



**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ**

**DEPARTMENT INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING
SCHOOL OF MECHANICS**

Διπλωματική Εργασία

«ΠΡΟΓΝΩΣΤΙΚΗ ΣΥΝΤΗΡΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΕΣ»

Νικόλαος Δημήτριος Λυκούρεσης (ΑΜ: 71445063)

Επιβλέπων: Γρηγόριος Νικολάου

Αθήνα, Οκτώβριος 2024

Εξεταστική Επιτροπή

1. Νικολάου Γρηγόριος

2. Βασιλείαδου Σουλτάνα

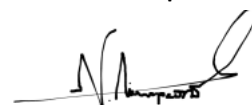
3. Δρόσσος Χρήστος

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Νικόλαος Δημήτριος Λυκούρεσης με αριθμό μητρώου 71445063 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών, του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



Ευχαριστίες – Αφιερώσεις

Θα ήθελα να εκφράσω την ειλικρινή ευγνωμοσύνη μου στον επιβλέποντα καθηγητή μου, Γρηγόρη Νικολάου, για την ανεκτίμητη υποστήριξη και καθοδήγησή που μου παρείχε σε όλη αυτή τη διατριβή. Επίσης θα ήθελα να τον ευχαριστήσω για τις γνώσεις και τα ερεθίσματα που άντλησα μέσω του μαθήματός του “Επιχειρηματική Ευφυΐα και Ανάλυση Μεγάλων Δεδομένων”.

Τέλος, εκφράζω την εγκάρδια εκτίμησή μου στην οικογένεια και τους ανθρώπους που ήταν δίπλα μου καθ' όλη τη διάρκεια αυτού του ταξιδιού.

Ημερομηνία 11/10/2024

Συγγραφέας Νικόλαος Δημήτριος Λυκούρεσης

ΠΕΡΙΛΗΨΗ

Η αιολική ενέργεια έχει αναδειχθεί ως βασικό στοιχείο στην παγκόσμια μετάβαση σε τρόπους ανανεώσιμης ενέργειας, ωστόσο η συντήρηση των ανεμογεννητριών, ιδιαίτερα σε θαλάσσια περιβάλλοντα, παρουσιάζει σημαντικές προκλήσεις. Η προγνωστική συντήρηση (PdM) σηματοδοτεί μια σημαντική πρόοδο στη διατήρηση της λειτουργικής αποτελεσματικότητας και τη μείωση των δαπανών για την παραγωγή ανανεώσιμης ενέργειας. Αυτή η διατριβή εξετάζει την εφαρμογή του PdM, χρησιμοποιώντας μεθοδολογίες μηχανικής μάθησης για την πρόβλεψη δυσλειτουργιών εξοπλισμού σε ανεμογεννήτριες. Η έρευνα επικεντρώνεται στη μελέτη μιας περίπτωσης αιολικών πάρκων, χρησιμοποιώντας δεδομένα SCADA από ανεμογεννήτριες για τη διαμόρφωση ενός μοντέλου μακροπρόθεσμης μνήμης (LSTM) για πρόβλεψη σφαλμάτων. Η έρευνα υπογραμμίζει τα οικονομικά, περιβαλλοντικά και λειτουργικά πλεονεκτήματα του PdM, αναδεικνύοντας την υπεροχή του έναντι των συμβατικών στρατηγικών συντήρησης. Προβλέποντας τις αστοχίες πριν από την εκδήλωσή τους, το PdM όχι μόνο μετριάξει τον χρόνο διακοπής λειτουργίας αλλά και παρατείνει τη μακροζωία των βασικών εξαρτημάτων, εξασφαλίζοντας έτσι βιώσιμη παραγωγή ενέργειας. Τα αποτελέσματα υπογραμμίζουν τη σημασία των τεχνολογικών καινοτομιών για την αύξηση της αξιοπιστίας και της αποτελεσματικότητας των συστημάτων αιολικής ενέργειας, συμβάλλοντας στους γενικούς στόχους της ενεργειακής βιωσιμότητας και της προστασίας του περιβάλλοντος.

Λέξεις Κλειδιά: Προγνωστική Συντήρηση, Ανεμογεννήτριες, Μηχανική Μάθηση, SCADA, LSTM, Πρόβλεψη Σφάλματος

ABSTRACT

Wind energy has emerged as a key element in the global transition to renewable energy modes, but maintaining wind turbines, particularly in marine environments, presents significant challenges. The integration of predictive maintenance (PdM) in wind turbines represents a significant advancement in maintaining operational efficiency and reducing costs in renewable energy production. This thesis investigates the application of PdM, leveraging machine learning techniques to predict equipment failures in wind turbines. The study focuses on a case analysis of wind farms, utilizing SCADA data from turbines to develop a Long Short-Term Memory (LSTM) model for fault prediction. The research highlights the economic, environmental, and operational benefits of PdM, showcasing its superiority over traditional maintenance methods. By predicting failures before they occur, PdM not only minimizes downtime but also extends the lifespan of critical components, ensuring sustainable energy production. The findings underscore the importance of technological advancements in enhancing the reliability and efficiency of wind energy systems, contributing to the broader goals of energy sustainability and environmental protection.

Keywords: Predictive Maintenance, Wind Turbines, Machine Learning, SCADA, LSTM, Fault Prediction

Πίνακας περιεχομένων

Εξεταστική Επιτροπή	ii
ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ	iii
Ευχαριστίες – Αφιερώσεις.....	iv
ΠΕΡΙΛΗΨΗ	v
ABSTRACT	vi
Πίνακας περιεχομένων.....	vii
Πίνακας Σχημάτων.....	ix
Πίνακας Πινάκων.....	x
Κεφάλαιο 1. Εισαγωγή.....	1
1.1 Δυσκολίες που εμφανίζονται στην συντήρηση	1
1.2 Οφέλη Προγνωστικής Συντήρησης σε Ανεμογεννήτριες	2
Κεφάλαιο 2. Θεωρητικό μέρος – Βιβλιογραφική έρευνα	4
2.1 Επισκόπηση Αιολικής Ενέργειας.....	4
2.1.1 Σημαντικότητα και ανάπτυξη της αιολικής ενέργειας σε παγκόσμια κλίμακα.....	4
2.1.2 Κύρια Στοιχεία και Τεχνολογίες Στις Ανεμογεννήτριες	6
2.1.3 Προκλήσεις στην συντήρηση ανεμογεννητριών.....	10
2.2 Προβλεπτική Συντήρηση	12
2.2.1 Ιστορική Εξέλιξη	12
2.2.2 Τρόπος λειτουργίας του PdM	13
2.2.3 Πλεονεκτήματα του PdM	15
2.2.4 Περιορισμοί και προκλήσεις.....	16
Κεφάλαιο 3. Μεθοδολογία.....	17
3.1 Δεδομένα για προγνωστικά μοντέλα.....	17
3.2 Επεξεργασία δεδομένων	17
3.2.1 Διαχείριση δεδομένων που λείπουν	17
3.2.2 Αλλαγή κλίμακας δεδομένων	18
3.2.3 Ανισοροπία συνόλου δεδομένων	18
3.2.4 Μοντέλα μηχανικής μάθησης.....	19
3.3 Artificial Neural Networks.....	20
3.4 Βελτιστοποίηση Υπερπαραμέτρων.....	26
3.5 Επιλογή Δεδομένων	27
3.6 Μετρητές αξιολόγησης	29
3.6.1 Μετρητές ταξινόμησης.....	30
Κεφάλαιο 4. Υλοποίηση.....	32
4.1 Εισαγωγή και δεδομένα	32
4.2 Προεπεξεργασία δεδομένων	36
4.3 Επιλογή χαρακτηριστικών	37

4.4	Επιλογή μοντέλου μηχανικής μάθησης – LSTM	41
Κεφάλαιο 5.	Αποτελέσματα – Συμπεράσματα – Μελλοντικές προσεγγίσεις	45
5.1	Εκπαίδευση LSTM με correlation matrix	45
5.2	Εκπαίδευση LSTM με PCA	48
5.3	Εκπαίδευση LSTM με αλλαγή στο time window και το resampling	50
5.3.1	Αλλαγή στο time window	50
5.3.2	Αλλαγή στο sampling	52
5.4	Συμπεράσματα και Μελλοντικές προσεγγίσεις.....	54
Βιβλιογραφικές Αναφορές		57

Πίνακας Σχημάτων

Εικόνα 1 Τεχνικά μέρη μιας ανεμογεννήτριας (Wind Turbine Facts Britannica, n.d.)	9
Εικόνα 2 Εξέλιξη των στρατηγικών συντήρησης (Evolution of Maintenance Download Scientific Diagram, n.d.).....	13
Εικόνα 3 Απεικόνιση υποπροσαρμογής, υπερπροσαρμογής και επιθυμητής προσαρμογής του μοντέλου στα δεδομένα (ML Underfitting and Overfitting - GeeksforGeeks, n.d.).....	20
Εικόνα 4 Ένα παράδειγμα μιας βασικής ανατροφοδότησης NN.(A Typical Artificial Neural Network (ANN) Download Scientific Diagram, n.d.).....	21
Εικόνα 5 Ξεδίπλωμα ενός RNN (Vijay Gadre, 2023)	22
Εικόνα 6 Σχηματική αναπαράσταση του κυτάρου LSTM τη χρονική στιγμή t.....	24
Εικόνα 7 Confusion matrices	29
Εικόνα 8 Το συνολικό ποσοστό εγγραφών ανά ανεμογεννήτρια στο σετ δεδομένων από την 1η φάρμα.	32
Εικόνα 9 Χρονοσειρά ενός από τα 22 σετ δεδομένων της wind farm A για τον αισθητήρα 48.	33
Εικόνα 10 Ποσοστό καταστάσεων εγγραφών επί του συνόλου των δεδομένων και για τις 5 ανεμογεννήτριες.	34
Εικόνα 11 Στάδια προτεινόμενης λύσης για μοντέλο προγνωστικής συντήρησης.....	35
Εικόνα 12 Correlation matrix για τα 81 διανύσματα χαρακτηριστικών του σετ δεδομένων.....	39
Εικόνα 13 Scatter plot των δύο από τις δέκα κύριες συνιστώσες του dataset.	40
Εικόνα 14 Η αρχιτεκτονική του μοντέλου μηχανικής μάθησης που χρησιμοποιήθηκε.	42
Εικόνα 15 Τα μεγέθη των επιμέρους επιπέδων και οι παράμετροι του μοντέλου μηχανικής μάθησης που χρησιμοποιήθηκε.	43
Εικόνα 16 Καμπύλες ακρίβειας ανά εποχές για training και validation σετ δεδομένων για correlation threshold 0.8.	46
Εικόνα 17 Καμπύλες loss ανά εποχές για training και validation σετ δεδομένων για correlation threshold 0.8.	47
Εικόνα 18 Confusion matrix για correlation threshold 0.8.....	47
Εικόνα 19 Καμπύλες ακρίβειας ανά εποχές για training και validation σετ δεδομένων για 40 κύριες συνιστώσες.	49
Εικόνα 20 Καμπύλες loss ανά εποχές για training και validation σετ δεδομένων για 40 κύριες συνιστώσες	49
Εικόνα 21 Confusion matrix για 40 κύριες συνιστώσες.	50
Εικόνα 22 Σύγκριση του ίδιου feature στην περίπτωση που έχει γίνει resampling για 3 ώρες και τα πακέτα έχουν time window μέγεθος 16 τιμών (πάνω) και 40 τιμών (κάτω).	51

Πίνακας Πινάκων

Πίνακας 1 Περιγραφή των διαφορετικών κατηγοριών για την στήλη status-ID των δεδομένων.....	34
Πίνακας 2 Αποτελέσματα πειραμάτων (finetuning) για επιλογή κατάλληλης τιμής στο κατώφλι correlation.....	46
Πίνακας 3 Αποτελέσματα πειραμάτων (finetuning) για επιλογή του κατάλληλου αριθμού κύριων συνιστωσών.....	48
Πίνακας 4 Αποτελέσματα πειραμάτων (finetuning) για επιλογή του κατάλληλου μεγέθους time window.....	52
Πίνακας 5 Αποτελέσματα πειραμάτων (finetuning) για επιλογή του κατάλληλου time resolution (resampling).....	54

Κεφάλαιο 1. Εισαγωγή

Τα αιολικά πάρκα αποτελούν σημαντικό παράγοντα στον τομέα της ενέργειας, ιδίως καθώς αυξάνεται η αναγκαιότητα για τη χρήση ανανεώσιμων πηγών ενέργειας. Η Ευρωπαϊκή Ένωση (ΕΕ) στοχεύει να μειώσει τις καθαρές εκπομπές αερίων θερμοκηπίου κατά τουλάχιστον 55% έως το 2030 σε σύγκριση με τα επίπεδα του 1990, ως μέρος της ευρύτερης στρατηγικής της για επίτευξη κλιματικής ουδετερότητας έως το 2050. Σημειώνεται πως από το 2022 υπήρξε αύξηση της αιολικής ισχύος κατά 33 GW. Παράλληλα η Ευρωπαϊκή Πράσινη Συμφωνία καθορίζει μια νέα στρατηγική ανάπτυξης με στόχο να μετατρέψει την ΕΕ σε μια κοινωνία ανεξάρτητη από εκπομπές αερίων θερμοκηπίου έως το 2050 (European Commission, n.d.). Οι αποτελεσματικές στρατηγικές συντήρησης είναι απαραίτητες για την ελαχιστοποίηση του λειτουργικού κόστους και τη μεγιστοποίηση της παραγωγής ενέργειας. Η έρευνα δείχνει ότι το κόστος λειτουργίας και συντήρησης (O&M) αποτελεί περίπου το 30% του κόστους ζωής των αιολικών πάρκων, υπογραμμίζοντας την ανάγκη για οικονομικά αποδοτικές προσεγγίσεις συντήρησης (Moros et al., 2024). Επιπλέον, η ενσωμάτωση μοντέλων προγνωστικής συντήρησης μπορεί να βελτιώσει την αποδοτικότητα χρησιμοποιώντας δεδομένα σε πραγματικό χρόνο για την πρόβλεψη βλαβών, ελαχιστοποιώντας έτσι το χρόνο διακοπής λειτουργίας και βελτιώνοντας τη δημιουργία εσόδων. Παράλληλα, ο περιορισμός του ανθρώπινου παράγοντα στις διαδικασίες συντήρησης είναι ζωτικής σημασίας, καθώς σχεδόν το 40% των βλαβών ανεμογεννητριών συνδέονται με ανθρώπινο λάθος (Golestani et al., 2024).

1.1 Δυσκολίες που εμφανίζονται στην συντήρηση

Η συντήρηση ειδικά των υπεράκτιων ανεμογεννητριών (OWT) αντιμετωπίζει πολλές προκλήσεις που επηρεάζουν την αποτελεσματικότητα και το κόστος της. Το υψηλό κόστος συντήρησης, που προκύπτει από την ανάγκη για εξειδικευμένο εξοπλισμό και υλικοτεχνικά μέσα, είναι σημαντικά μεγαλύτερο σε σύγκριση με τις χερσαίες αιολικές εγκαταστάσεις. Οι αποστάσεις από λιμάνια και παράκτιες περιοχές περιορίζουν την

προσβασιμότητα, οδηγώντας σε αυξημένες περιόδους αδράνειας και πρόσθετα έξοδα για τη μεταφορά τεχνικών και εξοπλισμού. Επιπλέον, οι δυσμενείς καιρικές συνθήκες, όπως οι υψηλές ταχύτητες ανέμου και τα μεγάλα κύματα, περιορίζουν τη δυνατότητα πρόσβασης για συντήρηση, συχνά προκαλώντας αναβολές στις προγραμματισμένες εργασίες. Τέλος, η πολυπλοκότητα των OWT, απαιτεί εξειδικευμένες γνώσεις και εξοπλισμό για τη συντήρηση (Shafiee, 2024).

Η κατανομή πόρων, αναφορικά με τη συχνότητα των επισκέψεων συντήρησης, μπορεί να επηρεάσει την οικονομική βιωσιμότητα και την αποτελεσματικότητα των διαδικασιών. Ενώ οι συχνές επισκέψεις ενδέχεται να είναι οικονομικά επιβαρυντικές, οι σπάνιες επισκέψεις ενδέχεται να οδηγούν σε αυξημένα ποσοστά αστοχίας και μακροχρόνιες περιόδους αδράνειας (Shafiee, 2024).

Τέλος, η εκπαίδευση και η ασφάλεια των τεχνικών είναι ζωτικής σημασίας, ιδιαίτερα υπό τις απαιτητικές συνθήκες που απαιτεί η διαδικασία της συντήρησης των ανεμογεννητριών.

1.2 Οφέλη Προγνωστικής Συντήρησης σε Ανεμογεννήτριες

Η προγνωστική συντήρηση στις ανεμογεννήτριες προσφέρει σημαντική προστιθέμενη αξία σε οικονομικές, κοινωνικές και περιβαλλοντικές διαστάσεις. Αξιοποιώντας προηγμένα μοντέλα μηχανικής μάθησης και μεγάλα σύνολα δεδομένων, ενισχύει την αξιοπιστία και την αποδοτικότητα της παραγωγής αιολικής ενέργειας. Αυτή η προσέγγιση όχι μόνο μειώνει το λειτουργικό κόστος, αλλά ελαχιστοποιεί επίσης τις περιβαλλοντικές επιπτώσεις και ενισχύει την κοινωνική αποδοχή της αξιοποίησης της αιολικής ενέργειας.

- **Οικονομικά Οφέλη**

Η εφαρμογή μοντέλων PdM, τα οποία βασίζονται σε αλγορίθμους μηχανικής μάθησης, μειώνουν σημαντικά το χρόνο αδράνειας και το κόστος συντήρησης, βελτιστοποιώντας τα χρονοδιαγράμματα συντήρησης και επεκτείνοντας τη διάρκεια ζωής των εξαρτημάτων των ανεμογεννητριών

(Nunnula et al., 2024). Για παράδειγμα, ένα υπεράκτιο αιολικό πάρκο ισχύος 1 GW θα μπορούσε να δημιουργήσει επιπλέον £1,4 εκατομμύρια σε έσοδα μέσω βελτιωμένου προγραμματισμού συντήρησης (Moros et al., 2024) .

- **Κοινωνικός Αντίκτυπος**

Η βελτίωση της αξιοπιστίας των ανεμογεννητριών μέσω της PdM συμβάλλει στη σταθερότητα της ενεργειακής παροχής, γεγονός που είναι ουσιώδες για την κοινωνική αποδοχή των ανανεώσιμων πηγών ενέργειας (Manna et al., 2024). Η μείωση των απρόσμενων βλαβών και παρεμβάσεων συντήρησης βελτιώνει επίσης την ασφάλεια και τις συνθήκες εργασίας για τις ομάδες συντήρησης (Moros et al., 2024).

- **Περιβαλλοντικά Οφέλη**

Η προβλεπτική συντήρηση συμβάλλει σημαντικά στην αποδοτική λειτουργία των ανεμογεννητριών, μεγιστοποιώντας έτσι την παραγωγή ενέργειας και μειώνοντας το αποτύπωμα του άνθρακα που συνδέεται με την παραγωγή ενέργειας (Nunnula et al., 2024). Μειώνοντας την ανάγκη για αντιδραστική συντήρηση, η οποία συχνά περιλαμβάνει μεταφορές και χρήση πόρων, η προβλεπτική συντήρηση συμβάλλει στην μείωση των περιβαλλοντικών επιπτώσεων που σχετίζονται με τη λειτουργία των αιολικών πάρκων (Rajaoarisoa et al., 2024).

Αν και η προγνωστική συντήρηση προσφέρει σημαντικά οφέλη, είναι σημαντικό να ληφθούν υπόψη η αρχική επένδυση και η πολυπλοκότητα που συνεπάγεται η εφαρμογή τέτοιων συστημάτων. Η ενσωμάτωση προηγμένων αναλύσεων δεδομένων και μοντέλων μηχανικής μάθησης απαιτεί σημαντική εξειδίκευση και πόρους, γεγονός που μπορεί να αποτελέσει πρόκληση για μικρότερους φορείς. Ωστόσο, καθώς η τεχνολογία εξελίσσεται και τα κόστη μειώνονται, η υιοθέτηση της προγνωστικής συντήρησης είναι πιθανό να εξαπλωθεί ευρύτερα, ενισχύοντας περαιτέρω την αξία της σε οικονομικές, κοινωνικές και περιβαλλοντικές διαστάσεις. Συνολικά, η προγνωστική συντήρηση αποτελεί μια σημαντική στρατηγική για τη βελτιστοποίηση της παραγωγής αιολικής ενέργειας και την υποστήριξη της βιώσιμης ανάπτυξης.

Κεφάλαιο 2. Θεωρητικό μέρος – Βιβλιογραφική έρευνα

2.1 Επισκόπηση Αιολικής Ενέργειας

2.1.1 Σημαντικότητα και ανάπτυξη της αιολικής ενέργειας σε παγκόσμια κλίμακα

Η αιολική ενέργεια έχει αναδειχθεί ως ουσιαστικό στοιχείο στην παγκόσμια μετάβαση προς τις ανανεώσιμες πηγές ενέργειας, προωθούμενη από την ικανότητά της να συμβάλει σημαντικά στην μείωση εκπομπών άνθρακα και να αντιμετωπίζει τις προκλήσεις που θέτει η κλιματική αλλαγή. Η επέκταση της αιολικής ενέργειας χαρακτηρίζεται από την αυξανόμενη συμβολή της στην παγκόσμια παραγωγή ηλεκτρικής ενέργειας και τη σημασία της στην επίτευξη των διεθνών κλιματικών στόχων. Η μετάβαση αυτή υποστηρίζεται από την τεχνολογική πρόοδο και τη στρατηγική εφαρμογή πολιτικών σε διάφορες γεωγραφικές περιοχές. Παρακάτω, εξετάζεται η σημασία και η ανάπτυξη της αιολικής ενέργειας σε παγκόσμια κλίμακα, αντλώντας πληροφορίες από ερευνητικές μελέτες.

2.1.1.1 Περιβαλλοντικά και οικονομικά οφέλη

Η αιολική ενέργεια χρησιμεύει ως βιώσιμη εναλλακτική λύση έναντι των ορυκτών καυσίμων, συμβάλλοντας στη μείωση των εκπομπών άνθρακα. Το έτος 2021, συνέβαλε στη μείωση περίπου 1.311 Mt εκπομπών CO₂ σε παγκόσμια κλίμακα, με προβλέψεις να δείχνουν πιθανή μείωση έως και 32.864 Mt έως το έτος 2100 (Long et al., 2023). Οι ανεμογεννήτριες παρουσιάζουν ελάχιστες περιβαλλοντικές επιπτώσεις, καθιστώντας τις κατάλληλες για ανάπτυξη σε γεωργικά περιβάλλοντα και απομακρυσμένες περιοχές όπου η δημιουργία γραμμών μεταφοράς είναι δύσκολη.

2.1.1.2 Παγκόσμια ανάπτυξη και ολοκλήρωση

Η παγκόσμια εγκατεστημένη αιολική ισχύς το 2021 ήταν 837 GW, παράγοντας περίπου 2186 TWh ηλεκτρικής ενέργειας. Η ικανότητα αυτή αναμένεται να επεκταθεί καθώς οι χώρες συνεχίζουν να επενδύουν στην αιολική ενέργεια για την επίτευξη των

κλιματικών στόχων. Αυτή η επέκταση ενισχύεται από χώρες όπως η Κίνα και η Ινδία, οι οποίες έχουν παρουσιάσει ετήσιους ρυθμούς ανάπτυξης άνω του 30% στον τομέα της αιολικής ενέργειας (Hassan et al., 2024). Η αιολική ενέργεια αναμένεται να αποτελέσει περισσότερο από το 30% της παγκόσμιας παραγωγής ηλεκτρικής ενέργειας μέχρι τα μέσα του 21ου αιώνα, διαδραματίζοντας έτσι κρίσιμο ρόλο στην επίτευξη των στόχων εξισορρόπησης άνθρακα (Long et al., 2023).

