

Διεθνές Πανεπιστήμιο της Ελλάδος

Τμήμα Μηχανικών Πληροφορικής, Υπολογιστών και Τηλεπικοινωνιών

Διπλωματική Εργασία

Πρόγραμμα Μεταπτυχιακών Σπουδών στη Ρομποτική

# Ανάπτυξη Τεχνικών Προβλεπτικής Συντήρησης σε Κινητήρες Βιομηχανικού Τύπου

Αλεξανδρής Νικόλαος

Διεθνές Πανεπιστήμιο της Ελλάδος

Θεσσαλονίκη, Οκτώβριος 2023

Εργασία που υποβλήθηκε στο Πρόγραμμα Μεταπτυχιακών Σπουδών στη Ρομποτική, του Διεθνούς Πανεπιστημίου της Ελλάδος, για τη μερική εκπλήρωση υποχρεώσεων για το Δίπλωμα Ειδίκευσης στη Ρομποτική.

Επιβλέπων: Σταύρος Βολογιαννίδης, Επίκουρος Καθηγητής

## Περίληψη

Η παρούσα εργασία αφορά την αξιολόγηση και σύγκριση διαφόρων μοντέλων μηχανικής μάθησης για εφαρμογές προβλεπτικής συντήρησης σε διάφορα σύνολα δεδομένων. Τα μοντέλα αξιολογήθηκαν χρησιμοποιώντας μετρικές απόδοσης, όπως ο μέσος τετραγωνικός σφάλμα (MSE), ο μέσος απόλυτος σφάλμα (MAE) και ο συντελεστής προσδιορισμού (R2).

Στο σύνολο δεδομένων του κινητήρα, το μοντέλο GRU ξεχώρισε με τα χαμηλότερα MSE και MAE. Στο σύνολο δεδομένων του δεξιού ρουλεμάν, τα μοντέλα LSTM και GRU επέδειξαν εξαιρετική ακρίβεια. Στο σύνολο δεδομένων του αριστερού ρουλεμάν, το LSTM και το GRU παρέδωσαν τα χαμηλότερα σφάλματα πρόβλεψης.

Ειδικότερα, στο ολοκληρωμένο σύνολο δεδομένων, το μοντέλο CNN ξεχώρισε με χαμηλό MSE και τον υψηλότερο R2. Επίσης, το μοντέλο GRU εμφάνισε καλή απόδοση με χαμηλό MSE και MAE.

Η επιλογή του καταλληλότερου μοντέλου εξαρτάται από τις απαιτήσεις κάθε εφαρμογής. Το μοντέλο GRU είναι κατάλληλο για ακριβείς προβλέψεις, ενώ το μοντέλο CNN είναι αποτελεσματικό στην καταγραφή προτύπων. Τελικά, το μοντέλο FNN είναι ευέλικτο και αξιόπιστο σε συγκεκριμένα σύνολα δεδομένων.

Συνοψίζοντας, αυτή η μελέτη παρέχει έναν συστηματικό πλαίσιο για την επιλογή του πιο κατάλληλου μοντέλου μηχανικής μάθησης για εφαρμογές προβλεπτικής συντήρησης, λαμβάνοντας υπόψη τις συγκεκριμένες απαιτήσεις και περιορισμούς κάθε εφαρμογής.

**Λέξεις-κλειδιά:** Προβλεπτική Συντήρηση, Μηχανική Μάθηση, Αξιολόγηση, Σύγκριση, Μετρικές Απόδοσης, MSE, MAE, R2, LSTM, GRU, CNN, FNN.

## **Abstract**

This research involves the evaluation and comparison of various machine learning models for predictive maintenance applications across different datasets. The models were assessed using performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R<sup>2</sup>).

In the motor dataset, the GRU model stood out with the lowest MSE and MAE. In the right bearing dataset, both LSTM and GRU models demonstrated exceptional accuracy. In the left bearing dataset, LSTM and GRU delivered the lowest prediction errors.

Specifically, in the integrated dataset, the CNN model excelled with low MSE and the highest R<sup>2</sup>. The GRU model also exhibited good performance with low MSE and MAE.

The choice of the most suitable model depends on the requirements of each application. The GRU model is suitable for precise predictions, while the CNN model is effective at capturing patterns. Ultimately, the FNN model is flexible and robust on specific datasets.

In summary, this study provides a comprehensive and systematic framework for selecting the most appropriate machine learning model for predictive maintenance applications, considering the specific requirements and constraints of the application.

**Keywords:** Predictive Maintenance, Machine Learning, Evaluation, Comparison, Performance Metrics, MSE, MAE, R-squared, LSTM, GRU, CNN, FNN.

## Πίνακας Περιεχομένων

Περίληψη .....	2
Abstract.....	3
Εισαγωγή .....	9
Στόχοι της Εργασίας .....	10
Διάρθρωση της Εργασίας.....	10
Θεωρητικό Υπόβαθρο.....	12
Εισαγωγή στην προληπτική συντήρηση .....	12
Χαρακτηριστικά υγείας εξοπλισμού .....	15
Εξαγωγή Χαρακτηριστικών στο πεδίο του Χρόνου.....	16
Εφαρμογή του προτύπου ISO .....	16
Παράμετροι Μέτρησης.....	17
Εξαγωγή συμπερασμάτων προληπτικής συντήρησης.....	21
Προσέγγιση Μηχανικής Μάθησης.....	21
Τεχνικές Μηχανικής Μάθησης .....	23
Τεχνικές Βαθείας Μάθησης .....	27
Μεθοδολογία.....	34
Σύνολο Δεδομένων .....	34
Ρύθμιση του περιβάλλοντος .....	34
Εκπαίδευση Μοντέλων Βαθείας Μάθησης .....	35
Τεχνικό Μέρος.....	38
Προεπεξεργασία δεδομένων για κάθε σύνολο δεδομένων .....	38
Εξερεύνηση δεδομένων για κάθε σύνολο δεδομένων .....	39
Μηχανική Χαρακτηριστικών για Κάθε Σύνολο Δεδομένων .....	44
Ενοποίηση δεδομένων για βελτιωμένη ανάλυση .....	45

Εφαρμογή της μηχανική χαρακτηριστικών, εξερεύνηση και προεπεξεργασία των δεδομένων των ενισχυμένων συνόλων και του ενσωματωμένου συνόλου δεδομένων.....	46
Διαχωρισμός δεδομένων για αξιολόγηση μοντέλου.....	54
Βελτίωση δεδομένων για μοντελοποίηση.....	55
Εκπαίδευση Μοντέλων Βαθιάς Μάθησης .....	56
Μοντέλο LSTM.....	56
Μοντέλο GRU.....	56
Μοντέλο CNN.....	57
Μοντέλο FNN .....	57
Αξιολόγηση και σύγκριση μοντέλων.....	58
Αποτελέσματα μεμονομένων μοντέλων ολοκληρωμένου συνόλου δεδομένων .....	58
Δεδομένα μοτέρ .....	58
Δεδομένα δεξιού ρουλεμάν.....	62
Δεδομένα αριστερού ρουλεμάν.....	66
Ενσωματωμένο σύνολο δεδομένων .....	70
Συνολική αξιολόγηση μοντέλων .....	74
Επιλογή μοντέλου.....	75
Συζήτηση- Συμπεράσματα .....	77
Επιλογή μοντέλου και απόδοση .....	77
Πρακτικές επιπτώσεις .....	78
Περιορισμοί και Μελλοντικές Κατευθύνσεις.....	79
Βιβλιογραφία.....	80

## Πίνακας Εικόνων

Εικόνα 1: Βασική διαδικασία Προβλεπτικής Συντήρησης.....	14
Εικόνα 2: Τύπος Συντήρησης.....	14
Εικόνα 3: Μαθηματική έκφραση μεγεθών δόνησης .....	18
Εικόνα 4: Σχέση μεταξύ μετατόπισης, ταχύτητας και επιτάχυνσης στο φάσμα συχνοτήτων ...	19
Εικόνα 5: Όρια σοβαρότητας κραδασμών σύμφωνα με το ISO 20816.....	21
Εικόνα 6: Απεικόνιση των δύο προσεγγίσεων μηχανικής μάθησης.....	22
Εικόνα 7: Γραμμική Παλινδρόμηση στην Μηχανική Μάθηση .....	24
Εικόνα 8: Αλγόριθμος SVM .....	25
Εικόνα 9: Αλγόριθμος kNN .....	27
Εικόνα 10: Μπλόκ διάγραμμα ανάπτυξης αλγορίθμου ANN .....	29
Εικόνα 11: Αλλαγές στην ταχύτητα και τη θερμοκρασία των κραδασμών RMS (Κινητήρας)	39
Εικόνα 12: Αλλαγές στην ταχύτητα και τη θερμοκρασία κραδασμών RMS (δεξί ρουλεμάν).	40
Εικόνα 13: Αλλαγές στην ταχύτητα και τη θερμοκρασία κραδασμών RMS (αριστερό ρουλεμάν).....	40
Εικόνα 14: Σχεδίαση box plot της ταχύτητας δόνησης RMS και της θερμοκρασίας του μοτέρ. ....	41
Εικόνα 15: Σχεδίαση box plot της ταχύτητας δόνησης RMS και της θερμοκρασίας του δεξιού ρουλεμάν.....	41
Εικόνα 16: Σχεδίαση box plot της ταχύτητας δόνησης RMS και της θερμοκρασίας του αριστερού ρουλεμάν .....	42
Εικόνα 17: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα κινητήρα. ....	59
Εικόνα 18: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα κινητήρα. ....	60
Εικόνα 19: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα κινητήρα. ....	61
Εικόνα 20: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα κινητήρα.....	62
Εικόνα 21: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα δεξιού ρουλεμάν.....	63

Εικόνα 22: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα δεξιού ρουλεμάν.....	64
Εικόνα 23: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα δεξιού ρουλεμάν.....	65
Εικόνα 24: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα δεξιού ρουλεμάν. ....	66
Εικόνα 25: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα αριστερού ρουλεμάν.....	67
Εικόνα 26: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα αριστερού ρουλεμάν.....	68
Εικόνα 27: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα αριστερού ρουλεμάν.....	69
Εικόνα 28: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα αριστερού ρουλεμάν. ....	70
Εικόνα 29: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα του ενσωματωμένου συνόλου.....	71
Εικόνα 30: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα του ενσωματωμένου συνόλου.....	72
Εικόνα 31: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα του ενσωματωμένου συνόλου.....	73
Εικόνα 32: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα του ενσωματωμένου συνόλου. ....	74

## Πίνακας Πινάκων

Πίνακας 1: Βασικά στατιστικά στοιχεία ταχύτητας RMS δόνησης και θερμοκρασίας του μοτέρ, του δεξιού και του αριστερού ρουλεμάν. ....	42
Πίνακας 2: Αριθμός τιμών που λείπουν σε κάθε μεταβλητή των συνόλων δεδομένων που αφορούν το μοτερ, το δεξί και το αριστερό ρουλεμάν. ....	47
Πίνακας 3: Βασικά στατιστικά στοιχεία των μεταβλητών του μοτέρ, του δεξιού και του αριστερού ρουλεμάν, μετά την εφαρμογή της μηχανικής χαρακτηριστικών και τον καθαρισμό των δεδομένων. ....	48
Πίνακας 4: Βασικά στατιστικά στοιχεία των μεταβλητών του ενσωματωμένου συνόλου δεδομένων, μετά την εφαρμογή της μηχανικής χαρακτηριστικών και τον καθαρισμό των δεδομένων. ....	51
Πίνακας 5: Δεδομένα μοτέρ.....	59
Πίνακας 6: Δεδομένα δεξιού ρουλεμάν.....	62
Πίνακας 7: Ενσωματωμένο σύνολο δεδομένων.....	70
Πίνακας 8: Συνολική αξιολόγηση μοντέλων. ....	74



# Εισαγωγή

Η τεχνητή νοημοσύνη και η μηχανική μάθηση αναδεικνύονται διαρκώς ως κορυφαίοι παίκτες στον κόσμο της τεχνολογίας, επηρεάζοντας σημαντικά πολλούς τομείς, συμπεριλαμβανομένου του βιομηχανικού τομέα [1]. Σε αυτό το πλαίσιο, η προβλεπτική συντήρηση αναδύεται ως μία από τις σημαντικότερες εφαρμογές της τεχνητής νοημοσύνης στη βιομηχανία [2].

Η προβλεπτική συντήρηση αποσκοπεί στο να παρέχει τη δυνατότητα πρόβλεψης προβλημάτων και συντηρητικών αναγκών σε βιομηχανικούς κινητήρες πριν αυτά συμβούν [3]. Ο κεντρικός στόχος είναι η μείωση των ανεπίτρεπτων διακοπών, η βελτίωση της απόδοσης και η αποφυγή ακριβών επισκευών. Αυτό έχει σημαντικές επιπτώσεις στην αποτελεσματικότητα και την οικονομία των βιομηχανικών διεργασιών [4].

Η προβλεπτική συντήρηση επιτρέπει στις βιομηχανίες να προγραμματίσουν τις συντηρητικές επεμβάσεις τους με βάση την κατάσταση των εξοπλισμών τους, αντί να βασίζονται σε προγραμματισμένες συντηρήσεις που μπορεί να είναι περιττές και δαπανηρές [5]. Αυτό οδηγεί σε μεγαλύτερη αξιοποίηση των πόρων, εξοικονόμηση χρόνου και χρημάτων, καθώς και σε μείωση του κινδύνου ατυχημάτων στον χώρο εργασίας [4].

Η εφαρμογή της τεχνητής νοημοσύνης και της μηχανικής μάθησης στην προβλεπτική συντήρηση έχει ανοίξει νέες προοπτικές για τη βιομηχανία. Χάρη στη συλλογή, την ανάλυση και την ερμηνεία δεδομένων, μπορούμε να προβλέψουμε πότε θα απαιτηθεί συντήρηση, ποια εξαρτήματα πρέπει να αντικατασταθούν, και πώς μπορούμε να βελτιώσουμε την απόδοση των μηχανημάτων [4][6].

Οι τεχνικές μηχανικής μάθησης, όπως οι αναδρομικοί νευρωνικοί δίκτυα (RNNs) [7], συμπεριλαμβανομένων των Long Short-Term Memory (LSTM) [7][8] και των Gated Recurrent Unit (GRU) [11], καθώς και τα συνελκτικά νευρωνικά δίκτυα (CNNs) [9] και τα νευρωνικά δίκτυα feedforward (FNNs) [10], έχουν αποδειχθεί αποτελεσματικά εργαλεία για την ανάλυση και την πρόβλεψη των προβλημάτων στους βιομηχανικούς κινητήρες.

Στο πλαίσιο της παρούσας εργασίας, θα εξετάσουμε πώς οι προαναφερθείσες τεχνικές μηχανικής μάθησης μπορούν να εφαρμοστούν αποτελεσματικά στην προβλεπτική συντήρηση βιομηχανικών κινητήρων. Προκειμένου να επιτύχουμε αυτόν τον στόχο, θα αναλύσουμε διάφορα μοντέλα μηχανικής μάθησης και θα τα υλοποιήσουμε σε πραγματικά σύνολα

δεδομένων βιομηχανικών κινητήρων. Τα συμπεράσματα αυτής της έρευνας θα συμβάλουν στην καλύτερη κατανόηση της αξίας της προβλεπτικής συντήρησης και της επίδρασης της τεχνητής νοημοσύνης στον τομέα αυτόν.

## Στόχοι της Εργασίας

Οι στόχοι της παρούσας διπλωματικής εργασίας περιλαμβάνουν:

- Ανασκόπηση των Τεχνικών Συντήρησης:** Μια λεπτομερής ανασκόπηση των τεχνικών συντήρησης που χρησιμοποιούνται σήμερα στη βιομηχανία, καθώς και των προκλήσεων που προκύπτουν στον τομέα της προβλεπτικής συντήρησης.
- Μηχανική Μάθηση και Προβλεπτική Συντήρηση:** Η αξιολόγηση της αποτελεσματικότητας των διαφόρων αρχιτεκτονικών μηχανικής μάθησης, συμπεριλαμβανομένων των LSTM, GRU, CNN και FNN, στην προβλεπτική συντήρηση βιομηχανικών κινητήρων.
- Πρακτική Υλοποίηση με Σύνολα Δεδομένων:** Η ανάπτυξη και υλοποίηση των μοντέλων μηχανικής μάθησης σε πραγματικά σύνολα δεδομένων βιομηχανικών κινητήρων.
- Συγκριτικά Συμπεράσματα:** Η ανάλυση και τα συμπεράσματα που προκύπτουν από τη σύγκριση των διαφόρων αρχιτεκτονικών μηχανικής μάθησης στην προβλεπτική συντήρηση βιομηχανικών κινητήρων.

## Διάρθρωση της Εργασίας

Η διπλωματική εργασία αυτή θα περιλαμβάνει πέντε ενότητες, κάθε μία από τις οποίες αντιπροσωπεύει τα διαφορετικά στάδια της έρευνας και της ανάλυσης. Στην πρώτη αυτή ενότητα παρουσιάζεται η εισαγωγή στο θέμα της προβλεπτικής συντήρησης σε βιομηχανικούς κινητήρες, καθώς και οι στόχοι και η διάρθρωση της εργασίας. Στη δεύτερη ενότητα, παρέχεται μια ανασκόπηση των τεχνικών συντήρησης και των αρχιτεκτονικών μηχανικής μάθησης που θα χρησιμοποιηθούν στην εργασία. Η τρίτη ενότητα περιγράφει τη μεθοδολογία που ακολουθήθηκε για την εφαρμογή των μοντέλων μηχανικής μάθησης στα δεδομένα της προβλεπτικής συντήρησης. Στην τέταρτη ενότητα παρουσιάζονται τα μοντέλα μηχανικής μάθησης που αναπτύχθηκαν και η πρακτική υλοποίησή τους στην προβλεπτική

συντήρηση. Τέλος, στην πέμπτη και τελευταία ενότητα παρουσιάζονται τα συμπεράσματα της έρευνας και συζητούνται τα αποτελέσματα και οι προοπτικές για μελλοντικές εφαρμογές στον τομέα της προβλεπτικής συντήρησης.

# Θεωρητικό Υπόβαθρο

## Εισαγωγή στην προληπτική συντήρηση

Η αναγκαιότητα των τελευταίων χρόνων για τη μείωση του χρόνου συντήρησης στον κλάδο της βιομηχανίας, έχει επιφέρει ραγδαία ανάπτυξη σε τεχνικές εξοικονόμησης και ελαχιστοποίησης του λειτουργικού κόστους. Κατά συνέπεια, αυτό έχει παρακινήσει τους ερευνητές να κατασκευάσουν συστήματα παρακολούθησης κατάστασης για τα πολύπλοκα μηχανήματα, καθώς επιτρέπουν σημαντική μείωση του κόστους συντήρησης και τον έγκαιρο εντοπισμό δυνητικά καταστροφικών βλαβών. Επιπρόσθετα, πέρα από τον έγκαιρο εντοπισμό της σοβαρότητας μιας βλάβης, παρέχεται και η αναγνώριση της κατάστασης εξαρτημάτων, χωρίς την αναγκαιότητα αποσυναρμολόγησης του μηχανήματος. Η παρακολούθηση της κατάστασης της μηχανής μπορεί να οριστεί ως ένα πεδίο τεχνικής δραστηριότητας, στο οποίο παρατηρούνται οι επιλεγμένες φυσικές παράμετροι που σχετίζονται με τη λειτουργία του μηχανήματος, με σκοπό τον προσδιορισμό της ακεραιότητας της μηχανής (συστήματος). Αφού εκτιμηθεί αυτό, οι δραστηριότητες συντήρησης μπορούν να προγραμματιστούν μόνο όταν χρειάζεται, γεγονός που οδηγεί στη βέλτιστη χρήση των πόρων. Ο λόγος πίσω από τη χρήση της παρακολούθησης της κατάστασης του μηχανήματος, είναι η δημιουργία ακριβών και ποσοτικών πληροφοριών για την τρέχουσα κατάσταση μιας μηχανής, προκειμένου να προγραμματιστεί βέλτιστα η συντήρηση, να επιτευχθεί η μέγιστη παραγωγικότητα και να αποφευχθούν απροσδόκητες βλάβες. Έχει επίσης ως αποτέλεσμα βελτιωμένη συντήρηση κινδύνου και αυξημένη αξιοπιστία του μηχανήματος.[12].

Με την εμφάνιση του Industry4.0, η έννοια της πρόγνωσης και της διαχείρισης υγείας έχει γίνει αναπόφευκτη τάση στο πλαίσιο των βιομηχανικών μεγάλων δεδομένων και της έξυπνης κατασκευής. Η τεχνική της προγνωστικής συντήρησης έχει προσελκύσει μεγάλο ενδιαφέρον τόσο στον ακαδημαϊκό χώρο όσο και στη βιομηχανία. Ο λόγος είναι ότι προσφέρει μια αξιόπιστη λύση για τον χειρισμό της κατάστασης της υγείας του βιομηχανικού εξοπλισμού. Το I4.0 και οι βασικές του τεχνολογίες, διαδραματίζουν ουσιαστικό ρόλο για να κάνουν τα βιομηχανικά συστήματα αυτόνομα και έτσι να κάνουν δυνατή την αυτοματοποιημένη συλλογή δεδομένων από βιομηχανικές μηχανές/εξαρτήματα. Με βάση τα συλλεγόμενα δεδομένα, αλγόριθμοι μηχανικής εκμάθησης μπορούν να εφαρμοστούν για αυτοματοποιημένο εντοπισμό και διάγνωση σφαλμάτων.

Για την επιθεώρηση της κατάστασης λειτουργίας ενός μηχανήματος, τοποθετούνται αισθητήρες με τέτοιο τρόπο ώστε να μπορούν να ελέγχονται και να παρακολουθούνται πληροφορίες, σχετικά με την κατάσταση της υγείας του. Το I4.0 στοχεύει στη δημιουργία της λεγόμενης «έξυπνης βιομηχανίας», δηλαδή μιας βιομηχανίας στην οποία κάθε στοιχείο του συστήματος επικοινωνεί και συνεργάζεται μεταξύ τους και με τους ανθρώπους, σε πραγματικό χρόνο, μέσω του Διαδικτύου Υπηρεσιών. Σε αυτή την προοπτική, η συνεχής παρακολούθηση της κατάστασης του μηχανήματος είναι επιτακτική. Υπάρχουν πολλές τεχνικές που χρησιμοποιούνται συνήθως για τη διάγνωση οποιουδήποτε ελαττώματος, που εμφανίζεται κατά τη λειτουργία του μηχανήματος: παρακολούθηση κραδασμών, θερμογραφία, ανάλυση υπερήχων και πολλές άλλες [13]. Ωστόσο, μόνο μερικές από αυτές τις τεχνικές επιτρέπουν μια συνεχή ανάλυση των παραμέτρων λειτουργίας των μηχανών. Ο συνεχής έλεγχος συγκεκριμένων παραμέτρων του μηχανήματος επιτρέπει την ανίχνευση του αρχικού σφάλματός του. Το θέμα των προγνωστικών, έχει μελετηθεί ευρέως και πολλές εργασίες είναι διαθέσιμες στην επιστημονική βιβλιογραφία[14][15][16].

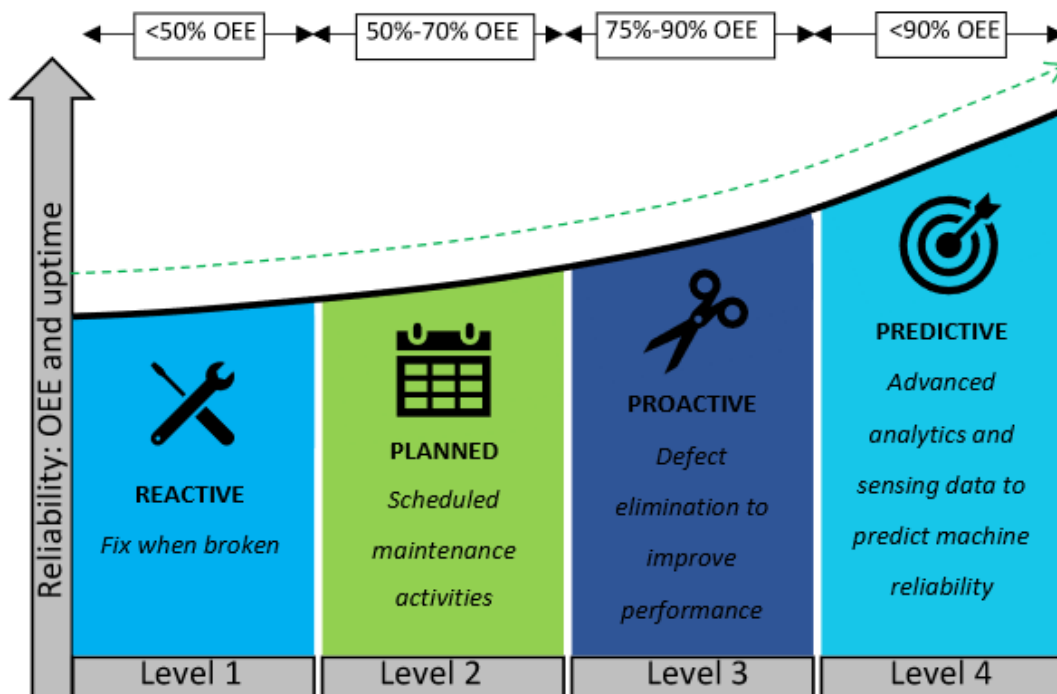
Οι τεχνικές για τις πολιτικές συντήρησης μπορούν να κατηγοριοποιηθούν στις ακόλουθες κύριες ταξινομήσεις .

1. **Αστοχία υλικού(R2F):** Είναι γνωστή ως διορθωτική συντήρηση ή μη προγραμματισμένη συντήρηση. Αποτελεί την απλούστερη τεχνική συντήρησης, η οποία εκτελείται μόνο όταν ο εξοπλισμός έχει αποτύχει. Μπορεί να οδηγήσει σε μεγάλο χρόνο διακοπής λειτουργίας του εξοπλισμού και σε υψηλό κίνδυνο δευτερογενών βλαβών και έτσι να δημιουργήσει πολύ μεγάλο αριθμό ελαττωματικών προϊόντων στην παραγωγή.
2. **Προγραμματισμένη Συντήρηση (PvM):** Επίσης γνωστή ως προγραμματισμένη συντήρηση ή συντήρηση βάσει χρόνου (TBM). Το PvM αναφέρεται σε περιοδικά εκτελούμενη συντήρηση, με βάση ένα προγραμματισμένο χρονοδιάγραμμα, προκειμένου να προβλεφθούν οι βλάβες. Μερικές φορές οδηγεί σε περιττή συντήρηση που αυξάνει το λειτουργικό κόστος. Ο κύριος στόχος εδώ είναι να βελτιωθεί η απόδοση του εξοπλισμού ελαχιστοποιώντας τις αστοχίες στην παραγωγή (2).
3. **Συντήρηση βάσει συνθηκών (CBM):** αυτή η μέθοδος συντήρησης βασίζεται στη συνεχή παρακολούθηση του μηχανήματος ή του εξοπλισμού ή στην υγεία της διεργασίας τους, η οποία μπορεί να πραγματοποιηθεί μόνο όταν είναι πραγματικά απαραίτητο. Το CBM συνήθως δεν μπορεί να προγραμματιστεί εκ των προτέρων.

4. **Προβλεπτική Συντήρηση (PdM):** Γνωστή ως Προβλεπτική Συντήρηση, όπου τα προγράμματα συντήρησης εκτελούνται μόνο όταν χρειάζεται. Βασίζεται στη συνεχή παρακολούθηση του εξοπλισμού ή του μηχανήματος, όπως το CBM. Χρησιμοποιεί εργαλεία πρόβλεψης για την κατανόηση της αναγκαιότητας συντήρησης, ώστε αυτή να μπορεί να προγραμματιστεί στοχευμένα. Επιπλέον, επιτρέπει την ανίχνευση αστοχίας σε πρώιμο στάδιο με βάση τα ιστορικά δεδομένα χρησιμοποιώντας μεθόδους μηχανικής μάθησης.



Εικόνα 1: Βασική διαδικασία Προβλεπτικής Συντήρησης



Εικόνα 2: Τύπος Συντήρησης

Η προβλεπτική συντήρηση, αποδείχθηκε ότι είναι μια από τις πιο ελπιδοφόρες στρατηγικές μεταξύ άλλων στρατηγικών συντήρησης που έχει την ικανότητα να επιτύχει αυτά τα χαρακτηριστικά. Οι τεχνικές που αξιοποιεί, προσφέρουν αποτελέσματα που μπορούν να προβλεφθούν με βάση ένα μοντέλο που έχει προετοιμαστεί και εκπαιδευτεί μέσω ιστορικών δεδομένων εισόδου και συμπεριφοράς εξόδου. Συμβατικές τεχνικές και τεχνικές μηχανικής μάθησης, στα πλαίσια της τεχνητής νοημοσύνης, που έχουν τεράστια πλεονεκτήματα καθώς έχουν την ικανότητα να χειρίζονται πολυμεταβλητά, υψηλών διαστάσεων δεδομένα και μπορούν να εξάγουν κρυφές σχέσεις σε πολύπλοκα, δυναμικά και χαοτικά περιβάλλοντα.

