

Predictive Fault Detection Using Machine Learning And Image Feature Extraction

Zahraa Mohammed Taha Issa , Hussein Najm Abd Ali

Abstract

Predictive fault detection is a key element of contemporary industrial maintenance that intends to detect the equipment failures ahead of severe operation breakdowns. The traditional maintenance methodologies, for example, reactive or preventive maintenance , cannot detect anomalies at the beginning stage, which results in more lead time and cost. With the evolution of the machine learning algorithms predictive maintenance is more efficient especially with the usage of image-based data. Extraction of image features enables capturing visual signals such as erosion, fractures, color fading and thermal deviations which are usually signs of incubating faults .Machine learning techniques such as Convolutional Neural Networks (CNNs), own, Helping Students (SVMs), Ensemble Classifiers, Naïve Bayes and Markov Models, among others, have demonstrated that can analyze these image features automatically for an accurate classification of equipment state. Noise Reduction, Normalization and Segmentation Preprocessing methods need to be employed to enhance feature quality and model performance .[38] As deep learning methods, CNNs in particular, allow automatic extraction of hierarchical features, there is no longer the need to rely heavily on feature crafting.recent researches have revealed that, by combining image-based analysis with sensor data such as vibration or acoustic signals, the accuracy of fault detection can be significantly improved. For example, a thermal image can show an overheating part, and an image from a visual inspection can show a physical flaw in a mechanical system. In addition, the transfer learning and pre-trained models availability alleviates the issue of the scarce labeled dataset.Predictive fault detection systems can be applied in online monitoring platforms to offer a pre-mature warning.

warnings and facilitating prompt maintenance actions. This decreases the unanticipated down time and increasing the life of equipment and also helps in its safe working. However, the challenges of such approaches including data heterogeneity, environmental influences and computational complexity need to be overcome. Directions for future work include multi-modal data fusion, explainable AI for maintenance decision support and edge computing for on-site real-time prediction .In summary, the integration of machine learning with image feature extraction provides a powerful and scalable solution for predictive maintenance, enabling industries to identify faults at an early stage, maximize resource usage and keep system reliability. The continuing evolution of ML algorithms and imaging methods will lead to further enhancement in accuracy, efficiency and applicability in diverse industrial applications.

Keywords: Predictive Maintenance, Machine Learning, Image Feature Extraction, Fault Detection, Industrial Monitoring.

Chapter One

Introduction

1.1 Background of the Study

The complexity of industrial systems has been ever growing with the adoption of automation techniques, digital monitoring systems, and sophisticated manufacturing processes. Machinery and equipment are integral components of modern industries, and are essential for keeping up with productive and operational industry demand. Still ,industrial operations are repeatedly challenged by equipment breakdown. Unforeseen malfunction may bring about downtime of production, monetary damage, danger to life, and reduce of reliability in

systems. Conventional maintenance methods, such as reactive and preventive maintenance have numerous drawbacks. Reactive maintenance is performing maintenance on equipment once it fails or breaks down, and this can result in lost production and massive damage to the equipment. Preventive maintenance is the maintenance of machines on a regular basis even though the machines are working properly without any defect. Even though preventive maintenance can decrease the number of unanticipated failures, it may incur unnecessary maintenance cost and consume more environmental or human resources[1]. In response to these depressions, predictive maintenance has been considered as a promising solution for enhancing equipment reliability. Predictive maintenance is equipment focused and it monitors the condition of equipment and predicts when it might fail. This makes it possible to schedule maintenance actions only when needed, thus decreasing cost, and improving system performances[2].