2.1.1.3 Προκλήσεις και δυναμική της αγοράς

Παρά τα πλεονεκτήματά της, η αιολική ενέργεια αντιμετωπίζει προκλήσεις λόγω της διαλείπουσας και απρόβλεπτης φύσης της, γεγονός που καθιστά δύσκολη την ενσωμάτωσή της στα ηλεκτρικά δίκτυα. Οι οικονομικές προκλήσεις επίσης περιλαμβάνουν την εξισορρόπηση του υψηλού αρχικού κόστους με μακροπρόθεσμα οφέλη, τα οποία μπορούν να αποτρέψουν τους υπεύθυνους λήψης αποφάσεων από το να επενδύσουν σε έργα αιολικής ενέργειας (Hassan et al., 2024). Η αγορά αιολικής ενέργειας παρουσιάζει σημαντική ευμεταβλητότητα και υψηλό βαθμό συγκέντρωσης, ενώ οι υπάρχουσες δομές των αγορών ηλεκτρικής ενέργειας συχνά ευνοούν τις παραδοσιακές πηγές ενέργειας. Αυτή η τάση ενδέχεται να περιορίσει τις απαραίτητες επενδύσεις και να επιβραδύνει την ανάπτυξη του τομέα, παρά τα κοινωνικά και περιβαλλοντικά οφέλη που η αιολική ενέργεια έχει τη δυνατότητα να προσφέρει.

2.1.1.4 Περιφερειακές ανισότητες και μελλοντικές προοπτικές

Η ανάπτυξη της αιολικής ενέργειας παρουσιάζει σημαντικές διαφοροποιήσεις μεταξύ διαφορετικών περιοχών. Η δημιουργία υποδομών αιολικής ενέργειας απαιτεί σημαντικές κεφαλαιουχικές δαπάνες, οι οποίες μπορούν να αποτελέσουν εμπόδιο, ιδίως για τις αναπτυσσόμενες χώρες με περιορισμένους προϋπολογισμούς. Η Ευρωπαϊκή Ένωση όμως, ιδιαίτερα χώρες όπως η Δανία και η Γερμανία, πρωτοπορεί στην ενσωμάτωση της αιολικής ενέργειας, ενώ τα κράτη της Ασίας και της Αμερικής επεκτείνουν τα χαρτοφυλάκιά τους στις ανανεώσιμες πηγές ενέργειας με ταχείς ρυθμούς (Hassan et al., 2024). Οι προοπτική ανάπτυξης για την αιολική ενέργεια είναι

ενθαρυντική, με την προβλεπόμενη μετατόπιση των αιολικών πόρων προς τις νότιες περιοχές να αναμένεται να ωφελήσει τις χώρες χαμηλότερου γεωγραφικού πλάτους, ενισχύοντας την εμπιστοσύνη τους στην ανάπτυξη του τομέα (Long et al., 2023) .

Παρόλο που η αιολική ενέργεια αποτελεί κεντρικό στοιχείο της παγκόσμιας ενεργειακής μετάβασης, η ανάπτυξή της καθορίζεται από τις δυνάμεις της αγοράς, τις περιφερειακές ανισότητες και τις τεχνολογικές προκλήσεις. Η αποτελεσματική αντιμετώπιση αυτών των προκλήσεων μέσω στοχευμένων πολιτικών και επενδύσεων είναι απαραίτητη για την πλήρη αξιοποίηση της δυναμικής της αιολικής ενέργειας στην αντιμετώπιση της κλιματικής αλλαγής και την επίτευξη των στόχων βιώσιμης ανάπτυξης.

2.1.2 Κύρια Στοιχεία και Τεχνολογίες Στις Ανεμογεννήτριες

Τα βασικά στοιχεία και τεχνολογίες στα αιολικά πάρκα περιλαμβάνουν ένα σύνολο συστημάτων που εξασφαλίζουν την αποδοτική παραγωγή ενέργειας, την αξιοπιστία και την ασφάλεια. Αυτά τα στοιχεία περιλαμβάνουν την φυσική υποδομή των ανεμογεννητριών, τα ηλεκτρικά συστήματα και τις προηγμένες μεθόδους κυβερνοασφάλειας, καθένα από τα οποία διαδραματίζει καθοριστικό ρόλο στη συνολική λειτουργία και βελτιστοποίηση των αιολικών πάρκων.

2.1.2.1 Κύρια Στοιχεία Ανεμογεννητριών

Ρότορας και Πτερύγια

Ο ρότορας αποτελεί θεμελιώδες μέρος της ανεμογεννήτριας, περιλαμβάνοντας τα πτερύγια και είναι υπεύθυνος για τη αξιοποίηση της αιολικής ενέργειας και τη μετατροπή της σε περιστροφική κίνηση. (Alipour & Rahimpour, 2024)

Τα πτερύγια έχουν αεροδυναμικό σχεδιασμό που στοχεύει στη μέγιστη συλλογή ενέργειας. Σε ορισμένες ανεμογεννήτριες, η γωνία τους μπορεί να ρυθμιστεί μέσω συστημάτων ελέγχου γωνίας, ώστε να βελτιστοποιείται η απόδοση υπό διαφορετικές συνθήκες ανέμου. (Alipour & Rahimpour, 2024)

Σύστημα Μετάδοσης Κίνησης

Το σύστημα αυτό μεταφέρει την περιστροφική ενέργεια από τον ρότορα στη γεννήτρια. Διασφαλίζει ότι η ενέργεια που συλλέγεται από τα πτερύγια μεταφέρεται αποδοτικά στη γεννήτρια για την παραγωγή ηλεκτρισμού. (Alipour & Rahimpour, 2024)

Κιβώτιο Ταχυτήτων

Το κιβώτιο ταχυτήτων βοηθάει στην ρύθμιση της ταχύτητας περιστροφής του ρότορα, ώστε να ανταποκρίνεται στις απαιτήσεις υψηλής ταχύτητας της γεννήτριας. Αυτό το στοιχείο είναι ιδιαίτερα σημαντικό στις ανεμογεννήτριες οριζόντιου άξονα (HAWT), όπου προσαρμόζει τη χαμηλή ταχύτητα περιστροφής των πτερυγίων στις ανάγκες της γεννήτριας. (Alipour & Rahimpour, 2024)

Γεννήτρια

Η γεννήτρια μετατρέπει τη μηχανική ενέργεια από το σύστημα μετάδοσης κίνησης σε ηλεκτρική ενέργεια. Ο τύπος της γεννήτριας που χρησιμοποιείται μπορεί να διαφέρει ανάλογα με τον σχεδιασμό και τις απαιτήσεις της ανεμογεννήτριας, επηρεάζοντας σημαντικά την αποδοτικότητα και την αξιοπιστία του συστήματος. (Alipour & Rahimpour, 2024)

Πύργος

Ο πύργος υποστηρίζει τον ρότορα και τη γόνδολα, ανεβάζοντάς τους σε ύψος όπου οι ταχύτητες του ανέμου είναι βέλτιστες. Σε ανεμογεννήτριες οριζόντιου άξονα (HAWTs), χρησιμοποιούνται συνήθως υψηλότεροι πύργοι για την πρόσβαση σε ισχυρότερους ανέμους, βελτιώνοντας έτσι τη συλλογή ενέργειας. (Alipour & Rahimpour, 2024)

Συστήματα Ελέγχου

Αυτά τα συστήματα διαχειρίζονται τη λειτουργία της ανεμογεννήτριας, συμπεριλαμβανομένης της εκκίνησης, της διακοπής και της βελτιστοποίησης της απόδοσης. Διασφαλίζουν ότι η ανεμογεννήτρια λειτουργεί με ασφάλεια και αποδοτικότητα υπό διαφορετικές συνθήκες ανέμου. (Alipour & Rahimpour, 2024)

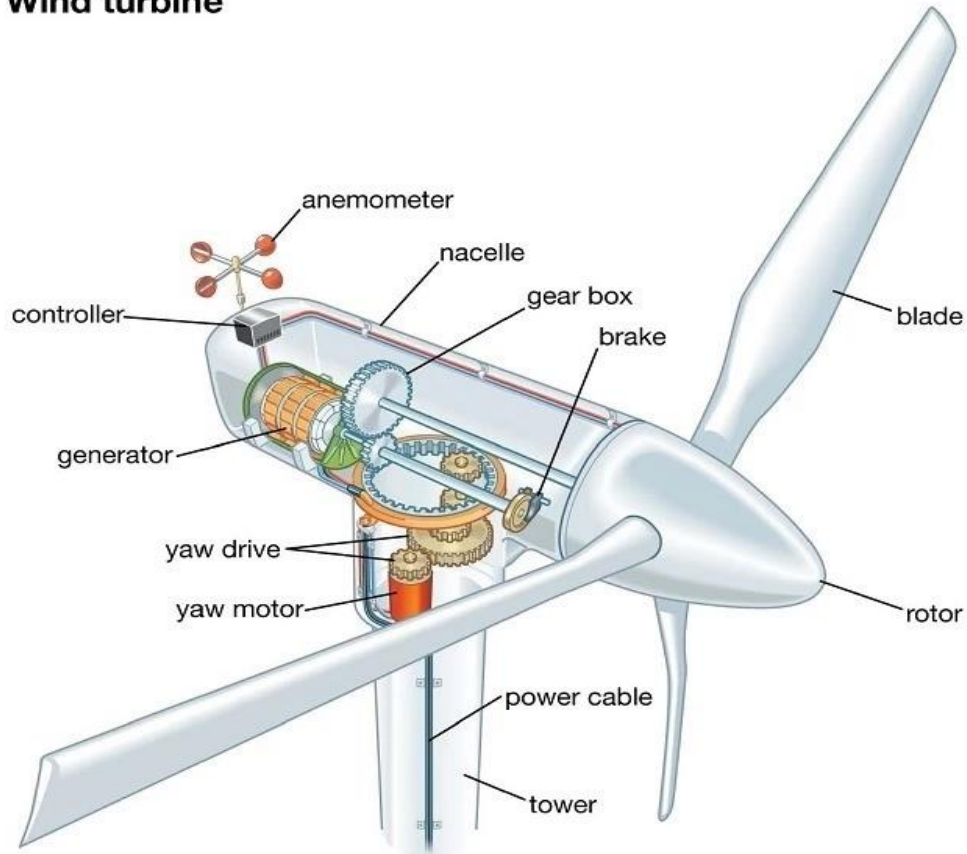
Σύστημα Περιστροφής (Yaw System)

Σε ανεμογεννήτριες οριζόντιου άξονα (HAWT), το σύστημα περιστροφής στρέφει τη γόνδολα ώστε ο ρότορας να είναι προσανατολισμένος προς την κατεύθυνση του ανέμου, μεγιστοποιώντας τη συλλογή ενέργειας. (Alipour & Rahimpour, 2024)

Γόνδολα (Nacelle)

Η γόνδολα στεγάζει το κιβώτιο ταχυτήτων, τη γεννήτρια και άλλα κρίσιμα στοιχεία στην κορυφή του πύργου, προσφέροντας προστασία από τις καιρικές συνθήκες. (Alipour & Rahimpour, 2024)

Wind turbine



© Encyclopædia Britannica, Inc.

Εικόνα 1 Τεχνικά μέρη μιας ανεμογεννήτριας (Wind Turbine Facts | Britannica, n.d.)

Αυτά τα στοιχεία λειτουργούν συνδυαστικά για να επιτρέπουν στις ανεμογεννήτριες να μετατρέπουν αποδοτικά την αιολική ενέργεια σε ηλεκτρική ισχύ, συμβάλλοντας στην παραγωγή ανανεώσιμης ενέργειας.

2.1.2.2 Κυβερνοασφάλεια και Μετάδοση Δεδομένων

Ασφάλεια Βιομηχανικών Συστημάτων Ελέγχου: Εξατομικευμένες στρατηγικές κυβερνοασφάλειας για την προστασία από μη εξουσιοδοτημένη πρόσβαση και κυβερνοαπειλές, διασφαλίζοντας τη σταθερή λειτουργία του συστήματος (Ma & Mashayekhy, 2020) .

Ασφαλής Μετάδοση Δεδομένων: Χρήση τεχνολογίας SIP για ασφαλή επικοινωνία στο σύστημα SCADA, αποτρέποντας διαρροές δεδομένων και ενισχύοντας την ασφάλεια του συστήματος (Ma & Mashayekhy, 2020).

2.1.3 Προκλήσεις στην συντήρηση ανεμογεννητριών

Η συντήρηση των ανεμογεννητριών περιλαμβάνει μια μυριάδα προκλήσεων που προκύπτουν από τα περίπλοκα λειτουργικά τους πλαίσια, την ανάγκη για εξελιγμένα συστήματα παρακολούθησης και τις μηχανικές καταπονήσεις που υπομένουν. Αυτά τα εμπόδια επιδεινώνονται από τις αυξανόμενες διαστάσεις και την πολυπλοκότητα των ανεμογεννητριών, ιδιαίτερα εκείνων που βρίσκονται σε υπεράκτια περιβάλλοντα. Ακολουθούν οι κύριες προκλήσεις που αναγνωρίζονται στη συντήρηση των ανεμογεννητριών, τεκμηριωμένες από στοιχεία από τα αναφερόμενα επιστημονικά άρθρα.

Μία από τις πρωταρχικές ανησυχίες είναι το υψηλό κόστος συντήρησης που συνδέεται με τα OWTs, το οποίο είναι σημαντικά υψηλότερο από αυτό των χερσαίων αιολικών πάρκων λόγω του εξειδικευμένου εξοπλισμού και της εφοδιαστικής που απαιτείται για υπεράκτιες δραστηριότητες. Τα ζητήματα προσβασιμότητας επιδεινώνουν περαιτέρω το κόστος αυτό, καθώς η απόσταση από τα υπεράκτια αιολικά πάρκα έως τους λιμένες ή τις ακτές οδηγεί σε αυξημένο χρόνο διακοπής λειτουργίας και υψηλότερα έξοδα που σχετίζονται με τα ταξίδια των τεχνικών και τη μεταφορά εξοπλισμού.

Οι δυσμενείς καιρικές συνθήκες, όπως οι υψηλές ταχύτητες ανέμου και τα σημαντικά ύψη κύματος, περιορίζουν την πρόσβαση των σκαφών εξυπηρέτησης και του προσωπικού, προκαλώντας συχνά καθυστερήσεις στις εργασίες συντήρησης. Ενώ καιρικά φαινόμενα όπως η βροχόπτωση και το χαλάζι μπορούν να συμβάλουν στη διάβρωση και την κόπωση των εξαρτημάτων του στροβίλου, ιδιαίτερα των λεπίδων. Αυτή η κατάσταση απαιτεί τακτικές επιθεωρήσεις και παρεμβάσεις συντήρησης για την αποφυγή παρατεταμένων ζημιών (Ren et al., 2021).

Η πολυπλοκότητα των OWTs, ιδιαίτερα με την εισαγωγή σταθερών και πλωτών θεμελίων, προσθέτει στη δυσκολία συντήρησης, απαιτώντας εξειδικευμένες γνώσεις και εξοπλισμό. Επιπλέον, ο πολλαπλασιασμός των αισθητήρων για την

παρακολούθηση της κατάστασης αυξάνει τόσο το κόστος όσο και την πολυπλοκότητα του συστήματος, παρουσιάζοντας προκλήσεις στη διαχείριση δεδομένων, συμπεριλαμβανομένης της συλλογής, του φιλτραρίσματος, της ανάλυσης και της αποθήκευσης.

Η ανίχνευση και διάγνωση σφαλμάτων παραμένει προβληματική, καθώς η διάκριση μεταξύ πραγματικών και ψευδών σφαλμάτων με τη χρήση ανάλυσης SCADA απαιτεί την ανάπτυξη προηγμένων αλγορίθμων και τεχνητής νοημοσύνης για την ενίσχυση της ακρίβειας. Η απαίτηση όμως για εξελιγμένους αλγόριθμους για την ανάλυση σημαντικών όγκων δεδομένων για την παρακολούθηση της κατάστασης μπορεί να επιβάλει σημαντικές υπολογιστικές απαιτήσεις. Υπάρχει λοιπόν ανάγκη για την ανάπτυξη πιο αποτελεσματικών αλγορίθμων που απαιτούν μικρότερο όγκο δεδομένων και υπολογιστικούς πόρους διευκολύνοντας παράλληλα την έγκαιρη ανίχνευση σφαλμάτων (Velandia-Cardenas et al., 2021).

Επιπλέον, οι δραστηριότητες συντήρησης πρέπει να εξισορροπούν προσεκτικά την κερδοφορία με την ελαχιστοποίηση των περιβαλλοντικών επιπτώσεων για την υποστήριξη της βιώσιμης ανάπτυξης της υπεράκτιας αιολικής ενέργειας. Η αποτελεσματική κατανομή των πόρων είναι κρίσιμη, καθώς οι συχνές επισκέψεις συντήρησης μπορεί να είναι δαπανηρές, ενώ οι σπάνιες επισκέψεις μπορεί να οδηγήσουν σε υψηλότερα ποσοστά αποτυχίας και μεγαλύτερους χρόνους διακοπής λειτουργίας.

Τέλος, οι δεξιότητες και η ασφάλεια των τεχνικών είναι υψίστης σημασίας στο σκληρό υπεράκτιο περιβάλλον εργασίας, απαιτώντας ολοκληρωμένη εκπαίδευση και εξοπλισμό για την εξασφάλιση αποτελεσματικής συντήρησης.

Αν και αυτές οι προκλήσεις είναι σημαντικές, η συνεχής πρόοδος της τεχνολογίας των ανεμογεννητριών παράλληλα με την επιδίωξη πιο βιώσιμων ενεργειακών λύσεων θα οδηγήσει στην εξέλιξη πιο αποτελεσματικών στρατηγικών συντήρησης.

2.2 Προβλεπτική Συντήρηση

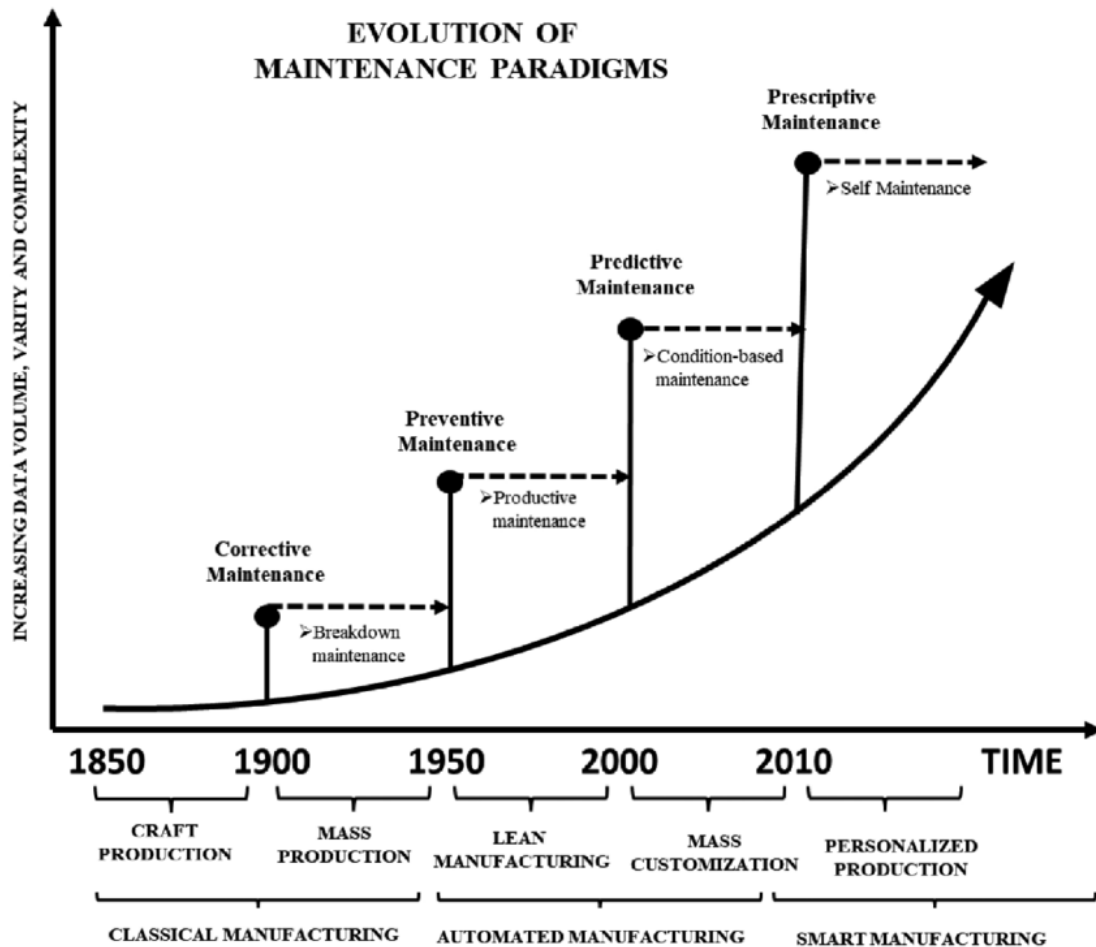
2.2.1 Ιστορική Εξέλιξη

Στα πρώτα στάδια της ανάπτυξης ανεμογεννητριών, οι στρατηγικές συντήρησης ήταν κυρίως αντιδραστικές. Η κύρια προσέγγιση ήταν η **διορθωτική συντήρηση**, η οποία συνεπαγόταν επισκευή ή αντικατάσταση εξαρτημάτων μόνο μετά την αστοχία τους (Peter Roór, 2019). Αυτή η μέθοδος είναι απλή και απαιτεί χαμηλότερο αρχικό κόστος, αλλά ενδέχεται να επιφέρει κλιμακωμένες μακροπρόθεσμες δαπάνες λόγω απρόβλεπτων χρόνων διακοπής λειτουργίας και επισκευών.

Με την ανάπτυξη της τεχνολογίας, η φιλοσοφία συντήρησης άρχισε να μετατοπίζεται προς πιο προληπτικές στρατηγικές. Η προγραμματισμένη ή **προληπτική συντήρηση** προέκυψε ως μια πιο προηγμένη προσέγγιση, όπου οι δραστηριότητες συντήρησης εκτελούνται σε τακτά χρονικά διαστήματα ανεξάρτητα από την κατάσταση του εξοπλισμού ((Peter Roór, 2019). Αυτή η μέθοδος στόχευε στην πρόληψη απροσδόκητων βλαβών και στην παράταση της διάρκειας ζωής των ανεμογεννητριών. Τα προγράμματα προληπτικής συντήρησης βασίζονταν συνήθως στις συστάσεις του κατασκευαστή και στην επιχειρησιακή εμπειρία. Η προληπτική συντήρηση, αν και πιο συστηματική, μπορεί να οδηγήσει σε περιττές δραστηριότητες συντήρησης.

