



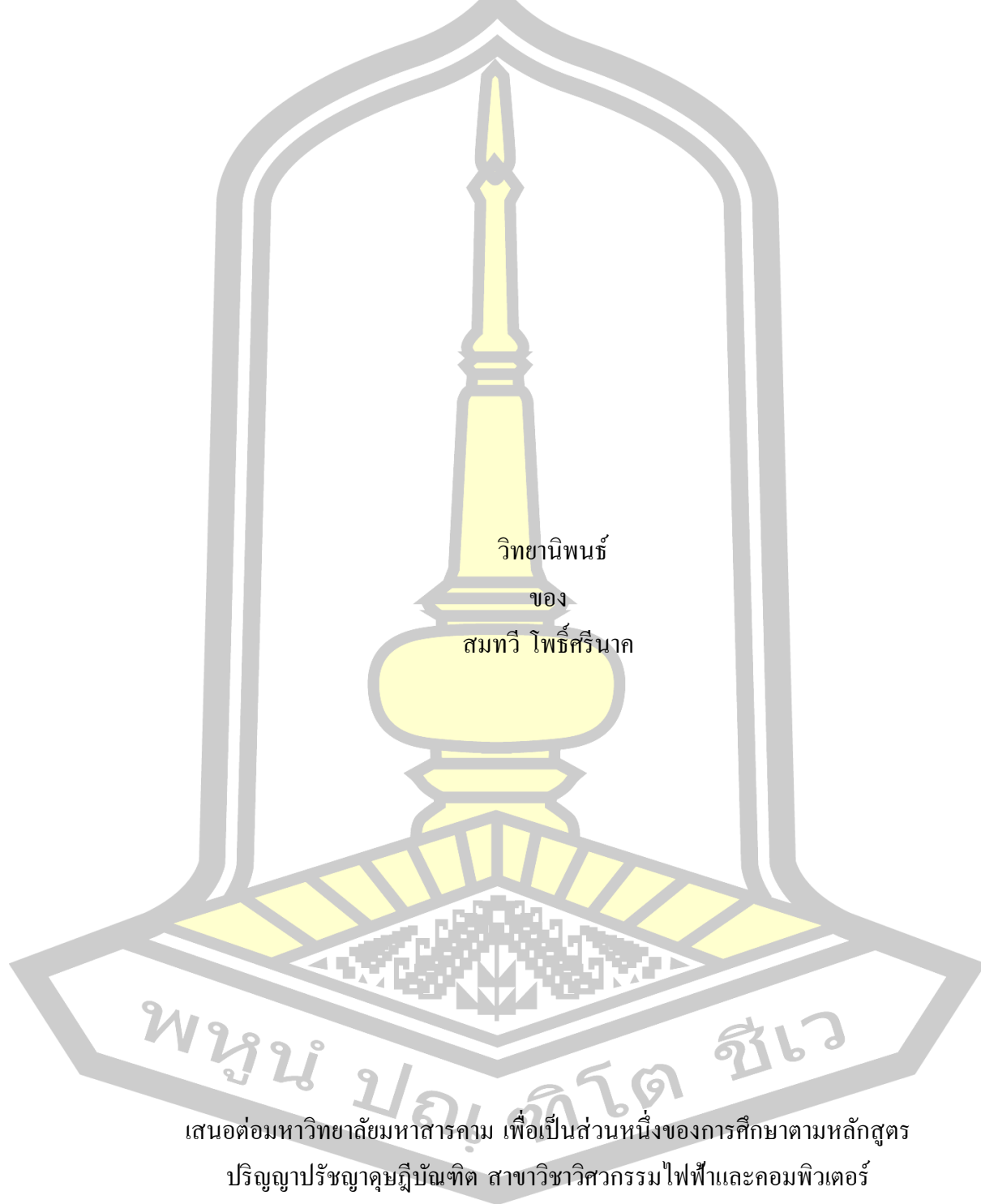
Optimized Machine Learning for Induction Motor Fault Diagnosis Using Vibration
and Frequency-Domain Features

Somdavee Bhosinak

A Thesis Submitted in Partial Fulfillment of Requirements for
degree of Doctor of Philosophy in Electrical and Computer Engineering
November 2025

Copyright of Mahasarakham University

การวิจัยความผิดปกติของมอเตอร์เหนี่ยวนำด้วยวิธีการปรับปรุงการเรียนรู้ของเครื่อง โดยใช้
ข้อมูลการสั่นสะเทือนและความถี่



เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์

พฤษภาคม 2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Optimized Machine Learning for Induction Motor Fault Diagnosis Using Vibration
and Frequency-Domain Features

Somdavee Bhosinak

A Thesis Submitted in Partial Fulfillment of Requirements
for Doctor of Philosophy (Electrical and Computer Engineering)

November 2025

Copyright of Mahasarakham University



The examining committee has unanimously approved this Thesis, submitted by Mr. Somdavee Bhosinak , as a partial fulfillment of the requirements for the Doctor of Philosophy Electrical and Computer Engineering at Mahasarakham University

Examining Committee

Chairman

(Assoc. Prof. Anan Kruesubthaworn
, Ph.D.)

Advisor

(Prof. Worawat Sa-Ngiamvibool ,
Ph.D.)

Co-advisor

(Assoc. Prof. Chonlatee Photong ,
Ph.D.)

Co-advisor

(Assoc. Prof. Niwat Angkawisittpan
, Ph.D.)

Committee

(Assoc. Prof. Supannika Wattana ,
Ph.D.)

External Committee

(Asst. Prof. Jagraphon Obma ,
Ph.D.)

Mahasarakham University has granted approval to accept this Thesis as a partial fulfillment of the requirements for the Doctor of Philosophy Electrical and Computer Engineering

(Assoc. Prof. Juckamas Laohavanich ,
Ph.D.)

Dean of The Faculty of Engineering

(Asst. Prof. Pondej Chaowarat , Ph.D.)

Dean of Graduate School

TITLE	Optimized Machine Learning for Induction Motor Fault Diagnosis Using Vibration and Frequency-Domain Features		
AUTHOR	Somdavee Bhosinak		
ADVISORS	Professor Worawat Sa-Ngiamvibool , Ph.D. Associate Professor Chonlatee Photong , Ph.D. Associate Professor Niwat Angkawisittpan , Ph.D.		
DEGREE	Doctor of Philosophy	MAJOR	Electrical and Computer Engineering
UNIVERSITY	Maharakham University	YEAR	2025

ABSTRACT

This research presents an optimized machine learning framework for predictive maintenance in industrial motor systems, utilizing vibration signal analysis and ensemble-based modelling to enhance fault detection accuracy and reliability. A publicly available dataset from Mendeley Data, consisting of vibration signals recorded from a 1.0 kW induction motor under eight operational conditions, was employed to evaluate model performance. The study integrates both time-domain and frequency-domain feature extraction using Fast Fourier Transform (FFT), enabling the identification of critical vibration patterns associated with motor degradation. Three experimental scenarios were designed to assess the impact of feature selection, data balancing, and hyperparameter tuning. Various machine learning algorithms—including Random Forest, Gradient Boosting, AdaBoost, Light Gradient Boosting Machine (LightGBM), and CatBoost—were trained and optimized using GridSearchCV to determine the most effective model configuration. Model performance was evaluated through multiple metrics such as Accuracy, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The results demonstrate that the optimized CatBoost model achieved the highest classification accuracy of 87.28%, outperforming other classifiers while maintaining computational efficiency and compact model size. These findings confirm that the integration of FFT-based features, data balancing, and hyperparameter optimization significantly enhances model robustness, making the proposed framework suitable for real-time industrial IoT applications. The study provides a reproducible and scalable approach for intelligent predictive maintenance, supporting the transition toward data-driven and autonomous maintenance strategies in smart manufacturing environments.

Keyword : predictive maintenance, industrial motor systems, ensemble models, FFT feature extraction, hyperparameter optimization, vibration analysis

ACKNOWLEDGEMENTS

The authors sincerely thank Professor Worawat Sa-Ngamvibool, Ph.D. who helped to bring this work together. Particularly thanks go to Committee and Mahasarakham University's staff for their ongoing support, technical help, and resource availability across the project life. This research was made possible in great part by their direction and encouragement.

Somdavee Bhosinak

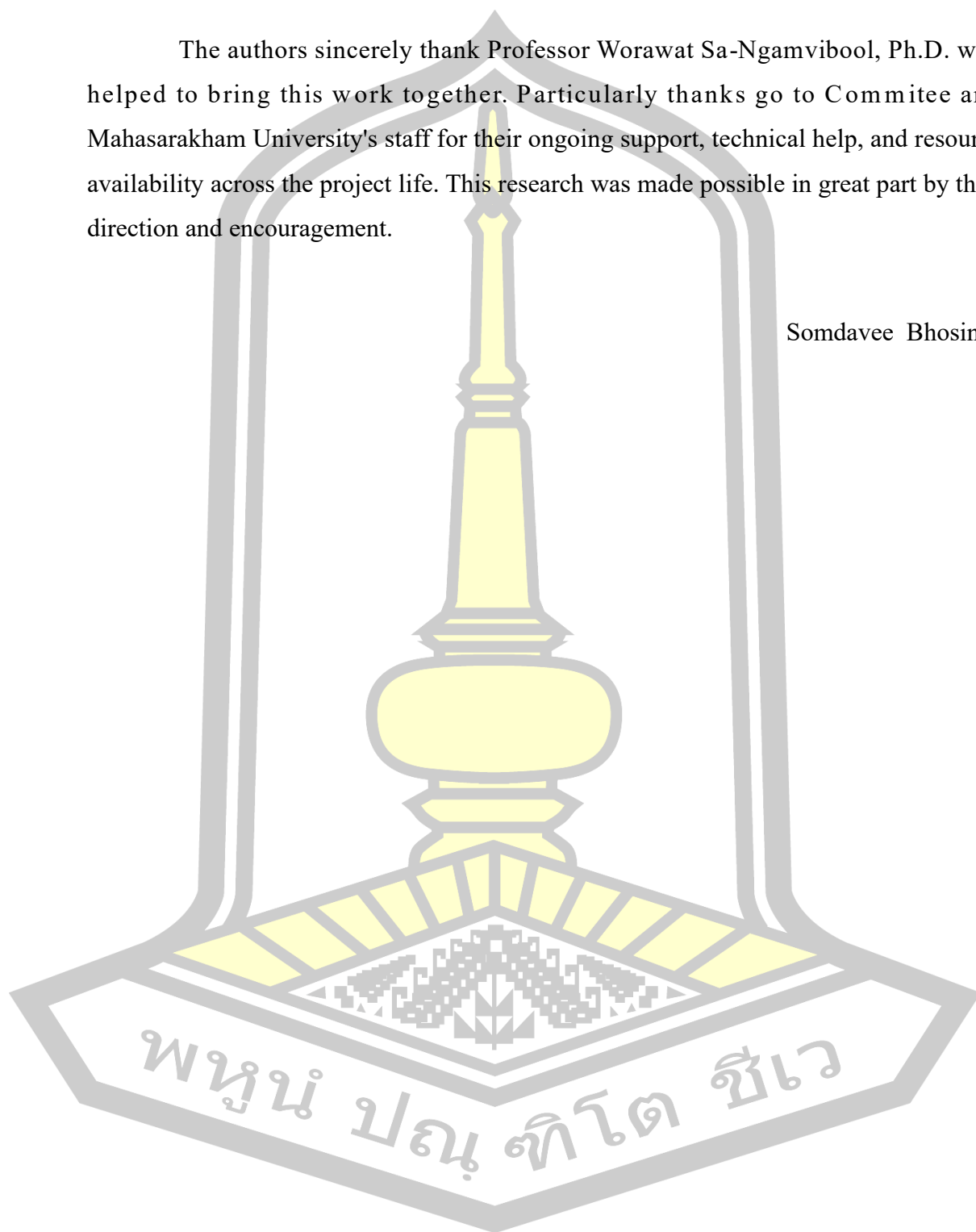
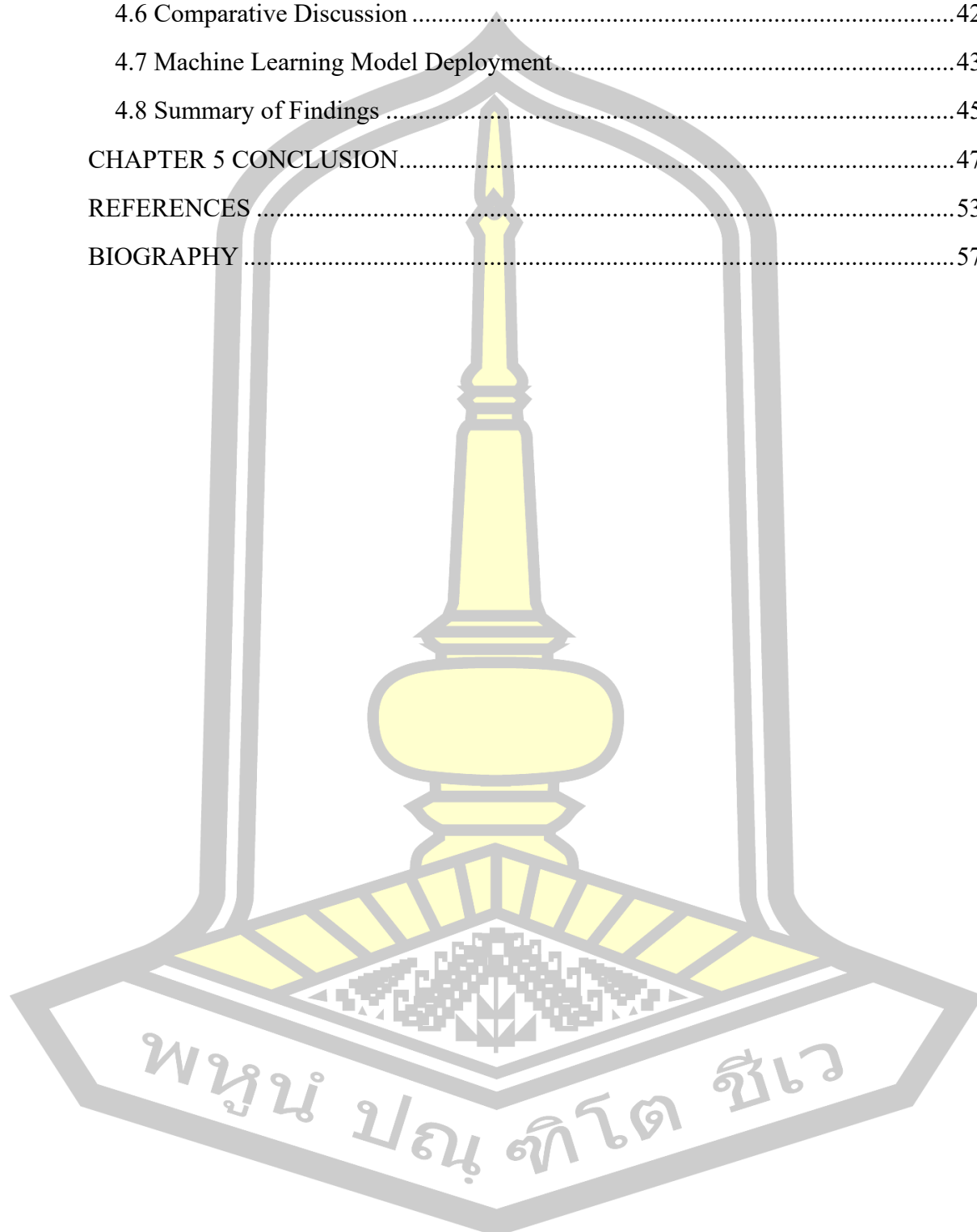


TABLE OF CONTENTS

	Page
ABSTRACT.....	D
ACKNOWLEDGEMENTS.....	E
TABLE OF CONTENTS.....	F
LIST OF TABLES	H
LIST OF FIGURES	I
CHAPTER 1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Purpose of the Study	3
1.3 Research Objectives.....	4
1.4 Scope of the Study.....	4
1.5 Significance of the Study.....	5
CHAPTER 2 THEORIES AND LITERATURE REVIEW	7
2.1 State of the Art	7
2.2 Operation of Industrial motor	20
2.3 Predictive Maintenance and Failure Prediction	21
2.4 The Utilization of Sensor Technology in Industrial Contexts	21
2.5 Fast Fourier Transform (FFT) in Predictive Maintenance.....	21
2.6 The Application of Deep Learning Techniques in Predictive Maintenance	24
CHAPTER 3 RESEARCH METHODOLOGY	28
3.1 Theoretical Framework.....	28
CHAPTER 4 RESULTS AND DISCUSSION	36
4.1 Dataset Characteristics.....	36
4.2 Setup 1: Baseline Multiclass Classification Using Time-Domain Features	38
4.3 Setup 2: Binary Classification with Time–Frequency Features	39
4.4 Setup 3: Optimized Binary Classification with Hyperparameter Tuning.....	40

4.5 Computational Efficiency: Inference Time and Model Size	41
4.6 Comparative Discussion	42
4.7 Machine Learning Model Deployment.....	43
4.8 Summary of Findings	45
CHAPTER 5 CONCLUSION.....	47
REFERENCES	53
BIOGRAPHY	57