## **Χαρακτηριστικά υγείας εξοπλισμού**

Η μηχανική χαρακτηριστικών [17][18], στοχεύει στην εξαγωγή πολλών βασικών χαρακτηριστικών παραμέτρων από τα αρχικά μετρούμενα σήματα, για να καταστήσει τη σχέση μεταξύ του σήματος και των καταστάσεων μηχανικής κατεργασίας σαφέστερη[19]. Διάφορες στρατηγικές για τον εντοπισμό αστοχιών έχουν αναπτυχθεί χρησιμοποιώντας εκτίμηση παραμέτρων βάσει μοντέλου, μεθόδους αναγνώρισης προτύπων και προσέγγιση εμπειρών συστημάτων που βασίζεται στη γνώση [19][20]. Γενικά, αυτές οι μέθοδοι λειτουργούν με ψηφιακά σήματα που λαμβάνονται από αισθητήρες οι οποίοι συλλαμβάνουν και κωδικοποιούν διάφορα φυσικά μεγέθη ως συναρτήσεις του χρόνου, όπως ταχύτητα, επιτάχυνση, πίεση, μετατόπιση κ.λπ. Ανάλογα με τον σκοπό της παρακολούθησης, οι σχετικές πληροφορίες που κωδικοποιούνται σε αυτά τα σήματα, πρέπει να εκφράζεται ως σύνολο μετρούμενων μεγεθών. Επομένως, πρέπει να εξαχθούν αρκετές τιμές από το σήμα που περιγράφει τη συνολική λειτουργία και συγκεκριμένα σφάλματα, με επαρκή ακρίβεια. Αυτές οι τιμές που ονομάζονται χαρακτηριστικά [21], μπορούν να προσδιοριστούν στον τομέα χρόνου (mean, root mean square, crest factor, standard deviation, kurtosis, skewness value), συχνότητας (power spectrum density, total spectrum energy, cepstrum ratio) ή στο πεδίο χρόνου-συχνότητας (continuous wavelet transform coefficients, discrete wavelet transform coefficients and wavelet packet transform coefficients)[22], απευθείας από το σήμα ή από τη μοντελοποίηση του συστήματος που βασίζεται σε αυτό το σήμα.

Μετά από αυτό το στάδιο εξαγωγής χαρακτηριστικών, πρέπει να εφαρμοστούν αξιόπιστοι αλγόριθμοι απόφασης, με σκοπό την ερμηνεία και ταξινόμηση των χαρακτηριστικών. Σε πολλές περιπτώσεις, τα χαρακτηριστικά μετρώνται ως συγκεκριμένες τιμές και διατάσσονται

σε διανυσματική μορφή, έτσι ώστε οι κλασικές προσεγγίσεις αναγνώρισης προτύπων και οι σύγχρονες προσεγγίσεις που βασίζονται σε νευρωνικά δίκτυα [23] ασαφή ή νευροασαφή συστήματα [24] ,να μπορούν να είναι ικανά για ταξινόμηση του χώρου χαρακτηριστικών. Παραδοσιακά, τα χαρακτηριστικά καθορίζονται χειροκίνητα μέσω ανθρώπινης εμπειρίας με σκοπό την επίβλεψη της εξαγωγής των σχετικών πληροφοριών, κάτι που απαιτεί εκτενή εξειδίκευση. Αυτό, καθιστά την μηχανική χαρακτηριστικών αποδεκτή ως το πιο κρίσιμο μέρος για την μέθοδο παρακολούθησης.

## **Εξαγωγή Χαρακτηριστικών στο πεδίο του Χρόνου**

Τα σήματα υποβάλλονται σε προεπεξεργασία, για την εξαγωγή χαρακτηριστικών που είναι σε θέση να επιβεβαιώσουν την ύπαρξη σφάλματος και να προσδιορίσουν τον τύπο, την κατάσταση και την τοποθεσία του. Η ορθή επιλογή χαρακτηριστικών είναι ένα ουσιαστικό βήμα στην διάγνωση σφαλμάτων [25]. Τα χαρακτηριστικά του τομέα χρόνου είναι τα πιο απλά χαρακτηριστικά, τα οποία μπορούν εύκολα να εξαχθούν, σε σύγκριση με τα χαρακτηριστικά που εξάγονται από τον τομέα συχνότητας και τον τομέα συχνότητας-χρόνου. Επίσης, οι λειτουργίες του τομέα χρόνου έχουν επαρκείς πληροφορίες για ένα σφάλμα. Μια συνοπτική μαθηματική περιγραφή των βασικών χαρακτηριστικών του χρόνου ακολουθούν παρακάτω:

1. Μέσος όρος του σήματος(RMS):
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

,όπου,  $x_i$  το στιγμιαίο πλάτος του σήματος δόνησης και το N ο αριθμός δειγμάτων.

2. Η τετραγωνική ρίζα του μέσου όρου των τετραγώνων του σήματος:
$$x_{rms} = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$$

3. Κύρτωση ως καμπύλη της κατανομής:
$$K = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \bar{x})^4}{\sigma^4}$$

,όπου,  $\sigma$  η τυπική απόκλιση.

## **Εφαρμογή του προτύπου ISO**

Στις περισσότερες βιομηχανίας που είναι εξοπλισμένες με μηχανές βαρέως τύπου, τα περιστρεφόμενα μέρη από τα οποία απαρτίζονται, είναι βασικά συστατικά για την σωστή λειτουργία τους. Τα δυνητικά σφάλματα, μπορεί να προκαλέσουν υποβάθμιση της ποιότητας λειτουργίας, καθώς και της απόδοσης της μηχανής και ολόκληρου του συστήματος. Αυτές οι



συνθήκες, σχετίζονται άμεσα με το κόστος συντήρησης της εγκατάστασης και πολύ σημαντικότερα με το επίπεδο ασφάλειας.

Ως μέρος της προβλεπτικής συντήρησης βιομηχανικών μηχανών, η εφαρμογή της παρακολούθησης των δονήσεων, είναι ένας από τους σημαντικούς και τους ακριβέστερους τρόπους αποφυγής και πρόληψης αστοχιών του συστήματος εγκατάστασης. Στις πρακτικές παρακολούθησης και ανάλυσης κραδασμών μηχανημάτων πεδίου, έχει αναπτυχθεί από τον Διεθνή Οργανισμό Τυποποίησης (ISO)[26], μια ποικιλία σχετικών μετρήσεων και προτύπων. Η έννοια της μέτρησης κραδασμών σε ένα μηχάνημα είναι η μόνιμη παρακολούθηση, η οποία είναι ένα σύστημα όπου ένα σύνολο οργάνων ελέγχει συνεχώς την κατάσταση της μηχανής σε περιορισμένο αριθμό. Όλα τα περιστρεφόμενα μηχανήματα, είναι συνήθως εξοπλισμένα με κάποιους ανιχνευτές εγγύτητας επαφής, ως ένδειξη κραδασμών. Αυτοί οι αισθητήρες, εγκαθίστανται συνήθως στα κύρια μέρη που απαρτίζουν το σύστημα, με σκοπό την ακριβέστερη και έγκυρη ενημέρωση της κατάστασής του. Τόσο τα δεδομένα δόνησης όσο και οι τάσεις, καταγράφονται και παρουσιάζονται συνεχώς.

Τα Διεθνή Πρότυπα (ISO), καθορίζουν γενικές συνθήκες και διαδικασίες για τη μέτρηση και την αξιολόγηση των κραδασμών, χρησιμοποιώντας μετρήσεις που γίνονται σε περιστρεφόμενα, μη περιστρεφόμενα και μη παλινδρομικά μέρη πλήρων μηχανών. Τα κριτήρια αξιολόγησης, διακρίνονται σε διαφορετικές κατηγορίες και παρουσιάζονται τόσο ως προς το μέγεθος των κραδασμών, όσο και ως προς την αλλαγή τους, εξάγοντας πληροφορίες σε χρονοσειρές. Ένα αξιόπιστο και αξιόλογο έγγραφο που καθορίζει γενικές κατευθυντήριες γραμμές για την μέτρηση και την αξιολόγηση των μηχανικώς κραδασμών, παρέχεται από το ISO 20816[26]. Το συγκεκριμένο, παρέχει μια επισκόπηση από ιστορικές συλλογές δεδομένων για κατηγοριοποιημένα μηχανήματα. Οι μετρήσεις αυτές, χρησιμοποιούνται για διάφορους σκοπούς, που κυμαίνονται από δοκιμές λειτουργικότητας και αποδοχής κατάσταση, έως προηγμένες πειραματικές δοκιμές, καθώς και διαγνωστικές και αναλυτικές έρευνες.

## **Παράμετροι Μέτρησης**

Υπάρχουν 3 βασικές κατηγορίες μεγεθών, που αποσκοπούν στην μέτρηση των κραδασμών

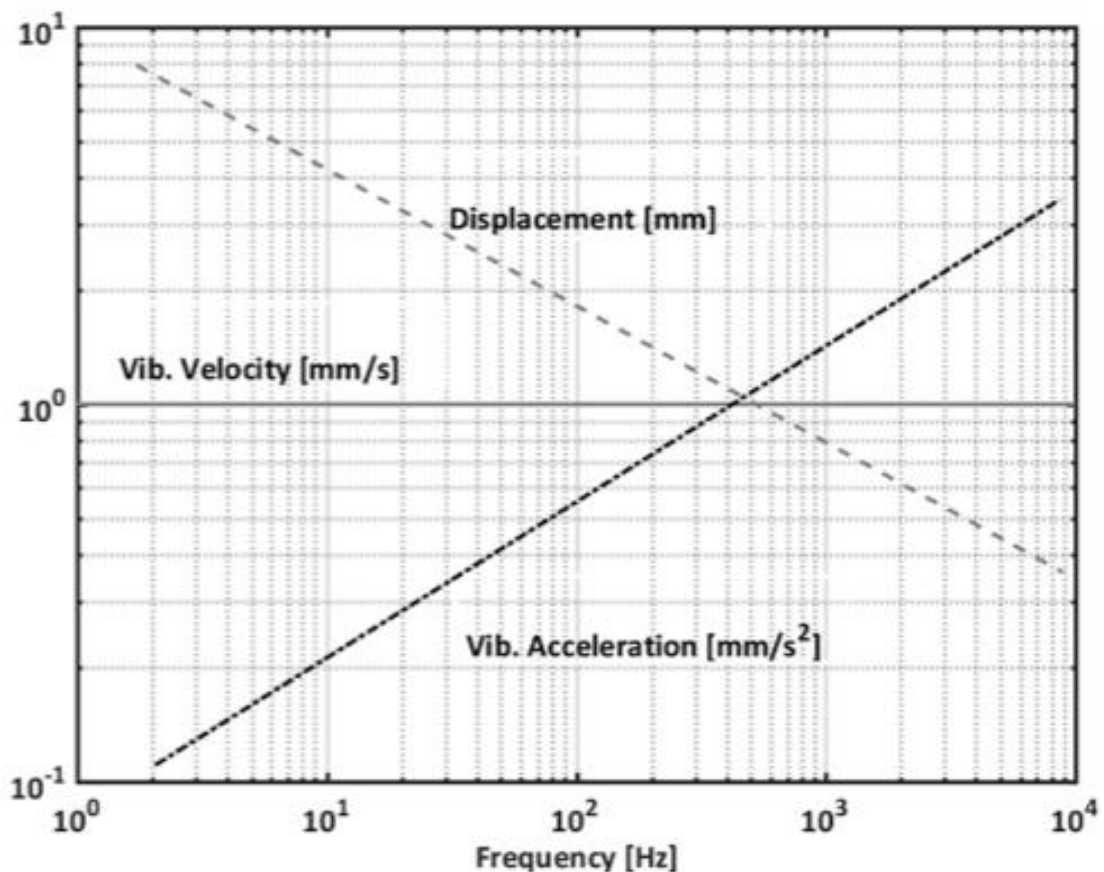
1. Μετατόπιση δόνησης, μετρούμενη σε μικρόμετρα

2. Ταχύτητα δόνησης, μετρούμενη σε χιλιοστά ανα δευτερόλεπτο
3. Επιτάχυνση δόνησης, μετρούμενη σε μέτρα ανα τετραγωνικό δευτερόλεπτο

Features of vibration	Definition	Function equation
Acceleration, $a$	Alteration of the velocity from vibration object.	$a(t) = \frac{dv(t)}{dt}$ (Eq. 1)
Velocity, $v$	Indicates how fast the object vibrates.	$v(t) = \int a(t)dt + C_1$ (Eq. 2)
Displacement, $x$	Indicates how large the object vibrates.	$x(t) = \int v(t)dt + C_2$ (Eq. 3)
Root mean Square, RMS	Root mean square of vibration amplitude values.	$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n a_i^2}$ (Eq. 4)

Εικόνα 3: Μαθηματική έκφραση μεγεθών δόνησης

Στην ανάλυση κραδασμών, η πραγματική κατάσταση του μηχανήματος αναλύεται χρησιμοποιώντας τον τομέα χρόνου. Όπως αναφέρθηκε, ένας δείκτης με βάση τα ενεργειακά χαρακτηριστικά των δονητικών σημάτων που λαμβάνονται από τους διάφορους αισθητήρες εγγύτητας, περιλαμβάνει την τιμή Root Mean Square (RMS). Η Εικόνα 4 απεικονίζει τη σχέση μεταξύ του πλάτους μετατόπισης, ταχύτητας και επιτάχυνσης και παρέχει μια χρήσιμη καθοδήγηση σχετικά με το ποια ποσότητα πρέπει να υιοθετηθεί για την ανάλυση της κατάστασης δόνησης που επηρεάζει τον υπό ανάλυση εξοπλισμό. Η μετατόπιση κραδασμών παρουσιάζει σημαντική σημασία σε πολύ χαμηλές συχνότητες ενώ τα πλάτη της επιτάχυνσης γίνονται πιο σημαντικά από τη μετατόπιση και την ταχύτητα για το φάσμα συχνοτήτων άνω του 1 kHz. Για συχνότητες που κυμαίνονται από 10 Hz έως 1 kHz, η ταχύτητα δόνησης είναι η ποσότητα που αποδίδει σημαντικά και ανιχνεύσιμα πλάτη. Δεδομένου ότι η πλειονότητα των ελαττωμάτων που επηρεάζουν τα περιστρεφόμενα μηχανήματα είναι στην περιοχή 10 Hz–1 kHz, η ταχύτητα δόνησης χρησιμοποιείται για την ανάλυση του σήματος [14][15].



Εικόνα 4: Σχέση μεταξύ μετατόπισης, ταχύτητας και επιτάχυνσης στο φάσμα συχνοτήτων

Η δόνηση ενός συγκεκριμένου μηχανήματος εξαρτάται από το μέγεθός του, τα χαρακτηριστικά του δονούμενου σώματος και του συστήματος στερέωσης και τον σκοπό για τον οποίο έχει σχεδιαστεί. Ως εκ τούτου, είναι απαραίτητο να ληφθούν υπόψη οι διάφοροι σκοποί και οι σχετικές περιστάσεις κατά τον καθορισμό περιοχών μέτρησης κραδασμών για διαφορετικούς τύπους μηχανών. Για όλα σχεδόν τα μηχανήματα, ανεξάρτητα από τον τύπο των ρουλεμάν που χρησιμοποιούνται, γενικά, οι μετρήσεις της ταχύτητας δόνησης ευρείας ζώνης σε δομικά μέρη, όπως τα περιβλήματα ρουλεμάν, χαρακτηρίζουν επαρκώς τις συνθήκες λειτουργίας των περιστρεφόμενων στοιχείων του άξονα, σε σχέση με την απρόσκοπτη λειτουργία τους. Αντίστοιχα, εξίσου επαρκείς για να χαρακτηριστούν ως συνθήκες απρόσκοπτης λειτουργίας είναι οι μετρήσεις που γίνονται σε μη περιστρεφόμενα μέρη. Υπάρχουν τύποι μηχανών, όπως αμοστρόβιλοι, αεριοστρόβιλοι και στροβιλοσυμπιεστές, οι οποίοι μπορούν να έχουν διάφορους τρόπους δόνησης στο εύρος στροφών λειτουργίας, για τους οποίους οι μετρήσεις σε δομικά μέλη, ενδέχεται να μην

χαρακτηρίζουν επαρκώς την κατάσταση λειτουργίας του μηχανήματος, αν και τέτοιες μετρήσεις είναι χρήσιμες. Τέτοιες μηχανές περιέχουν γενικά εύκαμπτα συστήματα και οι αλλαγές στην κατάσταση δόνησης μπορούν να ανιχνευθούν αποτελεσματικότερα και με μεγαλύτερη ακρίβεια, σε σχέση με μετρήσεις στα περιστρεφόμενα στοιχεία. Οι μηχανές που έχουν σχετικά άκαμπτα και/ή βαριά περιβλήματα σε σύγκριση με τη μάζα του ρότορα είναι τυπικές για εκείνες τις κατηγορίες μηχανών για τις οποίες προτιμώνται συχνά οι μετρήσεις κραδασμών του άξονα.

Αναγνωρισμένο εδώ και πολλά χρόνια είναι, ότι χρησιμοποιώντας τη μέτρηση  $V_{rms}$  ως χαρακτηριστικό της δονητικής απόκρισης ενός ευρέος φάσματος μηχανών, ήταν πολύ επιτυχημένη και συνεχίζει να είναι. Για απλές εναλλασσόμενες κυματομορφές που αποτελούνται από έναν διακριτό αριθμό αρμονικών συνιστωσών γνωστού πλάτους και φάσης, είναι δυνατόν, μέσω της ανάλυσης Fourier, να συσχετιστούν διάφορα θεμελιώδη μεγέθη (μετατόπιση, ταχύτητα, επιτάχυνση, κορυφή, κορυφή σε κορυφή, r.m.s., μέσος όρος, κ.λπ.) χρησιμοποιώντας αυστηρά καθορισμένες μαθηματικές σχέσεις. Η τιμή της ταχύτητας,

$V_{rms}$ , μπορεί να υπολογιστεί με τον τύπο:  $v_{rms} = \sqrt{\frac{1}{T} \int_0^T v^2(t) dt}$ , όπου  $v(t)$  η εξαρτώμενη μεταβλητή και  $T$  η περιοδική δειγματοληψία στον τομέα του χρόνου

Διάφορα πειράματα και δοκιμές που έχουν πραγματοποιηθεί, επιτρέπουν τον καθορισμό ενός χάρτη συνθηκών λειτουργίας, τα σημεία του οποίου συσχετίζουν την τιμή RMS της ταχύτητας δόνησης, το  $V_{rms}$ , με τις κατηγοριοποιημένες τιμές απρόσκοπτης λειτουργίας. Αξίζει να σημειωθεί ότι ο αναφερόμενος χάρτης συνθηκών λειτουργίας είναι ακριβής, καθώς το ελεγχόμενο εύρος στροφών δεν περιλαμβάνει κρίσιμες ταχύτητες (όπως έχει αποδειχθεί πειραματικά). Προκειμένου να εκτιμηθεί η σοβαρότητα και, ως εκ τούτου, η αποδοχή οποιασδήποτε συνθήκης λειτουργίας, θα πρέπει να παρέχεται μια τιμή καταφλίου, αφαιρούμενη από το πρότυπο ISO 20816-1. Το αποτέλεσμα της σύγκρισης μεταξύ της τρέχουσας τιμής  $V_{rms}$  και αυτής του ορίου, καθορίζει την αποδοχή της κατάστασης του κινητήρα που υφίσταται κραδασμούς. Μια τέτοια διαδικασία επιτρέπει την ποσοτική αξιολόγηση του τρόπου με τον οποίο διακρίνεται μια ελατωματική μηχανή. Η Εικόνα 5 απεικονίζει την κατάσταση λειτουργίας του κινητήρα (όσον αφορά το  $V_{rms}$ ).



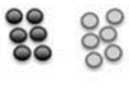

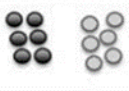

RMS Velocity [mm/s]	Class I Nom Pow. < 15 kW	Class II 15-75 kW	Class III > 75 kW Rigid Foundation	Class IV > 75 kW Soft Foundation
0.28	Good	Good	Good	Good
0.45	Good	Good	Good	Good
0.71	Good	Good	Good	Good
1.12	Satisfactory	Good	Good	Good
1.8	Satisfactory	Satisfactory	Good	Good
2.8	NOT Satisfactory	Satisfactory	Satisfactory	Good
4.5	NOT Satisfactory	NOT Satisfactory	Satisfactory	Satisfactory
7.1	Unacceptable	NOT Satisfactory	NOT Satisfactory	Satisfactory
11.2		Unacceptable	NOT Satisfactory	NOT Satisfactory
18.0			Unacceptable	NOT Satisfactory
28.0				Unacceptable
45.0				Unacceptable

Εικόνα 5: Όρια σοβαρότητας κραδασμών σύμφωνα με το ISO 20816.

## Εξαγωγή συμπερασμάτων προληπτικής συντήρησης

### Προσέγγιση Μηχανικής Μάθησης

Η τεχνολογία μηχανικής μάθησης (ML) έχει χρησιμοποιηθεί ευρέως σε διάφορους τομείς μηχανικής όπως η πολιτική μηχανική [27], η πρόβλεψη ανέμου [28], η παρακολούθηση της δομικής υγείας [29] και η κατασκευαστική μηχανική [30] στις μέρες μας. Ο σκοπός των τεχνικών μηχανικής μάθησης είναι να αναγνωρίζουν αυτόματα, σύνθετα μοτίβα σε ένα σύνολο δεδομένων, επιτρέποντας την εξαγωγή συμπερασμάτων ή την πρόβλεψη σε νέα σύνολα δεδομένων [31]. Οι τεχνικές μηχανικής μάθησης επιτρέπουν τον εντοπισμό ομοιογενών ομάδων στο σύνολο δεδομένων εισόδου. Όταν μια ομάδα δεδομένων είναι διαθέσιμη για κάθε περίπτωση (εποπτευόμενη μάθηση), οι τεχνικές μηχανικής εκμάθησης επιτρέπουν τη δημιουργία ενός ταξινομητή ή μιας συνάρτησης παλινδρόμησης, που προβλέπει τη συμμετοχή της κάθε κατηγορίας με ποσοστιαία αναλογία Εικόνα 5. Προκειμένου να διασφαλιστεί η καλή απόδοση των τεχνικών μηχανικής εκμάθησης σε ένα σύνολο δεδομένων, είναι αναγκαίο να εντοπιστούν, να αφαιρεθούν ή να ελαχιστοποιηθούν όλα τα ασαφή στοιχεία που παρεκκλίνουν στην αμερόληπτη κατάταξη τους. Επομένως, πριν από οποιαδήποτε διαδικασία μηχανικής μάθησης, θα πρέπει να επιβεβαιωθεί η αντιπροσωπευτικότητα του συνόλου δεδομένων εισόδου για τον πραγματικό πληθυσμό (δηλαδή, ο υπό μελέτη πληθυσμός), ο θόρυβος θα πρέπει να αφαιρεθεί, τα δεδομένα που λείπουν θα πρέπει να αντιμετωπιστούν και τα ρεαλιστικά δεδομένα, (δηλ. ο αριθμός των παραμέτρων/χαρακτηριστικών) πρέπει να προσαρμοστούν.

	Unsupervised Learning	Supervised Learning
<b>Dataset</b>	Unknown category membership 	Known category membership (black and grey) 
<b>Learning Process</b>	Clustering process (grouping of similar cases) 	Classification process (training for black and grey) 
<b>Output</b>	<b>Clusters</b> (homogeneous groups of cases) 	<b>Classifier</b> (black and grey) 

Εικόνα 6: Απεικόνιση των δύο προσεγγίσεων μηχανικής μάθησης

Η Μηχανική Μάθηση, είναι ευρέως ταξινομημένη σε δυο βασικές κατηγορίες, αλλά παρόλα αυτά, βασισμένη στην κατάσταση που κρίνεται να αντιμετωπιστεί, μπορούν να συνδυαστούν επιτυγχάνοντας το επιθυμητό αποτέλεσμα για την εκάστοτε εφαρμογή:

- Εποπτευόμενη Μάθηση: Εκπαιδεύει το σύστημα να αναγνωρίζει τη σχέση μεταξύ μεταβλητών προς αναγνώριση και μεταβλητών στόχων. Η βασική τμηματοποίηση αυτής της κατηγορίας είναι:
  - Προβλήματα ταξινόμησης
  - Προβλήματα υποβιβασμού
- Μη εποπτευόμενη Μάθηση: Η εκμάθηση του συστήματος γίνεται από το ίδιο με εσωτερικές τεχνικές, χωρίς να παρέχεται καμία πληροφορία από τα σύνολα δεδομένων που θα διαχειριστεί. Αξιόλογος είναι ότι τρόπος που δημιουργεί κρυφά μοτίβα και σχέσεις με τα δεδομένα κατεργασίας. Η τμηματοποίηση είναι η εξής:
  - Ομαδοποίηση
  - Μείωση διαστάσεων

Όλοι οι τύποι θεωρητικών κατανομών που χρησιμοποιούνται για την ανάπτυξη χρονοδιαγραμμάτων συντήρησης, εξακολουθούν να βασίζονται σε προσεγγιστικά στατιστικά μοντέλα κατανομής. Ο «μέσος χρόνος μέχρι την αστοχία» που προκύπτει από τέτοιες κατανομές δεν αντιπροσωπεύει απαραίτητα το χρόνο μέχρι την αστοχία του μηχανήματος. Εξακολουθεί να υπάρχει σημαντική αβεβαιότητα στην ερμηνεία των δεδομένων. Επεξηγείται στο [32], ότι η χρήση λανθασμένων μοντέλων για την εκτίμηση της αστοχίας ή της

κατανομής διάρκειας ζωής από λανθασμένα δεδομένα, έχει ως αποτέλεσμα μη βέλτιστο και μεροληπτικό κόστος αντικατάστασης..

Πληθώρα από μεθόδους Μηχανικής Μάθησης, έχουν εφαρμοστεί στην διαχείριση δεδομένων δόνησης. Στον τομέα της προβλεπτικής συντήρησης (PdM) της εκάστοτε μηχανής, αναφορικά κάποιες μέθοδοι είναι τα Τεχνητά Νευρωνικά Δίκτυα (ANN), Μηχανή Διανυσμάτων Υποστήριξης (SVM), Λογιστική Παλινδρόμηση (LR), Γραμμική Παλινδρόμηση, Κοντινότερος Γείτονας Κ (KNN), Δέντρο Αποφάσεων (DT) και πολλές ακόμη [33], [34]. Η επιλογή του καταλληλότερου, απλού και πιο αποτελεσματικού αλγορίθμου είναι ύψιστης σημασίας και ιδιαίτερης δυσκολίας . Οι αλγόριθμοι ML συνήθως απαιτούν τη συλλογή τεράστιων ποσοτήτων δεδομένων των σεναρίων κατάστασης αποτυχίας και των σεναρίων συνθηκών υγείας για την εκπαίδευση των μοντέλων τους.. Η ανάπτυξη αλγορίθμου ML καλύπτει την επιλογή ιστορικών δεδομένων, τα δεδομένα προεπεξεργασίας, την επιλογή μοντέλου, την εκπαίδευση μοντέλων, την επικύρωση μοντέλου και τη συντήρηση. Τα βήματα που εμπλέκονται στην ανάπτυξη , μπορούν να καθοριστούν ως είσοδος, εξαγωγή και επιλογή χαρακτηριστικών, χαρακτηριστικά, παραδοσιακές και σύγχρονες τεχνικές ML και έξοδος.

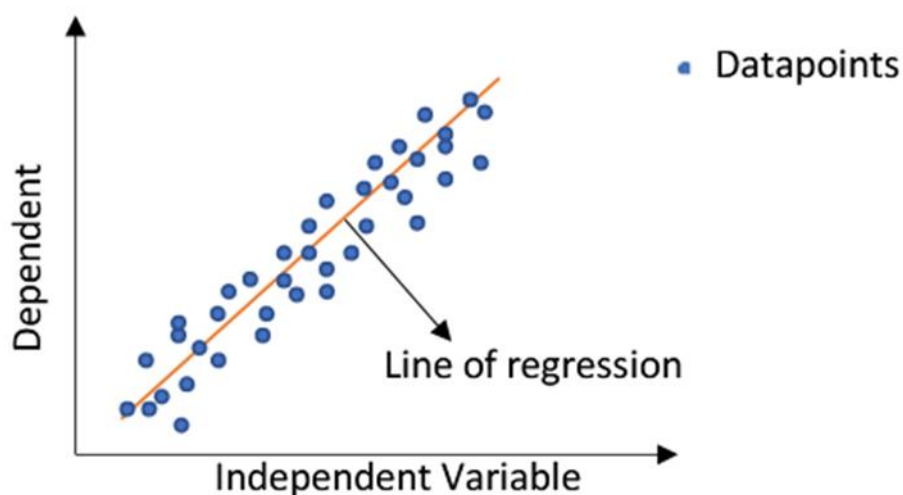
## **Τεχνικές Μηχανικής Μάθησης**

### **Linear Regression**

Η γραμμική παλινδρόμηση είναι ίσως η πιο γνωστή τεχνική και ο πιο κατανοητός αλγόριθμος στην στατιστική και την μηχανική μάθηση. Αναφέρεται σε έναν πολυμεταβλητό γραμμικό συνδυασμό συντελεστών παλινδρόμησης [35]. Η εκμάθηση ενός μοντέλου γραμμικής παλινδρόμησης, σημαίνει εκτίμηση των τιμών των συντελεστών που χρησιμοποιούνται στην αναπαράσταση, με τα δεδομένα που έχουμε διαθέσιμα.