Η **προβλεπτική συντήρηση (PdM)** αντιπροσωπεύει μια σημαντική εξέλιξη από αυτές τις προηγούμενες προσεγγίσεις. Σε αντίθεση με την προληπτική συντήρηση, η οποία βασίζεται σε σταθερά χρονοδιαγράμματα, το PdM αξιοποιεί δεδομένα σε πραγματικό χρόνο για να προβλέψει πότε ένα εξάρτημα είναι πιθανό να αποτύχει. Αυτή η προσέγγιση επιτρέπει τη συντήρηση να εκτελείται λίγο πριν από την αναμενόμενη βλάβη, βελτιστοποιώντας έτσι το πρόγραμμα συντήρησης και μειώνοντας τις περιττές παρεμβάσεις (Peter Roór, 2019). Η μετάβαση στην PdM

καθοδηγήθηκε από την πρόοδο στην τεχνολογία αισθητήρων, την ανάλυση δεδομένων και τη μηχανική μάθηση.



Εικόνα 2 Εξέλιξη των στρατηγικών συντήρησης (Evolution of Maintenance | Download Scientific Diagram, n.d.)

2.2.2 Τρόπος λειτουργίας του PdM

Στην Προγνωστική Συντήρηση (PdM), η αρχική φάση ξεκινά με την απόκτηση δεδομένων, από αισθητήρες που είναι τοποθετημένοι σε διάφορα μηχανήματα και συλλέγουν διαρκώς δεδομένα λειτουργίας σε πραγματικό χρόνο. Αυτοί οι αισθητήρες είναι απαραίτητοι για τον έλεγχο των παραμέτρων όπως η θερμοκρασία, η πίεση, οι κραδασμοί ή οι ακουστικές εκπομπές. Συλλογικά, οι πληροφορίες που λαμβάνονται από τους αισθητήρες μπορούν να παρέχουν μια ολοκληρωμένη εικόνα της κατάστασης λειτουργίας του εξαρτήματος. Το είδος των αισθητήρων που χρησιμοποιούνται και τα δεδομένα που συγκεντρώνονται βασίζονται σημαντικά στα

χαρακτηριστικά του εξοπλισμού και στους συγκεκριμένους τρόπους βλάβης που φιλοδοξεί να παρατηρήσει η ομάδα συντήρησης. Οι μετρήσεις από τους αισθητήρες, μεταφέρονται σε μια κατάλληλη βάση δεδομένων στην οποία αρχειοθετούνται ως ιστορικά δεδομένα. Από αυτήν τη βάση δεδομένων, αντλούνται πληροφορίες για πολύπλευρους σκοπούς και αναλύσεις. (Hansen, 2024)

Ένας τέτοιος σκοπός είναι η παρακολούθηση της κατάστασης (CM), όπου τα δεδομένα εξετάζονται λεπτομερώς για την αξιολόγηση της παρούσας κατάστασης του εξοπλισμού. Αυτή η ανάλυση εκτελείται συμβατικά με αλγόριθμους που διατυπώνονται για να εξακριβωθεί εάν υπάρχει απόκλιση από την φυσιολογική κατάσταση λειτουργίας. Με τον τρόπο αυτό μπορεί να διαπιστωθεί μια λειτουργία σφάλματος και, εάν αφεθεί χωρίς αντιμετώπιση, μπορεί να προχωρήσει σε ενδεχόμενη αποτυχία. Ο στόχος είναι, επομένως, ο εντοπισμός ζητημάτων πριν από την κλιμάκωσή τους σε δυσλειτουργίες του εξοπλισμού, παρακάμπτοντας έτσι δαπανηρές επισκευές και διακοπές λειτουργίας. (Hansen, 2024)

Μια άλλη θεμελιώδης πτυχή του PdM είναι η αξιολόγηση της εναπομένουσας ωφέλιμης ζωής (RUL) του εξοπλισμού. Αυτή η διαδικασία αξιοποιεί το ιστορικό των δεδομένων παράλληλα με εκείνα που λαμβάνει σε πραγματικό χρόνο, για την πρόβλεψη της διάρκειας για την οποία ένα κομμάτι εξοπλισμού μπορεί να λειτουργήσει προτού καταλήξει σε αποτυχία. Η εκτίμηση RUL είναι ζωτικής σημασίας για τη μακροπρόθεσμη στρατηγική συντήρησης, καθώς μπορεί να βοηθήσει στον προσδιορισμό του βέλτιστου προγραμματισμού για τη συντήρηση ή την αντικατάσταση εξαρτημάτων, αποφεύγοντας έτσι απρόβλεπτες βλάβες. Μπορεί επίσης να διευκολύνει τον εντοπισμό των στοιχείων που πρέπει να δοθεί προτεραιότητα για τη συντήρηση με βάση τα εκτιμώμενα επίπεδα φθοράς. (Hansen, 2024)

Κατανοώντας την παρούσα κατάσταση του εξοπλισμού και την αναμενόμενη διάρκεια λειτουργίας χωρίς προβλήματα, οι ομάδες συντήρησης μπορούν να οργανώσουν επισκευές και ενέργειες συντήρησης για την ελαχιστοποίηση των λειτουργικών διαταραχών. Στόχος είναι να εκτελούνται αυτές οι δραστηριότητες συντήρησης αποτελεσματικά, και την κατάλληλη στιγμή, μειώνοντας έτσι τον χρόνο διακοπής λειτουργίας και παρατείνοντας τη διάρκεια ζωής του εξοπλισμού. Αυτό

καταλήγει σε συνολικές βελτιώσεις στη διαθεσιμότητα και την αξιοπιστία του εξοπλισμού.

2.2.3 Πλεονεκτήματα του PdM

Τα πλεονεκτήματα του PdM είναι σημαντικά. Συγχρονίζοντας τα χρονοδιαγράμματα συντήρησης με τις πραγματικές απαιτήσεις του εξοπλισμού όπως ενημερώνονται από τις αξιολογήσεις CM και RUL, το PdM ενισχύει την κατανομή των πόρων και μετριάζει την πιθανότητα απρόβλεπτων βλαβών. Δίνει τη δυνατότητα στις ομάδες συντήρησης να συμμετέχουν προληπτικά και όχι αντιδραστικά, προωθώντας έτσι πιο αποτελεσματικές λειτουργίες και ενδεχομένως παρατείνοντας τη συνολική διάρκεια ζωής των μηχανημάτων. Αυτό το στρατηγικό πλαίσιο συντήρησης βελτιώνει τον προγραμματισμό και συμβάλλει σε μια πιο αξιόπιστη και ευέλικτη ροή εργασίας συντήρησης. (Hansen, 2024)

Όσον αφορά την προοπτική του PdM, αναμένεται να γίνει όλο και πιο ζωτικής σημασίας για τις βιομηχανικές λειτουργίες. Με τις εξελίξεις στην τεχνητή νοημοσύνη (AI) και το ML, τα συστήματα PdM αναπτύσσονται με ταχείς ρυθμούς. Η συγχώνευση μεθοδολογιών τεχνητής νοημοσύνης και βαθιάς μάθησης, όπως τα συνελκτικά και επαναλαμβανόμενα νευρωνικά δίκτυα, δίνει τη δυνατότητα στο PdM να διακρίνει περίπλοκα μοτίβα και να προβλέπει δυσλειτουργίες του εξοπλισμού. Τεχνολογίες όπως το Διαδίκτυο των πραγμάτων διευρύνουν τις παραμέτρους του PdM διευκολύνοντας την πιο οικονομική και απλή συλλογή και διάδοση επιχειρησιακών δεδομένων. Η συνεχιζόμενη ψηφιακή ανάπτυξη στον βιομηχανικό τομέα, που συχνά αναφέρεται ως Βιομηχανία 4.0, θέτει τις βάσεις για να αναδειχθεί το PdM ως μια πιο κυρίαρχη στρατηγική συντήρησης. Καθώς ένας μεγαλύτερος αριθμός βιομηχανιών αγκαλιάζει έξυπνες τεχνολογίες και πλαίσια λήψης αποφάσεων βάσει δεδομένων, η εξάρτηση από το PdM προβλέπεται να κλιμακωθεί (Hansen, 2024).

2.2.4 Περιορισμοί και προκλήσεις

Πολυπλοκότητα εφαρμογής: Η δημιουργία ενός συστήματος PdM απαιτεί σημαντικές επενδύσεις στην τεχνολογία και τις εξειδικευμένες γνώσεις, οι οποίες μπορεί να αποτελέσουν εμπόδιο για ορισμένους οργανισμούς (Nunes et al., 2023).

Χρήση δεδομένων: Αν και τα συστήματα PdM παράγουν εκτεταμένους όγκους δεδομένων, η αποτελεσματική εφαρμογή αυτών των δεδομένων για πληροφορίες με δυνατότητα δράσης παραμένει μια σημαντική πρόκληση, που συχνά απαιτεί προηγμένους αλγόριθμους ανάλυσης και μηχανικής μάθησης (Nunes et al., 2023).

Προσαρμογή ανά τομέα: Η αποτελεσματικότητα του PdM μπορεί να διαφέρει σε διάφορους κλάδους, απαιτώντας έτσι λύσεις που είναι ειδικά προσαρμοσμένες σε διαφορετικά λειτουργικά περιβάλλοντα (Nunes et al., 2023).

Συμπερασματικά, ενώ η προγνωστική συντήρηση παρουσιάζει σημαντικά πλεονεκτήματα έναντι αντιδραστικών και προληπτικών στρατηγικών, η μετάβαση στο PdM μπορεί να είναι περίπλοκη. Ωστόσο, τα μελλοντικά οφέλη όσον αφορά την αποτελεσματικότητα, τη μείωση του κόστους και τη βιωσιμότητα το καθιστούν μια ελκυστική επιλογή για τις σύγχρονες βιομηχανίες.

Κεφάλαιο 3. Μεθοδολογία

3.1 Δεδομένα για προγνωστικά μοντέλα

Οι ροές δεδομένων που παράγονται από τα συστήματα SCADA μπορούν να αξιοποιηθούν για την υποστήριξη στρατηγικών προληπτικής συντήρησης τροφοδοτώντας δεδομένα παρακολούθησης του συστήματος σε ένα μοντέλο πρόβλεψης. Ένα από τα πλεονεκτήματα της χρήσης δεδομένων SCADA για την κατασκευή μοντέλων πρόβλεψης είναι ότι υπάρχουν πολλά χαρακτηριστικά που καταγράφονται για κάθε ανεμογεννήτρια, πράγμα που σημαίνει ότι υπάρχουν πολλές επιλογές εισόδου από τις οποίες μπορεί να κατασκευαστεί ένα μοντέλο πρόβλεψης. Ωστόσο, τα δεδομένα SCADA περιέχουν συνήθως περιττά χαρακτηριστικά και, ως εκ τούτου, απαιτείται φιλτράρισμα χαρακτηριστικών, έτσι ώστε να απομένει μόνο ένα υποσύνολο αυτών που σχετίζονται με τον στόχο του μοντέλου πρόβλεψης (Canizo et al., 2017).

Τα σύνολα δεδομένων μπορούν να περιέχουν δύο τύπους δεδομένων: δεδομένα λειτουργίας και δεδομένα κατάστασης. Τα λειτουργικά δεδομένα αναφέρονται στην απόδοση λειτουργίας σε πραγματικό χρόνο της ανεμογεννήτριας και των εξαρτημάτων της, όπως η παραγόμενη ισχύς, η ταχύτητα του ρότορα, οι θερμοκρασίες των εξαρτημάτων και άλλα χαρακτηριστικά (Canizo et al., 2017). Τα δεδομένα κατάστασης περιλαμβάνουν κωδικούς κατάστασης και περιγραφές που καταγράφονται τη στιγμή που το σύστημα αντιμετωπίζει μια αλλαγή κατάστασης.

3.2 Επεξεργασία δεδομένων

3.2.1 Διαχείριση δεδομένων που λείπουν

Ενώ υπάρχουν αλγόριθμοι μηχανικής μάθησης όπως το eXtreme Gradient Boosting (XGBoost) που είναι σε θέση να χειριστούν δεδομένα που λείπουν, σε αυτήν την εργασία, η εστίαση είναι στις εφαρμογές του LSTM που δεν μπορεί να επεξεργαστεί το σύνολο δεδομένων με τιμές που λείπουν. Οι πιο δημοφιλείς προσεγγίσεις για την

αντιμετώπιση των τιμών αυτών είναι η διαγραφή τους και η αντικατάστασή τους με τον μέσο όρο ή την πιο κοινή τιμή.

3.2.2 Αλλαγή κλίμακας δεδομένων

Η κλιμάκωση δεδομένων αναφέρεται στην κανονικοποίηση του συνόλου δεδομένων ως μέρος της προετοιμασίας τους πριν από την εκτέλεση αλγορίθμων μηχανικής μάθησης. Οποιαδήποτε διαδικασία κανονικοποίησης χαρακτηρίζεται από την αλλαγή των τιμών των αριθμητικών στηλών σε μια κοινή κλίμακα, διατηρώντας ταυτόχρονα τις διαφορές στο εύρος τιμών. Ο κύριος στόχος της εφαρμογής αυτής της τεχνικής είναι να διασφαλιστεί ότι οι αλγόριθμοι που υπολογίζουν την απόσταση μεταξύ σημείων μέσω της Ευκλείδειας απόστασης θα υπολογίζουν όλα τα χαρακτηριστικά ισότιμα, αντί να κυριαρχεί στους υπολογισμούς ένα χαρακτηριστικό με μεγαλύτερες τιμές. Σε αυτή τη εργασία, θα γίνει χρήση του Z-score, επίσης γνωστή ως τυποποίηση, η οποία λειτουργεί αναπροσαρμόζοντας τα χαρακτηριστικά ώστε να έχουν τις ιδιότητες της κανονικής κατανομής με $\mu = 0$ και $\sigma = 1$. Η τυποποίηση υπολογίζεται σύμφωνα με τον ακόλουθο τύπο:

$$z = \frac{x - \mu}{\sigma},$$

όπου z είναι η τυποποιημένη τιμή ενός δείγματος x_i

3.2.3 Ανισορροπία συνόλου δεδομένων

Υψηλή ανισορροπία εμφανίζεται σε οποιαδήποτε πραγματική εφαρμογή όπου ο στόχος είναι η ανίχνευση σπάνιων αλλά κρίσιμων δειγμάτων. Υπάρχουν δύο κύριες στρατηγικές που χρησιμοποιούνται για την αντιμετώπιση σοβαρών ανισορροπιών. Αυτές διακρίνονται σε τεχνικές σε επίπεδο δεδομένων και σε επίπεδο αλγορίθμων. Οι προσεγγίσεις σε επίπεδο δεδομένων μεταβάλλουν την κατανομή των κλάσεων μέσω διαδικασιών δειγματοληψίας και οι μέθοδοι αυτές διαχωρίζονται σε κατηγορίες υποδειγματοληψίας και υπερδειγματοληψίας. Οι συνηθέστερες μέθοδοι υπερδειγματοληψίας που χρησιμοποιούνται είναι το SMOTE και το ADASYN. Από την άλλη, οι μέθοδοι σε επίπεδο αλγορίθμου προσαρμόζουν τον ταξινομητή ώστε να λαμβάνει υπόψη το μη ισορροπημένο σύνολο δεδομένων. Η συνηθέστερη προσέγγιση είναι η προσαρμογή του κόστους εσφαλμένης κλασματοποίησης ή του κατώτατου ορίου απόφασης. Σε αυτή την διπλωματική, χρησιμοποιούνται βάρη

τάξεων για να χειριστούν τις μη ισορροπημένες τάξεις. Το κύριο αποτέλεσμα της εφαρμογής των συντελεστών στάθμισης των τάξεων είναι να δοθεί μεγαλύτερη έμφαση στις μειονοτικές τάξεις και έτσι να επιτραπεί σε έναν αλγόριθμο να εκπαιδευτεί εξίσου καλά από όλες τις τάξεις.

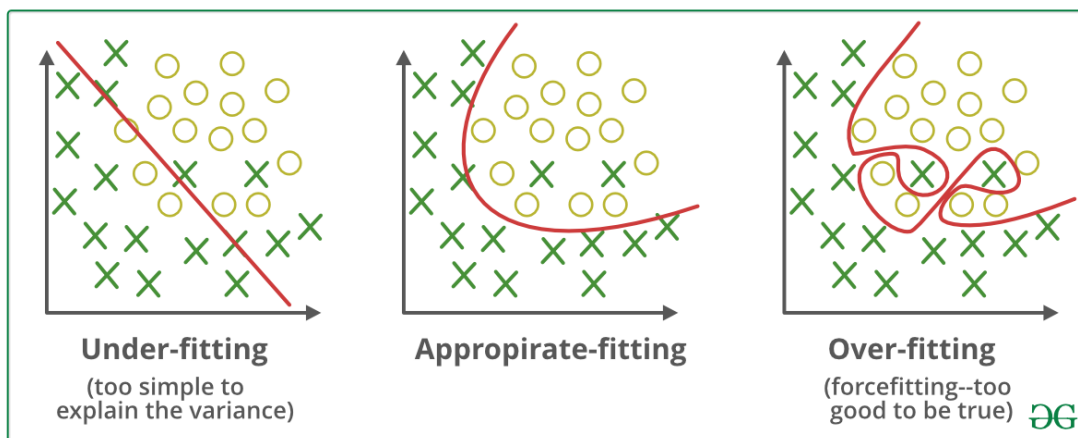
3.2.4 Μοντέλα μηχανικής μάθησης

Η μηχανική μάθηση αποτελεί έναν εξειδικευμένο τομέα στο ευρύτερο πεδίο της Τεχνητής Νοημοσύνης (AI). Οι αλγόριθμοι μηχανικής μάθησης διακρίνουν μοτίβα μέσα σε σύνολα δεδομένων μέσω ενός από τα τέσσερα καθιερωμένα παραδείγματα μάθησης: εποπτευόμενη, μη εποπτευόμενη, ημι-εποπτευόμενη και ενισχυτική μάθηση. Σε αυτό το πλαίσιο, χρησιμοποιείται η εποπτευόμενη μάθηση, όπου πρωταρχικός στόχος είναι η εξαγωγή ενός γενικού κανόνα που συνάγει τη μεταβλητή-στόχο, y , από τα δεδομένα εκπαίδευσης X .

Η θεμελιώδης πρόκληση στη μηχανική μάθηση έγκειται στην ανάγκη για το εκπαιδευμένο μοντέλο να λειτουργεί αποτελεσματικά σε νέα, προηγουμένως μη ορατά δεδομένα, ένα φαινόμενο που αναφέρεται ως γενίκευση. Κατά συνέπεια, ο απώτερος στόχος ενός αλγορίθμου μηχανικής μάθησης είναι να επιτύχει ένα ελάχιστο σφάλμα γενίκευσης.

Οι καθοριστικοί παράγοντες της απόδοσης ενός μοντέλου είναι η ικανότητά του να ελαχιστοποιεί το σφάλμα εκπαίδευσης και να διασφαλίζει ότι η διαφορά μεταξύ σφαλμάτων εκπαίδευσης και δοκιμών παραμένει ελάχιστη. Όταν πληρούνται και οι δύο αυτές προϋποθέσεις και οι προβλεπόμενες τιμές του μοντέλου ευθυγραμμίζονται στενά με τις πραγματικές τιμές, επιτυγχάνεται καλή προσαρμογή.

Ωστόσο, για την επίτευξη αυτού του τελικού στόχου εμφανίζονται συχνά δύο κύριες προκλήσεις κατά τη διάρκεια της εκπαίδευσης του μοντέλου, η υποπροσαρμογή και η υπερπροσαρμογή.



Εικόνα 3 Απεικόνιση υποπροσαρμογής, υπερπροσαρμογής και επιθυμητής προσαρμογής του μοντέλου στα δεδομένα (ML | Underfitting and Overfitting - GeeksforGeeks, n.d.)

Η υποπροσαρμογή εμφανίζεται όταν ένα μοντέλο αποτυγχάνει να μειώσει επαρκώς τα σφάλματα εκπαίδευσης ή δοκιμών. Η κυρίαρχη αιτία της ανεπαρκούς προσαρμογής είναι ένα μοντέλο που στερείται της πολυπλοκότητας που απαιτείται για την κατανόηση των υποκείμενων σύνθετων δεδομένων.

Αντίθετα, η υπερπροσαρμογή, γνωστή ως υψηλή διακύμανση, είναι ένα πιο συχνό φαινόμενο που συναντάται κατά τη διάρκεια της εκπαίδευσης μοντέλου. Η υπερπροσαρμογή χαρακτηρίζεται από αξιόπαινη απόδοση στο σύνολο εκπαίδευσης, σε αντίθεση με τις χαμηλές επιδόσεις κατά τη φάση δοκιμής νέων δεδομένων. Αυτό το ζήτημα προκύπτει όταν το μοντέλο διαθέτει επαρκή πολυπλοκότητα για να καταγράψει τις λεπτομέρειες του σετ εκπαίδευσης, αλλά αποτυγχάνει να γενικεύσει αποτελεσματικά στο σετ δοκιμών.

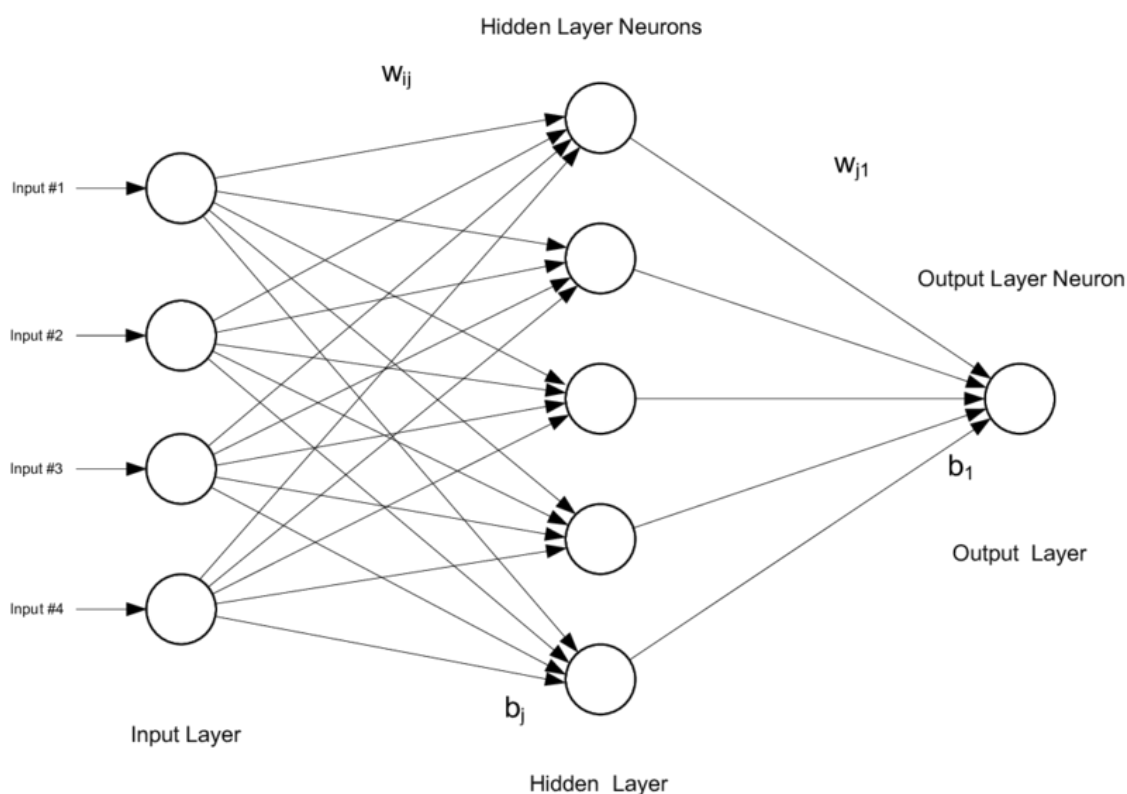
Ενώ η ανεπαρκής προσαρμογή μπορεί να διορθωθεί σχετικά εύκολα αυξάνοντας την πολυπλοκότητα του μοντέλου, η υπερβολική προσαρμογή αποτελεί μια πιο περίπλοκη πρόκληση.

