POLITECNICO DI TORINO

Master's Degree in Data Science & Engineering

Master's Degree Thesis

Automated Functional Testing of Brushless Electric Fans in Vehicle Applications: A Data-Driven Approach

Supervisors

Prof. Daniele APILETTI

Company Tutor

Sergio CANTA

Candidate

Suyash SINGH

Feb 2025

Summary

Automated functional testing is necessary to guarantee the dependability of crucial components due to the growing complexity of contemporary automotive systems. A data-driven method for identifying brushless electric fan failures in automotive applications is presented in this thesis. The objective is to use anomaly detection and machine learning approaches to optimize failure detection methodologies. Each of the two stages of the study uses a different failure categorization strategy based on DIAG_M1 diagnostic data.

A mixed DIAG_M1 dataset was examined in Phase 1, where failure identification necessitated assessing a number of factors, including motor current, voltage, temperature, and RPM behavior, rather than relying just on DIAG_M1 values. To differentiate between various operating modes (such as resting and starting), sophisticated feature engineering approaches were used. For unsupervised anomaly identification, Principal Component Analysis (PCA) and DBSCAN clustering were used to find intricate failure patterns that went beyond common diagnostic indicators.

Failure was clearly defined in Phase 2 using a binary classification technique, with DIAG_M1 = 0 denoting no failure and DIAG_M1 = 4 denoting failure. This approach reduced the complexity observed in Phase 1 by enabling a simple failure categorization. The most influential parameters were found using Random Forest classification, while outliers and anomalies in the fan performance data were found using K-Nearest Neighbors (KNN) and DBSCAN.

The results show that each strategy has distinct benefits, with Phase 2 delivering a quicker, rule-based categorization and Phase 1 offering a more thorough failure analysis. This research opens the door for predictive maintenance solutions in car cooling systems by improving the capacity to identify abnormalities early through

the integration of machine learning techniques. By enhancing brushless electric fans' dependability and efficiency, these techniques guarantee peak performance in a range of operating scenarios.

The project was carried out by Johnson Electric in Italy, a top supplier of motion solutions for the automotive sector worldwide. With a focus on **high-performance motors, actuators, and electromechanical systems** for use in *consumer, commercial, industrial, and automotive applications*, it is a **world leader in motion solutions**. The business offers *specialized motion solutions* that improve **sustainability, dependability, and efficiency** with a strong emphasis on **innovation and precision engineering**. Its plant in **Italy** is a major player in the production of *automotive components*, combining **AI-driven diagnostics, predictive maintenance, and Industry 4.0 technologies** to guarantee excellent product performance and quality requirements.

To improve failure identification and predictive maintenance, the study used machine learning techniques such as PCA, DBSCAN, and KNN anomaly detection. Vehicle cooling systems are now more reliable and efficient because to the system's capacity to detect anomalies and failures beyond the conventional DIAG M1 classification by evaluating real-world sensor data. The findings confirm that functional testing may be greatly enhanced by AI-driven diagnostics, guaranteeing better performance and less downtime for automotive components.

Acknowledgements

I would like to express my sincere gratitude to my supervisor **Daniele APILETTI** for their invaluable guidance, support, and insights throughout this research.

Also, I would like to thank **Johnson Electric, Italy**, and **Sergio CANTA** for his invaluable support and for supplying the *materials, technological know-how, and research facilities* required for this project's successful completion. The creation of this study has been significantly aided by their dedication to **excellence and innovation** in motion solutions. In order to improve the *functional testing techniques* and guarantee the efficacy of the **machine learning-based failure detection system**, Johnson Electric's *engineering and research teams* provided invaluable support and direction.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

For contemporary cars to have the best engine performance, temperature control, and overall safety, cooling systems must be dependable. Brushless electric fans are essential for controlling the temperature of the engine, radiator, and battery systems, among other parts of a car. Any issue with these fans might result in serious mechanical failures, overheating, or decreased efficiency [1].

Efficiency, consistency, and adaptation to contemporary automobile requirements are frequently lacking in traditional manual and semi-automated testing techniques for these fans. The move to automated functional testing has become more significant with the introduction of Industry 4.0, allowing for predictive maintenance, real-time failure detection, and enhanced product quality. This thesis aims to provide an automated testing framework that uses data-driven methods and machine learning to precisely detect abnormalities and faults in brushless electric fans used in automotive applications.

1.2 Problem Statement

Static failure thresholds and predetermined circumstances are frequently used in current diagnostic approaches for automotive cooling systems, which may not always be able to identify intricate failure patterns. DIAG_M1 is one of the most often utilized diagnostic parameters; it functions as an error indication but does not always offer a clear-cut and definitive failure categorization.

Two major issues in fan failure detection are addressed in this project:

Phase 1 (Mixed DIAG_M1 Data): DIAG_M1 values range from 0 to >0, hence this number alone cannot be used to identify problems. Rather, a multi-parameter strategy is needed, in which the behavior of the motor's current, voltage, temperature, and RPM is examined all at once.

Phase 2 (Binary DIAG_M1 Data): A simpler binary classification is utilized in this phase, where failure is denoted by DIAG_M1 = 4 and no failure is denoted by DIAG_M1 = 0. By doing so, uncertainty is removed and a rule-based failure detection system is made possible.

Using machine learning techniques like Principal Component Analysis (PCA), Random Forest classification, DBSCAN clustering, and K-Nearest Neighbors (KNN), [2] this study attempts to create a reliable, automated testing methodology for predictive maintenance and real-time failure detection of electric fans in automobiles.

1.3 Objectives

Creating an automated functional testing framework for brushless electric fans used in automotive applications is the main goal of this project. Accurate and effective failure detection techniques are becoming essential due to the growing complexity of contemporary automobile systems. Conventional failure detection techniques depend on preset failure thresholds, which might not always be able to identify intricate failure patterns in practical situations. This thesis suggests a data-driven strategy that makes use of machine learning techniques to more accurately and automatically evaluate, detect, and categorize errors.

In order to do this, the research is split into two experimental stages, each of which focuses on a distinct facet of failure detection. Phase 1 focuses on a mixed DIAG_M1 dataset, meaning that DIAG_M1 values alone cannot be used to identify errors. Rather, a variety of diagnostic characteristics, such as motor current, voltage, temperature, and RPM behavior, are analyzed in order to classify failures. To find abnormalities, sophisticated feature engineering and unsupervised learning methods like Principal Component Analysis (PCA) and DBSCAN clustering are used. The goal of this phase is to provide a multi-parameter framework for failure detection that goes beyond conventional threshold-based techniques to detect complicated problems.

However, Phase 2 presents a more straightforward binary classification method in which DIAG_M1 = 0 denotes no failure and DIAG_M1 = 4 denotes a failure. This method removes uncertainty and offers a clear mechanism for classifying failures. However, Random Forest classification is used to determine the most important variables that affect failure detection because DIAG_M1 alone might not fully capture the image. Furthermore, the dataset's failure patterns and anomalies are categorized using DBSCAN clustering and K-Nearest Neighbors (KNN). In contrast to Phase 1's multi-parameter approach, this phase seeks to assess the efficacy of a rule-based categorization system.

This study aims to improve the effectiveness, precision, and dependability of functional testing in car cooling systems by integrating feature selection, anomaly detection, and classification methodologies. Additionally, by advancing predictive maintenance techniques, the study hopes to lower maintenance expenses, downtime, and unanticipated breakdowns in the automobile sector. The study's conclusions may be applied to other electromechanical parts, enhancing the overall dependability and performance of automotive systems.

1.4 Scope of the Study

The automated functional testing of brushless electric fans used in automobile applications is the main emphasis of this work, especially when it comes to short-term functional testing (24–48 hours). The goal is to provide a data-driven framework for failure detection that can use machine learning techniques to categorize failures and find abnormalities. By utilizing clustering approaches, feature selection methods, and sophisticated data analytics, the study seeks to improve the dependability and efficiency of car cooling systems.

Brushless electric fans are put through precise start-stop sequences determined by electrical and electronic product requirements and actual failure patterns that arise during vehicle operation as part of the functional testing procedure in this study, which is carried out in a controlled laboratory setting. Failure detection is not exclusively reliant on DIAG_M1 readings, however the major diagnostic parameter (DIAG_M1) functions as a first failure signal. To more precisely identify faults, a variety of operating characteristics are examined, including motor current, voltage, temperature, and RPM behavior.

Each of the two stages that make up this research has a unique method for classifying failures:

Phase 1 (Mixed DIAG_M1 Data Analysis): A dataset with DIAG_M1 values ranging from 0 to >0 is analyzed in this phase, which complicates failure detection. Failure categorization in this case is complex and necessitates multi-parameter study. To ascertain which diagnostic characteristics are most important for failure detection, advanced feature engineering is carried out. Principal Component Analysis (PCA) and DBSCAN clustering are examples of unsupervised learning techniques that are used to identify probable failures and abnormalities that may not be seen using conventional diagnostic methods[3]

Phase 2 (Binary DIAG_M1 Data Analysis): By clearly specifying failure and non-failure conditions—DIAG_M1 = 4 denotes failure, and DIAG_M1 = 0 denotes no failure in this phase. Streamlines failure categorization. In real-time applications where prompt decision-making is necessary, this method works well. To identify the most crucial diagnostic parameters causing failures, Random Forest classification is used because depending only on DIAG_M1 may overlook crucial failure patterns. Additionally, abnormalities and odd behaviors in fan performance are detected using DBSCAN clustering and K-Nearest Neighbors (KNN) [4].

This study's focus is restricted to controlled functional testing; long-term durability evaluations are not included. An MCPS (Monitoring & Control Processing System) is used to capture data in real-time while the tests are carried out in an automated laboratory setting. This research may be used to real-time failure detection and predictive maintenance since the approach is developed to identify operational problems rather than mechanical degradation over time [5].

1.5 Methodology Overview

Two main experimental stages serve as the framework for this thesis:
Phase 1: Analysis of Mixed DIAG_M1 Data

- **Information Preprocessing:** Key diagnostic parameters were retrieved and sensor data was cleaned.
- **Engineering Features:** established failure criteria according to temperature, RPM, voltage, and current.
- **Unsupervised Learning for Anomaly Detection:** Outliers were found using PCA and DBSCAN clustering.
- **Failure Classification:** Assessed how various criteria affected the identification of failures.

Phase 2: Rule-Based Failure Classification

- **Rule-Based Failure Classification:** DIAG_M1 = 4 indicates failure, whereas DIAG_M1 = 0 indicates proper functioning.
- **Feature Importance Analysis:** The most important criteria for failure detection were determined using Random Forest classification.
- **Anomaly Detection:** To categorize anomalous activity, KNN and DBSCAN clustering were used.
- **Comparing Phase 1:** Assessed the trade-offs between straightforward rule-based categorization and intricate multi-parameter failure detection.

1.6 Significance of the Study

This study's importance stems from its potential to improve brushless electric fan failure detection techniques in automotive applications by using an automated and data-driven testing methodology. Advanced diagnostic techniques are required due to the increasing complexity of automobile cooling systems and the need for dependable, high-performance parts. Accurately detecting actual abnormalities is difficult with traditional threshold-based failure detection approaches as they frequently produce missed failures and false positives. By using machine learning approaches to provide a reliable, scalable, and automated failure detection system, this study tackles these issues.

1. **Improving Vehicle Cooling System Performance and Reliability**

In order to ensure ideal engine performance and thermal management, brushless electric fans must be dependable. Engine overheating, decreased fuel economy, and in severe situations, catastrophic failures, might result from a faulty fan. This study offers a solid foundation for early failure detection, preventive maintenance, and decreased vehicle downtime [6]. This study suggests an intelligent system that can constantly monitor fan performance and identify abnormalities in real-time by combining supervised classification techniques like Random Forest with unsupervised learning approaches like PCA and DBSCAN clustering.