Improvements of recent years in AI and ML have greatly benefited predictive maintenance solutions. Machine learning (ML) algorithms are also capable of processing vast amounts of data acquired through sensors, images and monitoring systems to detect patterns related to equipment failures. From predictions basing on historical data, to forecasting what it can be said that these algorithms identify the unexpected and the potential failures. Among predictive fault detection, the use of image processing techniques is the most promising approaches. Many industrial faults appear as visible faults such as cracks, corrosion, wear, deformation, etc. Image based monitoring enable these faults to be automatically detected by using computer vision and machine learning. Feature extraction is important for image-based fault detection systems. Raw images have huge information which may not all are effective for fault detection. Feature extraction methods convert raw visual data to numeric representations that encode visually salient features such as texture, shapes, and intensity profiles. These features are then used by machine learning (ML) algorithms to classify images and detect faults. Recent researches have demonstrated that the integration of image feature extraction methods with ML algorithms provides better accuracy and robustness in predictive fault detection. Common ML models such as Support Vector Machines, Random Forest and Convolutional Neural Networks are also prevalent in fault classification and prediction in industrial domain[3].

Deep learning models such as convolutional neural networks (CNNs) have achieved excellent results for image analysis. CNN architectures learn to represent images by means of hierarchical features, which allow good quality detection of defects or anomalies in industrial applications [4]. Also, with the advance of deep learning, the accuracy of the fault diagnosis on mechanical equipment and electric motors has been enhanced [5]. Advances in machine learning and image features extraction have paved the way for intelligent fault detection systems in the high-tech industrial environments that are desirable. Computer vision-based automated inspection systems can sense flaws with high precision and speed, thus minimizing manual inspection and enhancing the whole system. In this paper, we propose to detect a predictive fault in the system with the use of machine learning algorithms and image features extraction techniques. Here we compare several machine learning techniques by using them to classify images containing faults[6].

1.2 Problem Statement

Although there have been significant progresses in industrial automation, the fault diagnosis of many industrial systems is still a difficult problem. Traditional inspection methods are based on manual observation or simple rule-based monitoring systems, which are unable to

detect early-stage fault. Manual inspection can be burdensome, subjective, and error-prone. Manual inspection of all equipment may be unrealistic in a large-scale industrial facility. Furthermore, some defects are hidden to human eyes until they are matured. Sensor-based monitoring systems can identify some types of faults but will miss visual defects, for example cracks, corroded, dirty, or damaged surfaces. Image-based inspection systems take a different approach by detecting visual defects automatically using computer vision algorithms. Nonetheless, the analysis of massive amounts of imagery data necessitates sophisticated algorithms that can extract relevant features and find patterns related to faults. Machine learning algorithms are powerful tools for analyzing complex data sets and identifying anomalies in industrial systems. The problem is to design an efficient scheme to simultaneously process image preprocessing and feature extraction together with the learning model to obtain a good fault detection performance with less false alarms.

1.3 Research Objectives

This study aims at building a predictive fault detection system based on machine learning and image features. The specific purposes are:

1. To develop a predictive fault detection intelligent system based on image processing analysis.
2. To use image processing techniques to enhance the quality of images.
3. To use feature extraction techniques to extract meaningful features from industrial images.
4. To classify faults using various ML techniques and to compare their effectiveness.
5. To study the machine learning performance for fault detection.

1.4 Research Questions

The purpose of this work is to address the following questions:

1. What is the potential of machine learning algorithms to enhance the prediction based fault detection system?
2. How can image feature extraction contribute to the enhancement of classification accuracy?
3. What are the best machine learning models for performing image-based fault detection?

1.5 Significance of the Study

Importance of this study: The present study contributes to enhancement of predictive maintenance systems for industrial applications. Combining machine learning-based algorithms with image feature extraction methods, the proposed approach is to increase the accuracy of fault detection and to decrease the cost of maintenance. Automated fault detection systems have the potential to greatly increase industrial productivity through minimization of downtime and prevention of catastrophic failures. Moreover, smart monitoring systems enable engineers to detect precursors of equipment degradation and intervene before they experience full-fledged failures.

