# An IIoT Cloud Based Solution for Fault Diagnosis of Induction Motors

Author
Gilbert Delgado

A thesis submitted in partial fulfillment of the requirements for the degree of Master in Telecommunications Engineering

Advisor
Prof. Juan Manuel Rey López
PhD in Electronic Engineering

Co-advisor
Prof. Carlos Augusto Fajardo Ariza
PhD in Engineering

Universidad Industrial de Santander
Facultad de Ingenierías Fisicomecánicas
Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T)
Maestría en Ingeniería de Telecomunicaciones
2025

## Dedicatory

*Her, who gave me my heart to live.*
*Her, who gave my heart the care to live.*
*And her, to whom I gave my heart to live with.*

## Acknowledgments

I would like to express my sincere gratitude to my director, Juan Manuel Rey, for his invaluable guidance and continuous support throughout the development of this work. His expertise and clear vision were fundamental to shaping this project. I am also deeply thankful to my codirector, Carlos Fajardo Ariza, whose encouraging words and thoughtful insights provided motivation at every stage. Finally, I extend my appreciation to the enterprise Dautom, whose deep knowledge and experience in the industrial sector greatly enriched the practical and technical aspects of this research.

# Contents

# List of Tables

# List of Figures

# Appendices

# Abstract

**Title:** An IIoT Cloud Based Solution for Fault Diagnosis of Induction Motors [*]

**Author:** Gilbert Delgado [**]

**Keywords:** IIoT, Cloud, Fault Diagnosis, Induction Motors, Machine Learning

**Description:** Guided by the principles of Industry 4.0, enterprises are undergoing a digital transformation emphasizing interconnected, intelligent, and data-driven operations. In this context, industrial data has become a strategic asset for optimizing performance and enabling informed decision-making. Industrial Internet of Things (IIoT) technologies empower enterprises to centralize and leverage real-time data across all layers of the automation stack. Beyond supply chain management and direct sales applications, there is a growing demand for data-driven strategies in predictive maintenance (PdM) and fault diagnosis (FD) to mitigate process disruptions. Induction motors, essential components in industrial machinery, are susceptible to failures that can lead to production slowdowns, unplanned downtime, and increased operational costs. However, many existing approaches fail to fully account for real-time data flow and industrial constraints, limiting their effectiveness in real-world deployments. This work proposes a cloud-based solution for fault diagnosis in induction motors, integrating machine learning algorithms with advanced data preprocessing techniques. The solution leverages the Unified Namespace (UNS) paradigm, enabling seamless data storage, analysis, and visualization. This approach not only provides a practical framework adaptable to real industrial environments but also contributes to academic advancement in intelligent fault diagnosis by integrating emerging technologies such as IIoT, machine learning, and data-driven architectures.

---

[*] MSc Thesis

[**] Faculty of Physic-Mechanical Engineering. Department of Electrical, Electronic and Telecommunications Engineering (E3T). Advisor: Juan M. Rey López, PhD in Electronic Engineering. Co-advisor: Carlos A. Fajardo Ariza, PhD in Engineering

# Resumen

**Título:** Una solución IIoT basada en la nube para el diagnóstico de fallas en motores de inducción [*]

**Autor:** Gilbert Delgado [**]

**Palabras clave:** IIoT, Nube, Diagnóstico de Fallas, Motores de Inducción, Aprendizaje de Máquina.

**Descripción:** Guiadas por los principios de la Industria 4.0, las empresas están atravesando una transformación digital que enfatiza operaciones interconectadas, inteligentes y basadas en datos. En este contexto, los datos industriales se han convertido en un activo estratégico para optimizar el rendimiento y facilitar la toma de decisiones informadas. Las tecnologías del Internet Industrial de las Cosas (IIoT) permiten centralizar y aprovechar datos en tiempo real a lo largo de todos los niveles de la arquitectura de automatización industrial. Más allá de su aplicación en la gestión de la cadena de suministro y las ventas directas, existe una demanda creciente de estrategias basadas en datos para el mantenimiento predictivo (PdM por sus siglas en inglés) y el diagnóstico de fallos (FD por sus siglas en inglés), con el fin de mitigar interrupciones en los procesos. Los motores de inducción, componentes fundamentales en la maquinaria industrial, son propensos a fallos que pueden provocar desaceleraciones en la producción, paradas no planificadas y un aumento en los costos operativos. Sin embargo, muchos enfoques existentes no consideran adecuadamente el flujo de datos en tiempo real ni las características propias del entorno industrial, lo que limita su efectividad en implementaciones prácticas. Este trabajo propone una solución en la nube para el diagnóstico de fallas en motores de inducción, integrando algoritmos de aprendizaje automático con técnicas avanzadas de preprocesamiento de datos. La solución emplea el paradigma del *Unified Namespace* (UNS), lo que permite un almacenamiento, análisis y visualización de datos sin interrupciones. Esta propuesta no solo aporta un marco práctico adaptable a entornos industriales reales, sino que también contribuye al avance académico en el diagnóstico inteligente de fallas, al integrar tecnologías emergentes como IIoT, aprendizaje automático y arquitecturas orientadas a datos.

---

[*] Tesis de Maestría

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones. Director: Juan M. Rey López, Doctor en Ingeniería Electrónica. Co-director: Carlos A. Fajardo Ariza, Doctor en Ingeniería.

# Introduction

In 2008, Terminal 5 at London Heathrow Airport was inaugurated as the new hub for British Airways, following two decades of planning and a £4.3 billion investment. British Airways, the UK's second-largest airline after easyJet, experienced a turbulent launch, with at least 34 flights cancelled and thousands of passengers affected. The disruptions were largely attributed to technical failures, particularly in the intelligent baggage handling and transport system (Milmo, 2008). In industrial settings, motors are fundamental components of machinery, driving essential mechanical operations. Like all equipment, they are susceptible to faults, which can result in costly downtime, economic losses, and even personal injury (Jiang et al., 2022; S. Zhang et al., 2021). The growing interest in machine learning (ML) techniques represents a valuable opportunity to develop advanced predictive algorithms. Fault Diagnosis (FD), a closely related field, can harness the power of the Industrial Internet of Things (IIoT) to monitor and manage machinery in industrial environments, where machinery faults often lead to production slowdowns, increased downtime, and rising operational costs. FD is especially critical in manufacturing systems, where the timely detection of faults in rotating machines directly affects production safety, product quality, and maintenance efficiency (Fu et al., 2020). This work explores a solution for diagnosing faults in induction motors taking advantage of IoT technologies in industrial environments, integrating data acquisition with cloud-based data management.

Most state-of-the-art (SOTA) solutions are designed to address hypothetical scenarios and rely primarily on data collected under controlled laboratory conditions (Chakraborty et al., 2022; Tran et al., 2023). However, it is essential to evaluate the possibility of extending these tools to industrial applications, which have a lot of practical implications. The primary goal is to build a solution that integrates real-time data with processing and visualization. Achieving this requires breaking down the overall process into distinct stages and then integrating them into a cohesive system. This approach, however, presents several challenges in both the decomposition and integration phases.

The system begins by collecting real-time motor data, including vibration, temperature, and current measurements. This data is stored and visualized, making it easier to access and analyze. The next step involves pre-processing the stored data to train the ML algorithm responsible for diagnosing faults in the motor. Fault diagnosis is performed using the data, with the system continuously acquiring and updating information. All components are monitored in real time through an Industrial Internet of Things (IIoT) interface, enabling remote access and oversight from anywhere.

The proposed solution aims to make a valuable contribution to the research community by implementing an ML algorithm using real-time industrial data. It will consider factors such as the amount of data available, the constant flow of data, the sampling frequency, and normalization. The results will demonstrate the effectiveness and applicability of the solution, including the specific ML and pre-processing algorithms used.

# 1. State of the Art

Technology has made significant advances in Machine Learning (ML) models. These models can identify patterns and assess features that may not be evident to humans. This capacity has proven to be useful in various disciplines (Abualsauod, 2023; Angarita-Zapata et al., 2021; Jovicic et al., 2023; Sharma & Liu, 2021). In an IoT (Internet of Things) environment, interactions between devices are essential (Fathy et al., 2021; Garrido-Labrador et al., 2020). Therefore, predicting system failures becomes critical, especially in the industrial sector where unexpected downtime can result in significant economic losses (S. Zhang et al., 2021).

Diagnosis, prevention, and maintenance of electrical machines in constant operation have been studied extensively across various fields. Terms like Condition Monitoring (CM) (Aqueveque et al., 2021; Souza et al., 2021), Predictive Maintenance (PdM) (Huang et al., 2023; Rosati et al., 2023), and Fault Diagnosis (FD) (Li et al., 2021; Song et al., 2020) are commonly used in these studies. The objective of these techniques is to explore whether there is an effective way to monitor and maintain electrical machinery in use.

These terms have many applications in IoT technology, such as using DC motors in drones to predict falling risks (Hong, 2022) and diagnosing rotating machines in industries where temperature is crucial (Mokkamayan & Thankappan, 2023). In this work, we will concentrate on the induction motor, an essential device used in every factory across the global industry.

Fault diagnosis refers to the process of identifying and detecting faults in industrial machines and equipment (Fu et al., 2020). It plays a critical role in ensuring production safety, maintaining product quality, and managing maintenance costs. The development of the IIoT has provided opportunities for effective fault diagnosis by utilizing multi-source monitoring information and implementing data fusion techniques (Jiang et al., 2022).

A range of methods and algorithms have been proposed to enhance fault diagnosis, including Bayesian-optimized decision trees with ensemble classifiers, dynamic routing-based multimodal neural networks, and feature extraction based on oriented support vector machines (SVM) (Asad et al., 2023; Fu et al., 2020; Jiang et al., 2022). These approaches aim to improve the accuracy and efficiency of fault diagnosis by extracting relevant features, analyzing vibration and current signals, and enhancing the resolution of frequency spectra. By combining these techniques with IoT and cloud computing, real-time fault diagnosis and prediction can be achieved, ultimately enhancing the reliability, scalability, and performance of industrial automation systems.

Numerous studies on PdM and FD have primarily involved an experiment where electromechanical variables like current, voltage, and vibration are measured (Chang et al., 2022; Huang et al., 2023; Rubio et al., 2018; Tran et al., 2023). These variables are then used to extract features from a rotating machine that enables automated learning systems to diagnose and predict a possible failure (Jung et al., 2023).