3.3 Artificial Neural Networks

Τα Τεχνητά Νευρωνικά Δίκτυα (ANN) είναι αλγοριθμικές κατασκευές που προσπαθούν να μιμηθούν την λειτουργία του ανθρώπινου εγκεφάλου. Περιλαμβάνουν μια αλληλένδετη σειρά μεμονωμένων στοιχείων επεξεργασίας που

έχουν την ικανότητα να αποκτούν γνώση από την εμπειρία μέσω της αλλαγής των διασυνδέσεων τους. Τα ANN βασίζονται σε ένα σύνολο διασυνδεδεμένων μονάδων ή κόμβων, που αναφέρονται ως τεχνητοί νευρώνες. Σε αυτό το πλαίσιο, χρησιμοποιούνται νευρωνικά δίκτυα τροφοδοσίας (NN). Μια θεμελιώδης αρχιτεκτονική NN τροφοδοσίας απεικονίζεται στο Σχήμα 2.5, που περιλαμβάνει τρία στρώματα: επίπεδα εισόδου, κρυφά και εξόδου. Η επεξεργασία δεδομένων μπορεί να διαδοθεί σε πολλαπλά στρώματα μονάδων. Ωστόσο, σε αντίθεση με τα δίκτυα μακροπρόθεσμης βραχυπρόθεσμης μνήμης (LSTM), που εξετάζονται παρακάτω, δεν ενσωματώνονται διασυνδέσεις ανάδρασης στην αρχιτεκτονική. Τα ANN που αποτελούνται από ρυθμιζόμενα βάρη (w), παραμέτρους μεροληψίας (b) και συναρτήσεις ενεργοποίησης (f), υπολογίζουν τις εξόδους τους (y) με τον ακόλουθο τρόπο:

$$y = f(w_T X + b)$$



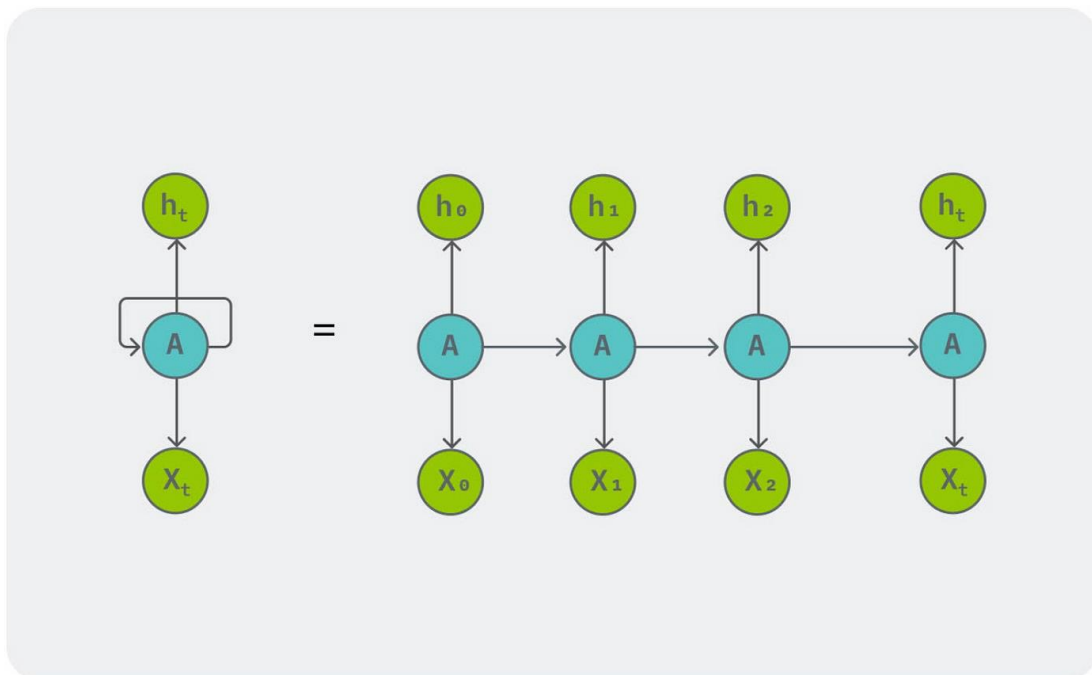
Εικόνα 4 Ένα παράδειγμα μιας βασικής ανατροφοδότησης NN. (A Typical Artificial Neural Network (ANN) | Download Scientific Diagram, n.d.)

Τα Τεχνητά Νευρωνικά Δίκτυα (ANN) μπορούν να εκπαιδευτούν μέσω ενός αλγορίθμου οπισθοδρόμησης, ο οποίος είναι μια ικανή τεχνική για τον υπολογισμό κλίσεων χρησιμοποιώντας έναν κανόνα αλυσίδας παραγώγων. Μόλις εξακριβωθούν οι διαβαθμίσεις, ένας αλγόριθμος βελτιστοποίησης εφαρμόζει τις διαβαθμίσεις για να καθοδηγήσει το μοντέλο.

3.3.1.1 RNNs

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) είναι μια κατηγορία νευρωνικών δικτύων που είναι ιδιαίτερα κατάλληλα για ανάλυση δεδομένων χρονοσειρών. Τα RNN έχουν σχεδιαστεί για να χειρίζονται ακολουθίες δεδομένων με διαφορετικά μήκη και μπορούν να συλλάβουν χρονικές εξαρτήσεις. Αυτό είναι απαραίτητο σε δεδομένα χρονοσειρών, όπως δεδομένα αισθητήρων ανεμογεννητριών, καθώς επιτρέπει στο μοντέλο να μαθαίνει μοτίβα που εκτείνονται σε διαφορετικά χρονικά βήματα. (Janakauskas et al., 2023)

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) προτάθηκαν το 1986 ως ένας τρόπος χειρισμού διαδοχικών δεδομένων, τα οποία διακρίνονται από τη σημασία της σειράς των πληροφοριών. Το θεμελιώδες χαρακτηριστικό ενός RNN είναι η ικανότητα να ξεδιπλώνεται σε μια αλυσίδα νευρωνικών δικτύων τροφοδοσίας, όπου κάθε στρώμα αντιστοιχεί σε μία αλληλουχία, όπως φαίνεται στην Εικόνα 5.



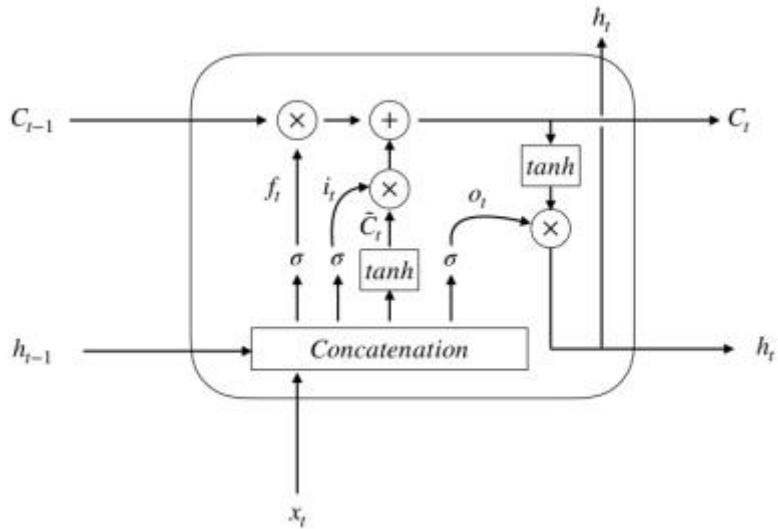
Εικόνα 5 Ξεδίπλωμα ενός RNN (Vijay Gadre, 2023)

Το ξεδίπλωμα ενός RNN έχει ως αποτέλεσμα την κοινή χρήση των παραμέτρων σε όλη τη βαθιά δομή του δικτύου, η οποία αποδίδει καλά για διάφορα μήκη ακολουθίας δεδομένων. Δεδομένου ότι τα RNNs ξεδιπλώνονται σε μια σειρά νευρωνικών δικτύων τροφοδοσίας, η οπισθοδιάδοση θα μπορούσε να είναι μια σωστή επιλογή εκπαίδευσης. Ωστόσο, λόγω του ότι τα RNNs είναι πολύ ευαίσθητα σε οποιεσδήποτε αλλαγές στις παραμέτρους, η τακτική οπισθοδιάδοση οδηγεί σε ένα εξαφανιζόμενο πρόβλημα κλίσης. Η λύση σε αυτό είναι η χρήση δικτύων LSTM, τα οποία είναι ένα από τα κύρια μοντέλα που χρησιμοποιούνται σε αυτό το έργο (Hochreiter & Schmidhuber, 1997).

3.3.1.2 LSTM Δίκτυα

Στη διπλωματική εργασία θα χρησιμοποιηθεί ένα μοντέλο βασισμένο σε δίκτυο LSTM. Τα LSTM (Long Short-Term Memory) είναι μια ειδική κατηγορία επαναλαμβανόμενων νευρωνικών δικτύων (RNNs) που εφαρμόζουν έναν αποτελεσματικό αλγόριθμο για τη διατήρηση του σφάλματος.(Udo & Muhammad, 2021).

Ένα από τα κύρια πλεονεκτήματα των LSTM σε σύγκριση με τα παραδοσιακά RNNs είναι η ικανότητά τους να διατηρούν χρονικές εξαρτήσεις μεταξύ των εισόδων για μεγαλύτερες περιόδους. Αντίθετα, τα RNNs συχνά περιορίζονται στην παρακολούθηση μόνο πρόσφατων εξαρτήσεων λόγω του φαινομένου της εκθετικής διάσπασης κλίσης, γνωστού και ως vanishing gradient problem. Αυτό το φαινόμενο δυσχεραίνει την εκτέλεση του αλγορίθμου backpropagation through time (BPTT) για μακρές ακολουθίες. Το παρακάτω σχήμα δείχνει πώς οργανώνεται εσωτερικά το δίκτυο LSTM(Udo & Muhammad, 2021).



Εικόνα 6 Σχηματική αναπαράσταση του κυττάρου LSTM τη χρονική στιγμή t .

Αποτελείται από τρεις πύλες (δηλαδή, την πύλη εισόδου, την πύλη εξόδου και η πύλη λήθης) διέπεται από τις ακόλουθες εξισώσεις.

Το X_t είναι το διάνυσμα εισόδου του παρόντος χρόνου, το οποίο είναι τα δεδομένα εισόδου στο μοντέλο LSTM τη χρονική στιγμή t . h_{t-1} είναι το διάνυσμα εξόδου παρελθοντικού χρόνου και το c_{t-1} αντιπροσωπεύει την κυτταρική κατάσταση του παρελθόντος. Στην Εικόνα 6 f_t και i_t είναι η πύλη λήθης και η πύλη εισόδου αντίστοιχα, οι οποίες χρησιμοποιούνται για τον έλεγχο της κατάστασης του κελιού του μοντέλου. Με άλλα λόγια, οι πύλες λήθης και οι πύλες εισόδου έχουν σχεδιαστεί για να περιορίζουν τη ροή πληροφοριών. Το σ είναι μια σιγμοειδής συνάρτηση που αποφασίζει ποιες τιμές θα ενημερωθούν στην κατάσταση κελιού και εξάγει έναν αριθμό μεταξύ 0 και 1 για κάθε αριθμό στην κατάσταση κελιού c_{t-1} . Όπου το 1 αντιπροσωπεύει τις τιμές που πρέπει οπωσδήποτε να διατηρήσει ενώ το 0 αντιπροσωπεύει τις τιμές που πρέπει να μην συγκρατήσει. Η πύλη λήθης ελέγχει τις πληροφορίες της προηγούμενης κυτταρικής κατάστασης c_{t-1} που μεταδίδονται στην παρούσα κυτταρική κατάσταση. Η διαδικασία μπορεί να εξηγηθεί με τον παρακάτω τύπο:

$$f_t = g(W_f \cdot [h_{t-1}, X_t] + b_f)$$

όπου $g(\cdot)$ είναι η συνάρτηση ενεργοποίησης που εκτελεί τη σιγμοειδή μη γραμμική συνάρτηση, W_f αντιπροσωπεύει τον πίνακα βάρους της πύλης λήθης, b_f είναι το

διάνυσμα μεροληψίας της πύλης λήθης και $[h_{t-1}, X_t]$ είναι ο συνδυασμός του διανύσματος εξόδου παρελθόντος χρόνου h_{t-1} και του διανύσματος εισόδου παρόντος χρόνου X_t . Η πύλη εισόδου ελέγχει τις παρούσες πληροφορίες εισόδου X_t που μεταδίδονται στην τρέχουσα κατάσταση κυψέλης c_t , που φαίνεται από

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$$

όπου W_i είναι ο πίνακας βάρους της πύλης εισόδου, b_i είναι το διάνυσμα μεροληψίας της πύλης εισόδου. Για να καταγραφεί η κατάσταση της τρέχουσας εισόδου, το c'_t , μπορεί να υπολογιστεί όπως φαίνεται στο

$$c'_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c)$$

όπου W_c και b_c είναι ο πίνακας βάρους και το διάνυσμα μεροληψίας, αντίστοιχα, \tanh είναι η υπερβολική εφαπτομένη συνάρτηση (η οποία κατανέμει τις τιμές της κατάστασης κελιού μεταξύ -1 και 1). Η παρούσα κατάσταση c_t μπορεί στη συνέχεια να ληφθεί συνδυάζοντας τόσο την πύλη λήθης όσο και την πύλη εισόδου, που περιγράφεται από

$$c_t = f_t * c_{t-1} + i_t * c'_t$$

Όπου το σύμβολο $*$ ορίζει τον πολλαπλασιασμό στοιχείων μεταξύ διανυσμάτων, η ροή πληροφοριών από την παρούσα κατάσταση c_t ελέγχεται από την πύλη εξόδου O_t στην τρέχουσα έξοδο, που περιγράφεται από

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)$$

όπου W_o είναι ο πίνακας βάρους, b_o είναι το διάνυσμα μεροληψίας. Τέλος, η πύλη εξόδου O_t και η τρέχουσα κατάσταση κυψέλης c_t καθορίζουν την έξοδο του μοντέλου LSTM όπως φαίνεται παρακάτω

$$h_t = O_t * \tanh(c_t)$$

3.4 Βελτιστοποίηση Υπερπαραμέτρων

Ένα σύνολο υπερπαραμέτρων που επηρεάζει την απόδοση ενός ANN είναι το μέγεθος παρτίδας και ο αριθμός των εποχών. Ο **αριθμός των εποχών**, ο οποίος αποτελείται από μία ή περισσότερες παρτίδες, είναι μια υπερπαραμέτρος που καθορίζει τον αριθμό των φορών που ο αλγόριθμος μάθησης θα διαδοθεί σε ολόκληρο το σύνολο δεδομένων εκπαίδευσης. Η λειτουργία ενός μοντέλου για πάρα πολλές εποχές θα οδηγήσει αναπόφευκτα σε υπερβολική προσαρμογή. Προκειμένου να αντιμετωπιστεί αυτός ο περιορισμός, χρησιμοποιείται μια τεχνική κανονικοποίησης γνωστή ως πρόωρη διακοπή. Η πρόωρη διακοπή διακόπτει τη διαδικασία εκμάθησης εάν το σφάλμα στο σύνολο επικύρωσης δεν ελαχιστοποιείται για έναν καθορισμένο αριθμό εποχών (J. Brownlee, 2019).

Το **μέγεθος παρτίδας** είναι μια υπερπαραμέτρος που καθορίζει τον αριθμό των δειγμάτων που πρέπει να επεξεργαστούν πριν από την ενημέρωση των παραμέτρων του εσωτερικού μοντέλου. Το μέγεθος της παρτίδας ρυθμίζεται κατά τη διάρκεια της διαδικασίας εκμάθησης του μοντέλου (J. Brownlee, 2019).

Εκτός από την πρόωρη διακοπή, μια άλλη τεχνική που χρησιμοποιείται σε αυτή τη διατριβή για την πρόληψη της υπερπροσαρμογής είναι η **εγκατάλειψη** (dropout). Το dropout είναι ένας υπερπαραμέτρος που χρησιμοποιείται κυρίως σε νευρωνικά δίκτυα, ιδιαίτερα σε μοντέλα βαθιάς μάθησης, για να μειώσει την υπερπροσαρμογή (overfitting) κατά την εκπαίδευση. Η παράμετρος αυτή λειτουργεί τυχαία απενεργοποιώντας ένα ποσοστό των νευρώνων σε κάθε βήμα εκπαίδευσης, μειώνοντας έτσι την αλληλεξάρτηση μεταξύ των νευρώνων. Το dropout αποτρέπει τα μοντέλα από το να προσαρμόζονται υπερβολικά στα δεδομένα εκπαίδευσης και τα ενισχύει στο να γενικεύουν καλύτερα σε νέα δεδομένα.

Η τιμή του dropout rate, δηλαδή το ποσοστό των νευρώνων που απενεργοποιούνται, είναι μια σημαντική παράμετρος που πρέπει να ρυθμιστεί σωστά. Συνήθως κυμαίνεται μεταξύ 0.1 και 0.5, με την τιμή 0.5 να είναι συνηθισμένη για κρυφά στρώματα. Οι χαμηλότερες τιμές διατηρούν περισσότερους ενεργούς νευρώνες και μπορεί να είναι πιο κατάλληλες για απλούστερα μοντέλα, ενώ υψηλότερες τιμές μπορούν να βοηθήσουν στη βελτίωση της γενίκευσης σε πιο σύνθετα μοντέλα. (Nitish Srivastava, 2014).

Ο αλγόριθμος ADAptive BOOSTing (AdaBoost) είναι μια τεχνική μηχανικής μάθησης που ανήκει στην κατηγορία των αλγορίθμων ενίσχυσης (boosting) και εισήχθη από τους Yoan Freund και Robert Schapire το 1996. Η βασική αρχή της ενίσχυσης είναι η δημιουργία ενός ισχυρού ταξινομητή από έναν συνδυασμό αδύναμων ταξινομητών, δηλαδή ταξινομητών που έχουν ελαφρώς καλύτερη απόδοση από μια τυχαία εικασία. Το AdaBoost λειτουργεί προσαρμοστικά, δίνοντας βάρη σε κάθε δείγμα εκπαίδευσης και αναπροσαρμόζοντας αυτά τα βάρη σε κάθε επόμενο γύρο μάθησης. Στόχος της αναπροσαρμογής είναι να δοθεί μεγαλύτερη έμφαση στα δείγματα που ταξινομούνται εσφαλμένα από τον προηγούμενο ταξινομητή, ώστε οι επόμενοι ταξινομητές να επικεντρωθούν στη βελτίωση αυτών των ταξινομήσεων. Με αυτόν τον τρόπο, ο αλγόριθμος ενισχύει διαδοχικά την απόδοση του συστήματος, επιτυγχάνοντας υψηλότερη ακρίβεια και αποτελεσματικότητα στη διαδικασία μάθησης, μετατρέποντας το σύνολο των αδύναμων ταξινομητών σε έναν ισχυρότερο μοντέλο (Yoan Freund, 1996).

Το μοντέλο AdaBoost λαμβάνει ως είσοδο ένα σύνολο εκπαίδευσης $\{(x_i, y_i)\}_{i=1}^N$, όπου $x_i \in \mathbb{R}^D$ και $y_i \in \{-1, +1\}$. Με βάση έναν ασθενή ταξινομητή, η $G_m(x) \in \{-1, +1\}$, μαζί με την ακόλουθη 0-1 συνάρτηση απώλειας I (Yoan Freund, 1996):

$$I(G_m(x), y) = \begin{cases} 0 & \text{if } f_m(x_i) = y_i \\ 1 & \text{if } f_m(x_i) \neq y_i \end{cases}$$

3.5 Επιλογή Δεδομένων

Η **μείωση διαστάσεων** μπορεί να χρησιμοποιηθεί για την επιτάχυνση της εκπαίδευσης μοντέλων επιτρέποντας τη χρήση απλούστερων μοντέλων. Ένα

αποτέλεσμα της μείωσης των διαστάσεων είναι ότι αφαιρεί περιττά ή πολύ συσχετισμένα χαρακτηριστικά και έτσι μειώνει το θόρυβο στα δεδομένα. Επομένως, μπορεί επίσης να χρησιμοποιηθεί για τη μείωση της σύγχυσης και τη βελτίωση της ακρίβειας του μοντέλου κατά την εργασία με σύνθετα δεδομένα.

Η διαδικασία επιλογής σχετικών χαρακτηριστικών και μείωσης της διάστασης μέσω τεχνικών όπως η **Ανάλυση Κύριων Συστατικών (PCA)** μπορεί να βελτιστοποιήσει σημαντικά την απόδοση του μοντέλου εστιάζοντας στα πιο χρήσιμα στοιχεία του συνόλου δεδομένων (Liu et al., 2023).

Μεγιστοποιώντας τη διακύμανση των δεδομένων, εντοπίζει τα κυρίαρχα χαρακτηριστικά σε ένα N-διάστατο σύνολο δεδομένων σε φθίνουσα σειρά μέσω ενός μετασχηματισμού. Έτσι, τα μετασχηματισμένα δεδομένα είναι γραμμικά ανεξάρτητα και αναφέρονται ως τα κύρια συστατικά (PCs). Τα κύρια συστατικά λαμβάνονται συνήθως μέσω της αποσύνθεσης μίας τιμής (SVD) του πίνακα συνδιακύμανσης S ($S = XX^T$) του αρχικού συνόλου δεδομένων X . Για ένα σύνολο δεδομένων X με διάσταση $(n \times p)$, όπου p είναι ο αριθμός των μεταβλητών και n είναι ο αριθμός των δειγμάτων, τα μετασχηματισμένα PCs, Z , υπολογίζονται από τον πίνακα συνδιακύμανσης όπου ικανοποιεί,

$$U^T S U = L$$

όπου L ($|l_1, |l_2, \dots, |l_p$) είναι οι ιδιοτιμές του S , οι οποίες μπορούν να λυθούν από τη χαρακτηριστική εξίσωση $|S - \lambda I| = 0$. Οι ιδιοτιμές $|l_1, |l_2, \dots, |l_p$ είναι επίσης οι διακυμάνσεις κάθε PC και το άθροισμα του L ισούται με το άθροισμα της διακύμανσης των αρχικών μεταβλητών.