Με απλή γραμμική παλινδρόμηση όταν έχουμε μία μόνο είσοδο, μπορούμε να χρησιμοποιήσουμε στατιστικά στοιχεία για να υπολογίσουμε τους συντελεστές. Αυτό απαιτεί να υπολογίζονται στατιστικές ιδιότητες από τα δεδομένα, όπως μέσους όρους, τυπικές αποκλίσεις, συσχετίσεις και συνδιακύμανση. Όλα τα δεδομένα πρέπει να είναι διαθέσιμα για τη διέλευση και τον υπολογισμό των στατιστικών.

Όταν υπάρχουν μία ή περισσότερες είσοδοι, μπορεί να χρησιμοποιηθεί μια διαδικασία βελτιστοποίησης των τιμών των συντελεστών, ελαχιστοποιώντας επαναληπτικά το σφάλμα του μοντέλου στα δεδομένα εκπαίδευσης. Αυτή η λειτουργία ονομάζεται Gradient Descent και λειτουργεί ξεκινώντας με τυχαίες τιμές για κάθε συντελεστή. Το άθροισμα των τετραγωνικών σφαλμάτων, υπολογίζεται για κάθε ζεύγος τιμών εισόδου και εξόδου. Χρησιμοποιείται ένας ρυθμός μάθησης ως παράγοντας κλίμακας και οι συντελεστές ενημερώνονται με κατεύθυνση προς την ελαχιστοποίηση του σφάλματος. Η διαδικασία επαναλαμβάνεται, έως ότου επιτευχθεί ένα ελάχιστο άθροισμα στο τετραγωνικό σφάλμα ή δεν είναι δυνατή η περαιτέρω βελτίωση.



Εικόνα 7: Γραμμική Παλινδρόμηση στην Μηχανική Μάθηση

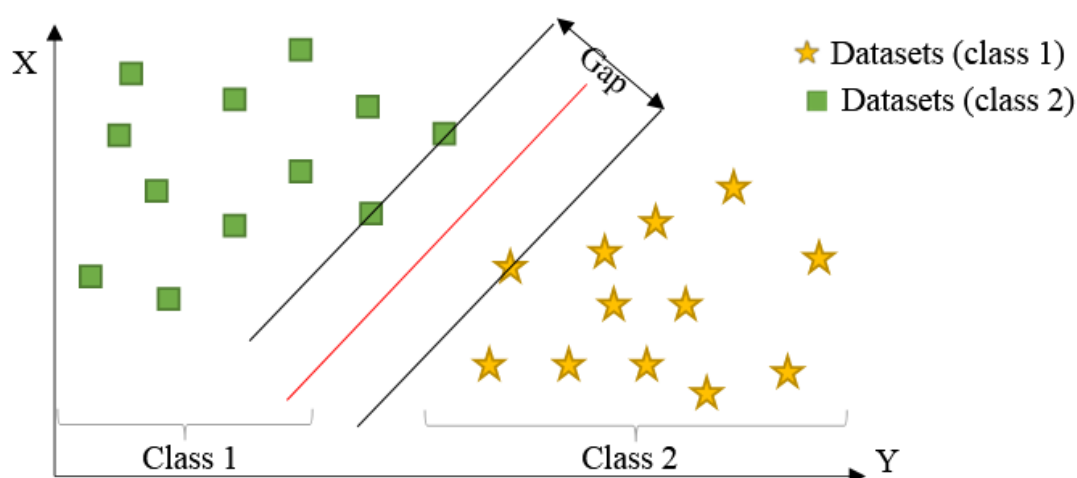
### **Support Vector Machine (SVM)**

Η SVM είναι πολύ γνωστή τεχνική ML που χρησιμοποιείται ευρέως τόσο για ανάλυση ταξινόμησης όσο και για ανάλυση παλινδρόμησης, λόγω της υψηλής ακρίβειάς της. Παρουσιάζεται ως ένας από τους δέκα αλγόριθμους στην ανάκτηση δεδομένων και θεωρείται επίσης ως μια από τις πιο ισχυρές και ακριβείς μεθόδους μεταξύ των γνωστών αλγορίθμων, καθώς είναι τεχνική που μπορεί να χρησιμοποιηθεί στην ανάλυση χρονοσειρών. Ορίζεται ως μια έννοια στατιστικής μάθησης με μια προσαρμοστική υπολογιστική μέθοδο και θεωρείται εποπτευόμενη τεχνική ML που μπορεί να εκτελέσει αναγνώριση προτύπων, ταξινόμηση και



ανάλυση παλινδρόμησης. Στο PdM του βιομηχανικού εξοπλισμού, τα SVM έχουν εφαρμοστεί ευρέως για τον προσδιορισμό μιας συγκεκριμένης κατάστασης με βάση το ληφθέν σήμα [36].

Η μέθοδος παρουσιάζει καλή απόδοση για τον εντοπισμό προβλημάτων σε περιστρεφόμενες μηχανές με βάση την ανάλυση κραδασμών [37]. Έχει εφαρμοστεί αποτελεσματικά σε πολλούς κλάδους της μηχανικής, όπως η διάγνωση σφαλμάτων και η εξόρυξη δεδομένων [38]. Ο αλγόριθμος διαχωρίζει τα δεδομένα που αναλύονται, επιλέγοντας ένα υπερεπίπεδο ικανό να διαφοροποιεί τα δεδομένα από δύο διακριτές κλάσεις. Το καλύτερο υπερεπίπεδο για SVM δίνεται με το μεγαλύτερο περιθώριο μεταξύ των δύο κλάσεων. Το περιθώριο ορίζεται ως το μέγιστο πλάτος της περιοχής που δεν περιέχει δεδομένα και είναι επίσης παράλληλο με το υπερεπίπεδο. Ο αλγόριθμος έχει μερικά μειονεκτήματα όπως ο αργός ρυθμός εκμάθησης για μεγάλο όγκο δεδομένων εκπαίδευσης, καθώς και ότι τα SVM αναπτύχθηκαν αρχικά για δυαδική ταξινόμηση, η οποία απαιτεί εξωγενείς προσαρμογές για προβλήματα πολλαπλών τάξεων.



Εικόνα 8: Αλγόριθμος SVM

Η συγκεκριμένη τεχνική έχει απασχολήσει πολλούς ερευνητές και επαγγελματίες για προβλεπτική συντήρηση (PdM). Για παράδειγμα, ο Konar [39] χρησιμοποίησε το SVM για την ανίχνευση σφαλμάτων ενός τριφασικού κινητήρα επαγωγής χρησιμοποιώντας σήματα δονήσεων κατά την εκκίνηση. Τα αποτελέσματα έδειξαν ότι ο ταξινομητής SVM έδωσε

εξαιρετικά αποτελέσματα αφού είναι πολύ απλός και εύκολος στην εφαρμογή του σε σύγκριση με άλλες μεθόδους όπως τα τεχνητά νευρωνικά δίκτυα (ANN). Ομοίως, ο Hu [40] πρότεινε μια νέα μέθοδο για τη διάγνωση σφαλμάτων που βασίζεται σε έναν Βελτιωμένο Μετασχηματισμό Πακέτων Κυμάτων ( Improved Wavelet Package Transform ), μια τεχνική αξιολόγησης εξ αποστάσεως και SVM. Τα βέλτιστα χαρακτηριστικά χρησιμοποιούνται ως είσοδος στο SVM με σκοπό τον εντοπισμό διαφορετικών, μη φυσιολογικών περιπτώσεων. Η προτεινόμενη μέθοδος χρησιμοποιείται για την ανίχνευση σφαλμάτων σε ρουλεμάν κύλισης και τα αποτελέσματα δοκιμών έδειξαν ότι το SVM μπορούσε να διαχωρίσει αξιόπιστα διαφορετικές συνθήκες σφάλματος και να προσδιορίσει τη σοβαρότητα τους με μεγάλη ακρίβεια. Βιβλιογραφικά η μέθοδος βρίσκει ενδιαφέρον σε προβλήματα ανάπτυξης προβλεπτικής συντήρησης.

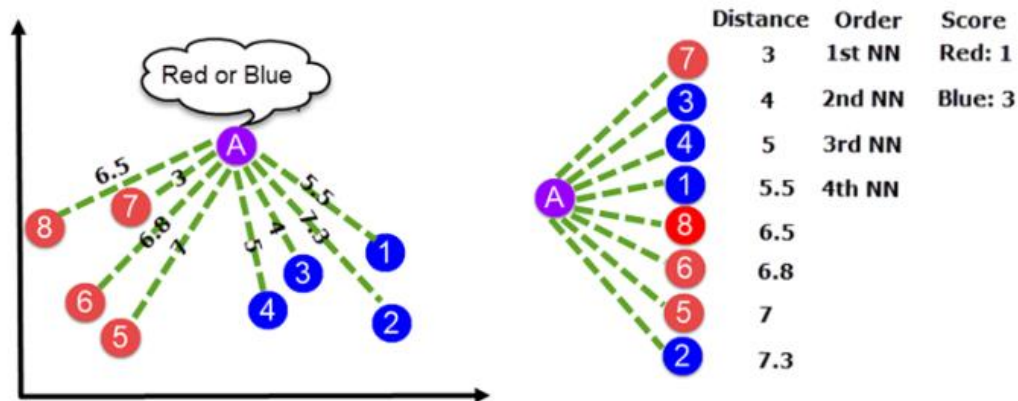
### **k Nearest Neighbor (kNN)**

Ο ταξινομητής kNN είναι ένας ταξινομητής που χρησιμοποιείται συνήθως μεταξύ των ταξινομητών ML λόγω της ευκολίας εφαρμογής και της απλότητάς του. Ο ταξινομητής kNN [41] είναι ένας αλγόριθμος που προσπαθεί να ταξινομήσει τα νέα δείγματα σύμφωνα με τα πιο παρόμοια γειτονικά του. Αυτό συμβαίνει για ένα συγκεκριμένο σύνολο εκπαίδευσης με ταξινομημένα δείγματα  $T = \{(a_1, b_1), (a_2, b_2), \dots, (a_N, b_N)\}$ , όπου  $a_i$  είναι το διάνυσμα χαρακτηριστικών των μη επισημασμένων δεδομένων και  $b_i$  είναι τα πραγματικά χαρακτηριστικά, όπου  $b_i = c_1, c_2, \dots, c_K$  και  $i = 1, 2, \dots, N$ . Για ένα δείγμα εκπαίδευσης  $(a, b)$ , ο αλγόριθμος kNN αναζητά τις  $k$  πλησιέστερες περιπτώσεις στο  $a$  με βάση μια υπολογισμένη Ευκλείδεια απόσταση. Η γειτονιά που περιέχει αυτά τα  $k$  δείγματα αντιπροσωπεύεται από  $N_k(a)$ . Στη συνέχεια, η ετικέτα του δείγματος δοκιμής  $a$  μπορεί να βρεθεί σύμφωνα με τους ακόλουθους κανόνες:  $y = \underset{j}{\operatorname{argmax}} D(b_i = c_j)$ ,  $i = 1, 2, \dots, N$ ;  $J = 2, 3, \dots, k$ , όπου  $D$  είναι η συνάρτηση προσδιορισμού.

Η κατάσταση που εκφράζεται μαθηματικά παραπάνω μπορεί εύκολα να εξηγηθεί σε τρία βήματα για το  $k = 4$ , το οποίο φαίνεται στην Εικόνα 9. Στο πρώτο βήμα, η απόσταση μεταξύ ενός νέου δείγματος που περιβάλλεται από δείγματα με καθορισμένες κατηγορίες μετράται ένα προς ένα. Στο δεύτερο βήμα, οι αποστάσεις εγγύτητας ταξινομούνται από το πλησιέστερο προς το πιο απομακρυσμένο. Στο τρίτο βήμα, η βαθμολόγηση γίνεται σύμφωνα με τη λίστα

του πλησιέστερου γείτονα και το νέο δείγμα περιλαμβάνεται στην ομάδα με την υψηλότερη βαθμολογία.

Η τεχνική αυτή μπορεί να εφαρμοστεί με ιδιαίτερα καλά αποτελέσματα για δεδομένα δονήσεων που είναι προκαθορισμένα σε timestamps.



Εικόνα 9: Αλγόριθμος kNN

## Τεχνικές Βαθιάς Μάθησης

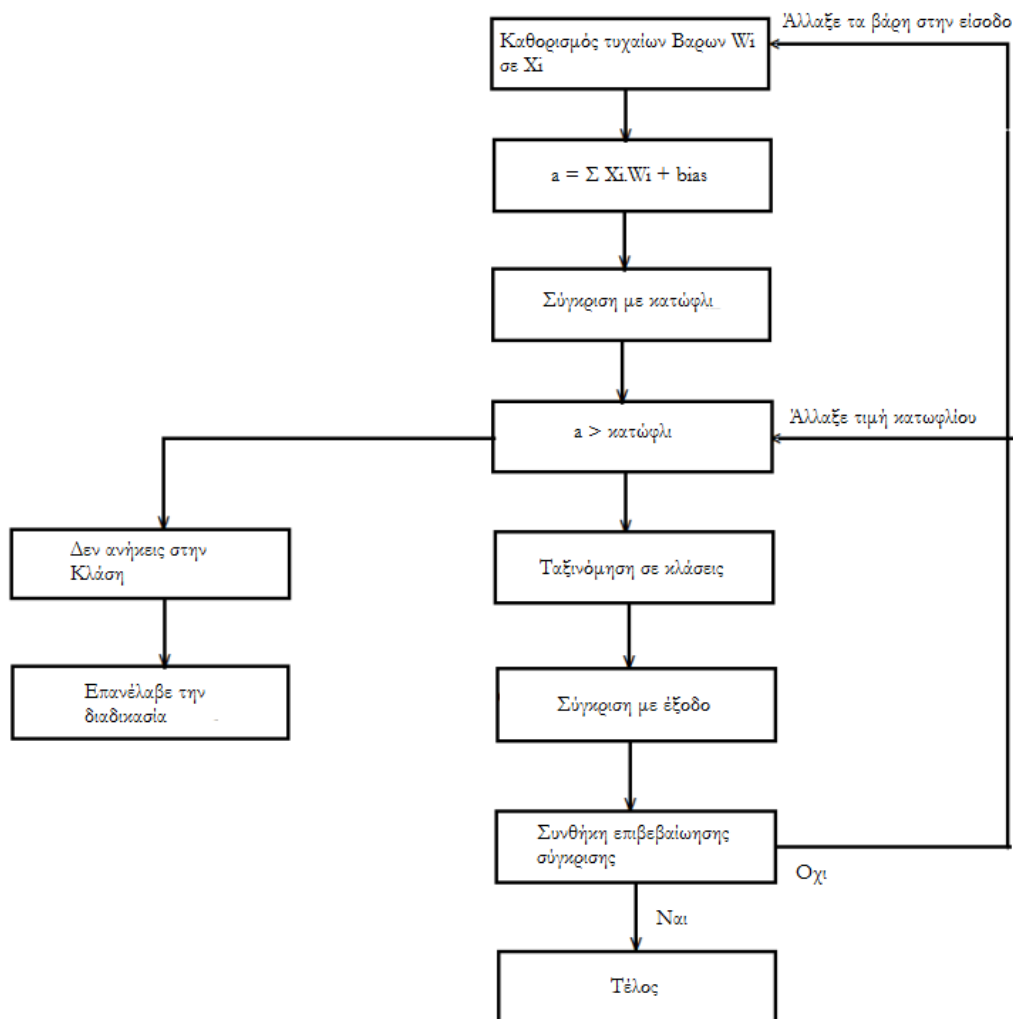
### Artificial Neural Network (ANN)

Το τεχνητό νευρωνικό δίκτυο (ANN) έχει σχεδιαστεί για να μιμείται τον ανθρώπινο εγκέφαλο για να σκέφτεται μόνο του και να αναλαμβάνει δράση χωρίς να είναι ρητά προγραμματισμένο. Ακριβώς όπως ένας ανθρώπινος εγκέφαλος, το μοντέλο ANN [42] αποτελείται από τεχνητούς νευρώνες ή κόμβους ως βασικά δομικά στοιχεία. Οι κόμβοι είναι διατεταγμένοι σε επίπεδα: στρώμα εισόδου, στρώμα εξόδου και ένα ή περισσότερα κρυφά στρώματα, όπου το κρυφό στρώμα χρησιμοποιείται για την αύξηση της ακρίβειας της ταξινόμησης. Η λειτουργία του αλγορίθμου μπορεί να εξηγηθεί από τα παρακάτω βήματα.

1. Ο αλγόριθμος εκχωρεί βάρη ή προτεραιότητες σε καθεμία από τις μεταβλητές εισόδου.
2. Στη συνέχεια πολλαπλασιάζει το βάρος με τη μεταβλητή εισόδου, χι και προσθέτει την απόκλιση σε αυτό. Το ίδιο πράγμα γίνεται σε κάθε κόμβο και το άθροισμα αυτών των τιμών μεταδίδεται στο επόμενο επίπεδο. Η εκμάθηση προχωρά καθώς τα δεδομένα κινούνται σε κάθε επίπεδο.

3. Τα δεδομένα στο τέλος του κόμβου εξόδου, συγκρίνονται με μια καθορισμένη τιμή κατωφλίου και εάν η υπολογισμένη τιμή υπερβαίνει το όριο, το δiάνυσμα εισόδου ταξινομείται στην προβλεπόμενη κλάση ή αλλιώς μεταβιβάζεται για έλεγχο για άλλες κλάσεις.
4. Η υπολογισμένη κλάση συγκρίνεται με την πραγματική έξοδο και οποιοδήποτε σφάλμα διορθώνεται μεταβάλλοντας τον αριθμό των επιπέδων και τον αριθμό των νευρώνων.
5. Η διαδικασία συνεχίζεται έως ότου επιτευχθεί μια ακριβής προσαρμογή και, ως εκ τούτου, το μοντέλο εκπαιδευτεί.

Το μπλοκ διάγραμμα της λειτουργίας ενός αλγόριθμου νευρικού δικτύου τροφοδοσίας προς τα εμπρός πίσω βρόχου είναι όπως φαίνεται στην Εικόνα 10.



Εικόνα 10: Μπλόκ διάγραμμα ανάπτυξης αλγορίθμου ANN

Ένας μεμονωμένος νευρώνας σε ένα δίκτυο, λαμβάνει ένα μόνο κομμάτι δεδομένων εισόδου και εκτελεί κάποιο μετασχηματισμό δεδομένων για να παράγει ένα μόνο κομμάτι αναφοράς εξόδου στο νευρωνικό. Αυτά είναι τα πρώτα μοντέλα δικτύου, που ονομάζονταν **perceptrons**, είναι σχετικά απλά συστήματα που χρησιμοποιούν πολλούς συνδυασμούς απλών συναρτήσεων, όπως τον υπολογισμό της κλίσης μιας γραμμής. Ενώ αυτοί οι μετασχηματισμοί είναι πολύ απλοί μεμονωμένα, μαζί καταλήγουν σε περίπλοκη συμπεριφορά για ολόκληρο το σύστημα και την συναρμολόγηση μεγάλου αριθμού μετασχηματισμών για τον υπολογισμό προηγμένων συναρτήσεων. Στην προσπάθεια να καταπολεμηθεί αυτό το ζήτημα με τη συναρμολόγηση των perceptrons μαζί, οι ερευνητές αντιμετώπισαν ένα άλλο

πρόβλημα: λίγοι υπολογιστές εκείνη την εποχή ήταν σε θέση να αποθηκεύσουν αρκετά δεδομένα για να εκτελέσουν αυτά τα προγράμματα.

Στις αρχές της δεκαετίας του 2000, όταν οι καινοτομίες στο υλικό υπολογιστών επέτρεψαν πιο περίπλοκες τεχνικές μοντελοποίησης, οι ερευνητές είχαν ήδη αναπτύξει συστήματα νευρωνικών δικτύων που συνδύαζαν πολλά στρώματα νευρώνων, συμπεριλαμβανομένων των συνελκτικών νευρωνικών δικτύων (CNN), των πολυστρωματικών perceptrons (MLP) και των επαναλαμβανόμενων νευρικών δικτύων (RNN). Αυτές οι αρχιτεκτονικές ονομάζονται βαθιά νευρωνικά δίκτυα, επειδή έχουν πολλά στρώματα νευρώνων που συνδυάζονται για να δημιουργήσουν μια «βαθιά» στοίβα νευρώνων. Κάθε μία από αυτές τις αρχιτεκτονικές βαθιάς μάθησης έχει τα δικά της σχετικά δυνατά σημεία:

- Τα δίκτυα MLP αποτελούνται από πολυεπίπεδα perceptrons. Τείνουν να είναι καλοί στην επίλυση απλών εργασιών, όπως η εφαρμογή ενός φίλτρου σε κάθε pixel σε μια φωτογραφία. Αποτελούνται από ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Το σήμα εισόδου διαδίδεται προς τα εμπρός ανα επίπεδα, μέχρι να φτάσει στην έξοδο του δικτύου. Η εκπαίδευση του νευρωνικού γίνεται με εποπτευόμενο τρόπο και, παρόλο που υπάρχουν και άλλοι, ο συγκεκριμένος αλγόριθμος είναι ο πιο συνηθισμένος αλγόριθμος εκμάθησης.
- Τα δίκτυα CNN έχουν σχεδιαστεί για να επεξεργάζονται δεδομένα, εφαρμόζοντας την ίδια λειτουργία συνέλιξης σε μια είσοδο. Αυτό καθιστά απλούστερη και πιο αποτελεσματική την επεξεργασία, η οποία γενικά αποδίδει πολύ υψηλές διαστάσεις και απαιτεί μεγάλη επεξεργασία. Τα CNN μπορούν να μοντελοποιήσουν σχέσεις βάσει χρόνου, μεταξύ αισθητήρων, χρησιμοποιώντας δισδιάστατα φίλτρα. Το κύριο πλεονέκτημά τους είναι ότι με το μοίρασμα βάρους, μειώνουν τους απαιτούμενους πόρους εκπαίδευσης και την πολυπλοκότητα του μοντέλου, αλλά έχουν περιορισμένη μνήμη
- Τα RNN υιοθετήθηκαν ευρέως στην επεξεργασία φυσικής γλώσσας επειδή ενσωματώνουν έναν βρόχο στις συνδέσεις μεταξύ των νευρώνων, επιτρέποντας στις πληροφορίες να παραμένουν σε μια αλυσίδα. Όταν η αλυσίδα των νευρώνων σε ένα RNN «αναπτύσσεται», γίνεται ευκολότερο να δούμε ότι αυτά τα μοντέλα αποτελούνται από πολλά αντίγραφα του ίδιου νευρώνα, καθένα από τα οποία περνάει πληροφορίες στον διάδοχό του. Οι νευρώνες που δεν είναι οι πρώτοι ή οι τελευταίοι

σε ένα αναπτυγμένο RNN αναφέρονται μερικές φορές ως «κρυμμένα» στρώματα δικτύου. Ο πρώτος και ο τελευταίος νευρώνας ονομάζονται στρώματα «εισόδου» και «εξόδου», αντίστοιχα. Τα ποιά συνηθισμένα RNN είναι τα LTSMs και GRUs που επιτυγχάνουν αποτελέσματα SotA. Και τα δυο εξίσου παρουσιάζουν παρόμοια εξαιρετικά αποτελέσματα σε εφαρμογές με σκοπό την ανίχνευση και την πρόγνωση ανωμαλιών. Το LSTM παρουσιάζει ένα προβάδισμα. Τα τυπικά RNN αποδεικνύεται ότι είναι τα βέλτιστα για εργασίες που περιλαμβάνουν ακολουθίες.

Ένα ενδιαφέρον άρθρο, δημοσιευμένο από τους Gustavo Scalabrini Sampaio και λοιποί [43], παρουσιάζει την μέθοδο MLP για πρόβλεψη αποτυχίας κινητήρα με τεχνικές νευρωνικών δικτύων. Το μοντέλο ANN συγκρίθηκε ως προς τις τιμές του δείκτη απόδοσης RMSE με άλλες τεχνικές μηχανικής εκμάθησης: Regression Tree, Random Forest και Support Vector Machine. Τα αποτελέσματα της εκπαίδευσης και των συγκριτικών δοκιμών ήταν ικανοποιητικά και έδειξαν ότι το μοντέλο ANN ήταν ανώτερο από τις άλλες τεχνικές. Η επιτυχία της εκπαίδευσης και της δοκιμής του νευρωνικού δικτύου MLP που περιγράφηκε σε αυτή την ερευνητική εργασία για την πρόβλεψη της αστοχίας του κινητήρα, δείχνει ότι αυτή η τεχνική θα μπορούσε να εφαρμοστεί στη συντήρηση που βασίζεται σε βιομηχανικές συνθήκες. Παρόλη την καλή απόδοση του συγκεκριμένου ANN για το σύνολο δεδομένων που ενσωματώθηκε σε αυτό το έγγραφο, αποδεικνύεται ότι τα MLP με αλγόριθμο εκμάθησης backpropagation μπορεί να έχουν κάποιες αδυναμίες.

Ένα ακόμη ενδιαφέρον άρθρο που δημοσιεύτηκε από τους Shao Haidong και λοιποί, το [44] παρουσιάζει μια μεθοδολογία βελτιστοποίησης AE (Autoencoder) για τη διάγνωση σφαλμάτων περιστρεφόμενων μηχανημάτων. Πρώτον, δημιούργησαν μια νέα συνάρτηση απώλειας με σκοπό την εύρεση του σφάλματος, για τη βελτίωση της εκμάθησης χαρακτηριστικών. Δεύτερον, βελτιστοποίησαν τις βασικές παραμέτρους του μοντέλου για να το προσαρμόσουν στα χαρακτηριστικά σήματος. Αυτό το μοντέλο εφαρμόστηκε για τη διάγνωση σφαλμάτων του κιβωτίου ταχυτήτων και των ρουλεμάν κυλίνδρων.

Μια ενδιαφέρουσα τεχνική με εφαρμογές PdM είναι η βαθιά ενίσχυση της εκμάθησης. Οι Chi Zhang και λοιποί, στο [45] χρησιμοποιεί για εκμάθηση HI(Health Indicator ) την μέθοδο DRL (Deep Reinforcement Learning), όπου ξεπερνά τα feed-forward δίκτυα, αλλά σαφέστατα παρουσιάζει χαμηλότερη απόδοση σε σύγκριση με CNN και LSTM. Αυτή η τεχνική συνίσταται στη μεταφορά της γνώσης που αποκτάται από ένα σύνολο δεδομένων σε ένα άλλο

σύνολο δεδομένων. Η διαδικασία βρίσκει ανταπόκριση στην επαναχρησιμοποίηση ενός, ή του πλήρους προεκπαιδευμένου μοντέλου προσαρμόζοντάς το στις νέες απαιτήσεις. Ενώ αυτή η προσέγγιση μερικές φορές απαιτεί επανεκπαίδευση του μοντέλου, απαιτεί λιγότερα δεδομένα και χρόνο. Το πείραμα διεξήχθη με ένα σύνολο δεδομένων NASA C-MAPSS (Commercial Modular AeroPropulsion System Simulation), καθώς είναι ένα τυπικό σημείο αναφοράς που χρησιμοποιείται ευρέως στην κοινότητα PHM [46]. Πρόκειται για ένα σύνολο δεδομένων προσομοίωσης κινητήρα Turbofan, όπου περιέχει μετρήσεις δονήσεων, που προσομοιώνουν την λειτουργία κινητήρων turbofan υπό διαφορετικές συνθήκες. Αυτό το σημείο αναφοράς έχει τέσσερα σύνολα δεδομένων (FD001 ~ FD004), που αποτελούνται από δεδομένα αισθητήρα (συμπεριλαμβανομένων αισθητήρων δόνησης) και δεδομένα κατάστασης λειτουργίας.