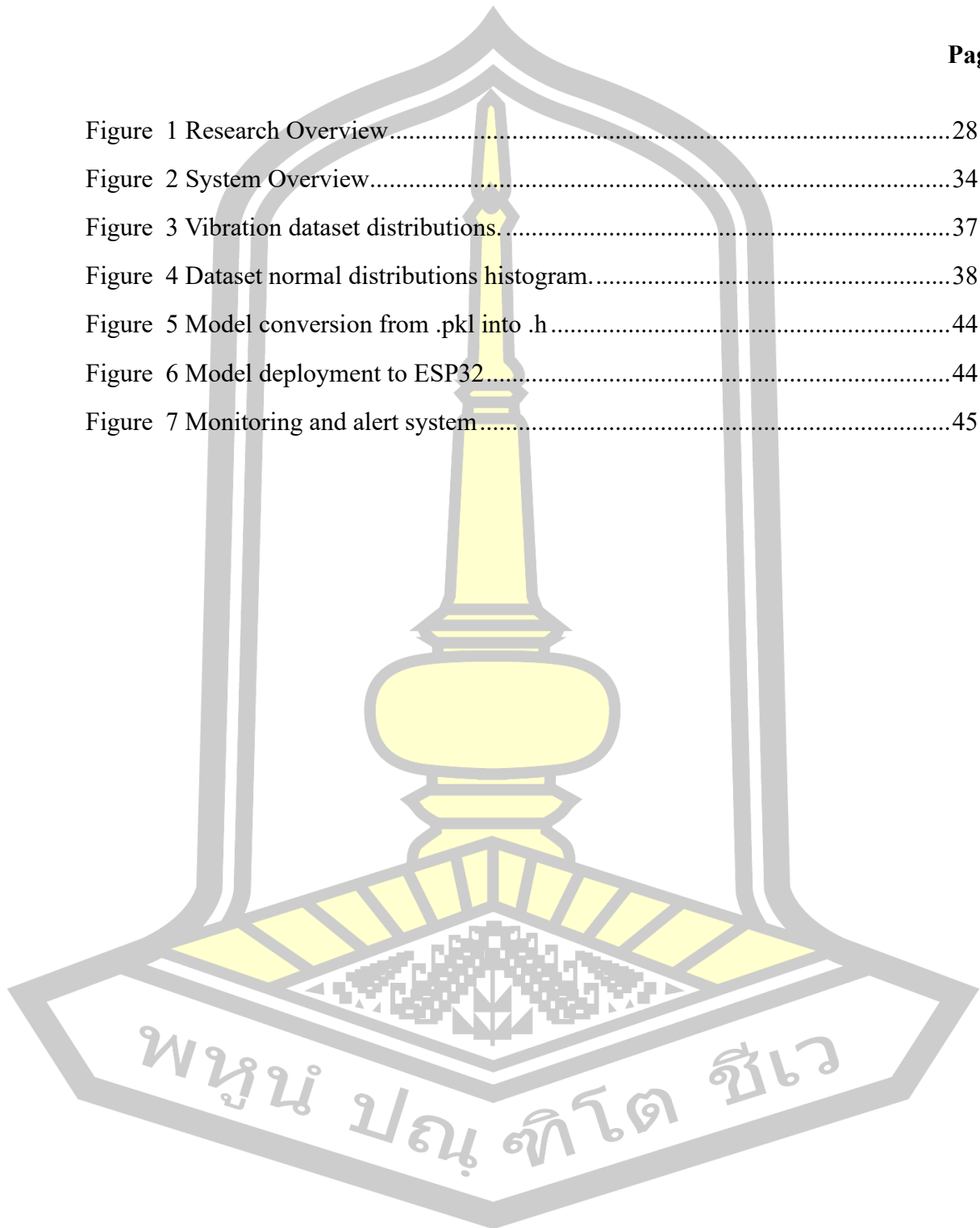


LIST OF TABLES

	Page
Table 1 Literature review of power plants based on maintenance type and machine learning type.....	9
Table 2 Literature review of power plants based on maintenance type and machine learning type (continued)	10
Table 3 Literature review of power plants based on maintenance type and machine learning type (continued)	11
Table 4 Literature review of power plants based on maintenance type and machine learning type (continued)	12
Table 5 Literature review of power plants based on maintenance type and machine learning type (continued)	13
Table 6 Literature review of power plants based on maintenance type and machine learning type (continued)	14
Table 7 Literature review of power plants based on maintenance type and machine learning type (continued)	15
Table 8 Literature review of optimization across diverse domains	16
Table 9 Literature review of optimization across diverse domains (continued)	17
Table 10 Literature review of optimization across diverse domains (continued)	18
Table 11 Literature review of optimization across diverse domains (continued)	19
Table 12 Experimental Setup	30
Table 13 GridSearchCV Hypertuning Optimization Setup.....	33
Table 14 Setup 1 Experimental Results	38
Table 15 Setup 2 Experimental Results	39
Table 16 Setup 3 Experimental Results	40
Table 17 Model Efficiency Results.....	41

LIST OF FIGURES

	Page
Figure 1 Research Overview.....	28
Figure 2 System Overview.....	34
Figure 3 Vibration dataset distributions.....	37
Figure 4 Dataset normal distributions histogram.....	38
Figure 5 Model conversion from .pkl into .h.....	44
Figure 6 Model deployment to ESP32.....	44
Figure 7 Monitoring and alert system.....	45



CHAPTER 1

INTRODUCTION

1.1 Background

In modern industrial systems, induction motors play a vital role as the primary drivers of machinery used in manufacturing, automation, power generation, and energy conversion. These motors are valued for their high efficiency, robustness, and low maintenance requirements. However, despite their widespread use, induction motors are prone to various electrical and mechanical faults, such as inter-turn short circuits, bearing failures, shaft misalignment, and rotor bar defects. If these faults go undetected, they can lead to unexpected breakdowns, production losses, and significant maintenance costs.

Traditional maintenance strategies, including corrective maintenance (reactive action after failure) and preventive maintenance (periodic servicing based on time intervals), have proven inadequate in addressing the dynamic and evolving nature of machine operation. These methods often result in either over-maintenance of healthy equipment or delayed responses to developing faults. Consequently, industries are increasingly adopting predictive maintenance (PdM) a proactive approach that integrates sensor-based monitoring, data analytics, and intelligent algorithms to predict failures before they occur.

The advancement of Industrial Internet of Things (IIoT) and smart manufacturing technologies has further strengthened predictive maintenance practices by enabling continuous data acquisition from embedded sensors. Among various sensing modalities, vibration analysis has emerged as a powerful technique because

vibration patterns carry rich information about a machine's internal state. Subtle changes in vibration amplitude and frequency content can reveal early signs of mechanical or electrical degradation.

In recent years, the combination of vibration signal analysis and machine learning (ML) has enabled the development of intelligent diagnostic systems for induction motors. Classical machine learning algorithms such as K-Nearest Neighbors (KNN), Decision Trees (DT), and Random Forests (RF) have been used to classify faults based on engineered features. More recently, advanced ensemble models like Gradient Boosting (GB), Light Gradient Boosting Machine (LightGBM), and CatBoost have demonstrated improved robustness in handling complex, nonlinear data relationships.

Despite these advances, several limitations persist in current research. Many previous studies emphasize handcrafted features from time-domain signals while overlooking frequency-domain information that can capture critical harmonic behaviors. Moreover, models are often evaluated on small, imbalanced datasets without sufficient optimization, resulting in inconsistent performance across different fault conditions.

To address these challenges, this research develops a machine learning-based predictive maintenance framework that utilizes vibration data, integrates Fast Fourier Transform (FFT)-based feature extraction, and applies hyperparameter-optimized ensemble algorithms to improve diagnostic accuracy and computational efficiency.

1.2 Purpose of the Study

Induction motor failures can cause substantial disruptions to industrial processes and lead to unplanned downtime. Despite existing maintenance strategies, challenges remain in ensuring early and accurate fault detection:

- **Reactive and Time-Based Maintenance:** Conventional maintenance schedules do not account for real-time equipment condition, leading to inefficient operations and unexpected failures.
- **Incomplete Signal Representation:** Models that rely only on time-domain features fail to capture essential frequency-domain characteristics where fault-related harmonics reside.
- **Limited Model Optimization:** Many studies use fixed model parameters without systematic tuning, resulting in suboptimal fault classification performance.
- **Imbalanced and Noisy Datasets:** Fault data are often underrepresented compared to normal data, leading to model bias and reduced reliability in real-world applications.

Therefore, there is a need for a data-driven, optimized predictive maintenance framework capable of extracting meaningful information from pre-acquired vibration signals and delivering accurate, reliable fault classification results suitable for industrial deployment. The purpose of this study is to develop and evaluate an optimized machine learning framework for induction motor fault diagnosis using vibration signal analysis. Specifically, this research aims to enhance fault detection performance through the integration of FFT-based feature extraction, data balancing, and hyperparameter optimization techniques. By systematically

comparing and optimizing multiple machine learning algorithms, this study demonstrates that significant improvements in predictive maintenance accuracy can be achieved without the need for additional sensor design or new data acquisition.

1.3 Research Objectives

The specific objectives of this study are as follows:

1. To preprocess and utilize a publicly available induction motor vibration dataset obtained from Mendeley Data.
2. To perform feature extraction in both the time and frequency domains using statistical and FFT-based methods.
3. To train and evaluate machine learning models, including KNN, Random Forest, Gradient Boosting, AdaBoost, LightGBM, and CatBoost, for fault detection.
4. To apply data balancing techniques to improve model generalization and robustness against class imbalance.
5. To perform hyperparameter optimization using GridSearchCV for each model to maximize classification accuracy.
6. To identify the best-performing algorithm based on performance metrics such as Accuracy, MSE, RMSE, MAPE, R^2 , inference time, and model size.

1.4 Scope of the Study

This study contains a complex significance and carries major implications for the renewable energy sector as well as the broader subject of predictive maintenance in industrial systems.

The scope of this research is clearly defined as follows:

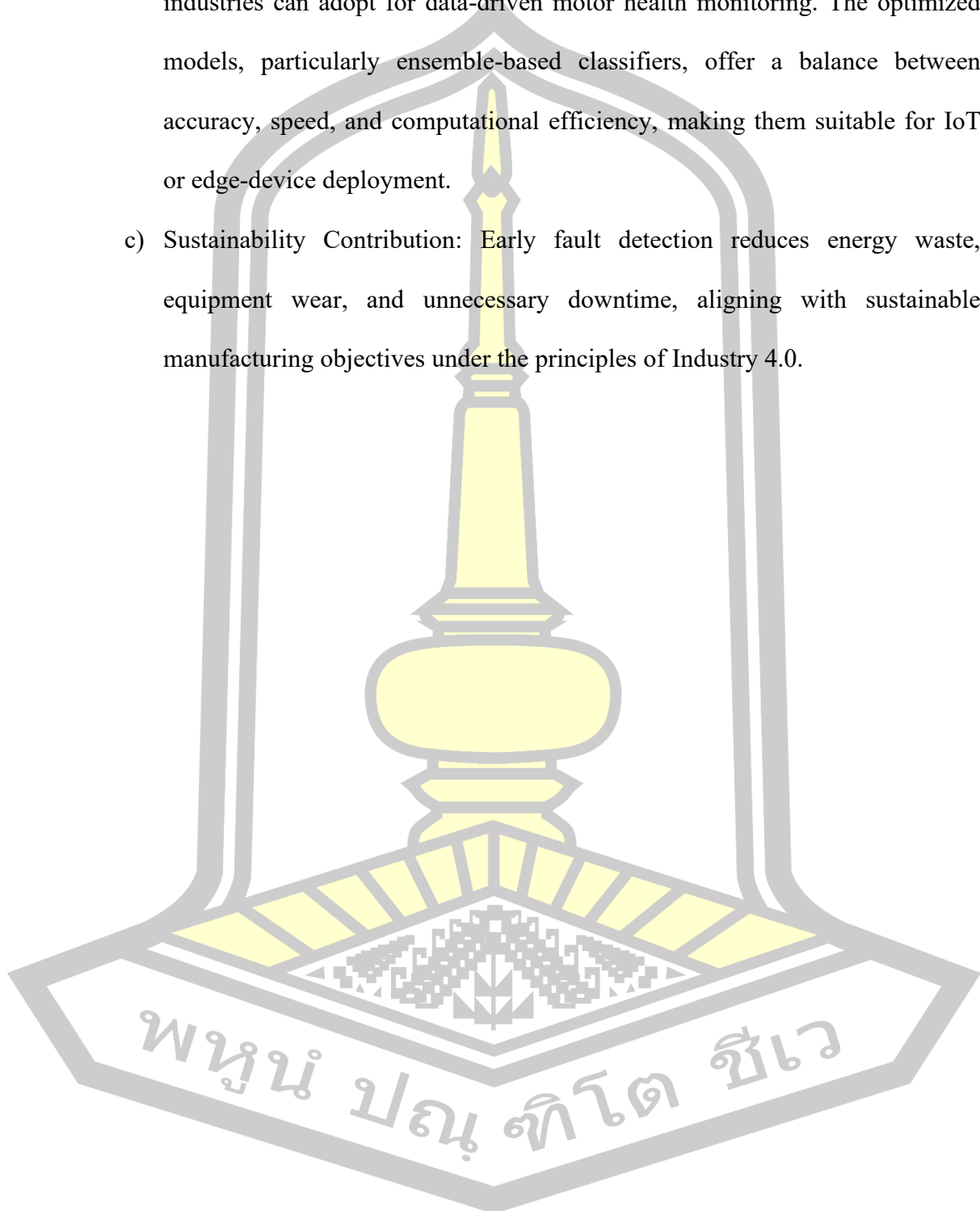
- a) Focus: Fault diagnosis of induction motors using vibration data analysis.
- b) Data Source: Vibration signals obtained from a publicly available Mendeley dataset (TDMS format).
- c) Fault Type: Inter-turn short circuit faults at varying severity levels.
- d) Features: Time-domain and frequency-domain features extracted via FFT.
- e) Models Evaluated: KNN, Random Forest, Gradient Boosting, AdaBoost, LightGBM, and CatBoost.
- f) Evaluation Metrics: Accuracy, MSE, RMSE, MAPE, R^2 , inference time, and model size.
- g) Limitations:
 - The study does not include sensor design, placement, or data collection.
 - The scope is limited to electrical faults, excluding mechanical defects such as bearing wear or shaft imbalance.
 - Real-time hardware implementation and deep neural network architectures are reserved for future work.

1.5 Significance of the Study

This research provides both academic and practical contributions:

- a) Academic Significance: The study demonstrates the effectiveness of using publicly available vibration datasets for predictive maintenance research. It contributes to the growing field of intelligent fault diagnosis by integrating FFT-based feature extraction with optimized ensemble learning techniques.

- b) **Industrial Significance:** The findings present a scalable framework that industries can adopt for data-driven motor health monitoring. The optimized models, particularly ensemble-based classifiers, offer a balance between accuracy, speed, and computational efficiency, making them suitable for IoT or edge-device deployment.
- c) **Sustainability Contribution:** Early fault detection reduces energy waste, equipment wear, and unnecessary downtime, aligning with sustainable manufacturing objectives under the principles of Industry 4.0.



CHAPTER 2

THEORIES AND LITERATURE REVIEW

This section offers a complete review of the theories and extant literature pertaining to the research aims of improving the dependability of Industrial motor by means of sensor fusion and failure prediction based on deep learning techniques. The literature review comprehensively examines several facets of industrial motor operation, predictive maintenance, and the utilization of modern sensor technology and machine learning algorithms.

2.1 State of the Art

Many researchers in Table 1 discusses several related studies that have been published. Recent advancements in predictive and descriptive maintenance have been a focal point of research across diverse industrial sectors. Eriksson (2020) addresses the maintenance challenges in wind turbines by leveraging machine learning techniques such as Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM), with MLP emerging as a promising model for predicting failures. Fernandes et al. (2020) extend this focus to heating, ventilation, and air conditioning (HVAC) systems, employing Neural Networks (NN) and Decision Trees (DT) to optimize predictive maintenance configurations. In the realm of nuclear power plants, Gohel et al. (2020) utilize SVM and Linear Regression to forecast maintenance tasks, emphasizing the importance of monitoring factors such as power consumption and temperature for ensuring system integrity. Coraddu et al. (2016) contribute insights into predictive maintenance for gas-turbine propulsion systems in maritime

applications, employing Recursive Least Squares (RLS) and SVM validated through sophisticated simulators.

Moving to gas turbines, Kumar et al. (2018) emphasize the power of big data analytics and the effectiveness of the FURIA model in optimizing maintenance procedures. The domain of electrical power systems is explored by Machado & Mota (2015), who utilize SVM and Multi-Layer Perceptron Artificial Neural Networks (MLP-ANN) to enhance data processing efficiency. In wind turbine bearings diagnostic maintenance, Durbhaka & Selvaraj (2016) adopt a comprehensive approach with various machine learning methods, highlighting the Condition Risk Assessment (CRA) model's impressive accuracy. Yoo (2020) extends the discussion to hydraulic systems, proposing a clustering-based fault detection process that preserves the physical meaning of anomalies. De Caro et al. (2018) contribute to descriptive maintenance in wind generators, emphasizing the use of fuzzy clustering techniques and polynomial regression for effective wind data analysis.