Μετά τη λήψη αυτών των ιδιοτιμών, μπορούν να υπολογιστούν τα αντίστοιχα ιδιοδιανύσματα $U = \{u_i\}$, όπου u_i είναι στήλη U , $u_i = (u_{1i}, u_{2i}, \dots, u_{pi})$, $i = 1, \dots, p$. Το ιδιοδιάνυσμα U αναφέρεται ως φορτίσεις, που αντιπροσωπεύουν τις συσχετίσεις μεταξύ των μεταβλητών και των PCs. Η σχέση μεταξύ των PCs, Z (z_1, z_2, \dots, z_p) και του αρχικού συνόλου δεδομένων X ($n \times p$) εκφράζεται ως $Z = UX$. Έχει αποδειχθεί ότι, με τη διατήρηση των PCs q ($q < p$), η διαστασιολόγηση των δεδομένων μπορεί να μειωθεί σημαντικά, θυσιάζοντας μόνο μικρή μεταβλητότητα δεδομένων (Y. Wang et al., 2018)

3.6 Μετρητές αξιολόγησης

Οι διευκρινισμένοι αλγόριθμοι εκπαιδεύονται και εφαρμόζονται στο σύνολο αξιολόγησης, μετά από ποια φάση αξιολογούνται σύμφωνα με τρία κριτήρια: ακρίβεια, ανάκληση και βαθμολογία F1. Οι μετρήσεις υπολογίζονται με βάση τιμές στον πίνακα σύγχυσης που απεικονίζεται στην Εικόνα 7, ο οποίος δεν είναι ένα ρητό μέτρο απόδοσης καθαυτό, αλλά μάλλον μια σύνθεση των αποτελεσμάτων πρόβλεψης σε ένα ζήτημα ταξινόμησης. Κατά συνέπεια, τα διαγώνια κελιά εμφανίζουν την ποσότητα των περιπτώσεων για τις οποίες η πραγματική ετικέτα αντιστοιχεί στις εκτιμώμενες ετικέτες, ενώ τα μη διαγώνια κελιά απεικονίζουν εκείνα που έχουν επισημανθεί εσφαλμένα από τον εκπαιδευμένο ταξινομητή (Emin Elmar, 2019). Στην Εικόνα 7, οι συντομογραφίες υποδηλώνουν τα εξής:

- Πραγματικά θετικά (TP): Ποσότητα αποτυχιών που προβλέπονται με ακρίβεια.
- Εσφαλμένα θετικά (FP): Ποσότητα προβλεπόμενων βλαβών που δεν ανταποκρίνονται στην πραγματικότητα
- Πραγματικά αρνητικά (TN): Ποσότητα μη αποτυχιών που προβλέπονται σωστά ότι δεν συμβαίνουν.
- Εσφαλμένα αρνητικά (FN): Ποσότητα μη αποτυχιών που προβλέπονται εσφαλμένα ότι συμβαίνουν.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Εικόνα 7 Confusion matrices

3.6.1 Μετρητές ταξινόμησης

Αυτές οι μετρήσεις χρησιμοποιούνται όταν ο στόχος του μοντέλου είναι να προβλέψει κατηγορικά αποτελέσματα, όπως εάν μια τουρμπίνα θα αποτύχει ή όχι (αστοχία έναντι μη αστοχίας).

Ευστοχία - Accuracy

- Ορισμός: Η ευστοχία είναι ο λόγος των σωστά προβλεπόμενων περιπτώσεων (τόσο των αποτυχιών όσο και των μη αποτυχιών) προς τον συνολικό αριθμό των περιπτώσεων.
- Τύπος: $Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Samples}$
- Περίπτωση χρήσης: Αν και χρήσιμη, η ακρίβεια μπορεί να είναι παραπλανητική σε μη ισορροπημένα σύνολα δεδομένων (π.χ. περισσότερα συμβάντα μη αποτυχίας παρά αποτυχίας).

Ακρίβεια - Precision

- Ορισμός: Η ακρίβεια μετρά το ποσοστό των προβλεπόμενων αποτυχιών που ήταν πραγματικές αποτυχίες (πόσες από τις προβλεπόμενες θετικές ήταν σωστές).
- Τύπος: $Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$
- Περίπτωση χρήσης: Η υψηλή ακρίβεια είναι κρίσιμη όταν τα ψευδώς θετικά αποτελέσματα (εσφαλμένη πρόβλεψη βλάβης) είναι δαπανηρά, όπως περιττές ενέργειες συντήρησης.

Ανάκληση (ευαισθησία) - Recall (Sensitivity)

- Ορισμός: Η ανάκληση (ή ευαισθησία) μετρά το ποσοστό των πραγματικών αστοχιών που προβλέφθηκαν σωστά από το μοντέλο (πόσα από τα πραγματικά θετικά αναγνωρίστηκαν σωστά).

- Τύπος: $Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$
- Περίπτωση χρήσης: Η υψηλή ανάκληση είναι σημαντική όταν η απώλεια μιας πραγματικής βλάβης (ψευδώς αρνητικά) θα μπορούσε να οδηγήσει σε δαπανηρά ή επικίνδυνα αποτελέσματα, όπως απροσδόκητες διακοπές λειτουργίας στροβίλων.

F1- Score

- Ορισμός: Το F1-score είναι ο αρμονικός μέσος όρος ακρίβειας και ανάκλησης, παρέχοντας ισορροπία μεταξύ των δύο, ειδικά σε περιπτώσεις μη ισορροπημένων συνόλων δεδομένων.
- Τύπος: $F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$
- Περίπτωση χρήσης: Το F1-score είναι χρήσιμο όταν τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά είναι σημαντικά, εξισορροπώντας την αντιστάθμιση μεταξύ ακρίβειας και ανάκλησης.

Περιοχή κάτω από την καμπύλη ROC (AUC-ROC)

- Ορισμός: Η καμπύλη ROC απεικονίζει το πραγματικό θετικό ποσοστό (ανάκληση) έναντι του ψευδώς θετικού ποσοστού και η AUC μετρά τη συνολική απόδοση του μοντέλου σε όλα τα όρια ταξινόμησης.
- Περίπτωση χρήσης: Το AUC-ROC είναι μια καλή μέτρηση για την αξιολόγηση μοντέλων κατά την αντιμετώπιση μη ισορροπημένων δεδομένων, καθώς μετρά την ικανότητα του μοντέλου να διακρίνει μεταξύ κλάσεων.

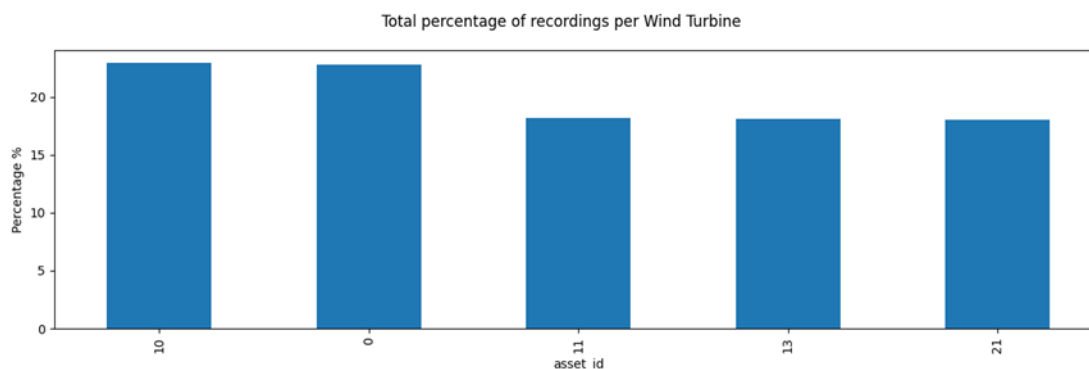
Κεφάλαιο 4. Υλοποίηση

4.1 Εισαγωγή και δεδομένα

Στόχος της παρούσας διπλωματικής αποτελεί η υλοποίηση ενός προβλήματος προβλεπτικής συντήρησης σε ανεμογεννήτριες. Για τον σκοπό αυτό επιλέχθηκε ένα σετ δεδομένων που είναι δημόσια διαθέσιμο (*Wind Turbine SCADA Data For Early Fault Detection, 2024*) και περιγράφεται αναλυτικά στην δημοσίευση.

Αρχικά, ακολουθεί μια ανάλυση των δεδομένων που αξιοποιούνται στην υλοποίηση της διπλωματικής εργασίας. Στα πλαίσια της διπλωματικής αξιοποιήθηκαν τα δεδομένα μόνο από την 1η φάρμα (Wind Farm A), που περιέχει δεδομένα από 5 ανεμογεννήτριες σε μια φάρμα στην Πορτογαλία.

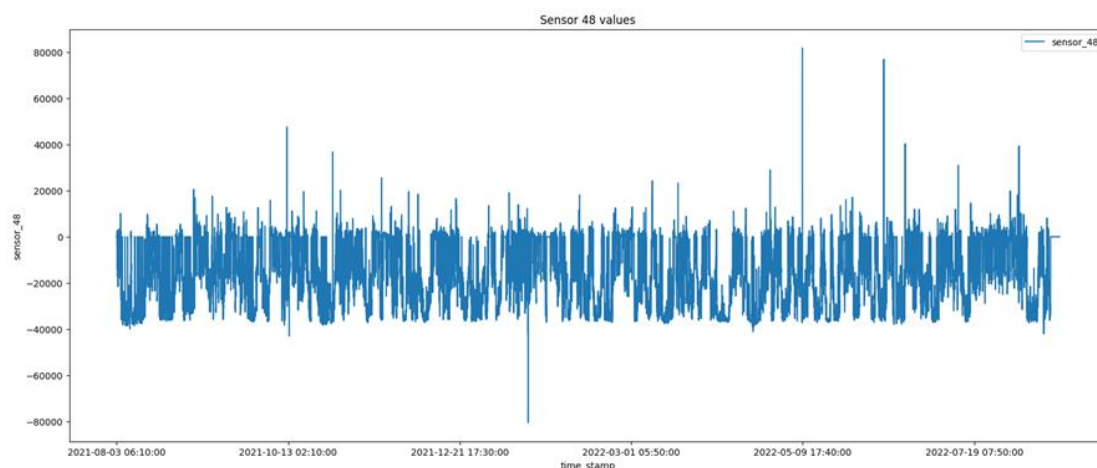
Τα δεδομένα της 1ης φάρμας αποτελούνται από 22 επιμέρους σετ δεδομένων για τις 5 ανεμογεννήτριες, σχετικά ίσα μοιρασμένα ανά ανεμογεννήτρια, όπως φαίνεται στην Εικόνα 8. Πιο συγκεκριμένα, 22,91% των μετρήσεων αφορούν την ανεμογεννήτρια 10, 22,77% αφορούν την 0, 18,22% αφορούν την 11, 18,09% αφορούν την 13 και 18,01% αφορούν την 21. Οι μετρήσεις είναι δεδομένα που προκύπτουν από της καταγραφές χρονοσειρών από σύστημα ελέγχου SCADA (Supervisory Control And Data Acquisition) με χρονική ανάλυση 10 λεπτών.



Εικόνα 8 Το συνολικό ποσοστό εγγραφών ανά ανεμογεννήτρια στο σετ δεδομένων από την 1η φάρμα.

Ακόμη, το σετ της Wind Farm A διαθέτει 86 χαρακτηριστικά (features) που περιλαμβάνουν μετρήσεις από αισθητήρες, στατιστικά μεγέθη αυτών των μετρήσεων (μέση τιμή, μέγιστο, ελάχιστο, τυπική απόκλιση), καθώς επίσης και τον χρόνο καταγραφής, ένα ID για την ανεμογεννήτρια, μια τιμή που υποδεικνύει εάν

ανήκει στο σετ εκπαίδευσης ή όχι και ένα status-ID για την κατάσταση στην οποία εκείνη την στιγμή βρίσκεται η ανεμογεννήτρια, η οποία θα αναλυθεί περαιτέρω στην συνέχεια. Ένα παράδειγμα από την χρονοσειρά του αισθητήρα 48 σε ένα από τα 22 σετ δεδομένων, φαίνεται στην Εικόνα 9.



Εικόνα 9 Χρονοσειρά ενός από τα 22 σετ δεδομένων της wind farm A για τον αισθητήρα 48.

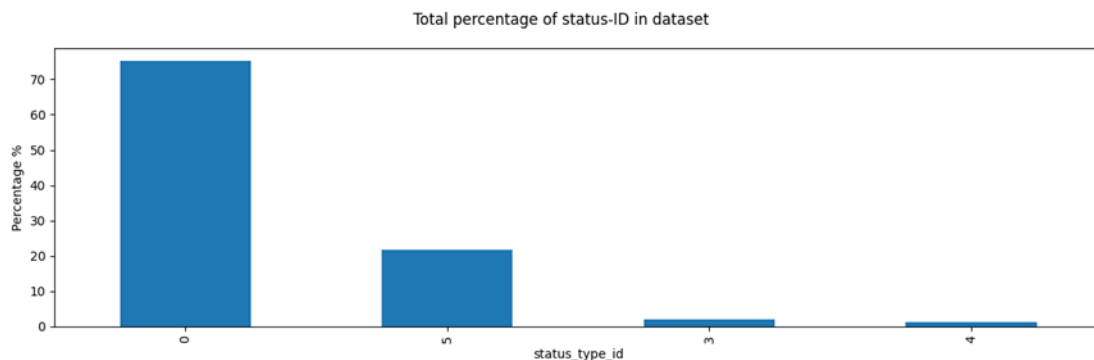
Τα 22 σετ της wind Farm A είναι χωρισμένα σε 11 σετ που παρουσιάζουν φυσιολογική λειτουργία οι ανεμογεννήτριες και 11 σετ που παρουσιάζουν κάποια ελαττωματική συμπεριφορά. Η κατάσταση στην οποία βρίσκεται κάθε φορά η ανεμογεννήτρια έχει τιμές από 0 έως 5 (status-ID). Πιο αναλυτικά, στον Πίνακα 1 φαίνεται σε τι αντιστοιχεί η κάθε μια από τις καταστάσεις αυτές.

Status-ID	Περιγραφή	Θεωρείται φυσιολογική κατάσταση
0	Κανονική λειτουργία χωρίς περιορισμούς	Ναι
1	Μειωμένη παραγωγή ενέργειας λόγω περιορισμών ισχύος	Όχι
2	Αδράνεια ανεμογεννήτριας και αναμονή για λειτουργία	Ναι
3	Έπισκευή της ανεμογεννήτριας / η	Όχι

	ομάδα επισκευών βρίσκεται στο πεδίο	
4	Δεν λειτουργεί λόγω σφάλματος ή για άλλους λόγους	Όχι
5	Άλλες καταστάσεις π.χ. έλεγχος συστήματος, στήσιμο, συσσώρευση πάγου ή έκτακτη κατάσταση	Όχι

Πίνακας 1 Περιγραφή των διαφορετικών κατηγοριών για την στήλη status-ID των δεδομένων

Στην Εικόνα 10 φαίνεται η κατανομή των διαφορετικών καταστάσεων στο σύνολο των 22 σετ δεδομένων και για τις 5 ανεμογεννήτριες. Πιο αναλυτικά, τα ποσοστά είναι 75,09% για την κατάσταση 0, 21,76% για την κατάσταση 5, 1,98% για την κατάσταση 3 και 1,17% για την κατάσταση 4.

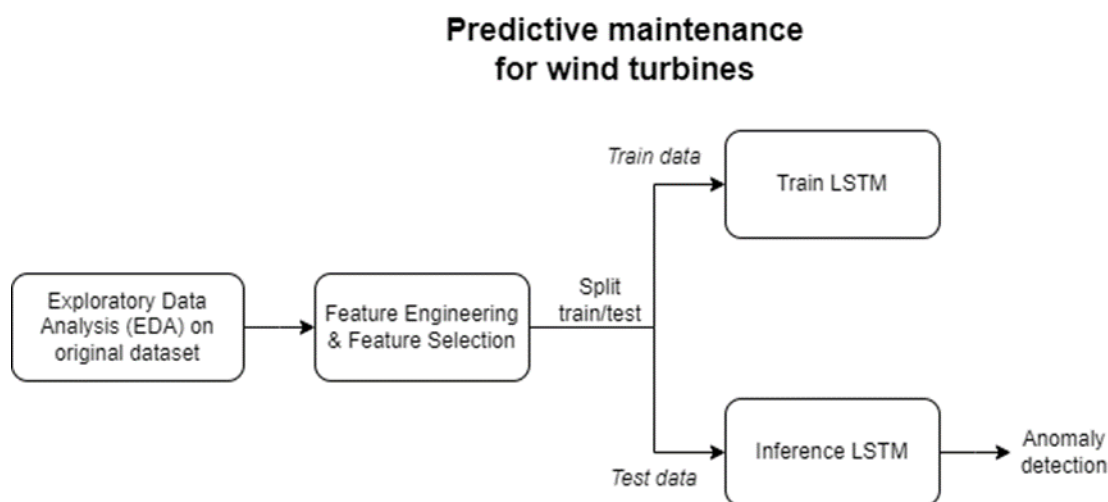


Εικόνα 10 Ποσοστό καταστάσεων εγγραφών επί του συνόλου των δεδομένων και για τις 5 ανεμογεννήτριες.

Έπειτα από την ανάλυση των υπαρχόντων δεδομένων και βάσει της βιβλιογραφίας (Chen et al., 2021; Qian et al., 2019; X. Wang et al., 2022), η δομή της λύσης που ακολουθείται παρουσιάζεται στην Εικόνα 11. Αρχικά, ένα στάδιο ανάλυσης των δεδομένων για να εξεταστούν τα διάφορα στοιχεία και οι κατανομές των δεδομένων, όπως αυτά που αναφέρθηκαν παραπάνω, οδηγεί σε μια καλή κατανόηση του σετ δεδομένων που είναι απαραίτητη. Έπειτα, ακολουθεί το στάδιο του feature engineering και feature selection όπου γίνεται μια επεξεργασία που οδηγεί στην δημιουργία νέων δεδομένων ή στην επιλογή (φιλτράρισμα) των ήδη υπαρχόντων, προκειμένου να οριστικοποιηθεί το τελικό σετ δεδομένων που θα χρησιμοποιηθεί για το μοντέλο μηχανικής μάθησης που θα εκπαιδευτεί. Τέλος, αφού χωριστούν τα

δεδομένα σε train, validation και test set γίνεται η εκπαίδευση και η αξιολόγηση του μοντέλου μηχανικής μάθησης.

Στην προκειμένη περίπτωση, το μοντέλο που επιλέχθηκε είναι το LSTM (Long Short-Term Memory), ένα είδος RNN (Recurrent Neural Network) ιδανικό για δεδομένα χρονοσειρών (σειριακά δεδομένα, ενώ διατηρεί long-term memory), το οποίο εκπαιδεύεται και αξιολογείται στα δεδομένα που αναλύθηκαν προηγουμένως. Ακολουθούν τρία υποκεφάλαια στα οποία αναλύεται το κάθε ένα από τα στάδια που αναφέρθηκαν (προεπεξεργασία δεδομένων, εξαγωγή και επιλογή χαρακτηριστικών, LSTM) και πώς υλοποιήθηκαν στην παρούσα διπλωματική εργασία.



Εικόνα 11 Στάδια προτεινόμενης λύσης για μοντέλο προγνωστικής συντήρησης

Τέλος, πριν την ανάλυση των επιμέρους σταδίων, πρέπει να δοθεί ιδιαίτερη προσοχή στην διατύπωση του προβλήματος, ώστε να γνωρίζει κανείς ποιο είναι το ζητούμενο που επιδιώκει η συγκεκριμένη υλοποίηση. Το πρόβλημα του PdM σε ανεμογεννήτριες έγκειται στην δημιουργία ενός μοντέλου που στην έξοδό του θα προβλέπει μελλοντικά σφάλματα με βάση την τωρινή (ή/και παρελθοντική) κατάσταση (π.χ. αισθητήρες) μιας ανεμογεννήτριας. Συνεπώς, αναπτύσσεται ένα μοντέλο που στην είσοδό του θα λαμβάνει σειριακά δεδομένα από συνεχόμενες χρονικές στιγμές και στην έξοδό του θα παράγει την κατάσταση του συστήματος την αμέσως επόμενη χρονική στιγμή.

Έτσι, το σετ δεδομένων μπορεί να απλοποιηθεί ως εξής. Μετατρέποντας την πληροφορία του status-ID σε δυαδική (binary) μορφή δίνεται η τιμή 0 όταν η λειτουργία της ανεμογεννήτριας είναι φυσιολογική και η τιμή 1 όταν εμφανίζεται κάποιο σφάλμα. Αυτή η τιμή αποτελεί και την ετικέτα (label) του κάθε διάνυσματος στο σετ δεδομένων, που είναι απαραίτητο για ένα πρόβλημα επιβλεπόμενης (supervised) μηχανικής μάθησης. Έτσι, με βάση τα νούμερα που παρουσιάστηκαν παραπάνω, στο σύνολο του σετ 75,09% των εγγραφών παρουσιάζουν φυσιολογική λειτουργία των ανεμογεννητριών και 24,91% περιέχει κάποιο σφάλμα λειτουργίας.

4.2 Προεπεξεργασία δεδομένων

Προκειμένου να μπορέσει να εκπαιδευτεί κάποιο μοντέλο μηχανικής μάθησης, πρέπει πρώτα να προηγηθεί μια κατάλληλη διαδικασία προεπεξεργασίας των δεδομένων. Στο στάδιο αυτό εφαρμόζεται μια σειρά από επεξεργασίες στα σετ δεδομένων, ώστε να ακολουθήσει στην συνέχεια η επιλογή χαρακτηριστικών (features), με σκοπό να εκπαιδευτεί το μοντέλο LSTM.

Αρχικά, όπως έχει ήδη αναφερθεί, τα 22 επιμέρους σετ δεδομένων εμπεριέχουν 86 χαρακτηριστικά, εκ των οποίων 5 είναι περιγραφικά (status_id, time_stamp, asset_id, id, train_test). Επίσης τα δεδομένα από τους αισθητήρες 46 και 49 διαπιστώθηκε ότι αποτελούνται από μία συνεχόμενη τιμή. Έτσι, τα συγκεκριμένα αφαιρούνται από το σετ των δεδομένων καταλήγοντας σε ένα υποσύνολο, που αποτελείται από τα υπόλοιπα 79 χαρακτηριστικά και είναι αυτά που προέρχονται από αισθητήρες ή στατιστικά μεγέθη αυτών. Πάνω σε αυτά θα πραγματοποιηθούν μια σειρά από επεξεργασίες, ενώ από τα περιγραφικά χαρακτηριστικά το διάνυσμα του status_id θα χρησιμοποιηθεί ως ground truth διάνυσμα καθώς χαρακτηρίζει την κατάσταση λειτουργίας σε κάθε σημείο καταγραφής (data entry).

Προκειμένου να μειωθούν οι ανάγκες σε υπολογιστική ισχύ, αλλά και να μπορέσει να δημιουργηθεί ένα μοντέλο πρόγνωσης σφαλμάτων το οποίο έχει τη δυνατότητα να ανιχνεύει μοτίβα στις μετρήσεις σε μεγαλύτερο χρονικό βάθος, πραγματοποιείται εκ νέου δειγματοληψία (resampling) στα δεδομένα με χρόνο δειγματοληψίας τις 3 ώρες (Jancauskas et al., 2023). Επιπλέον, η ίδια διαδικασία δειγματοληψίας ακολουθείται

και στο ground truth διάνυσμα, προκειμένου να υπάρχει αντιστοιχία μεταξύ των μετρήσεων και των ετικετών τους (labels).

Στη συνέχεια, γίνεται έλεγχος στο σετ δεδομένων για missing values, δηλαδή εγγραφές που δεν είχαν καμία τιμή (NaN). Επειδή η ασυνέχεια στα δεδομένα δεν είναι επιθυμητή, επιλέγεται να γεμίσουν οι τιμές αυτές αντί να απορριφθούν. Έτσι, εφαρμόζεται data imputation με mean strategy για να αντικατασταθούν οι τιμές που λείπουν με τις μέσες τιμές των αντίστοιχων πεδίων. Στα δεδομένα που προέρχονται από την wind farm A, παρατηρούνται 40 τέτοιες εγγραφές συνολικά, 20 εκ των οποίων είναι στον sensor_14_avg και 20 στον sensor_22_avg.