Bearing faults and incipient short-circuit faults are two common failures in induction machines (motors). A recent study presented in (W. J. Lee et al., 2021) demonstrated the effectiveness of a deep learning–based monitoring approach for detecting bearing faults, which significantly outperformed traditional ML methods. Similarly, Cunha et al. (2021) proposed a reliable method for detecting incipient short-circuit faults. Their approach utilizes the voltage signal induced by the motor axial flux. X. R. Zhang et al. (2021) highlighted the importance of FD in safeguarding the machine against overcurrent and overvoltage. However, as the complexity of industrial systems and the volume of data increase, enhancing FD capabilities becomes imperative. In this context, algorithms such as support vector machines (SVMs) play a critical role by extracting relevant features from real-time motor data to improve fault detection accuracy and reliability.

On the other hand, Zayed et al. (2023) proposes a Digital Twins (DT) model to collect vast amounts of data for posterior feature selection with a proposed a ML-based framework. Similarly, Souza et al. (2021) used unsupervised artificial neural networks, specifically Auto Encoders (AE), reducing dimensionality and identify the parameters that best represent data. After collecting data, ML algorithms analyze the featured data to diagnose and predict with varying degrees of emphasis.

Fault diagnosis on induction motors can be achieved using various techniques and approaches. One effective approach is using Long Short-Term Memory (LSTM) networks, which have shown promising results in diagnosing motor failure and predicting the probability of failure within a specified time frame (Kizito et al., 2021). Another promising method is the Bayesian-optimized decision tree with ensemble classifiers, which has been found to perform well in motor bearing fault diagnosis (Jiang et al., 2022). Convolutional Neural Networks (CNN) have also been widely used in this domain. For instance, CNN-based deep learning models have been employed to develop data-driven condition monitoring systems for detecting faults in induction motors (Irgat et al., 2022). In addition, the Dynamic Routing-based Multimodal Neural Network (DRMNN) has been introduced for data fusion-based diagnosis, outperforming other fusion strategies in terms of accuracy and robustness (Fu et al., 2020). Finally, custom CNN architectures have shown excellent performance in fault detection, surpassing transfer learning-based models (Chakraborty et al., 2022).

Chakraborty et al. (2022) reported strong performance in their proposed algorithm. However, these results were achieved under controlled laboratory conditions and still require validation in real-world industrial environments. Deploying such systems in industry remains challenging due to the complexity introduced by noisy data. To address this, Jarwar et al. (2023) proposed a noisy encoder utilizing AI (NEAT) along with AE to improve the robustness and reliability of the FD algorithm.

Tran et al. (2023) have proposed a new fault recognition and correction (FRC) system for induction motors (IMs) that utilizes IoT and deep learning. The system uses vibration signals during the motor's operation to extract bearing fault characteristics, which are fed into a deep neural network (DNN) model designed to distinguish four bearing conditions: normal conditions, inner bearing faults, outer bearing faults, and cyber-attacks. The approach has also been tested for its robustness against false data injection (FDI) attacks—cybersecurity threats in which attackers manipulate sensor data to mislead diagnostic systems. Experimental data was used to create training and testing datasets, with 70% allocated for training and 30% for testing. The model's performance was evaluated using a categorical cross-entropy loss function and the corrected linear unit as the activation function. According to the authors, the proposed method outperforms existing fault recognition techniques, and its performance could be further enhanced by incorporating additional sensor data such as temperature and current. Tran et al. (2023) claims that future research may focus on optimizing the deep learning architecture and fine-tuning training parameters to improve the accuracy, robustness, and efficiency of the anomaly detection and correction framework.

Despite the remarkable progress in ML and fault diagnosis techniques, most existing solutions remain confined to controlled environments or limited-scale deployments. Real-world industrial systems generate large volumes of heterogeneous data across multiple sources, including vibration, temperature, current, and operational states, making seamless data integration a key challenge. Furthermore, the need for scalable, remote-accessible solutions is growing as enterprises adopt cloud architectures and IIoT frameworks. In this context, the development of a cloud-based fault diagnosis system for induction motors addresses a critical industrial need. By enabling centralized data collection, real-time analytics, and integration with diverse sensors and enterprise systems, such a solution offers a practical, scalable, and efficient approach to predictive maintenance. This work is particularly relevant for industries seeking to minimize downtime, enhance decision-making, and transition toward intelligent, interconnected operations in line with the principles of Industry 4.0.

# 2. Objectives

## 2.1. General Objective

To design an Induction Motor Fault Diagnosis solution based on a cloud IIoT framework.

## 2.2. Specific Objectives

To consolidate a database with data gathered from motor sensing in a real application.

To preprocess the database to facilitate the training process of models for Induction Motor fault diagnosis.

Select an ML algorithm for the fault diagnosis of induction motors.

To deploy the solution approach on a IIoT environment on cloud.

## 3. Methodology

This project is structured around three core concepts: data, FD, and IIoT. Accordingly, the methodology is divided into three main stages to develop a cloud-based solution for the fault diagnosis of induction motors:

- **Data Selection and State-of-the-Art Review:** This stage involves identifying and analyzing relevant datasets that contain motor condition data, as well as conducting a comprehensive literature review of current methods and technologies used in induction motor fault diagnosis.
- **Machine Learning Model Selection and Implementation:** In this phase, appropriate machine learning algorithms are selected, designed, and trained to detect and classify faults effectively. The process includes data preprocessing, model architecture selection, and performance evaluation using established metrics.
- **Deployment of an IIoT Cloud-Based Solution:** The final stage focuses on the integration of the trained models within an IIoT infrastructure. This includes real-time data acquisition from edge devices, cloud-based processing, and visualization interfaces for remote monitoring and decision-making support.

The following chapters will elaborate on each of these stages, detailing the specific methodologies, tools, and results obtained throughout the development of the proposed solution.

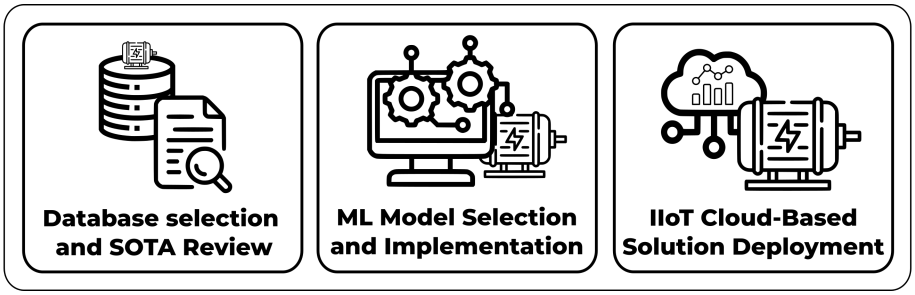


**Figure 1.** *Methodology summary.*

# 4. Data selection and State of the Art

In machine learning projects, particularly those involving classification, data is a foundational resource that requires careful selection and preparation to ensure high performance. This stage focuses on selecting a suitable Fault Diagnosis of Induction Motors (FDoIM) database based on essential characteristics relevant to the project. Once selected, the data undergoes preprocessing to extract crucial features, thereby enhancing the performance of predictive algorithms.

## 4.1. Identifying relevant variables for FD of IM

A comprehensive SOTA review on induction motor FD identified 17 key references (Asad et al., 2023; Chakraborty et al., 2022; Cunha et al., 2021; Fu et al., 2020; Irgat et al., 2022; Jaen-Cuellar et al., 2023; Jiang et al., 2022; Kizito et al., 2021; Kovilpillai & Jayanthy, 2022; W. J. Lee et al., 2021; Ling et al., 2023; Rajput et al., 2023; Rubio et al., 2018; Tran et al., 2021, 2023; Vermesan et al., 2022; X. R. Zhang et al., 2021) aligned with the Cloud-Based Solution framework. This review emphasized critical sensed variables, each with advantages and limitations.

- **Vibration:** the most widely utilized variable in machine learning-based fault detection for induction motors, primarily due to its strong correlation with common mechanical faults such as unbalance, misalignment, and bearing failures.

- **Acoustic Signals:** Acoustic signals offer diagnostic capabilities comparable to vibration analysis and can be effective in detecting mechanical anomalies. However, their adoption in industrial environments remains limited due to high noise, which can significantly compromise signal reliability and accuracy.

- **Electrical Signals (Current and Voltage):** Initially leveraged for Motor Current Signature Analysis (MCSA), this non-invasive approach has demonstrated effectiveness in real-time fault detection, making it highly valuable for industrial applications.

- **Temperature:** Temperature sensors provide critical insights into motor health by detecting overheating, insulation breakdown, and excessive friction in bearings. While temperature changes may lag early fault symptoms detected by vibration or current analysis, thermal monitoring is essential for long-term asset health management and preventive maintenance strategies.

Among all sensed variables, vibration has proven to be the most relevant for FD of induction motors, often providing sufficient information to address the classification problem effectively. Because of this, after selecting the database, our analysis will focus exclusively on vibration data. However, the inclusion of additional sensor variables, such as temperature and current, remains essential for developing a more robust and scalable solution. Scalability is a critical consideration in real-world industrial environments, where systems must adapt to varying conditions and operational complexities.

### 4.2. Database search and comparison

After identifying the most relevant variables for effective FD in induction motors, the next step involved searching for databases that contained these characteristics. The selection process emphasized datasets that offered comprehensive and representative information for model training and evaluation. Three main publicly available datasets were selected and summarized in Table 1.

**KAIST Dataset:** Compiled at the Korea Advanced Institute of Science and Technology and hosted by Elsevier, this dataset contains time-series data from rotating machines operating under various conditions. Faults such as bearing defects, misalignment, and imbalance were physically induced in controlled laboratory experiments, producing realistic and high-fidelity signals suitable for condition monitoring and FD tasks (Jung et al., 2023).

**MAFAULDA Dataset:** This database offers 1,951 multivariate time-series samples collected using the Spectra Quest Machinery Fault Simulator (MFS). The dataset includes a wide range of fault scenarios designed to support the evaluation of fault detection and classification techniques (Ribeiro, 2020).

**BBAR Dataset:** Developed through experimental tests on three-phase induction motors, this dataset includes both electrical and mechanical signals. It focuses on rotor faults, particularly broken bars with varying severities, and provides a valuable benchmark for studying rotor-related anomalies (Treml et al., 2020).