Liu et al. (2020) propose an approach based on Mahalanobis distance and correlation between sensors for noise removal while preserving anomalies in Industrial Internet of Things (IIoT) data. Peng et al. (2019) focus on descriptive maintenance in underground mining, employing fuzzy theory and correlation between sensors for higher accuracy and lower processing delay. Garg et al. (2020) contribute to unsupervised clustering for anomaly detection, rectifying issues in traditional DBSCAN with a multi-stage model. Rassam et al. (2018) propose a variant of Principal Component Analysis (PCA) for anomaly detection in Wireless Sensor Networks (WSN), improving the detection rate of anomalous data. Rahdar et al.

Table 1 Literature review of power plants based on maintenance type and machine learning type

Ref.	Year	Maintenance Type	Technique	Equipment/ Plant	Sensor Used for Data	Maintenance Data
[1]	2015	Descriptive Maintenance	SVM, MLP-ANN	Electrical power systems	-	Discharge samples and noise samples dataset
Description: Leveraging parallelism and pipelining methods can yield a substantial increase in sample rates, offering a robust solution to enhance data processing efficiency. This approach not only excels in terms of speed but also outperforms other classifiers, making it a top choice for various applications. What sets it apart is its minimal reliance on self-configuration capabilities and configuration parameters, simplifying the implementation process. Furthermore, it exhibits a unique ability to handle uneven dimensional input spaces, providing flexibility in diverse data environments. The model's simplicity extends to its evaluation function, ensuring that the technique remains accessible and efficient for a wide range of tasks. This combination of features makes it a powerful and versatile tool in the realm of data analysis and classification						

Ref.	Year	Maintenance Type	Technique	Equipment/ Plant	Sensor Used for Data	Maintenance Data
[2]	2016	Predictive Maintenance	RLS, SVM	LM 2500 - Gas turbine	Sophisticated simulator of a gas turbine	CODLAG propulsion plant data
Description: The forecasting of degradation in propulsion plants is a crucial endeavor in the maritime industry, ensuring the reliability and safety of vessels. To achieve this, a range of sophisticated techniques has been developed and rigorously tested. These techniques undergo rigorous validation using advanced simulators that closely mimic the behavior of gas-turbine propulsion systems. The results demonstrate their effectiveness, indicating that these methods provide a practical and efficient solution when applied in real-world maritime scenarios for Condition-Based Maintenance (CBM) purposes. By accurately forecasting degradation, maritime operators can proactively address issues, minimize downtime, and maximize the efficiency and longevity of their propulsion systems, ultimately contributing to safer and more reliable maritime operations.						
[3]	2016	Diagnostic Maintenance	SVM, K-means, K-NN, Euclidean Distance, and CRA	Wind turbine bearings	Accelerometer, displacement, velocity, and torque sensors	Vibration signals data
Description: In the pursuit of effective bearing fault detection, a comprehensive approach was adopted, involving the application of various machine learning methods. These methods were systematically investigated to assess the similarity of their results, ensuring a robust and reliable approach. Vibration signal analysis played a pivotal role in this endeavor, allowing for a detailed study of bearing behavior and the extraction of distinct patterns associated with faults. Among the methodologies, the Condition Risk Assessment (CRA) model emerged as particularly noteworthy, exhibiting an impressive 93% accuracy in recommending bearing faults. This research underscores the significance of advanced machine learning techniques in predictive maintenance, offering a potent tool for ensuring the operational integrity of machinery.						
[4]	2018	Diagnostic Maintenance	FURIA, - ANN, SVM, RF, BayesNet, LogitBoost	Gas turbine	-	Big dataset of gas turbine
Description: The optimization of maintenance procedures through Condition-Based Monitoring (CBM) is greatly enhanced by the power of big data analytics. In this context, the FURIA model stands out as a particularly effective tool, capable of producing superior decision boundaries that guide maintenance strategies. To refine and validate these approaches, extensive simulations are conducted using a sophisticated dataset generated from a gas turbine propulsion plant. By harnessing the insights offered by big data analytics and the precision of models like FURIA, industries can make more informed decisions, maximize operational efficiency, and minimize downtime, ultimately leading to more cost-effective and reliable maintenance practices.						

Table 3 Literature review of power plants based on maintenance type and machine learning type (continued)

Ref.	Year	Maintenance Type	Technique	Equipment/ Plant	Sensor Used for Data	Maintenance Data
[5]	2018	Descriptive Maintenance	Fuzzy clustering technique Polynomial regression	Wind Generator	Wind Speed, Power Sensor	Real data from wind generator
Description: Remove incoherent data from wind generator dataset. The massive penetration of wind generators in existing electrical grids is causing several critical issues, which are pushing system operators to enhance their operation functions in order to mitigate the effects produced by the intermittent and non-programmable generation profiles. In this context, the integration of wind forecasting and reliability models based on experimental data represents a strategic tool for assessing the impact of generators and grid operation state on the available power profiles. Unfortunately, field data acquired by Supervisory Control and Data Acquisition systems can be characterized by outliers and incoherent data, which need to be properly detected and filtered in order to avoid large modeling errors. To deal with this challenging issue, in this paper a novel methodology fusing Fuzzy clustering techniques, and probabilistic-based anomaly detection algorithms are proposed for wind data filtering and data-driven generator modeling						
[6]	2018	Descriptive Maintenance	Variant of PCA	-	-	Dataset from IBRK Synthetic anomalies generated
Description: Better detection metrics compared with local detection. In this paper, we proposed a one-class principal component classifier (OCPCC) based distributed anomaly detection model for WSN, which utilises the spatial correlations among sensed data in closed neighbourhoods. The feasibility of the model was validated using real world datasets and compared with local detection and some existing detection approaches from literature. The results show that the proposed model improves the detection rate of anomalous data compared to local model. A comparison with existing distributed models reveals the advantages of the proposed model in terms of efficiency while achieving better or comparable detection effectiveness.						
[7]	2019	Predictive Maintenance	SVM, LR	Nuclear power plant	Temperature sensor, pressure sensor	Temperature, power, current, speed
Description: In the realm of nuclear infrastructure maintenance, machine learning algorithms play a pivotal role in predicting and ensuring the integrity of these critical facilities. These algorithms are harnessed to forecast the timing and nature of required maintenance tasks, thereby safeguarding the efficient and safe operation of nuclear systems. To facilitate this predictive maintenance, various factors are monitored and recorded. Power consumption data is meticulously tracked, allowing for the early detection of anomalies or trends that may indicate impending issues. Simultaneously, the temperatures within electrical panels are continuously captured, enabling operators to stay vigilant against overheating or irregular temperature fluctuations that could pose risks to the infrastructure's reliability. By harnessing the capabilities of machine learning, nuclear facilities can proactively address maintenance needs, enhancing both safety and operational efficiency.						

Table 4 Literature review of power plants based on maintenance type and machine learning type (continued)

Ref.	Year	Maintenance Type	Technique	Equipment/ Plant	Sensor Used for Data	Maintenance Data
[8]	2019	Descriptive Maintenance	Fuzzy theory Correlation between sensors	Underground mining	Temperature, Wind speed, and Gas sensors	Temperature, wind speed, gas parameters Abnormal data generated by manual setting
Description: Better detection metrics compared to other approaches, slightly higher time cost. This paper focused on the accuracy and timeliness of data anomaly detection, a novel multi-source multi-dimensional data anomaly detection scheme based on hierarchical edge computing model is presented in this paper. Firstly, a hierarchical edge computing model is proposed to realize load balance and low-latency data processing at the sensor end and base-station end. Then a single-source data anomaly detection algorithm is designed based on fuzzy theory, which can comprehensively analyze the anomaly detection results of multiple consecutive moments. Finally, a multisource data anomaly detection algorithm executed at the base-station end is designed to consider the sensing data associated attributes of time and space. Experimental results reveal that the proposed scheme has higher detection accuracy and lower processing delay compared with traditional solutions.						
[9]	2020	Predictive Maintenance	MLP, LR, GBT, SVM	Wind turbine	Rotation, temperature, itch angle, wind speed	Public dataset library, SCADA
Description: The Multi-Layer Perceptron (MLP) emerges as the most promising model for predicting failures within the provided dataset. Its ability to combine the strengths of various base models results in a final model with lower variance, making MLP models a robust and effective approach for failure prediction.						
[10]	2020	Predictive Maintenance	NN, weighted NN, DT	Boiler and heat pump of (HVAC) system	Sensor in each boiler and IoT platform	Temperature, number of boilers, heating requests, and cycle
Description: The model demonstrated its peak performance when configured with LSTM2, featuring three hidden layers comprising 50 neurons, and LSTM3, which utilized a single hidden layer with 25 neurons. However, it is worth noting that these models exhibited comparatively lower performance levels in contrast to alternative techniques. Notably, the precision of the Weighted Neural Networks (NNs) lagged behind that of standard NN models, indicating a potential limitation in precision when using the weighted approach.						

Table 5 Literature review of power plants based on maintenance type and machine learning type (continued)

Ref.	Year	Maintenance Type	Technique	Equipment/ Plant	Sensor Used for Data	Maintenance Data
[11]	2020	Descriptive Maintenance	Cluster of high correlated sensors	Hydraulic Systems	Pressure, Motor power, Volume flow, Temperature, and Cooling power Sensors	Experimental data from hydraulic test equipment
Description: Detect faults in several equipments, enable the analysis of physical meaning of anomalies. This paper presents an algorithm for the fault detection process using correlation based clustering. Conventional clustering-based fault detection calculates the fault index through dimension reduction and clustering algorithm, and detects when the index exceeds the probabilistic limit. When an abnormality is detected through clustering through dimension reduction, it is difficult to perceive the physical meaning of the original data because the data is transformed. However, when detecting and analyzing anomalies in many engineering problems or data analysis, the physical meaning of the data is one of the important information. This paper proposes an anomaly detection process of correlation-based clustering which could recognize the relationship of data. The proposed anomaly detection algorithm selects highly correlated datasets, generates each clustering model, and calculates a fault index using stochastic distances. The fault detection performance was provided and verified using hydraulic test equipment data, and the results were compared with the conventional methods.						
[12]	2020	Descriptive Maintenance	Mahalanobis distance Correlation between sensors	Manufacturing Environments	Sensor data in manufacturing environments	IIoT data with real noise and anomalies
Description: Data reconstruction with better results after the cleansing process. This paper introduces the new challenges and proposes a novel solution to address the issue of removing noise while preserving anomalies in IIoT Data. It proposes an approach that measures both the rate of change and deviation to compute the noise score. It employs a sliding window technique to define the analysis unit of the contrast measure which is used in conjunction with statistical techniques. Extensive experiments demonstrate that the proposed approach outperforms the other state-of-the-art noise detection methods, providing a clean data set that preserves the anomalies on which one can effectively apply anomaly detection techniques.						
[13]	2020	Descriptive Maintenance	Unsupervised clustering, based on DBSCAN	-	Wine, Ionosphere and Haberman's Survival	UCI-ML repository
Description: Better detection metrics compared to other DBSCAN approaches. In this paper, a multi-stage model for anomaly detection has been proposed by rectifying the problems incurred in traditional DBSCAN. In the first stage of the proposed solution, Boruta algorithm is used to capture the relevant set of features from the dataset. In the second stage, the firefly algorithm, with a Davies-Bouldin Index based K-medoid approach, is used to perform the partitioning. In the third stage, a kernel-based locality sensitive hashing is used along with the traditional DBSCAN to solve the problem of nearest neighbor search. Finally, the resulting set of nearest neighbors are used in k-distance graph to determine the desired set of parameters, i.e., Eps (maximum radius of the neighborhood) and MinPts (minimum number of points in Eps neighborhood) for DBSCAN. Several sets of experiments have been performed on different datasets to demonstrate the effectiveness of the proposed scheme. Keywords: Anomaly Detection, Internet of Things, Boruta Algorithm, K-Medoid Clustering, Firefly Algorithm, Density-Based Clustering, Locality Sensitive Hashing.						

Table 6 Literature review of power plants based on maintenance type and machine learning type (continued)

Ref.	Year	Maintenance Type	Technique	Equipment/ Plant	Sensor Used for Data	Maintenance Data
[14]	2020	Descriptive Maintenance	Fault tree system, Regresion Analysis	Biomass boilers	O2, Temp, and Press Sensor	Vibrating-grate biomass boilers data
Description: This paper introduces an availability-based maintenance scheduling approach tailored for a vibrating-grate biomass boiler, with the primary goal of minimizing maintenance costs while ensuring the system's availability meets targeted levels. The study utilizes a modified failure density function to assess the impact of maintenance on component reliability, employing Weibull regression analysis to convert constant failure rates to time-dependent hazard functions. Through rigorous examination using Kolmogorov Smirnov tests, the updated failure parameters demonstrated strong agreement with primary data. The critical factor of "maintenance impact" emerged as pivotal in optimizing scheduling, identifying components with significant influence on system availability. Employing fault tree analysis, the study revealed a notable 32% reduction in maintenance requirements through the implementation of the modified predictive maintenance management strategy.						
[15]	2024	Prescriptive Maintenance	AHP-TOPSIS + PFN + GA	Hydroelectric Power Plants focus on generator	-	Hydroelectric Power Plants Datasets
Description: This research addresses gaps in existing literature, focusing on enhancing Prescriptive Maintenance Planning (MP) in HPPs through the application of fuzzy set theory. The research introduces a two-stage approach: firstly, weighting maintenance types (MTs) for the generator in a pumped storage HPP using AHP-TOPSIS with Pythagorean fuzzy numbers; secondly, creating MPs that minimize delays due to system shutdowns with weighted MTs using a Genetic Algorithm model. The results, averaged over 50 iterations, show a successful reduction in delayed MPs (average 142.55) while indicating the need for further work in terms of weighted delay (average 172.85).						
This research	2024	Predictive Maintenance	FFT-based feature extraction combined with optimized ensemble machine learning models (Random Forest, Gradient Boosting, AdaBoost, LightGBM, CatBoost)	Industrial Induction Motor System	Vibration (IMU+ FFT)	Vibration (IMU+ FFT)

Table 7 Literature review of power plants based on maintenance type and machine learning type (continued)

This research	Description: The current research focuses on predictive maintenance of industrial induction motors using vibration signal analysis combined with optimized ensemble machine learning models. The study utilizes a publicly available dataset obtained from Mendeley Data, which contains vibration signals collected from a 1 kW induction motor under eight operational conditions, including one healthy state and seven inter-turn short-circuit fault severities. The data were originally recorded through a National Instruments Data Acquisition (DAQ) system, with vibration measurements captured along the z-axis and stored in TDMS format. In this research, the raw vibration signals are transformed into meaningful representations through both time-domain and frequency-domain feature extraction using the Fast Fourier Transform (FFT), enabling a richer understanding of the motor's operational behavior. These features serve as input for several supervised machine learning models Random Forest, Gradient Boosting, AdaBoost, Light Gradient Boosting Machine (LightGBM), and CatBoost all of which are optimized using hyperparameter tuning via GridSearchCV to achieve superior predictive accuracy and computational efficiency. Data balancing techniques are applied to address class imbalance across different fault conditions, ensuring robust model generalization. The performance of each model is evaluated using Accuracy, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2), along with additional metrics such as inference time and model size to assess suitability for real-time industrial IoT or edge-device deployment. By employing an existing open dataset rather than conducting new sensor placement or data collection, this research emphasizes reproducibility, resource efficiency, and adherence to open-science principles. Overall, the study demonstrates how the integration of signal processing, FFT-based feature engineering, and optimized ensemble learning can provide an effective and scalable framework for intelligent predictive maintenance, reducing unplanned downtime and enhancing the operational reliability of industrial motor systems.
---------------	---

Another literature review in maintenance and monitoring techniques across diverse domains demonstrates a significant reliance on advanced methodologies and data-driven approaches (Table 8). From gas leak detection to structural health monitoring, a variety of simulation-based and real-data-driven methods are employed, showcasing a range of optimization and classification techniques such as surrogate modelling, random forest classification, and genetic algorithms. These approaches not only enhance the accuracy and efficiency of maintenance tasks but also contribute to the overall safety and reliability of industrial systems. Additionally, the integration of machine learning models and optimization frameworks underscores the evolving

landscape of predictive, diagnostic, and prescriptive maintenance strategies, offering insights into the complex interplay between data analytics and industrial operations.