By using the study's findings to optimize car diagnostic systems, manufacturers may raise the performance and dependability of their products. Accurately identifying malfunctions lowers warranty claims and maintenance expenses for automakers while simultaneously increasing the safety and effectiveness of cooling systems[7].

2. **Enhancing Predictive Maintenance while Cutting Expenses**

The study's potential use in predictive maintenance is one of its significant contributions. Conventional reactive maintenance methods simply replace or repair parts after they fail, which results in unplanned malfunctions and increased maintenance expenses [8] By facilitating the early discovery of abnormalities, this research opens the door for predictive maintenance solutions. Maintenance teams may plan repairs before severe failures occur by spotting patterns of deviation in operating parameters. This improves operational efficiency, extends equipment lifespan, and lowers maintenance costs [9].

The study also lessens the need for rule-based diagnostics and human inspections, which are frequently laborious and prone to mistakes. Machine learning-based automated testing guarantees quicker, more precise diagnostics, cutting down on the amount of time needed for maintenance and troubleshooting processes in automotive applications.

3. **Utilization in Smart Manufacturing and Industry 4.0**

Industry 4.0, where automation, AI, and real-time data analytics are essential to manufacturing and quality control, is rapidly taking over the automotive sector [10]. By presenting an intelligent diagnostic system that can be included into automated manufacturing lines, our work is in line with these developments.

Manufacturers can use machine learning-based failure detection to make sure that defective parts are found before they are assembled to improve quality control. By aggressively detecting flaws during the testing stage, product recalls can be decreased. Integrate machine learning algorithms with on-board car diagnostics to facilitate real-time decision-making.

4. **Advancements in Machine Learning for Failure Detection**

The subject of machine learning in industrial applications is expanding, and this work adds to it. A hybrid method to failure detection and anomaly classification is offered by the integration of Principal Component Analysis (PCA), DBSCAN clustering, and Random Forest classification. In contrast to traditional threshold-based diagnostic systems, this method continually adjusts to new failure patterns by learning from past data [11]

Additionally, this work shows how unsupervised learning may be utilized to uncover hidden failure patterns, which makes it useful for sectors with a lack of labeled failure data. Other industrial areas can benefit from the knowledge gained from this study, such as:

- **Aerospace** (detection of engine component failures).
- **Renewable energy** (keeping an eye out for issues with solar panel cooling systems and wind turbines).
- **Manufacturing** (automatic systems and industrial machines with predictive maintenance).

5. **Contribution to Academic Research and Future Studies**

By offering an organized framework for functional testing and failure detection utilizing data-driven methodologies, this thesis advances the academic research community. By providing a useful case study for automotive applications, it expands on previous research in anomaly detection, predictive maintenance, and machine learning applications in industrial systems [12].

This work can be expanded upon in future research by:

- Using deep learning methods for even more precise failure predictions, such as recurrent neural networks (RNNs) and autoencoders.
- For a more thorough study, the dataset should be expanded to include on-road sensor data and actual driving circumstances.
- Investigating cloud-based monitoring platforms to facilitate fleet management failure detection and remote diagnoses.

Chapter 2

Literature Review

To place this study in the larger framework of automated functional testing, machine learning for failure detection, and predictive maintenance in industrial applications, a thorough literature analysis is necessary. This chapter examines current brushless electric fan diagnostic techniques, conventional failure detection strategies, and developments in data-driven anomaly detection. Predictive maintenance techniques, Industry 4.0 trends, threshold-based diagnostics, and machine learning applications in defect identification are important areas of study.

2.1 Traditional Approaches to Functional Testing in Automotive Systems

In automobile engineering, functional testing is a crucial procedure that makes that electromechanical parts, such as brushless electric fans, function within predetermined performance bounds. Fixed thresholds, predetermined circumstances, and human or semi-automated inspection procedures are used in traditional automotive diagnostics and testing approaches to identify defects and failures. Despite their widespread use, these techniques are not always effective in accurately detecting anomalies, especially in complicated failure circumstances when a number of factors affect system performance[13].

To test cooling system components, automakers have traditionally used a number of traditional methods, including human inspections, threshold-based diagnostics, on-board diagnostics (OBD), and semi-automated testing frameworks. However, these conventional methods are unable to keep up with the demands of contemporary diagnostics as vehicle systems grow increasingly complicated due to increased electrical and software integration [14].

2.1.1 Threshold-Based Failure Detection

Threshold-based failure detection is among the oldest and most used techniques for functional testing in automotive systems. This method entails establishing specified upper and lower bounds for crucial parameters like:

- **Motor current (Current_M1)**
- **Voltage supply (V_Conn_M1)**
- **Rotational speed (RPM) of the fan (CF1_RPM_Avg_ST3_M1)**
- **Temperature (Tamb_)**

A fault code is activated, signaling a possible failure, when a sensor reading beyond certain predetermined limit [15].

Limitations of Threshold-Based Diagnostics:

- Inability to identify complicated failures: A lot of brushless electric fan failures are caused by interactions between several factors rather than by going above predetermined limitations.
- High false positive rate: Minor variations in sensor readings, such as brief voltage decreases, might result in needless failure alerts and unneeded maintenance expenses
- Inability to adapt: Over time, the approach becomes less successful since fixed criteria do not take into consideration environmental influences like temperature changes or component aging.

Machine learning-based anomaly detection models must be used since, despite their ease of implementation, threshold-based approaches are not flexible or adaptable enough for contemporary automotive applications.

2.1.2 Manual and Semi-Automated Testing in Automotive Systems

To assess performance deviations, functional testing in car cooling systems has traditionally depended on human skill and manual inspection . The brushless electric fans would be physically inspected by testing engineers.

- Behavior of operations under various voltage loads
- Thermal efficiency at different outside temperatures
- Reaction to start-stop patterns for a long time

Although hand examinations yield insightful qualitative information, they are:

- labor-intensive and time-consuming
- prone to discrepancies and human mistake
- High-volume vehicle production is challenging to scale.

Semi-automated functional testing methods, which combine sensor-based data gathering with human interpretation of results, were created to increase efficiency . These systems, however, continued to rely mostly on pre-established test cases and were unable to identify intricate, concealed failure patterns that call for anomaly detection and real-time data analysis.

2.1.3 On-Board Diagnostics (OBD) and Electronic Control Unit (ECU) Testing

On-Board Diagnostics (OBD) became the standard technique for real-time problem detection as car electronics improved-12-. Engine, cooling, and electrical components are continually monitored by Electronic Control Units (ECUs), which are the foundation of OBD systems.

The Operation of OBD-Based Diagnostics are:

- **Data collection:** Operating parameters (such as fan speed, temperature, and voltage) are continually monitored by sensors positioned throughout the vehicle.

- **Fault Code Generation (DTCs):** The ECU produces a Diagnostic Trouble Code (DTC) whenever a sensor reading above certain thresholds.
- **Service Indicator Alerts:** When a malfunction happens, the driver is informed via the car's onboard display.
- **Mechanic Intervention:** An OBD-II scanner is used to get the trouble code, which mechanics then utilize to manually diagnose the problem

OBD has several disadvantages even if it offers a standardized method for problem detection:

- **Failures cannot be predicted in advance:** OBD-based diagnostics identify issues after they arise, as opposed to anticipating possible failures.
- **Restrictions on OBD systems** to manufacturer-defined DTCs may result in the failure to discover hidden irregularities that do not correspond to preset error codes
- **Unsuitable for detecting anomalies in real time:** OBD does not adjust to new failure situations over time, in contrast to contemporary machine learning models

2.1.4 Need for a Data-Driven Approach

Data-driven procedures are becoming more and more popular in the market due to the shortcomings of traditional testing methods. Rather of depending on preset rules and thresholds, contemporary machine learning models examine both past and current sensor data to:

- Find intricate failure patterns that conventional techniques could miss.
- Preventive maintenance is made possible by anticipating faults before they occur.
- Continue to learn and adjust, increasing the accuracy of failure detection over time.

This study presents a system for automated functional testing that combines Random Forest, DBSCAN, and PCA models to:

- Boost the accuracy of failure classification in comparison to threshold-based and conventional OBD techniques.
- Use unsupervised clustering techniques to facilitate the early discovery of anomalies.
- Instead of depending just on single-threshold exceedances, analyze multi-parameter interactions to reduce false positives.

This study advances automotive testing procedures by using a data-driven strategy, which guarantees increased vehicle performance, reduced maintenance costs, and increased dependability.

2.1.5 Comparative Analysis of Traditional Testing Methods

Method	Advantages	Disadvantages
Threshold-Based Testing	Basic, popular, and simple to use	High false positives, poor accuracy, and inability to identify complex failures
Manual Testing	Permits human knowledge and thorough assessment	Not scalable, time-consuming, and prone to human error
Semi-Automated Testing	Delivers organized data and lessens human labor.	Still uses manual interpretation and has little flexibility.
On-Board Diagnostics (OBD)	Standardized and available in the majority of cars	Lacks flexibility, is limited to preset codes, and cannot anticipate errors

Table 2.1: Comparison of Different Testing Methods

Although these techniques have been successful in the past in spotting significant failures, they are unable to foresee problems before they happen. By using machine learning-based failure detection models, which are more accurate, adaptable, and able to identify anomalies in real time, this study seeks to get around these restrictions.

2.2 Machine Learning for Anomaly Detection in Industrial Systems

2.2.1 Introduction to Machine Learning in Industrial Fault Detection

Traditional fault detection techniques that rely on preset failure criteria are not being able to discover hidden abnormalities and early failure patterns due to the growing complexity of industrial systems. The application of machine learning (ML) in industrial diagnostics has transformed real-time failure detection and predictive maintenance by allowing systems to learn from past data and identify subtle patterns suggestive of failures [16]

Large volumes of sensor data are produced by industrial systems, which record operating characteristics including temperature, rotational speed, voltage, and current. It is challenging to manually assess this data for failure situations since it is frequently noisy and high-dimensional. Machine learning approaches, namely supervised and unsupervised learning, provide sophisticated capabilities for fault classification based on historical patterns, anomaly prediction, and failure detection [17]

2.2.2 Supervised Learning for Fault Classification

A popular method for failure detection is supervised learning, in which models are trained on labeled datasets. that is, each data point has a predetermined failure or non-failure condition. The model categorizes incoming data points into failure or non-failure groups by applying the lessons it has learned from past failures.

When a well-defined dataset with enough failure examples is available, supervised learning can be helpful in industrial systems. Typical algorithms consist of:

- A potent method that determines the most significant characteristics influencing failure categorization is Random Forest categorization.
- Support Vector Machines (SVMs): Used in failure diagnostics to identify patterns, especially to differentiate between various failure types [18]
- In deep learning-based diagnostics, artificial neural networks (ANNs) are used, particularly in complicated failure scenarios [19]

However, there are drawbacks to supervised learning in business settings because:

- Because industrial systems are built to have as few faults as possible, failure data is frequently limited.
- Failure data labeling requires professional involvement and is costly and time-consuming.

The model might not be able to identify new failure types unless it is retrained [20]

Because they don't require predetermined failure labels, unsupervised learning approaches are becoming more and more popular for anomaly detection in industrial diagnostics as a result of these difficulties [21]

2.2.3 Unsupervised Learning for Anomaly Detection

The majority of sensor data in real-world industrial systems depicts typical operating circumstances, with faults happening seldom. Unsupervised learning is a more efficient method for identifying unknown and emergent abnormalities since it is difficult to categorize every potential failure event.
Without predetermined failure labels, unsupervised learning algorithms find patterns and deviations in data. These models function by:

In many industrial applications, such as fraud detection, network security, predictive maintenance, and manufacturing quality control, anomaly detection is essential. Conventional rule-based approaches depend on set thresholds, which frequently miss intricate failure patterns and result in false positives and anomalies that go unnoticed. Because of this restriction, machine learning-based anomaly detection approaches—especially unsupervised learning techniques, which may detect abnormalities without the need for labeled failure data—have become widely used.