**Table 1.** *Summary of databases for Fault Diagnosis on Induction Motors*

| Features | KAIST | MAFAULDA | | BBAR |
|---|---|---|---|---|
| Variables | Vibration (g) | Vibration (g) | | Vibration (g) |
| | Acoustic (Pa) | Acoustic (Pa) | | |
| | Temperature (°C) | Frequency (RPM) | | Current (A) |
| | Current (A) | | | Voltage (V) |
| Sampling Rate | 51.2 kHz acoustic | 50kHz | | 55.6 kHz electrical |
| | 25.6 kHz else | | | 7.6 kHz vibration |
| Operating conditions | Normal | Normal | | Normal |
| | Unbalance | Unbalance | | 1 broken bar |
| | Misalign | Misalign | Horizontal | 2 broken bars |
| | Inner race fault | | Vertical | 3 broken bars |
| | Outer race fault | Underhang Bearing | Cage fault | 4 broken bars |
| | Ball fault | | Outer race | |
| | | Overhang Bearing | Ball fault | |
| Variations | Load (Nm) | | | Load (Torque) |
| | Speed (RPM) | Speed (RPM) | | |
| | Fault Severity | Fault Severity | | |
| Data size | 7.63 GB | 13 GB | | 6.86 GB |
| Public | Yes | Yes | | Yes |
| Total classes | 6 | 10 | | 4 |

### 4.3. Database selection and characterization

When comparing the characteristics of the datasets, the KAIST dataset stands out for its comprehensive range of electro-mechanical variables and a sufficiently diverse set of fault classes. In contrast, the BBAR dataset, while offering a robust collection of sensed variables, is limited by a narrower variety of fault classes. The MAFAULDA dataset includes a larger number of fault classes, but these are highly specific to its particular data collection setup. Overall, the KAIST dataset emerges as the most suitable option, offering a diverse array of sensed variables, a generalized set of fault classes, and operational variations such as changes in load and speed. These attributes collectively enhance model generalization.

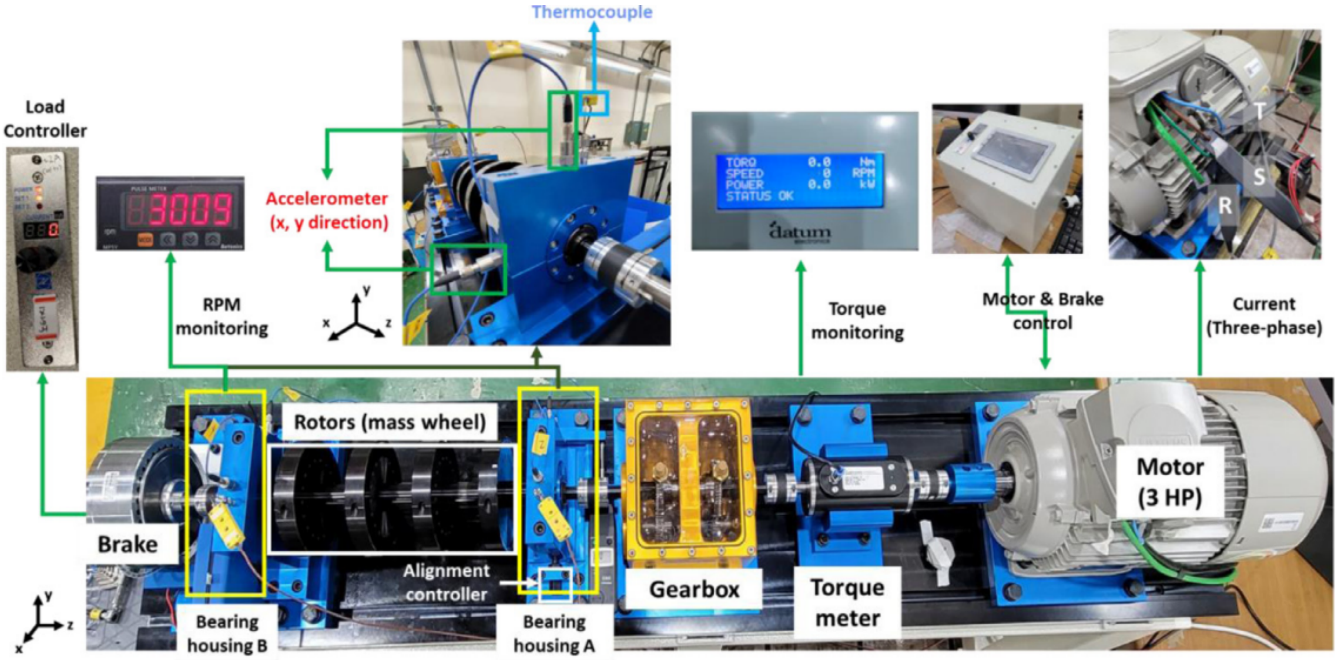


**Figure 2.** *KAIST Dataset Testbed.*

#### KAIST Dataset Overview (Jung et al., 2023)

Figure 2 shows the KAIST testbed, which consists of:

- **Three-phase induction motor:** This is a four-pole AC motor manufactured by SIEMENS, with a power rating of 3 HP. It operates at 380 V, 60 Hz, with a rated speed of 1770 rpm.

- **Gearbox:** The gearbox increases the rotational speed by a factor of 2.07, allowing the motor to reach speeds of up to 3663 rpm. To avoid signal overlapping with the driving frequency of 60 Hz, the dataset was collected at an operating speed of 3010 rpm.
- **Bearing Houses A & B:** Each housing has four accelerometers (two per housing) following the vibration installation guidelines specified in ISO 10816-1:1995
- **Hysteresis brake:** This component is used to apply load to the system.
- **Torque Meter:** A torque meter from Datum Electronics M425 was utilized to measure the applied load.

**Table 2.** *Dataset symbology*

| Symbol | Meaning | Concept | Types |
|---|---|---|---|
| L | Load | Indicates the load applied to the machine during operation | - |
| Nm | Newton meter | Unit of applied load | - |
| FaultType | Fault type | Specifies the machine's condition | • Normal<br>• BPFI (Bearing Pass Fault - Inner Race)<br>• BPFO (Bearing Pass Fault - Outer Race)<br>• Misalignment<br>• Unbalance |
| SS | Fault Severity | Defines the severity of the fault conditions | • 0.3, 1 3 mm for BPFI BPFO<br>• 0.1, 0.3 0.5 mm for Misalignment<br>• 583, 1169, 1751 3318 mg for Unbalance |

***Operating conditions and File format description:*** The dataset follows a structured file naming convention

*LNm_FaultType_SS*

which is explained in Table 2.

***Sensor data and data shape:*** The KAIST dataset is organized into three main folders, each containing specific types of sensor data. The dataset was collected for 120 seconds in normal state, and for 60 seconds in faulty state. A more detailed table can be found on Appendix A.

***Domains:*** In ML for fault diagnosis, a domain refers to the operating conditions under which data is collected (e.g., speed, load), while a label indicates the machine's actual state (e.g., normal, bearing fault). The same fault can occur across different domains, but models trained in one may not perform well in others due to changes in data distribution. In this case we define Domains 0, 1 and 2 as pointed by Usman et al. (2024) in Table 3.

**Table 3.** *Considered Domains for KAIST dataset*

| Domain | Load (Nm) | Vibration Data Files (#) |
|---|---|---|
| 0 | 0 | 5 |
| 1 | 2 | 5 |
| 2 | 4 | 5 |

Domain 0 corresponds to a no-load condition (0 Nm) under which five vibration data files were collected. Within this domain, the classification task involves distinguishing between five health states of the motor: Class 0 (Normal) indicates standard operation without faults. Class 1 (BPFI) and Class 2 (BPFO) represent bearing faults occurring at the inner and outer races, respectively. Class 3 (Misalignment) refers to a mechanical issue where the motor shaft is not aligned correctly, and Class 4 (Unbalance) reflects an uneven mass distribution in the rotating system. This domain is chosen due to its relative simplicity, allowing the focus to remain on the integration and validation of the IIoT-based diagnostic architecture, rather than solely on model complexity or adaptation across operating conditions.

# 5. ML Model Selection and Implementation

In this project, Fault Diagnosis (FD) is performed using a Machine Learning (ML) model trained to classify different fault types. The following sections outline the process of model selection, training, and testing.

## 5.1. Selection of Pre-Processing Technique and ML model

A comprehensive state-of-the-art (SOTA) review was carried out to analyze previous studies that utilized the selected dataset. Out of 16 identified references (Anbalagan et al., 2023; Aydin & Seker, 2023; Dehnavi & Shafiee, 2023; Esmaili & Cristaldi, 2024; Hou et al., 2024; Jalonen et al., 2024; Jung et al., 2023; H. Lee et al., 2023; Liu et al., 2023; Ni et al., 2023; Usman et al., 2024; Wang et al., 2023; Wang et al., 2024; Wu et al., 2024; X. Zhang et al., 2024; Zhu et al., 2023), nine were deemed highly relevant to the scope of this project. From these studies, critical elements were identified, including input variables, data preprocessing techniques, diagnosed fault types, feature extraction methods, core algorithms employed, and reported performance metrics. These insights informed the development of a set of evaluation criteria to assess the methodological soundness and applicability of each study in relation to the project's objectives.

### *5.1.1. Criteria Definition*

Table 4 outlines the five general criteria used to guide the selection of the foundational SOTA article for this work. Each criterion includes key subcomponents for evaluating algorithms suited for fault diagnosis in induction motors.

### *5.1.2. Comparison Table*

After defining the evaluation criteria, each selected article was assigned a numerical score from 0 to 5 based on its methodological rigor and relevance to the project's objectives. This scoring process facilitated the identification of the most suitable approach for implementation, specifically within the fault diagnosis component of the proposed cloud-based architecture. A detailed comparative summary of the reviewed studies is presented in Appendix B.

**Table 4.** *Criteria for Algorithm Evaluation*

| Main Criterion | Subcriterion | Description |
|---|---|---|
| **Input Requirements** | **Data Availability** | Evaluates required data types (e.g., current, voltage, vibration, temperature) and acquisition feasibility. |
| | **Data Preprocessing** | Assesses preprocessing needs such as feature extraction, normalization, and denoising; favors minimal preprocessing. |
| **Classification Output** | **Types of Faults Diagnosed** | Determines the algorithm's capability to identify various faults (e.g., rotor, stator, bearing) and severity levels. |
| | **Domain Variations** | Assesses fault detection across different machine states or operating conditions. |
| **Algorithm Characteristics** | **Algorithm Type** | Specifies if the algorithm is data-driven, physics-based, or hybrid. |
| | **Performance Metrics** | Evaluates the accuracy and other relevant performance indicators. |
| **Ease of Use and Implementation** | **Complexity** | Prioritizes simplicity for easier deployment and use. |
| | **Real-time Computing** | Checks feasibility for real-time implementation based on computational efficiency. |
| **Compatibility with IIoT Systems** | **Integration with IIoT Tools** | Evaluates compatibility with industrial IoT platforms, protocols, and tools. |
| | **Scalability** | Assesses performance when handling increased data or system complexity. |

Based on this analysis, the preprocessing strategy and fault diagnosis model proposed by Usman et al. (2024) were adopted for this work.