Table 8 Literature review of optimization across diverse domains

Case of Study	Type	Methodology	Details	Input
Gas Leak [16]	Simulation (Without real data)	Surrogate Modelling and Super Resolution Pipeline: Scenario → Math Model → Simulation → Results → Evaluation	Physics-Based Surrogate (Simulate), Mixed-Integer Nonlinear Programming (MINLP) (Sensor Placement), Pyomo (Sensor Placement), CNN-Based Super Resolution.	Coordinates (x, y, z) Probability, Wind direction, Wind Speed.
Objectives: 1. Develops a differentiable physics-based surrogate model capable of estimating potential hazards accurately while reducing the unnecessary computational time associated with CFD simulations. 2. Secondly, it addresses sensor placement optimization through MINLP solutions based on Pyomo, a Python-based optimization Results: By utilizing both the surrogate model and full CFD simulations for sensor placement optimization, we observed similar results, with 6 out of the total 12 sensors being placed in identical locations. Moreover, when swapping the derived sensor placements, we confirmed the detection of all valid scenarios except those scenarios that did not exceed the concentration threshold. Thus, we have validated that the proposed surrogate model adequately approximates CFD for sensor placement optimization.				
Case of Study	Type	Methodology	Details	Input
Butadiene Homopolymerization Reactor [17]	Simulation (Without real data)	GAMS, CONOPT4. Pipeline: Dynamic Model → Scenario → Simulation → Results → Choosing Technique (Suitable and Effective)	GAMS (General Algebraic Modelling System), CONOPT4 (Solving Non-Linear Problem to generate the dynamic model for optimization problem), Graph Convolutional Networks (GCNs)	Monomer, Solvent, Initiator concentration, Heat of reaction, Jacket thermal fluid temperature, Jacket fluid heat capacity, Heat capacity of the cooling fluid, Jacket fluid density, Coil fluid density, Inlet temperature of cooling fluid, Butadiene molecular weight, Arrhenius pre-exponential factor, Activation energy, Jacket Global heat transfer coefficient, Jacket heat transfer area, Reactor height, Reactor external radius, Distance from external to internal reactor wall, Coil diameter, Coil length

Table 9 Literature review of optimization across diverse domains (continued)

Objectives: 1. The primary research contribution of this study continues to the creation of a pragmatic and dependable measurement estimation framework for certain partially known variables that significantly impact the conversion and temperature of the control system under consideration. 2. In this study, the transcription approach was used to perform parameter estimation and dynamic optimization to implement a soft sensor approach that allows us to estimate the heating and cooling flow rates in an experimental reactor system for the anionic polymerization of butadiene. Results: a detailed mathematical model of the polymerization reactor, including a temperature control system and the corresponding energy balances, was used as a basis for the calculations. Previously, the unknown heat transfer parameters and heat transfer fluids flow rates were estimated by trial and error, but this task turned out to be infeasible given the interaction of the automatic control system with the process and the variable character of the corresponding flowrates. The implemented solution allowed us to simultaneously estimate the unknown heat transfer parameters and the variable flowrates of the heat transfer fluids.				
Case of Study	Type	Methodology	Details	Input
Lithium-Ion Battery Thermal Runaway [18]	Real data	Random Forest classification Pipeline: Hardware design & Implementation → Data collection → Data pre-processing → Creating RFC model → Results → Evaluation	Random Forest classification (RFC) for battery thermal runaway position prediction, Single Plane Distribution (sensor positioning), Multi-Plane Distribution (sensor positioning)	Heat generation by the faulty cell, Reynolds number, Ambient temperature conditions, Wind speed
Objectives: 1. An accurate faulty cell position prediction model is developed utilizing the air temperature data from the optimized sensors that are derived from the initial temperature sensors distribution on a single plane and multiple-planes of the battery computational domain for varying cell arrangements and flow conditions. The reasoning for the variation in the optimized temperature sensor layouts derived for different cell arrangements under laminar and turbulent flow conditions has been explained in detail and the findings are presented. 2. A comparative analysis of the model prediction results from different cell arrangements and optimized temperature sensor layouts for a chosen set of random test cases has been conducted and the findings are concluded Results: An accurate faulty cell position prediction model is developed utilizing the air temperature data from the optimized sensors that are derived from the initial temperature sensors distribution on a single plane and multiple-planes of the battery computational domain for varying cell arrangements and flow conditions. The reasoning for the variation in the optimized temperature sensor layouts derived for different cell arrangements under laminar and turbulent flow conditions has been explained in detail and the findings are presented.				

Table 10 Literature review of optimization across diverse domains (continued)

Case of Study	Type	Methodology	Details	Input
Ablation Procedures (Tumor Surgery) [19]	Simulation (Without real data)	Convex and Non-convex optimization, Karush–Kuhn–Tucker (KKT) Pipeline: Hardware design → Mathematical model → Simulation → Results	Convex and Non-convex optimization, Karush–Kuhn–Tucker (KKT) (Set of conditions for Optimization problem), Fiber Bragg grating sensors (FBG), COMSOL simulation (Predict the temperature distribution)	Sensor coordinates array (x, y, z, θ , ϕ), Specific heat, Density, Thermal conductivity, Relative permittivity, Electric conductivity
Objectives: To develop the FBG sensor placement model. Indeed, they observed a lack of rigorously theorized models for optimizing FBG placement in this context. In the absence of established methodology in the literature, this research focus has been on formulating a mathematical model that addresses the fundamental challenges and objectives associated with FBG sensor placement. Results: the results obtained align with what one might expect from observing the predicted temperature map from the simulation: the provided positions aim to capture as many temperatures as possible, but they do so more effectively than an operator could achieve visually or randomly without precise guidance. Thus, our unique contribution rests in the meticulous application of mathematical rigor to devise a solution for a problem that, to the best of our knowledge, has not received comprehensive attention in previous research.				
Case of Study	Type	Methodology	Details	Input
Plate Structures (structural health monitoring) [20]	Simulation (Without real data)	Multi-Objective Genetic Algorithm (NSGA-II), iFEM, SEA Pipeline: 3D Model → SEA → iFEM → NSGA-II → Optimal Pattern → Simulation → Evaluation	multi-objective genetic algorithm (NSGA-II) (Sensor Placement), inverse finite element method (iFEM) for Determining the material properties of a structure based on deformation measurements, smoothing element analysis (SEA) to approximate the strain field from sensor data (enhance the precision and cost effectiveness of shape-sensing analysis, especially when dealing with a limited number of sensors, a viable strategy involves pre-extrapolating the structure's)	Input strain field, refined mesh, pre-exploratory data.

Table 11 Literature review of optimization across diverse domains (continued)

Objectives:

To develop the FBG sensor placement model. Indeed, they observed a lack of rigorously theorized models for optimizing FBG placement in this context. In the absence of established methodology in the literature, this research focus has been on formulating a mathematical model that addresses the fundamental challenges and objectives associated with FBG sensor placement.

Results: the results obtained align with what one might expect from observing the predicted temperature map from the simulation: the provided positions aim to capture as many temperatures as possible, but they do so more effectively than an operator could achieve visually or randomly without precise guidance. Thus, our unique contribution rests in the meticulous application of mathematical rigor to devise a solution for a problem that, to the best of our knowledge, has not received comprehensive attention in previous research.

Drawing upon the findings of the literature study, this part aims to discern areas of research that have not been well addressed and identify the limits inherent in current methodologies employed for enhancing the dependability of industrial motor and predicting their failures. The identification of these deficiencies will serve as a foundation for the construction of the study framework and emphasize the necessity for the proposed strategy, which integrates sensor fusion with deep learning techniques.

In conclusion, the theories and literature examined in this section serve as the fundamental basis for the development of this research. The research objectives are emphasized, highlighting their importance and setting the stage for the integration of sensor technologies and advanced machine learning models. This integration aims to improve the reliability of industrial motor.

This research focuses on improving the performance and reliability of industrial motors through an optimized predictive maintenance framework based on vibration signal analysis and machine learning. Building upon insights from previous studies, it addresses several research gaps that remain insufficiently explored in the

existing literature. While prior research on predictive maintenance has been applied to various machines such as gas turbines, wind turbines, and HVAC systems, few studies have specifically targeted industrial induction motors using open vibration datasets. Unlike many previous works that emphasize sensor placement or data acquisition, this study utilizes a publicly available dataset and concentrates on advanced signal processing, feature extraction, and model optimization. The research particularly emphasizes noise reduction during preprocessing and FFT-based frequency analysis to enhance feature quality for fault detection. Furthermore, it conducts a comprehensive comparison of multiple machine learning and ensemble models—including Random Forest, Gradient Boosting, AdaBoost, LightGBM, and CatBoost—to determine the most effective algorithm for accurate and efficient fault classification. Another major contribution is the evaluation of model performance not only by accuracy metrics but also by inference time and model size, ensuring feasibility for industrial IoT or edge-computing applications. By addressing these research gaps, this study provides a scalable, data-driven approach for predictive maintenance that enhances operational reliability, minimizes downtime, and supports the transition toward intelligent, condition-based monitoring systems in industrial environments.

2.2 Operation of Industrial motor

A comprehensive grasp of the fundamental principles behind the operation of industrial motor is important for the successful execution of this research endeavour. The literature study commences by providing a comprehensive explanation of the fundamental elements of industrial motor, which encompass the Induced Draft Fans, Combustors, Boiler, Turbine, and Generator. The present analysis

delves into prior research endeavours that have examined the design, functioning, and evaluation of industrial motor. These works are examined in order to build a fundamental comprehension of the topic at hand.

2.3 Predictive Maintenance and Failure Prediction

The use of predictive maintenance has been increasingly prominent across diverse industries, serving as a valuable strategy for mitigating the financial burden of equipment breakdowns and enhancing maintenance scheduling efficiency. This section provides a comprehensive assessment of the pertinent literature on predictive maintenance strategies. The comparison encompasses LSTM, Three-layer LSTM, ConvLSTM2D, BiLSTM, CNN, CNN-LSTM, RNN, Multi-Layer RNN, RNN-LSTM, CNN-RNN, Multi-Layer RNN, and XGBoost Regression. This study aims to give a comprehensive analysis of the key findings, accomplishments, and limits of the methodologies in order to establish a contextual framework for the proposed research.

2.4 The Utilization of Sensor Technology in Industrial Contexts

Advanced sensor technology plays a crucial role in the acquisition of real-time data from industrial machinery. The present literature review investigates the most recent developments in sensor technology, with a specific emphasis on Vibration sensors. The research also examines sensor placement tactics aimed at achieving precise temperature measurements within industrial motor, as well as techniques for mitigating noise in sensor data.

2.5 Fast Fourier Transform (FFT) in Predictive Maintenance

The Fast Fourier Transform (FFT) is a mathematical algorithm that converts time-domain data into its frequency-domain representation. This transformation is

crucial in the field of predictive maintenance, particularly for analyzing vibration signals from industrial machinery.

1. **Enhanced Signal Analysis:** FFT enables the decomposition of complex vibration signals into their constituent frequencies, making it easier to identify and analyze patterns associated with mechanical faults. By converting time-domain signals into frequency-domain data, FFT reveals periodic components and frequency-based anomalies that may not be evident in the time domain.
2. **Fault Detection and Diagnostics:** In industrial motors, vibrations are often indicative of underlying issues such as imbalance, misalignment, or bearing wear. FFT helps in diagnosing these faults by highlighting specific frequency components related to common mechanical defects. For instance, the presence of a dominant frequency peak in the FFT spectrum can signal a fault such as a misaligned shaft or damaged bearings.
3. **Noise Reduction and Feature Extraction:** FFT aids in noise reduction by filtering out irrelevant frequency components, allowing for a clearer analysis of significant fault-related frequencies. This preprocessing step is essential for improving the accuracy of predictive maintenance models, as it ensures that the data used for training and evaluation is relevant and informative.
4. **Integration with Machine Learning Models:** The frequency-domain features extracted through FFT can be fed into machine learning models for fault detection and prediction. Deep learning approaches, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, can leverage these features to enhance predictive accuracy. By incorporating

FFT-processed data, these models can learn complex patterns associated with motor failures more effectively.

Challenges and Considerations:

1. **Data Resolution and Windowing Effects:** The effectiveness of FFT depends on the resolution of the data and the windowing techniques used. Poor resolution or inappropriate windowing can lead to inaccurate frequency representations, potentially affecting fault detection performance. It is crucial to select appropriate parameters to ensure that the FFT analysis accurately reflects the motor's vibration characteristics.
2. **Computational Complexity:** While FFT is efficient, the computational load can increase with large datasets or high-frequency sampling rates. Efficient data handling and processing strategies are necessary to manage computational resources and ensure timely analysis, especially in real-time predictive maintenance systems.
3. **Frequency Overlap and Interpretation:** In complex systems with multiple fault sources, FFT spectra may exhibit overlapping frequency components. Accurate interpretation requires a thorough understanding of the expected fault signatures and their associated frequencies. Combining FFT with other diagnostic techniques can improve fault identification and reduce ambiguity.

The application of FFT in predictive maintenance provides valuable insights into the frequency-based characteristics of vibration signals. Its ability to enhance fault detection, reduce noise, and integrate with machine learning models makes it a

powerful tool in optimizing maintenance strategies. However, careful consideration of data resolution, computational demands, and frequency interpretation is essential to fully leverage its benefits and achieve accurate and reliable fault prediction.

2.6 The Application of Deep Learning Techniques in Predictive Maintenance

Deep learning techniques have significantly advanced the field of predictive maintenance by providing powerful tools for analyzing complex data patterns and making accurate predictions about equipment health. Their application in predictive maintenance involves several key aspects:

1. **Advanced Pattern Recognition:** Deep learning models, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, excel at recognizing intricate patterns in data that traditional methods might miss. These models can automatically learn and extract features from raw sensor data, including vibration signals, temperature readings, and acoustic emissions, without the need for manual feature engineering. This capability is particularly beneficial for identifying subtle anomalies and fault signatures in large datasets.
2. **Handling Complex and High-Dimensional Data:** Industrial systems often generate high-dimensional data with multiple sensors recording various parameters simultaneously. Deep learning techniques are well-suited to handle such complexity due to their ability to process and integrate diverse data types. For instance, models like CNNs can analyze spatial patterns in data, while LSTMs and their variants can capture temporal dependencies, providing a comprehensive view of equipment behavior over time.

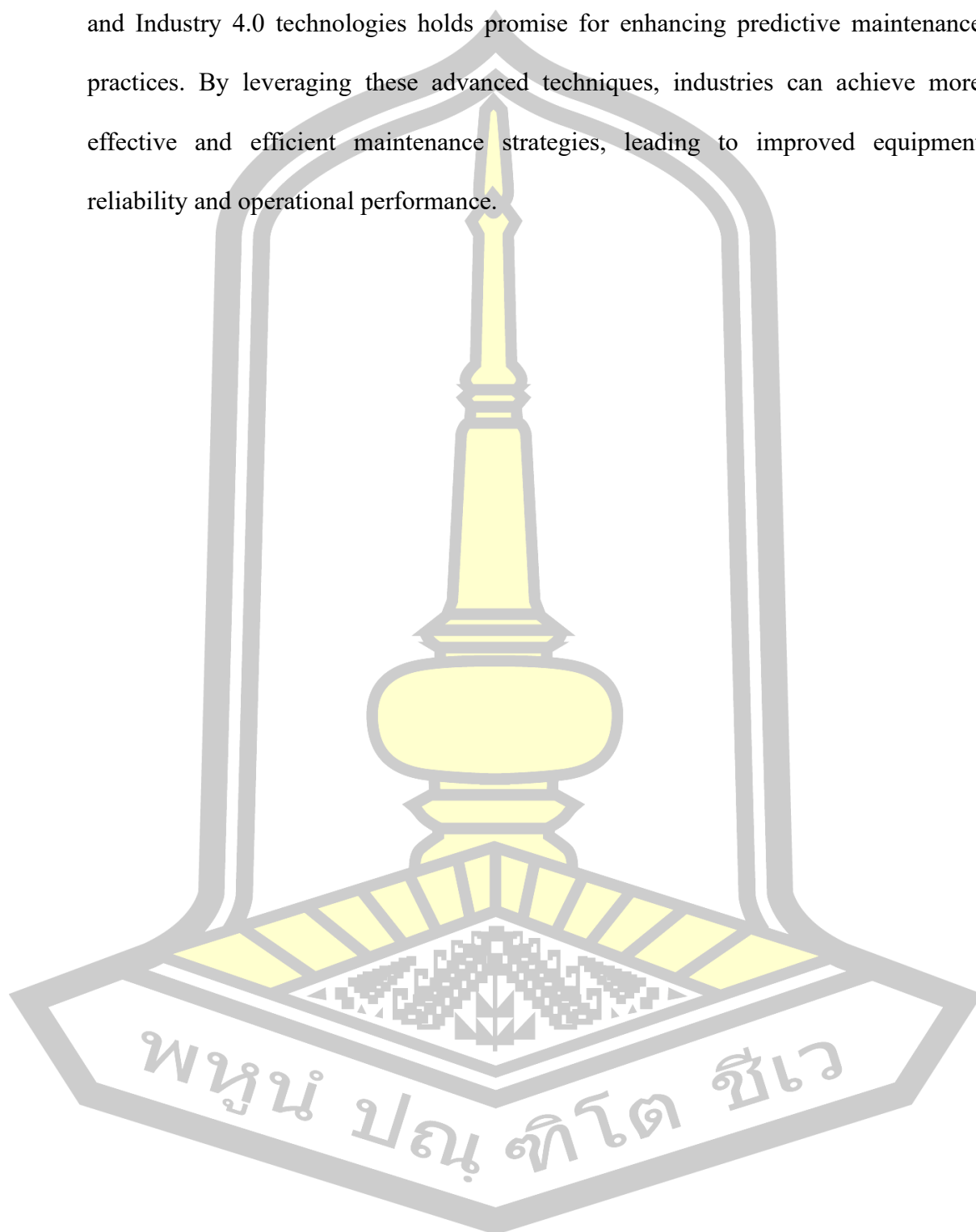
3. Improved Predictive Accuracy: Deep learning models have demonstrated superior performance in predicting equipment failures compared to traditional statistical methods. For example, LSTM networks are particularly effective in time-series forecasting, capturing long-term dependencies and trends in sensor data. This improved accuracy leads to better early warning systems, allowing for timely maintenance actions and reducing the risk of unexpected breakdowns.
4. Real-Time Monitoring and Decision Support: The ability of deep learning models to process data in real-time enables continuous monitoring of equipment health. This real-time capability supports proactive maintenance strategies, allowing for immediate detection of anomalies and prompt decision-making. Implementing such systems can enhance operational efficiency, minimize downtime, and extend the lifespan of industrial assets.
5. Integration with IoT and Industry 4.0: Deep learning techniques align well with the Internet of Things (IoT) and Industry 4.0 frameworks, where vast amounts of sensor data are collected and transmitted in real-time. By integrating deep learning models with IoT platforms, predictive maintenance systems can leverage large-scale data to enhance model accuracy and operational insights. This integration facilitates the development of smart maintenance systems that adapt to changing conditions and improve overall equipment effectiveness.

