B1%CF%82”.
- [34] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, “The future of forecasting for renewable energy,” *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 9, no. 2. John Wiley and Sons Ltd, Mar. 01, 2020. doi: 10.1002/wene.365.
- [35] H. Böök and A. V. Lindfors, “Site-specific adjustment of a NWP-based photovoltaic production forecast,” *Solar Energy*, vol. 211, pp. 779–788, Nov. 2020, doi: 10.1016/j.solener.2020.10.024.
- [36] “<https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/>”.

- [37] D. Watari *et al.*, “Multi-time scale energy management framework for smart PV systems mixing fast and slow dynamics,” *Appl Energy*, vol. 289, May 2021, doi: 10.1016/j.apenergy.2021.116671.
- [38] “ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών Improving Photovoltaic Energy Yield Forecasting Accuracy using Neural Networks.”
- [39] M. W. B. Azlinah, M. Bee, and W. Yap, “Supervised and Unsupervised Learning for Data Science Unsupervised and Semi-Supervised Learning Series Editor: M. Emre Celebi.” [Online]. Available: <http://www.springer.com/series/15892>
- [40] “https://en.wikipedia.org/wiki/Random_forest”.
- [41] “<https://www.ibm.com/topics/decision-trees>”.
- [42] L. Breiman, “Random Forests,” 2001.
- [43] B. Peter Bühlmann, “ANALYZING BAGGING,” 2002.
- [44] P. Probst, M. N. Wright, and A. L. Boulesteix, “Hyperparameters and tuning strategies for random forest,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 3. Wiley-Blackwell, May 01, 2019. doi: 10.1002/widm.1301.
- [45] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, and A. J. Tallón-Ballesteros, “The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis,” 2023, pp. 344–353. doi: 10.1007/978-3-031-42536-3_33.
- [46] “https://scikit-learn.org/stable/getting_started.html”.
- [47] A. Gholamy, V. Kreinovich, and O. Kosheleva, “A Pedagogical Explanation A Pedagogical Explanation Part of the Computer Sciences Commons,” 2018. [Online]. Available: https://scholarworks.utep.edu/cs_techrephttps://scholarworks.utep.edu/cs_techrep/1209
- [48] “<https://pandas.pydata.org/docs/>”.
- [49] “<https://numpy.org/doc/>”.