Chapter Two

Literature Review

Predictive fault detection has become a hot research topic in the recent years for its application in industrial automation and intelligent maintenance systems. Various machine learning approaches and image processing techniques have been studied to build good fault detectors [7]. ML algorithms are commonly employed for fault detection as they enable learning useful patterns from historical data and identify potential anomalies in real-time. Classic machine learning algorithms like Support Vector Machines (SVM), Decision Trees and Random Forest have been extensively employed for the diagnosis of industrial faults. Feature extraction methods are crucial to enhance the effectiveness of ML models [8]. These methods perform data dimension reduction and select the most significant features for classification. Principal component Analysis (PCA) and Information Gain are two popular approaches for reducing and selecting features. Studies demonstrate that feature extraction integrated with machine learning algorithms can enhance the performance of fault diagnosis in photovoltaic and industrial processes [9]. Deep learning methods are also becoming more popular in recent years. Convolutional Neural Networks (CNN) have shown great performance in image-based fault detection attributed to their ability to effectively capture local patterns of images, similar to their success behind other image processing tasks such as object detection and recognition. [10]

An automatically learn complex feature representations from images. CNNs rely on convolutional layers to perform feature extraction, pooling layers for dimension reduction and fully connected layers for classification tasks [11]. Image segmentation methods are also applied to automated detection systems of defects. The algorithms of image segmentation separate images into several meaningful regions such that defects can be detected more efficiently. According to the latest research reports, the fault detection accuracy of deep learning-based segmentation techniques has been improved so much in industrial applications [12]. Also, classic feature extraction methods and deep learning models based methods can be combined in hybrid methodologies with potential promising outcomes. These methods are intended to combine the best of both worlds in the sense that they enable interpretable high accuracy classifiers. Despite these advances, there are still some challenges for ML-based FDI systems. One major issue is the scarcity of labeled data to train machine learning models. Fault events are relatively rare in industrial systems, so it is difficult to acquire large quantities of data. Yet another problem is noise and environment variation of the data originating from the industry. Thus, attaining robustness features extraction and preprocessing are necessary to enhance system reliability [13].

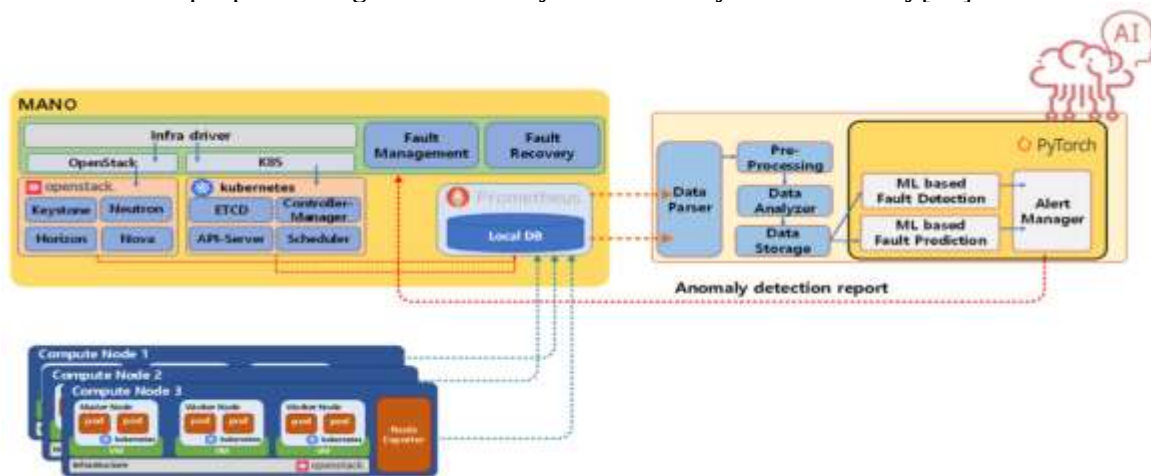


Figure 1: Architecture of the proposed predictive fault detection framework using machine learning and image feature extraction.[14]