Ο Giovanni Martinez-Arellano και Svetan Ratchev παρουσίασαν έναν ταξινομητή χρησιμοποιώντας τη Bayesian αναζήτηση και το CNN για το AD. Χρησιμοποίησαν αρχικά ένα μικρό σύνολο δεδομένων για να εκπαιδεύσουν το μοντέλο και στη συνέχεια χρησιμοποίησαν το μοντέλο για να ταξινομήσουν τα υπόλοιπα δεδομένα. Το μοντέλο χρησιμοποιεί μοντελοποίηση αβεβαιότητας για να αναλύσει τις παρατηρήσεις που δεν μπορούν να ταξινομηθούν σωστά λόγω της υψηλής εντροπίας. Τέλος, επιλέγει τους κορυφαίους 100 με την υψηλότερη εντροπία για να ρωτήσει έναν τεχνικό γνώσης τομέα, ζητώντας του/της να τους επισημάνει για να επανεκπαιδεύσει το μοντέλο με αυτά τα νέα δεδομένα. Αυτή η διαδικασία ακολουθείται μέχρι το μοντέλο να επιτύχει καλή ακρίβεια. Αυτή η εργασία είναι ένα παράδειγμα του τρόπου χρήσης δύο ενδιαφέρουσες τεχνικές στον τομέα του PdM για την αντιμετώπιση του προβλήματος των ανεπαρκών ετικετοποιημένων δεδομένων, ερωτώντας τους τεχνικούς τομέα και δείχνοντάς τους τις περιπτώσεις από τις οποίες το μοντέλο μπορεί να μάθει τα περισσότερα.

Μια ακόμη εφαρμογή με CNN, δημοσιεύτηκε επίσης από τους Jaun Pablo και Cabezas Rodriguez [47], όπου η συγκεκριμένη έρευνα περιορίζεται στην μελέτη των απλών δικτύων ταξινόμησης και τις πιθανές εφαρμογές τους στη διάγνωση σφαλμάτων αποκλειστικά σε ρουλεμάν, μέσω σημάτων δόνησης. Τα σύνολα δεδομένων προέρχονται από το Case Western Reserve University Bearing Data Center (CWR) καθώς και από το Society of Machine Failure Prevention Technology (MFPT). Ακόμη ο Xiaojie Guo και λοιποί [48] δημοσίευσαν ένα σχετικά άρθρο με παρόμοια τεχνική για διάγνωση και πρόγνωση κατάστασης και αποτυχίας σε περιστρεφόμενα ρουλεμάν, σε βαριές βιομηχανικές μηχανές. Τα αποτελέσματα των



πειραμάτων με δεδομένα από ρουλεμάν, καταδεικνύουν την υπεροχή του προτεινόμενου μοντέλου ADCNN σε άλλες μεθόδους διάγνωσης σφαλμάτων, όπως τα παραδοσιακά DCNN. Το προτεινόμενο μοντέλο επιτυγχάνει υψηλό βαθμό ακρίβειας και προσφέρει μια διαδικασία αυτόματης εξαγωγής χαρακτηριστικών που είναι πρακτική και βολική για χρήση στη διάγνωση σφαλμάτων περιστρεφόμενης μηχανής. Αξίζει να σημειωθεί ότι ορισμένα έξυπνα μοντέλα μάθησης είναι επίσης ανταγωνιστικά όσον αφορά τη γρήγορη ταχύτητα εκπαίδευσης και την υψηλή ακρίβεια, που βασίζονται στην καλά σχεδιασμένη εξαγωγή χαρακτηριστικών.

Ένα άρθρο από τον Olgun Aydin και λοιποί [49] ,περιγράφει το LSTM το οποίο εφαρμόζεται στο Apache Spark που προσφέρει ένα κατανεμημένο περιβάλλον επεξεργασίας δεδομένων μεγάλης κλίμακας προκειμένου να προβλέψει την τρέχουσα κατάσταση ζωής ενός κινητήρα. Προηγούμενες έρευνες που στοχεύουν στην πρόβλεψη των συνθηκών συντήρησης βασίζονταν κυρίως σε κρυφά μοντέλα Markov ή παραδοσιακά τεχνητά νευρωνικά δίκτυα. Αντί για τα κοινά μοντέλα, το LSTM που εργάζεται σε κατανεμημένο περιβάλλον, είναι μια τεχνολογία αιχμής για την πρόβλεψη της τρέχουσας κατάστασης του κινητήρα. Ακόμη ένα άρθρο με εφαρμογή LSTM είναι από τον Fabrizio De Vita [50] και λοιποί, όπου στόχος είναι να προβλεφθεί το RUL (Remaining Useful Life), και συγκεκριμένα να εκτιμηθεί για ένα σύνολο βιομηχανικών κινητήρων, με στόχο τη μείωση των περιόδων εκτός σύνδεσης λόγω συντήρησης. Χρησιμοποιήθηκε εποπτευόμενη προσέγγιση, όπου και χρειάστηκε ένα σύνολο δεδομένων που να περιέχει ολόκληρο το «ιστορικό» ενός συνόλου κινητήρων, μέχρι την αποτυχία τους. Λόγω της πολυπλοκότητας της συλλογής σε πραγματικές συνθήκες, επιλέχθηκε ένα σύνολο δεδομένων από την NASA μέσω του εργαλείου C-MAPSS που προσομοιώνει την υποβάθμιση των κινητήρων. Το μοντέλο που κατασκευάστηκε σε αυτήν την έρευνα, κατέληξε να είναι το καλύτερο όσον αφορά το RMSE και το ποσοστό σφάλματος, σε σύγκριση με μοντέλα όπως SVM, DNN.

## **Μεθοδολογία**

### **Σύνολο Δεδομένων**

Το σύνολο δεδομένων που χρησιμοποιήθηκε στην εργασία αποτελείται από δεδομένα που συλλέχθηκαν με τη χρήση αισθητήρων δόνησεων και θερμοκρασίας. Αυτοί οι αισθητήρες κατέγραψαν διαφορομετρικές δονήσεις, συμπεριλαμβανομένης της μετατόπισης δόνησης σε μικρόμετρα, της ταχύτητας δόνησης σε χιλιοστά ανά δευτερόλεπτο και της επιτάχυνσης δόνησης σε μέτρα ανά τετραγωνικό δευτερόλεπτο. Επιπλέον, καταγράφηκε και η θερμοκρασία σε βαθμούς Κελσίου.

Τα δεδομένα που προέκυψαν από αυτούς τους αισθητήρες περιλάμβαναν μετρήσεις δονήσεων για τον κινητήρα, το δεξί ρουλεμάν κύλισης και το αριστερό ρουλεμάν κύλισης. Αυτά τα δεδομένα αποτέλεσαν τη βάση για την εκπαίδευση και την αξιολόγηση των μοντέλων προβλεπτικής συντήρησης. Η ποικιλία των μετρικών δονήσεων και η συλλογή δεδομένων από διάφορα σημεία του μηχανισμού επέτρεψαν την πιο ολοκληρωμένη ανάλυση της κατάστασης των βιομηχανικών κινητήρων και τη δημιουργία αξιόπιστων προβλέψεων για τη συντήρησή τους.

### **Ρύθμιση του περιβάλλοντος**

Σε αυτήν την ανάλυση προγνωστικής συντήρησης, το υπολογιστικό περιβάλλον δημιουργήθηκε χρησιμοποιώντας το Google Colab, μια πλατφόρμα φορητών υπολογιστών Python που βασίζεται σε σύννεφο. Το Google Colab προσφέρει το πλεονέκτημα της απρόσκοπτης ενοποίησης με το Google Drive, παρέχοντας ένα συνεργατικό και επεκτάσιμο περιβάλλον για ανάλυση δεδομένων, μηχανική μάθηση και εργασίες βαθιάς μάθησης.

Το Google Colab, με συντομογραφία Google Collaboratory, χρησιμεύει ως μια ελεύθερα προσβάσιμη πλατφόρμα που προσφέρει ένα περιβάλλον Notebook Jupyter άμεσα προσβάσιμο μέσω ενός προγράμματος περιήγησης ιστού. Αξιοποιώντας τη δύναμη της υποδομής cloud της Google, παρέχει υπολογιστικούς πόρους υψηλής ποιότητας, συμπεριλαμβανομένης της επιτάχυνσης GPU, η οποία είναι ζωτικής σημασίας για την αποτελεσματική εκπαίδευση σύνθετων μοντέλων μηχανικής εκμάθησης.

Τα σημειωματάρια Colab διαμορφώθηκαν με τις απαραίτητες βιβλιοθήκες και τις εξαρτήσεις που χρησιμοποιούνται συνήθως στην ανάλυση δεδομένων και τη μηχανική μάθηση. Βιβλιοθήκες κλειδιών, όπως οι NumPy, Pandas, Matplotlib και Scikit-Learn, χρησιμοποιήθηκαν για χειρισμό δεδομένων, οπτικοποίηση και μοντελοποίηση. Επιπλέον, πλαίσια βαθιάς μάθησης όπως το TensorFlow και το Keras χρησιμοποιήθηκαν για τη δημιουργία και την εκπαίδευση μοντέλων νευρωνικών δικτύων.

Για πρόσβαση και χειρισμό του συνόλου δεδομένων για προγνωστική συντήρηση, το Google Drive ενσωματώθηκε απευθείας στο περιβάλλον Colab. Αυτό επέτρεψε την εύκολη εισαγωγή συνόλων δεδομένων, αποθήκευση σημείων ελέγχου μοντέλων και εξαγωγή αποτελεσμάτων ανάλυσης. Η απρόσκοπτη ενσωμάτωση με το Google Drive απλοποιεί τη διαχείριση δεδομένων, καθιστώντας το μια βολική επιλογή για συλλογικά έργα δεδομένων.

Το Google Colab παρέχει πρόσβαση σε ισχυρούς πόρους υλικού, συμπεριλαμβανομένων των Μονάδων Επεξεργασίας Γραφικών (GPU) και των Μονάδων Επεξεργασίας Τενσογράφου (TPU). Αυτοί οι επιταχυντές υλικού επιταχύνουν σημαντικά την εκπαίδευση μοντέλων μηχανικής μάθησης και βαθιάς μάθησης, επιτρέποντας την αποτελεσματική επεξεργασία μεγάλων συνόλων δεδομένων και την εκπαίδευση σύνθετων μοντέλων.

Με τη ρύθμιση του περιβάλλοντος στο Google Colab, εξασφαλίστηκε ένας ευέλικτος, επεκτάσιμος και πλούσιος σε πόρους χώρος εργασίας για τη διεξαγωγή ολοκληρωμένης ανάλυσης πρόβλεψης συντήρησης. Αυτό το περιβάλλον επέτρεψε την απρόσκοπτη εκτέλεση της προεπεξεργασίας δεδομένων, της ανάπτυξης μοντέλων και της αξιολόγησης, οδηγώντας τελικά σε πολύτιμες ιδέες και προγνωστικά μοντέλα για την αξιολόγηση της υγείας των μηχανημάτων.

## **Εκπαίδευση Μοντέλων Βαθιάς Μάθησης**

Η μεθοδολογία εκπαίδευσης των μοντέλων βαθιάς μάθησης που ακολουθήθηκε για την εκπαίδευση των μοντέλων βαθιάς μάθησης περιλαμβάνει τα παρακάτω:

1. Προεπεξεργασία Δεδομένων για Κάθε Σύνολο Δεδομένων: Κάθε σύνολο δεδομένων υποβλήθηκε σε ένα σύνολο προεπεξεργαστικών διαδικασιών για την προετοιμασία του για την εκπαίδευση των μοντέλων βαθιάς μάθησης. Οι κύριες διαδικασίες προεπεξεργασίας περιλαμβάνουν:

- a. Φόρτωση Συνόλων Δεδομένων: Τα δεδομένα φορτώθηκαν από τις αντίστοιχες πηγές, συμπεριλαμβάνοντας πληροφορίες σχετικά με τη λειτουργία των βιομηχανικών κινητήρων.
  - b. Χειρισμός Ελλειπόντων Δεδομένων: Τα ενδεχόμενα ανεπίττωτα απουσιάζοντα δεδομένα αντιμετωπίστηκαν με τρόπους όπως η απομάκρυνση των σχετικών εγγραφών ή η αντικατάστασή τους με κατάλληλες τιμές (π.χ., μέση τιμή).
  - c. Εξασφάλιση Κατάλληλων Τύπων Δεδομένων: Τα δεδομένα μετατράπηκαν σε κατάλληλους τύπους (π.χ., ακέραιοι αριθμοί, πραγματικοί αριθμοί) για την εκπαίδευση των μοντέλων.
  - d. Εξερεύνηση Δεδομένων για Κάθε Σύνολο Δεδομένων: Πραγματοποιήθηκε λεπτομερής εξερεύνηση των δεδομένων προκειμένου να κατανοηθεί η κατανομή των δεδομένων, οι σχέσεις μεταξύ των μεταβλητών, και η παρουσία ενδεχόμενων ανωμαλιών ή εκτυπώσεων.
  - e. Οπτικοποίηση Δεδομένων με την Πάροδο του Χρόνου: Τα δεδομένα οπτικοποιήθηκαν για να αποτυπώσουν τις αλλαγές στη λειτουργία των βιομηχανικών κινητήρων με την πάροδο του χρόνου, προσφέροντας έτσι βαθύτερη κατανόηση των δεδομένων.
  - f. Υπολογισμός Βασικών Στατιστικών: Για κάθε σύνολο δεδομένων, υπολογίστηκαν βασικές στατιστικές μετρικές όπως ο μέσος όρος, η διακύμανση και τα ποσοστιαία διαγράμματα προκειμένου να παρέχονται κατανοητές πληροφορίες για τα δεδομένα.
2. Μηχανική Χαρακτηριστικών για Κάθε Σύνολο Δεδομένων: Κάθε σύνολο δεδομένων υποβλήθηκε σε διαδικασίες εξαγωγής χαρακτηριστικών (feature engineering) προκειμένου να επιλεγούν και να δημιουργηθούν χαρακτηριστικά που θα χρησιμοποιηθούν κατά την εκπαίδευση των μοντέλων.
  3. Ενοποίηση Δεδομένων για Βελτιωμένη Ανάλυση: Στη συνέχεια, ενοποιήσαμε τα δεδομένα από τα διάφορα σύνολα δεδομένων προκειμένου να δημιουργήσουμε ένα ενιαίο σύνολο δεδομένων που θα χρησιμοποιηθεί για την εκπαίδευση και την αξιολόγηση των μοντέλων.
  4. Διαχωρισμός Δεδομένων για Αξιολόγηση Μοντέλου: Τα δεδομένα διαχωρίστηκαν σε σύνολα εκπαίδευσης, επικύρωσης (validation), και δοκιμής (test) προκειμένου να αξιολογηθεί η απόδοση των μοντέλων βαθιάς μάθησης.

5. Βελτίωση Δεδομένων για Μοντελοποίηση: Κατά τη διάρκεια της εκπαίδευσης των μοντέλων, προσαρμόσαμε τις διάφορες παραμέτρους και υπερπαραμέτρους προκειμένου να βελτιώσουμε την απόδοσή τους και να εξασφαλίσουμε αξιόπιστες προβλέψεις.
6. Εκπαίδευση Μοντέλων Βαθιάς Μάθησης: Τα μοντέλα βαθιάς μάθησης, συμπεριλαμβανομένων των αναφερθέντων LSTM, GRU, CNN και FNN, εκπαιδεύτηκαν χρησιμοποιώντας τα προεπεξεργασμένα δεδομένα και την ενοποιημένη συλλογή δεδομένων.

Με αυτήν τη μεθοδολογία εκπαίδευσης, αναμένουμε να επιτύχουμε αξιόπιστες προβλέψεις για την προβλεπτική συντήρηση των βιομηχανικών κινητήρων και να αξιολογήσουμε την απόδοση των μοντέλων σε πραγματικές συνθήκες.

## Τεχνικό Μέρος

### Προεπεξεργασία δεδομένων για κάθε σύνολο δεδομένων

Αυτή η ενότητα εστιάζει στην αρχική προεπεξεργασία κάθε μεμονωμένου συνόλου δεδομένων. Ο στόχος αυτού του βήματος είναι η προετοιμασία των δεδομένων για επακόλουθη ανάλυση και μοντελοποίηση. Οι βασικές εργασίες που εκτελούνται σε αυτό το βήμα προεπεξεργασίας δεδομένων περιλαμβάνουν:

- **Φόρτωση συνόλων δεδομένων:** Τα σύνολα δεδομένων που χρησιμοποιούνται σε αυτήν την ανάλυση, δηλαδή «big\_fan\_motor», «big\_fan\_right\_bearing» και «big\_fan\_left\_bearing», φορτώνονται στο περιβάλλον χρησιμοποιώντας τη συνάρτηση `pd.read_csv()`. Αυτή η λειτουργία χρησιμοποιείται για την ανάγνωση δεδομένων από αρχεία CSV και την αποθήκευσή τους σε `pandas DataFrames`, διευκολύνοντας την περαιτέρω ανάλυση.
- **Χειρισμός ελλειπόντων δεδομένων:** Μια κρίσιμη πτυχή της προεπεξεργασίας δεδομένων είναι η αντιμετώπιση τιμών που λείπουν, που δηλώνονται ως τιμές NaN (Not-a-Number) ή μηδενικές τιμές. Τα ελλείποντα δεδομένα ενδέχεται να οδηγήσουν σε σφάλματα και μεροληψία στα μοντέλα μηχανικής εκμάθησης. Για να μετριαστεί αυτό το ζήτημα, ο κώδικας ελέγχει συστηματικά για τιμές που λείπουν σε κάθε σύνολο δεδομένων χρησιμοποιώντας τη συνάρτηση `.isnull().sum()`. Τυχόν σειρές που περιέχουν τιμές που λείπουν εξαλείφονται από τα σύνολα δεδομένων χρησιμοποιώντας τη συνάρτηση `.dropna()`, διασφαλίζοντας ότι τα δεδομένα παραμένουν πλήρη και κατάλληλα για ανάλυση.
- **Εξασφάλιση κατάλληλων τύπων δεδομένων:** Η συνέπεια των δεδομένων είναι ζωτικής σημασίας για την αποτελεσματική ανάλυση. Για να διασφαλιστεί ότι οι τύποι δεδομένων των στηλών ευθυγραμμίζονται με την προβλεπόμενη ανάλυση, εφαρμόζονται συγκεκριμένοι μετασχηματισμοί. Συγκεκριμένα, η στήλη 'Ωρα' μετατρέπεται σε μορφή ημερομηνίας, χρησιμοποιώντας τη συνάρτηση `pd.to_datetime()`. Αυτή η μετατροπή είναι ιδιαίτερα σημαντική όταν ασχολούμαστε με δεδομένα χρονοσειρών, καθώς επιτρέπει την ουσιαστική χρονική ανάλυση.

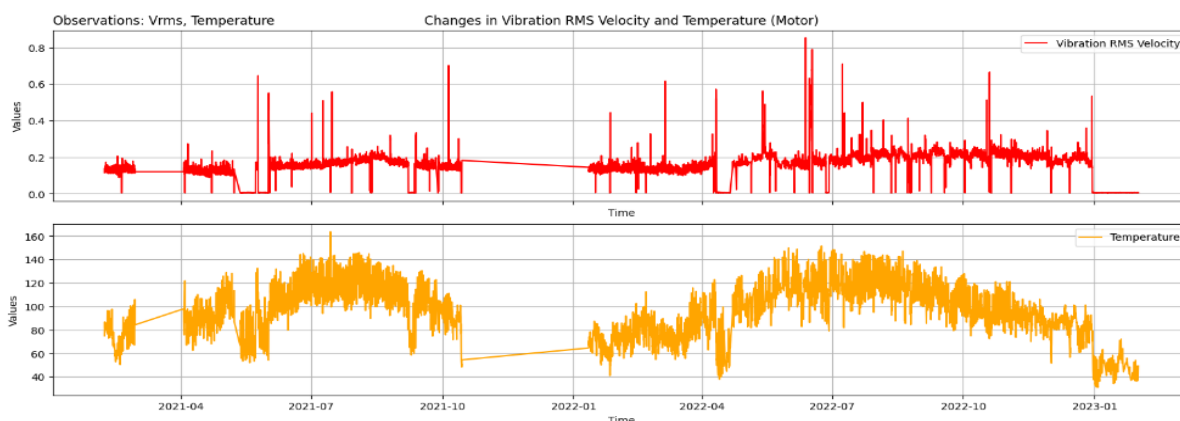
Με την συστηματική εκτέλεση αυτών των βημάτων προεπεξεργασίας για κάθε σύνολο δεδομένων, τα δεδομένα βελτιώνονται και αντιμετωπίζονται αποτελεσματικά πιθανά ζητήματα που σχετίζονται με τιμές και τύπους δεδομένων που λείπουν. Αυτό θέτει μια ισχυρή

βάση για τις επόμενες φάσεις εξερεύνησης δεδομένων, μηχανικής χαρακτηριστικών και ανάπτυξης μοντέλων.

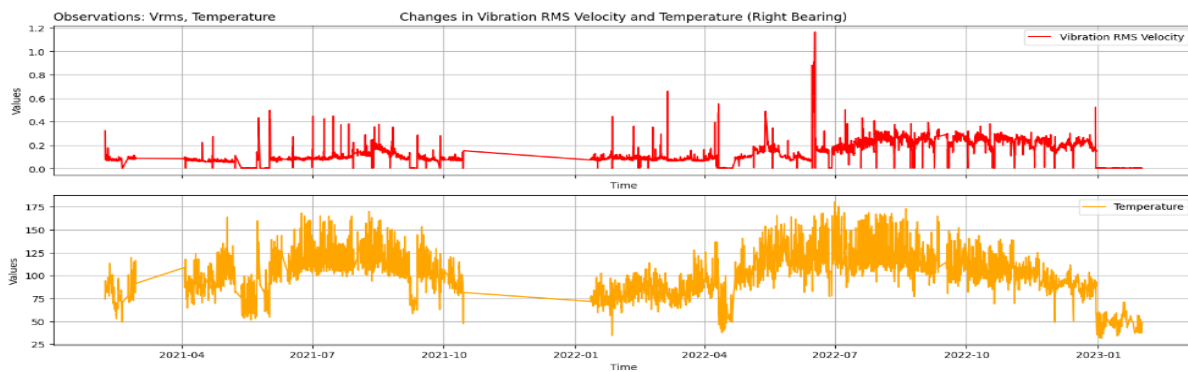
## Εξερεύνηση δεδομένων για κάθε σύνολο δεδομένων

Στη φάση εξερεύνησης δεδομένων, η εστίαση μετατοπίζεται προς την απόκτηση βαθύτερης κατανόησης κάθε συνόλου δεδομένων. Αυτό το κρίσιμο βήμα επιτρέπει στους αναλυτές να αποκαλύπτουν μοτίβα, να ανιχνεύουν ακραίες τιμές και να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την επιλογή χαρακτηριστικών και τη μηχανική. Ακολουθεί μια επισκόπηση των εργασιών που εκτελέστηκαν κατά τη διάρκεια αυτής της φάσης:

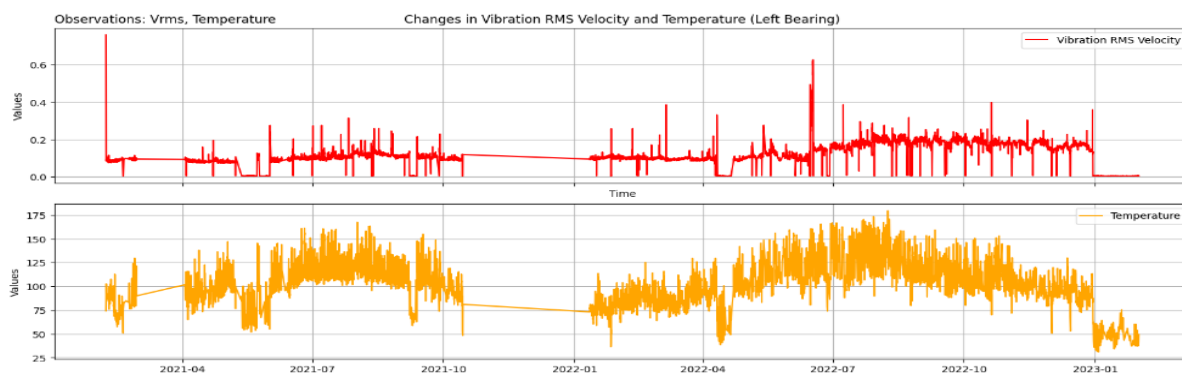
- **Οπτικοποίηση δεδομένων με την πάροδο του χρόνου:** Για καθένα από τα τρία σύνολα δεδομένων, ο κώδικας ξεκινά τη διαδικασία εξερεύνησης οπτικοποιώντας την εξέλιξη συγκεκριμένων μεταβλητών με την πάροδο του χρόνου. Οι μεταβλητές που παρουσιάζουν ιδιαίτερο ενδιαφέρον για την οπτικοποίηση είναι η «Ταχύτητα RMS δόνησης» και η «Θερμοκρασία». Αυτές οι μεταβλητές παίζουν ζωτικό ρόλο στη διάγνωση μηχανικών προβλημάτων και στην παρακολούθηση της απόδοσης του μηχανήματος. Για την οπτικοποίηση των χρονικών αλλαγών, ο κώδικας χρησιμοποιεί τη βιβλιοθήκη Matplotlib για τη δημιουργία ενημερωτικών γραφικών χρονοσειρών.



Εικόνα 11:Αλλαγές στην ταχύτητα και τη θερμοκρασία των κραδασμών RMS (Κινητήρας)



Εικόνα 12: Αλλαγές στην ταχύτητα και τη θερμοκρασία κραδασμών RMS (δεξί ρουλεμάν)

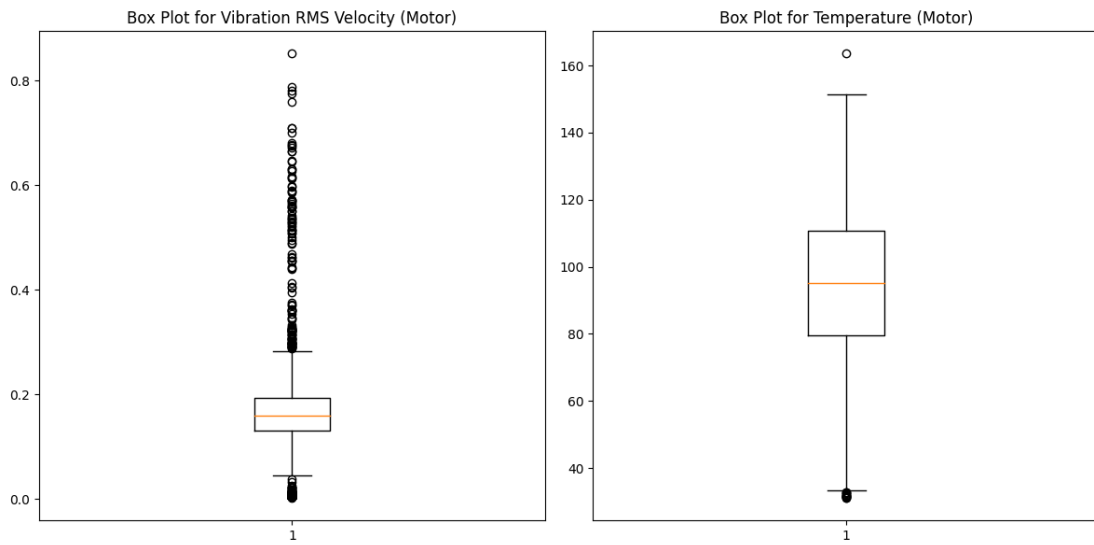


Εικόνα 13: Αλλαγές στην ταχύτητα και τη θερμοκρασία κραδασμών RMS (αριστερό ρουλεμάν)

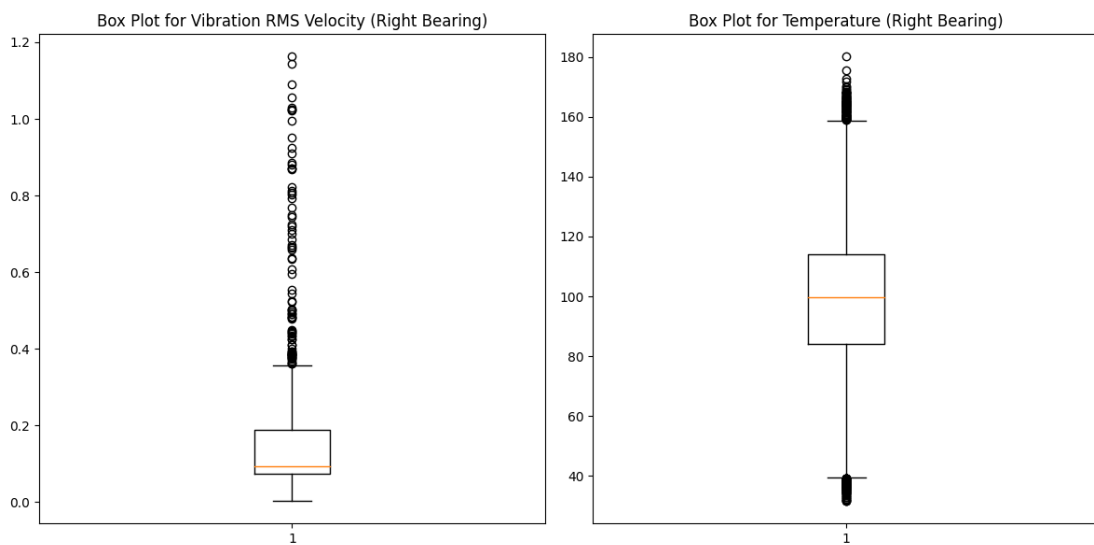
Τα γραφήματα χρονοσειρών που δημιουργούνται παρέχουν μια οπτική αναπαράσταση του τρόπου με τον οποίο αυτές οι μεταβλητές κυμαίνονται με την πάροδο του χρόνου. Αυτή η οπτική εξερεύνηση βοηθά στον εντοπισμό τάσεων, εποχικότητας και πιθανών ανωμαλιών που μπορεί να επηρεάσουν την επακόλουθη ανάλυση.