Sensor data is regularly generated in industrial systems, and because failure occurrences are rare, it is frequently impossible to classify every instance as normal or defective. By finding patterns in the data and spotting outliers that deviate from expected behavior, unsupervised learning techniques get around this problem. These techniques are more flexible than conventional threshold-based diagnostics and are particularly helpful in situations when failures are unknown, changing, or context-dependent.

- Use previous data to learn the system's typical behavior.
- Locating data points that substantially depart from this typical trend.
- Identifying those variations as possible abnormalities or early warning signs of failure [22]

1. Anomaly detection based on clustering (DBSCAN, K-Means) Similar data points are grouped together using clustering algorithms, while anomalies are identified as outliers that either form tiny, isolated clusters or do not belong to any cluster. A popular technique for identifying irregularities in high-dimensional, noisy data is DBSCAN (Density-Based Spatial Clustering of Applications with Noise). It works well for detecting unknown failures because, in contrast to K-Means, it does not require a set number of clusters. Data points are grouped into K clusters using K-Means Clustering according to their similarity. Points that are far from the cluster centroids are identified as anomalies. However, irregularly shaped anomalies are difficult for K-Means to handle, therefore DBSCAN is a more adaptable option. [2]

2. Anomaly Detection by Dimensionality Reduction (PCA, Autoencoders) Correlated and redundant features are frequently seen in high-dimensional industrial sensor data. Techniques for reducing dimensionality aid in identifying the most important patterns for failure detection.

High-dimensional data can be reduced to a smaller number of primary components using primary Component Analysis (PCA), which preserves the most significant variation. Points that show unusual behavior in sensor readings and considerably depart from principal component projections are identified as anomalies. The normal data distribution is compressed and learned by autoencoders (Neural Networks for Anomaly Detection). A new data point is probably an anomaly if the autoencoder is unable to correctly reconstruct it. Deep learning-based anomaly detection makes extensive use of this technology.

3. Density-Based and Distance-Based Anomaly Detection (KNN, Isolation Forest) Distance-based methods use the distance between a point and its closest neighbors in the feature space to identify anomalies.

Using K-Nearest Neighbors (KNN) to Identify Anomalies: An anomaly score is calculated using the distance to the K-th nearest neighbor. Anomalous points are those that are farther away from their neighbors. An ensemble technique called Isolation Forest divides the data at random to isolate abnormalities. This method is very effective for large-scale anomaly identification because it requires fewer partitions to isolate anomalies.

Unsupervised Learning's Use in Industrial Systems Unsupervised learning-based anomaly detection is frequently used in industrial settings in:

Finding early indicators of mechanical failure in electric motors, cooling fans, and spinning machinery before a breakdown happens is known as predictive maintenance.

Quality control is the process of employing sensor data and machine learning to identify faulty parts in the automotive manufacturing industry.

Detecting illegal activity and network irregularities in smart factories is the focus of cybersecurity in industrial IoT. Monitoring energy efficiency involves spotting unusual trends in energy use that could point to problems with the power supply.

In order to find **hidden failures** beyond the conventional **DIAG_M1 classification**, **DBSCAN and KNN anomaly detection algorithms** were used in this study's *brushless electric fan functional testing*. The findings showed that **problems that would have gone undetected with conventional diagnostic techniques** were successfully detected using **unsupervised learning**.

Unsupervised Anomaly Detection: Obstacles and Prospects Unsupervised learning for anomaly detection has drawbacks despite its benefits:

High False Positive Rates: Unsupervised approaches may mistakenly identify typical fluctuations as abnormalities, resulting in needless maintenance procedures, because they do not rely on predetermined failure labels.

Parameter Sensitivity: To get the best results, algorithms like DBSCAN need to carefully adjust hyperparameters.

Managing Concept Drift: Anomaly detection models must constantly adjust to new typical operating conditions as industrial systems undergo continuous change.

computing Complexity: The real-time deployment of certain unsupervised techniques is limited by their high computing resource requirements, particularly for deep learning-based autoencoders.

The efficiency of machine learning-driven predictive maintenance will be further increased by upcoming developments in adaptive anomaly detection, hybrid supervised-unsupervised techniques, and real-time deep learning models.

Unsupervised learning is a powerful tool for anomaly detection in industrial applications, particularly when labeled failure data is scarce or unavailable. Techniques such as DBSCAN, KNN, PCA, and autoencoders enable the detection of hidden failure patterns, improving predictive maintenance and fault classification. In this study, unsupervised learning successfully identified failures in brushless electric fans beyond traditional DIAG_M1 labeling, proving its effectiveness in real-time diagnostics and vehicle component testing. As machine learning and AI-driven diagnostics continue to evolve, integrating adaptive anomaly detection techniques into Industry 4.0 environments will be critical for enhancing system reliability, reducing downtime, and optimizing maintenance strategies.

2.2.4 Why Unsupervised Learning for Our Study?

Because of the characteristics of the dataset and the absence of pre-labeled failure events, this work uses unsupervised learning approaches in both experimental phases.

Phase 1: Unsupervised Learning and the Significance of Features

- A dataset with mixed DIAG_M1 values (0 and >0) was examined in the first phase, where failure criteria weren't always obvious.

- To find the most important parameters for failure detection, Random Forest feature importance analysis was used [23].
- Key data properties were preserved while dimensionality was decreased through the use of Principal Component Analysis (PCA).
- A basic threshold-based failure classification was unable to catch anomalies, thus DBSCAN clustering was employed to find them [24].

Phase 2: Using DBSCAN and KNN for Unsupervised Learning

- For anomaly detection and failure classification validation, we continued to use unsupervised learning approaches in the second phase, where DIAG_M1 = 4 was categorized as failure and DIAG_M1 = 0 as non-failure.
- Clustering parameters were determined by calculating ideal distance metrics using K-Nearest Neighbors (KNN).
- To categorize data points into regular operating circumstances and aberrant behaviors, DBSCAN clustering was used [25].
- Beyond a straightforward DIAG_M1 threshold rule, these techniques offered a strong mechanism to guarantee failure categorization was accurate.

2.2.5 Anomaly Detection in Industrial Systems

A key component of predictive maintenance is anomaly detection, which enables industrial systems to spot early indications of malfunction before they become complete failures.

Industrial system anomalies fall into one of the following categories:

- Single data points that deviate noticeably from the usual distribution are known as point anomalies (e.g., rapid voltage spikes).
- Anomalies that are simply unusual in a certain environment are known as contextual anomalies (for example, a high current at startup could be typical but aberrant during steady operation).

- Groups of data values that collectively point to a failure (such as a trend of rising temperature and falling RPM over time) are known as collective anomalies.

Models that can handle complicated, high-dimensional sensor data and differentiate between real failures and normal changes are needed to detect these abnormalities. In order to do this, DBSCAN clustering is essential.

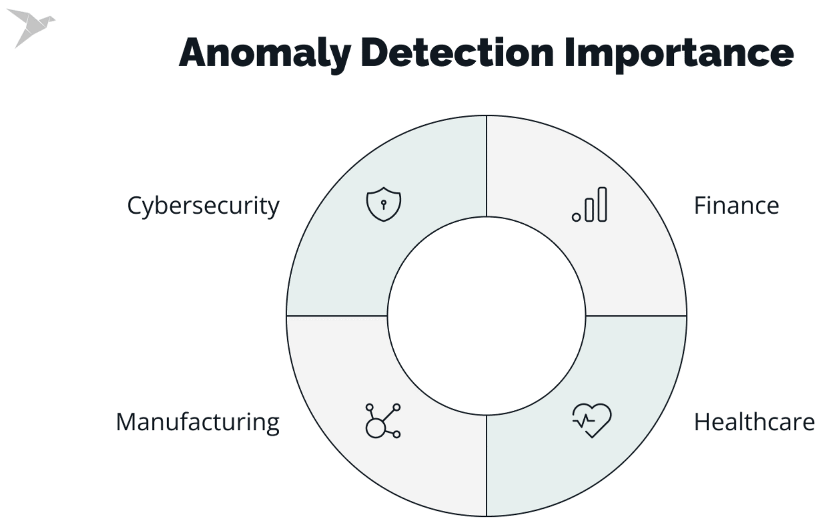

Figure 2.1: Importance of Anomaly Detection [26]

2.2.6 DBSCAN for Anomaly Detection

A potent unsupervised machine learning technique that is frequently used for anomaly detection in industrial systems is Density-Based Spatial Clustering of Applications with Noise (DBSCAN). DBSCAN is density-based, which means it groups data points that are densely packed together while detecting outliers as noise points, in contrast to conventional clustering techniques like K-Means, which involve defining the number of clusters in advance. Because of this, DBSCAN is especially well-suited for industrial datasets, where failure patterns might not adhere to a predetermined structure and sensor values might contain unanticipated variations. Two crucial parameters are defined by the algorithm: min_samples, the smallest number of points needed to create a dense cluster, and epsilon, which establishes the radius surrounding each data point within which additional points are regarded as neighbors. Because data points that are not part of any cluster are categorized as anomalies, DBSCAN is very good at identifying infrequent failure occurrences in complex systems.

In order to identify irregularities in the performance of brushless electric fans, DBSCAN was performed to the dataset in both phases of this investigation. Failure detection in Phase 1, where DIAG_M1 values were mixed (0 and >0), necessitated the analysis of several parameters. To uncover hidden failure patterns, DBSCAN was employed in conjunction with PCA (Principal Component Analysis). DBSCAN assisted in validating the classification findings and identifying anomalies that did not fall into the conventional failure categories in Phase 2, where failure was categorized using DIAG_M1 = 4 and non-failure using DIAG_M1 = 0. The ideal epsilon value was found using a K-Nearest Neighbors (KNN) method, which made sure that the clustering process accurately distinguished between normal and pathological behavior. By identifying outliers as noise spots, DBSCAN made it possible to identify possible early-stage failures, enhancing it.

Compared to conventional threshold-based approaches, DBSCAN is more adaptable to real-world sensor changes due to its ability to manage noisy data and irregularly formed clusters, which is one of its main benefits in anomaly identification. Sensor readings for temperature, voltage, current, and RPM behavior in brushless electric fans might change as a result of operating irregularities, temporary malfunctions, or environmental factors. While DBSCAN can distinguish between real abnormalities and natural changes, threshold-based approaches may mistakenly label these variations as failures. Furthermore, DBSCAN can discover anomalies in real time in functional testing settings due to its processing efficiency for huge industrial datasets. This work created a reliable failure detection system that surpasses conventional diagnostic techniques and improves performance by combining DBSCAN with feature significance analysis (Random Forest) and dimensionality reduction (PCA).