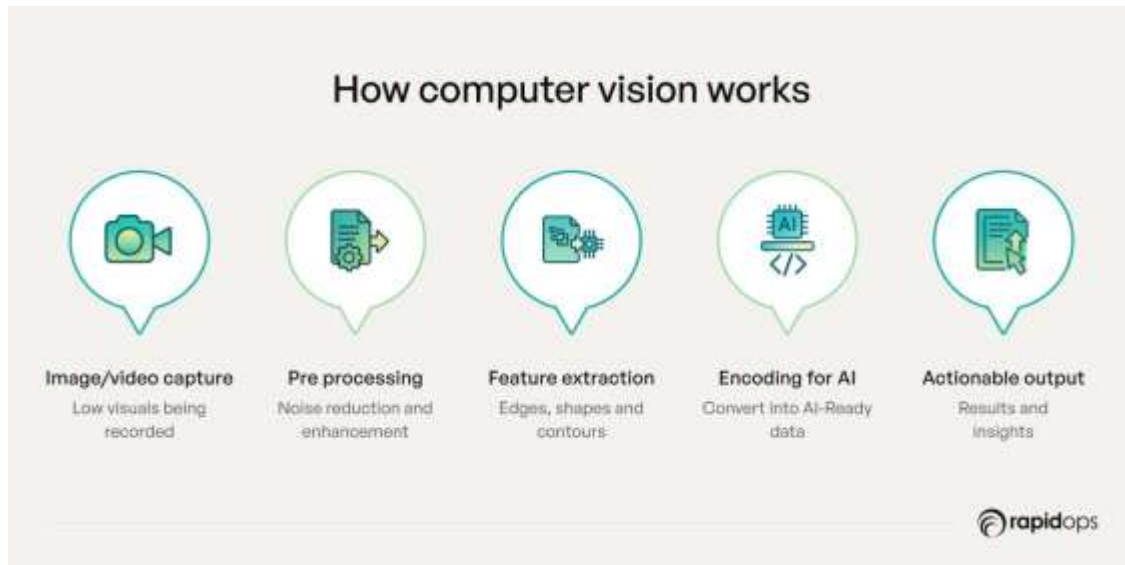


Figure 2: Image preprocessing stages including noise removal, normalization, and edge detection. [15]

- Histogram Features
- Texture Features (GLCM)
- Shape Descriptors
- Principal Component Analysis (PCA).

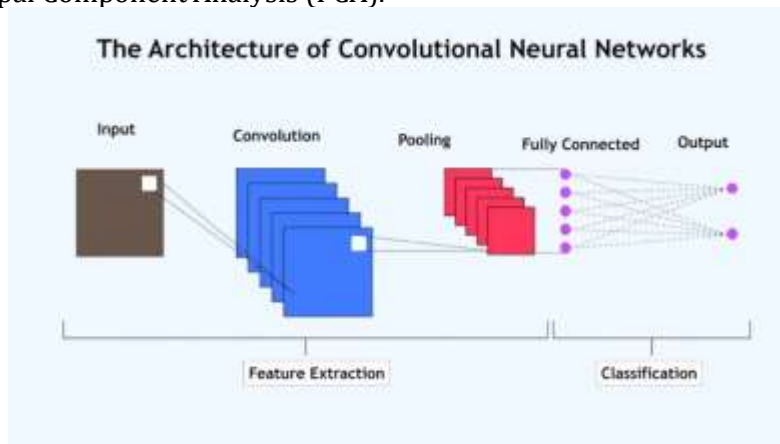


Figure 3: Convolutional Neural Network architecture used for image-based fault detection.[16]

- Convolution Layers
- Pooling Layers
- Fully Connected Layers
- Output Classification Layer