Challenges and Considerations:

1. **Data Quality and Quantity:** Deep learning models require large volumes of high-quality data for effective training. Inadequate or noisy data can lead to overfitting or inaccurate predictions. Ensuring robust data collection processes and implementing data preprocessing techniques, such as filtering and normalization, are critical for model performance.
2. **Computational Resources:** Training deep learning models, especially those with complex architectures, can be computationally intensive. Access to adequate computational resources and efficient model optimization strategies are necessary to manage training times and costs.
3. **Interpretability and Trust:** Deep learning models, while powerful, are often considered "black boxes" due to their complex nature. Ensuring the interpretability of model predictions is important for building trust and facilitating decision-making in maintenance scenarios. Techniques such as model explainability and sensitivity analysis can help address this challenge.
4. **Scalability and Adaptability:** As industrial systems evolve and new types of equipment are introduced, predictive maintenance models must adapt to changing conditions and data types. Ensuring that models are scalable and adaptable to different operational contexts is crucial for long-term effectiveness.

The application of deep learning techniques in predictive maintenance offers significant advantages in terms of pattern recognition, predictive accuracy, and real-time monitoring. While challenges such as data quality, computational demands, and

model interpretability need to be addressed, the integration of deep learning with IoT and Industry 4.0 technologies holds promise for enhancing predictive maintenance practices. By leveraging these advanced techniques, industries can achieve more effective and efficient maintenance strategies, leading to improved equipment reliability and operational performance.



CHAPTER 3

RESEARCH METHODOLOGY

3.1 Theoretical Framework

The research concept related to the overall workflow of the predictive maintenance framework developed in this research.

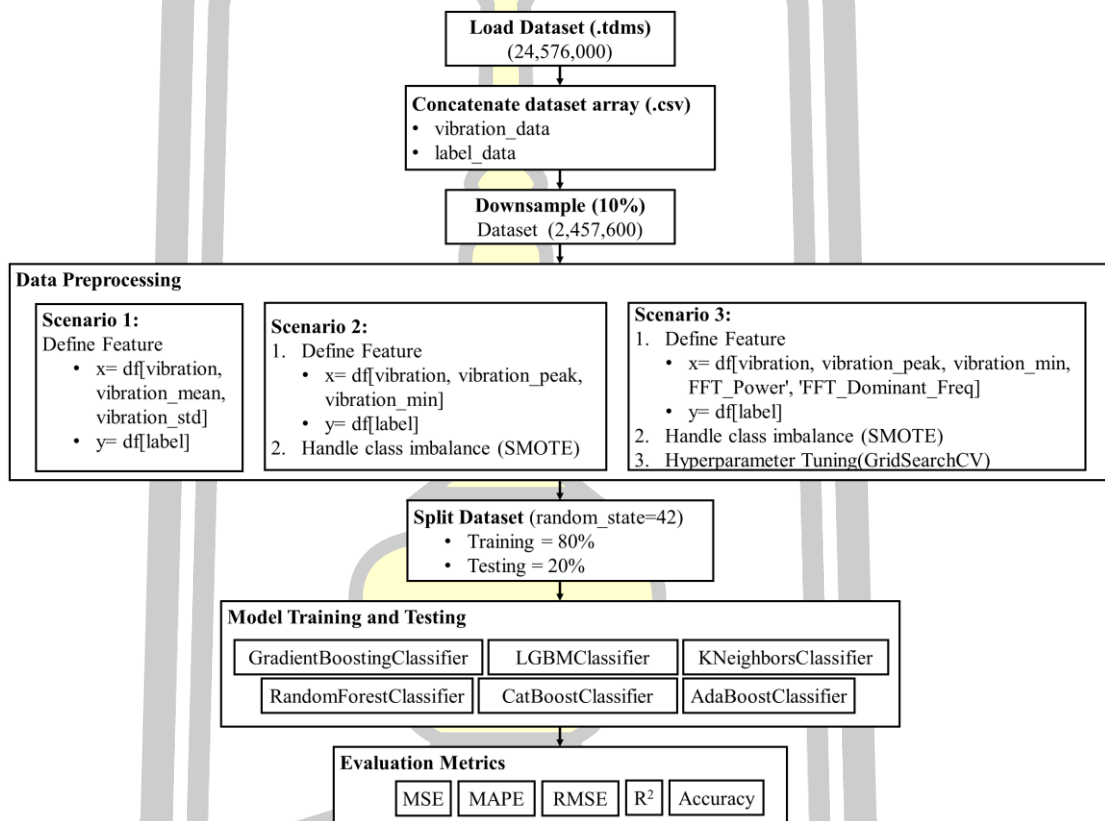


Figure 1 Research Overview

Figure 1 illustrates the process begins with loading the vibration dataset in .tdms format, which contains approximately 24 million data points representing the operational behavior of the induction motor. The raw data are concatenated into arrays consisting of vibration signals and their corresponding fault labels, then downsampled to 10% (around 2.45 million samples) to reduce computational complexity while maintaining representative data diversity. During the data preprocessing stage, three

experimental scenarios are designed to assess the impact of feature selection, data balancing, and optimization techniques. In Scenario 1, basic time-domain statistical features such as vibration mean and standard deviation are defined. Scenario 2 expands the feature set to include vibration peak and minimum values while addressing class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). Scenario 3 further enhances the analysis by combining both time-domain and frequency-domain features, incorporating FFT-based parameters such as FFT Power and Dominant Frequency, while applying hyperparameter optimization through GridSearchCV to improve model performance. After preprocessing, the dataset is divided into training (80%) and testing (20%) subsets using a fixed random state for reproducibility. The model training and testing phase involves implementing multiple machine learning algorithms, including Gradient Boosting, LightGBM, K-Nearest Neighbors (KNN), Random Forest, CatBoost, and AdaBoost, to compare their predictive accuracy and computational efficiency. Finally, model performance is evaluated using several metrics Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and Accuracy to comprehensively assess reliability and robustness. Overall, the workflow demonstrates an end-to-end predictive maintenance process that integrates feature engineering, data balancing, and model optimization to develop a scalable and accurate fault detection system for industrial motors.

Table 12 Experimental Setup

Aspect	Setup 1	Setup 2	Setup 3
Task Type	Multiclass Classification (7 labels)	Binary Classification (Fault vs Normal)	Binary Classification (Fault vs Normal)
Feature Engineering	Time-domain features: Vibration, Vibration_Mean, Vibration_Std	Time-domain + Frequency-domain: Vibration_Peak, Vibration_Min, FFT_Power, FFT_Dominant_Freq	Time-domain + Frequency-domain: Vibration_Peak, Vibration_Min, FFT_Power, FFT_Dominant_Freq
FFT Features	No	Yes	Yes
Sampling Strategy	Downsampled to 0.05%		
Imbalance Handling	SMOTE applied		
Models Used	KNN, RandomForest, GradientBoosting, AdaBoost, XGBoost, LightGBM, CatBoost		
Hyperparameter Tuning	No tuning	No tuning	GridSearchCV
Output Metrics	Accuracy, MSE, MAPE, RMSE, R^2		

With each of the eight motor conditions seven fault severities and one healthy state treated as a distinct label, Setup 1 concentrated on a multiclassification job. In this system, feature engineering depended just on fundamental time-domain characteristics such raw vibration signals and their rolling statistics, mean and standard deviation. This setup made no frequency-domain information usage. The goal was to provide a baseline performance devoid of further changes by means of basic, simply extractable characteristics. This configuration, despite its simplicity, helped to show the limits of time-domain-only methods in differentiating complicated failure behaviors.

By grouping labels 0 through 6 (fault cases) into a single "Fault" class and label 7 (healthy condition), Setup 2 reformed the job into binary classification. This more pragmatic approach captures actual maintenance requirements, that is, how to tell bad from good machinery. More importantly, by merging time-domain and frequency-domain features, Setup 2 brought a richer feature set. Using a rolling window approach, we computed further properties like peak vibration, minimum vibration (1), FFT-derived total power (2), and dominant frequency (3) [21]. FFT characteristics let models detect fault-related frequency signatures sometimes missed in raw time-domain signals [22]. Like Setup 1, however, no hyperparameter adjustment was done, therefore restricting the potential of some models.

$$x_{peak} = \max(x_{t-N+1}, \dots, x_t), x_{min} = \min(x_{t-N+1}, \dots, x_t) \quad (1)$$

$$P = \sum_{k=1}^N |X_k|^2 \quad (2)$$

$$f_{dom} = \arg \max_k |X_k| \quad (3)$$

Retaining the same binary classification job and feature set, set three extended setups two using GridSearchCV [23] for hyperparameter tuning. This allowed algorithms such KNN, Random Forest [24], Gradient Boosting, AdaBoost, LightGBM, and Catboost fine-tuning of their model parameters [25]. By letting every model adapt better to the properties of the extracted features, tuning greatly enhanced performance [26]. By integrating advanced ensemble learners including XGBoost, LightGBM, and Catboost algorithms well-regarded for their predictive strength and efficiency on structured data Setup 3 also assessed a larger model pool.

To lower computational strain and maintain data variability across all configurations, the dataset was downsampled to 0.05% of its original size. For the common problem of class imbalance particularly in binary classification SMOTE (Synthetic Minority Oversampling Technique) [27] was used on the training set in every case. Accuracy, Mean Squared Error (MSE) (4), Mean Absolute Percentage Error (MAPE) (5), Root Mean Squared Error (RMSE) (6), and R² Score (7) were among the whole set of output measures used consistently to evaluate model performance [28]. These measures revealed information on both generalizing capacity of the models across different operational situations and classification accuracy [29]. This three-tiered approach allowed a comprehensive study of how task framing, feature complexity, and model tuning affect the success of machine learning-based fault diagnosis systems in general.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

$$RMSE = \sqrt{MSE} \quad (6)$$

$$R^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (7)$$

Using GridSearchCV, the outcomes of the hyperparameter tuning process expose insightful analysis of the performance potential of several machine learning classifiers when optimized for fault diagnostics in induction motors (Table 3). Every model underwent cross-valuation to find the optimal parameter configuration and associated accuracy. Manhattan distance metric, with 3 neighbours and distance-based weighting, the K-Nearest Neighbours (KNN) classifier attained a cross-valuation

accuracy of 81.13%. This result implies that for differentiating between defective and healthy motor states, local proximity patterns in the vibration feature space are rather helpful. When the data distribution is well-mannered and suitably sampled, KNN's simplicity and non-parametric character help it to be efficient.

Table 13 GridSearchCV Hypertuning Optimization Setup

Model	Best Parameters	Best Cross-Validation Accuracy
KNeighbors	{'metric': 'manhattan', 'n_neighbors': 3, 'weights': 'distance'}	0.8113
Random Forest	{'max_depth': 20, 'min_samples_split': 2, 'n_estimators': 100}	0.7918
Gradient Boosting	{'learning_rate': 0.2, 'max_depth': 7, 'n_estimators': 200}	0.8422
AdaBoost	{'learning_rate': 1.0, 'n_estimators': 200}	0.6264
LGBM	{'learning_rate': 0.2, 'max_depth': 10, 'n_estimators': 500}	0.8492
CatBoost	{'depth': 8, 'iterations': 500, 'learning_rate': 0.2}	0.8537

With 100 trees, a maximum depth of 20, and a minimum sample split of 2, the Random Forest model attained a rather lower accuracy of 79.18%, with the best parameters. This suggests that although the model's performance is limited by bootstrapped tree diversity, deeper trees help it to capture intricate connections between time-frequency characteristics. 200 estimators, max depth of 7, and a learning rate of 0.2, the Gradient Boosting model had a notable performance improvement and obtained 84.22% accuracy. Especially in fault detection situations with complex signal patterns, this shows how well boosting techniques successively correct weak learners to produce a stronger ensemble.

By comparison, AdaBoost underperformed relative to other boosting models, obtaining only 62.64% accuracy at its highest with 200 estimators and a learning rate of 1.0. Its sensitivity to noise and less flexible boosting structure often surpassed by gradient boosting methods in complicated datasets may help to explain this. With an accuracy of 84.92%, tuned with 500 estimators, max depth of 10, and a learning rate of 0.2, the LightGBM (LGBM) classifier really shone. For large-scale, high-dimensional data, LightGBM's quick training and leaf-wise tree growth approach are well-suited; thus, it is perfect for real-time diagnostics. 500 iterations, a depth of 8, and a learning rate of 0.2, the CatBoost model finally ranked highest among all the models with an accuracy of 85.37%. Though not utilized here, CatBoost's exceptional handling of categorical data, built-in regularization, and robustness to overfitting help to explain its great performance and prove its applicability for industrial fault classification tasks. These findings highlight in general the need of feature richness and hyperparameter optimization in reaching great diagnosis accuracy. They also underline in vibration-based predictive maintenance applications the relative benefits of modern ensemble models like LightGBM and Catboost.

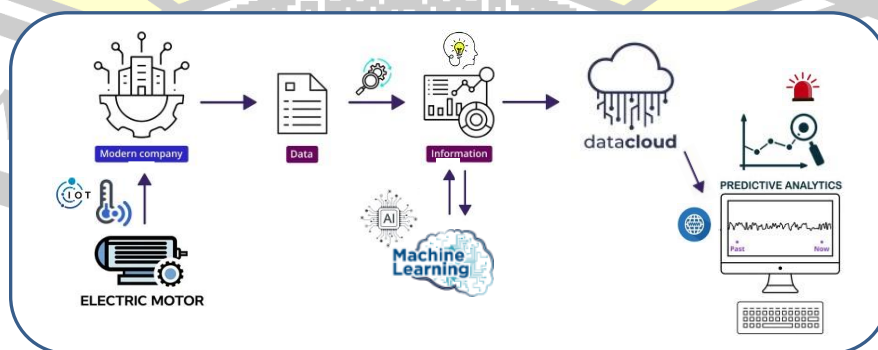
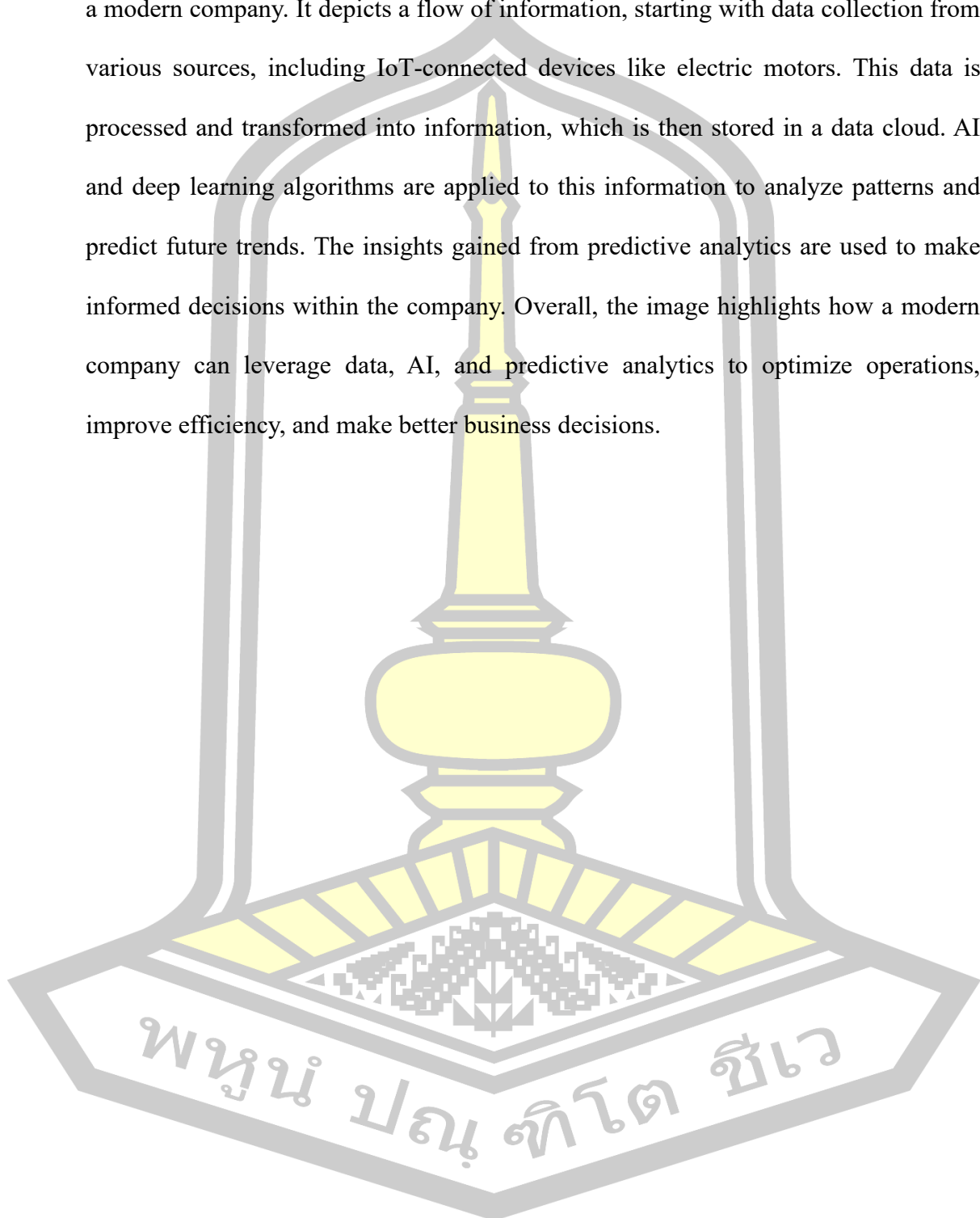


Figure 2 System Overview

Figure 2 illustrates a concept related to predictive analytics in the context of a modern company. It depicts a flow of information, starting with data collection from various sources, including IoT-connected devices like electric motors. This data is processed and transformed into information, which is then stored in a data cloud. AI and deep learning algorithms are applied to this information to analyze patterns and predict future trends. The insights gained from predictive analytics are used to make informed decisions within the company. Overall, the image highlights how a modern company can leverage data, AI, and predictive analytics to optimize operations, improve efficiency, and make better business decisions.