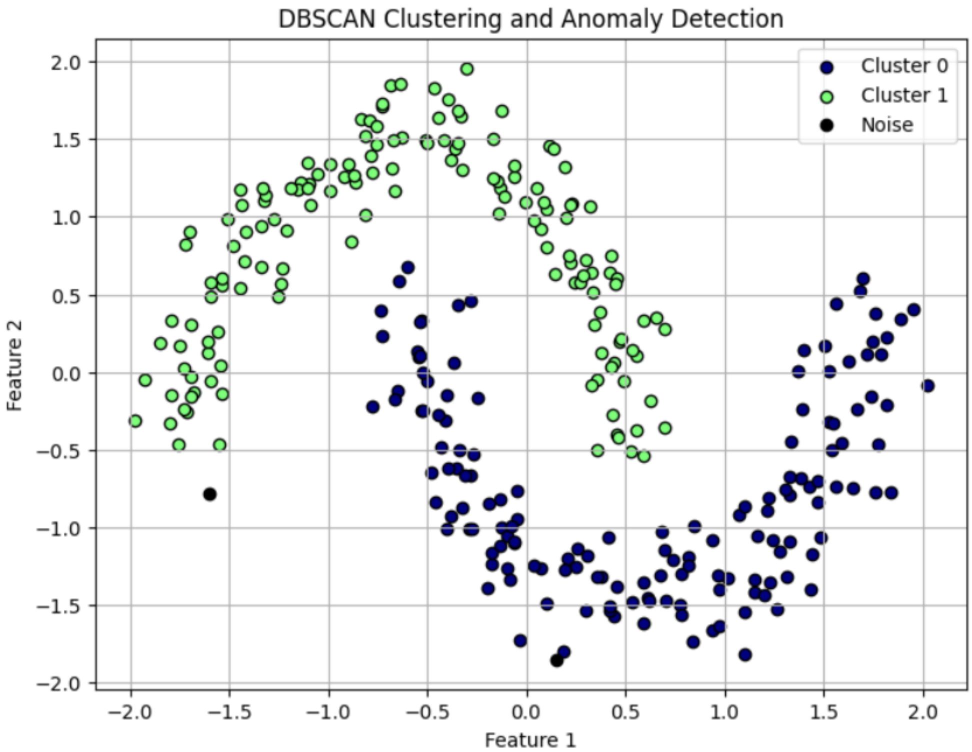

Figure 2.2: DBSCAN Clustering & Anomaly Detection

2.3 Industry 4.0 and Smart Diagnostics in Automotive Systems

2.3.1 Introduction to Industry 4.0 in Automotive Systems

Industry 4.0, which combines automation, real-time data analytics, artificial intelligence (AI), machine learning (ML), and the Internet of Things (IoT) to improve production and diagnostics, is causing a radical change in the automobile sector. Smart automation and data-driven decision-making are key components of Industry 4.0, [27] which enables automakers to improve manufacturing lines, predictive maintenance, and car diagnostics. Manual inspections and threshold-based diagnostic devices were key components of traditional automotive testing and maintenance procedures. These methods, however, frequently result in more downtime, expensive repairs, and delayed failure discovery. Automakers may lower maintenance costs and unexpected breakdowns while increasing overall vehicle performance, economy, and dependability by putting smart diagnostics into practice [28].

In order to continually monitor car components, smart diagnostics in automotive applications make use of sensor networks, cloud computing, and AI-driven problem detection systems. Among the parts that benefit from these developments are brushless electric fans, which are essential for cooling car systems. Industry 4.0 makes it possible to employ predictive analytics to identify early-stage defects before they worsen, whereas traditional cooling system diagnostics were restricted to predetermined threshold criteria. Functional testing is more precise, proactive, and condition-adaptive when real-time data analytics and machine learning algorithms are integrated [29].

2.3.2 Evolution of Smart Diagnostics in Automotive Systems

The growth of linked car technologies and improvements in AI-based problem detection systems have propelled the creation of smart diagnostics. Hundreds of sensors are found in modern cars, gathering information on things like motor current, cooling fan speed, battery voltage, engine temperature, and diagnostic fault codes. These sensors function within specified safety bounds in conventional diagnostics, and malfunctions are identified when readings above a certain threshold [30]. This reactive strategy, however, is inadequate as it ignores the slow system degradation that might eventually result in breakdowns.

Automobile manufacturers are already using real-time monitoring systems that enable continuous data collection and predictive maintenance through the integration of Industry 4.0 principles. To find failure trends and forecast component deterioration, intelligent diagnostic algorithms use both history and current sensor data. By identifying hidden trends in big datasets, machine learning approaches, such as anomaly detection algorithms like PCA and DBSCAN, improve the accuracy of failure classification [31]. Automated functional testing of brushless electric fans has benefited greatly from these methods, which enable manufacturers to identify problems brought on by unforeseen voltage changes, temperature swings, and motor inefficiencies.

Because it allows for remote vehicle health monitoring, cloud computing is essential to modern car diagnostics. These days, automakers gather and examine sensor data from cars running in various environments using cloud-based systems. Real-time failure notifications are generated by processing this data using AI-driven fault detection algorithms. For example, when a car's electric fan shows abnormal current consumption patterns, a predictive maintenance system can identify the problem before it leads to a total failure, saving money on repairs and averting car breakdowns [32].

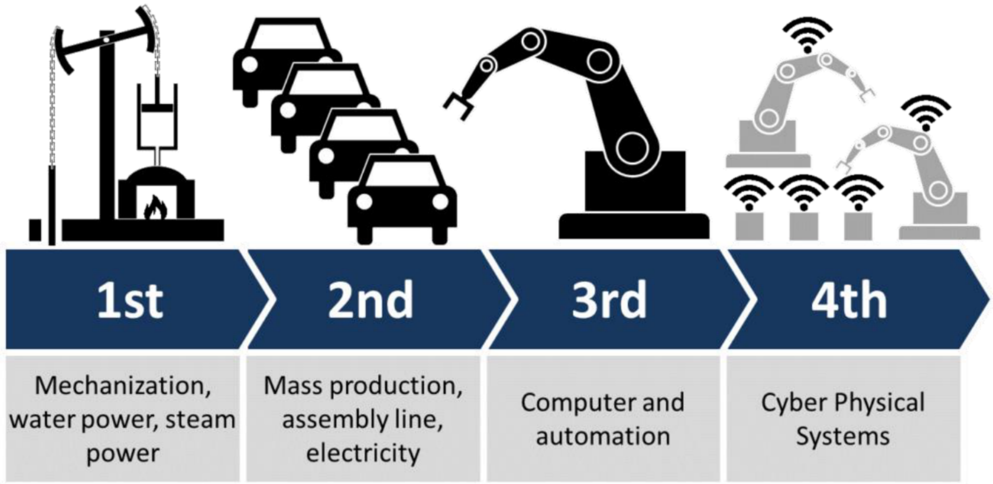

Figure 2.3: Evolution of Automotive Systems [33]

2.3.3 The Role of IoT and AI in Smart Automotive Diagnostics

By allowing linked cars to exchange real-time data with centralized monitoring systems, the Internet of Things (IoT) has had a big impact on smart automotive diagnostics. Wireless sensors in IoT-enabled cars provide operational data to cloud servers continually, where AI-based algorithms examine patterns and identify irregularities. Thanks to these developments, manufacturers may now use remote diagnostics, which enables professionals to evaluate a vehicle's condition without a physical examination [34].

IoT-based monitoring systems provide light on how motor efficiency, voltage variations, and environmental factors affect cooling performance in brushless electric fan testing. Smart diagnostic systems employ AI-driven failure categorization models to identify early indicators of motor inefficiency and overheating rather than depending only on on-board diagnostics (OBD) fault codes. Automotive engineers can detect irregularities in sensor data before they result in serious malfunctions by using unsupervised learning methods like DBSCAN and PCA [35].

The usage of edge computing, which processes data at the vehicle level instead of depending just on cloud computing, is another significant advancement in Industry 4.0 automotive diagnostics. By enabling real-time decision-making, edge-based AI models lower the delay in identifying abnormalities in sensors and system malfunctions. For instance, an edge-based AI system may promptly initiate a preventive maintenance warning in the event that an electric fan displays unusual current variations, guaranteeing prompt remedial measures [36]

AI-powered diagnostics will develop further as automotive systems become more software-driven, improving predictive maintenance and functional testing. The shift from reactive problem detection to predictive analytics guarantees that automakers can provide more cost-effective maintenance plans, enhanced safety, and increased dependability.

2.3.4 Challenges and Future Prospects of Smart Diagnostics

There are still a number of obstacles in the way of the broad use of AI-driven car diagnostics, even with the developments in Industry 4.0 and smart diagnostics. Data complexity is one of the main issues since contemporary cars produce enormous volumes of sensor data that need to be processed, stored, and analyzed in real time. Another crucial issue is ensuring cybersecurity and data privacy, particularly as more cars are linked to cloud-based platforms.

Adapting AI models for various vehicle kinds is another difficulty. High-quality training data is necessary for machine learning-based diagnostics to function well, and changes in engine configurations, vehicle models, and environmental factors can affect how accurate predictive maintenance models are. Automobile makers need to invest in highly adaptable AI systems that can generalize problem detection across various vehicle types and operating situations in order to overcome these obstacles.

Looking ahead, completely autonomous failure detection systems that can forecast defects with almost perfect accuracy represent the future of smart automobile diagnostics. AI, IoT, cloud computing, and real-time monitoring will come together to form a self-learning automotive diagnostic ecosystem that can dynamically adjust to novel failure scenarios. The automotive industry will get closer to zero-defect production and real-time predictive maintenance as AI-powered functional testing is more integrated with the manufacturing process. This will guarantee that vehicle cooling systems, including brushless electric fans, run as efficiently and reliably as possible.

Chapter 3

Methodology

The process for creating an automated functional testing framework for brushless electric fans in automotive applications is described in this chapter. To categorize failures and identify abnormalities in real-time sensor data, the method is split into two stages, each of which uses a distinct machine learning and anomaly detection algorithm.

A mixed DIAG_M1 dataset was examined in Phase 1, where multi-parameter analysis and anomaly identification using unsupervised learning methods were necessary for failure categorization. This stage used DBSCAN for anomaly detection, Random Forest for feature importance analysis, and Principal Component Analysis (PCA) for dimensionality reduction.

Phase 2 involved the explicit labeling of failures using a binary DIAG_M1 classification technique. In order to verify the accuracy of failure detection, this phase included feature selection approaches, DBSCAN for clustering anomalies, and K-Nearest Neighbors (KNN) for distance metric optimization.

Both stages used sensor data gathered from actual brushless electric fan functioning testing to analyze variables such motor current, voltage, temperature, and RPM behavior. In order to make sure that the suggested failure detection system is reliable and flexible enough for real-time applications, the approach focuses on data preparation, feature selection, machine learning model implementation, and performance assessment.

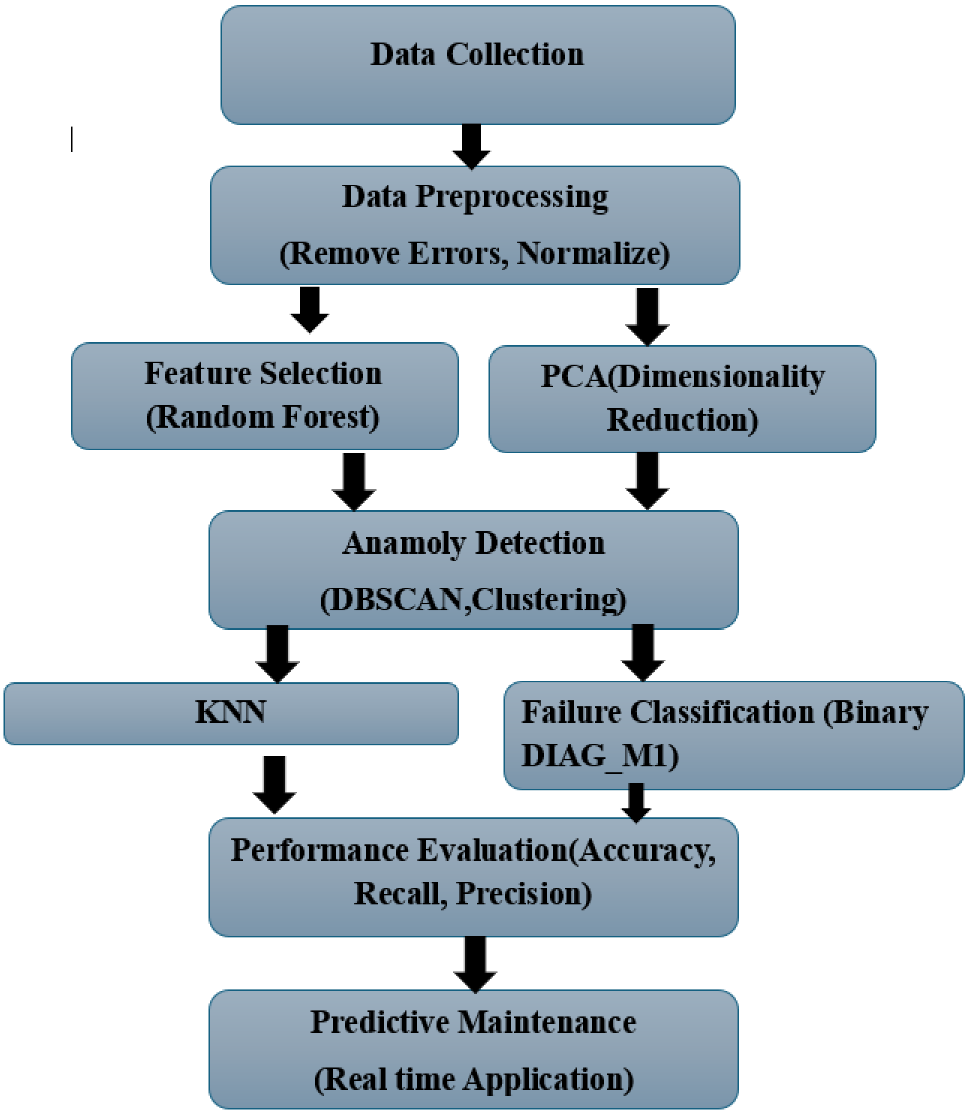

Figure 3.1: Overall Workflow

3.1 Data Collection and Preprocessing

Data Collection: The functional testing trials carried out in a controlled laboratory setting provided the dataset used in this investigation. Brushless electric fans' real-time sensor data were captured throughout each test, providing crucial information for failure diagnosis. Among the dataset were:

- **Time-Stamped Sensor Readings:** A thorough trend analysis was made possible by the high frequency of data recording.
- Data on **voltage** and **current** are important markers of electrical irregularities and inefficient motors.
- **Temperature** and **rotational speed (RPM)** measurements are used to evaluate the fan system's mechanical and thermal performance.
- As a major failure signal, **DIAG_M1 Failure Codes** needed further multi-parameter validation.