**Υπολογισμός Βασικών Στατιστικών:** Συμπληρώνοντας την οπτική εξερεύνηση, ο κώδικας υπολογίζει βασικά στατιστικά στοιχεία για βασικές μεταβλητές σε κάθε σύνολο δεδομένων. Αυτά τα στατιστικά στοιχεία περιλαμβάνουν βασικά μέτρα όπως ο μέσος όρος, η τυπική απόκλιση, το ελάχιστο, το μέγιστο και τα τεταρτημόρια. Συγκεκριμένα, οι μεταβλητές «Vibration RMS Velocity» και «Temperature» υποβάλλονται σε στατιστική ανάλυση. Υπολογίζοντας αυτά τα στατιστικά στοιχεία χρησιμοποιώντας τη βιβλιοθήκη Pandas, ο κώδικας παρουσιάζει τα ευρήματα σε οργανωμένη μορφή.

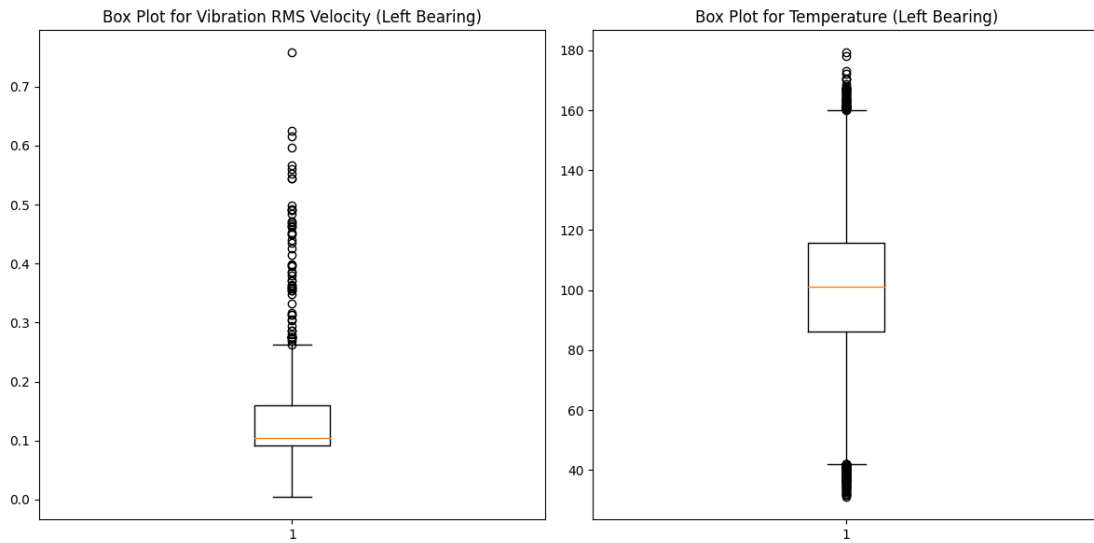




Εικόνα 14: Σχεδίαση box plot της ταχύτητας δόνησης RMS και της θερμοκρασίας του μοτέρ.



Εικόνα 15: Σχεδίαση box plot της ταχύτητας δόνησης RMS και της θερμοκρασίας του δεξιού ρουλεμάν.



Εικόνα 16: Σχεδίαση box plot της ταχύτητας δόνησης RMS και της θερμοκρασίας του αριστερού ρουλεμάν

Η παρουσία ακραίων τιμών, τα σημεία δεδομένων που αποκλίνουν σημαντικά από το αναμενόμενο εύρος, μπορεί να επηρεάσει την απόδοση και την ακρίβεια των μοντέλων μηχανικής εκμάθησης. Για τον εντοπισμό πιθανών ακραίων τιμών, ο κώδικας κατασκευάζει διαγράμματα πλαισίου για τις βασικές μεταβλητές. Τα διαγράμματα πλαισίου (box plots) παρέχουν μια σαφή αναπαράσταση της κατανομής δεδομένων, βοηθώντας στον εντοπισμό ακραίων τιμών. Αυτή η οπτική επιθεώρηση διευκολύνει τον εντοπισμό ακραίων τιμών που μπορεί να απαιτούν ιδιαίτερη προσοχή κατά τα επόμενα στάδια επεξεργασίας και μοντελοποίησης δεδομένων.

Στον Πίνακα 1 παρουσιάζονται τα βασικά στατιστικά στοιχεία της ταχύτητας RMS δόνησης και της θερμοκρασίας του μοτέρ, του δεξιού και του αριστερού ρουλεμάν.

Πίνακας 1: Βασικά στατιστικά στοιχεία ταχύτητας RMS δόνησης και θερμοκρασίας του μοτέρ, του δεξιού και του αριστερού ρουλεμάν.

		Μοτέρ	Δεξίρουλεμάν	Αριστερόρουλεμάν
Ταχύτητα	Καταμέτρηση	12.130	10.571	10.989

RMS δόνησης	Μέσοςόρος	0,153	0,128	0,116
	Τυπική απόκλιση	0,070	0,093	0,059
	Ελάχιστο	0,001	0,002	0,004
	Μέγιστο	0,852	1.1163	0,758
Θερμοκρασία	Καταμέτρηση	12.130	10.571	10.989
	Μέσοςόρος	94,436	98,972	100,757
	Τυπική απόκλιση	22,525	24,566	24,614
	Ελάχιστο	31,0	31,5	30,9
	Μέγιστο	163,6	180,2	179,3

Η διερευνητική φάση παρέχει πολύτιμες πληροφορίες για κάθε σύνολο δεδομένων. Με την οπτικοποίηση των τάσεων δεδομένων, τον υπολογισμό στατιστικών στοιχείων και τον εντοπισμό πιθανών ακραίων τιμών, οι αναλυτές είναι καλύτερα εξοπλισμένοι για να λαμβάνουν τεκμηριωμένες αποφάσεις σε επόμενες φάσεις ανάλυσης και μοντελοποίησης δεδομένων.

Τα διαγράμματα πλαισίου (Σχήματα: «Σχήμα πλαισίου για ταχύτητα δόνησης RMS» και «Θερμοκρασία γραφικού πλαισίου») αντιπροσωπεύουν οπτικά την κατανομή των βασικών μεταβλητών, βοηθώντας στον εντοπισμό ακραίων τιμών και ασυνήθιστων μοτίβων δεδομένων. Αυτά τα ευρήματα είναι απαραίτητα για την προεπεξεργασία δεδομένων και την επιλογή μοντέλου.

Επιπλέον, τα βασικά στατιστικά στοιχεία για κάθε σύνολο δεδομένων προσφέρουν μια σύνοψη των κεντρικών τάσεων και μεταβλητότητας των μεταβλητών «Ταχύτητα RMS δόνησης» και «Θερμοκρασία». Αυτά τα στατιστικά στοιχεία παρέχουν ουσιαστικό πλαίσιο για την κατανόηση των χαρακτηριστικών του συνόλου δεδομένων.

## Μηχανική Χαρακτηριστικών για Κάθε Σύνολο Δεδομένων

Σε αυτό το στάδιο, η εστίαση μετατοπίζεται προς τη μηχανική χαρακτηριστικών, μια διαδικασία απαραίτητη για τη βελτίωση των δεδομένων και τον εμπλουτισμό της χρησιμότητάς της για μοντέλα μηχανικής μάθησης. Η μηχανική χαρακτηριστικών περιλαμβάνει τη δημιουργία νέων χαρακτηριστικών ή τον μετασχηματισμό υπαρχόντων, ενισχύοντας τελικά την ικανότητα ενός μοντέλου να κατανοεί και να προβλέπει μοτίβα. Αυτή η διαδικασία είναι εγγενώς προσαρμοσμένη στα μοναδικά χαρακτηριστικά των δεδομένων, με στόχο να ενδυναμώσει μοντέλα με πλουσιότερες πληροφορίες χωρίς πλεονασμό.

Επιλογή μεγέθους παραθύρου για κυλιόμενες στατιστικές: Στην αρχή, μια βασική παράμετρος, το 'window\_size', μπαίνει στο παιχνίδι. Αυτή η παράμετρος υπαγορεύει το εύρος των κυλιόμενων στατιστικών υπολογισμών. Τα κυλιόμενα στατιστικά περιλαμβάνουν τον υπολογισμό βασικών στατιστικών μετρήσεων, όπως ο μέσος όρος και η τυπική απόκλιση, σε ένα δυναμικά κινούμενο παράθυρο σημείων δεδομένων. Η επιλογή του 'μέγεθος παραθύρου' είναι μια απόφαση με αποχρώσεις, που εξαρτάται από τη χρονική δυναμική του συνόλου δεδομένων και τα συγκεκριμένα μοτίβα που απαιτούν προσοχή. Ένα μεγαλύτερο μέγεθος παραθύρου μπορεί να καταγράφει εκτεταμένες τάσεις, ενώ ένα μικρότερο μέγεθος παραθύρου είναι καλύτερα κατάλληλο για την καταγραφή παροδικών διακυμάνσεων.

Κυλιόμενες πληροφορίες για την ταχύτητα δόνησης RMS: Σε ολόκληρο το σύνολο δεδομένων, ο κώδικας αναλαμβάνει τον υπολογισμό των κυλιόμενων στατιστικών για τη μεταβλητή «Ταχύτητα RMS δόνησης», επιδιώκοντας να αποκαλύψει διαφοροποιημένες πληροφορίες:

- **Κυλιόμενος μέσος όρος:** Ο υπολογισμός του κυλιόμενου μέσου για την «Ταχύτητα RMS δόνησης» ενσωματώνει τις εξελισσόμενες τάσεις και τάσεις που υπάρχουν στα δεδομένα.
- **Κυλιόμενη τυπική απόκλιση:** Παράλληλα, υπολογίζεται η κυλιόμενη τυπική απόκλιση, καταγράφοντας διακυμάνσεις και αποκλίσεις από τη μέση τιμή.

Κυλιόμενες πληροφορίες για τη θερμοκρασία: Ταυτόχρονα, ο κώδικας επεκτείνει το ίδιο σύνολο κυλιόμενων στατιστικών στη μεταβλητή «Θερμοκρασία»:

- **Κυλιόμενος μέσος όρος:** Το κυλιόμενο μέσο για «Θερμοκρασία» χαρακτηρίζει τις σταδιακές μετατοπίσεις και τις περιοδικότητες της θερμοκρασίας.

- **Κυλιόμενη τυπική απόκλιση:** Συμπληρωματική της μέσης τιμής, η τυπική απόκλιση κύλισης παρέχει ορατότητα στις διακυμάνσεις της θερμοκρασίας με την πάροδο του χρόνου.

Αυτή η προσπάθεια μηχανικής χαρακτηριστικών είναι καθοριστική για τη δημιουργία χαρακτηριστικών που ενσωματώνουν τη χρονική δυναμική της «Ταχύτητας RMS δόνησης» και της «Θερμοκρασίας» σε όλο το σύνολο δεδομένων. Με τον υπολογισμό κυλιόμενων στατιστικών χωρίς διαχωρισμό συνόλων δεδομένων, διασφαλίζεται ότι η ουσία της εξέλιξης των δεδομένων αποτυπώνεται πλήρως. Αυτά τα κατασκευασμένα χαρακτηριστικά, χωρίς επανάληψη, δίνουν τη δυνατότητα στα επόμενα μοντέλα μηχανικής εκμάθησης να αποκρυπτογραφούν περίπλοκα μοτίβα δεδομένων και να οδηγούν σε πιο ακριβείς προβλέψεις.

## **Ενοποίηση δεδομένων για βελτιωμένη ανάλυση**

Σε αυτή τη φάση της ροής εργασιών μηχανικής μάθησης, συνδυάζονται τα τρία σύνολα δεδομένων σε ένα ενιαίο.

Η ενοποίηση δεδομένων συνεπάγεται τη διάταξη των σημείων δεδομένων σε ένα χρονολογικό πλαίσιο, όλα ενορχηστρωμένα από μια κοινή χρονική σήμανση. Οι χρονικές σημάνσεις συγχρονίζουν τα δεδομένα από τς τρία επιμέρους σύνολα δεδομένων. Ο κώδικας ταξινομώντας σχολαστικά κάθε σύνολο δεδομένων με βάση τη μοναδική στήλη χρονικής σήμανσης («Ωρα»).

Το σύνολο δεδομένων του μοτέρ ενσωματώνεται σε αύξουσα χρονολογική σειρά, δημιουργώντας ένα απρόσκοπτο χρονικό των γεγονότων. Τα σύνολα δεδομένων του δεξιού και αριστερού ρουλεμάν ακολουθούν το παράδειγμά του.

Με τα τρία σύνολα δεδομένων να είναι πλέον συντονισμένα στο ίδιο χρονικό πλαίσιο, ο κώδικας ξεκινά τη σύντηξη, συγχωνεύοντάς τα σε ένα ολοκληρωμένο σύνολο χρησιμοποιώντας τη στήλη 'Ωρα' ως κατευθυντήριο χαρακτηριστικό. Η συγχώνευση συνόλων δεδομένων μέσω χρονικών σημάνσεων είναι χαρακτηριστικό της αντιμετώπισης δεδομένων χρονοσειρών ή δεδομένων που ξεδιπλώνονται με την πάροδο του χρόνου. Ο κώδικας χρησιμοποιεί τη συνάρτηση `pd.merge_asof()`, γνωστή για την ικανότητά της να ενώνει σύνολα δεδομένων με αντιστοίχιση χρονικών σημάνσεων. Α

Τέλος, για να διασφαλιστεί ότι η σαφήνεια κυριαρχεί, μετονομάζονται οι στήλες μέσα στο συγχωνευμένο σύνολο δεδομένων. Κάθε μεταβλητή λαμβάνει ένα όνομα που είναι ταυτόχρονα μοναδικό και διευκρινιστικό, εξαλείφοντας κάθε ασάφεια και δίνοντας έναν αέρα σαφήνειας και διάκρισης στο ενσωματωμένο σύνολο δεδομένων.

## **Εφαρμογή της μηχανικής χαρακτηριστικών, εξερεύνηση και προεπεξεργασία των δεδομένων των ενισχυμένων συνόλων και του ενσωματωμένου συνόλου δεδομένων**

Σε αυτή τη φάση της ροής εργασιών μηχανικής μάθησης, η εστίαση γίνεται πιο έντονο στην εξερεύνηση και την προεπεξεργασία δεδομένων για το βελτιωμένο σύνολο δεδομένων, που περιλαμβάνει όχι μόνο τα μεμονωμένα σύνολα δεδομένων αλλά και το ενσωματωμένο σύνολο δεδομένων. Αυτή η σχολαστική διαδικασία έχει εξέχουσα σημασία για την αποκάλυψη περιπλοκών δεδομένων, την αντιμετώπιση πιθανών ζητημάτων δεδομένων και την προετοιμασία του συνόλου δεδομένων με στόχο τη μοντελοποίησή τους.

Αρχικά, εντοπίστηκαν και διαχειρίστηκαν οι τιμές που έλειπαν, καθώς τα δεδομένα που λείπουν, εάν δεν αντιμετωπιστούν, μπορούν να ρίξουν μια σκιά στην ευρωστία των αναλύσεων και των μοντέλων μηχανικής μάθησης. Ο κώδικας ενορχηστρώνει τις ακόλουθες στρατηγικές σχετικά με τα δεδομένα που λείπουν:

Σάρωση για τιμές που λείπουν: Πρώτα και κύρια, ο κώδικας χρησιμοποιεί την συνάρτηση `.isnull().sum()`, αποκαλύπτοντας την έκταση των δεδομένων που λείπουν σε κάθε στήλη του ενσωματωμένου δεδομένων. Αυτή η προσεκτική εξέταση ρίχνει φως στις συγκεκριμένες στήλες που περιέχουν κενά δεδομένων.

Λεπτομερής χειρισμός κενών: Οπλισμένος με τη γνώση για το πού λείπουν τα δεδομένα, ο κώδικας επικαλείται τακτικά τη συνάρτηση `.dropna()` για να αφαιρέσει τις σειρές που βρίσκονται στη βάση αυτών των αόριστων τιμών. Ωστόσο, είναι σημαντικό να υπογραμμιστεί ότι αυτή η απόφαση λαμβάνεται μετά από μια συνετή αξιολόγηση του αντίκτυπου των ελλιπών δεδομένων στην ακρίβεια του συνόλου δεδομένων.

Εμβάθυνση στην ουσία των δεδομένων μέσω της στατιστικής: Η κατανόηση της ουσίας του συνόλου δεδομένων είναι επιτακτική για την αποτελεσματική χρήση μοντέλων μηχανικής

μάθησης. Ο κώδικας αναλαμβάνει ένα στατιστικό ταξίδι χρησιμοποιώντας τη συνάρτηση `.describe()`, χρησιμοποιώντας το στο ενσωματωμένο σύνολο δεδομένων. Αυτό το ταξίδι αριθμητικής εξερεύνησης ξεδιπλώνει βασικά στατιστικά στοιχεία: τη μέση, τυπική απόκλιση, ελάχιστο, 25ο εκατοστημόριο, διάμεσο (50ο εκατοστημόριο), 75ο εκατοστημόριο και τις περίφημες μέγιστες τιμές για κάθε αριθμητική στήλη. Αυτά τα στατιστικά στοιχεία χρησιμεύουν ως πυξίδα που καθοδηγεί την κατανόηση των κεντρικών τάσεων και επισημαίνει την παρουσία ακραίων στοιχείων.

Ο Πίνακας 2 παρουσιάζει τον αριθμό των τιμών που λείπουν σε κάθε μεταβλητή των συνόλων δεδομένων που αφορούν το μοτερ, το δεξί και το αριστερό ρουλεμάν.

Πίνακας 2: Αριθμός τιμών που λείπουν σε κάθε μεταβλητή των συνόλων δεδομένων που αφορούν το μοτερ, το δεξί και το αριστερό ρουλεμάν.

	Μοτέρ	Δεξίρουλεμάν	Αριστερόρουλεμάν
Χρόνος	0	0	0
Ταχύτητα RMS δόνησης	0	0	0
Θερμοκρασία	0	0	0
Μέσος όρος κύλισης ταχύτητας RMS δόνησης	4	4	4
Δόνηση RMS Velocity Rolling Std	4	4	4
Μέσος όρος κυλιόμενης θερμοκρασίας	4	4	4
Θερμοκρασία Rolling Std:	4	4	4

Όσον αφορά τον αριθμό των δεδομένων που λείπουν στο σύνολο δεδομένων που δημιουργήθηκε και ενσωματώνει τα παραπάνω τρία σύνολα δεδομένων, αυτός διαμορφώνεται ως εξής:

- Time: 0
- Motor\_Vibration: 0
- Right\_Bearing\_Temperature: 0

- Vibration RMS Velocity Rolling Mean\_x: 4
- Vibration RMS Velocity Rolling Std\_x: 4
- Temperature Rolling Mean\_x: 4
- Temperature Rolling Std\_x: 4
- Left\_Bearing\_Vibration: 1830
- Left\_Bearing\_Temperature: 1830
- Vibration RMS Velocity Rolling Mean\_y: 1833
- Vibration RMS Velocity Rolling Std\_y: 1833
- Temperature Rolling Mean\_y: 1833
- Temperature Rolling Std\_y: 1833
- Vibration RMS Velocity: 1477
- Temperature: 1477
- Vibration RMS Velocity Rolling Mean: 1480
- Vibration RMS Velocity Rolling Std: 1480
- Temperature Rolling Mean: 1480
- Temperature Rolling Std: 1480

Ο Πίνακας 3 παρουσιάζει τα βασικά στατιστικά στοιχεία των μεταβλητών του μοτέρ, του δεξιού και του αριστερού ρουλεμάν, μετά την εφαρμογή της μηχανικής χαρακτηριστικών και τον καθαρισμό των δεδομένων.

Πίνακας 3: Βασικά στατιστικά στοιχεία των μεταβλητών του μοτέρ, του δεξιού και του αριστερού ρουλεμάν, μετά την εφαρμογή της μηχανικής χαρακτηριστικών και τον καθαρισμό των δεδομένων.

		Μοτέρ	Δεξίρουλεμ άν	Αριστερόρουλεμ άν
Ταχύτητα RMS δόνησης	Αριθμός	12.126	10.567	10.985
	Μέσοςόρος	0,153	0,128	0,116



	Τυπική απόκλιση	0,07	0,093	0,053
	Ελάχιστο	0,001	0,002	0,004
	25°εκατοστημόριο	0,131	0,075	0,092
	Διάμεσος	0,159	0,093	0,104
	75°εκατοστημόριο	0,193	0,188	0,16
	Μέγιστο	0,852	1,163	0,758
Θερμοκρασία	Αριθμός	12.126	10.567	10.985
	Μέσος όρος	94,452	98,992	100,776
	Τυπική απόκλιση	22,512	24,55	24,596
	Ελάχιστο	31,0	31,5	30,9
	25°εκατοστημόριο	79,7	84,1	86,3
	Διάμεσος	95,1	99,8	101,2
	75°εκατοστημόριο	110,8	114,0	115,8
	Μέγιστο	136,6	180,2	179,3
Μέσος όρος κύλισης ταχύτητας RMS δόνησης	Αριθμός	12.126	10.567	10.985
	Μέσος όρος	0,153	0,128	0,116
	Τυπική απόκλιση	0,065	0,089	0,057
	Ελάχιστο	0,0042	0,004	0,004
	25°εκατοστημόριο	0,132	0,077	0,092
	Διάμεσος	0,157	0,093	0,104
	75°εκατοστημόριο	0,193	0,189	0,16

	Μέγιστο	0,763	1.096	0,59
Δόνηση RMS Velocity Rolling Std	Αριθμός	12.126	10.567	10.985
	Μέσοζόρος	0,014	0,013	0,00
	Τυπική απόκλιση	0,025	0,027	0,017
	Ελάχιστο	0,0	0,0	0,0
	25°εκατοστημόριο	0,006	0,003	0,003
	Διάμεσος	0,009	0,006	0,004
	75°εκατοστημόριο	0,013	0,011	0,007
	Μέγιστο	0,364	0,545	0,301
	Μέσοζόροςκυλιόμενηςθερμοκρασίας	Αριθμός	12.126	10.567
Μέσοζόρος		94,446	98,985	100,769
Τυπική απόκλιση		22,0579	23,723	23,759
Ελάχιστο		31,44	32,38	32,08
25°εκατοστημόριο		80,025	84,52	86,92
Διάμεσος		95,26	100,34	101,54
75°εκατοστημόριο		110,9	114,5	116,08
Μέγιστο		145,28	168,56	169,88
Θερμοκρασία Rolling Std	Αριθμός	12.126	10.567	10.985
	Μέσοζόρος	4,129	5,415	5,542
	Τυπική απόκλιση	2,947	4,599	4,532
	Ελάχιστο	0,084	0,071	0,071
	25°εκατοστημόριο	1,929	2,151	2,010
	Διάμεσος	3,414	4,081	4,386

	75°εκατοστημό ριο	5,786	7,233	7,772
	Μέγιστο	23,866	37,641	33,540

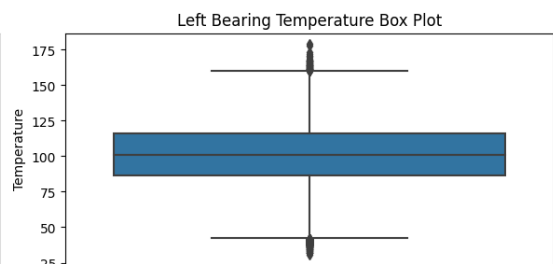
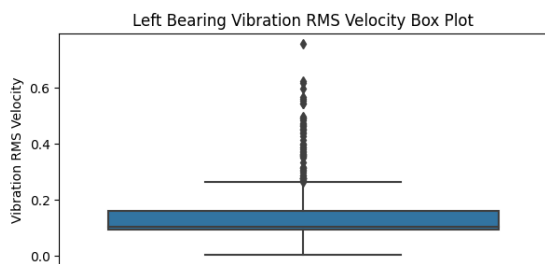
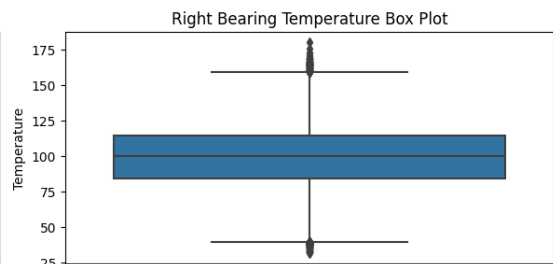
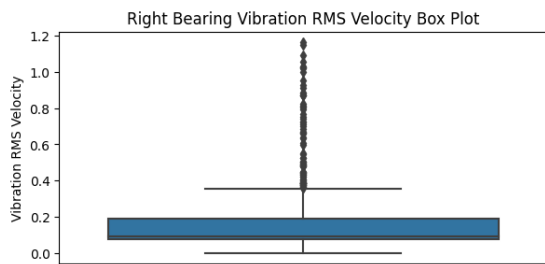
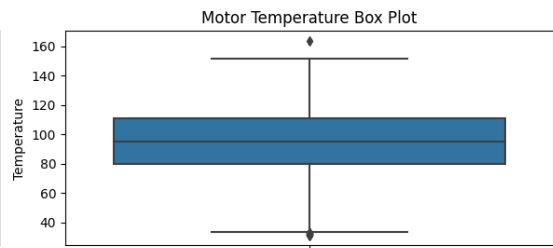
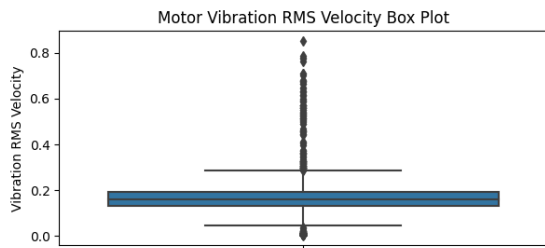
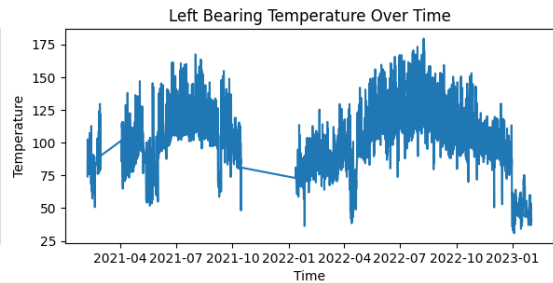
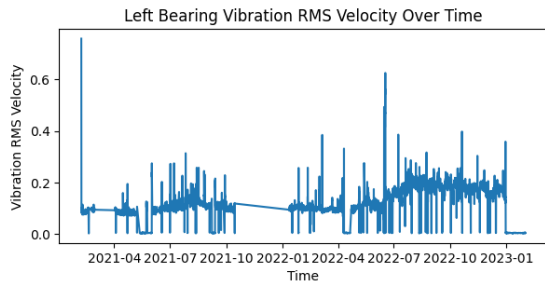
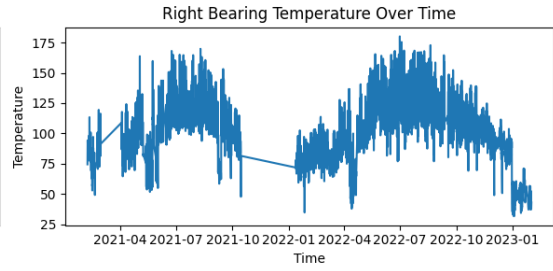
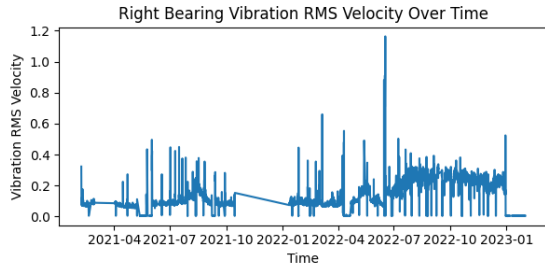
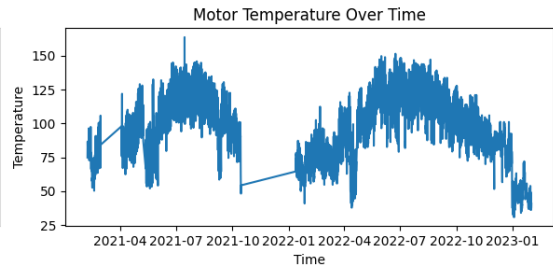
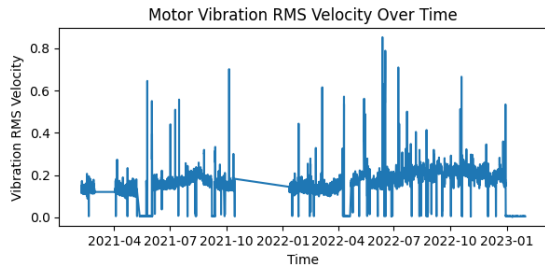
Ο Πίνακας 4 παρουσιάζει τα βασικά στατιστικά στοιχεία των μεταβλητών του ενσωματωμένου δυνόλου δεδομένων, μετά την εφαρμογή της μηχανικής χαρακτηριστικών και τον καθαρισμό των δεδομένων.