Τέλος, η διαδικασία προεπεξεργασίας ολοκληρώνεται με την κανονικοποίηση των δεδομένων. Προκειμένου να εξασφαλιστεί η ομαλή εκπαίδευση του μοντέλου και να αποφευχθεί η απότομη αλλαγή των παραγώγων, πραγματοποιείται κανονικοποίηση των δεδομένων σε τιμές μεταξύ 0 και 1. Με τον τρόπο αυτόν επιτυγχάνεται η γρηγορότερη και ομαλότερη σύγκλιση του μοντέλου μηχανικής μάθησης προς το επιθυμητό σημείο. Έχοντας ολοκληρώσει τη διαδικασία της ανάλυσης και επεξεργασίας των δεδομένων, ακολουθεί η επιλογή των χαρακτηριστικών για τη διαμόρφωση της κατάλληλης εισόδου στο μοντέλο μηχανικής μάθησης.

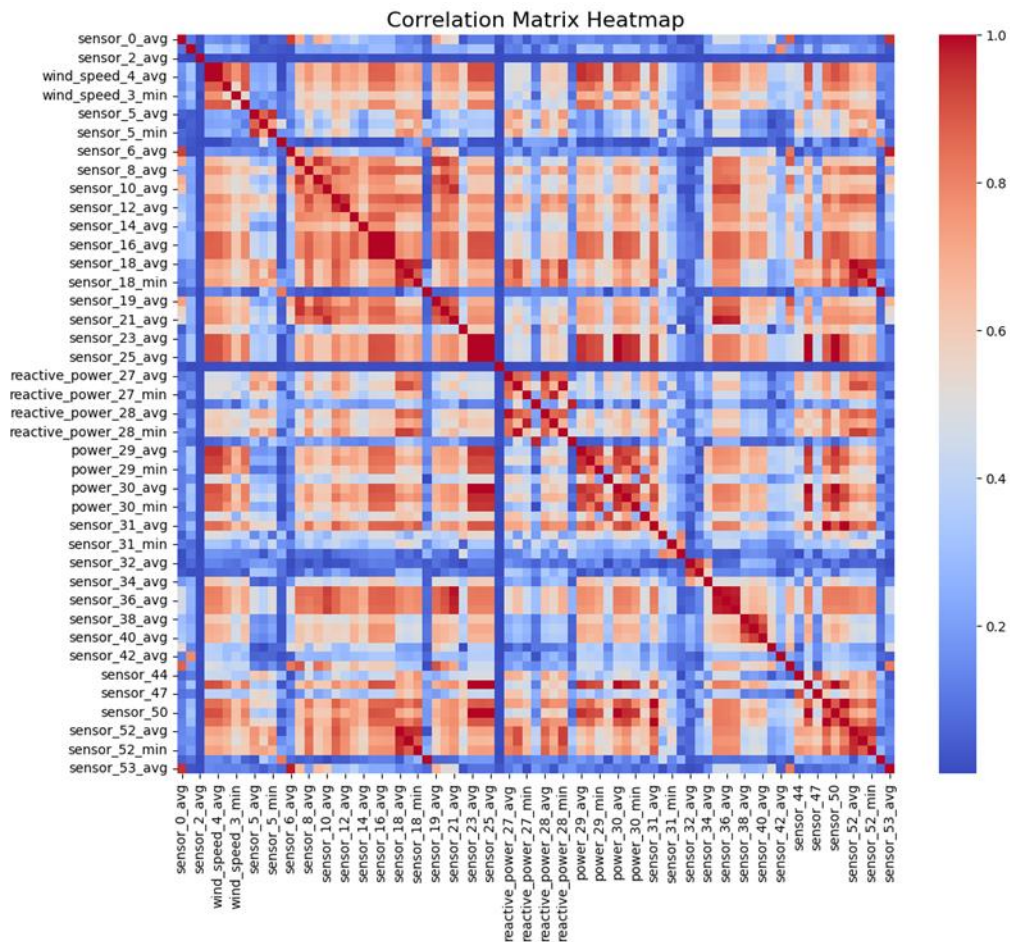
4.3 Επιλογή χαρακτηριστικών

Η επιλογή χαρακτηριστικών είναι μια τεχνική που ακολουθείται για την εκπαίδευση μοντέλων μηχανικής μάθησης και αφορά στον εντοπισμό και την επιλογή των πιο σχετικών χαρακτηριστικών που θα χρησιμοποιηθούν ως μεταβλητές εισόδου για το μοντέλο. Μέσω της διαδικασίας αυτής βελτιώνεται η απόδοση του μοντέλου, καθώς χαρακτηριστικά που έχουν περιττή ή μη σχετική πληροφορία αφαιρούνται, μειώνοντας την πιθανότητα για overfitting και αυξάνοντας την ικανότητα γενίκευσης του μοντέλου. Επιπλέον, μειώνεται η υπολογιστική πολυπλοκότητα του μοντέλου καθώς μειώνεται ο όγκος των δεδομένων που καλείται να επεξεργαστεί. Τέλος, το γεγονός ότι τα δεδομένα εισόδου μειώνονται, οδηγεί σε αυξημένη κατανόηση (interpretability) της λειτουργίας του μοντέλου, κάτι που είναι ιδιαίτερα σημαντικό.

Για την επίτευξη της διαδικασίας αυτής υπάρχει ένα πλήθος τεχνικών που μπορούν να ακολουθηθούν. Στην παρούσα διπλωματική επιλέχθηκαν δύο από τις κυριότερες μεθόδους και εφαρμόστηκαν σε μια ξεχωριστή σειρά πειραμάτων με σκοπό να γίνει σύγκριση των αποτελεσμάτων τόσο σε ό,τι αφορά την ακρίβεια του μοντέλου μηχανικής μάθησης, όσο και σχετικά με το μέγεθος των δεδομένων εισόδου σε κάθε περίπτωση.

Η πρώτη μέθοδος που αξιοποιήθηκε ήταν ο υπολογισμός του πίνακα συσχέτισης (correlation matrix), ο οποίος φαίνεται στην Εικόνα 12 για τα 79 διανύσματα που αποτελούν το σετ δεδομένων για την εκπαίδευση του μοντέλου LSTM. Στην περίπτωση αυτή γίνεται υπολογισμός της συσχέτισης μεταξύ όλων των ζευγών διανυσμάτων των χαρακτηριστικών που εμπεριέχονται στο σετ δεδομένων. Οι τιμές που λαμβάνει το μέγεθος της συσχέτισης είναι από -1 έως 1, με το 1 να σημαίνει ότι τα δύο διανύσματα έχουν μια θετική σχέση που τα συνδέει, το -1 να σημαίνει ότι έχουν μια αρνητική σχέση, ενώ το 0 σημαίνει ότι είναι ασυσχέτιστα μεταξύ τους. Στην Εικόνα 12 έχει υπολογιστεί ο correlation matrix, αλλά φαίνονται οι απόλυτες τιμές, καθώς αυτό που ελέγχεται είναι η ύπαρξη ή μη κάποιας συσχέτισης μεταξύ δύο διανυσμάτων που υποδεικνύει περιττή πληροφορία. Δεν έχει σημασία αν είναι θετική ή αρνητική η συσχέτιση. Τέλος, είναι απαραίτητο να τονιστεί ότι η συγκεκριμένη τεχνική χρησιμοποιείται για διανύσματα που εμφανίζουν γραμμικές σχέσεις μεταξύ τους. Σε περίπτωση μη γραμμικών συσχετίσεων ακολουθούνται διαφορετικές τεχνικές.

Ένα διάνυσμα συγκρινόμενο με τον εαυτό του έχει τέλεια συσχέτιση και έτσι προκύπτει παντού η τιμή της μονάδας στην κύρια διαγώνιο του πίνακα. Επίσης, ο συγκεκριμένος πίνακας είναι συμμετρικός. Έτσι, προχωρώντας στην διαδικασία της επιλογής χαρακτηριστικών, απομονώνεται το άνω τριγωνικό μέρος και γίνεται ο εξής έλεγχος. Γίνεται η επιλογή ενός κατωφλίου συσχέτισης ($corr_{threshold}$) και εντοπίζονται τα χαρακτηριστικά που τα διανύσματά τους έχουν συσχέτιση μεγαλύτερη από το κατώφλι αυτό. Στη συνέχεια, εκείνα αφαιρούνται από τα τελικά διανύσματα που θα αξιοποιηθούν για την εκπαίδευση του μοντέλου μηχανικής μάθησης. Για την επιλογή της τιμής αυτής πραγματοποιήθηκαν μια σειρά από πειράματα (finetuning) των οποίων τα αποτελέσματα θα παρουσιαστούν στο Κεφάλαιο 5.

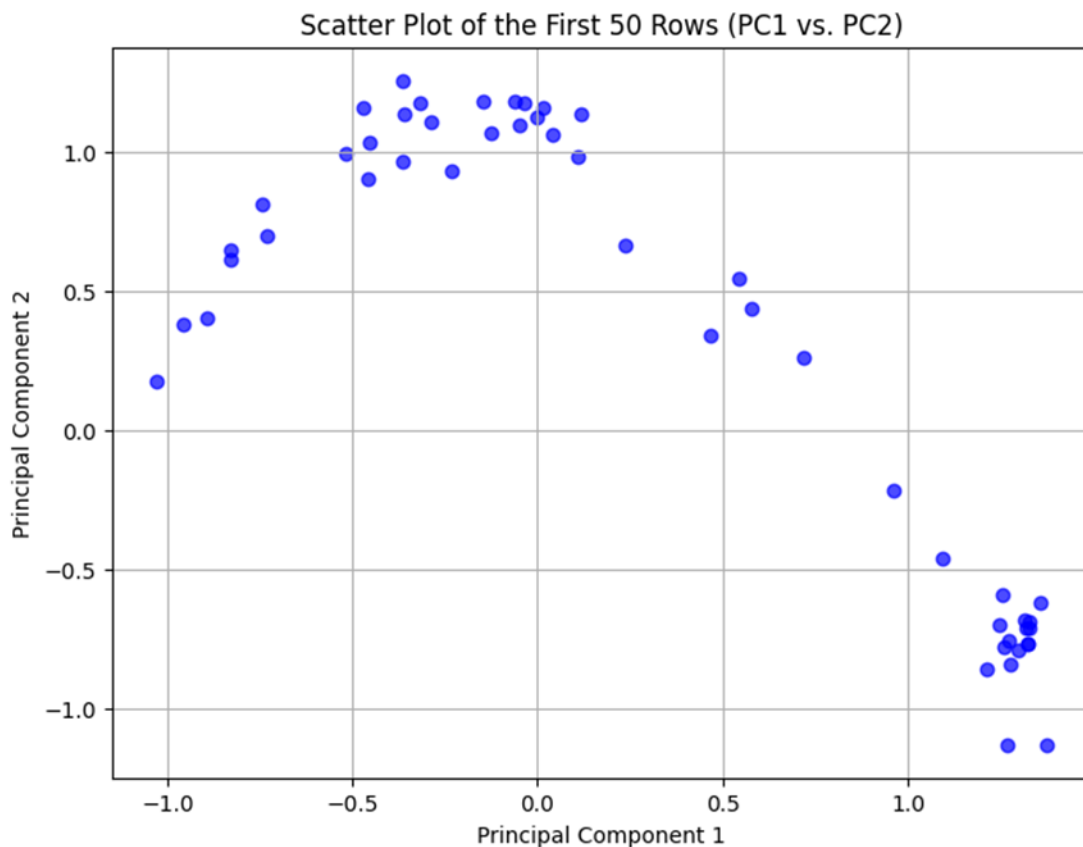


Εικόνα 12 Correlation matrix για τα 81 διανύσματα χαρακτηριστικών του σετ δεδομένων.

Η δεύτερη μέθοδος επιλογής χαρακτηριστικών που εφαρμόστηκε οδηγώντας σε μείωση της διαστασιμότητας των δεδομένων εισόδου του μοντέλου που θα εκπαιδευτεί είναι η μέθοδος Ανάλυσης Κύριων Συνιστωσών ή Principal Component Analysis (PCA). Σκοπός του PCA είναι η μείωση των χαρακτηριστικών εισόδου του νευρωνικού δικτύου, ανιχνεύοντας και επιλέγοντας τις κυριότερες συνιστώσες ανάμεσα στα χαρακτηριστικά. Αυτό επιτυγχάνεται με την μετατροπή των δεδομένων σε ένα καινούριο διανυσματικό χώρο στον οποίο τα νέα χαρακτηριστικά καλούνται κύριες συνιστώσες και είναι ασυσχέτιστα μεταξύ τους. Στον νέο αυτό χώρο, τα χαρακτηριστικά αντιπροσωπεύουν την κατεύθυνση των διανυσμάτων της μέγιστης διακύμανσης μεταξύ των αρχικών δεδομένων. Η μετατροπή των δεδομένων στον νέο χώρο γίνεται ως εξής. Πρώτα, πραγματοποιείται κανονικοποίηση των δεδομένων εισόδου και στην συνέχεια υπολογίζεται ο πίνακας συνδιακύμανσης (covariance) των κανονικοποιημένων δεδομένων. Με βάση τον νέο πίνακα, γίνεται υπολογισμός των

ιδιοτιμών και ιδιοδιανυσμάτων και επιλέγονται τα k ιδιοδιανύσματα με τις μεγαλύτερες ιδιοτιμές, με βάση τον καινούριο αριθμό χαρακτηριστικών που επιδιώκουμε για το σύστημά μας. Ο καινούργιος διανυσματικός χώρος προκύπτει από τον πολλαπλασιασμό των k ιδιοδιανυσμάτων με τον αρχικό πίνακα των δεδομένων, από τον οποίο προκύπτουν οι κύριες συνιστώσες.

Μια από τις βασικές προκλήσεις του αλγορίθμου PCA για την μείωση της διάστασης των δεδομένων και την επιλογή των επικρατέστερων συνιστωσών αποτελεί η κατάλληλη επιλογή του αριθμού των συνιστωσών. Η προσέγγιση που ακολουθήθηκε για την αντιμετώπιση αυτού του προβλήματος αφορά την υλοποίηση πολλαπλών πειραμάτων (finetuning) για την εύρεση της βέλτιστης τιμής των συνιστωσών. Τα αποτελέσματα των πειραμάτων θα παρουσιαστούν στο Κεφάλαιο 5.



Εικόνα 13 Scatter plot των δύο από τις δέκα κύριες συνιστώσες του dataset.

4.4 Επιλογή μοντέλου μηχανικής μάθησης – LSTM

Στην παρούσα διπλωματική, όπως έχει προηγουμένως αναφερθεί, αξιοποιείται το μοντέλο LSTM, ένα είδος ανατροφοδοτούμενου νευρωνικού δικτύου (RNN). Η επιλογή του μοντέλου οφείλεται στην ικανότητά του να λειτουργεί με δεδομένα ακολουθιών, όπως οι χρονοσειρές, αλλά και στην διαχείριση συσχετισμών, καθώς μπορεί να κρατάει πληροφορίες από το παρελθόν, προκειμένου να προβλέψει τιμές σε επόμενες χρονικές στιγμές.

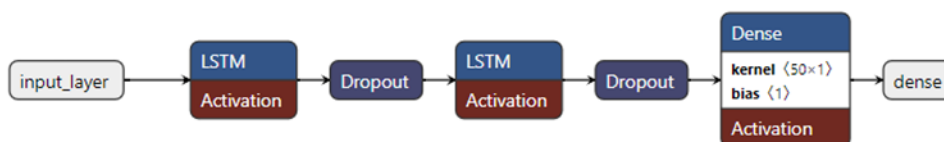
Για την εκπαίδευση ενός νευρωνικού δικτύου τύπου LSTM, είναι απαραίτητη η κατάλληλη προετοιμασία των δεδομένων. Πιο συγκεκριμένα, πραγματοποιήθηκε επιμερισμός των δεδομένων σε πακέτα, διότι πρέπει τα δεδομένα που εισάγονται στο μοντέλο να έχουν χρονική συνέχεια μεταξύ τους. Ωστόσο, τα πακέτα μεταξύ τους δεν είναι απαραίτητο να είναι συνεχόμενα χρονικά. Στην παρούσα υλοποίηση τα δίνουμε με τυχαία σειρά, αφού πρώτα τα έχουμε ανακατέψει (shuffle).

Το μέγεθος του πακέτου που επιλέγεται για τα δεδομένα εισόδου ενός δικτύου τύπου LSTM είναι ένα παραμετροποιήσιμο μέγεθος που μπορεί να επηρεάσει τα αποτελέσματα της εκπαίδευσης. Η φύση του προβλήματος παίζει μεγάλη σημασία στην επιλογή της χρονικής ανάλυσης των δεδομένων εισόδου. Ιδανικά, πρέπει το μέγεθος των πακέτων να είναι τέτοιο, ώστε να μπορεί να αποτυπώσει σημαντικά μοτίβα ή τάσεις που εμφανίζονται επαναλαμβανόμενα με την πάροδο του χρόνου στα δεδομένα.

Στην περίπτωση που έχουμε σχετικά απότομες αλλαγές στην διάρκεια μιας μέρας πρέπει τα πακέτα εισόδου να περιλαμβάνουν δεδομένα μιας ημέρας ή λιγότερα. Δηλαδή, για τα δεδομένα μετά την εκ νέου δειγματοληψία (ανά 3 ώρες) το μέγεθος των πακέτων πρέπει να είναι έως 8 δείγματα. Όμως, στα συγκεκριμένα δεδομένα παρατηρείται ότι είναι πολύ μικρό ένα τέτοιο δείγμα και δεν είναι αρκετό για να αποκαλύψει πιθανά μοτίβα. Έτσι, επιλέγεται ένα μεγαλύτερο μέγεθος κατά την υλοποίηση της διπλωματικής εργασίας.

Πιο συγκεκριμένα, το κάθε πακέτο περιλαμβάνει συνολικά 16 καταμετρήσεις από το σετ δεδομένων. Με αυτόν τον τρόπο δημιουργούνται τα δεδομένα εισόδου στο μοντέλο, τα οποία διαθέτουν συνολική χρονική διάρκεια 48 ωρών, δεδομένου ότι έχει γίνει εκ νέου δειγματοληψία των δεδομένων ανά 3 ώρες. Αυτή η διαδικασία, δίνει τη δυνατότητα στο μοντέλο LSTM να καταγράψει τα δεδομένα των τελευταίων δύο ημερών, προκειμένου να μπορέσει να ανιχνεύσει την κατάσταση της εκάστοτε ανεμογεννήτριας την επόμενη χρονική στιγμή, δηλαδή 3 ώρες μετά την τελευταία καταγραφή. Η επιλογή των 16 χρονικών στιγμών ως είσοδο στο μοντέλο έγινε με σκοπό να δοθεί το μεγαλύτερο χρονικό διάστημα, κρατώντας παράλληλα τις υπολογιστικές απαιτήσεις χαμηλές, καθώς όσο αυξάνεται το μέγεθος των πακέτων, τόσο αυξάνεται η πολυπλοκότητα του μοντέλου και κατ' επέκταση αυξάνονται οι υπολογιστικές απαιτήσεις και οι απαιτήσεις σε μνήμη. Μετά το διαχωρισμό των δεδομένων εισόδου σε πακέτα, ακολουθεί η ανάθεση στο καθένα από αυτά της αντίστοιχη ετικέτας (labeling). Προκειμένου να μπορεί το μοντέλο να προβλέπει πιθανά σφάλματα των ανεμογεννητριών στο μέλλον, ως ετικέτα για κάθε πακέτο δεδομένων επιλέχθηκε το `status_id` της ανεμογεννήτριας την αμέσως επόμενη χρονική στιγμή. Επομένως, κάθε πακέτο περιλαμβάνει συνολικά 16 σημεία από $t-48h$ έως $t0h$ και ως ετικέτα ορίζεται η κατάσταση της ανεμογεννήτριας τη χρονική στιγμή $t+3h$.

Έπειτα από τη δημιουργία των ζευγών δεδομένων και ετικετών, έγινε ο διαχωρισμός του σετ δεδομένων σε σετ εκπαίδευσης και επικύρωσης (train/val) με ποσοστό 80/20%, αντίστοιχα. Η αρχιτεκτονική του μοντέλου LSTM που χρησιμοποιήθηκε οπτικοποιείται στην Εικόνα 14, ενώ τα μεγέθη του κάθε επιπέδου καθώς και οι εκπαιδεύσιμες παράμετροι φαίνονται στην Εικόνα 15.



Εικόνα 14 Η αρχιτεκτονική του μοντέλου μηχανικής μάθησης που χρησιμοποιήθηκε.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 40, 64)	19,200
dropout (Dropout)	(None, 40, 64)	0
lstm_1 (LSTM)	(None, 50)	23,000
dropout_1 (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51

Total params: 42,251 (165.04 KB)
Trainable params: 42,251 (165.04 KB)
Non-trainable params: 0 (0.00 B)

Εικόνα 15 Τα μεγέθη των επιμέρους επιπέδων και οι παράμετροι του μοντέλου μηχανικής μάθησης που χρησιμοποιήθηκε.

Η εκπαίδευση γίνεται για 30 εποχές με batch size 8. Ως μετρική για τη σύγκριση και την αξιολόγηση των εκπαιδύσεων επιλέχθηκε η μετρική accuracy, ενώ η εκπαίδευση έγινε με συνάρτηση σφάλματος την binary crossentropy η οποία χρησιμοποιείται σε ανάλογες περιπτώσεις δυαδικής κατηγοριοποίησης, καθώς και με χρήση του αλγορίθμου Adam για τη βελτιστοποίηση του μοντέλου. Ο αλγόριθμος αυτός είναι πολύ αποδοτικός και έχει μικρές απαιτήσεις σε μνήμη. Οι διάφορες επιλογές που γίνονται για τις υπερπαραμέτρους, αξιολογούνται στο validation σετ.

Όπως είναι αναμενόμενο, και έχει ήδη αναφερθεί στην αρχή του Κεφαλαίου 4, υπάρχει ανισορροπία των δεδομένων σε περιπτώσεις πρόβλεψης σφαλμάτων. Αυτό οφείλεται στο γεγονός ότι κατά τη διάρκεια της καταγραφής της λειτουργίας των ανεμογεννητριών, η πλειοψηφία των δεδομένων αφορά σε κανονική λειτουργία των μηχανών, ενώ τα σφάλματα εμφανίζονται με πολύ μικρότερη συχνότητα. Προκειμένου να αντιμετωπιστεί αυτό το φαινόμενο και να μην αποκτήσει το μοντέλο προκατάληψη (bias) ως προς την επικρατούσα κλάση, χρησιμοποιήθηκαν βάρη, με σκοπό η κάθε κλάση να επηρεάζει διαφορετικά τη συνάρτηση κόστους. Τα βάρη που επιλέχθηκαν είναι τα εξής: 1 για την κλάση 0 (κανονική λειτουργία) και 4 για την κλάση 1 (σφάλμα στη λειτουργία), που προέκυψαν από την αναλογία των δεδομένων εκπαίδευσης.