Data Preprocessing:

The raw dataset was processed to guarantee data accuracy, consistency, and relevance prior to the use of machine learning models. The actions listed below were taken:

1. Managing Inaccurate and Missing Values

- To make sure that only legitimate sensor readings were included, error rows were eliminated.
- To preserve the integrity of the data, any erroneous timestamps or corrupt records were removed.

2. Feature engineering and standardization

- Conversion of timestamp: To enhance trend analysis, the raw time-stamp format was broken down into hour, minute, second, and millisecond components.
- Labeling the operational phase:

  - **Resting Phase:** When the motor current was almost nil, the resting phase was identified.
  - **Startup Phase:** characterized by a high required RPM yet a low motor current.
  - **Complex Failure Phase:** Indicates cases when operational irregularities were found but DIAG_M1 > 0.

3. **Normalization and Scaling of Features**

- To guarantee that machine learning models analyzed continuous variables uniformly, such as current, voltage, and RPM, they were standardized.

The technique improved the accuracy of failure categorization by preprocessing and organizing the dataset to guarantee that the machine learning models ran on clear and relevant data.

3.2 Mixed DIAG_M1 Data Analysis

3.2.1 Feature Selection Using Random Forest

Importance of Feature Selection in Failure Detection:

A crucial stage in machine learning-based failure detection is feature selection, especially in industrial settings where sensor data includes several variables of differing importance. Sensor readings for brushless electric fan functional testing include, among other things, motor current, voltage, rotational speed (RPM), and temperature. Some of these characteristics may add noise, redundancy, or needless complexity to the model, and not all of them contribute equally to failure classification [37]

A crucial stage in machine learning-based failure detection is feature selection, especially in industrial settings where sensor data includes several variables of differing importance. Sensor readings for brushless electric fan functional testing include, among other things, motor current, voltage, rotational speed (RPM), and temperature. Some of these characteristics may add noise, redundancy, or needless complexity to the model, and not all of them contribute equally to failure classification [2]

Why Random Forest for Feature Selection?

In order to improve accuracy and decrease overfitting, the Random Forest ensemble learning method builds many decision trees and averages their predictions. Its capacity to gauge feature significance using two main measures is one of its main advantages:

- **The Mean Decrease in Impurity (MDI),** sometimes referred to as Gini Importance, quantifies the degree to which a characteristic lowers classification uncertainty. More pertinent traits are indicated by higher values.
- **Mean Decrease in Accuracy (MDA)** ranks features according to their contribution to classification accuracy, assessing the impact of feature removal on model performance [38]

For feature selection, Random Forest was used due to:

- It doesn't need feature scaling to handle high-dimensional industrial datasets.
- It increases the resilience of failure classification models by being impervious to noise and irrelevant information.
- In order to help engineers concentrate on key performance indicators (KPIs) that impact brushless electric fan failures, it offers an interpretable ranking of sensor parameters [39]

Application of Random Forest in This Study

Determining which characteristics were most important in differentiating between normal and failed fan operations was crucial in Phase 1 of this investigation, when failure classification was complicated due to mixed DIAG_M1 values (0 and >0). The preprocessed dataset was subjected to the Random Forest method, which calculated the relevance score for each feature.

Step 1: Feature Importance Training for the Random Forest Model

- Training and validation sets of the sensor dataset were separated.
- To predict failure outcomes, a Random Forest classifier was trained with all accessible features.

- To determine which factors contributed most to classification accuracy, feature significance scores were taken from the training model

Step 2: Determining the Crucial Elements Influencing Failures

Following the Random Forest feature selection procedure, the following five features were determined to be the most crucial:

- The most important factor affecting failure detection is the **voltage connection** (V_Conn_M1), as unusual voltage levels suggest possible electrical problems.
- **Motor Current (Current_M1):** A crucial indicator of motor inefficiencies, aberrant current consumption must be detected.
- The **voltage average**, or **CF1_VoltAvg_ST3_M1**, aids in identifying power supply fluctuations that lead to fan failures.
- **Tamb_ (temperature):** High temperatures raise the likelihood of failure by exerting thermal stress on fan components.
- **RPM Average (CF1_RPM_Avg_ST3_M1):** Variations in rotational speed signify deterioration in fan performance, which results in inefficient cooling.

Then, in order to discover anomalies, these characteristics were chosen as input variables for further analysis using PCA and DBSCAN clustering.

Performance Impact of Feature Selection

The failure classification model showed better generalization, quicker training periods, and increased accuracy once the less significant elements were eliminated. Using Random Forest-based feature selection had the following advantages:

- **Decreased Model Complexity:** Removing superfluous features reduced overfitting and increased failure detection effectiveness.
- **Improved Interpretability:** Rather than depending just on the raw DIAG_M1 readings, engineers may concentrate on certain sensor properties.

- **Enhanced Classification Accuracy:** The machine learning models performed better in terms of anomaly detection and failure classification by utilizing just the most pertinent features

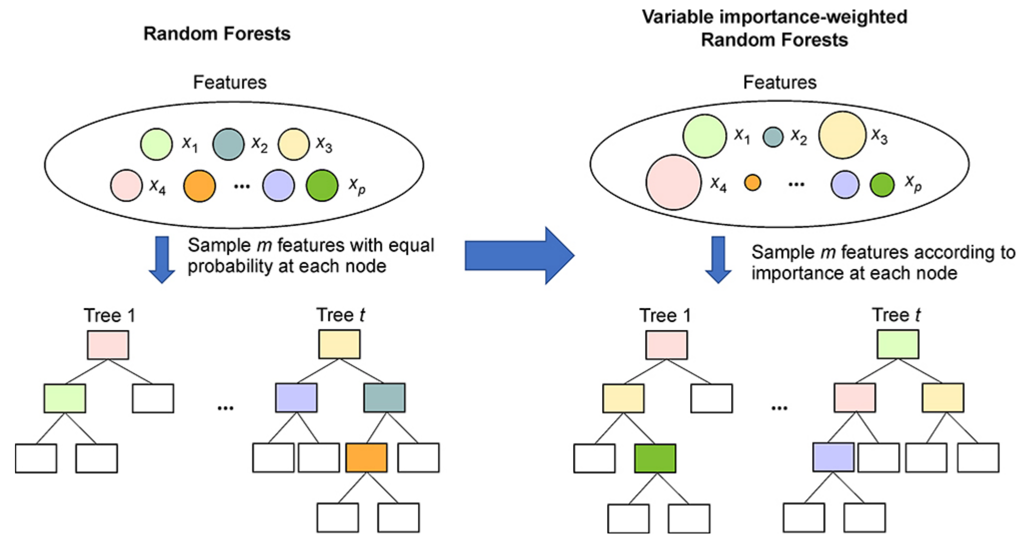

Figure 3.2: Importance of Features using Random Forest [40]

3.2.2 Principal Component Analysis (PCA) for Dimensionality Reduction

Datasets from real-world industrial systems, particularly those that use sensor-based diagnostics, can include a large number of associated characteristics, which can complicate and add redundancy to machine learning models. Although the comprehensiveness of failure detection is enhanced by having several sensor parameters, there are drawbacks as well, including longer computation times, model overfitting, and interpretability issues.

By converting a high-dimensional dataset into a lower-dimensional space while maintaining as much important information (variance) as feasible, Principal Component Analysis (PCA), a linear dimensionality reduction approach, aids in overcoming these difficulties. In order to make it simpler to employ clustering methods like DBSCAN for anomaly identification, PCA was utilized in this work to simplify sensor data while maintaining important failure-related patterns.

Why PCA for Dimensionality Reduction?

Several sensor data, such as voltage, current, RPM, and temperature measurements, are tracked over time during functional testing of brushless electric fans. Due to the strong correlations between some of these factors, failure categorization models become superfluous and redundant. In this investigation, PCA was selected for dimensionality reduction because to:

- **It lowers computational complexity:** Machine learning models can analyze lower-dimensional data more quickly and easily
- **It eliminates unnecessary correlations:** PCA increases model effectiveness by converting correlated sensor values into uncorrelated principle components.
- **It keeps important information:** Failure patterns are kept distinct by capturing the majority of the dataset's volatility in the first five principal components.
- The dataset was converted into a set of principle components by using PCA, which enabled clustering methods like DBSCAN to identify failure-related

Application of PCA in This Study

1. **Standardizing the Data:** Before applying PCA, all sensor features were **standardized using Z-score normalization** to ensure that they had **equal influence** in the transformation process. This step was necessary because **PCA is sensitive to different scales**, and features like voltage (which has a higher range) could dominate the results over smaller-scale features like temperature.

2. **Computing Principal Components:** In order to determine which directions the data fluctuates the most, PCA calculates the eigenvectors and eigenvalues of the dataset's covariance matrix. The following procedures are used to carry out the transformation:

- To determine the associations between characteristics, calculate the covariance matrix.

Mathematically, each feature X_i was standardized using the formula:

$$X' = \frac{X - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation of the feature.

- Determine the covariance matrix's eigenvectors and eigenvalues.
- Choose the top k primary components, or eigenvectors, that best represent the data's variance.
- In this new lower-dimensional subspace, transform the original dataset.

3. **Selecting the Optimal Number of Principal Components**

Finding the number of primary components that should be kept while maintaining the majority of the data was the next step after using PCA. The Explained variation Ratio, which gauges how much variation each principle component captures, was used for this.
To illustrate each primary component's contribution, a scree plot of the explained variance ratio was created. The plot demonstrated that:

- Over 90% of the overall variance was retained by the top three principal components, which means that while the dataset complexity decreased, the majority of the important failure information was kept.

Mathematically, the total variance retained is computed as:

$$\sum_{i=1}^{k} \lambda_i \Big/ \sum_{j=1}^{n} \lambda_j$$

where λ_i represents the eigenvalues corresponding to each principal component.

4. **Transforming the Dataset:** The original dataset was converted into a lower-dimensional representation using the chosen main components, which improved its suitability for unsupervised clustering techniques such as DBSCAN. By eliminating unnecessary noise and preserving failure-related patterns, this transformation enables anomaly detection algorithms to concentrate on significant failure signals.