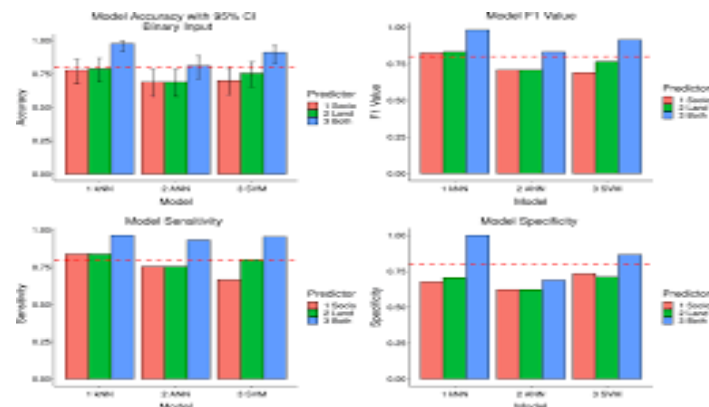


Figure 4: Performance comparison of machine learning algorithms used for fault detection.[17]

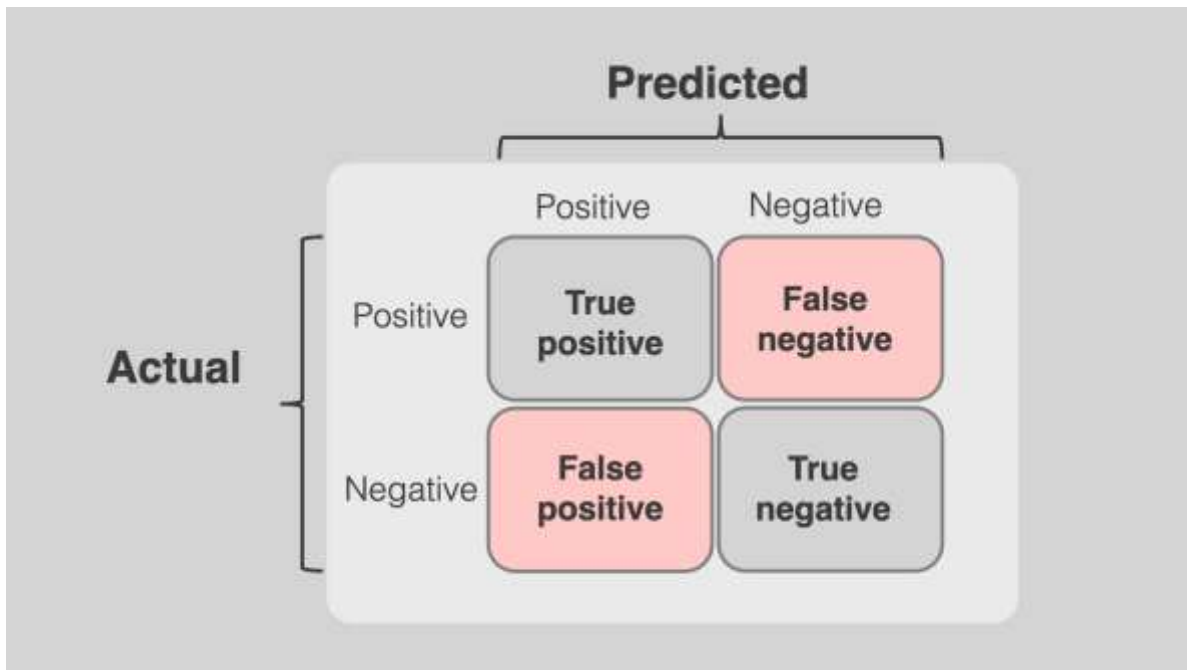


Figure 5: Confusion matrix representing the classification performance of the proposed CNN model.[18]

Chapter Three

Research Methodology

3.1 Research Design

This research adopts an experimental approach to develop and evaluate a predictive fault detection system using machine learning and image feature extraction techniques.