### 5.1.3. Selection Result

***Preprocessing:*** For this study, the Fast Fourier Transform (FFT) was selected as the primary preprocessing technique because it effectively captures the key characteristics of vibration signals in the frequency domain, which are essential for fault diagnosis in induction motors (Asad et al., 2023; Jiang et al., 2022; W. J. Lee et al., 2021; Zhao et al., 2020). To further enhance frequency resolution and suppress noise, the Hilbert transform was applied beforehand to extract the signal envelope, as recommended by Usman et al. (2024).

***ML Model:*** Recent research has explored various ML techniques for classifying induction motor faults, with Convolutional Neural Networks (CNNs) proving particularly effective (Anbalagan et al., 2023; Chakraborty et al., 2022; W. J. Lee et al., 2021). CNNs efficiently extract key features from FFT-transformed vibration signals, eliminating the need for time-domain feature engineering while accurately classifying critical motor faults.

## 5.2. Model Training and Performance Evaluation

The model was designated as the model for Domain 0 of the KAIST dataset. It was developed using the Keras library in Python and follows the CNN architecture outlined by Usman et al. (2024). The data preprocessing pipeline involved several key steps: window segmentation, z-normalization, and the extraction of frequency-domain features using the FFT and the Hilbert Transform. To ensure a robust evaluation, K-fold cross-validation was utilized, partitioning the dataset into 70% for training, 15% for validation, and 15% for testing. The model achieved a macro-average accuracy of 98.5% on Domain 0 of the KAIST dataset.

Table 5 presents the classification performance of the trained model using k-fold cross-validation. The metrics reported include precision, recall, and F1-score for each class, as well as overall accuracy, macro average, and weighted average. Each value is expressed as the mean ± standard deviation across all folds, demonstrating the model's robustness and consistency across different data partitions.

**Table 5.** *Performance metrics for the model*

| | Precision | Recall | F1-Score |
|---|---|---|---|
| Class 0 | 0.9815 ± 0.0041 | 0.9751 ± 0.0054 | 0.9783 ± 0.0040 |
| Class 1 | 1.0000 ± 0.0000 | 1.0000 ± 0.0000 | 1.0000 ± 0.0000 |
| Class 2 | 1.0000 ± 0.0000 | 1.0000 ± 0.0000 | 1.0000 ± 0.0000 |
| Class 3 | 1.0000 ± 0.0000 | 1.0000 ± 0.0000 | 1.0000 ± 0.0000 |
| Class 4 | 0.9750 ± 0.0062 | 0.9816 ± 0.0046 | 0.9783 ± 0.0047 |
| Accuracy | 0.9913 ± 0.0017 | 0.9913 ± 0.0017 | 0.9913 ± 0.0017 |
| Macro Avg | 0.9913 ± 0.0017 | 0.9913 ± 0.0017 | 0.9913 ± 0.0017 |
| Weighted Avg | 0.9913 ± 0.0017 | 0.9913 ± 0.0017 | 0.9913 ± 0.0017 |

Figure 3 illustrates the final aggregated confusion matrix obtained from the model's predictions on the test sets across all folds. Each cell indicates the percentage of instances of the actual class (rows) predicted as each class (columns). The strong diagonal pattern confirms the model's high classification accuracy, with minimal misclassifications observed, between classes 0 and 4.

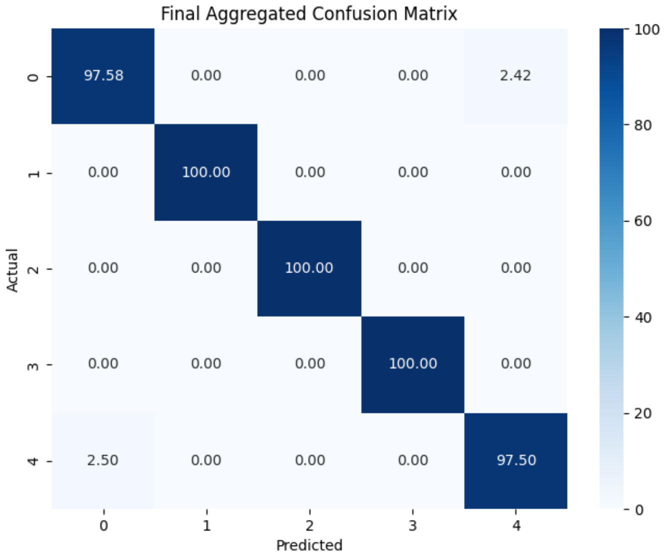


**Figure 3.** *Confusion matrix resulting from 5 folds.*

# 6. IIoT Cloud Based Solution Deployment

The final objective of this project is to implement and validate a cloud-based IIoT solution for the FD of induction motors in an industrial environment. This section describes the development and deployment process of the system architecture, including the necessary tools and services to operate a complete FD pipeline, from data ingestion to model inference.

## 6.1. Industry Standard UNS Architecture Definition

In the realm of industrial data management, the Unified Namespace (UNS) has become a crucial architectural concept for organizing and contextualizing data throughout an enterprise. By providing a standardized naming and structuring convention, the UNS ensures that data is accessible, discoverable, and relevant to users based on their roles and areas of interest.

### *6.1.1. General UNS Hierarchy*

Implementing a UNS architecture for FD applications enables streamlined access to both raw and processed data from various production lines and operational levels. Below is an example of a generalized UNS hierarchy designed for the FD of IM.

```
Enterprise/Site/Area/
|- Line/
|  |- Edge/
|  |  |- AssetModel/
|  |  |  |- sensor1          # Time-series data from sensors
|  |  |  |- ...              # Other sensor variables
|  |- Analysis/
|  |  |- Data/
|  |  |  |- raw/             # Unprocessed sensor data
|  |  |  |- processed/       # Transformed signals
|  |  |- Diagnosis/
|  |  |  |- prediction/      # Fault classification
```

**Figure 4.** *General Hierarchical data structure for UNS model.*

This hierarchical model organizes data sources and analytic layers in a structured and scalable way. At the top level, the hierarchy begins with the Enterprise/Site/Area, representing the organizational and physical layout of the facility. Each Line contains an Edge node, responsible for acquiring and publishing raw sensor data under the AssetModel namespace. The asset models include time-series data from various sensors attached to industrial assets such as induction motors. Parallel to the edge layer, the Analysis section houses processed data and diagnostic results. The Data folder stores both raw and transformed sensor signals, while the Diagnosis folder contains outputs from fault classification models. This hierarchical structure provides clear context regarding the origin of the data, the time it was collected, and the processing methods employed, ensuring traceability and facilitating seamless integration across industrial systems. Moreover, its modular design supports scalability, allowing new assets (e.g., additional motors or lines) or new physical locations (e.g., other areas or sites) to be easily integrated into the hierarchy without compromising data organization or semantic clarity.

### *6.1.2. Specific Architecture*

In this project, the UNS is a crucial framework for organizing and contextualizing data throughout all stages of the fault diagnosis pipeline. Below is a project-specific example of the UNS structure implemented for a wind turbine located in Bucaramanga:

```
Factory UIS/Bucaramanga/Wheat/
|- Mill 1/
|  |- Edge/
|  |  |- MotorModel/
|  |  |  |- vibration         # Edge sensor data
|  |  |  |- temperature
|  |  |  |- current
|  |  |  |- ...
|  |- Analysis/
|  |  |- Vibration/
|  |  |  |- raw_vector/     # Vibration data formatted
|  |  |  |- fft_vector/     # Frequency-domain representation
|  |  |- Diagnosis/
|  |  |  |- prediction/     # CNN Fault prediction
```

**Figure 5.** *Example UNS hierarchy for Mill 1 at the Wheat processing line in UIS Bucaramanga.*

### *6.1.3. MQTT Data Topics*

MQTT is a lightweight publish–subscribe protocol widely used in industry due to its efficiency and scalability. Its topic-based structure supports the Unified Namespace (UNS) model by enabling organized, real-time data exchange between devices and systems, making it key for IIoT implementations (HiveMQ, 2022).

- **Vibration Sensor Data**
  `(Factory UIS/Bucaramanga/Wheat/Mill 1/Edge/MotorModel/Vibration)`
  Raw time-series data acquired directly from the vibration sensors installed on the mill. It serves as the starting point for the fault diagnosis process, capturing essential mechanical behavior for early detection.

- **Preprocessed Vibration Vector**
  `(Factory UIS/Bucaramanga/Wheat/Mill 1/Analysis/Vibration/Raw_Vector)`
  Structured data vector generated from the raw vibration signals after segmentation and normalization. This step prepares the data for frequency analysis and model inference.

- **FFT Feature Vector**
  `(Factory UIS/Bucaramanga/Wheat/Mill 1/Analysis/Vibration/FFT_Vector)`
  Frequency-domain representation of the raw vector, obtained through a Fast Fourier Transform (FFT). This vector captures frequency-related features crucial for detecting anomalies and is used as input to the 1D CNN model.

- **Fault Prediction Result**
  `(Factory UIS/Bucaramanga/Wheat/Mill 1/Analysis/Diagnosis/Prediction)`
  Output from the CNN-based fault diagnosis model, including fault classification results and, optionally, confidence scores or severity indicators.

## 6.2. UNS Based Architecture Tool Selection

### 6.2.1. Base Architecture

The UNS based architecture is implemented through layers. A general scheme is presented in Figure 6

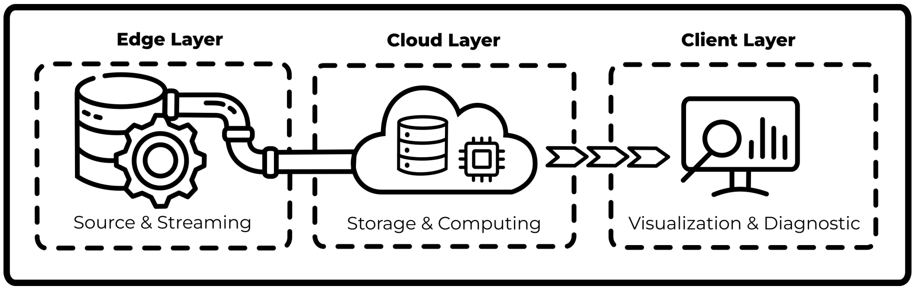


**Figure 6.** *UNS based architecture layers.*

***Edge Layer***: Data is generated and initially processed at the edge layer, near equipment such as induction motors. This layer includes sensors that monitor parameters like vibration, temperature, and electrical current. These sensors send data to edge devices, such as industrial PCs, PLCs, or gateways, which handle tasks like signal conditioning, filtering and protocol conversion (e.g., OPC UA or MQTT). The primary objective is to enable real-time data acquisition while minimizing the volume of raw data transmitted to the cloud, thereby reducing latency and optimizing bandwidth usage.