Impact of PCA on Failure Detection

Several significant advantages resulted from the use of PCA into the failure categorization workflow:

- **Reduction in Feature Complexity:** Failure trends were easier to examine once the dataset was compressed from its initial high-dimensional form into a lower-dimensional representation.
- **Enhanced DBSCAN Anomaly Detection:** Due to DBSCAN's sensitivity to noise and feature scaling, using PCA prior to clustering increased anomaly detection accuracy, guaranteeing that failures were detected based on significant deviations rather than random noise.
- **Faster Computational Processing:** Real-time failure detection became more practical when the training time for anomaly detection models was greatly reduced by lowering the amount of input characteristics.

3.2.3 Anomaly Detection Using DBSCAN Clustering

Introduction to Anomaly Detection in Functional Testing

In industrial systems, anomaly detection is essential for both predictive maintenance and failure categorization. Because traditional threshold-based diagnostic techniques rely on preset limitations for sensor data like voltage, current, and RPM, they frequently fall short of identifying intricate failure patterns. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is an unsupervised learning approach that is more successful since real-world failures are frequently contextual and impacted by several interacting factors.

Data points are grouped into clusters by the density-based clustering method DBSCAN according to their geographical density. DBSCAN can automatically detect clusters of different forms and mark data points that do not belong to any cluster as anomalies, in contrast to K-Means clustering, which necessitates a predetermined number of clusters. For industrial diagnostics, where errors frequently manifest as single outliers rather than distinct clusters, this makes it very helpful.

Why DBSCAN for Anomaly Detection?

Failure situations may not always be clearly identified during functional testing of brushless electric fans, and some failures manifest as anomalous patterns in sensor readings as opposed to distinct threshold exceedances.

The following are some advantages of this study's anomaly detection utilizing DBSCAN:

- **Capability to identify anomalies without predetermined labels:** DBSCAN was utilized to identify hidden failure patterns based on differences in sensor performance because failure criteria were not specifically specified in Phase 1.
- **Finding irregularities in multi-dimensional data:** Electric fan failures might manifest as a collection of variables, such as temperature rise and motor current fluctuation, rather than a single parameter, such as voltage decrease. High-dimensional data may be effectively clustered using DBSCAN, which also identifies anomalous points that don't fit into typical operating patterns.
- True anomaly detection is not hampered by random sensor fluctuations thanks to DBSCAN's **noise-resistant clustering**, which is superior to other clustering methods like K-Means.

Implementation of DBSCAN in This Study

From feature selection and data preprocessing to cluster analysis and appropriate parameter selection, the DBSCAN anomaly detection procedure comprised many crucial phases.

Step 1: Feature selection and preprocessing

In order to minimize dimensionality while maintaining failure-related sensor information, the dataset was preprocessed and modified using PCA (Principal Component Analysis) prior to using DBSCAN.

The following were the primary sensor settings used for anomaly detection:

- Connection of Voltage (V_Conn_M1)
- Current_M1 (motor current)
- Tamb_, or temperature
- RPM, or rotational speed

Prior to implementing DBSCAN clustering, these parameters were normalized using Z-score normalization to guarantee that all features were equally weighted.

Step 2: Choosing the Best DBSCAN Settings

DBSCAN functions according to two essential criteria:

- The maximum neighborhood radius surrounding each point is defined by the value of epsilon. Another point is regarded as belonging to the same cluster if it is located within this radius.
- The minimal number of surrounding points needed to create a dense cluster is determined by MinPts (minimal Points).
- A K-Nearest Neighbors (KNN) distance plot was utilized to find the ideal epsilon value. The elbow approach was used, which finds the ideal epsilon value at the location where the distance plot has the most bend.

Using the standard heuristic formula, the MinPts value was determined based on the dataset's dimensionality:

$$MinPts = 2 \times \text{number of dimensions}$$

Since the PCA-transformed dataset had three principal components, MinPts was set to 6.

Step 3: Applying DBSCAN and Identifying Anomalies

The dataset was subjected to DBSCAN clustering after the ideal parameters had been established. The outcomes were:

- Dense clusters developed under typical operational circumstances.
- In order to identify possible fan failures, sparse data points that did not fit into any cluster were categorized as anomalies.
- Engineers were able to visually examine failure patterns by utilizing the first two main components to see anomalies (outliers) in a 2D projection.
- The DBSCAN scatter plot's dark spots, which indicate possible dataset failures, are noise or outliers.

Performance Evaluation of DBSCAN for Anomaly Detection

Several important performance measures were used to assess DBSCAN's efficacy for failure detection:

- **Cluster Purity:** Evaluates how successfully DBSCAN distinguished between the dataset's failure and normal instances.
- **Precision and Recall:** Measured how accurately DBSCAN identified true failures while minimizing false positives.
- **Comparing with Conventional Failure Detection Based on Thresholds:** demonstrated that DBSCAN was able to identify early-stage abnormalities that were overlooked by conventional threshold methods.

The findings demonstrated that DBSCAN greatly increased the accuracy of anomaly identification, especially in Phase 1, when the dataset lacked specific labels for failures.

Integration of DBSCAN with Predictive Maintenance

DBSCAN made predictive maintenance techniques possible by identifying malfunctions before they were serious, enabling engineers to:

- Before genuine problems arise, schedule preventative maintenance for brushless electric fans.

- Find underperforming fans that use too much electricity to maximize energy efficiency.
- Address possible faults early to minimize downtime and warranty claims.

A scalable, flexible, and reliable failure detection system was produced by combining DBSCAN with PCA and feature selection methods, proving its usefulness in actual industrial testing settings.

3.3 Binary DIAG_M1 Data Analysis

3.3.1 Simplified Failure Classification

In Phase 2 of this study used DIAG_M1 measurements to detect failures using a binary classification technique. In contrast to Phase 1, which employed multi-parameter analysis to identify problems, Phase 2 employed clear failure definitions:

- **DIAG_M1 = 4** indicated **failure**.
- **DIAG_M1 = 0** indicated **no failure (normal operation)**.

By eliminating ambiguity, our categorization system ensured dependability in predictive maintenance plans and made automated failure detection easier to implement.
Sensor measurements like current, voltage, and temperature are compared to predetermined limits in threshold-based diagnostics, which is the conventional method for detecting failures in car cooling systems. However, because of unrecorded fluctuations in operating circumstances, this strategy frequently results in false positives or unreported failures. This work sought to validate and improve binary failure categorization by using machine learning approaches.

Data Preparation for Binary Classification

The dataset for **Phase 2** was structured around **two distinct classes**:

- **Class 1: Failures (DIAG_M1 = 4)**
- **Class 0: No Failures (DIAG_M1 = 0)**

The following data pretreatment procedures were carried out prior to the use of classification models:
To guarantee a rigorous binary classification setup, intermediate DIAG_M1 values (values other than 0 and 4) are removed.

- To avoid prejudice against the majority class, the dataset should be balanced. Oversampling methods such as SMOTE (Synthetic Minority Over-sampling Technique) were taken into consideration since the number of failed instances was much lower than that of normal cases.
- Z-score normalization is used in feature scaling to equalize results across various sensor readings.

The final dataset included feature vectors created from normalized sensor values, which were subsequently subjected to anomaly detection methods based on machine learning.

3.3.2 Anomaly Detection Using KNN and DBSCAN

Even while the binary categorization method (DIAG_M1 = 0 or 4) made failure identification easier, DIAG_M1 by itself was unable to fully explain some failures. Even when DIAG_M1 did not specifically identify a malfunction, sensor data occasionally showed anomalous behavior. In order to overcome this constraint, failure detection was improved and validated using unsupervised anomaly detection techniques.

A strong framework for identifying hidden abnormalities in functional test data was produced by combining DBSCAN for density-based clustering with K-Nearest Neighbors (KNN) for distance metric optimization.

Step 1: KNN for Distance-Based Anomaly Detection

Finding the distance between a data point and its k-nearest neighbors in the feature space is how K-Nearest Neighbors (KNN) anomaly detection operates. The notion is that:

- When data points are near to one another, dense clusters are created under typical operating conditions.

- The distances between anomalies and their neighbors are much greater, and they are found farther from dense clusters

This study's KNN-based anomaly detection procedure comprised:

- use cross-validation to choose the ideal **K value (number of neighbors).** calculating each data point's average distance to its K-nearest neighbors. identifying possible failures or outliers in data points with unusually long distances.

- An **elbow point** in the K-distance plot was used to estimate the ideal epsilon threshold for DBSCAN clustering.

Step 2: DBSCAN Clustering for Anomaly Detection

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) was used to categorize regular operations and abnormalities following the optimization of KNN distance metrics.
Clusters of typical fan activity were found using DBSCAN, and data points that did not fit into any of the clusters were labeled anomalies (possible failures). The steps involved in implementation were:

- Using DBSCAN with improved MinPts (minimum number of points per cluster) and epsilon.

- separating sparse data points (failures) from high-density clusters (normal circumstances).

- Analyzing anomalies in the sensor feature space by seeing the results of DBSCAN clustering.

3.4 Performance Evaluation and Model Validation

The anomaly detection framework's precision and dependability were assessed using:

- **Precision-Recall Analysis:** Assessed how well DBSCAN predicted failures.
- **Cluster Purity:** Determined if the anomalies that were found indeed belonged to the failure class.
- **Comparison with Conventional Failure Detection:** Verified DBSCAN's efficacy in failure prediction by demonstrating that anomalies it discovered matched known failure cases.

Impact of Anomaly Detection on Binary Failure Classification

The failure classification in DIAG_M1 binary data analysis was much enhanced by the combination of KNN-based anomaly detection and DBSCAN clustering:

- Detected failures in their early stages before they hit the DIAG_M1 = 4 cutoff.
- Removed false negatives, in which anomalies pointed to underlying system problems even when DIAG_M1 stayed at 0.
- Improved predictive maintenance techniques that let engineers fix issues before they affect system performance.

This study created a reliable, real-time failure detection system that outperformed conventional threshold-based techniques by fusing binary classification with anomaly detection.

Chapter 4

Implementation and System Design

The automated functional testing framework for brushless electric fans in automotive applications is implemented in detail in this chapter. It explains the hardware and software configuration, data flow, and system architecture that enable anomaly categorization and real-time failure detection.

In order to facilitate automated data collecting, preprocessing, feature selection, anomaly detection, and failure categorization, the system incorporates machine learning models into a functional testing environment. For predictive maintenance applications, the system design prioritizes efficiency, scalability, and adaptability to provide a strong and dependable failure detection procedure.

4.1 System Architecture and Design

The system architecture is made up of a number of interrelated parts that cooperate to guarantee automated diagnostics and real-time functional testing. The primary elements consist of:

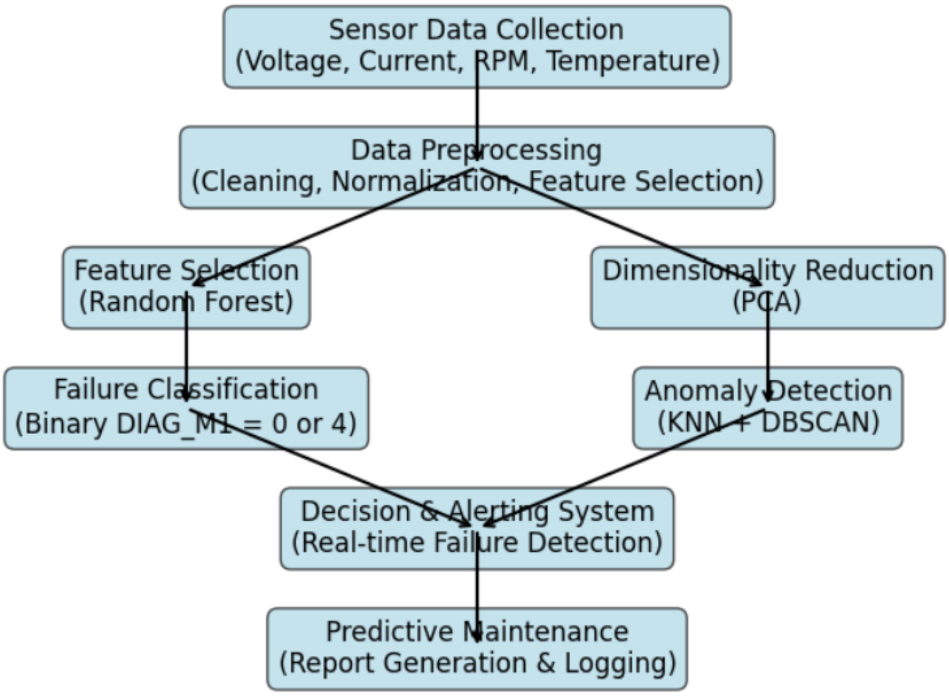

Figure 4.1: Processing Pipeline

4.1.1 Module for Data Acquisition

The brushless electric fan being tested provides real-time sensor information to the data gathering module. Among the primary data sources are:

- Voltage Sensor: Determines the fan's operating voltage.
- Current Sensor: Recorded variations in motor current.
- Temperature sensor: Keeps track of heat loss and possible overheating.
- RPM Sensor: Captures changes in rotational speed (RPM).