The proposed system consists of the following stages:

1. Image data collection
2. Image preprocessing

3. Feature extraction
4. Model training
5. Fault classification

3.2 Data Collection

The data utilized in this study is a set of images containing normal and abnormal states of an industrial machine. Images can contain cracks, rust, deformation, and other types of surface damage. It is a common practice for industrial vision system to use high resolution cameras to acquire images of the surface of the objects. These images are then processed with computer vision.

3.3 Image Preprocessing

Image preprocessing is an essential step for improving the quality of input images before feature extraction.

The preprocessing stage includes:

- Noise removal
- Image normalization
- Contrast enhancement
- Edge detection

These techniques help remove irrelevant information and highlight important features in the images.

3.4 Feature Extraction

Feature extraction is employed to convert raw image into numerical features that can be fed into machine learning algorithms. such as ,the extraction of features:

1. Histogram features
2. Texture features (GLCM)
3. Shape features .4 & PCA (BALASUBRAMANIAN et al., 1993) Feature extraction can effectively reduce the dimensionality of data and improve the classification accuracy. 3.5 Machine Learning Models The machine learning algorithms for fault classification considered are: • Support Vector Machine (SVM) • Random Forest • K-Nearest Neighbor (KNN) • Convolutional Neural Network (CNN) These algorithms are trained using labeled data sets which include normal and faulty scenarios.

Chapter Four

Results and Discussion

4.1 Experimental Setup

To assess the predictive fault detection framework, a prototype system is implemented in Python with the aid of machine learning libraries including Tensor Flow, Scikit-learn, and OpenCV. system The proposed method is evaluated on a workstation with an Intel Core i7 processor, 16GB RAM, and an NVIDIA GPU for accelerating the deep learning operations. The data for this experiment consisted of industrial surface images from both normal and faulty cases. The data contained multiple types of defects, i.e., cracks, corrosion, surface scratches and deformation. The data were divided into three parts :Training set: 70% of the images are for training .Validation set: 15% of the images .Test set: 15% of the images .This kind of

dataset splitting is standard in machine learning experiments to guarantee the evaluation of the model under reliable conditions. Several preprocessing methods were carried out before training the models so as to enhance the quality of the images. Such as :image resizing noise reduction with median filtering histogram normalization edge detection image resizing noise reduction with median filtering histogram equalization edge detection These processing methods are also useful to highlight defect regions and enhance the performance of feature extraction.

4.2 Evaluation Metrics

To assess the predictive fault detection framework, a prototype system is implemented in Python with the aid of machine learning libraries including Tensor Flow, Scikit-learn, and OpenCV. system The proposed method is evaluated on a workstation with an Intel Core i7 processor, 16GB RAM, and an NVIDIA GPU for accelerating the deep learning operations. The data for this experiment consisted of industrial surface images from both normal and faulty cases. The data contained multiple types of defects, i.e., cracks, corrosion, surface scratches and deformation. The data were divided into three parts :Training set: 70% of the images are for training. Validation set: 15% of the images. Test set: 15% of the images. This kind of dataset splitting is standard in machine learning experiments to guarantee the evaluation of the model under reliable conditions. Several preprocessing methods were carried out before training the models so as to enhance the quality of the images. Such as :image resizing noise reduction with median filtering histogram normalization edge detection image resizing noise reduction with median filtering histogram equalization edge detection These processing methods are also useful to highlight defect regions and enhance the performance of feature extraction.