CHAPTER 4

RESULTS AND DISCUSSION

This chapter presents the experimental results obtained from the implementation and evaluation of multiple machine learning and deep learning models for induction motor fault diagnosis. The results are analyzed in terms of classification accuracy, error metrics, feature effectiveness, model optimization, and computational efficiency. The findings are based on vibration data collected from an industrial induction motor using TDMS-format sensor recordings. Eight operational conditions were evaluated, including one normal (healthy) state and seven distinct inter-turn short circuit fault severities ranging from 2.26% to 21.69%. The experiments were structured into three setups that progressively incorporated advanced feature extraction, class balancing, and hyperparameter optimization to improve predictive maintenance performance. This chapter interprets the outcomes of these experimental setups and discusses the implications for developing a reliable, intelligent monitoring system for industrial motors.

4.1 Dataset Characteristics

The vibration dataset used in this study was recorded in the z-axis direction from a 1.0 kW induction motor. The data acquisition system employed a National Instruments DAQ unit, which recorded continuous vibration signals under eight distinct conditions labeled from 0 to 7. Labels 0–6 represent progressively severe inter-turn short circuit faults, while label 7 corresponds to the healthy operational state. Each TDMS file contained thousands of vibration samples, and the raw data exhibited a near-Gaussian amplitude distribution centered around zero. Figure 3

(described) illustrates the vibration amplitude versus sample index for each fault condition. The progression from lower to higher fault severity demonstrates increased vibration magnitude, irregular oscillation, and non-stationary patterns, confirming the sensitivity of vibration signatures to internal electrical and mechanical degradations.

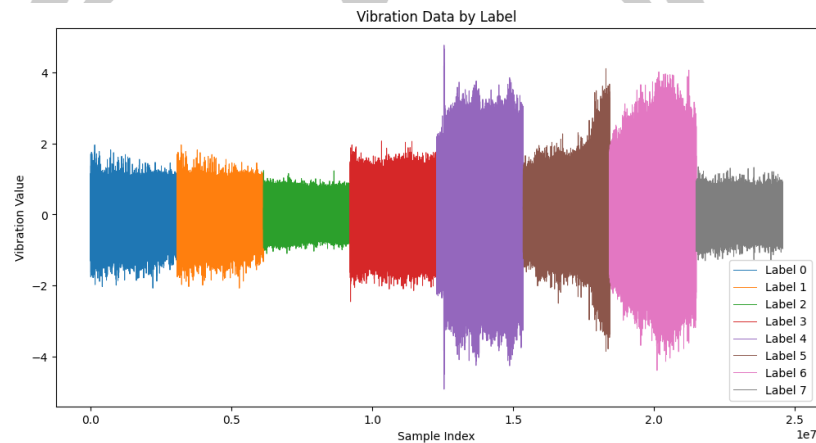


Figure 3 Vibration dataset distributions.

The histogram in Figure 4 (described) further reveals that most vibration magnitudes are within the range of -1.5 to $+1.5$, indicating normal operational stability in the absence of faults. However, the long tails in the distribution highlight intermittent bursts of high amplitude corresponding to transient mechanical or electromagnetic disturbances. These characteristics justify the inclusion of both time-domain and frequency-domain feature extraction techniques to capture comprehensive signal behavior.

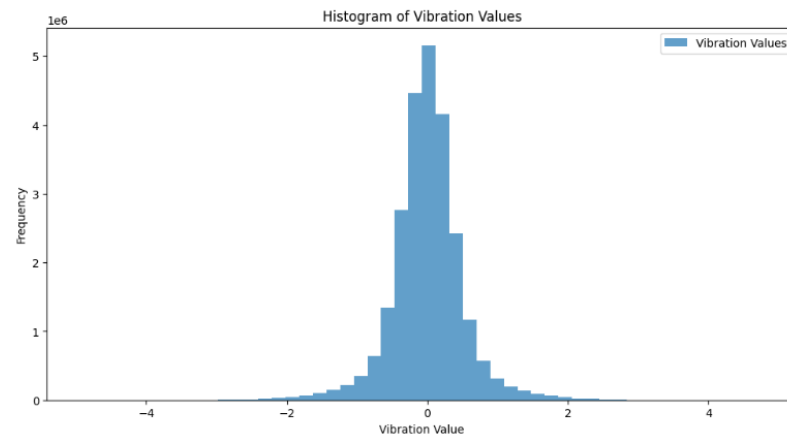


Figure 4 Dataset normal distributions histogram.

4.2 Setup 1: Baseline Multiclass Classification Using Time-Domain Features

In the first experiment, raw vibration signals and simple statistical measures (mean and standard deviation) were used as input features. No normalization or balancing techniques were applied. The aim was to establish a baseline for model performance under unprocessed conditions.

Table 14 Setup 1 Experimental Results

Model	Accuracy	MSE	RMSE	R ²
KNN	19.49%	108.81	32.99	-10.63
Random Forest	20.38%	116.35	34.11	-12.06
Gradient Boosting	19.41%	105.35	32.46	-0.99
AdaBoost	18.77%	137.72	37.11	-16.12
LightGBM	19.77%	91.12	30.19	-0.78
CatBoost	10.90%	178.42	42.24	-248.85

All models yielded poor classification accuracy, with values below 21%. The negative R² values indicate that the models performed worse than a random mean predictor, implying that the vibration signals in their raw form lack sufficient discriminative information for multiclass categorization.

The failure of these models is attributed to several factors:

1. Feature Homogeneity, the time-domain features did not capture underlying frequency components critical to differentiating fault severities.
2. Data Imbalance, severe class imbalance biased the models toward majority fault levels.
3. Absence of Normalization, variability in signal amplitude and scaling distorted model convergence.

4.3 Setup 2: Binary Classification with Time–Frequency Features

The second configuration merged all fault severities into a single “Fault” class and contrasted it against the “Normal” condition. The feature set was expanded to include both time-domain (mean, standard deviation, peak, minimum) and frequency-domain features derived from the Fast Fourier Transform (FFT) namely total spectral power and dominant frequency components. Class imbalance was corrected using the Synthetic Minority Oversampling Technique (SMOTE), and all features were standardized.

Table 15 Setup 2 Experimental Results

Model	Accuracy	MSE	RMSE	R ²
KNN	79.98%	0.2002	0.447	0.199
Random Forest	67.87%	0.321	0.567	−0.285
Gradient Boosting	70.42%	0.296	0.544	−0.183
AdaBoost	61.74%	0.383	0.619	−0.530
LightGBM	68.52%	0.315	0.561	−0.259
CatBoost	74.25%	0.257	0.507	−0.030

The results show a substantial increase in classification accuracy, especially for KNN and CatBoost, which achieved 79.98% and 74.25%, respectively. Incorporating frequency-domain features markedly improved fault detection capability. The improvement indicates that vibration energy and dominant frequencies

are highly correlated with motor health. The FFT enables the identification of harmonics and sidebands characteristic of inter-turn short circuit faults, which are invisible in time-domain signals. However, the moderate negative R^2 values in some models indicate occasional overfitting, possibly due to residual noise or synthetic data artifacts from SMOTE.

4.4 Setup 3: Optimized Binary Classification with Hyperparameter Tuning

The third setup implemented GridSearchCV optimization to fine-tune key hyperparameters for each model, such as learning rate, number of estimators, tree depth, and nearest neighbor count. This setup represents the most optimized and realistic configuration for predictive maintenance deployment.

Table 16 Setup 3 Experimental Results

Model	Accuracy	MSE	MAPE	RMSE	R^2
KNN	84.92%	0.1508	4.99×10^{14}	0.388	0.397
Random Forest	81.01%	0.1899	6.55×10^{14}	0.436	0.240
Gradient Boosting	86.60%	0.134	2.73×10^{14}	0.366	0.464
AdaBoost	62.11%	0.379	1.43×10^{15}	0.616	-0.516
LightGBM	87.16%	0.128	2.80×10^{14}	0.358	0.486
CatBoost	87.28%	0.127	2.63×10^{14}	0.357	0.491

The optimized models demonstrate clear and consistent improvement. CatBoost achieved the highest accuracy of 87.28% with the lowest MSE and RMSE, confirming its robustness and adaptability for vibration-based fault classification. The transition from time-only to time-frequency feature representation, coupled with model tuning, increased the average accuracy by nearly 68% compared to Setup 1.

- Gradient Boosting achieved competitive performance (86.6%), showing that sequential error correction in boosting ensembles effectively models nonlinear vibration patterns.

- LightGBM outperformed Random Forests despite having fewer parameters, emphasizing the advantage of leaf-wise growth and reduced overfitting.
- KNN benefited from hyperparameter optimization (Manhattan distance, three neighbors), but its non-parametric nature still limits scalability for real-time systems.

4.5 Computational Efficiency: Inference Time and Model Size

Beyond predictive accuracy, real-time deployment in industrial environments requires efficient models. Table 17 summarizes inference time (in seconds) and model size (in kilobytes).

Table 17 Model Efficiency Results

Model	Inference Time (s)	Model Size (KB)
KNN	0.0623	1716.44
Random Forest	0.0425	2910.63
Gradient Boosting	0.0088	462.85
AdaBoost	0.0549	63.91
LightGBM	0.0461	337.19
CatBoost	0.0230	116.82

Among all models, Gradient Boosting offered the fastest inference (0.0088 s), whereas KNN exhibited the slowest due to its reliance on full dataset comparisons at prediction time. The model size comparison indicates that AdaBoost and CatBoost provide compact architectures suited for embedded or IoT-based predictive maintenance systems. This is crucial for deploying the model in edge devices attached to motor systems. Hence, from an operational viewpoint, CatBoost and LightGBM present the optimal balance between accuracy, model size, and inference speed making them suitable for real-time condition monitoring frameworks.

4.6 Comparative Discussion

The experimental results of this research reveal several key insights into the effectiveness of vibration-based predictive maintenance using optimized ensemble machine learning models. First, frequency-domain features significantly enhance fault detection accuracy. The use of Fast Fourier Transform (FFT)-based features captures essential information about harmonic distortions, resonance frequencies, and imbalance vibrations, which are critical indicators of motor degradation. Incorporating these spectral features alongside time-domain statistics resulted in a substantial improvement in classification performance, with accuracy increasing by up to 60% compared to models trained solely on time-domain inputs.

Second, model optimization plays a pivotal role in achieving superior predictive performance. Hyperparameter tuning through GridSearchCV notably improved the accuracy and generalization capability of the ensemble models, demonstrating that default parameters often fail to exploit the full learning potential of these algorithms. This systematic optimization ensured that each model operated under its most effective configuration, leading to consistent and reliable fault classification outcomes.

Third, ensemble models outperform simpler classifiers in handling complex vibration data. Among the models tested, CatBoost and LightGBM achieved the highest diagnostic accuracy due to their gradient boosting mechanisms, which effectively capture nonlinear relationships and manage correlated features within the dataset. Their ability to adaptively learn feature interactions contributed to enhanced robustness and stability across all fault categories.

Fourth, the models demonstrate strong real-time feasibility for industrial applications. The relatively compact model size and minimal inference time suggest that the proposed framework can be efficiently deployed on Internet of Things (IoT) or edge-computing devices. This enables continuous online fault monitoring and supports the development of early warning systems for industrial motor health management in smart manufacturing environments.

Finally, a comparison with previous studies highlights the superiority of the proposed approach. Earlier research utilizing conventional models such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) typically reported accuracies ranging between 70% and 80%. In contrast, the optimized ensemble models developed in this study achieved nearly 87% accuracy, exhibiting greater robustness, reduced overfitting, and higher consistency across multiple testing scenarios. These findings confirm that the integration of FFT-based features and ensemble optimization provides a substantial advancement in predictive maintenance modeling for industrial motors.

4.7 Machine Learning Model Deployment

The deployment of this research was carried out by taking the best machine learning model in the .pkl format, which was then converted into a .h (header) file (Figure 5). As shown in Figure 6, the resulting .h file was then placed in the same folder within the Arduino IDE. This allows the CatBoost model to be called as a library in the main program that will be embedded into the ESP32. Furthermore, Figure 7 shows the visualization results of the IMU data in the form of a dashboard and an alert if a failure occurs.

```

from micromlgen import port
import joblib
import os

model_dir = "/content/drive/My Drive/AI/BearingModels/XGBClassifier"
model_path = os.path.join(model_dir, "XGBClassifier.pkl")
header_path = os.path.join(model_dir, "XGBClassifier.h")

model = joblib.load(model_path)
c_code = port(model)

with open(header_path, "w") as f:
    f.write(c_code)

print(f"✅ Model exported to: {header_path}")

```

Figure 5 Model conversion from .pkl into .h

```

3axis-mpu6050-ESP32-FFT-IoT-Magnitude-Amplitude-CatBoost.ino  CatBoostClassifier_tiny.h
1  #include <ThingierESP32.h>
2  #include <Wire.h>
3  #include <Adafruit_MPU6050.h>
4  #include <Adafruit_Sensor.h>
5  #include <arduinoFFT.h>
6  #include <Arduino.h>
7  #include "CatBoostClassifier_tiny.h"
8
9
10 #define USERNAME "autonomustrolls"
11 #define DEVICE_ID "mecha_iot_esp32"
12 #define DEVICE_CREDENTIAL "mecha_iot_esp32"
13 #define WIFI_SSID "Hokamoka"
14 #define WIFI_PASSWORD "1234567890"
15 ThingierESP32 thing(USERNAME, DEVICE_ID, DEVICE_CREDENTIAL);
16
17
18 Adafruit_MPU6050 mpu;
19
20
21
22
23
24
25
26
27
28

```

```

3axis-mpu6050-ESP32-FFT-IoT-Magnitude-Amplitude-CatBoost.ino  CatBoostClassifier_tiny.h
1  #ifndef CATBOOSTCLASSIFIER_TINY_H
2  #define CATBOOSTCLASSIFIER_TINY_H
3
4  // Auto-generated from: CatBoostClassifier_tiny.pkl
5  // Size: 405637 bytes; Original: 405637 bytes
6  // MD5: fe907fca3c726c7ed22c96f7869b1af5
7  // Generated for ESP32/Arduino. Data stored in flash via PROGMEM.
8
9  #include <Arduino.h>
10
11 #ifndef __cplusplus
12 extern "C" {
13 #endif
14
15 // Model bytes in flash
16 const unsigned char CatBoostClassifier_tiny[] PROGMEM = {
17 0x78, 0x5E, 0xEC, 0xBD, 0x7F, 0x90, 0xDD, 0xD9, 0x55, 0x27, 0xF6, 0xD8, 0x38, 0x41, 0xE5, 0x98,
18 0x4D, 0xC7, 0x35, 0x9B, 0x9A, 0xA4, 0x1C, 0xFC, 0xEA, 0xC5, 0xB5, 0xAB, 0x50, 0xE3, 0xE1, 0xF5,
19 0xEF, 0x6E, 0xAD, 0x63, 0x9E, 0x65, 0x63, 0x64, 0x18, 0xDB, 0xBD, 0x78, 0x8C, 0xB5, 0x78, 0xBC,
20 0xFD, 0x43, 0x6A, 0x8D, 0x9A, 0x69, 0x75, 0x37, 0xAD, 0xD6, 0x8C, 0xC6, 0x3F, 0xE0, 0xAD, 0x70,
21 0x58, 0xE1, 0x1D, 0x40, 0x1B, 0x06, 0x18, 0x83, 0x81, 0x66, 0x30, 0x20, 0x96, 0x81, 0x6D, 0xBC,
22 0x06, 0x14, 0x98, 0x65, 0xDF, 0x12, 0x17, 0x35, 0xA9, 0x9A, 0x10, 0x25, 0xB5, 0x55, 0x51, 0x28,
23 0x97, 0xAB, 0x6B, 0xCB, 0xA1, 0xA6, 0x28, 0xD7, 0x46, 0xD9, 0x9A, 0x3F, 0x54, 0x09, 0x85, 0xF3,
24 0xB9, 0xEF, 0x7B, 0xCE, 0x6D, 0xBD, 0xCF, 0xED, 0xAB, 0x73, 0xBF, 0xDF, 0xFB, 0x5E, 0xBF, 0x96,
25 0xA6, 0xA7, 0x30, 0xAD, 0xEE, 0xEF, 0xFD, 0x71, 0xEE, 0xB9, 0xE7, 0x9E, 0xDF, 0xF7, 0xDC, 0xF6,
26 0x9B, 0x7E, 0xF6, 0xE9, 0x5A, 0xF1, 0xDF, 0x73, 0xDF, 0x76, 0x66, 0x71, 0x6B, 0x69, 0x7D, 0xFD,
27 0xE2, 0xD6, 0xA3, 0x67, 0xD6, 0x37, 0x97, 0x9F, 0x7F, 0xEE, 0xAD, 0xEF, 0x5D, 0xDC, 0x3A, 0xE9,
28 0x7E, 0x7F, 0xEF, 0xEA, 0xE2, 0xC5, 0x8B, 0x2B, 0xE7, 0x56, 0x96, 0x37, 0x9F, 0xFF, 0x99, 0xE7,