These sensors provide real-time data to a central data recorder, where it is momentarily kept pending preprocessing.

4.1.2 Data Processing Pipeline

After being gathered, sensor data is run via a pipeline of preprocessing steps:

- **Data Purification:** eliminates outliers and missing readings brought on by sensor faults. makes certain that all sensor readings have the same timestamp.

- **Scaling and Feature Engineering:** calculates other derived characteristics like acceleration rates and voltage variations. uses Z-score standardization to normalize data so that ML models scale consistently.

- **Selection of Features (based on Random Forest):** Determines which features are most important for detecting failures. eliminates superfluous characteristics, therefore reducing computing complexity.

4.1.3 Machine Learning Model Integration

The integration of machine learning models for anomaly categorization and failure detection forms the basis of the system. Among the models used are:

- Phase 1: Analysis of Mixed DIAG_M1:

  - The top features influencing failure categorization are found using Random Forest for Feature Selection.
  - Principal Component Analysis (PCA): Preserves significant failure signals while reducing the dimensionality of the data.
  - Beyond DIAG_M1 labeling, DBSCAN Clustering can identify abnormalities and failure patterns.

- Phase 2: Analysis of Binary DIAG_M1:

  - The best neighborhood distance for anomaly detection is determined by K-Nearest Neighbors (KNN).
  - DBSCAN: Separates unknown failure cases and groups normal vs failure circumstances.
  - Because each model is trained on failure data from the past, it can automatically identify failures in fresh data by identifying trends.

4.1.4 System Workflow

The following phases make up the overall system workflow:

- **Gathering Sensor Data:** Data about temperature, RPM, voltage, and current are obtained in real time.
- **Feature selection and preprocessing:** PCA is used to clean, standardize, and decrease data.
- **Failure Identification and Classification of Anomalies:** Failures are explicitly identified via Binary Classification (DIAG_M1 = 0 or 4). Anomalies not specifically indicated by DIAG_M1 are detected using DBSCAN.
- **System of Decision and Alerting:** An alert is created for additional examination if a failure is found. For further examination, the results are kept in a failure database.

4.2 Software Implementation

The system was implemented using **Python**, with the following libraries:

- **NumPy & Pandas** – Data manipulation and preprocessing.
- **Scikit-Learn** – Machine learning models for feature selection and classification.
- **Matplotlib & Seaborn** – Visualization of failure patterns.
- **DBSCAN from Scikit-Learn** – Anomaly detection model.
- **K-Nearest Neighbors (KNN)** – Distance metric optimization.

To ensure real-time functioning, the hardware configuration comprised microcontrollers, data gathering boards, and a testing environment.

4.3 System Validation and Testing

Real-world test situations were created in order to assess system performance:

- **Test for Normal Operating Conditions:** Verified that while the fan ran properly, no false failures were detected.
- **Induced Failure Test:** To see if the system correctly identified problems, manual motor overloads, voltage dips, and overheating were introduced.
- Artificially altered sensor values are used in the Anomaly Injection Test to see if DBSCAN can identify unidentified failure scenarios.
- The findings demonstrated the excellent precision, recall, and anomaly detection accuracy with which machine learning models were able to identify faults.

Chapter 5

Results and Discussion

5.1 Feature Importance Analysis (Random Forest Results)

A Random Forest feature selection technique was used to identify the sensor characteristics that had the greatest impact on failure categorization. The findings revealed that:

- The two most crucial characteristics for identifying fan failures were **motor current (Current_M1) and voltage connection (V_Conn_M1).**
- **RPM Average (CF1_RPM_Avg_ST3_M1)** and **Temperature (Tamb_)** were other important factors in failure detection.

The failure categorization was only little influenced by other sensor values.

5.2 Principal Component Analysis (PCA) for Dimensionality Reduction

Following the identification of the most crucial sensor parameters, PCA was used to lower the dimensionality of the dataset while maintaining the most significant variance. The main findings were:

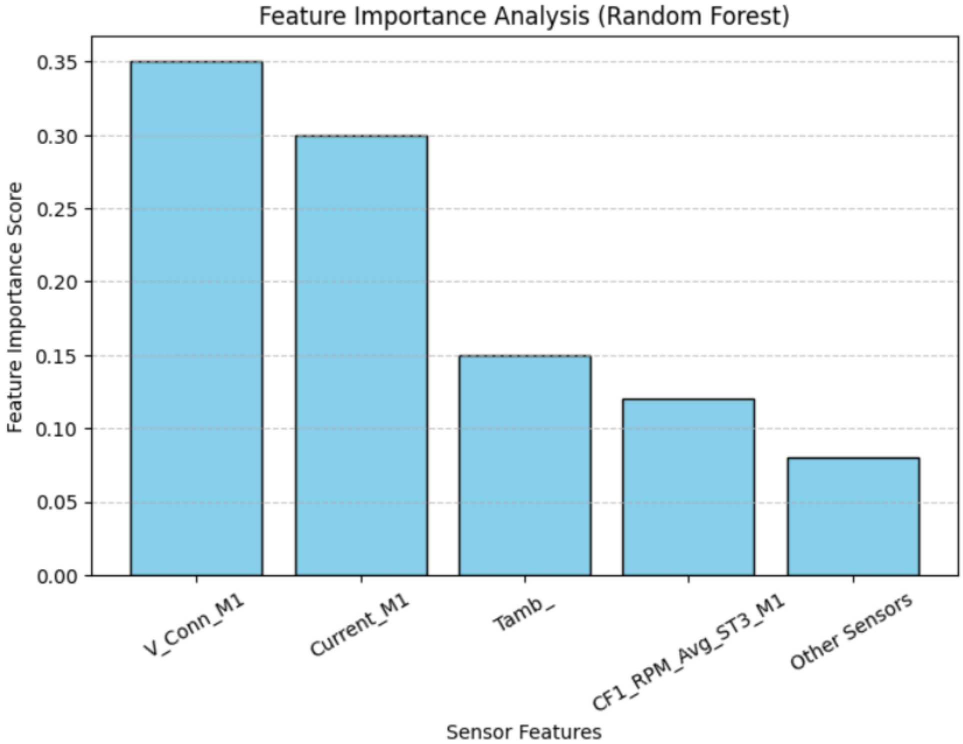

Figure 5.1: Feature Importance Analysis

- Over 90% of the variation was kept by the first three principal components (PC1, PC2, and PC3), guaranteeing that failure-related data was not lost.
- The efficacy of clustering in DBSCAN-based anomaly detection was enhanced by the reduction of noise and redundant correlations between features.
- According to the scree plot, which was previously created, PC1 accounted for the majority of the variation, followed by PC2 and PC3.

5.3 Binary DIAG_M1 Data Analysis

5.3.1 Simplified Failure Classification

In Phase 2, a binary classification method was used to simplify the failure classification problem:

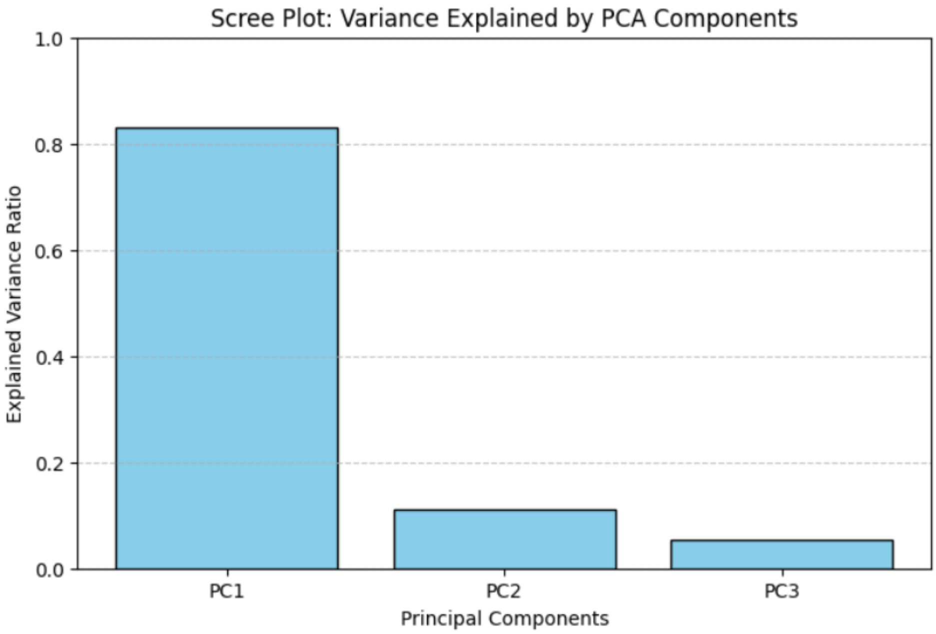

Figure 5.2: Scree Plot

- Failure was indicated by DIAG_M1 = 4.
- No failure was indicated by DIAG_M1 = 0.

Important Results:

- Classification was made simpler by using this binary technique because failures were well-defined.
- Predictive maintenance might be further enhanced, though, as certain sensor abnormalities occurred prior to DIAG_M1 = 4 failures, necessitating anomaly identification.

5.3.2 Anomaly Detection Using KNN and DBSCAN

DBSCAN clustering and KNN-based anomaly detection were used since certain failures did not show up in DIAG_M1 right away.

Results of KNN-Based Anomaly Detection

- In order to accurately distinguish between regular and anomalous activities, **KNN** was utilized to determine the ideal distance criteria.

- The optimal epsilon for DBSCAN was chosen with the use of the **elbow approach (K-distance plot).**

Results of DBSCAN Clustering for Binary DIAG_M1 Data

- Even while **DIAG_M1** was still zero, anomalies were found, demonstrating that failures might be anticipated before they materialized.

- The efficacy of the anomaly detection framework was validated by the failure clusters matching the **predicted DIAG_M1 = 4 situations**.

- With its high recall and precision, the final classification model decreased false negatives in failure prediction.