Τέλος, κατά τη διάρκεια της εκπαίδευσης του νευρωνικού δικτύου παρατηρήθηκαν περιπτώσεις overfitting, γεγονός που σημαίνει ότι το μοντέλο προσαρμόζεται υπερβολικά στα δεδομένα εκπαίδευσης, με αποτέλεσμα να αποδίδει καλά σε αυτά,

αλλά να μην μπορεί να ανταποκριθεί σε νέα δεδομένα και να γενικεύσει. Πιθανές αιτίες αυτού του φαινομένου είναι η πολυπλοκότητα του μοντέλου, καθώς και τα λίγα δεδομένα εκπαίδευσης. Προκειμένου να αντιμετωπιστεί το overfitting, προστέθηκαν στο μοντέλο επίπεδα Dropout τα οποία απενεργοποιούν τυχαία με ποσοστό 40% νευρώνες του δικτύου κατά τη διάρκεια της εκπαίδευσης. Αυτή η διαδικασία είναι ιδιαίτερα σημαντική, καθώς αποτρέπει το μοντέλο από το να εξαρτάται υπερβολικά από συγκεκριμένους νευρώνες, ενώ παράλληλα το τελικό μοντέλο είναι μια ένωση πολλών μικρότερων εκπαιδευμένων μοντέλων. Επιπλέον, εφαρμόστηκε και η τεχνική του early stopping, με βάση την οποία η εκπαίδευση του μοντέλου σταματάει, μόλις το σφάλμα των δεδομένων επικύρωσης σταματά να μειώνεται. Στην προκειμένη περίπτωση, επιλέχθηκε η εκπαίδευση να σταματά έπειτα από 5 συνεχόμενες εποχές κατά τις οποίες το validation loss δεν έχει μειωθεί. Στην συνέχεια ακολουθεί το Κεφάλαιο 5 στο οποίο παρουσιάζονται τα αποτελέσματα των επιμέρους πειραμάτων, τα συμπεράσματα που προέκυψαν από αυτά καθώς και ορισμένες μελλοντικές προσεγγίσεις.

Κεφάλαιο 5. Αποτελέσματα – Συμπεράσματα – Μελλοντικές προσεγγίσεις

Στο παρόν κεφάλαιο παρουσιάζονται τα αποτελέσματα των πειραμάτων μαζί με τα αντίστοιχα συμπεράσματα, καθώς και ορισμένες μελλοντικές προσεγγίσεις που θα μπορούσαν να ακολουθηθούν.

5.1 Εκπαίδευση LSTM με correlation matrix

Όποια διανύσματα έχουν correlation πάνω από την τιμή του correlation threshold θεωρείται ότι έχουν overlapping πληροφορία και άρα αφαιρούμε το ένα από τα δύο διανύσματα. Συνολικά στην αρχή είχαμε 81 διανύσματα. Στον Πίνακα 2 φαίνεται η τιμή του κατωφλίου και για κάθε μια από αυτές τις τιμές φαίνεται η αντίστοιχη διάσταση των δεδομένων εισόδου που αξιοποιήθηκαν για την εκπαίδευση του LSTM δικτύου. Επίσης, φαίνεται η απόδοση του δικτύου για την περίπτωση baseline όπου δεν έχει εφαρμοστεί μείωση της διαστασιμότητας στα δεδομένα.

Το δίκτυο έχει την αρχιτεκτονική που έχει ήδη αναφερθεί στις Εικόνες 14 και 15, ενώ η εκπαίδευση είναι για 30 εποχές με early stopping για να αποφευχθεί το overfitting. Τα δεδομένα στα πειράματα αυτά έχουν γίνει resample σε διαστήματα τριών ωρών και το χρονικό παράθυρο που έχει επιλεγεί είναι μεγέθους 16 δειγμάτων, που αντιστοιχεί σε 48 ώρες.

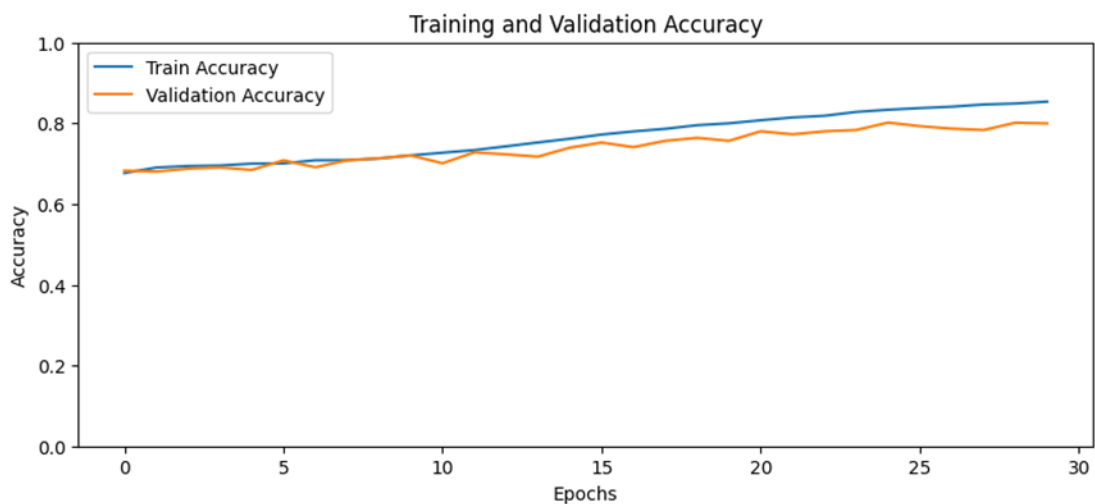
Αναφορικά με τα αποτελέσματα Precision, Recall και F1-score αφορούν στην κατηγορία 1 δηλαδή στην περίπτωση των σφαλμάτων, καθώς αυτό που καλείται να κάνει το σύστημα είναι να εντοπίζει εγκαίρως και με ακρίβεια σφάλματα στην λειτουργία των ανεμογεννητριών προτού εμφανιστούν. Όσο πιο κοντά στην μονάδα βρίσκονται οι τιμές αυτές τόσο το καλύτερο. Με έντονα γράμματα φαίνεται το πείραμα που είχε τα καλύτερα αποτελέσματα, δηλαδή αυτό με correlation threshold ίσο με 0.8.

Correlation threshold	Input data (dimension)	Precision	Recall	Accuracy	F1-score
-----------------------	------------------------	-----------	--------	----------	----------

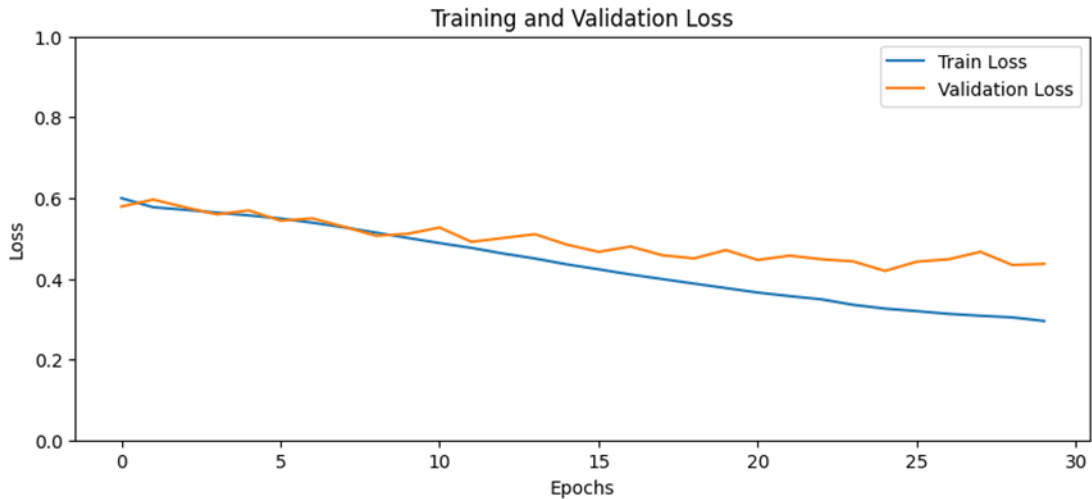
Baseline (3h resampling)	66,496 x 79	0.56	0.57	0.78	0.57
0.5	66,496 x 9	0.53	0.77	0.77	0.63
0.6	66,496 x 13	0.54	0.77	0.78	0.64
0.7	66,496 x 16	0.44	0.65	0.71	0.52
0.8	66,496 x 26	0.58	0.78	0.80	0.66
0.9	66,496 x 36	0.55	0.79	0.79	0.65

Πίνακας 2 Αποτελέσματα πειραμάτων (finetuning) για επιλογή κατάλληλης τιμής στο κατώφλι correlation.

Με βάση τον Πίνακα 2 φαίνεται ότι όσο το κατώφλι μικραίνει, τόσο χειρότερη είναι και η απόδοση του συστήματος, καθώς απορρίπτονται εν δυνάμει χρήσιμα δεδομένα για το μοντέλο. Έτσι, η κατάλληλη τιμή κατωφλίου είναι εκείνη του 0.8 όπου έχει λίγο καλύτερη απόδοση μάλιστα από το baseline, που σημαίνει ότι όντως υπάρχει περιττή πληροφορία στα αρχικά δεδομένα. Στις Εικόνες 16 και 17 φαίνονται οι καμπύλες της ακρίβειας και του loss του δικτύου τόσο για το training όσο και για το validation set, ενώ στην Εικόνα 18 φαίνεται ο confusion matrix.

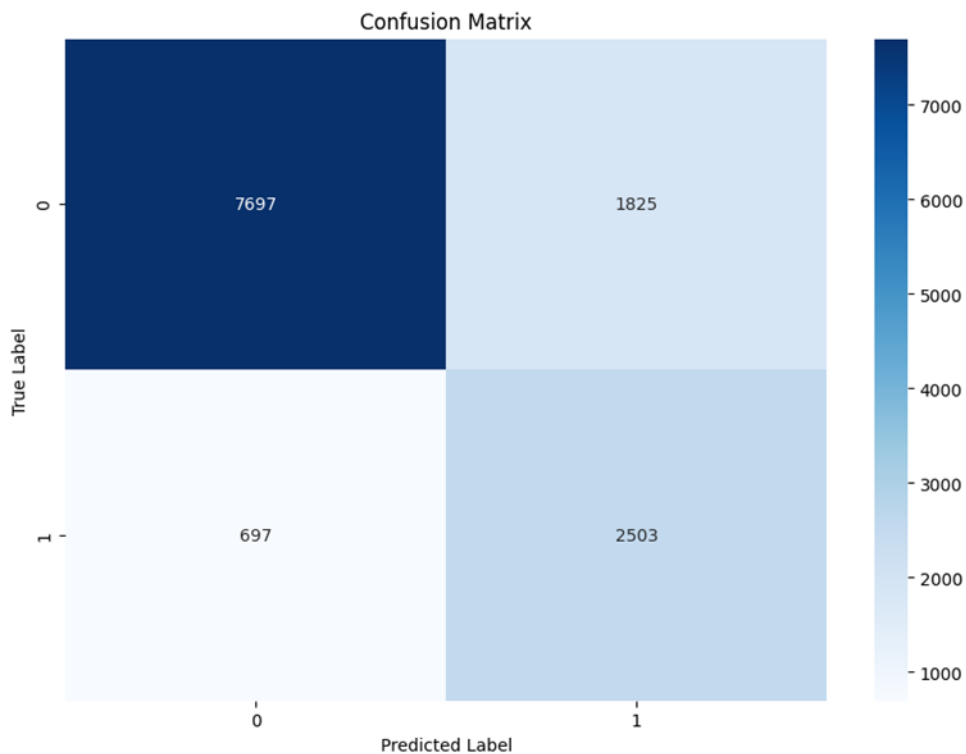


Εικόνα 16 Καμπύλες ακρίβειας ανά εποχές για training και validation σετ δεδομένων για correlation threshold 0.8.



Εικόνα 17 Καμπύλες loss ανά εποχές για training και validation σετ δεδομένων για correlation threshold 0.8.

Στις Εικόνες 16 και 17 βλέπουμε ότι υπάρχει ένα trend μείωσης του training loss. Συνεχίζοντας την εκπαίδευση για περισσότερες εποχές ίσως να μειωνόταν περεταίρω πριν γίνει overfit (θα σταματήσει λόγω early stopping). Όμως, επειδή τα πειράματα αυτά απαιτούσαν μεγάλη υπολογιστική ισχύ επιλέχθηκε για λόγους αποτελεσματικότητας και ταχύτητας η τιμή των 30 εποχών για όλα τα πειράματα.



Εικόνα 18 Confusion matrix για correlation threshold 0.8.

5.2 Εκπαίδευση LSTM με PCA

Για την ανάλυση των δεδομένων με την μέθοδο PCA πραγματοποιούμε μια σειρά πειραμάτων για την εύρεση του βέλτιστου αριθμού κύριων συνιστωσών. Για να το πετύχουμε αυτό, υλοποιούμε τον αλγόριθμο PCA για 4 διαφορετικές τιμές κύριων συνιστωσών και εκπαιδεύουμε 4 διαφορετικά μοντέλα LSTM, ώστε να ελεγχθεί σε ποια περίπτωση έχουμε καλύτερα αποτελέσματα. Στον Πίνακα 3 παρουσιάζονται συγκεντρωτικά τα αποτελέσματα των πειραμάτων ξεκινώντας από το αρχικό πείραμα όπου τα δεδομένα έχουν 81 χαρακτηριστικά (χωρίς μείωση της διαστασιμότητας) και στη συνέχεια με τα αποτελέσματα για PCA με 10, 20, 30 και 40 κύριες συνιστώσες αντίστοιχα. Η αρχιτεκτονική του δικτύου παραμένει ίδια όπως και στα προηγούμενα πειράματα. Με έντονα γράμματα φαίνεται το πείραμα που είχε τα καλύτερα αποτελέσματα, δηλαδή αυτό με κύριες συνιστώσες τις 40.

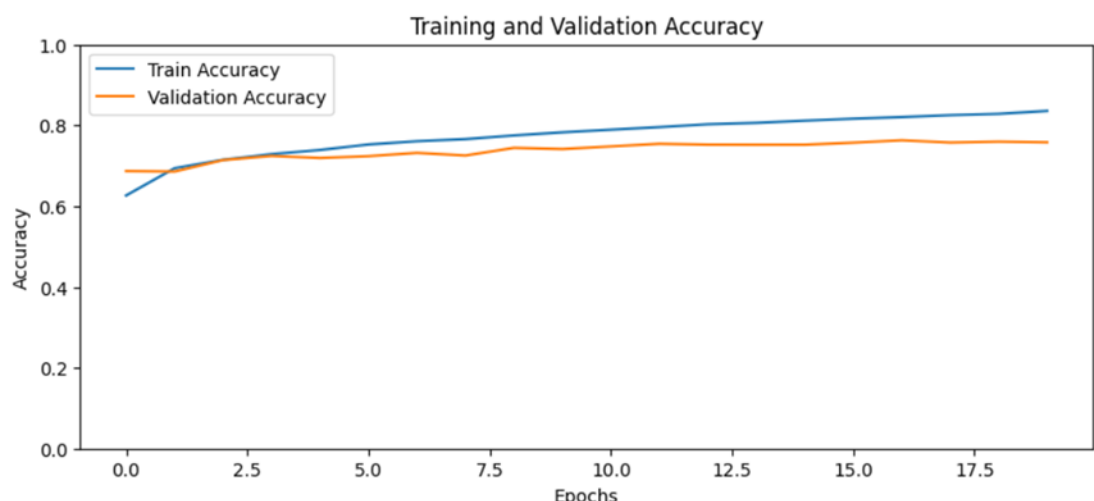
Number of components	Input data (dimension)	Precision	Recall	Accuracy	F1-score
Baseline (3h resampling)	66,496 x 79	0.56	0.57	0.78	0.57
10	66,496 x 10	0.54	0.73	0.78	0.61
20	66,496 x 20	0.53	0.77	0.77	0.62
30	66,496 x 30	0.52	0.78	0.76	0.60
40	66,496 x 40	0.57	0.69	0.79	0.63

Πίνακας 3 Αποτελέσματα πειραμάτων (finetuning) για επιλογή του κατάλληλου αριθμού κύριων συνιστωσών.

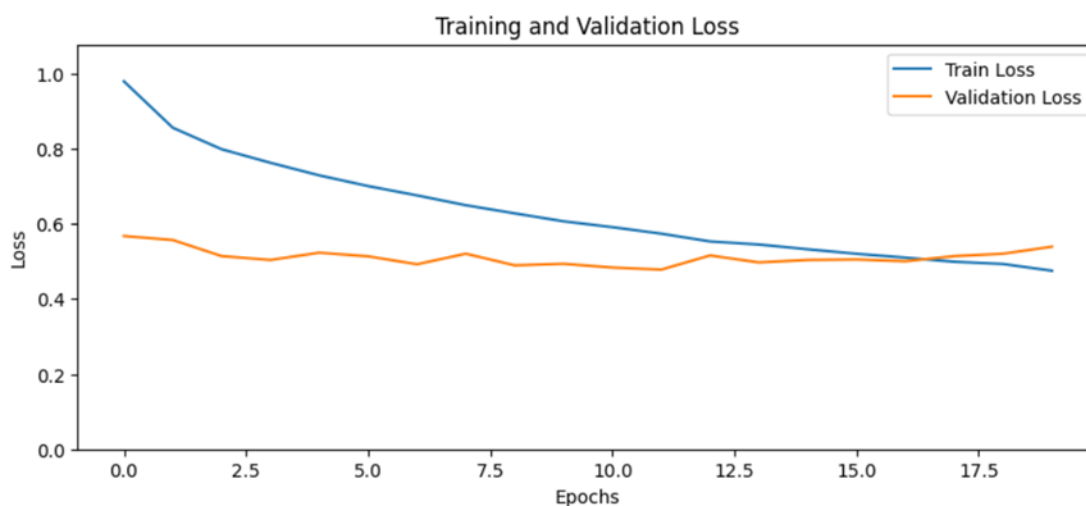
Με βάση τον Πίνακα 3 φαίνεται ότι η καλύτερη επίδοση του μοντέλου προκύπτει όταν στην είσοδο του δίνονται δεδομένα με τον μεγαλύτερο αριθμό συνιστωσών.

Αυτό οφείλεται στο γεγονός ότι όσο περισσότερες οι κύριες συνιστώσες, τόσο περισσότερη πληροφορία εμπεριέχεται στα δεδομένα εισόδου.

Στις Εικόνες 19 και 20 φαίνονται οι καμπύλες της ακρίβειας και του loss του δικτύου τόσο για το training όσο και για το validation set, ενώ στην Εικόνα 21 φαίνονται τα αποτελέσματα του confusion matrix.



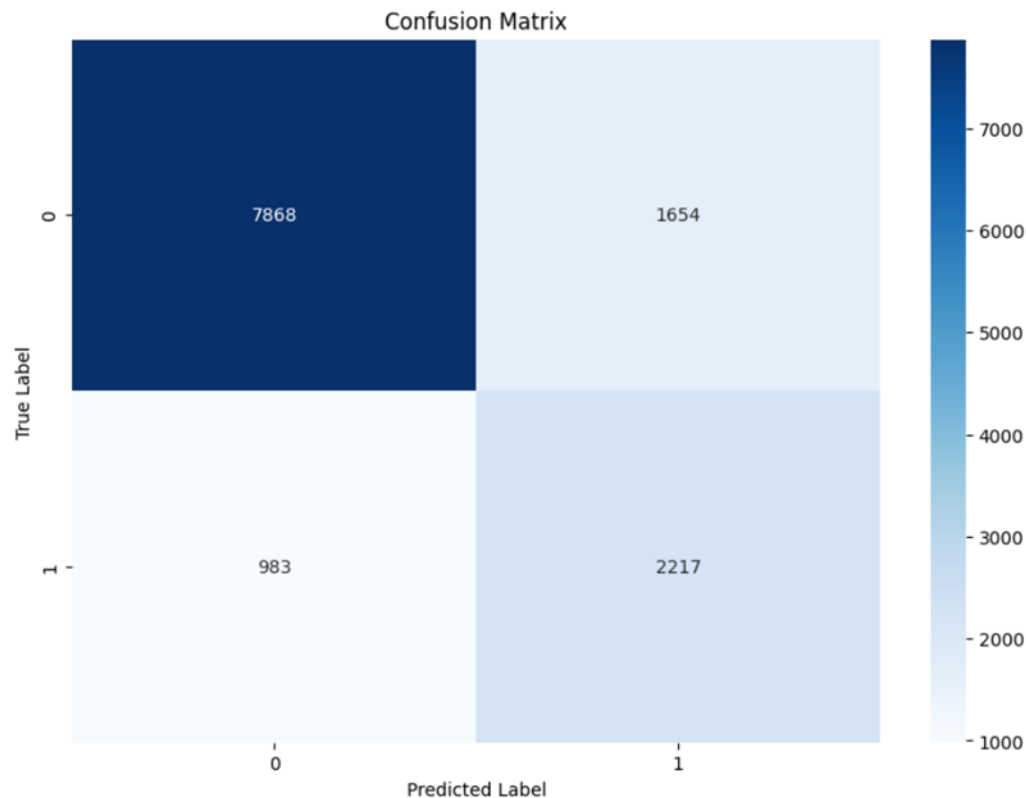
Εικόνα 19 Καμπύλες ακρίβειας ανά εποχές για training και validation σετ δεδομένων για 40 κύριες συνιστώσες.



Εικόνα 20 Καμπύλες loss ανά εποχές για training και validation σετ δεδομένων για 40 κύριες συνιστώσες

Στις Εικόνες 19 και 20 βλέπουμε ότι με χρήση του early stopping έγινε παύση της διαδικασίας του training στις 20 εποχές. Αυτό οφείλεται στο γεγονός ότι στο

validation loss παρατηρήθηκε τάση αύξησης της καμπύλης, πράγμα που σημαίνει πιθανή έναρξη overfitting στα δεδομένα εκπαίδευσης.



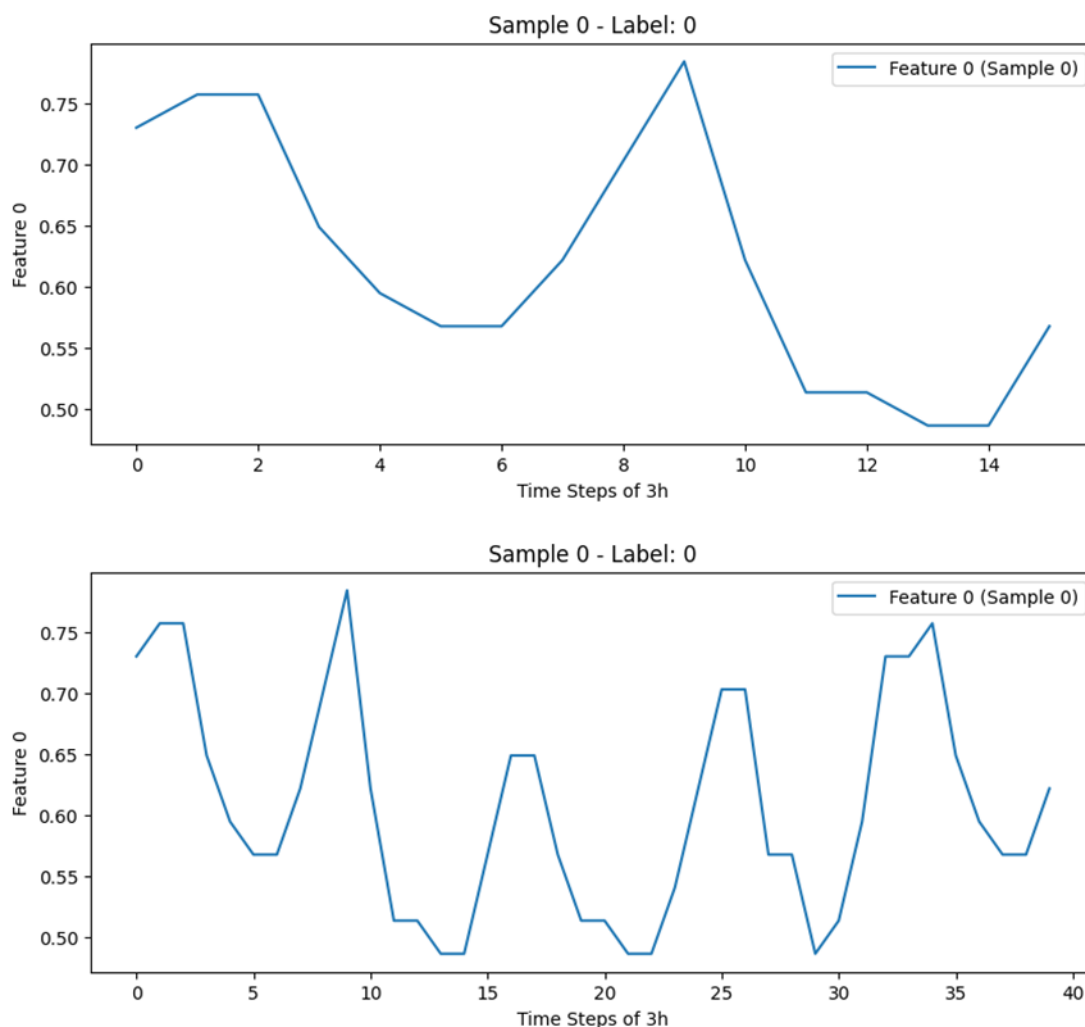
Εικόνα 21 Confusion matrix για 40 κύριες συνιστώσες.