***Cloud Layer***: The cloud layer serves as the backbone for data management and advanced analytics. It stores incoming edge data in scalable databases that can handle large volumes. This layer supports machine learning and analytics for fault detection and prediction. Typically deployed on cloud or hybrid infrastructures, its primary role is to centralize data processing and enable enterprise-level diagnostics and decision-making.

***Client Layer***: The client layer serves as the interface for users, providing tools for visualization, interaction, and decision-making. It offers dashboards, real-time alerts, and diagnostic reports based on cloud analytics. Users, including operators and engineers, access these interfaces through SCADA systems, HMIs, or mobile/web applications to monitor equipment health and make informed maintenance decisions.

### *6.2.2. Tool Selection*

After research of the highlighted tools in both industry and research worlds, a stack (set of tools) was designed to serve as part of the solution. Figure 7 illustrates the distribution of selected tools across the edge, cloud, and client layers. Table 6 summarizes these tools and their general functions within each architectural layer.

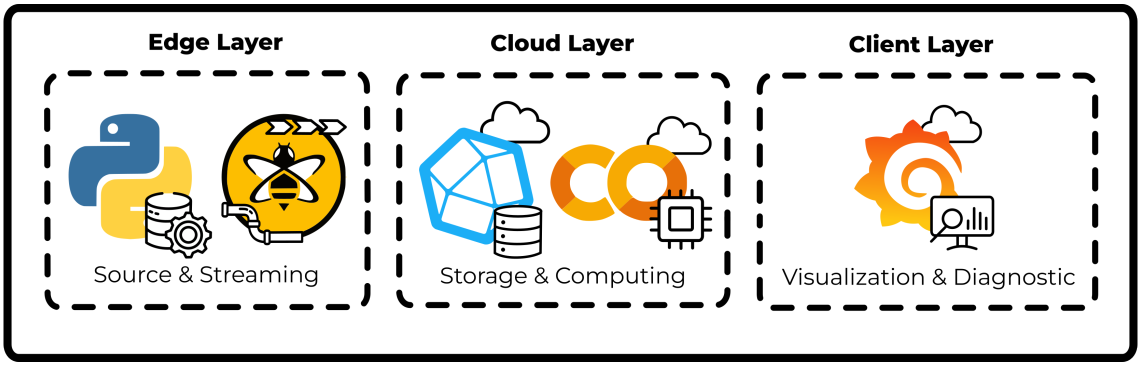


**Figure 7.** *Architecture tools by layer.*

**Table 6.** *Summary of tools and their general functions by architectural layer*

| Layer | Component | General Function |
|---|---|---|
| Edge | Python | Simulates edge devices and performs local data processing and publishing to communication protocols. |
| | HiveMQ Broker | Manages message exchange between devices using MQTT for efficient and scalable communication. |
| Cloud | InfluxDB | Stores and manages high-frequency time-series data for further analysis. |
| | Google Colab | Provides a scalable and accessible environment for executing cloud-based analytics and machine learning tasks. |
| Client | Grafana | Visualizes time-series data through real-time dashboards for monitoring and decision support. |

## 6.3. Implementation and Model Integration

The proposed solution follows a three-layer architecture: Edge, Cloud, and Client, designed to simulate, process, and visualize induction motor data for fault diagnosis in an IIoT environment as shown in Figure 8.

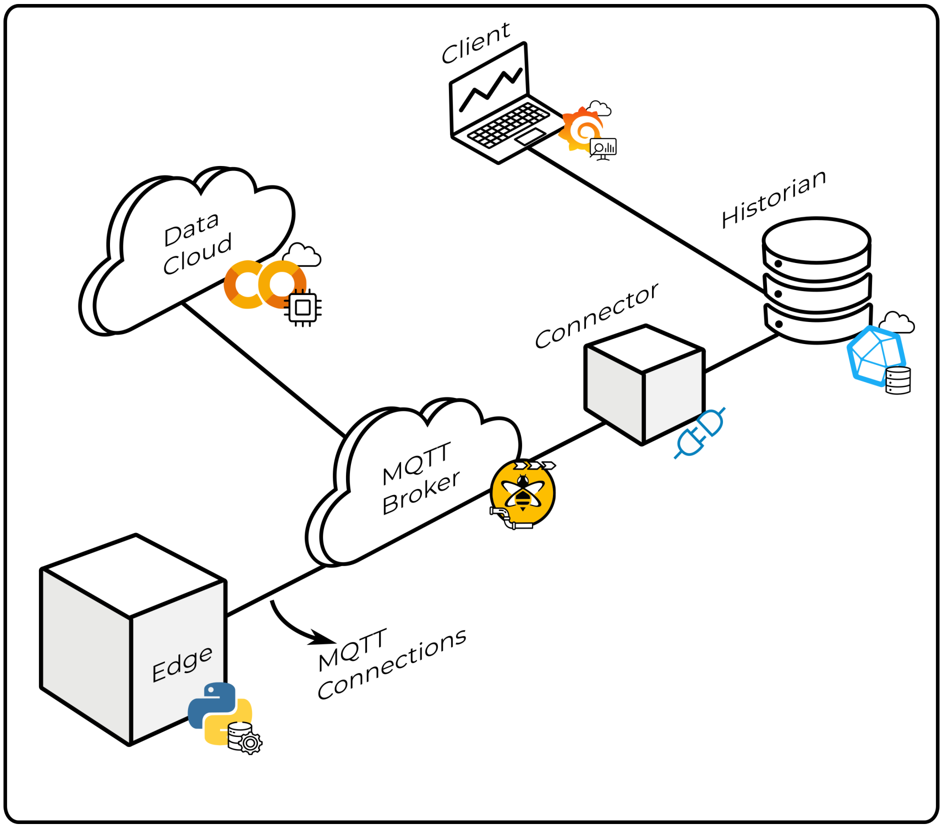


**Figure 8.** *Final Architecture Schema*

### *Solution Workflow*

**Edge Layer:** The system begins with a Python-based simulation that generates sensor data streams representing various induction motor behaviors, including both normal and fault conditions. This approach enables realistic testing without the need for physical hardware (Azad et al., 2025; Deghanimohammadabadi et al., 2021). The data is transmitted through HiveMQ, a scalable MQTT broker that ensures high-throughput and low-latency communication. HiveMQ operates as a Unified Namespace (UNS), enabling structured, hierarchical organization of data streams and decoupled communication between devices and systems (HiveMQ, 2022; H. Lee et al., 2019).

**Cloud Layer:** The MQTT data is collected by the Telegraf plugin and written to InfluxDB, a time-series database optimized for high-frequency telemetry data such as vibration signals (Tripathi et al., 2023; Zhu et al., 2023). For computational tasks, Google Colab is employed as a cloud-based development environment. Its support for GPU/TPU resources allows for scalable deployment and testing of machine learning models aimed at fault detection (Ali et al., 2020). Since it operates over the internet, Colab accesses the HiveMQ broker remotely, ensuring seamless integration into the data pipeline.

**Client Layer:** Visualization and monitoring are handled by Grafana, an open-source platform well-suited for time-series analytics. It integrates directly with InfluxDB to display real-time dashboards, performance trends, and machine learning diagnostic results (Hindi et al., 2024). This interface enables engineers to explore data intuitively and respond to motor conditions through actionable insights and alerts.

# 7. Results

The implementation was deployed within a Docker container, incorporating a technology stack that includes InfluxDB, Telegraf, and Grafana. Each component was configured with dedicated endpoints to ensure seamless interaction and access to their respective functionalities. A real-time simulation environment was established using a local Python script that generates and publishes batches of vibration data to the HiveMQ broker. These batches consist of segmented time-series data stored in separate CSV files, each representing a specific fault class from the selected dataset. This setup enables end-to-end connectivity across the system. As shown in Figure 9, raw simulated vibration signals are captured and processed through an FFT transformation, and the resulting diagnostic predictions, generated by a machine learning model deployed in the Google Colab, are visualized.

| | Last | Min | Max |
|---|---|---|---|
| sensor1 {host="7a21871dc89f", topic="Edge/Process/vibration"} | 0.0770 g | -1.4 g | 1.96 g |
| sensor2 {host="7a21871dc89f", topic="Edge/Process/vibration"} | -0.066 g | -2.8 g | 2.31 g |
| sensor3 {host="7a21871dc89f", topic="Edge/Process/vibration"} | 0.0751 g | -0.48 g | 0.444 g |
| sensor4 {host="7a21871dc89f", topic="Edge/Process/vibration"} | 0.00325 g | -0.25 g | 0.218 g |

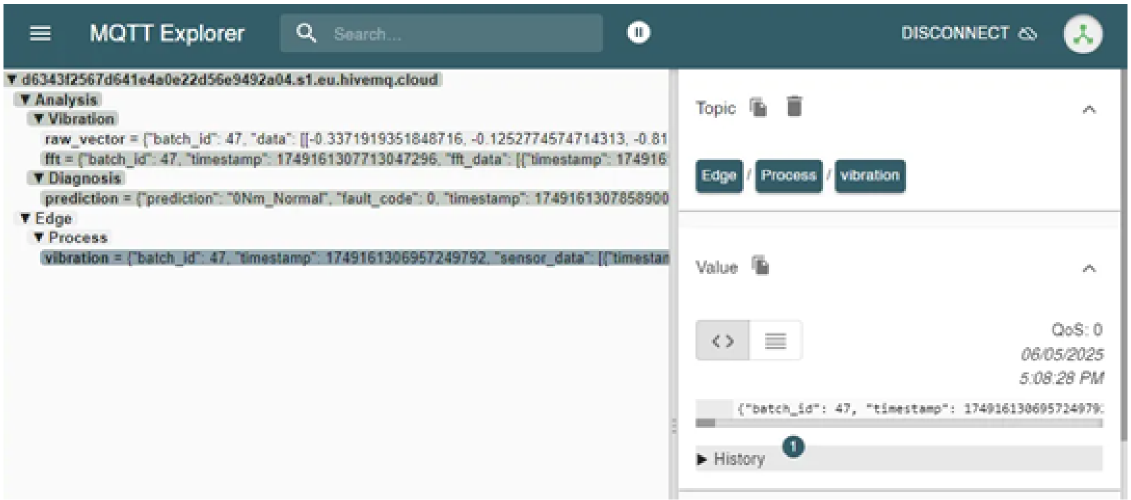
| | Min | Max | Last |
|---|---|---|---|
| sensor1 {host="7a21871dc89f", topic="Analysis/Vibration/fft"} | 0.504 g | 20.2 g | 20.2 g |
| sensor2 {host="7a21871dc89f", topic="Analysis/Vibration/fft"} | 0.00203 g | 87.9 g | 87.9 g |
| sensor3 {host="7a21871dc89f", topic="Analysis/Vibration/fft"} | 0.000934 g | 42.3 g | 42.3 g |
| sensor4 {host="7a21871dc89f", topic="Analysis/Vibration/fft"} | 0.370 g | 58.0 g | 58.0 g |

fault_code {host="7a21871dc89f", topic="Analysis/Diagnosis/prediction"}

**Figure 9.** *Cloud Based Solution Dashboard*

### 7.1. Layers

***Edge Layer:*** A python script publishes the messages while the broker was monitored using MQTT Explorer. To simplify topic representation, the main path (Factory UIS/Bucaramanga/Wheat/Mill1) was omitted while preserving the internal topic structure for consistency and readability.