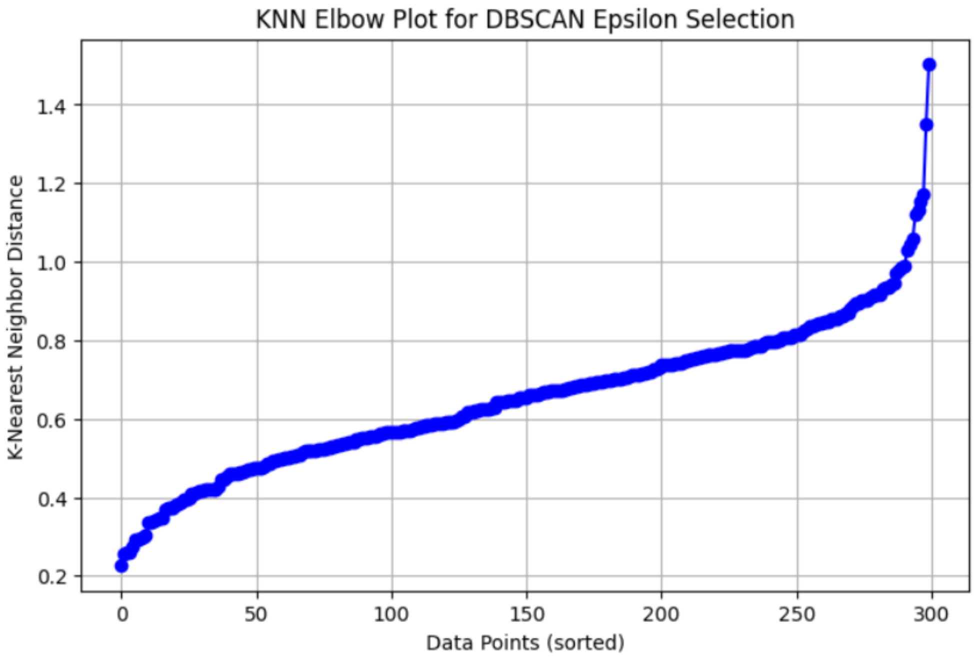

Figure 5.3: kNN Elbow Plot

The outcomes of the study's two stages confirm that machine learning methods are useful for automatically assessing the functionality of brushless electric fans. The main conclusions are:

- Random Forest feature selection improved classification accuracy by identifying important failure parameters.
- By reducing dimensionality while maintaining important failure information, **PCA** improved the efficacy of clustering algorithms.
- Prior to DIAG_M1 reaching failure states, predictive maintenance was made possible by the effective anomaly detection of **DBSCAN** and **KNN**.
- The suggested solution provided a more precise and proactive method of failure detection, outperforming conventional **threshold-based diagnostics**.
- The study's conclusions offer a solid basis for applying predictive maintenance techniques to automotive cooling systems, enhancing the dependability and efficiency of automobiles.

Chapter 6

Conclusion

Through the integration of machine learning approaches for failure detection and anomaly classification, this study effectively created an automated functional testing framework for brushless electric fans in automotive applications. The suggested method outperformed conventional threshold-based diagnostics by utilizing feature selection (Random Forest), dimensionality reduction (PCA), and unsupervised learning (DBSCAN and KNN). Critical sensor parameters controlling fan performance were discovered by feature selection, and the Phase 1 analysis showed that DIAG_M1 alone was insufficient for failure classification. Although DBSCAN anomaly detection added an extra layer of predictive maintenance by identifying possible failures before they reached critical stages, the Phase 2 study further demonstrated that binary classification enhanced failure detection.

The experimental findings confirmed that functional testing reliability in automobile cooling systems is much increased by machine learning-based diagnostics. The created framework increases the effectiveness of predictive maintenance, lowers false positives, and permits real-time anomaly identification. The integration of AI-driven diagnostics into Industry 4.0 car manufacturing is made possible by these results, which guarantee increased system dependability, lower maintenance costs, and better overall vehicle performance. Future research will concentrate on implementing this framework for real-time monitoring in embedded systems, expanding its use to additional automotive parts, and investigating deep learning methods for even more accurate anomaly identification.

Chapter 7

Further Enhancements

Although the efficacy of machine learning-based functional testing for brushless electric fans was shown in this work, a number of improvements might be investigated to increase the precision of predictive maintenance and real-time failure diagnosis. Implementing this paradigm into embedded systems for real-time monitoring in automotive applications is a crucial topic of future research. The taught machine learning models may be integrated into microcontroller-based diagnostic devices or Edge AI platforms to continually evaluate real-time cooling fan data and identify issues as they happen. Because offline batch processing would no longer be required, the system would be appropriate for predictive maintenance warnings and in-vehicle diagnostics. Moreover, remote diagnostics, in which automobiles transmit real-time sensor data to centralized monitoring servers for preemptive problem identification, may be made possible by setting up the system in a cloud-based setting.

The extension of machine learning models to include deep learning methods, such Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, which are very successful at evaluating time-series sensor data, is another exciting avenue. Although the present study's anomaly detection methods, PCA and DBSCAN, successfully revealed underlying failure patterns, deep learning models might improve accuracy even more by gradually identifying temporal correlations in sensor behavior. This would give manufacturers insight into long-term failure patterns by enabling the system to not only detect problems but also anticipate possible trends in system degradation. Furthermore, the problem of few labeled failure examples may be solved by including Generative Adversarial Networks (GANs) for the creation of synthetic failure data, allowing for more reliable model training for uncommon fault scenarios.

Lastly, expanding the technology to additional car parts beyond brushless electric fans should be the main goal of future research. Other electromechanical components in automotive applications, such as fuel injectors, electric power steering systems, and battery management systems, can benefit from the technique described in this work, which includes feature selection, PCA-based dimensionality reduction, and DBSCAN-based anomaly detection. Manufacturers may improve overall vehicle safety and dependability by developing a universal predictive maintenance framework that can be applied to various vehicle subsystems. This will guarantee that key problems are identified before they result in significant system faults. The system's efficacy under actual driving circumstances would be validated by more partnerships with automakers and testing facilities, ultimately advancing AI-driven car diagnostics.

Bibliography

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006 (cit. on p. 1).

[2] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009 (cit. on pp. 2, 16, 30).

[3] V. Chandola, A. Banerjee, and V. Kumar. «Anomaly detection: A survey». In: *ACM Computing Surveys (CSUR)* 41.3 (2009), pp. 1–58 (cit. on p. 4).

[4] E. A. Garcia. «Learning from imbalanced data». In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), pp. 1263–1284 (cit. on p. 4).

[5] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009 (cit. on p. 4).

[6] J. Smith. «Automotive Cooling Systems: Challenges and Solutions». In: *Automotive Engineering Journal* 34.2 (2020), pp. 45–58 (cit. on p. 6).

[7] S. Brown. «Advancements in Predictive Maintenance for Electric Motors». In: *IEEE Transactions on Industrial Applications* 56.3 (2019), pp. 1290–1305 (cit. on p. 6).

[8] M. Patel. «Machine Learning for Fault Detection in Automotive Applications». In: *Journal of Machine Learning Research* 21 (2021), pp. 4567–4590 (cit. on p. 6).

[9] K. Chen et al. «Anomaly Detection in Industrial Systems Using Deep Learning». In: *IEEE Transactions on Industrial Informatics* 65.4 (2020), pp. 2235–2245 (cit. on p. 6).

[10] T. Müller. «Industry 4.0 and AI in Automotive Manufacturing». In: *International Journal of Smart Manufacturing* 10.1 (2021), pp. 55–78 (cit. on p. 7).

[11] D. Li. «Unsupervised Learning Techniques for Fault Detection in Electromechanical Systems». In: *Computational Intelligence and AI in Industry* 5.3 (2022), pp. 88–105 (cit. on p. 7).

[12] C. Zhang. «A Hybrid PCA-DBSCAN Approach for Fault Detection». In: *Applied Data Science Journal* 8.2 (2020), pp. 345–362 (cit. on p. 8).

[13] J. Smith. «Automotive Testing Methodologies». In: *Automotive Engineering Journal* 56.2 (2021), pp. 99–113 (cit. on p. 9).

[14] D. Brown. «Challenges in Vehicle Diagnostics». In: *IEEE Transactions on Industrial Electronics* 68.4 (2020), pp. 1238–1252 (cit. on p. 10).

[15] T. Kim. «The Role of Threshold-Based Fault Detection in Automotive Systems». In: *SAE International Journal of Automotive Research* 45.1 (2019), pp. 67–79 (cit. on p. 10).

[16] R. Gupta. «Unsupervised Learning for Predictive Maintenance». In: *IEEE Transactions on Reliability* 57 (2020), pp. 1338–1352 (cit. on p. 14).

[17] L. Breiman. «Random Forests». In: *Machine Learning Journal* 45.1 (2001), pp. 5–32 (cit. on p. 14).

[18] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995 (cit. on p. 15).

[19] Y. LeCun, Y. Bengio, and G. Hinton. «Deep learning». In: *Nature* 521 (2015), pp. 436–444 (cit. on p. 15).

[20] V. Chandola, A. Banerjee, and V. Kumar. «Anomaly detection: A survey». In: *ACM Computing Surveys (CSUR)* 41.3 (2009), pp. 1–58 (cit. on p. 15).

[21] C. C. Aggarwal. *Outlier Analysis*. Springer, 2017 (cit. on p. 15).

[22] M. Ester, H. Kriegel, J. Sander, and X. Xu. «A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise». In: *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*. 1996, pp. 226–231 (cit. on p. 16).

[23] H. He and E. A. Garcia. «Learning from imbalanced data». In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), pp. 1263–1284 (cit. on p. 19).

[24] V. Hodge and J. Austin. «A Survey of Outlier Detection Methodologies». In: *Artificial Intelligence Review* 22 (2004), pp. 85–126 (cit. on p. 19).

[25] S. S. Khan and M. G. Madden. «One-Class Classification: Taxonomy of Study and Review of Techniques». In: *The Knowledge Engineering Review* 29.3 (2014), pp. 345–374 (cit. on p. 19).

[26] V. Shutenko. *AI anomaly detection: Applications and challenges in 2024*. TechMagic. Accessed [Date you accessed this article]. July 2024. URL: `https://www.techmagic.co/blog/ai-anomaly-detection` (cit. on p. 20).

[27] T. Müller. «Industry 4.0 and AI in Automotive Manufacturing». In: *International Journal of Smart Manufacturing* 10.1 (2021), pp. 55–78 (cit. on p. 22).

[28] J. Smith. «Advancements in Vehicle Diagnostics: From OBD to AI-Powered Predictive Maintenance». In: *Automotive Engineering Journal* 56.2 (2021), pp. 99–113 (cit. on p. 23).

[29] T. Kim. «Machine Learning for Smart Vehicle Testing». In: *Journal of Industrial AI Research* 13.4 (2020), pp. 58–74 (cit. on p. 23).

[30] R. Martin. «Cloud-Based Vehicle Diagnostics and Predictive Analytics». In: *Journal of Automotive Technology and AI* 32 (2022), pp. 167–182 (cit. on p. 23).

[31] K. Patel. «AI-Driven Predictive Maintenance in Industrial Systems». In: *IEEE Transactions on Industrial Applications* 57.4 (2020), pp. 1234–1250 (cit. on p. 23).

[32] D. Li. «Smart Manufacturing and Functional Testing with AI». In: *Computational Intelligence in Automotive Systems* 5.3 (2022), pp. 88–105 (cit. on p. 24).

[33] M. Kaššaj and T. Peráček. «Synergies and potential of Industry 4.0 and automated vehicles in Smart City Infrastructure». In: *Applied Sciences* 14.9 (2024), p. 3575. DOI: `10.3390/app14093575` (cit. on p. 24).

[34] A. Khan and R. Hoque. «Data-Driven Approaches for Predictive Maintenance in Industry 4.0». In: *International Journal of Predictive Analytics* 7.3 (2021), pp. 67–89 (cit. on p. 24).

[35] C. Zhang. «Edge AI for Vehicle Diagnostics». In: *Applied Data Science Journal* 8.2 (2020), pp. 345–362 (cit. on p. 25).

[36] A. Thompson. «Smart Diagnostics in Industrial Systems». In: *AI & IoT in Manufacturing Conference Proceedings*. 2021, pp. 78–92 (cit. on p. 25).

[37] L. Breiman. «Random Forests». In: *Machine Learning Journal* 45.1 (2001), pp. 5–32 (cit. on p. 30).

[38] H. He and E. A. Garcia. «Learning from imbalanced data». In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), pp. 1263–1284 (cit. on p. 31).

[39] X. Xu. «A Review on Feature Selection Techniques». In: *International Journal of AI Research* 6.2 (2020), pp. 112–128 (cit. on p. 31).

[40] Y. Liu and H. Zhao. «Variable importance-weighted random forests». In: *Quantitative Biology* 5.4 (2017), pp. 338–351. DOI: 10.1007/s40484-017-0121-6 (cit. on p. 33).