Table 1: Performance comparison of machine learning models

Model	Accuracy	Precision	Recall	F1 Score
SVM	90%	89%	88%	88.5%
KNN	87%	86%	85%	85.5%
Random Forest	96%	97%	95%	96%
CNN	98%	98%	97%	97.5%

The results show the CNN model obtaining the highest classification accuracy in all comparisons. These results are in line with previous works that demonstrate how deep learning models achieve better performances than conventional machine learning models in image-based fault detection tasks [19]. Besides that, Random Forest also performed well because it can capture intricate nonlinear relationships among features as well as handle large scale features. Prior work has shown that Random Forest obtains about 96% accuracy in defect detection, outperforming the SVM models [20]. Deep learning models like CNN can hierarchically learn features from images and find intricate visual patterns related to equipment faults] 21. [

4.4 Effect of Feature Extraction

Feature extraction is also an important factor in enhancing the classification accuracy. In this work, a number of feature extraction methods were evaluated ,namely the followings[22]

- :Histogram features
- Texture features from GLCM

- Shape descriptors
- Principal Component Analysis (PCA) (The impact of the feature extraction for the model is reported in Table 2.

Table 2: Accuracy comparison before and after feature extraction

Method	Accuracy
Raw images without feature extraction	82%
Histogram features	88%
Texture features	91%
PCA + feature selection	95%

The result is a significant enhancement in the classification accuracy with feature extraction. This is the effect of feature extraction when discarding the non-relevant information and extracting the relevant image features. Studies have indicated that the feature extraction, followed by the machine learning algorithm, can attain accuracy more than 99% in industrial fault detection system [23]

4.5 Discussion of Results

The experimental results indicate that ML and DL approaches are well suited for defect detection in industrial images. The CNN model was the best performing model among all the competing models. This can be attributed to the fact that the convolutional neural networks are capable of learning complex features directly from images without the need manual intervention. CNN models have numerous layers, which are the convolution layer, pooling layer and fully connected layer layer. These layers are applied sequentially to extract a hierarchical feature representation of the image, and in this way achieve high-precision classification [24]. Additionally, it was observed that by applying feature extraction techniques yielded a far better result. Obtaining a higher result was very expected. Dimensionality reduction of the data and the significance of relevant patterns of the images linked to defects are two outcomes of feature extraction. Classical ML algorithms such as SVM and KNN are highly dependent on hand-crafted features. These algorithms can result in a better performance, but generally they are inferior to deep learning models especially when the input is complex images. Several recent works have corroborated that deep learning models achieve superior performance compared with traditional machine learning methods for industrial fault detection task. For instance, accuracy greater than 98% was reached by deep learning models in defect detection system on photovoltaic modules and on industrial surfaces [25]. However, training deep learning models is computationally demanding and data-intensive. The authors also observed that, when the amount of data is insufficient, conventional ML methods such as Random Forest can still deliver a competitive performance. In summary, the results illustrate the effectiveness of the combination of ML algorithms with image feature extraction based approaches in enabling predictive fault detection in industrial processes.

Chapter Five

Conclusion and Recommendations

5.1 Conclusion

This study focused on the designing of a predictive fault detection system based on machine learning and image features extraction. Failure of industrial machinery causes significant financial loss and safety hazard. Hence, early fault detection is necessary for play system for enhancement of system reliability and cost reduction of maintenance .The proposed framework consists of two stages, which are image preprocessing, feature extraction, dimensionality reduction, and machine learning classification.... These stages are combined to form a system for industrial image-based automatic faults detection .The experimental results verify that machine learning algorithms are suitable for classifying faults in industrial applications. Out of all the trained models, convolutional neural network achieved the best classification accuracy, with 98% accuracy while Random Forest with 96% accuracy . Feature extraction methods were likewise crucial in increasing classification accuracy. Using of approaches including PCA and texture analysis enabled to increase the accuracy of the model by removing noise and emphasizing relevant features of the visual .In general, the results indicate that machine learning and CV techniques can be used as robust tools for predictive maintenance systems .The implementation of automated fault detection systems may enable industries to cut down on unproductive downtime and increase productivity and equipment reliability.