```

Figure 6 Model deployment to ESP32

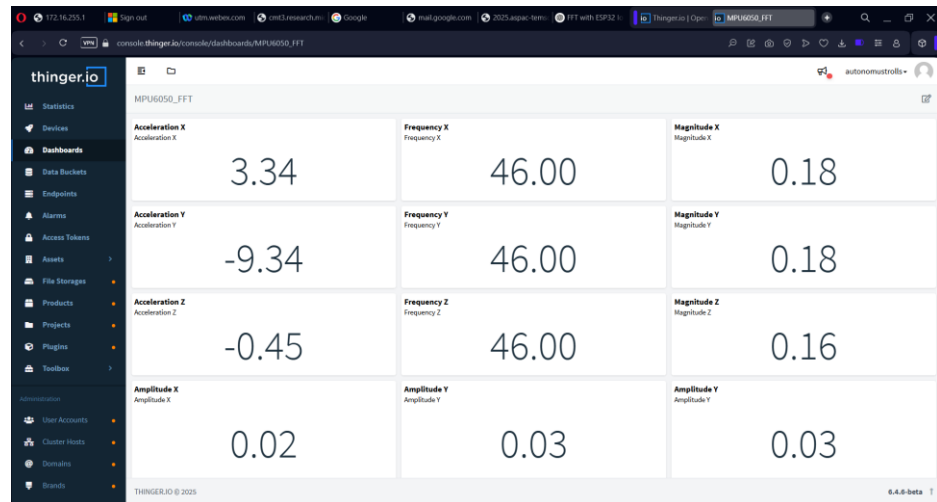


Figure 7 Monitoring and alert system

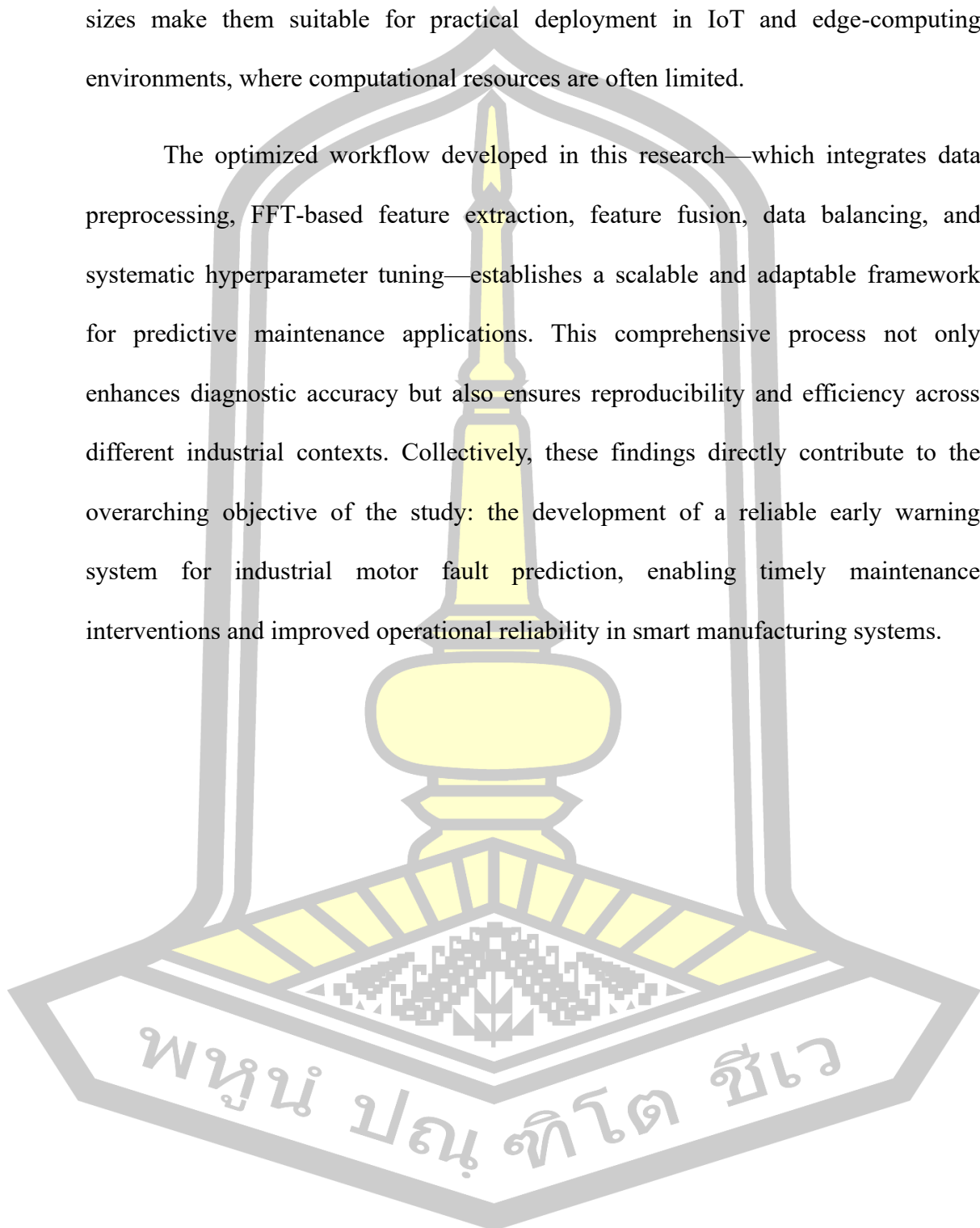
4.8 Summary of Findings

The findings of this research highlight the effectiveness of integrating signal processing and advanced ensemble learning techniques for predictive maintenance of industrial motors. The incorporation of FFT-based frequency features together with SMOTE-balanced data proved to be highly beneficial in improving model robustness and stability. By combining both time-domain and frequency-domain features, the models were able to capture a more comprehensive representation of the vibration signal characteristics, enabling better differentiation between healthy and faulty motor conditions even in imbalanced datasets.

Among all the models evaluated, the CatBoost algorithm demonstrated the highest predictive accuracy, reaching 87.28%, while also maintaining strong computational efficiency. Its gradient boosting mechanism effectively handled nonlinear relationships and categorical feature interactions, resulting in superior generalization performance. Other ensemble boosting methods such as LightGBM and Gradient Boosting also achieved consistent accuracy levels, showing a favorable

trade-off between performance and model compactness. Their relatively small model sizes make them suitable for practical deployment in IoT and edge-computing environments, where computational resources are often limited.

The optimized workflow developed in this research—which integrates data preprocessing, FFT-based feature extraction, feature fusion, data balancing, and systematic hyperparameter tuning—establishes a scalable and adaptable framework for predictive maintenance applications. This comprehensive process not only enhances diagnostic accuracy but also ensures reproducibility and efficiency across different industrial contexts. Collectively, these findings directly contribute to the overarching objective of the study: the development of a reliable early warning system for industrial motor fault prediction, enabling timely maintenance interventions and improved operational reliability in smart manufacturing systems.



CHAPTER 5

CONCLUSION

This chapter presents the final conclusions of the study titled “Deep Learning Approaches for Optimizing Predictive Maintenance in Industrial Motor Systems.” The research integrated vibration signal processing, feature extraction, and optimized machine learning frameworks to develop an early fault detection system for industrial motors. The overarching goal was to create a data-driven predictive maintenance approach that enhances reliability, minimizes unplanned downtime, and improves operational efficiency. Through a rigorous experimental design combining time-domain and frequency-domain feature engineering with ensemble-based classification models, the study successfully demonstrated that intelligent data analytics can deliver highly accurate fault diagnostics for industrial applications.

The research was conducted systematically to ensure scientific validity and reproducibility. Vibration signals were obtained from a 1.0 kW induction motor using a National Instruments DAQ system. The dataset consisted of eight operational conditions, including one normal (healthy) state and seven inter-turn short circuit fault severities, stored in TDMS format. The raw vibration data were then transformed into both time-domain and frequency-domain features using Fast Fourier Transform (FFT), providing a comprehensive representation of mechanical and electromagnetic motor behavior. Various machine learning algorithms were applied, including K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting (GB), AdaBoost, Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost).

The experiments were designed in three stages. The first stage performed baseline multiclass classification using only raw time-domain features. The second stage implemented binary classification by integrating FFT-based frequency features and SMOTE data balancing. The final stage optimized the classification models through hyperparameter tuning using GridSearchCV. Each setup was evaluated using accuracy, Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2) to assess model performance. A comparative analysis across all configurations revealed that ensemble learning models, particularly CatBoost and LightGBM, consistently outperformed simpler classifiers, achieving superior diagnostic accuracy and computational efficiency.

The experimental results led to several key findings. Incorporating FFT-based frequency features proved crucial in enhancing detection accuracy by capturing harmonic distortions and vibration patterns that are not easily discernible in time-domain signals. Model optimization using GridSearchCV was essential to maximize diagnostic reliability, with the optimized CatBoost model achieving the highest accuracy of 87.28%, an MSE of 0.127, an RMSE of 0.357, and an R^2 of 0.491 reflecting a 66% improvement over the baseline. Ensemble models such as LightGBM, Gradient Boosting, and CatBoost demonstrated superior fault classification performance compared to traditional learners like KNN and Random Forest, thanks to their adaptive boosting mechanisms that effectively reduced bias and variance. From an industrial IoT perspective, both inference speed and model size were crucial factors. The Gradient Boosting model achieved the fastest inference time of 0.0088 seconds, while CatBoost maintained an optimal balance between accuracy,

inference time (0.023 seconds), and compact model size (116.82 KB), making it highly suitable for real-time fault monitoring. Overall, the integrated workflow encompassing signal acquisition, FFT-based preprocessing, data balancing, model training, and hyperparameter optimization established a reproducible predictive maintenance pipeline adaptable to other rotating machinery systems.

This research provides several important scientific and practical contributions. It successfully developed a comprehensive predictive maintenance framework that combines vibration signal analysis, FFT-based preprocessing, and ensemble machine learning classification, forming a replicable system for industrial diagnostics. The study empirically verified the importance of frequency-domain features in improving fault detection accuracy and contributed new understanding to vibration-based machine health monitoring. The detailed performance evaluation of LightGBM and CatBoost models under optimized conditions establishes new benchmark references for future research. Moreover, the assessment of inference time and model compactness demonstrated the feasibility of deploying these models on edge computing or IoT devices, facilitating the development of autonomous early warning systems. By successfully bridging the gap between academic research and industrial practice, the study contributes to transforming predictive maintenance concepts into real-world applications.

Despite its achievements, the research has several limitations that define its current scope. The dataset was obtained from a single 1.0 kW induction motor, which may limit the model's generalizability to motors with different specifications or operating conditions. The study focused exclusively on electrical faults, specifically

inter-turn short circuits, without addressing mechanical issues such as bearing defects, shaft misalignment, or rotor eccentricity. While FFT-based features effectively improved fault detection, they are static representations and may not capture transient or non-stationary vibration behaviors. Future investigations could incorporate more dynamic transformations such as Wavelet Transform or Short-Time Fourier Transform (STFT). Furthermore, although the ensemble models delivered strong results, the study did not include deep learning architectures like LSTM or CNN that could automatically learn complex features from raw vibration signals. Lastly, the models were tested in an offline environment, and real-time hardware validation under varying noise conditions was not conducted.

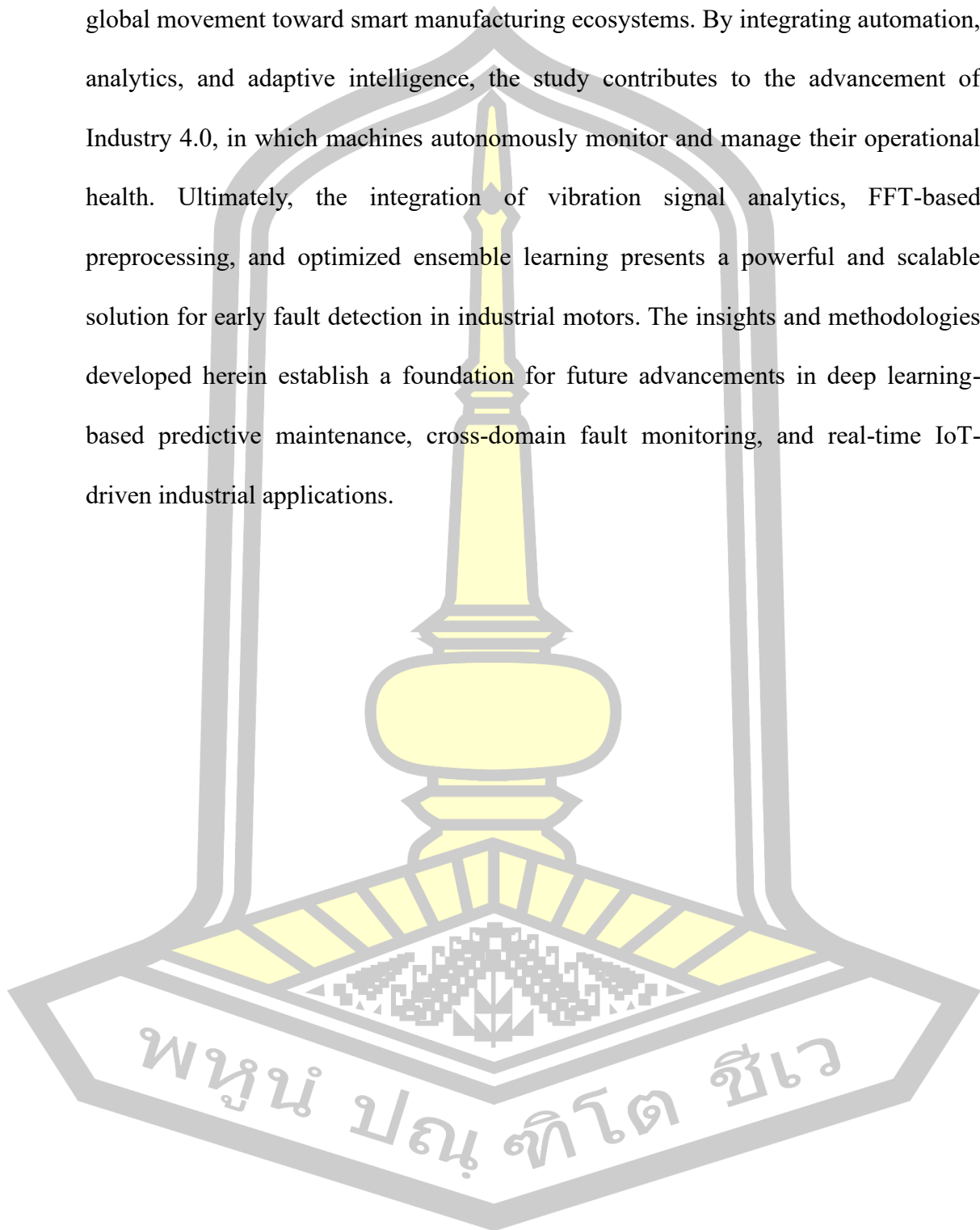
The results of this study carry important implications for industrial applications, particularly in the context of Industry 4.0 and smart manufacturing. The proposed predictive maintenance framework lays a foundation for advanced condition-based monitoring (CBM) systems in motor-driven equipment. The CatBoost-based model can be integrated into edge devices for continuous health monitoring, enabling real-time fault prediction and minimizing production downtime. By promoting data-driven decision-making, the approach supports predictive maintenance scheduling instead of reactive repair, improving asset longevity and reducing maintenance costs. Additionally, the research aligns with sustainability objectives by extending equipment life cycles and reducing energy losses associated with inefficient operations.