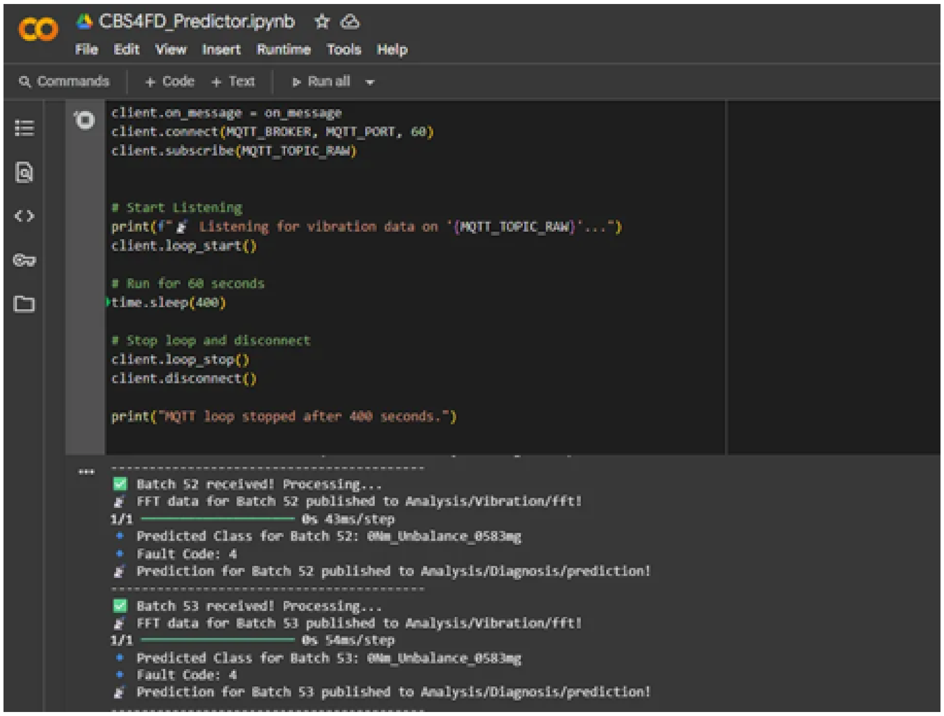
```
▼ d6343f2567d641e4a0e22d56e9492a04.s1.eu.hivemq.cloud
  ▼ Analysis
    ▼ Vibration
      raw_vector = {"batch_id": 47, "data": [[-0.3371919351848716, -0.1252774574714313, -0.81
      fft = {"batch_id": 47, "timestamp": 1749161307713047296, "fft_data": [{"timestamp": 174916
    ▼ Diagnosis
      prediction = {"prediction": "0Nm_Normal", "fault_code": 0, "timestamp": 1749161307858900
  ▼ Edge
    ▼ Process
      vibration = {"batch_id": 47, "timestamp": 1749161306957249792, "sensor_data": [{"timestan
```

```
QoS: 0
06/05/2025
5:08:28 PM
{"batch_id": 47, "timestamp": 174916130695724979
```

**Figure 10.** *Broker Topic Messages*

***Data Layer:*** InfluxDB was configured with the Telegraf plugin to capture and store data transmitted by the broker. Telegraf was responsible for parsing messages that conformed to the predefined format, extracting timestamps and vibration signal values for time-series storage. A Google Colab notebook was developed to receive the raw signal vector and return the corresponding diagnostic result. This process utilized a pre-trained CNN model, which was previously uploaded and integrated into the workflow. The resulting diagnosis was then published back to the broker and subsequently stored in InfluxDB for further analysis and visualization.

**Figure 11.** *Vibration data in InfluxDB UI*



```
client.on_message = on_message
client.connect(MQTT_BROKER, MQTT_PORT, 60)
client.subscribe(MQTT_TOPIC_RAW)


# Start Listening
print(f"📡 Listening for vibration data on '{MQTT_TOPIC_RAW}'...")
client.loop_start()

# Run for 60 seconds
time.sleep(400)

# Stop loop and disconnect
client.loop_stop()
client.disconnect()

print("MQTT loop stopped after 400 seconds.")
```

```
----------------------------------------
✅ Batch 52 received! Processing...
📡 FFT data for Batch 52 published to Analysis/Vibration/fft!
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 43ms/step
🔹 Predicted Class for Batch 52: 0Nm_Unbalance_0583mg
🔹 Fault Code: 4
📡 Prediction for Batch 52 published to Analysis/Diagnosis/prediction!
----------------------------------------
✅ Batch 53 received! Processing...
📡 FFT data for Batch 53 published to Analysis/Vibration/fft!
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 54ms/step
🔹 Predicted Class for Batch 53: 0Nm_Unbalance_0583mg
🔹 Fault Code: 4
📡 Prediction for Batch 53 published to Analysis/Diagnosis/prediction!
----------------------------------------
```

**Figure 12.** *Google Colab Notebook*

***Client Layer:*** At the client layer, a Grafana dashboard was designed and connected directly to InfluxDB, enabling real-time visualization of key variables such as raw vibration data, the signal’s FFT, and the diagnostic results generated by the CNN model.

**Figure 13.** *Grafana Dashboard*

## 8. Conclusions

This project presented the design and development of a cloud-based architecture for the fault diagnosis of induction motors, grounded in the foundational principles of Industry 4.0. The proposed solution offers a systematic framework for the deployment of machine learning algorithms within industrial environments, ensuring technical scalability, reliability, and integration across heterogeneous systems. By incorporating key technologies, such as edge computing, time-series data historians, cloud services, and advanced visualization platforms, the architecture enables continuous monitoring of motor conditions and real-time detection of anomalies.

At the core of the design is the implementation of the Unified Namespace, which standardizes and contextualizes industrial data, providing a consistent and interoperable structure for information exchange across the enterprise. This facilitates a seamless flow of data from acquisition to insight, supporting predictive maintenance strategies and improving operational responsiveness.

This project demonstrated that a cloud-based architecture designed for the fault diagnosis of induction motors, aligned with Industry 4.0 principles, can serve as a blueprint for implementing intelligent systems in industrial settings. By combining technologies such as edge computing, time series data historians, cloud services, and advanced visualization tools, the solution enables real-time condition monitoring and early fault detection. Central to this design is the Unified Namespace, which organizes and contextualizes data across the enterprise, ensuring interoperability and consistent information flow. Beyond the technical benefits, this type of solution has a significant industrial impact, enhancing operational visibility, reducing unplanned downtime, and improving the effectiveness of maintenance actions. As a result, industries can move toward more predictive, data-driven operations that increase productivity, reduce costs, and ultimately drive digital transformation at scale.

# Bibliography

Abualsauod, E. H. (2023). Machine learning based fault detection approach to enhance quality control in smart manufacturing. *Production Planning and Control*. https://doi.org/10.1080/09537287.2023.2175736

Ali, I., Khan, A., & Waleed, M. (2020). A google colab based online platform for rapid estimation of real blur in single-image blind deblurring. *Proceedings of the 12th International Conference on Electronics, Computers and Artificial Intelligence*, 1–6. https://doi.org/10.1109/ECAI50035.2020.9223244

Anbalagan, S., Agarwal, D., Natarajan, B., & Srinivasan, B. (2023). Foundational models for fault diagnosis of electrical motors. *2023 IEEE International Conference on Power Electronics, Smart Grid, and Renewable Energy*. https://doi.org/10.1109/PESGRE58662.2023.10404206

Angarita-Zapata, J. S., Maestre-Gongora, G., & Calderín, J. F. (2021). A bibliometric analysis and benchmark of machine learning and automl in crash severity prediction: The case study of three colombian cities. *Sensors*, *21*(24), 8401. https://doi.org/10.3390/s21248401

Aqueveque, P., Radrigan, L., Pastene, F., Morales, A., & Guerra, E. (2021). Data-driven condition monitoring of mining mobile machinery in non-stationary operations using wireless accelerometer sensor modules. *IEEE Access*, *9*, 17365–17381. https://doi.org/10.1109/ACCESS.2021.3051583

Asad, B., Raja, H. A., Vaimann, T., Kallaste, A., Pomarnacki, R., & Hyunh, V. (2023). A current spectrum-based algorithm for fault detection of electrical machines using low-power data acquisition devices. *Electronics*, *12*(7), 1746. https://doi.org/10.3390/electronics12071746

Aydin, I., & Seker, S. (2023). Trend analysis and continuous wavelet transform for bearing fault detection in induction motors. *14th International Conference on Electrical and Electronics Engineering*. https://doi.org/10.1109/ELECO60389.2023.10416018

Azad, S. M., Suresh, P., & Sireesha, R. (2025). Machine learning approaches for process automation: A case study using q-learning and supervised learning. *2025 4th International Conference on Power, Control and Computing Technologies*, 564–567. https://doi.org/10.1109/ICPC2T63847.2025.10958721

Chakraborty, S., Javed, F. M., Shamrat, M., Ahammad, R., Billah, M. M., Kabir, M., & Hosen, R. (2022). Bearing fault detection based on internet of things using convolutional neural network. *International Journal of Advanced Computer Science and Applications*, *13*(4).