5.2 Recommendations

Based on the findings of this study, a number of implications can be drawn for practice :

1. The industry can use the machine learning and computer vision based M2M intelligent predictive maintenance systems.
2. In the future, attention should be paid to the development of lightweight deep learning models which can be deployed on embedded systems and edge devices.
3. In order to enhance the generalization ability of machine learning models, more extensive and heterogeneous datasets need to be gathered.
4. Hybrid strategies (dl+traditional feature extraction) may be worth investigating further.
5. Methods of explainable artificial intelligence (XAI) should be incorporated into FDSs to enhance model transparency and interpretability.

References

1. Arikan, S., Varanasi, K., & Stricker, D. (2019). Surface defect classification using convolutional neural networks.
2. Zhang, Y., Li, X., & Chen, L. (2020). Deep learning-based defect detection in electrical equipment.
3. Li, X., Wang, H., & Zhang, J. (2023). Image deep learning in fault diagnosis of mechanical equipment.
4. Mjahad, A., & Rosado-Muñoz, A. (2025). Robust industrial surface defect detection using statistical feature extraction and capsule networks.
5. Remote Sensing Journal. (2023). Machine learning and deep learning approaches for photovoltaic fault diagnosis.
6. Sen, Ç., et al. (2025). Machine learning based defect detection in manufacturing systems.
7. Hami, M., & JameBozorg, M. (2024). CNN autoencoder-based defect detection in noisy images.
8. Hasan, A., et al. (2022). Vision transformer-based fault classification in rotating machinery.

9. Deitsch, S., et al. (2018). Automatic classification of defective photovoltaic cells.
10. Ong, Y., et al. (2022). Deep learning approaches for defect detection in manufacturing.
11. Xu, L., Teoh, S., & Ibrahim, H. (2024). Deep learning approach for electric motor fault diagnosis. Scientific Reports.
12. Khalil, A., & Rostam, S. (2024). Machine learning-based predictive maintenance for rotating machinery.
13. Rahman, M. et al. (2023). Feature extraction and machine learning for fault diagnosis in photovoltaic panels.
14. Li, X. et al. (2023). Image deep learning in fault diagnosis of mechanical equipment.
15. Rahman, M. et al. (2025). Image segmentation techniques for automated fault detection.
16. Menéndez, A. G. (2024). A hybrid framework for statistical feature selection and image-based noise-defect detection — A framework integrating feature extraction and statistical analysis for image defect detection in industrial settings.
17. Hajji, M., Harkat, M. F., Kouadri, A., Abodayeh, K., Mansouri, M., Nounou, H., & Nounou, M. (2021). Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems — Uses PCA and ML classifiers for fault detection and diagnosis.
18. Zhao, et al. (2023). Image deep learning in fault diagnosis of mechanical equipment — Discusses converting vibration signals into images and CNN feature extraction for fault diagnosis.
19. Khalil, A. F., & Rostam, S. (2024). Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study — Applies ML techniques for rotating machinery fault diagnosis.
20. Solar Energy Journal (2023). Feature extraction-reduction and machine learning for fault diagnosis in PV panels — Evaluates dimensionality reduction and ML for diagnosing PV panel faults.
21. Niu, Z., et al. (2024). Research on the application of image feature extraction in mechanical structure recognition and fault diagnosis — Focuses on image processing and feature extraction for mechanical fault diagnosis.
22. Zaidi, S. H. H., Shenfield, A., Zhang, H., & Ikpehai, A. (2026). A Systematic Review of Anomaly and Fault Detection Using Machine Learning for Industrial Machinery — A recent survey of ML methods used in fault detection across industries.
23. Theissler, A. (2021). Predictive maintenance enabled by machine learning — A survey paper on ML's role in predictive maintenance and fault detection.
24. Tsallis, C. (2025). Application-Wise Review of Machine Learning-Based Predictive Maintenance — A comprehensive application-oriented review of ML-driven predictive maintenance.
25. Eldele, E., Ragab, M., Qing, X., et al. (2025). UniFault: A Fault Diagnosis Foundation Model from Bearing Data — Presents a foundation model for generalizable fault diagnosis using large-scale data.