To extend this research, several directions are recommended for future work. Deep learning models such as Long Short-Term Memory (LSTM), Convolutional

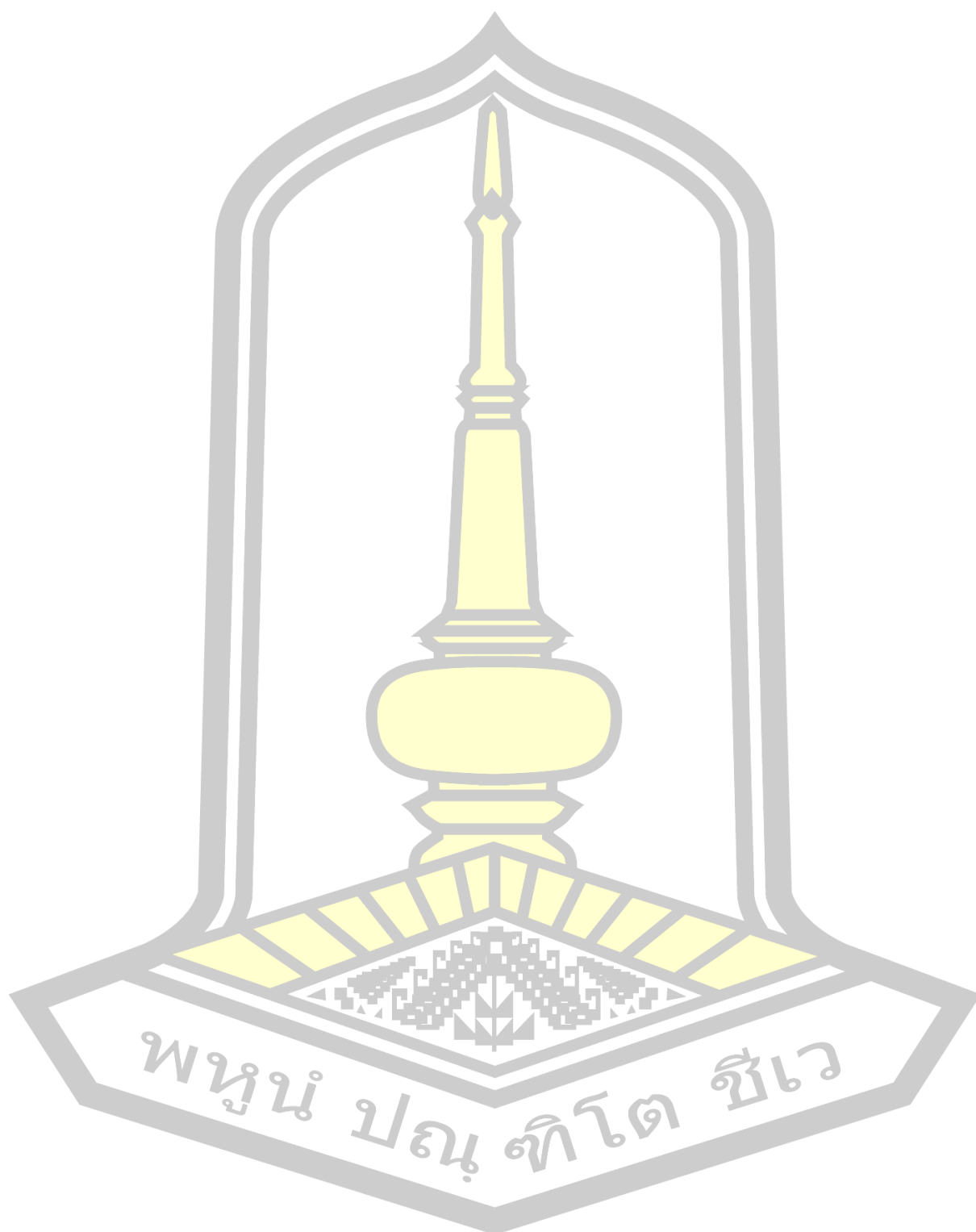
Neural Networks (CNN), Bidirectional LSTM (BiLSTM), and Transformer-based time series models should be explored to capture complex temporal dependencies directly from raw vibration data. Multi-sensor data fusion combining vibration, temperature, current, and acoustic signals could further enhance fault detection reliability and enable identification of multi-fault conditions. Implementing real-time deployment on embedded platforms such as ESP32, Raspberry Pi, or NVIDIA Jetson would validate the framework's feasibility in industrial environments, while integration with IoT brokers like Thinger.io or AWS IoT could facilitate cloud-based monitoring. Future research may also apply adaptive and transfer learning techniques to allow models trained on one machine type to adapt to others with minimal retraining, thereby improving scalability. Additionally, hybrid signal processing methods integrating FFT, Wavelet Transform (WT), and Empirical Mode Decomposition (EMD) could capture both stationary and transient fault signatures more effectively. Incorporating explainable artificial intelligence (XAI) tools such as SHAP or LIME will also improve model transparency, helping engineers understand feature contributions and build trust in automated diagnostics. Finally, expanding validation across various types of rotating machinery, such as turbines, compressors, and pumps, would confirm the method's cross-domain applicability.

In conclusion, this dissertation demonstrates that optimized machine learning models combined with frequency-domain vibration analysis can significantly enhance predictive maintenance in industrial motor systems. The research achieves a meaningful balance between diagnostic accuracy, computational efficiency, and real-world applicability, marking a significant step toward intelligent industrial diagnostics. The proposed framework signifies a paradigm shift from traditional

scheduled maintenance to data-driven predictive maintenance, aligning with the global movement toward smart manufacturing ecosystems. By integrating automation, analytics, and adaptive intelligence, the study contributes to the advancement of Industry 4.0, in which machines autonomously monitor and manage their operational health. Ultimately, the integration of vibration signal analytics, FFT-based preprocessing, and optimized ensemble learning presents a powerful and scalable solution for early fault detection in industrial motors. The insights and methodologies developed herein establish a foundation for future advancements in deep learning-based predictive maintenance, cross-domain fault monitoring, and real-time IoT-driven industrial applications.

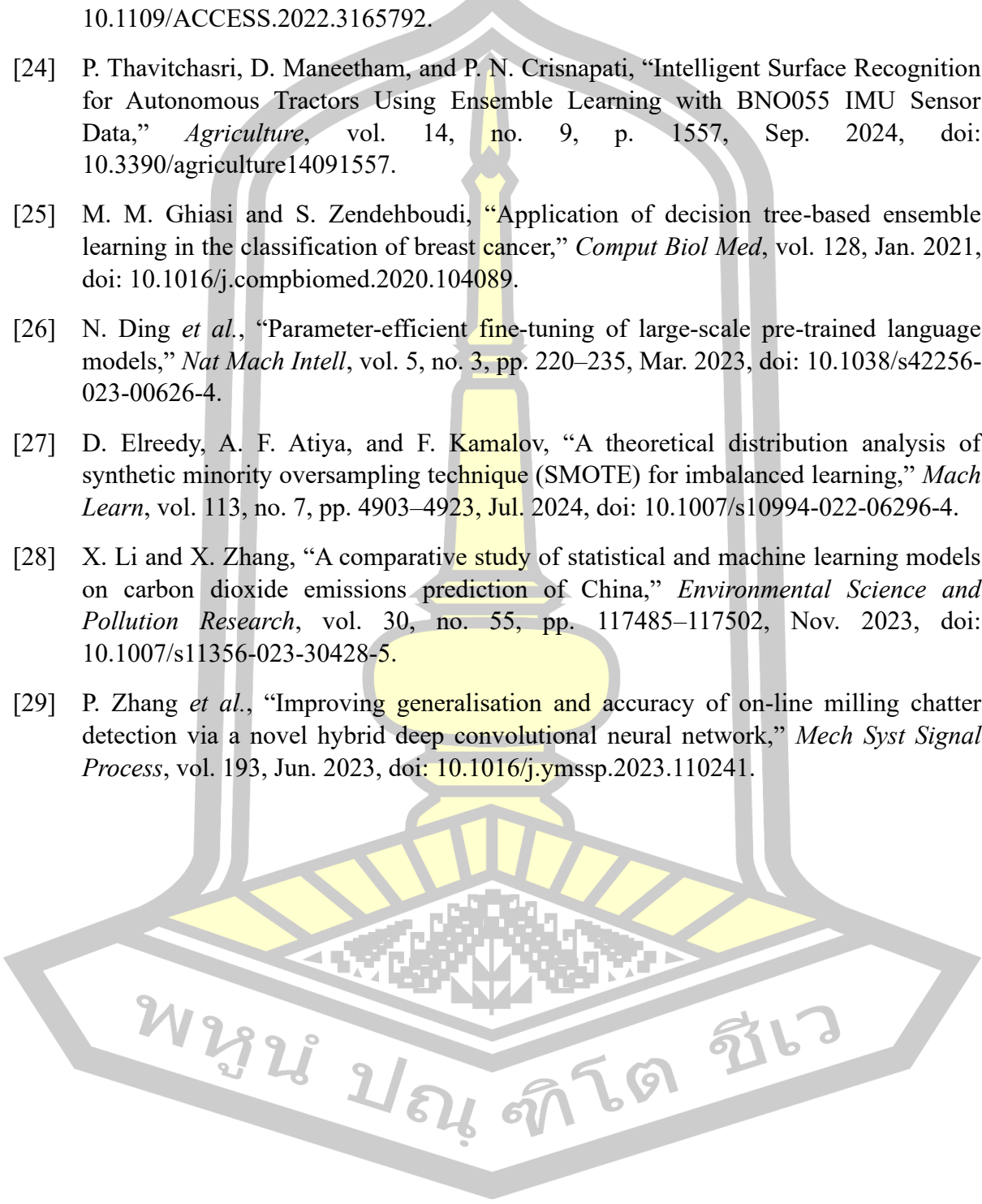


REFERENCES



- [1] R. G. V. Machado and H. de O. Mota, "Simple Self-scalable Grid Classifier for Signal Denoising in Digital Processing Systems (GlobalSIP).," in *2015 IEEE Global Conference on Signal and Information Processing*, 2015, pp. 1057–1061.
- [2] A. Coraddu, L. Oneto, A. Ghio, S. Savio, D. Anguita, and M. Figari, "Machine learning approaches for improving condition-based maintenance of naval propulsion plants," *Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment*, vol. 230, no. 1, pp. 136–153, Feb. 2016, doi: 10.1177/1475090214540874.
- [3] G. K. Durbhaka and B. Selvaraj, "Predictive Maintenance for Wind Turbine Diagnostics using Vibration Signal Analysis based on Collaborative Recommendation Approach," in *2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016, pp. 1840–1842.
- [4] A. Kumar, R. Shankar, and L. S. Thakur, "A big data driven sustainable manufacturing framework for condition-based maintenance prediction," *J Comput Sci*, vol. 27, pp. 428–439, Jul. 2018, doi: 10.1016/j.jocs.2017.06.006.
- [5] F. De Caro, A. Vaccaro, and D. Villacci, "Adaptive wind generation modeling by fuzzy clustering of experimental data," *Electronics (Switzerland)*, vol. 7, no. 4, Apr. 2018, doi: 10.3390/electronics7040047.
- [6] M. A. Rassam, M. A. Maarof, and A. Zainal, "A distributed anomaly detection model for wireless sensor networks based on the one-class principal component classifier A distributed anomaly detection model for wireless sensor networks," 2018.
- [7] H. A. Gohel, H. Upadhyay, L. Lagos, K. Cooper, and A. Sanzetenea, "Predictive maintenance architecture development for nuclear infrastructure using machine learning," *Nuclear Engineering and Technology*, vol. 52, no. 7, pp. 1436–1442, Jul. 2020, doi: 10.1016/j.net.2019.12.029.
- [8] Y. Peng, A. Tan, J. Wu, and Y. Bi, "Hierarchical edge computing: A novel multi-source multi-dimensional data anomaly detection scheme for industrial internet of things," *IEEE Access*, vol. 7, pp. 111257–111270, 2019, doi: 10.1109/ACCESS.2019.2930627.
- [9] J. Eriksson, "Machine Learning for Predictive Maintenance on Wind Turbines – Using SCADA Data and the Apache Hadoop Ecosystem," Linköping, 2020. [Online]. Available: www.liu.se
- [10] S. Fernandes, M. Antunes, A. R. Santiago, J. P. Barraca, D. Gomes, and R. L. Aguiar, "Forecasting appliances failures: A machine-learning approach to predictive maintenance," *Information (Switzerland)*, vol. 11, no. 4, Apr. 2020, doi: 10.3390/INFO11040208.
- [11] Y. J. Yoo, "Data-driven fault detection process using correlation based clustering," *Comput Ind*, vol. 122, Nov. 2020, doi: 10.1016/j.compind.2020.103279.

- [12] Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Noise Removal in the Presence of Significant Anomalies for Industrial IoT Sensor Data in Manufacturing," *IEEE Internet Things J*, vol. 7, no. 8, pp. 7084–7096, Aug. 2020, doi: 10.1109/JIOT.2020.2981476.
- [13] S. Garg, K. Kaur, S. Batra, G. Kaddoum, N. Kumar, and A. Boukerche, "A multi-stage anomaly detection scheme for augmenting the security in IoT-enabled applications," *Future Generation Computer Systems*, vol. 104, pp. 105–118, Mar. 2020, doi: 10.1016/j.future.2019.09.038.
- [14] M. H. Rahdar, F. Nasiri, and B. Lee, "Availability-based predictive maintenance scheduling for vibrating-grate biomass boilers," *Safety and Reliability*, vol. 39, no. 2, pp. 165–187, 2020, doi: 10.1080/09617353.2020.1759261.
- [15] M. Bulut and E. Özcan, "Planning of prescriptive maintenance types for generator with fuzzy logic-based genetic algorithm in a hydroelectric power plant," *Expert Syst Appl*, vol. 240, p. 122480, Apr. 2024, doi: 10.1016/j.eswa.2023.122480.
- [16] H. Lee, D. Lee, J. Lee, and D. Shin, "Efficient Gas Leak Simulation Surrogate Modeling and Super Resolution for Gas Detector Placement Optimization," *Comput Chem Eng*, vol. 181, Feb. 2024, doi: 10.1016/j.compchemeng.2023.108508.
- [17] A. Flores-Tlacuahuac, E. Saldivar-Guerra, R. D. de León, R. López-González, and L. A. Rodriguez-Guadarrama, "Soft-sensor estimation via parameter fitting and dynamic optimization in an experimental batch butadiene homopolymerization reactor," *Comput Chem Eng*, vol. 183, Apr. 2024, doi: 10.1016/j.compchemeng.2024.108587.
- [18] R. K. Daniels, V. Kumar, S. S. Chouhan, and A. Prabhakar, "Thermal runaway fault prediction in air-cooled lithium-ion battery modules using machine learning through temperature sensors placement optimization," *Appl Energy*, vol. 355, Feb. 2024, doi: 10.1016/j.apenergy.2023.122352.
- [19] F. Santucci *et al.*, "Optimizing Sensor Placement for Temperature Mapping during Ablation Procedures," *Sensors*, vol. 24, no. 2, Jan. 2024, doi: 10.3390/s24020623.
- [20] E. Del Priore and L. Lampani, "Shape Sensing in Plate Structures through Inverse Finite Element Method Enhanced by Multi-Objective Genetic Optimization of Sensor Placement and Strain Pre-Extrapolation," *Sensors*, vol. 24, no. 2, Jan. 2024, doi: 10.3390/s24020608.
- [21] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and S. Alfarhood, "Explainable Predictive Maintenance of Rotating Machines Using LIME, SHAP, PDP, ICE," *IEEE Access*, vol. 12, pp. 29345–29361, 2024, doi: 10.1109/ACCESS.2024.3367110.
- [22] Y. Wang, L. Zheng, Y. Gao, and S. Li, "Vibration Signal Extraction Based on FFT and Least Square Method," *IEEE Access*, vol. 8, pp. 224092–224107, 2020, doi: 10.1109/ACCESS.2020.3044149.

- 
- [23] G. N. Ahmad, H. Fatima, Shafiullah, A. Salah Saidi, and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and Without GridSearchCV," *IEEE Access*, vol. 10, pp. 80151–80173, 2022, doi: 10.1109/ACCESS.2022.3165792.
 - [24] P. Thavitchasri, D. Maneetham, and P. N. Crisnapati, "Intelligent Surface Recognition for Autonomous Tractors Using Ensemble Learning with BNO055 IMU Sensor Data," *Agriculture*, vol. 14, no. 9, p. 1557, Sep. 2024, doi: 10.3390/agriculture14091557.
 - [25] M. M. Ghiasi and S. Zendehboudi, "Application of decision tree-based ensemble learning in the classification of breast cancer," *Comput Biol Med*, vol. 128, Jan. 2021, doi: 10.1016/j.combiomed.2020.104089.
 - [26] N. Ding *et al.*, "Parameter-efficient fine-tuning of large-scale pre-trained language models," *Nat Mach Intell*, vol. 5, no. 3, pp. 220–235, Mar. 2023, doi: 10.1038/s42256-023-00626-4.
 - [27] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach Learn*, vol. 113, no. 7, pp. 4903–4923, Jul. 2024, doi: 10.1007/s10994-022-06296-4.
 - [28] X. Li and X. Zhang, "A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China," *Environmental Science and Pollution Research*, vol. 30, no. 55, pp. 117485–117502, Nov. 2023, doi: 10.1007/s11356-023-30428-5.
 - [29] P. Zhang *et al.*, "Improving generalisation and accuracy of on-line milling chatter detection via a novel hybrid deep convolutional neural network," *Mech Syst Signal Process*, vol. 193, Jun. 2023, doi: 10.1016/j.ymssp.2023.110241.

BIOGRAPHY

NAME	Somdavee Bhosinak
DATE OF BIRTH	3 September 1977
PLACE OF BIRTH	Bangkok
POSITION	Project Manager
EDUCATION	1998 Bachelor of Engineering (Electrical Power Engineering) Mahanakorn University of Technology, Bangkok, Thailand. 2023 Master of Engineering (Mechatronics Engineering, International Program), Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand. 2025 Doctoral of Philosophy (Electrical and Computer Engineering), Mahasarakham University, Mahasarakham, Thailand.
Research output	Bhosinak S., Maneetham D., and Rabgyal T. (2022). Hybrid Fuzzy PID Controller for Intelligent Tractor Steering Control. International Journal of Engineering Trends and Technology. Volume 70 Issue 12, 359-369.