Πίνακας 4: Βασικά στατιστικά στοιχεία των μεταβλητών του ενσωματωμένου συνόλου δεδομένων, μετά την εφαρμογή της μηχανικής χαρακτηριστικών και τον καθαρισμό των δεδομένων.

	Αριθμός	M.O.	T.A.	Min	25°εκ.	Διάμεσος	75°εκ.	Max
Motor_Vibration	9.995	0,151	0,071	0,001	0,13	0,156	0,188	0,852
Right_Bearing_Temperature	9.995	94,154	23,030	31,0	78,9	94,9	111,1	151,4
Vibration RMS Velocity Rolling Mean_x	9.995	0,151	0,066	0,004	0,131	0,155	0,188	0,763
Vibration RMS Velocity Rolling Std_x	9.995	0,014	0,025	0,0	0,006	0,009	0,013	0,364
Temperature Rolling Mean_x	9.995	94,116	22,567	31,44	79,1	95,0	111,15	145,28
Temperature Rolling Std_x	9.995	4,024	2,895	0,084	1,864	3,31	5,662	23,866
Left_Bearing_Vibration	9.995	0,127	0,094	0,002	0,075	0,092	0,186	1,163
Left_Bearing_Temperature	9.995	98,549	24,495	31,5	83,7	99,3	113,5	175,5
Vibration RMS	9.995	0,127	0,090	0,004	0,076	0,092	0,187	1,096

Velocity Rolling Mean_y								
Vibration RMS Velocity Rolling Std_y	9.995	0,012	0,027	0,0	0,003	0,006	0,011	0,545
Temperature Rolling Mean_y	9.995	98,529	23,711	32,38	84,12	99,78	114,04	168,56
Temperature Rolling Std_y	9.995	5,248	4,459	0,071	2,101	3,972	6,985	37,64
Vibration RMS Velocity	9.995	0,115	0,06	0,004	0,091	0,103	0,156	0,758
Temperature	9.995	100,29 9	24,815	30,9	85,65	100,6	115,6	179,3
Vibration RMS Velocity Rolling Mean	9.995	0,115	0,057	0,004	0,092	0,103	0,156	0,59
Vibration RMS Velocity Rolling Std	9.995	0,008	0,017	0,0	0,003	0,004	0,007	0,301
Temperature Rolling Mean	9.995	100,20 2	23,979	32,08	86,16	100,82	115,73	169,88
Temperature Rolling Std	9.995	5,361	4,393	0,071	1,937	4,251	7,549	32,403

Η οπτικοποίηση αναδεικνύεται ως ο καλύτερος σύμμαχος στην αποκάλυψη πληροφοριών από δεδομένα. Ο κώδικας αξιοποιεί τη δύναμη της βιβλιοθήκης matplotlib για να δημιουργήσει συναρπαστικές οπτικές αφηγήσεις, συμπεριλαμβανομένης της ευγλωττίας των γραμμικών πλοκών και της διάκρισης των πλοκών:



## Διαχωρισμός δεδομένων για αξιολόγηση μοντέλου

Η διαδικασία μηχανικής μάθησης παίρνει μια ρεαλιστική τροπή, εστιάζοντας στη συνετή διαίρεση του συνόλου δεδομένων σε διακριτά υποσύνολα που προορίζονται για εκπαίδευση, επικύρωση και δοκιμή. Αυτός ο συστηματικός διαχωρισμός είναι μια θεμελιώδης πρακτική που είναι απαραίτητη για την ακριβή αξιολόγηση της απόδοσης του μοντέλου και την προστασία από τους κινδύνους της υπερβολικής τοποθέτησης. Ο κώδικας εννοχηστρώνει αυτήν τη λειτουργία διαχωρισμού δεδομένων και στα τέσσερα σύνολα δεδομένων—σύνολο δεδομένων μοτέρ, σύνολο δεδομένων δεξιού ρουλεμάν, σύνολο δεδομένων αριστερού ρουλεμάν και το ενσωματωμένο σύνολο δεδομένων—με μεθοδική ακρίβεια:

Πριν ξεκινήσει η διαίρεση, είναι επιτακτική ανάγκη να προσδιορίσετε το North Star - τη μεταβλητή στόχο που θα προσπαθήσουν να προβλέψουν τα μοντέλα μηχανικής μάθησης. Στον κώδικα, το North Star ορίζεται αναμφίβολα ως 'Motor\_Vibration', υποδηλώνοντας ότι η αποστολή των μοντέλων είναι να προβλέπουν τα επίπεδα δόνησης κινητήρα. Πιο αναλυτικά, το σύνολο δεδομένων αναλύεται προσεκτικά σε δύο βασικά στοιχεία:

- **Χαρακτηριστικά (X):** Αυτές είναι οι απαραίτητες μεταβλητές εισόδου που χρησιμεύουν ως δομικά στοιχεία για προβλέψεις. Τα χαρακτηριστικά εξάγονται με την επιλεκτική εξαίρεση της μεταβλητής στόχου και τυχόν εξωτερικών στηλών από το σύνολο δεδομένων. Σε αυτόν τον κώδικα, τα χαρακτηριστικά δημιουργούνται χρησιμοποιώντας τη λειτουργία `combinationd_data.drop(columns=['Time', target_variable])`, απομονώνοντας αποτελεσματικά τα προγνωστικά στοιχεία.
- **The North Star (y):** Αυτό το στοιχείο είναι το lodestar που καθοδηγεί τα μοντέλα στο ταξίδι τους. Σε αυτόν τον κώδικα, η μεταβλητή στόχος, `y`, ορίζεται σχολαστικά ως `συνδυασμένα_δεδομένα[target_variable]`, προσδιορίζοντάς την ως τον φάρο που θα προσπαθήσουν να μιμηθούν τα μοντέλα.

Στην συνέχεια, το σύνολο δεδομένων υφίσταται μια υπολογισμένη κατάτμηση, με αποτέλεσμα τρεις διακριτές περιοχές:

- **Train set:** Η μερίδα του λέοντος των δεδομένων (70%) προορίζεται για τα γήπεδα προπόνησης (`X_train` και `y_train`). Μέσα σε αυτούς τους λόγους τα μοντέλα λαμβάνουν την εκπαίδευσή τους, βελτιώνοντας την προγνωστική τους ικανότητα. Ο

κώδικας χρησιμοποιεί τη συνάρτηση `train_test_split` με `test_size=0.3`, κατανέμοντας επιδέξια τα δεδομένα.

- **Validation set:** Μια μικρότερη αλλά όχι λιγότερο σημαντική περιοχή (15%) εμφανίζεται ως όαση επικύρωσης (`X_val` και `y_val`). Εδώ ελέγχονται τα μοντέλα υπό συνθήκες προσομοίωσης πραγματικών σεναρίων και όπου οι υπερπαράμετροι συντονίζονται προσεκτικά. Ο κώδικας τμηματοποιεί περαιτέρω τα δεδομένα εκπαίδευσης χρησιμοποιώντας `train_test_split` με `test_size=0,5`, δημιουργώντας αυτό το κεντρικό έδαφος επικύρωσης.
- **Testing set:** Η υπόλοιπη έκταση (15%) γίνεται το όριο δοκιμής (`X_test` και `y_test`). Στέκεται σε απόσταση, αδιατάρακτη κατά τη διάρκεια της εκπαίδευσης και του συντονισμού μοντέλων, που προορίζεται αποκλειστικά για την αξιολόγηση της ικανότητας των τελικών μοντέλων σε αόρατα δεδομένα. Αυτός ο βασικός διαχωρισμός ξετυλίγεται ταυτόχρονα με τη δημιουργία του συνόλου επικύρωσης.

## Βελτίωση δεδομένων για μοντελοποίηση

Σε αυτή τη φάση της ροής εργασιών μηχανικής μάθησης, η εστίαση είναι στην προεπεξεργασία δεδομένων, ένα ρεαλιστικό βήμα που στοχεύει στη βελτίωση του συνόλου δεδομένων ώστε να το καταστήσει ευνοϊκό για αποτελεσματική εκπαίδευση μοντέλων. Η προεπεξεργασία δεδομένων περιλαμβάνει διάφορες βασικές εργασίες, συμπεριλαμβανομένου του καθαρισμού δεδομένων και της κλιμάκωσης χαρακτηριστικών. Είναι σημαντικό να σημειωθεί ότι αυτά τα βήματα προεπεξεργασίας εφαρμόζονται με συνέπεια και στα τέσσερα σύνολα δεδομένων: μοτέρ, δεξιού ρουλεμάν, αριστερού ρουλεμαν και ενσωματωμένου συνόλου δεδομένων.

Μια βασική πτυχή της προεπεξεργασίας δεδομένων είναι η κλιμάκωση χαρακτηριστικών, και συγκεκριμένα η τυποποίηση. Η τυποποίηση είναι μια μεθοδική τεχνική που χρησιμοποιείται για την επανακλιμάκωση αριθμητικών χαρακτηριστικών, με αποτέλεσμα έναν μέσο όρο 0 και μια τυπική απόκλιση 1 για κάθε χαρακτηριστικό. Αυτή η κανονικοποίηση είναι ιδιαίτερα σημαντική για αλγόριθμους μηχανικής μάθησης που είναι ευαίσθητοι στην κλίμακα των χαρακτηριστικών εισόδου, όπως τα νευρωνικά δίκτυα.

Η μέθοδος μετασχηματισμού χρησιμοποιείται για την κλίμακα των δεδομένων εκπαίδευσης, επικύρωσης και δοκιμής ( $X_{train}$ ,  $X_{val}$  και  $X_{test}$ ) με βάση τον μέσο όρο και την τυπική απόκλιση που προέρχονται από τα δεδομένα εκπαίδευσης. Αυτό διασφαλίζει ότι όλα τα χαρακτηριστικά μοιράζονται μια συνεπή κλίμακα.

Η τυποποίηση ενισχύει τη διαδικασία εκπαίδευσης του μοντέλου διασφαλίζοντας ότι τα χαρακτηριστικά εισόδου βρίσκονται σε ίσους όρους ανταγωνισμού. Αυτό, με τη σειρά του, βοηθά στην επίτευξη καλύτερης σύγκλισης και απόδοσης του μοντέλου.

## Εκπαίδευση Μοντέλων Βαθιάς Μάθησης

Αυτή η ενότητα εμβαθύνει στην κρίσιμη φάση της εκπαίδευσης του μοντέλου στο πλαίσιο της προγνωστικής συντήρησης. Τέσσερα χαρακτηριστικά μοντέλα μηχανικής μάθησης, συγκεκριμένα τα LSTM (Long Short-Term Memory | Μακροπρόθεσμη Μνήμη), GRU (Gated Recurrent Unit | Περιφρασμένη Επαναλαμβανόμενη Μονάδα), CNN (Convolutional Neural Network | Συνελκτικό Νευρωνικό Δίκτυο) και FNN (Νευρωνικό Δίκτυο Feedforward), αξιοποιούνται για το κεντρικό έργο της πρόβλεψης της κατάστασης υγείας του εξαρτήματα μηχανημάτων. Αυτή η περιεκτική ανάλυση στοχεύει να αναλύσει την αποτελεσματικότητα κάθε μοντέλου στον χειρισμό των διαφορετικών συνόλων δεδομένων.

### Μοντέλο LSTM

Η μακροπρόθεσμη μνήμη, μια παραλλαγή των επαναλαμβανόμενων νευρωνικών δικτύων (RNN), διαμορφώνεται σχολαστικά μέσω του Keras. Η αρχιτεκτονική του μοντέλου περιλαμβάνει ένα στρώμα LSTM διακοσμημένο με 64 μονάδες και μια λειτουργία ενεργοποίησης ReLU, ακολουθούμενη από ένα πυκνό στρώμα εξόδου. Αυτό το μοντέλο στη συνέχεια καταρτίζεται χρησιμοποιώντας τον βελτιστοποιητή Adam με μέσο τετραγωνικό σφάλμα (MSE) που χρησιμεύει ως συνάρτηση απώλειας.

### Μοντέλο GRU

Παρόμοια με το μοντέλο LSTM, η αρχιτεκτονική Gated Recurrent Unit (GRU) αντικατοπτρίζει την αντίστοιχή της με 64 μονάδες και μια λειτουργία ενεργοποίησης ReLU.



## Μοντέλο CNN

Ενώ τα Συνελκτικά Νευρωνικά Δίκτυα συσχετίζονται συμβατικά με δεδομένα εικόνας, σε αυτή τη διακριτή περίπτωση, χρησιμοποιείται ένα 1D CNN για να αποκαλύψει εντοπισμένα μοτίβα που βρίσκονται μέσα στα δεδομένα εισόδου. Το μοντέλο CNN διαθέτει ένα συνελκτικό στρώμα που περιέχει 64 φίλτρα, μέγεθος πυρήνα 3 και λειτουργία ενεργοποίησης ReLU. Το διαδέχεται ένα στρώμα max-pooling και ένα πυκνό στρώμα εξόδου.

## Μοντέλο FNN

Το Feedforward Neural Network (FNN) χαρακτηρίζεται από ένα μοναχικό κρυφό στρώμα, που φιλοξενεί 64 μονάδες και μια λειτουργία ενεργοποίησης ReLU. Μετά από αυτό, υπάρχει ένα πυκνό στρώμα εξόδου.

Η ουσία της εκπαίδευσης μοντέλων υλοποιείται μέσω επαναληπτικών βελτιώσεων παραμέτρων που στοχεύουν στην ελαχιστοποίηση της απώλειας, με το μέσο τετράγωνο σφάλμα (MSE) να αποτελεί την κυρίαρχη μέτρηση σε αυτό το πλαίσιο. Ακολουθεί μια βαθιά εξερεύνηση της διαδικασίας εκπαίδευσης μοντέλου:

- **Εποχές και Μέγεθος παρτίδας (batch size)** : Η εκπαίδευση των μοντέλων εκτυλίσσεται σε έναν προκαθορισμένο αριθμό εποχών, οι οποίοι έχουν οριστεί στοχαστικά στις 50 εντός του κώδικα. Κάθε εποχή αποτελεί ένα πλήρες πέρασμα ολόκληρου του συνόλου των δεδομένων εκπαίδευσης μέσω του μοντέλου για διάδοση προς τα εμπρός και προς τα πίσω για ενημέρωση των παραμέτρων. Ένα μέγεθος παρτίδας 32 έχει επιλεγεί με σύνεση, υποδηλώνοντας την ποσότητα των σημείων δεδομένων που υποβάλλονται σε επεξεργασία σε κάθε επανάληψη.
- **Παράμετρος Verbose** : Για τους σκοπούς της σχολαστικής παρακολούθησης της εκπαίδευσης, η παράμετρος περιφρονητική είναι ο φρουρός με μια διαμορφωμένη τιμή 2. Αυτή η έξυπνη διαμόρφωση γεννά ζωτικής σημασίας ενημερώσεις προόδου που περιλαμβάνουν μετρήσεις απώλειας για κάθε εποχή.

## **Αξιολόγηση και σύγκριση μοντέλων**

Με την κορύφωση της φάσης αυστηρής εκπαίδευσης, κάθε μοντέλο υπομένει μια αυστηρή διαδικασία αξιολόγησης στο σύνολο δεδομένων επικύρωσης ( $X_{val}$ ). Το χωνευτήριο της αξιολόγησης χρησιμεύει ως το βαρόμετρο για τη μέτρηση της τάσης του μοντέλου να γενικεύει σε αόρατα δεδομένα. Δύο βασικές μετρήσεις, το μέσο τετράγωνο σφάλμα (MSE) και το μέσο απόλυτο σφάλμα (MAE), υποβάλλονται σε σχολαστικούς υπολογισμούς για κάθε μοντέλο στα δεδομένα επικύρωσης. Αυτές οι μετρήσεις, εγγενείς στον κόσμο της πρόβλεψης, ποσοτικοποιούν ανελλιπώς την ακρίβεια της πρόβλεψης και το μέγεθος των σφαλμάτων.

Μετά από τη συνετά ενορχηστρωμένη εκπαίδευση και την ενδεδειγμένη αξιολόγηση δεδομένων επικύρωσης, προκύπτει μια ένθερμη συγκριτική ανάλυση. Αυτή η σχολαστική αξιολόγηση συνεπάγεται μια διεξοδική εξέταση των μετρήσεων απόδοσης που παρουσιάζονται από κάθε μοντέλο. Αυτό το χωνευτήριο ελέγχου ανοίγει το δρόμο για την αναγνώριση του μοντέλου που ανεβαίνει πάνω από τα αντίστοιχα του. Τα κριτήρια αξιολόγησης, σε αυτόν τον κώδικα, περιλαμβάνουν μια τριάδα βαθμολογιών MSE, MAE και R-squared ( $R^2$ ).

## **Αποτελέσματα μεμονομένων μοντέλων ολοκληρωμένου συνόλου δεδομένων**

Εξετάζοντας την απόδοση κάθε μοντέλου μηχανικής εκμάθησης ξεχωριστά, εστιάζοντας στα τρία ξεχωριστά σύνολα δεδομένων: δεδομένα μοτέρ, δεδομένα δεξιού ρουλεμάν και δεδομένα αριστερού ρουλεμάν. Αυτά τα σύνολα δεδομένων αντιπροσωπεύουν διαφορετικά στοιχεία του μηχανήματος και τα μοντέλα αξιολογούνται ξεχωριστά για καθένα από αυτά. Επιπλέον, το ενσωματωμένο σύνολο δεδομένων, το οποίο συνδυάζει πληροφορίες από όλα τα στοιχεία, αξιολογείται επίσης ανεξάρτητα.

### **Δεδομένα μοτέρ**

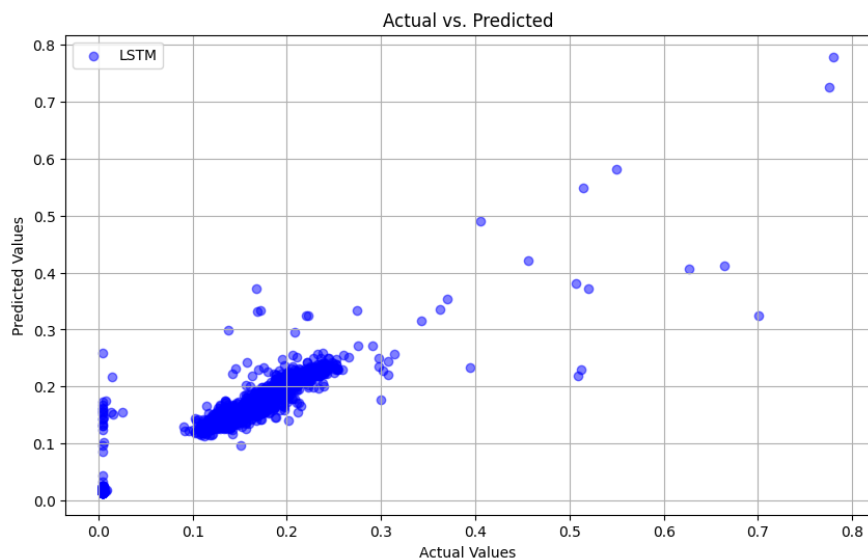
Το σύνολο δεδομένων μοτέρ αντιπροσωπεύει ένα από τα κρίσιμα στοιχεία του μηχανήματος υπό ανάλυση. Αυτό το σύνολο δεδομένων ασχολείται κυρίως με την παρακολούθηση της κατάστασης υγείας και λειτουργίας του κινητήρα. Τα μοντέλα μηχανικής εκμάθησης

εκπαιδεύονται και αξιολογούνται σε αυτό το σύνολο δεδομένων για να κάνουν προβλέψεις που σχετίζονται με την κατάσταση του κινητήρα.

Πίνακας 5: Δεδομένα μοτέρ

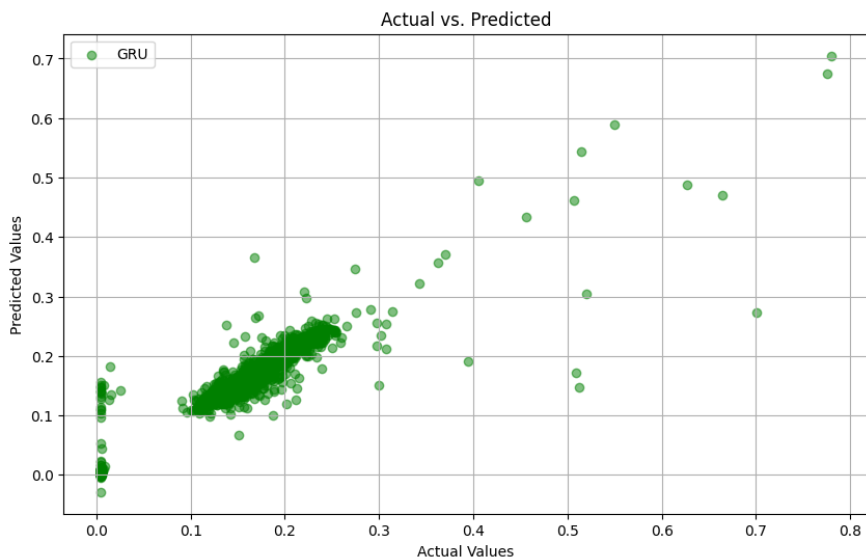
Μοντέλο	MSE	MAE	R <sup>2</sup>
LSTM	0,0009	0,0145	0,8417
GRU	0,0008	0,0122	0,8598
CNN	0,0012	0,0140	0,7790
FNN	0,0008	0,0126	0,8577

Το μοντέλο LSTM εμφανίζει ένα μέτρια χαμηλό MSE, υποδεικνύοντας ότι οι προβλέψεις του είναι γενικά κοντά στις πραγματικές τιμές. Η τιμή MAE 0,0145 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου έχουν απόλυτο σφάλμα 0,0145 μονάδες. Η βαθμολογία R2 0,8417 υποδηλώνει ότι το μοντέλο LSTM αντιπροσωπεύει περίπου το 84,17% της διακύμανσης στα δεδομένα κινητήρα, υποδεικνύοντας ένα καλό επίπεδο επεξηγηματικής ισχύος.



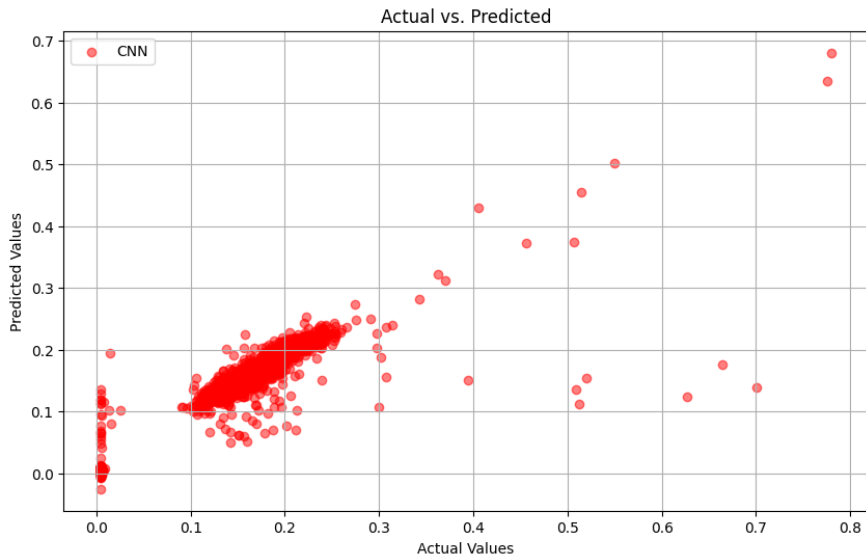
Εικόνα 17: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα κινητήρα.

Το μοντέλο GRU αποδίδει αρκετά καλά στα δεδομένα κινητήρα με ακόμη χαμηλότερο MSE και MAE από το μοντέλο LSTM. Το MSE 0,0008 υποδεικνύει ότι τα τετράγωνα σφάλματα πρόβλεψης του μοντέλου είναι ελάχιστα. Το MAE 0,0122 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0122 μονάδες από τις πραγματικές τιμές. Η υψηλή βαθμολογία R2 0,8598 σημαίνει ότι το μοντέλο GRU καταγράφει ένα σημαντικό μέρος (περίπου 85,98%) της διακύμανσης των δεδομένων.



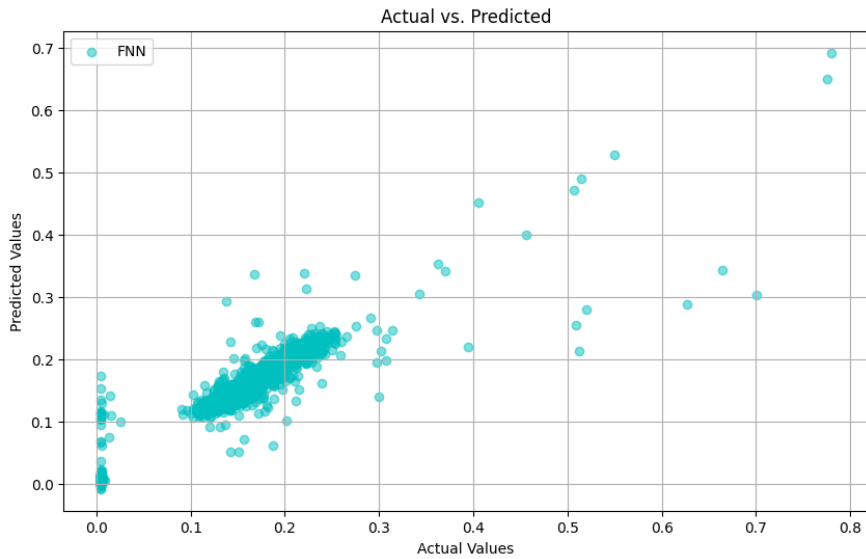
Εικόνα 18:Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα κινητήρα.

Το μοντέλο CNN, ενώ εξακολουθεί να παρέχει λογικές προβλέψεις, παρουσιάζει ελαφρώς υψηλότερο MSE και MAE σε σύγκριση με τα μοντέλα LSTM και GRU. Το MSE 0,0012 δείχνει ότι οι προβλέψεις του μοντέλου έχουν ελαφρώς μεγαλύτερα τετράγωνα σφάλματα. Το MAE 0,0140 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν κατά 0,0140 μονάδες από τις πραγματικές τιμές. Η βαθμολογία R2 0,7790 υποδηλώνει ότι το μοντέλο CNN εξηγεί περίπου το 77,90% της διακύμανσης στα δεδομένα κινητήρα.



Εικόνα 19: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα κινητήρα.

Το μοντέλο FNN, όπως και το μοντέλο GRU, εμφανίζει χαμηλό MSE και MAE. Το MSE 0,0008 υποδεικνύει ότι τα τετράγωνα σφάλματα πρόβλεψης του μοντέλου είναι ελάχιστα, παρόμοια με το μοντέλο GRU. Το MAE 0,0126 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0126 μονάδες από τις πραγματικές τιμές. Η βαθμολογία R2 0,8577 σημαίνει ότι το μοντέλο FNN καταγράφει ένα σημαντικό μέρος (περίπου 85,77%) της διακύμανσης στα δεδομένα κινητήρα.



Εικόνα 20: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα κινητήρα.