Chang, Y., Chen, J., Wu, W., Pan, T., Zhou, Z., & He, S. (2022). Intelligent fault quantitative identification for industrial internet of things (IIoT) via a novel deep dual reinforcement learning model accompanied with insufficient samples. *IEEE Internet of Things Journal*, *9*(20), 19811–19822. https://doi.org/10.1109/JIOT.2022.3168317

Cunha, R. G. C., da Silva, E. T., & de Sá Medeiros, C. M. (2021). Machine learning and multiresolution decomposition for embedded applications to detect short-circuit in induction motors. *Computers in Industry*, *129*, 103461. https://doi.org/10.1016/j.compind.2021.103461

Dehghanimohammadabadi, M., Belsare, S., & Thiesing, R. (2021). Simulation-optimization of digital twin. *Proceedings - Winter Simulation Conference*. https://doi.org/10.1109/WSC52266.2021.9715412

Dehnavi, V. S., & Shafiee, M. (2023). Inner and outer bearing fault diagnosis of electrical motors using a proposed algorithm and vibration signals. *14th International Conference on Information and Knowledge Technology*, 175–180. https://doi.org/10.1109/IKT62039.2023.10433018

Esmaili, P., & Cristaldi, L. (2024). Health indicator effectiveness in localized fault diagnosis: Rolling bearing elements. *2024 IEEE International Workshop on Metrology for Industry 4.0 and IoT*, 552–556. https://doi.org/10.1109/METROIND4.0IOT61288.2024.10584223

Fathy, Y., Jaber, M., & Brintrup, A. (2021). Learning with imbalanced data in smart manufacturing: A comparative analysis. *IEEE Access*, *9*, 2734–2757. https://doi.org/10.1109/ACCESS.2020.3047838

Fu, P., Wang, J., Zhang, X., Zhang, L., & Gao, R. X. (2020). Dynamic routing-based multimodal neural network for multi-sensory fault diagnosis of induction motor. *Journal of Manufacturing Systems*, *55*, 264–272. https://doi.org/10.1016/j.jmsy.2020.04.009

Garrido-Labrador, J. L., Puente-Gabarri, D., Ramírez-Sanz, J. M., Ayala-Dulanto, D., & Maudes, J. (2020). Using ensembles for accurate modelling of manufacturing processes in an IoT data-acquisition solution. *Applied Sciences*, *10*(13), 4606. https://doi.org/10.3390/app10134606

Hindi, I., Alyaman, M., AboZenah, A., Zaid, A., & Shrara, M. (2024). Smart alarm IoT system: Monitoring elevator traffic and meteorological data on job sites using mqtt and influxdb integrated with grafana. *2024 15th International Conference on Information and Communication Systems*. https://doi.org/10.1109/ICICS63486.2024.10638309

HiveMQ. (2022). *What is a unified namespace (UNS) and why does it matter?* Retrieved 2022, from https://www.hivemq.com/blog/what-is-unified-namespace-uns-iiot-industry-40/

Hong, J.-K. (2022). Vibration prediction of flying IoT based on LSTM and GRU. *Electronics*, *11*(7), 1052. https://doi.org/10.3390/electronics11071052

Hou, W., Xiao, G., Liu, X., Jiang, L., & Jing, W. (2024). 1d-dctn: 1-d deformable convolutional transformer network for multi-signal fault diagnosis. *Chinese Control Conference*, 5050–5057. https://doi.org/10.23919/CCC63176.2024.10661826

Huang, B., Dou, H., Luo, Y., Li, J., Wang, J., & Zhou, T. (2023). Adaptive spatiotemporal transformer graph network for traffic flow forecasting by IoT loop detectors. *IEEE Internet of Things Journal*, *10*(2), 1642–1653. https://doi.org/10.1109/JIOT.2022.3209523

Irgat, E., Çinar, E., Ünsal, A., & Yazıcı, A. (2022). An IoT-based monitoring system for induction motor faults utilizing deep learning models. *Journal of Vibration Engineering and Technologies*, *1*, 1–11. https://doi.org/10.1007/s42417-022-00769-5

Jaen-Cuellar, A. Y., Elvira-Ortiz, D. A., Saucedo-Dorantes, J. J., & del Rio, S. J. (2023). Statistical machine learning strategy and data fusion for detecting incipient ITSC faults in IM. *Machines*, *11*(7), 720. https://doi.org/10.3390/machines11070720

Jalonen, T., Al-Sa'd, M., Kiranyaz, S., & Gabbouj, M. (2024). Real-time vibration-based bearing fault diagnosis under time-varying speed conditions. *Proceedings of the IEEE International Conference on Industrial Technology*. https://doi.org/10.1109/ICIT58233.2024.10540813

Jarwar, M. A., Khowaja, S. A., Dev, K., Adhikari, M., & Hakak, S. (2023). Neat: A resilient deep representational learning for fault detection using acoustic signals in IIoT environment. *IEEE Internet of Things Journal*, *10*(4), 2864–2871. https://doi.org/10.1109/JIOT.2021.3109668

Jiang, J. H., Liu, Y. Q., Xu, C. Z., & Shen, H. T. (2022). Research on motor bearing fault diagnosis based on the adaboost algorithm and the ensemble learning with bayesian optimization in the industrial internet of things. *Security and Communication Networks*, *2022*, 4569954. https://doi.org/10.1155/2022/4569954

Jovicic, E., Primorac, D., Cupic, M., & Jovic, A. (2023). Publicly available datasets for predictive maintenance in the energy sector: A review. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3295113

Jung, W., Kim, S. H., Yun, S. H., Bae, J., & Park, Y. H. (2023). Vibration, acoustic, temperature, and motor current dataset of rotating machine under varying operating conditions for fault diagnosis. *Data in Brief*, *48*, 109049. https://doi.org/10.1016/j.dib.2023.109049

Kizito, R., Scruggs, P., Li, X. P., Devinney, M., Jansen, J., & Kress, R. (2021). Long short-term memory networks for facility infrastructure failure and remaining useful life prediction. *IEEE Access*, *9*, 67585–67594. https://doi.org/10.1109/ACCESS.2021.3077192

Kovilpillai, J. J. A., & Jayanthy, S. (2022). A transfer learning approach for effective motor fault identification of industrial machines used in tile manufacturing. *National Academy Science Letters*, *45*(5), 405–409. https://doi.org/10.1007/s40009-022-01145-3

Lee, H., Lee, Y., Jo, M., Nam, S., Jo, J., & Lee, C. (2023). Enhancing diagnosis of rotating elements in roll-to-roll manufacturing systems through feature selection approach considering overlapping data density and distance analysis. *Sensors*, *23*(18), 7857. https://doi.org/10.3390/S23187857

Lee, H., Lim, J., & Kwon, T. ". (2019). MQTLS: toward secure MQTT communication with an untrusted broker. *ICTC 2019 - 10th International Conference on ICT Convergence*, 53–58. https://doi.org/10.1109/ICTC46691.2019.8940001

Lee, W. J., Xia, K., Denton, N. L., Ribeiro, B., & Sutherland, J. W. (2021). Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery. *Journal of Intelligent Manufacturing*, *32*(2), 393–406. https://doi.org/10.1007/s10845-020-01578-x

Li, Y., Song, Y., Jia, L., Gao, S., Li, Q., & Qiu, M. (2021). Intelligent fault diagnosis by fusing domain adversarial training and maximum mean discrepancy via ensemble learning. *IEEE Transactions on Industrial Informatics*, *17*(4), 2833–2841. https://doi.org/10.1109/TII.2020.3008010

Ling, Y., Fu, D., Jiang, P., Sun, Y., Yuan, C., Huang, D., Lu, J., & Lu, S. (2023). Lightweight and intelligent model based on enhanced sparse filtering for rotating machine fault diagnosis. *Transactions of the Institute of Measurement and Control*. https://doi.org/10.1177/01423312231185702

Liu, H., Liu, Z., Luo, Z., Chen, J., Liu, H., & Zhang, Y. (2023). Contrastive learning based fault diagnosis of motor with multi-modal feature fusion. *Proceedings - 2023 China Automation Congress*, 2486–2491. https://doi.org/10.1109/CAC59555.2023.10451894

Milmo, D. (2008). *Passengers fume in the chaos of terminal 5's first day*. Retrieved 2008, from https://www.theguardian.com/environment/2008/mar/28/travelandtransport.theairlineindustry

Mokkamayan, M., & Thankappan, S. P. (2023). An optimized deep learning approach for predicting the electric motor temperature using IoT sensors. *Electric Power Components and Systems*. https://doi.org/10.1080/15325008.2023.2227194

Ni, Q., Ji, J. C., Halkon, B., Feng, K., & Nandi, A. K. (2023). Physics-informed residual network (piresnet) for rolling element bearing fault diagnostics. *Mechanical Systems and Signal Processing*, *200*, 110544. https://doi.org/10.1016/j.ymssp.2023.110544

Rajput, D. S., Meena, G., Acharya, M., & Mohbey, K. K. (2023). Fault prediction using fuzzy convolution neural network on iot environment with heterogeneous sensing data fusion. *Measurement: Sensors*, *26*, 100701. https://doi.org/10.1016/j.measen.2023.100701

Ribeiro, F. (2020). *Mafaulda*. Retrieved 2020, from https://www02.smt.ufrj.br/~offshore/mfs/page_01.html

Rosati, R., Romeo, L., Cecchini, G., Tonetto, F., Viti, P., Mancini, A., & Frontoni, E. (2023). From knowledge-based to big data analytic model: A novel IoT and machine learning based decision support system for predictive maintenance in industry 4.0. *Journal of Intelligent Manufacturing*, *34*(1), 107–121. https://doi.org/10.1007/s10845-022-01960-x

Rubio, E. M., Dionísio, R. P., & Torres, P. M. B. (2018). Predictive maintenance of induction motors in the context of industry 4.0. *International Journal of Mechatronics and Applied Mechanics*, *2018*(4), 238–242.