5.3 Εκπαίδευση LSTM με αλλαγή στο time window και το resampling

5.3.1 Αλλαγή στο time window

Προκειμένου να δοκιμαστεί η ικανότητα του LSTM να ανιχνεύει μοτίβα στα δεδομένα του παρελθόντος, δοκιμάστηκε η αλλαγή του χρονικού παραθύρου με βάση το οποίο δημιουργούνται τα πακέτα δεδομένων. Πιο συγκεκριμένα στο παρόν πείραμα το χρονικό παράθυρο μεγάλωσε από 16 χρονικά σημεία, τα οποία αντιστοιχούν σε 2 ημέρες, σε 40 χρονικά σημεία, τα οποία αντιστοιχούν σε 5 ημέρες. Σκοπός του πειράματος είναι να ελεγχθεί η ικανότητα του μοντέλου να προβλέπει σφάλματα, ανάλογα με το χρονικό παράθυρο το οποίο λαμβάνει ως είσοδο.

Στην Εικόνα 22 γίνεται η σύγκριση των δεδομένων εισόδου του μοντέλου της περίπτωσης 1 με time window 16 τιμών και time window 40 τιμών για το ίδιο feature:



Εικόνα 22 Σύγκριση του ίδιου feature στην περίπτωση που έχει γίνει resampling για 3 ώρες και τα πακέτα έχουν time window μέγεθος 16 τιμών (πάνω) και 40 τιμών (κάτω).

Για το συγκεκριμένο πείραμα επιλέχθηκε να γίνει σύγκριση τριών διαφορετικών μοντέλων. Το πρώτο αφορά το base line μοντέλο στο οποίο δεν έχει γίνει μείωση της διαστασιμότητας των χαρακτηριστικών, το δεύτερο μοντέλο αφορά σε μείωση της διαστασιμότητας με χρήση του καλύτερου κατωφλίου συσχέτισης (threshold=0.8) και το τρίτο αφορά τη χρήση 40 κύριων συνιστωσών με PCA, όπως υπολογίστηκαν στο προηγούμενο κεφάλαιο. Τα συγκριτικά αποτελέσματα φαίνονται στον Πίνακα 4.

Case	Time window	Input data (dimension)	Precision	Recall	Accuracy	F1-score
Baseline (3h resampling)	16	66,496 x 79	0.56	0.57	0.78	0.57
Correlation threshold 0.8	16	66,496 x 26	0.58	0.78	0.80	0.66
PCA 40	16	66,496x40	0.57	0.69	0.79	0.63
Baseline (3h resampling)	40	66,496 x 81	0.62	0.64	0.81	0.63
Correlation threshold 0.8	40	66,496 x 26	0.58	0.72	0.80	0.64
PCA 40	40	66,496 x 40	0.63	0.61	0.81	0.62

Πίνακας 4 Αποτελέσματα πειραμάτων (finetuning) για επιλογή του κατάλληλου μεγέθους time window.

Από τις μετρήσεις στον πίνακα παρατηρούμε ότι η αλλαγή του χρονικού παραθύρου έχει επίπτωση στις μετρικές του μοντέλου. Πιο συγκεκριμένα, βλέπουμε ότι το baseline μοντέλο (χωρίς εξαγωγή χαρακτηριστικών), παρουσιάζει αρκετά βελτιωμένα αποτελέσματα σε σύγκριση με τα αντίστοιχα μοντέλα με χρονικό παράθυρο 16 χρονικών σημείων. Συνεπώς, η αύξηση του χρονικού παραθύρου, μπορεί να συμβάλλει στην ικανότητα του μοντέλου να βρίσκει μεγαλύτερη συσχέτιση μεταξύ των δεδομένων και να προβλέπει με μεγαλύτερη ικανότητα πιθανά σφάλματα. Είναι σημαντικό, όμως, να τονιστεί ότι η αύξηση του χρονικού παραθύρου δημιουργεί φυσικούς περιορισμούς, καθώς αυξάνεται η πολυπλοκότητα και η απαίτηση σε υπολογιστική μνήμη.

5.3.2 Αλλαγή στο sampling

Σε αυτό το κεφάλαιο γίνεται αλλαγή του sampling rate στα δεδομένα, ώστε να ελεγχθεί η επίδραση του συγκριτικά με τις άλλες μεθόδους. Πιο συγκεκριμένα, γίνεται resampling των δεδομένων ανά διαστήματα 6 ωρών και τα αποτελέσματα του

μοντέλου συγκρίνονται με την περίπτωση με sampling τριών ωρών. Κατά τη διάρκεια του πειράματος δοκιμάστηκε και η εκπαίδευση του μοντέλου με τα αρχικά δεδομένα, χωρίς resampling και με καταγραφές ανά δέκα λεπτά. Σε αυτήν την περίπτωση, όμως, προκειμένου να είναι άμεσα συγκρίσιμα τα δεδομένα χρειάζεται να έχουν το ίδιο συνολικό χρονικό παράθυρο. Αυτό έχει ως αποτέλεσμα να δημιουργούνται δεδομένα με διαστάσεις 1,144,714 x 79, γεγονός το οποίο απαιτεί πολύ μεγάλη υπολογιστική ισχύ και συνεπώς παραλήφθηκε. Έτσι, γίνεται σύγκριση μόνο μεταξύ resampling τριών και έξι ωρών.

Όπως αναφέρθηκε στο κεφάλαιο 5.3.1 με resampling τριών ωρών και χρονικό παράθυρο 2 ημερών χρειαζόμαστε 16 σημεία καταγραφής. Αντίστοιχα, για resampling έξι ωρών χρειαζόμαστε 8 σημεία για να δημιουργήσουμε χρονικό παράθυρο 2 ημερών και να είναι συγκρίσιμα. Η σύγκριση στον παρακάτω πίνακα αφορά το baseline μοντέλο (no dimensionality reduction), correlation threshold 0.8 και pca με 40 συνιστώσες.

Case	Time window	Input data (dimension)	Precision	Recall	Accuracy	F1-score
Baseline (original data)	288	1,144,714 x 79	-	-	-	-
Baseline (3h resampling)	16	66,496 x 81	0.56	0.57	0.78	0.57
Correlation threshold 0.8	16	66,496 x 26	0.58	0.78	0.80	0.66
PCA 40	16	66,496 x 40	0.57	0.69	0.79	0.63
Baseline (6h resampling)	8	33,253 x 81	0.48	0.60	0.74	0.53
Correlation threshold 0.8	8	33,253 x 26	0.51	0.54	0.75	0.52

PCA 40	8	33,253 x 40	0.57	0.52	0.78	0.54
Baseline (6h resampling)	16	33,253 x 81	0.59	0.49	0.78	0.54
Correlation threshold 0.8	16	33,253 x 26	0.53	0.55	0.76	0.54
PCA 40	16	33,253 x 40	0.57	0.51	0.78	0.54

Πίνακας 5 Αποτελέσματα πειραμάτων (finetuning) για επιλογή του κατάλληλου time resolution (resampling).

Το sampling ανά μεγαλύτερα χρονικά διαστήματα μειώνει την υπολογιστική πολυπλοκότητα και επιτρέπει τη δημιουργία μεγαλύτερων χρονικών παραθύρων και συνεπώς μεγαλύτερη πρόσβαση του μοντέλου σε δεδομένα του παρελθόντος. Γι' αυτό γίνονται και δοκιμές με 6h resampling και 16 data points για το time window, που αντιστοιχεί σε 4 ημέρες.

Από τα αποτελέσματα του πίνακα, παρατηρούμε ότι η αλλαγή του sampling rate σε 6 ώρες οδηγεί σε μικρότερες τιμές συγκριτικά με το sampling των 3 ωρών. Αυτό ενδεχομένως οφείλεται στη μείωση της πληροφορίας που δέχεται το μοντέλο. Συνεπώς, η αύξηση του sampling rate με σκοπό να δίνονται μεγαλύτερα time windows στο μοντέλο δεν αποδίδει ικανοποιητικά και είναι προτιμότερο να γίνεται μικρότερο resampling.

5.4 Συμπεράσματα και Μελλοντικές προσεγγίσεις

Συνοψίζοντας στην παρούσα διπλωματική δοκιμάστηκε η εκπαίδευση ενός μοντέλου βαθιάς μηχανικής μάθησης για την πρόγνωση των σφαλμάτων σε ανεμογεννήτριες. Για την επίτευξη του σκοπού χρησιμοποιήθηκαν δεδομένα SCADA με καταγραφές δεδομένων από διάφορους αισθητήρες πάνω σε ανεμογεννήτριες. Προκειμένου να βρεθεί ο βέλτιστος τρόπος εκπαίδευσης μοντέλων και πρόβλεψης σφαλμάτων έγιναν πειράματα σε τρεις βασικούς άξονες.

Ο πρώτος άξονας αφορά την επαναδειγματοληψία των δεδομένων. Αυτή η μέθοδος είναι απαραίτητη για τον εξής λόγο. Προκειμένου να μπορεί το μοντέλο να ανιχνεύει τάσεις στα δεδομένα και να γενικεύει, είναι σημαντικό να μπορεί το μοντέλο να εκπαιδευτεί με όσο το δυνατόν περισσότερες πληροφορίες από το παρελθόν. Αυτή η ανάγκη, όμως, αυξάνει την πολυπλοκότητα και την ανάγκη σε υπολογιστική ισχύ. Συνεπώς, είναι απαραίτητο να υπάρξει επαναδειγματοληψία των δεδομένων, χωρίς να χαθεί χρήσιμη πληροφορία. Από τα πειράματα προκύπτει ότι η επαναδειγματοληψία των 3 ωρών αυξάνει τις μετρικές των μοντέλων και αποδίδει καλύτερα από την επαναδειγματοληψία 6 ωρών, γεγονός που οφείλεται σε πληροφορία που χάνεται.

Ο δεύτερος άξονας αφορά την επιλογή των βέλτιστων χαρακτηριστικών που θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου. Για το σκοπό αυτό έγιναν δοκιμές με correlation matrix και Principal Component Analysis. Επιπλέον, καθένας από τους αλγορίθμους αυτούς χρειάζεται fine tuning σε παραμέτρους οι οποίες επηρεάζουν το αποτέλεσμα. Για τον αλγόριθμο correlation matrix έπειτα από πειράματα που πραγματοποιήθηκαν για διάφορες τιμές του , παρατηρήθηκε ότι το κατώφλι με τιμή 0.8 αποδίδει τα καλύτερα αποτελέσματα. Από την άλλη μεριά, ο αλγόριθμος PCA χρησιμοποιεί τον αριθμό k των κύριων συνιστωσών τις οποίες υπολογίζει. Για να βρεθεί ο βέλτιστος αριθμός συνιστωσών, πραγματοποιήθηκαν πειράματα με διαφορετικές τιμές. Από τα αποτελέσματα προκύπτει ότι το κατώφλι με τιμή 0.8 δίνει το βέλτιστο αποτέλεσμα.

Ο τρίτος άξονας αφορά την δημιουργία πακέτων δεδομένων με χρήση ενός χρονικού παραθύρου για την εκπαίδευση των μοντέλων. Αυτά τα χρονικά παράθυρα είναι υπεύθυνα για την επιλογή των αντίστοιχων καταγραφών που θα χρησιμοποιηθούν ως είσοδος για την πρόβλεψη που παράγει το μοντέλο. Το χρονικό παράθυρο καθορίζει πόσες καταγραφές και πόσο πίσω στο παρελθόν μπορεί να κοιτάξει το μοντέλο για να πραγματοποιήσει πρόβλεψη. Από τα πειράματα προκύπτει ότι η αύξηση του παραθύρου, η οποία επιτρέπει στο μοντέλο να εντοπίσει τάσεις στα δεδομένα σε μεγαλύτερο χρονικό διάστημα, οδηγεί σε αύξηση της ικανότητας του μοντέλου να προβλέπει σφάλματα. Φυσικά η αύξηση του χρονικού παραθύρου

ενέχει τον περιορισμό ότι δεν μπορεί να γίνει αυθαίρετα λόγω περιορισμών υπολογιστικής μνήμης.

Από τα συνολικά πειράματα, προκύπτει ότι η καλύτερη μέθοδος για την πρόβλεψη των σφαλμάτων είναι η επαναδειγματοληψία των δεδομένων ανά 3 ώρες, η χρήση κατωφλιού με τιμή 0.8 και η δημιουργία όσο το δυνατόν μεγαλύτερου χρονικού παραθύρου, όσο δηλαδή επιτρέπει το hardware που χρησιμοποιείται. Στην συγκεκριμένη διπλωματική, έγιναν δοκιμές μέχρι και του χρονικού παραθύρου 40 καταγραφών των 3 ωρών που αντιστοιχεί σε 5 ημέρες.

Σε μελλοντικές έρευνες, μπορούν να διερευνηθούν αρκετές μέθοδοι για την ενίσχυση της προληπτικής συντήρησης στις ανεμογεννήτριες. Πέρα από τη χρήση μοντέλων LSTM, θα ήταν χρήσιμο να διερευνηθεί η απόδοση εναλλακτικών μοντέλων μηχανικής μάθησης, όπως το Random Forest και οι αλγόριθμοι ομαδοποίησης, για την αξιολόγηση της αποτελεσματικότητάς τους στην πρόβλεψη αποτυχιών. Επιπλέον, ο συνδυασμός πολλαπλών τεχνικών για την επιλογή χαρακτηριστικών, θα μπορούσε να οδηγήσει σε πιο ισχυρή απόδοση του μοντέλου. Μια άλλη κατεύθυνση είναι η πιθανή χρήση μοντέλων όπως το Prophet του Facebook για την πρόβλεψη χρονοσειρών, τα οποία μπορούν να διερευνηθούν για την πρόβλεψη μελλοντικών τιμών αισθητήρων ή λειτουργικών τάσεων στις ανεμογεννήτριες. Αυτές οι μεθοδολογίες θα μπορούσαν να προσφέρουν καλύτερη γενίκευση, ταχύτερο υπολογισμό και βελτιωμένη ακρίβεια, συμβάλλοντας σε πιο αποτελεσματικά και αξιόπιστα συστήματα προληπτικής συντήρησης. Επιπλέον, μελλοντικές μελέτες θα μπορούσαν να διερευνήσουν την επεκτασιμότητα του αναπτυγμένου μοντέλου σε διαφορετικούς τύπους ανεμογεννητριών και περιβαλλοντικών συνθηκών, καθώς επίσης και την δοκιμή μεγαλύτερων χρονικών παραθύρων. Με την ενσωμάτωση αυτών των προσεγγίσεων, η προσαρμοστικότητα του μοντέλου σε μεταβαλλόμενες συνθήκες λειτουργίας θα μπορούσε να βελτιωθεί σημαντικά.

Βιβλιογραφικές Αναφορές

1. Alipour, F., & Rahimpour, M. R. (2024). Components of Wind Turbines (Rotors, Blades, Drive Trains, Gearboxes, Generators, etc.). In *Encyclopedia of Renewable Energy, Sustainability and the Environment* (pp. 343–360). Elsevier. <https://doi.org/10.1016/B978-0-323-93940-9.00228-0>
2. European Commission. (n.d.). https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.
3. Golestani, N., Arzaghi, E., Abbassi, R., & Garaniya, V. (2024). Modelling error chains in offshore wind energy systems: Examining the interplay of human performance and machine state. *Ocean Engineering*, 308, 118157. <https://doi.org/10.1016/j.OCEANENG.2024.118157>
4. Hansen, H. H. (2024). *Predictive maintenance for power plants*. Technical University of Denmark.
5. Hassan, Q., Viktor, P., J. Al-Musawi, T., Mahmood Ali, B., Algburi, S., Alzoubi, H. M., Khudhair Al-Jiboory, A., Zuhair Sameen, A., Salman, H. M., & Jaszczur, M. (2024). The renewable energy role in the global energy Transformations. *Renewable Energy Focus*, 48, 100545. <https://doi.org/10.1016/j.ref.2024.100545>
6. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
7. Jankauskas, M., Serackis, A., Šapurov, M., Pomarnacki, R., Baskys, A., Hyunh, V. K., Vaimann, T., & Zakis, J. (2023). Exploring the Limits of Early Predictive Maintenance in Wind Turbines Applying an Anomaly Detection Technique. *Sensors* 2023, Vol. 23, Page 5695, 23(12), 5695. <https://doi.org/10.3390/S23125695>
8. Liu, X., Zhang, Y., Zhang, Y., & Deng, C. (2023). Analysis of SCADA Data Preprocessing Methods for Wind Power Farms. *2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2)*, 3578–3583. <https://doi.org/10.1109/EI259745.2023.10512521>

9. Long, Y., Chen, Y., Xu, C., Li, Z., Liu, Y., & Wang, H. (2023). The role of global installed wind energy in mitigating CO2 emission and temperature rising. *Journal of Cleaner Production*, 423, 138778.
<https://doi.org/10.1016/j.jclepro.2023.138778>
10. Ma, W., & Mashayekhy, L. (2020). Truthful Computation Offloading Mechanisms for Edge Computing. *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, 199–206. <https://doi.org/10.1109/CSCloud-EdgeCom49738.2020.00043>
11. Manna, D., Mitici, M., & Dalla Vedova, M. D. L. (2024). System-level Probabilistic Remaining Useful Life Prognostics and Predictive Inspection Planning for Wind Turbines. *PHM Society European Conference*, 8(1), 13.
<https://doi.org/10.36001/phme.2024.v8i1.3991>
12. Moros, D., Berrabah, N., Searle, K. D., & Ashton, I. G. (2024). Maintenance & failure data analysis of an offshore wind farm. *Journal of Physics: Conference Series*, 2767(6), 062006. <https://doi.org/10.1088/1742-6596/2767/6/062006>
13. Nunes, P., Santos, J., & Rocha, E. (2023). Challenges in predictive maintenance – A review. *CIRP Journal of Manufacturing Science and Technology*, 40, 53–67. <https://doi.org/10.1016/J.CIRPJ.2022.11.004>
14. Nuvvula, R. S. S., Kumar, P. P., S, T., Ahammed, S. R., R, H., J M, B., & Ali, A. (2024). Machine Learning-Driven Predictive Maintenance Framework for Anomaly Detection and Prognostics in Wind Farm Operations. *2024 12th International Conference on Smart Grid (IcSmartGrid)*, 284–289.
<https://doi.org/10.1109/icSmartGrid61824.2024.10578083>
15. Peter Poór, D. Ž. J. B. (2019). *Historical Overview of Maintenance Management Strategies: Development from Breakdown Maintenance to Predictive Maintenance in Accordance with Four Industrial Revolutions*.
16. Rajaoarisoa, L., Randrianandraina, R., & Sayed-Mouchaweh, M. (2024). Predictive maintenance model-based on multi-stage neural network systems for wind turbines. *2024 International Conference on Artificial*

- Intelligence, Computer, Data Sciences and Applications (ACDSA)*, 1–7.
<https://doi.org/10.1109/ACDSA59508.2024.10467452>
17. Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J. E., & Jiang, Z. (2021). Offshore wind turbine operations and maintenance: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 144, 110886.
<https://doi.org/10.1016/j.rser.2021.110886>
 18. Shafiee, M. (2024). Extending the Lifetime of Offshore Wind Turbines: Challenges and Opportunities. *Energies*, 17(16), 4191.
<https://doi.org/10.3390/en17164191>
 19. Velandia-Cardenas, C., Vidal, Y., & Pozo, F. (2021). Wind Turbine Fault Detection Using Highly Imbalanced Real SCADA Data. *Energies*, 14(6), 1728. <https://doi.org/10.3390/en14061728>
 20. *wind turbine Facts | Britannica*. (n.d.). Retrieved September 29, 2024, from <https://www.britannica.com/facts/wind-turbine>
 21. A typical artificial neural network (ANN) | Download Scientific Diagram. (n.d.). Retrieved October 1, 2024, from https://www.researchgate.net/figure/A-typical-artificial-neural-network-ANN_fig1_305325563
 22. Canizo, M., Onieva, E., Conde, A., Charramendieta, S., & Trujillo, S. (2017). Real-time predictive maintenance for wind turbines using Big Data frameworks. 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), 70–77.
<https://doi.org/10.1109/ICPHM.2017.7998308>
 23. Chen, H., Liu, H., Chu, X., Liu, Q., & Xue, D. (2021). Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network. *Renewable Energy*, 172, 829–840.
<https://doi.org/10.1016/j.renene.2021.03.078>
 24. Emin Elmar. (2019). Predictive Maintenance of Wind Generators based on AI Technique. University of Waterloo.
 25. Evolution of maintenance | Download Scientific Diagram. (n.d.). Retrieved October 1, 2024, from https://www.researchgate.net/figure/Evolution-of-maintenance_fig2_373874317

26. J. Brownlee. (2019). "What is the difference between a batch and an epoch in a neural network?" .
27. ML | Underfitting and Overfitting - GeeksforGeeks. (n.d.). Retrieved October 1, 2024, from <https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/>
28. Nitish Srivastava, G. H. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*.
29. Qian, P., Tian, X., Kanfoud, J., Lee, J., & Gan, T.-H. (2019). A Novel Condition Monitoring Method of Wind Turbines Based on Long Short-Term Memory Neural Network. *Energies*, 12(18), 3411. <https://doi.org/10.3390/en12183411>
30. Udo, W., & Muhammad, Y. (2021). Data-Driven Predictive Maintenance of Wind Turbine Based on SCADA Data. *IEEE Access*, 9, 162370–162388. <https://doi.org/10.1109/ACCESS.2021.3132684>
31. Wang, X., Zheng, Z., Jiang, G., He, Q., & Xie, P. (2022). Detecting Wind Turbine Blade Icing with a Multiscale Long Short-Term Memory Network. *Energies*, 15(8), 2864. <https://doi.org/10.3390/en15082864>
32. Wang, Y., Ma, X., & Qian, P. (2018). Wind Turbine Fault Detection and Identification Through PCA-Based Optimal Variable Selection. *IEEE Transactions on Sustainable Energy*, 9(4), 1627–1635. <https://doi.org/10.1109/TSTE.2018.2801625>
33. Wind Turbine SCADA Data For Early Fault Detection (April 17, 2024).
34. Yoav Freund, R. E. S. (1996). Experiments with a New Boosting Algorithm.