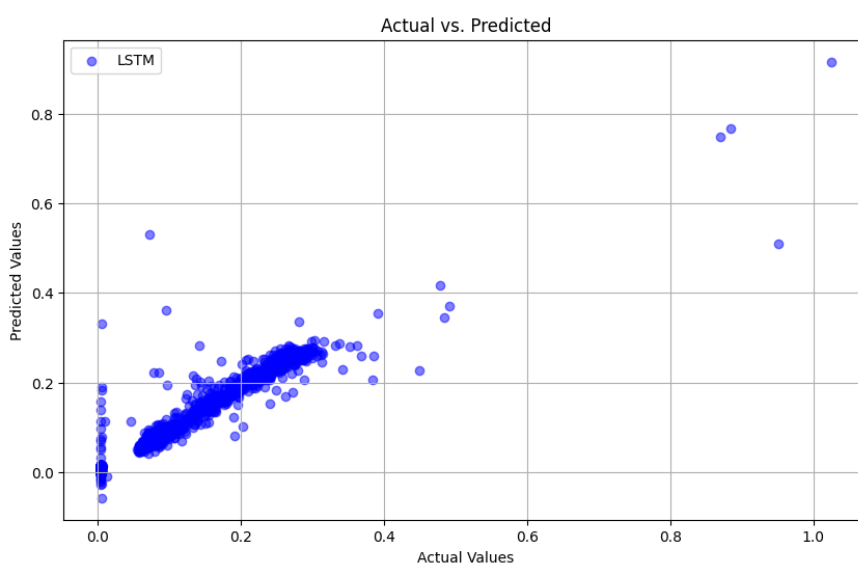
### Δεδομένα δεξιού ρουλεμάν

Το σύνολο δεδομένων δεξιού ρουλεμάν εστιάζει στην παρακολούθηση της κατάστασης και της απόδοσης του σωστού εξαρτήματος ρουλεμάν εντός του μηχανήματος. Κάθε μοντέλο μηχανικής εκμάθησης εκπαιδεύεται και αξιολογείται σε αυτό το σύνολο δεδομένων για τη δημιουργία προβλέψεων που σχετίζονται με την υγεία και τη λειτουργική κατάσταση του σωστού ρουλεμάν.

Πίνακας 6: Δεδομένα δεξιού ρουλεμάν

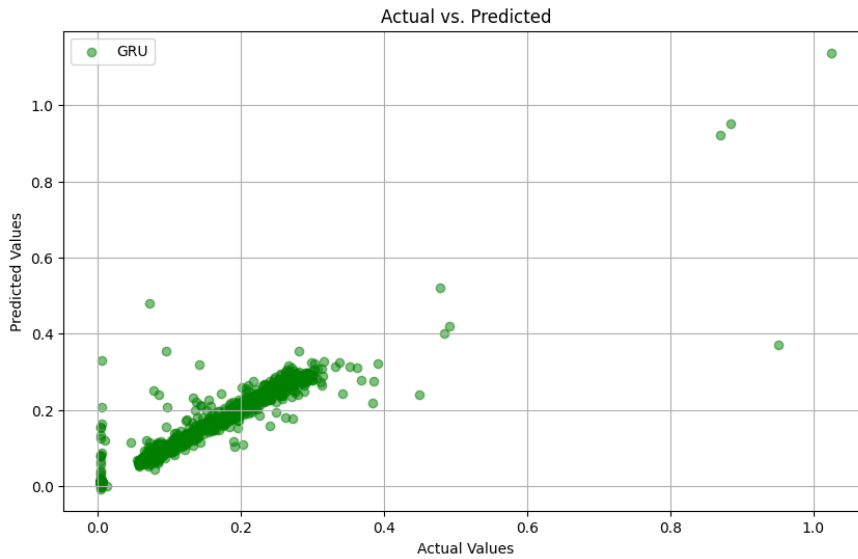
Μοντέλο	MSE	MAE	R <sup>2</sup>
LSTM	0,0009	0,0137	0,8982
GRU	0,0009	0,0119	0,8984
CNN	0,0008	0,0124	0,9094
FNN	0,0015	0,0229	0,8293

Το μοντέλο LSTM επιτυγχάνει αξιόπαινα αποτελέσματα όταν εφαρμόζεται στα Δεδομένα Δεξιάς Έδρασης. Εμφανίζει ένα μέτρια χαμηλό MSE, που σημαίνει ότι οι προβλέψεις του ευθυγραμμίζονται στενά με τις πραγματικές τιμές. Η τιμή MAE 0,0137 υποδεικνύει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0137 μονάδες από τις πραγματικές τιμές. Η υψηλή βαθμολογία R2 0,8982 υπογραμμίζει την ικανότητα του μοντέλου να αντιπροσωπεύει περίπου το 89,82% της διακύμανσης στα Δεδομένα Δεξιάς Έδρασης, υποδεικνύοντας ισχυρή προγνωστική απόδοση.



Εικόνα 21: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα δεξιού ρουλεμάν.

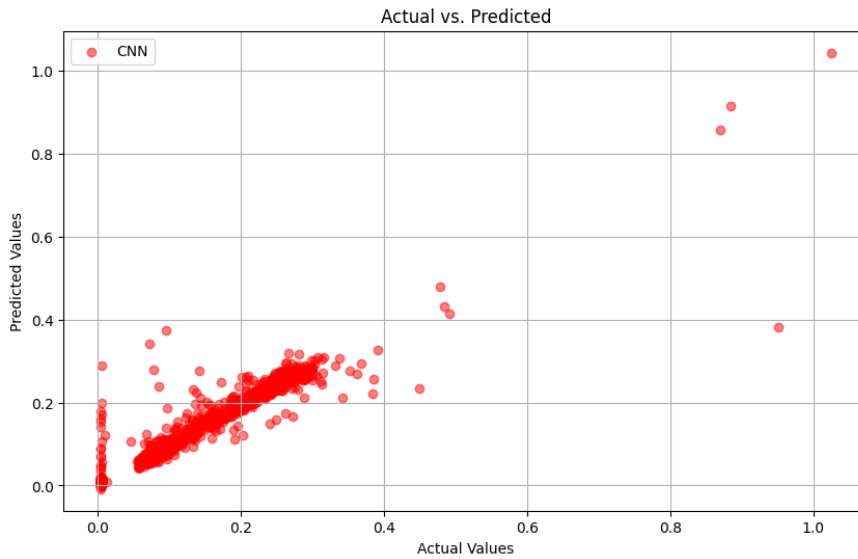
Το μοντέλο GRU στο Right Bearing Data ανταγωνίζεται στενά το μοντέλο LSTM όσον αφορά την απόδοση. Εμφανίζει χαμηλό MSE, υπονοώντας ότι τα τετράγωνα σφάλματα πρόβλεψής του είναι ελάχιστα. Η τιμή MAE 0,0119 υποδεικνύει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0119 μονάδες από τις πραγματικές τιμές, επιδεικνύοντας υψηλή ακρίβεια. Η εξαιρετικά υψηλή βαθμολογία R2 0,8984 υπογραμμίζει την αξιοσημείωτη ικανότητα του μοντέλου να εξηγεί περίπου το 89,84% της διακύμανσης στα Δεδομένα Δεξιάς Έδρασης, υπογραμμίζοντας τις ισχυρές προγνωστικές του ικανότητες.



Εικόνα 22: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα δεξιού ρουλεμάν.

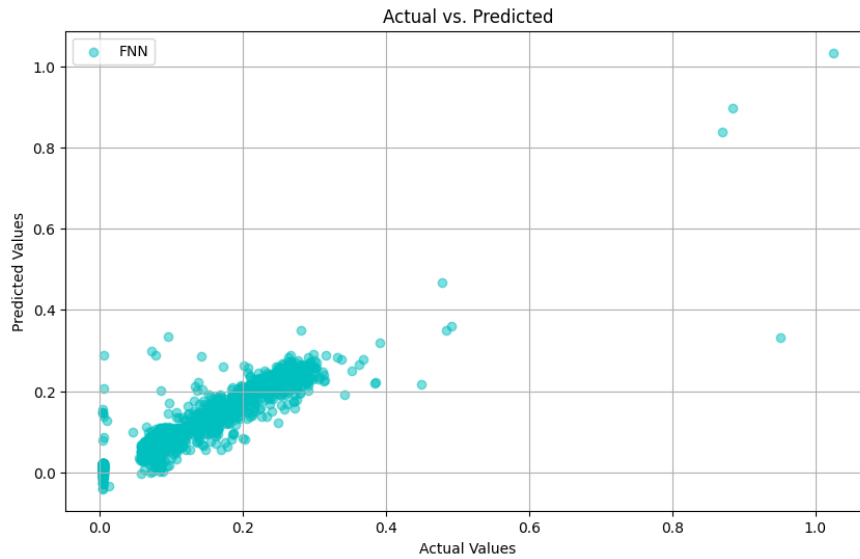
Το μοντέλο του CNN αποδεικνύεται ότι έχει ισχυρές επιδόσεις στο πλαίσιο των Δεδομένων Σωστά Φέροντα. Διαθέτει χαμηλό MSE, υποδεικνύοντας ότι οι προβλέψεις του έχουν ελάχιστα τετράγωνα σφάλματα. Η τιμή MAE 0,0124 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0124 μονάδες από τις πραγματικές τιμές, επιδεικνύοντας υψηλή ακρίβεια. Η εξαιρετική βαθμολογία R2 0,9040 υποδηλώνει την εξαιρετική ικανότητα του μοντέλου να εξηγεί περίπου το 90,40% της διακύμανσης στα Δεδομένα Δεξιάς Έδρασης, υπογραμμίζοντας την αποτελεσματικότητά του στην καταγραφή προγνωστικών μοτίβων.





Εικόνα 23: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα δεξιού ρουλεμάν.

Το μοντέλο FNN, ενώ εξακολουθεί να παρέχει αρκετά καλές προβλέψεις, παρουσιάζει ελαφρώς υψηλότερο MSE και MAE σε σύγκριση με τα μοντέλα LSTM, GRU και CNN. Το MSE 0,0015 δείχνει ότι τα τετράγωνα σφάλματα πρόβλεψης του μοντέλου είναι κάπως μεγαλύτερα. Το MAE 0,0229 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν κατά 0,0229 μονάδες από τις πραγματικές τιμές. Η βαθμολογία R2 0,8293 αντικατοπτρίζει την ικανότητα του μοντέλου να εξηγεί περίπου το 82,93% της διακύμανσης στα Δεδομένα Δεδομένων Δεξιάς, υποδηλώνοντας ένα αξιοπρεπές επίπεδο προγνωστικής απόδοσης.



Εικόνα 24: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα δεξιού ρουλεμάν.

### Δεδομένα αριστερού ρουλεμάν

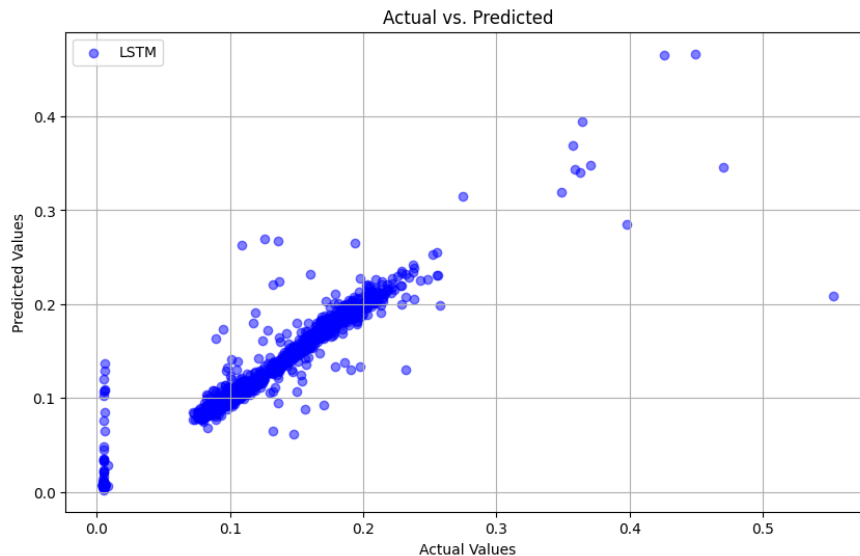
Το σύνολο δεδομένων αριστερού ρουλεμάν επικεντρώνεται στην παρακολούθηση της κατάστασης και της απόδοσης του αριστερού εξαρτήματος ρουλεμάν εντός του μηχανήματος. Κάθε μοντέλο μηχανικής εκμάθησης εκπαιδεύεται και αξιολογείται σε αυτό το σύνολο δεδομένων για τη δημιουργία προβλέψεων σχετικά με την υγεία και την κατάσταση λειτουργίας του αριστερού ρουλεμάν.

Πίνακας 8: Δεδομένα αριστερού ρουλεμάν

Μοντέλο	MSE	MAE	R <sup>2</sup>
LSTM	0,0003	0,0067	0,9195
GRU	0,0003	0,0075	0,9327
CNN	0,0003	0,0093	0,9087
FNN	0,0003	0,0094	0,9135

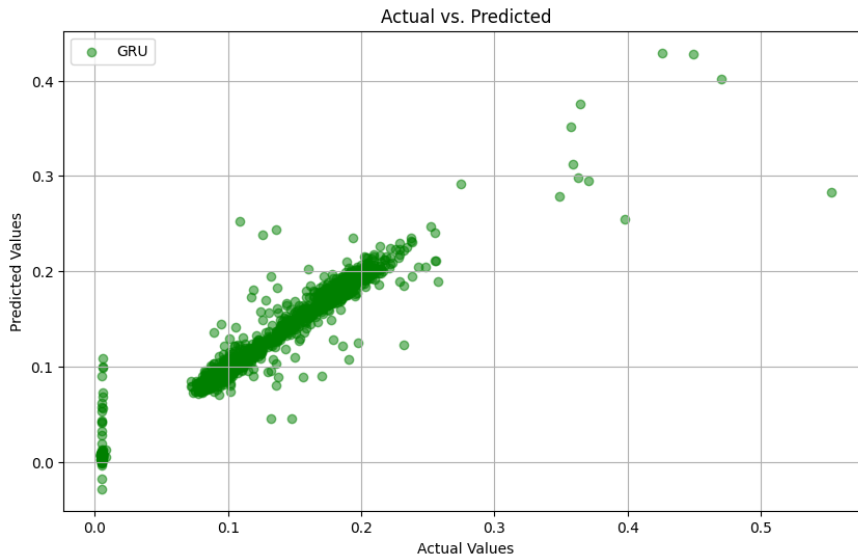
Το μοντέλο LSTM παρουσιάζει εξαιρετικές επιδόσεις στα δεδομένα του αριστερού ρουλεμάν. Δείχνει εξαιρετικά χαμηλό MSE, υποδηλώνοντας ότι οι προβλέψεις του είναι εντυπωσιακά

κοντά στις πραγματικές τιμές. Η τιμή MAE 0,0067 δείχνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν κατά ένα εντυπωσιακά χαμηλό 0,0067 μονάδες από τις πραγματικές τιμές. Η υψηλή βαθμολογία R2 0,9195 υπογραμμίζει την ικανότητα του μοντέλου να αντιπροσωπεύει περίπου το 91,95% της διακύμανσης στα Δεδομένα Αριστερής Έδρασης, υπογραμμίζοντας τις ισχυρές προγνωστικές του ικανότητες.



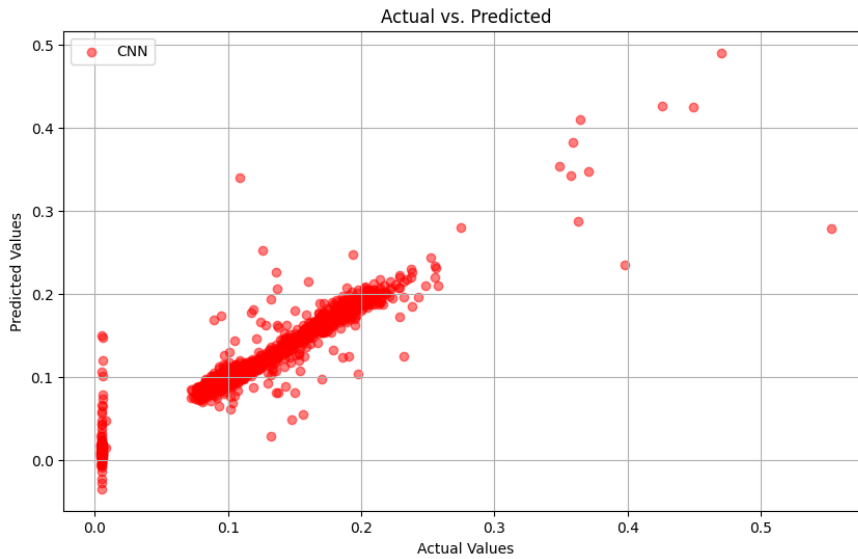
Εικόνα 25: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα αριστερού ρουλεμάν.

Το μοντέλο GRU παρουσιάζει επίσης ισχυρή απόδοση όταν εφαρμόζεται στα Δεδομένα Αριστερής Ρουλεμάν. Εμφανίζει χαμηλό MSE, υποδεικνύοντας ότι τα τετράγωνα σφάλματα πρόβλεψής του είναι ελάχιστα. Η τιμή MAE 0,0075 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν κατά μόλις 0,0075 μονάδες από τις πραγματικές τιμές, επιδεικνύοντας υψηλή ακρίβεια. Η εξαιρετικά υψηλή βαθμολογία R2 0,9327 υπογραμμίζει την εξαιρετική ικανότητα του μοντέλου να εξηγή περίπου το 93,27% της διακύμανσης στα Δεδομένα Αριστερής Ρουλεμάν, υπογραμμίζοντας τις ισχυρές προγνωστικές του ικανότητες.



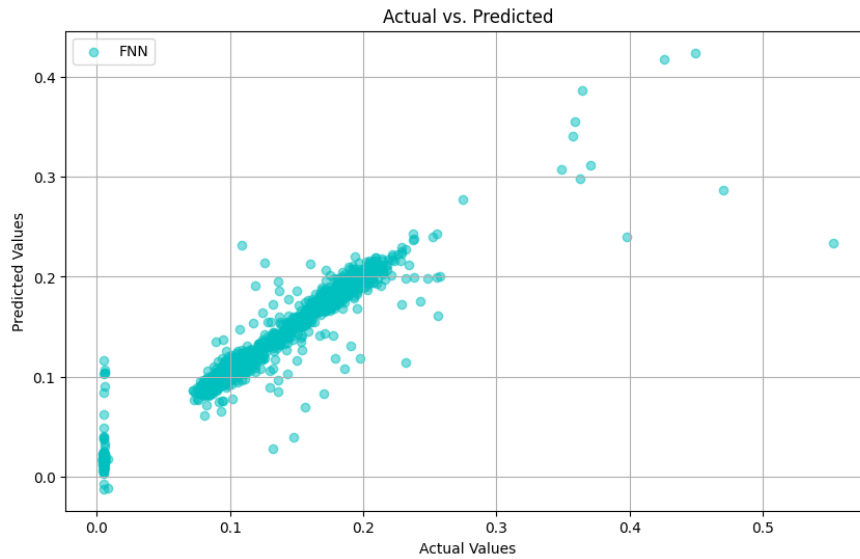
Εικόνα 26: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα αριστερού ρουλεμάν.

Το μοντέλο του CNN αποδίδει αξιοθαύμαστα στο πλαίσιο του Left Bearing Data. Επιτυγχάνει χαμηλό MSE, υποδεικνύοντας ότι οι προβλέψεις του χαρακτηρίζονται από ελάχιστα τετράγωνα σφάλματα. Η τιμή MAE 0,0093 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0093 μονάδες από τις πραγματικές τιμές, επιδεικνύοντας υψηλή ακρίβεια. Η ισχυρή βαθμολογία R2 0,9087 αντικατοπτρίζει την εντυπωσιακή ικανότητα του μοντέλου να εξηγεί περίπου το 90,87% της διακύμανσης στα Δεδομένα Αριστερής Ρουλεμάν, υπογραμμίζοντας την αποτελεσματικότητά του στην καταγραφή προγνωστικών μοτίβων.



Εικόνα 27: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα αριστερού ρουλεμάν.

Το μοντέλο FNN, ενώ εξακολουθεί να παρέχει ισχυρές προβλέψεις, εμφανίζει ελαφρώς υψηλότερες τιμές MSE και MAE σε σύγκριση με τα μοντέλα LSTM, GRU και CNN. Το MSE 0,0003 υποδεικνύει ότι τα τετράγωνα σφάλματα πρόβλεψης του μοντέλου είναι κάπως μεγαλύτερα. Το MAE 0,0094 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν κατά 0,0094 μονάδες από τις πραγματικές τιμές. Ωστόσο, το μοντέλο FNN υπερέρχει με υψηλή βαθμολογία R2 0,9135, υποδηλώνοντας την ουσιαστική του ικανότητα να εξηγεί περίπου το 91,35% της διακύμανσης στα Δεδομένα Αριστερής Έδρασης, υποδεικνύοντας ισχυρή προγνωστική απόδοση.



Εικόνα 28: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα αριστερού ρουλεμάν.

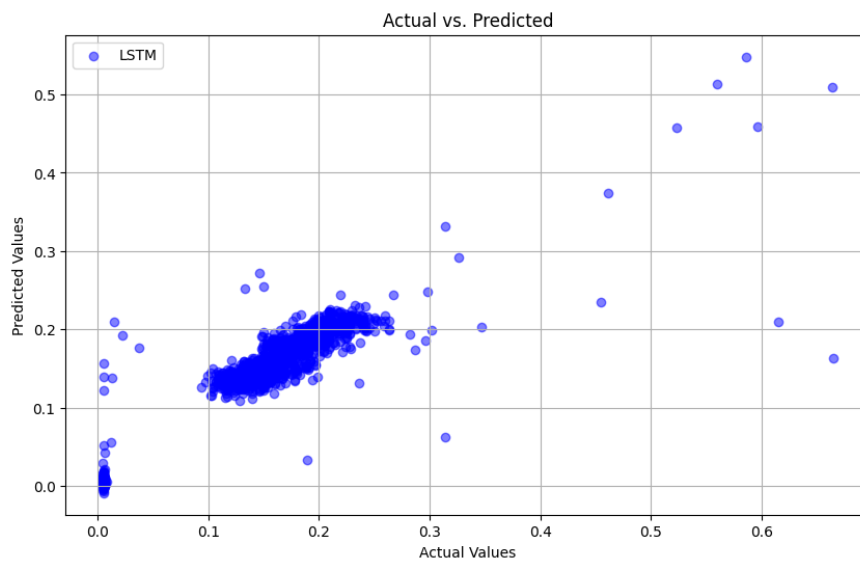
### Ενσωματωμένο σύνολο δεδομένων

Το ενσωματωμένο σύνολο δεδομένων είναι ένα ενοποιημένο σύνολο δεδομένων που συνδυάζει πληροφορίες από τα στοιχεία του κινητήρα, του δεξιού ρουλεμάν και του αριστερού ρουλεμάν εντός του μηχανήματος. Κάθε μοντέλο μηχανικής εκμάθησης εκπαιδεύεται και αξιολογείται σε αυτό το ολοκληρωμένο σύνολο δεδομένων για τη δημιουργία προβλέψεων που σχετίζονται με τη συνολική κατάσταση της υγείας και της λειτουργίας του μηχανήματος.

Πίνακας 7: Ενσωματωμένο σύνολο δεδομένων

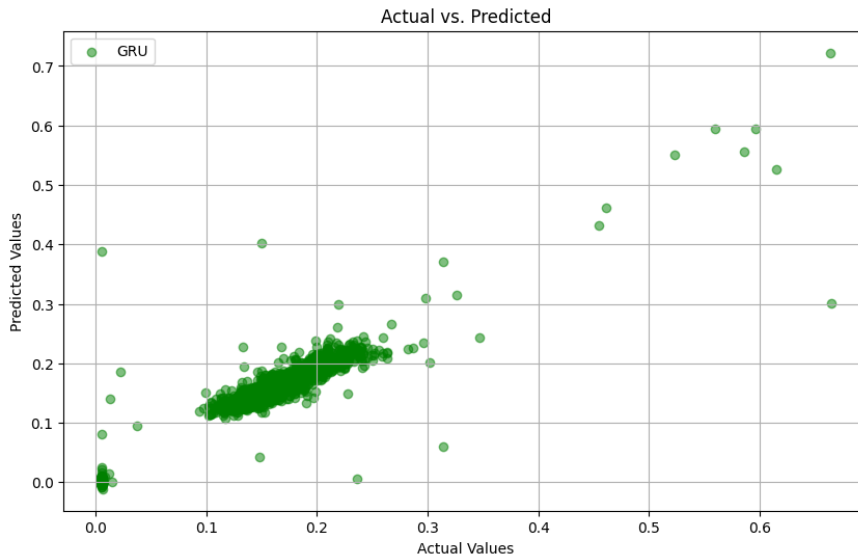
Μοντέλο	MSE	MAE	R <sup>2</sup>
LSTM	0,0008	0,0148	0,8329
GRU	0,0006	0,0122	0,8845
CNN	0,0004	0,0114	0,9200
FNN	0,0005	0,0123	0,9025

Το μοντέλο LSTM, όταν εφαρμόζεται στο ολοκληρωμένο σύνολο δεδομένων, επιδεικνύει σταθερή απόδοση. Πετυχαίνει ένα αρκετά χαμηλό MSE, υποδεικνύοντας ότι οι προβλέψεις του είναι κοντά στις πραγματικές τιμές. Η τιμή MAE 0,0148 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν κατά 0,0148 μονάδες από τις πραγματικές τιμές. Η βαθμολογία R2 0,8329 αντικατοπτρίζει την ικανότητα του μοντέλου να εξηγεί περίπου το 83,29% της διακύμανσης στο ολοκληρωμένο σύνολο δεδομένων, υποδεικνύοντας την ικανότητά του να καταγράφει σημαντικά μοτίβα πρόβλεψης.



Εικόνα 29: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το LSTM στα δεδομένα του ενσωματωμένου συνόλου.

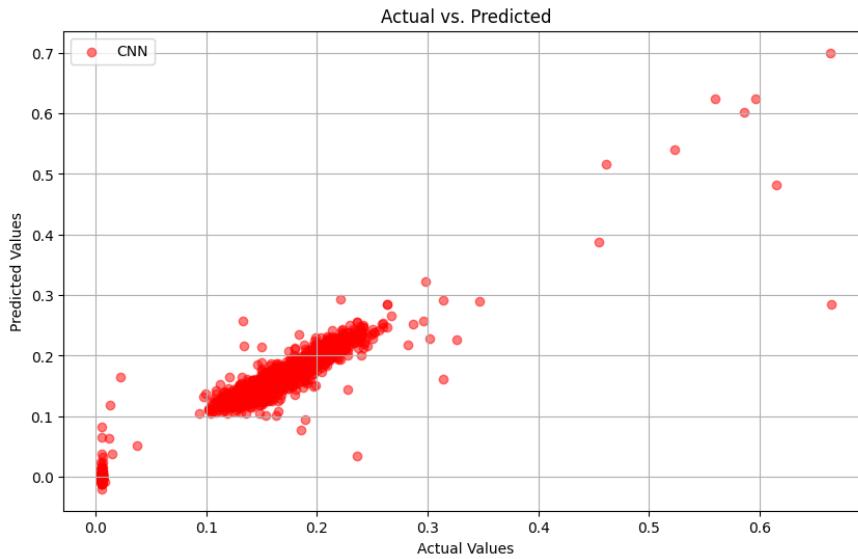
Το μοντέλο GRU παρουσιάζει ισχυρή απόδοση όταν εφαρμόζεται στο ολοκληρωμένο σύνολο δεδομένων. Παρέχει χαμηλό MSE, υποδηλώνοντας ελάχιστα τετράγωνα σφάλματα πρόβλεψης. Η τιμή MAE 0,0122 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0122 μονάδες από τις πραγματικές τιμές, επιδεικνύοντας υψηλή ακρίβεια. Η αξιοσημείωτη βαθμολογία R2 0,8845 υπογραμμίζει την ισχυρή ικανότητα του μοντέλου να εξηγεί περίπου το 88,45% της διακύμανσης στο Ολοκληρωμένο σύνολο δεδομένων, υπογραμμίζοντας την αποτελεσματικότητά του στην καταγραφή σχετικών προγνωστικών μοτίβων.



Εικόνα 30: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το GRU στα δεδομένα του ενσωματωμένου συνόλου.

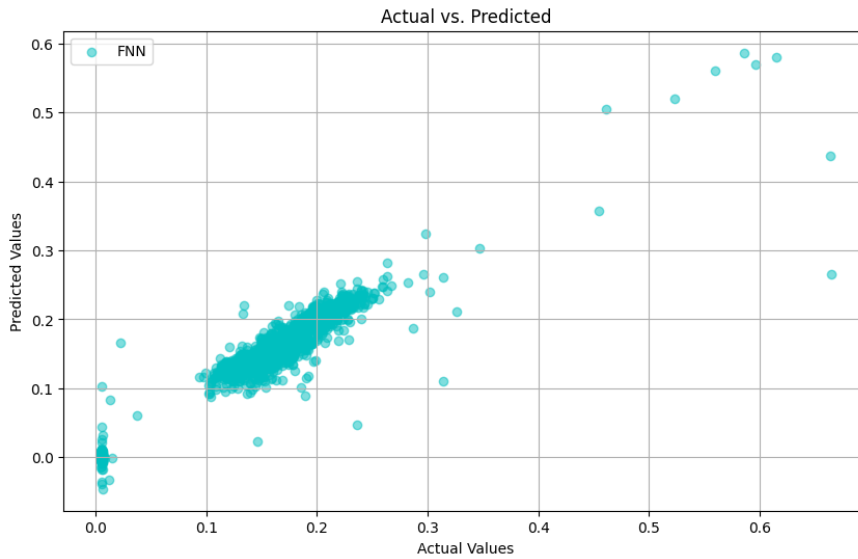
Το μοντέλο του CNN υπερέχει στο πλαίσιο του Ολοκληρωμένου Συνόλου Δεδομένων. Επιτυγχάνει χαμηλό MSE, υποδεικνύοντας ότι οι προβλέψεις του χαρακτηρίζονται από ελάχιστα τετράγωνα σφάλματα. Η τιμή MAE 0,0114 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν μόνο κατά 0,0114 μονάδες από τις πραγματικές τιμές, επιδεικνύοντας υψηλή ακρίβεια. Η ισχυρή βαθμολογία R2 0,9200 αντικατοπτρίζει την εξαιρετική ικανότητα του μοντέλου να εξηγεί περίπου το 92,00% της διακύμανσης στο Ολοκληρωμένο σύνολο δεδομένων, υπογραμμίζοντας την αποτελεσματικότητά του στην καταγραφή προγνωστικών μοτίβων.