Sharma, P., & Liu, H. (2021). A machine-learning-based data-centric misbehavior detection model for internet of vehicles. *IEEE Internet of Things Journal*, *8*(6), 4991–4999. https://doi.org/10.1109/JIOT.2020.3035035

Song, Y., Li, Y., Jia, L., & Qiu, M. (2020). Retraining strategy-based domain adaption network for intelligent fault diagnosis. *IEEE Transactions on Industrial Informatics*, *16*(9), 6163–6171. https://doi.org/10.1109/TII.2019.2950667

Souza, M. L. H., da Costa, C. A., de Oliveira Ramos, G., & da Rosa Righi, R. (2021). A feature identification method to explain anomalies in condition monitoring. *Computers in Industry*, *133*, 103528. https://doi.org/10.1016/j.compind.2021.103528

Tran, M.-Q., Amer, M., Dababat, A., Abdelaziz, A. Y., Dai, H. J., Liu, M. K., & Elsisi, M. (2023). Robust fault recognition and correction scheme for induction motors using an effective IoT with deep learning approach. *Measurement*, *207*, 112398. https://doi.org/10.1016/j.measurement.2022.112398

Tran, M.-Q., Elsisi, M., Mahmoud, K., Liu, M.-K., Lehtonen, M., & Darwish, M. M. F. (2021). Experimental setup for online fault diagnosis of induction machines via promising IoT and machine learning: Towards industry 4.0 empowerment. *IEEE Access*, *9*, 115429–115441. https://doi.org/10.1109/ACCESS.2021.3105297

Treml, A., Andrade, R., Suetake, M., & Afonso, N. (2020). Experimental database for detecting and diagnosing rotor broken bar in a three-phase induction motor. | ieee dataport. https://ieee-dataport.org/open-access/experimental-database-detecting-and-diagnosing-rotor-broken-bar-three-phase-induction

Tripathi, P., Miraz, M. H., & Joshi, S. (2023). Comparative analysis of MongoDB and InfluxDB for time series data management in iot environments. *Proceedings - 2023 International Conference on Computing, Networking, Telecommunications and Engineering Sciences Applications*, 39–42. https://doi.org/10.1109/CONTESA61248.2023.10384962

Usman, M., Komatsu, T., Htun, K. M., Liu, Z., & Beck, A. (2024). Benchmarking sensor modalities with few-shot domain adaptation for cross-domain fault diagnosis. *IEEE International Conference on Automation Science and Engineering*, 3219–3224. https://doi.org/10.1109/CASE59546.2024.10711623

Vermesan, O., Coppola, M., Bahr, R., Bellmann, R. O., Martinsen, J. E., Kristoffersen, A., Hjertaker, T., Breiland, J., Andersen, K., Sand, H. E., & Lindberg, D. (2022). An intelligent real-time edge processing maintenance system for industrial manufacturing, control, and diagnostic. *Frontiers in Chemical Engineering*, *4*, 900096. https://doi.org/10.3389/fceng.2022.900096

Wang, X., Lv, Y., & Liu, Y. (2023). Intelligent fault diagnosis based on vibration and acoustic-monitored data fusion for rolling bearings. *2023 IEEE International Conference on Industrial Engineering and Engineering Management*, 1356–1360. https://doi.org/10.1109/IEEM58616.2023.10406566

Wang, X., Liu, Z., Dai, M., Li, W., & Tang, J. (2024). Time-shift denoising combined with dwt-enhanced condition domain adaptation for motor bearing fault diagnosis via current signals. *IEEE Sensors Journal*. https://doi.org/10.1109/JSEN.2024.3455099

Wu, H., Lv, Y., Yuan, R., Li, B., Yang, X., & Zhu, W. (2024). Self-matching chirplet extraction transform: A novel tool for dense multicomponent signals analysis and machinery fault diagnosis. *IEEE Sensors Journal*, *24*(7), 10904–10919. https://doi.org/10.1109/JSEN.2024.3365887

Zayed, S. M., Attiya, G., El-Sayed, A., Sayed, A., & Hemdan, E. E. D. (2023). An efficient fault diagnosis framework for digital twins using optimized machine learning models in smart industrial control systems. *International Journal of Computational Intelligence Systems*, *16*(1), 1–18. https://doi.org/10.1007/s44196-023-00241-6

Zhang, S., Ye, F., Wang, B., & Habetler, T. (2021). Semi-supervised bearing fault diagnosis and classification using variational autoencoder-based deep generative models. *IEEE Sensors Journal*, *21*(5), 6476–6486. https://doi.org/10.1109/JSEN.2020.3040696

Zhang, X. R., Rane, K. P., Kakaravada, I., & Shabaz, M. (2021). Research on vibration monitoring and fault diagnosis of rotating machinery based on internet of things technology. *Nonlinear Engineering*, *10*(1), 245–254. https://doi.org/10.1515/nleng-2021-0019

Zhang, X., Li, C., Han, C., Li, S., Feng, Y., Wang, H., Cui, Z., & Gryllias, K. (2024). A personalized federated meta-learning method for intelligent and privacy-preserving fault diagnosis. *Advanced Engineering Informatics*, *62*, 102781. https://doi.org/10.1016/j.aei.2024.102781

Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R., & Chen, X. (2020). Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. *ISA Transactions*, *107*, 224–255. https://doi.org/10.1016/j.isatra.2020.08.010

Zhu, X., Nie, X., & Liu, J. (2023). Time series database optimization based on influxdb. *Proceedings - 2023 International Conference on Power, Electrical Engineering, Electronics and Control*, 879–885. https://doi.org/10.1109/PEEEC60561.2023.00172

# Appendixes

## Appendix A. Sensor data and data shape

| | Acoustic | Current & Temp | Vibration |
|---|---|---|---|
| Samplig Frequency | 51.2kHz | 25.6kHz | 25.6kHz |
| Load | 0Nm | 0, 2, 4 Nm | 0, 2, 4 Nm |
| Operating Conditions | Normal<br>BPFI<br>BPFO | Normal<br>BPFI<br>BPFO<br>Misalign<br>Unbalance | Normal<br>BPFI<br>BPFO<br>Misalign<br>Unbalance |
| Severity | 0.3mm<br>1mm | All* | All* |
| Measures | 1 Acoustic | 2 Temp, 3 Current** | 4: X & Y from housings A & B |
| Format | .mat | .tdms | .mat |
| # Files | 5 | 47 | 45 |

*All severity variations are listed here

- **BPFI & BPFO:** Fault severity is classified based on crack width, measured at 0.3 mm, 1 mm, and 3 mm.
- **Misalignment:** Shaft misalignment severity, categorized by displacement distances of 0.1 mm, 0.3 mm, and 0.5 mm.
- **Unbalance:** Unbalance severity defined by the added mass on the rotor disk, with values of 583 mg, 1169 mg, 1751 mg, and 3318 mg.

**Some files had missing measurings (null)

# Appendix B. Article Methodology Comparison Table

| | 20% | | 20% | | 30% | | 10% | | 20% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Input Requirements** | | **Classification Output** | | **Algorithm Characteristics** | | **Ease of Use and Implementation** | | **Compatibility with IIoT Systems** | | |
| | Data availability | Data PP | Types of faults | Variations | Description | Performance | Complexity | Real-Time computing | Integration | Scalability | **Final Score** |
| PIResNet (Ni et al., 2023) | Vibration | Normalization and resampling. Modal-property-dominant-generated layer | Inner race<br>Outer race<br>Ball | - | 1 Dim - CNN and Residual Network. | Accuracy: 97.79% | Confined to the proposed physic feature extraction block. | 32040 parameters plus logarithm and cepstrum filters | - | Complex feature extraction | |
| **Score** | 3.5 | | 3 | | 5 | | 2 | | 3 | | 3.6 |
| (P. Zhu et al., 2024) | Vibration | Window sampling. Time-frequency feature extraction | Inner race<br>Outer race<br>Misalignment | Different loads 0Nm, 2Nm, 4Nm | Filtering, CNN, Domain adaptation | More than 99.5% accuracy for domain transfer tasks | Complex loss function to enhance interpretability | - | Good noise robustness but not much for KAIST | Time-frequency features are easily applied to different datasets | |
| **Score** | 3.5 | | 4 | | 5 | | 3 | | 4 | | 4.1 |
| (Usman et al., 2024) | 3 Current sensors and 4 vibration 25.6kHz | Segmentation, Z-score, FFT, Hilbert transform | Inner race<br>Shaft misalignment<br>Unbalanced rotor | Load conditions | CNN based network for domain adaptation | Same domain Accuracy: 97%<br>Cross domain 52% | Simple data preprocessing | Few convolutional kernels | Domain adaptation | Considered for different dataset | |
| **Score** | 4.5 | | 4.5 | | 4.5 | | 4 | | 4 | | 4.35 |
| (Hou et al., 2024) | Vibration, current | Continuous Wavelet Transform | Inner race<br>Outer race<br>Unbalance<br>Misalignment | - | Embedding layer, 1-D Deformable , with a multi-head attention layer | PP: 99.83% | Confined to the multi-head attention layer | Model's training time is lengthy | Simplification is required to achieve real-time diagnosis | Not as effective when applied to practical problems | |
| **Score** | 4.5 | | 4 | | 5 | | 2 | | 2 | | 3.8 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (Jalonen et al., 2024) | 4 vibration signals | Down sampling using finite impulse response anti-aliasing lowpass filter | Outer race<br>Inner race<br>Ball | - | Lightweight CNN model | Quantified by accuracy, precision, recall and F1-score<br>98.4% | Very simple implementation | Runs in real-time with processing durations five time less than adquisition | Robust to noise with high performance across various SNR | Confined to fixed noise levels and to diagnosisn isolated faults | |
| **Score** | 4 | | 3 | | 4 | | 4.5 | | 4.5 | | 3.95 |
| (Liu et al., 2023) | Vibration, Temperature, 3phase current, 25.6kHz, 0Nm load | Data PP module with Kalman filtering, Feature self-fusion module | 15 classes of motor states Inner, outer, misalignment, rotor imbalance | Fault severity for every fault | Classification module based on multi-head attention network | Contrastive learning: 0.93 | Three different components with autoencoders, and attention-based | - | Complex for online applications | Multi-head attention module becomes complex with more sensors | |
| **Score** | 5 | | 4.5 | | 4 | | 2 | | 2 | | 3.7 |
| (Dehnavi & Shafiee, 2023) | Vibration | Notch filter, Wavelet transformation, Window selection | Inner and outer race faults | - | General algorithm | Mostly 99% with best performance using static window size | Very simple and complete set of feature extraction | - | Statistical features need less calculation, good for online computing | - | |
| **Score** | 3.5 | | 2 | | 4 | | 4.5 | | 3 | | 3.35 |
| (Wang et al., 2023) | Vibration and acoustic | Resampling, Wavelet Packet transform, energy vectors | Inner and outer race faults | - | KNN, SVM, DT | KNN multimodal feature fusion Accuracy 99.72% | Very simple algorithm and feature fusion method | - | - | KNN complexity grows as classes increase | |
| **Score** | 4 | | 2 | | 3.5 | | 3.5 | | 2 | | 3 |
| (Anbalagan et al., 2023) | Vibration | - | Bearing, misalignment, unbalance at constant speed | Bearing faults at varying speeds | CNN based backbone model, and fine tuning | 95% accuracy with only 15% of data used to train the backbone | Basic CNN architecture | Lot of filtering, performance need to be measured | Perfect for industrial applications | Methodology replicable | |
| **Score** | 3 | | 4.5 | | 4.5 | | 4 | | 5 | | 4.25 |