Εικόνα 31: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το CNN στα δεδομένα του ενσωματωμένου συνόλου.

Το μοντέλο FNN, ενώ εξακολουθεί να παρέχει ισχυρές προβλέψεις, εμφανίζει ελαφρώς υψηλότερες τιμές MSE και MAE σε σύγκριση με τα μοντέλα LSTM, GRU και CNN. Το MSE 0,0005 δείχνει ότι τα τετράγωνα σφάλματα πρόβλεψης του μοντέλου είναι κάπως μεγαλύτερα. Το MAE 0,0123 υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις του μοντέλου αποκλίνουν κατά 0,0123 μονάδες από τις πραγματικές τιμές. Ωστόσο, το μοντέλο FNN διατηρεί υψηλό σκορ R2 0,9025, υποδηλώνοντας την ουσιαστική του ικανότητα να εξηγεί περίπου το 90,25% της διακύμανσης στο Ολοκληρωμένο σύνολο δεδομένων, υποδεικνύοντας ισχυρή προγνωστική απόδοση.



Εικόνα 32: Γραφική παράσταση της πραγματικής έναντι της προβλεπόμενης τιμής για το FNN στα δεδομένα του ενσωματωμένου συνόλου.

## Συνολική αξιολόγηση μοντέλων

Η συνολική αξιολόγηση μοντέλου αξιολογεί την απόδοση τεσσάρων διαφορετικών μοντέλων μηχανικής εκμάθησης —Μακροπρόθεσμης Βραχυπρόθεσμης Μνήμης (LSTM), Περιφερόμενης Επαναλαμβανόμενης Μονάδας (GRU), Συνελικτικού Νευρωνικού Δικτύου (CNN) και Νευρωνικού Δικτύου Feedforward (FNN)—σε διάφορα σύνολα δεδομένων, συμπεριλαμβανομένων Κινητήρας, δεξιό ρουλεμάν, αριστερό ρουλεμάν και το ενσωματωμένο σύνολο δεδομένων.

Πίνακας 8: Συνολική αξιολόγηση μοντέλων.

Μοντέλο	MSE	MAE	R <sup>2</sup>
LSTM	0,0008	0,0145	0,8499
GRU	0,0008	0,0121	0,8676
CNN	0,0008	0,0120	0,8619
FNN	0,0009	0,0149	0,8535

Το μοντέλο LSTM παρέχει σταθερά αξιόπαινη απόδοση σε όλα τα σύνολα δεδομένων. Επιδεικνύει την επάρκειά του στη σύλληψη διαδοχικών μοτίβων σε δεδομένα χρονοσειρών. Επιτυγχάνει χαμηλές τιμές MSE σε όλα τα σύνολα δεδομένων, υποδεικνύοντας ακριβείς προβλέψεις. Οι τιμές MAE υποδηλώνουν ότι οι προβλέψεις του μοντέλου αποκλίνουν λογικά ελάχιστα από τις πραγματικές τιμές. Επιπλέον, οι βαθμολογίες στο R2 τονίζουν την ικανότητά του να εξηγεί ένα σημαντικό μέρος της διακύμανσης σε κάθε σύνολο δεδομένων.

Το μοντέλο GRU επιδεικνύει σταθερά ισχυρή απόδοση σε όλα τα σύνολα δεδομένων. Όπως το μοντέλο LSTM, υπερέχει στην καταγραφή διαδοχικών εξαρτήσεων σε δεδομένα χρονοσειρών. Επιτυγχάνει χαμηλές τιμές MSE και MAE, υποδεικνύοντας την ικανότητά του να κάνει ακριβείς προβλέψεις. Οι υψηλές βαθμολογίες στο τετράγωνο R υπογραμμίζουν την αποτελεσματικότητά του στην εξήγηση της διακύμανσης σε κάθε σύνολο δεδομένων, με ιδιαίτερα αξιοσημείωτη απόδοση στα σύνολα δεδομένων Left Bearing και Right Bearing.

Το μοντέλο του CNN υπερέχει στην καταγραφή τοπικών προτύπων εντός των δεδομένων, ακόμη και σε σύνολα δεδομένων που δεν είναι εγγενώς χωρικά. Επιτυγχάνει τις χαμηλότερες τιμές MSE και MAE στο ολοκληρωμένο σύνολο δεδομένων, υποδεικνύοντας ακριβείς προβλέψεις. Επιπλέον, εμφανίζει την υψηλότερη βαθμολογία R-squared, υποδηλώνοντας την ισχυρή του ικανότητα να εξηγεί τη διακύμανση δεδομένων, ιδιαίτερα στο Integrated Dataset.

Το μοντέλο FNN προσφέρει ανταγωνιστική απόδοση σε όλα τα σύνολα δεδομένων. Αν και μπορεί να μην καταγράφει διαδοχικές εξαρτήσεις τόσο αποτελεσματικά όσο τα μοντέλα LSTM και GRU, διατηρεί σταθερά χαμηλές τιμές MSE και MAE, υποδεικνύοντας ακριβείς προβλέψεις. Το μοντέλο FNN αποδίδει εξαιρετικά καλά στο σύνολο δεδομένων Left Bearing όσον αφορά τη βαθμολογία R-squared.

## **Επιλογή μοντέλου**

Κατά την επιλογή του καταλληλότερου μοντέλου για εργασίες πρόβλεψης συντήρησης, είναι σημαντικό να ληφθούν υπόψη οι συγκεκριμένες απαιτήσεις της εφαρμογής και οι αντισταθμίσεις μεταξύ διαφορετικών μετρήσεων αξιολόγησης. Ακολουθούν τα βασικά στοιχεία για την επιλογή μοντέλου με βάση τη Συνολική Αξιολόγηση Μοντέλου:

- Χαμηλότερο MSE: Το μοντέλο CNN επιτυγχάνει το χαμηλότερο MSE στο ολοκληρωμένο σύνολο δεδομένων, καθιστώντας το εξαιρετική επιλογή όταν η ελαχιστοποίηση των σφαλμάτων πρόβλεψης αποτελεί κορυφαία προτεραιότητα.
- Χαμηλότερο MAE: Το μοντέλο GRU εμφανίζει το χαμηλότερο MAE στο σύνολο δεδομένων Right Bearing, υποδεικνύοντας την υψηλότερη ακρίβεια στις προβλέψεις.
- Υψηλότερη βαθμολογία στο R2: Το μοντέλο CNN επιδεικνύει την υψηλότερη βαθμολογία στο R2 στο ολοκληρωμένο σύνολο δεδομένων, τονίζοντας την ικανότητά του να εξηγεί τη διακύμανση δεδομένων και να καταγράφει αποτελεσματικά μοτίβα πρόβλεψης.
- Ευελιξία: Το μοντέλο FNN, ενώ είναι ανταγωνιστικό, παρουσιάζει ευελιξία και είναι ιδιαίτερα ισχυρό στο σύνολο δεδομένων αριστερού ρουλεμάν όσον αφορά τη βαθμολογία R2.

Τελικά, η επιλογή του καταλληλότερου μοντέλου θα πρέπει να ευθυγραμμίζεται με τους συγκεκριμένους στόχους και τους περιορισμούς της εργασίας πρόβλεψης συντήρησης. Η Συνολική Αξιολόγηση Μοντέλου παρέχει μια ολοκληρωμένη κατανόηση των δυνατών σημείων και των τομέων αριστείας κάθε μοντέλου σε διαφορετικά σύνολα δεδομένων, βοηθώντας στη λήψη μιας τεκμηριωμένης απόφασης σχετικά με την επιλογή μοντέλου για εφαρμογές πρόβλεψης υγείας μηχανημάτων πραγματικού κόσμου.

## Συζήτηση- Συμπεράσματα

Η προγνωστική συντήρηση, που καθοδηγείται από αλγόριθμους μηχανικής μάθησης, διαδραματίζει κεντρικό ρόλο στις σύγχρονες βιομηχανικές ρυθμίσεις, διασφαλίζοντας λειτουργική απόδοση, εξοικονόμηση κόστους και πρόληψη απρογραμμάτιστων αστοχιών εξοπλισμού. Σε αυτήν την ανάλυση, εφαρμόσαμε ένα ποικίλο σύνολο μοντέλων μηχανικής εκμάθησης σε δεδομένα αισθητήρων από εξαρτήματα μηχανών, με στόχο να προβλέψουμε την κατάσταση της υγείας τους και να εντοπίσουμε πιθανές ανάγκες συντήρησης. Η συζήτηση και τα συμπεράσματα παρακάτω συνοψίζουν τα βασικά μας ευρήματα και ιδέες:

### Επιλογή μοντέλου και απόδοση

Η ανάλυσή μας περιελάμβανε τέσσερα διαφορετικά μοντέλα μηχανικής εκμάθησης, το καθένα προσαρμοσμένο για να χειρίζεται τα συγκεκριμένα χαρακτηριστικά των συνόλων δεδομένων που σχετίζονται με δεδομένα κινητήρα, δεδομένα δεξιού ρουλεμάν, δεδομένα αριστερού ρουλεμάν και ενσωματωμένο σύνολο δεδομένων. Αυτά τα μοντέλα είναι τα LSTM, GRU, CNN και FNN.

Δεδομένα κινητήρα: Το μοντέλο CNN αναδείχθηκε ως το μοντέλο με τις καλύτερες επιδόσεις για δεδομένα κινητήρα, παρουσιάζοντας το χαμηλότερο μέσο τετράγωνο σφάλμα (MSE) και το μέσο απόλυτο σφάλμα (MAE). Αυτό το μοντέλο πέτυχε επίσης μια αξιόπαινη βαθμολογία R2, υποδηλώνοντας την ικανότητά του να εξηγεί τη διακύμανση στα δεδομένα. Ωστόσο, τα μοντέλα LSTM και GRU επέδειξαν επίσης ισχυρές προγνωστικές ικανότητες.

Δεδομένα δεξιάς ρουλεμάν: Το μοντέλο του CNN συνέχισε την εντυπωσιακή του απόδοση σε δεδομένα δεξιών ρουλεμάν, με το χαμηλότερο MSE, MAE και την υψηλότερη βαθμολογία R2. Και τα δύο μοντέλα LSTM και GRU εμφάνισαν αξιοσημείωτη προγνωστική ακρίβεια, τονίζοντας την καταλληλότητά τους για αυτό το σύνολο δεδομένων.

Δεδομένα αριστερού ρουλεμάν: Το μοντέλο GRU διέπρεψε στην πρόβλεψη της κατάστασης υγείας των αριστερών ρουλεμάν, επιτυγχάνοντας την υψηλότερη βαθμολογία R2. Ωστόσο, όλα τα μοντέλα επέδειξαν ισχυρές προγνωστικές ικανότητες για αυτό το σύνολο δεδομένων.

Ολοκληρωμένο σύνολο δεδομένων: Το μοντέλο CNN ξεπέρασε και πάλι τα άλλα στη μοντελοποίηση του ενσωματωμένου συνόλου δεδομένων, επιδεικνύοντας την ευελιξία του

στην καταγραφή πολύπλοκων αλληλεπιδράσεων μεταξύ των στοιχείων του μηχανήματος. Όλα τα μοντέλα συνεισέφεραν πολύτιμες πληροφορίες για τα ενσωματωμένα δεδομένα, με ισχυρές βαθμολογίες R2 σε όλους τους τομείς.

Συνολική Αξιολόγηση Μοντέλου: Για να παρέχουμε μια ενοποιημένη εικόνα της απόδοσης του μοντέλου, υπολογίσαμε τα συνολικά αποτελέσματα συγκεντρώνοντας μετρήσεις (MSE, MAE, R2) από κάθε μοντέλο και σύνολο δεδομένων, αποδίδοντας ίση βαρύτητα στην απόδοση κάθε συνόλου δεδομένων. Το μοντέλο του CNN επέδειξε με συνέπεια ανώτερη απόδοση σε όλα τα σύνολα δεδομένων, επιβεβαιώνοντας την αποτελεσματικότητά του σε εργασίες πρόβλεψης συντήρησης.

## Πρακτικές επιπτώσεις

Τα ευρήματα αυτής της ανάλυσης έχουν σημαντικές πρακτικές επιπτώσεις για τις στρατηγικές πρόβλεψης συντήρησης σε βιομηχανικά περιβάλλοντα:

- **Επιλογή μοντέλου:** Η επιλογή του μοντέλου μηχανικής εκμάθησης θα πρέπει να προσαρμόζεται στα συγκεκριμένα χαρακτηριστικά των δεδομένων και στην εργασία πρόβλεψης συντήρησης. Ενώ το μοντέλο του CNN διέπρεψε σε αυτήν την ανάλυση, τα μοντέλα LSTM και GRU αποδείχθηκαν επίσης εξαιρετικά αποτελεσματικά στην καταγραφή διαδοχικών μοτίβων σε δεδομένα χρονοσειράς.
- **Έγκαιρη συντήρηση:** Οι ακριβείς προβλέψεις που παρέχονται από αυτά τα μοντέλα επιτρέπουν στις ομάδες συντήρησης να προγραμματίζουν προληπτικά τις επισκευές και τις αντικαταστάσεις, ελαχιστοποιώντας το χρόνο διακοπής λειτουργίας και μειώνοντας το λειτουργικό κόστος.
- **Κατανομή πόρων:** Εντοπίζοντας εξαρτήματα μηχανημάτων που απαιτούν άμεση προσοχή, οι οργανισμοί μπορούν να κατανεύμουν τους πόρους αποτελεσματικά, βελτιστοποιώντας το κόστος εργασίας και υλικού.
- **Ενοποίηση δεδομένων:** Η ενσωμάτωση δεδομένων από πολλαπλά στοιχεία μηχανημάτων σε ένα ενιαίο σύνολο δεδομένων διευκολύνει μια ολιστική άποψη της υγείας του συστήματος, βοηθώντας στον έγκαιρο εντοπισμό ανωμαλιών και πιθανών ζητημάτων σε όλο το σύστημα.

## Περιορισμοί και Μελλοντικές Κατευθύνσεις

Είναι σημαντικό να αναγνωρίσουμε ορισμένους περιορισμούς αυτής της ανάλυσης. Η απόδοση των μοντέλων μηχανικής εκμάθησης μπορεί να διαφέρει ανάλογα με την ποιότητα και την ποσότητα των διαθέσιμων δεδομένων. Επιπλέον, παράγοντες όπως η βαθμονόμηση του αισθητήρα και οι περιβαλλοντικές συνθήκες μπορούν να επηρεάσουν την ακρίβεια των προβλέψεων. Οι μελλοντικές κατευθύνσεις για τη βελτίωση των στρατηγικών προγνωστικής συντήρησης περιλαμβάνουν:

- **Feature Engineering:** Διερεύνηση προηγμένων τεχνικών μηχανικής χαρακτηριστικών για την εξαγωγή πιο ουσιαστικών πληροφοριών από δεδομένα αισθητήρων.
- **Μέθοδοι συνόλου:** Εφαρμογή μεθόδων συνόλου για συνδυασμό προβλέψεων από πολλαπλά μοντέλα για βελτιωμένη ακρίβεια και ευρωστία.
- **Παρακολούθηση σε πραγματικό χρόνο:** Ενσωμάτωση ροών δεδομένων σε πραγματικό χρόνο για συνεχή παρακολούθηση και άμεση απόκριση σε αναδυόμενα ζητήματα.
- **Επεξηγησιμότητα:** Ανάπτυξη ερμηνεύσιμων μοντέλων για την παροχή πληροφοριών σχετικά με τους παράγοντες που συμβάλλουν στις προβλέψεις, ενισχύοντας την εμπιστοσύνη στις αποφάσεις συντήρησης που βασίζονται στη μηχανική μάθηση.

Συμπερασματικά, η προγνωστική συντήρηση που υποστηρίζεται από μοντέλα μηχανικής μάθησης προσφέρει τεράστιες δυνατότητες για βελτιστοποίηση βιομηχανικών λειτουργιών και μείωση του κόστους. Ενώ το μοντέλο του CNN αναδείχθηκε ως το κορυφαίο σε αυτήν την ανάλυση, η καταλληλότητα των μοντέλων μπορεί να διαφέρει σε διαφορετικά σύνολα δεδομένων και σενάρια. Αγκαλιάζοντας αυτά τα ευρήματα και βελτιώνοντας συνεχώς τις στρατηγικές προγνωστικής συντήρησης, οι οργανισμοί μπορούν να αξιοποιήσουν τη δύναμη των γνώσεων που βασίζονται σε δεδομένα για να εξασφαλίσουν την αξιοπιστία και τη μακροζωία των κρίσιμων εξαρτημάτων του μηχανήματος.

## Βιβλιογραφία

- [1] Raheem, F. και Iqbal, N. (2022) ‘Artificial Intelligence and machine learning for the Industrial Internet of Things (iiot)’, *Industrial Internet of Things*, σσ. 1–20.
- [2] Kamel, H. (2022) ‘Artificial Intelligence for Predictive Maintenance’, *Journal of Physics: Conference Series*, 2299(1), σ. 012001.
- [3] Pashami, S. κ.α. (2023) Explainable predictive maintenance, arXiv.org. Διαθέσιμο: <https://arxiv.org/abs/2306.05120> (Ημ. Επίσκεψης: Σεπτέμβριος 2023).
- [4] Zonta, T. κ.α. (2020) ‘Predictive maintenance in the industry 4.0: A systematic literature review’, *Computers & Industrial Engineering*, 150, σ. 106889.
- [5] Preis, S.J. (2022) ‘Predictive analytics for equipment maintenance operations’, *Handbook of Research on Digital Transformation, Industry Use Cases, and the Impact of Disruptive Technologies*, σ.σ. 340–358.
- [6] Gautam, S., Noureddine, R. και Solvang, W.D. (2023) ‘Machine Learning and IIoT application for predictive maintenance’, *Advanced Manufacturing and Automation XII*, σ.σ. 257–265.
- [7] Tan, J.P. κ.α. (2022) ‘A performance review of recurrent neural networks long short-term memory (LSTM)’, 2022 3rd International Conference for Emerging Technology (INCET) [Preprint].
- [8] J. P. Tan, A. L. A. Ramos, M. V. Abante, R. L. Tadeo και R. R. Lansigan, "A Performance Review of Recurrent Neural Networks Long Short-Term Memory (LSTM)," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, σσ. 1-5.
- [9] Li, D. κ.α. (2022) ‘A privacy preservation framework for feedforward-designed Convolutional Neural Networks’, *Neural Networks*, 155, σσ. 14–27.
- [10] Zhang, Z., Feng, F. και Huang, T. (2022) ‘FNNS: An effective feedforward neural network scheme with random weights for processing large-scale datasets’, *Applied Sciences*, 12(23), σ. 12478.



- [11] F. R. Alharbi και D. Csala, "Short-Term Wind Speed and Temperature Forecasting Model Based on Gated Recurrent Unit Neural Networks," 2021 3rd Global Power, Energy and Communication Conference (GPECOM), Antalya, Turkey, 2021, σσ. 142-147.
- [12] Garcia, R.F. κ.α. (2013) 'Expert condition monitoring on hydrostatic self-levitating bearings', *Expert Systems with Applications*, 40(8), σ.σ. 2975–2984.
- [13] A. K. S. Jardine, D. Lin, και D. Banjevic, 'A review on machinery diagnostics and prognostics implementing condition-based maintenance', *Mech. Syst. Signal Process.*, τ. 20, τχ. 7, σσ. 1483–1510, Οκτωβρίου 2006.
- [14] P. Girdhar και C. Scheffer, '1 - Predictive maintenance techniques: Part 1 predictive maintenance basics', στο *Practical Machinery Vibration Analysis and Predictive Maintenance*, P. Girdhar και C. Scheffer, Επιμ. Oxford: Newnes, 2004, σσ. 1–10.
- [15] P. Girdhar και C. Scheffer, '2 - Predictive maintenance techniques: Part 2 vibration basics', στο *Practical Machinery Vibration Analysis and Predictive Maintenance*, P. Girdhar και C. Scheffer, Επιμ. Oxford: Newnes, 2004, σσ. 11–28.
- [16] D. A. Tobon-Mejia, K. Medjaher, και N. Zerhouni, 'CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks', *Mech. Syst. Signal Process.*, τ. 28, σσ. 167–182, Απριλίου 2012.
- [17] I. Guyon και A. Elisseeff, 'An Introduction to Variable and Feature Selection', σ. 26.
- [18] M. Kudo και J. Sklansky, 'Comparison of algorithms that select features for pattern classifiers', *Pattern Recognit.*, τ. 33, τχ. 1, σσ. 25–41, Ιανουαρίου 2000.
- [19] Y. Fu κ.α., 'Machining vibration states monitoring based on image representation using convolutional neural networks', *Eng. Appl. Artif. Intell.*, τ. 65, σσ. 240–251, Οκτωβρίου 2017.
- [20] D. Barschdorff, 'COMPARISON OF NEURAL AND CLASSICAL DECISION ALGORITHMS', στο *Fault Detection, Supervision and Safety for Technical Processes 1991*, A. Isermann και B. Freyermuth, Επιμ. Oxford: Pergamon, 1992, σσ. 409–415.
- [21] T. Wagner, 'Review of 'Pattern Recognition Principles' (Tou, J. T., and Gonzalez, R. C.; 1974)', *IEEE Trans. Inf. Theory*, τ. 22, τχ. 5, σσ. 632–633, Σεπτεμβρίου 1976.

- [22] S. Qian και D. Chen, ‘Joint time-frequency analysis’, *IEEE Signal Process. Mag.*, τ. 16, τχ. 2, σσ. 52–67, Μαρτίου 1999.
- [23] C. M. Bishop και P. of N. C. C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.
- [24] C.-T. Lin και C. S. G. Lee, *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*. USA: Prentice-Hall, Inc., 1996.
- [25] T. W. Rauber, F. de Assis Boldt, και F. M. Varejão, ‘Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis’, *IEEE Trans. Ind. Electron.*, τ. 62, τχ. 1, σσ. 637–646, Ιανουαρίου 2015.
- [26] B. Iso, ‘BS ISO 20816-1:2016, Mechanical vibration — Measurement and evaluation of machine vibration - Part 1: General guidelines’, σ. 46, 2016.
- [27] A. T. A. Dantas, M. Batista Leite, και K. de Jesus Nagahama, ‘Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks’, *Constr. Build. Mater.*, τ. 38, σσ. 717–722, Ιανουαρίου 2013.
- [28] J. Hu, J. Wang, και G. Zeng, ‘A hybrid forecasting approach applied to wind speed time series’, *Renew. Energy*, τ. 60, σσ. 185–194, Δεκεμβρίου 2013.
- [29] Y. Guo, Z. Yang, K. Liu, Y. Zhang, και W. Feng, ‘A compact and optimized neural network approach for battery state-of-charge estimation of energy storage system’, *Energy*, τ. 219, σ. 119529, Μαρτίου 2021.
- [30] K. Yu, J. Liang, B. Qu, Y. Luo, και C. Yue, ‘Dynamic Selection Preference-Assisted Constrained Multiobjective Differential Evolution’, *IEEE Trans. Syst. Man Cybern. Syst.*, τ. 52, τχ. 5, σσ. 2954–2965, Μαΐου 2022.
- [20][31] M. Caixinha και S. Nunes, ‘Machine Learning Techniques in Clinical Vision Sciences’, *Curr. Eye Res.*, τ. 42, τχ. 1, σσ. 1–15, Ιανουαρίου 2017.
- [32] C. Bunea και T. Bedford, ‘The effect of model uncertainty on maintenance optimization’, *IEEE Trans. Reliab.*, τ. 51, τχ. 4, σσ. 486–493, Δεκεμβρίου 2002.

- [33] J. Lee, E. Lapira, B. Bagheri, και H. Kao, 'Recent advances and trends in predictive manufacturing systems in big data environment', *Manuf. Lett.*, τ. 1, τχ. 1, σσ. 38–41, Οκτωβρίου 2013.
- [34] W. Zhang, D. Yang, και H. Wang, 'Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey', *IEEE Syst. J.*, τ. 13, τχ. 3, σσ. 2213–2227, Σεπτεμβρίου 2019.
- [35] J. Zenisek, F. Holzinger, και M. Affenzeller, 'Machine learning based concept drift detection for predictive maintenance', *Comput. Ind. Eng.*, τ. 137, σ. 106031, Νοεμβρίου 2019.
- [36] B. Dong, C. Cao, και S. E. Lee, 'Applying support vector machines to predict building energy consumption in tropical region', *Energy Build.*, τ. 37, τχ. 5, σσ. 545–553, Μαΐου 2005.
- [37] J. Xu, C. Xu, B. Zou, Y. Y. Tang, J. Peng, και X. You, 'New Incremental Learning Algorithm With Support Vector Machines', *IEEE Trans. Syst. Man Cybern. Syst.*, τ. 49, τχ. 11, σσ. 2230–2241, Νοεμβρίου 2019.
- [38] J. Zheng, H. Pan, και J. Cheng, 'Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines', *Mech. Syst. Signal Process.*, τ. 85, σσ. 746–759, Φεβρουαρίου 2017.
- [39] P. Konar και P. Chattopadhyay, 'Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs)', *Appl. Soft Comput.*, τ. 11, τχ. 6, σσ. 4203–4211, Σεπτεμβρίου 2011.
- [40] Q. Hu, Z. He, Z. Zhang, και Y. Zi, 'Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble', *Mech. Syst. Signal Process.*, τ. 21, τχ. 2, σσ. 688–705, Φεβρουαρίου 2007.
- [41] Goyal, R., Chandra, P. και Singh, Y. (2014) 'Suitability of KNN regression in the development of interaction-based software fault prediction models', *IERI Procedia*, 6, σσ. 15–21.
- [42] V. Kavana και M. Neethi, 'Fault Analysis and Predictive Maintenance of Induction Motor Using Machine Learning', στο 2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Δεκεμβρίου 2018, σσ. 963–966.

- [43] G. Scalabrini Sampaio, A. R. de A. Vallim Filho, L. Santos da Silva, και L. Augusto da Silva, 'Prediction of Motor Failure Time Using An Artificial Neural Network', *Sensors*, τ. 19, τχ. 19, Art. τχ. 19, Ιανουαρίου 2019.
- [44] H. Shao, H. Jiang, H. Zhao, και F. Wang, 'A novel deep autoencoder feature learning method for rotating machinery fault diagnosis', *Mech. Syst. Signal Process.*, τ. 95, σσ. 187–204, Οκτωβρίου 2017.
- [45] C. Zhang, C. Gupta, A. Farahat, K. Ristovski, και D. Ghosh, 'Equipment Health Indicator Learning Using Deep Reinforcement Learning', στο *Machine Learning and Knowledge Discovery in Databases*, Cham, 2019, σσ. 488–504.
- [46] Saxena, A. κα. (2008) 'Damage propagation modeling for aircraft engine run-to-failure simulation', 2008 International Conference on Prognostics and Health Management, Οκτώβριος 2008, σσ. 1–9.
- [47] Rodríguez, C. και Droguett, L. (1970) Repositorio Académico - universidad de chile, Repositorio Académico de la Universidad de Chile. Διαθέσιμο: <https://repositorio.uchile.cl/handle/2250/170996> (Ημ. Επίσκεψης: Σεπτέμβριος 2023).
- [48] Guo, X., Chen, L. και Shen, C. (2016) 'Hierarchical adaptive deep convolution neural network and its application to Bearing Fault diagnosis', *Measurement*, 93, σσ. 490–502.
- [49] O. Aydin και S. Guldamlasioglu, 'Using LSTM networks to predict engine condition on large scale data processing framework', στο 2017 4th International Conference on Electrical and Electronic Engineering (ICEEE), Απριλίου 2017, σσ. 281–285.
- [50] D. Bruneo και F. De Vita, 'On the Use of LSTM Networks for Predictive Maintenance in Smart Industries', στο 2019 IEEE International Conference on Smart Computing (SMARTCOMP), Ιουνίου 2019, σσ. 241–248.