International Journal of Mechatronics and Applied Mechanics
2026, Issue 24, Vol. I

INTELLIGENT FAULT DIAGNOSIS OF ELECTROMECHANICAL EQUIPMENT BASED ON IOT AND DEEP LEARNING

Xiali Yue
Department of Intelligent Manufacturing, Xinxiang Vocational and Technical College,
Xinxiang 453000, Henan, China

Abstract - Fault diagnosis of electromechanical equipment plays an important role in ensuring the reliability and safety of industrial systems. Traditional machine learning methods often struggle to capture complex nonlinear relationships in high-dimensional data, especially under imbalanced conditions. To address these challenges, this study proposes a hybrid framework that combines artificial neural network (ANN)-based feature extraction with ensemble machine learning models for fault diagnosis. First, an ANN is employed to learn high-level representations from the original input features. The extracted features are then fed into multiple machine learning models, including random forest (RF), decision tree (DT), and k-nearest neighbors (KNN), to perform classification. To further improve performance, the best-performing models are selected and combined using a logistic regression-based adaptive weighting strategy. In addition, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to alleviate the class imbalance problem. Experimental results conducted on the publicly available AI4I 2020 predictive maintenance dataset, which is a synthetic dataset designed to simulate industrial electromechanical equipment operating conditions, demonstrate that the proposed framework achieves superior performance compared with individual models. In particular, the ensemble model (ANN+RF+KNN) achieves the best results in terms of F1-score, indicating a better balance between precision and recall. The findings show that combining deep feature learning with ensemble strategies can significantly enhance fault diagnosis performance.

Keywords: Electromechanical equipment; Fault diagnosis; Artificial neural network; Ensemble learning.

1. Introduction

With the rapid development of industrial automation and intelligent manufacturing [1-3], electromechanical equipment has become a key component in modern production systems. These machines often operate under complex conditions for long periods, making them prone to various types of faults. Unexpected failures can lead to production interruptions, increased maintenance costs, and even safety risks. Therefore, developing effective fault diagnosis methods for electromechanical equipment has become an important research topic.

In recent years, the integration of the Internet of Things (IoT) has enabled real-time monitoring of equipment through various sensors [4-6]. Large amounts of operational data, such as temperature, rotational speed, and torque, can now be collected continuously.

This provides a solid data foundation for intelligent fault diagnosis. At the same time, Machine Learning (ML) methods have been widely applied to analyze such data and predict equipment failures [7, 8]. Traditional ML models, including decision trees, random forests, and k-nearest neighbors, have shown good performance in many prediction tasks due to their simplicity and efficiency.

However, these methods still have some limitations. First, most traditional models rely heavily on manually designed features, which may not fully capture the complex relationships in the data. Second, their performance can be limited when dealing with high-dimensional or nonlinear data. To address these issues, deep learning methods, especially Artificial Neural Networks (ANNs) [9, 10], have been introduced. ANNs are capable of automatically learning useful representations from raw data, which helps improve prediction accuracy.

dx.doi.org/10.17683/ijomam/issue24.24

Despite these advantages, using deep learning alone also has challenges. Deep models often require large amounts of data and may lack interpretability. In addition, classical ML models sometimes perform better on structured data and small datasets. Therefore, combining deep learning with traditional machine learning methods has become a promising direction.

In this paper, we propose an intelligent fault diagnosis framework for electromechanical equipment based on IoT and deep learning. The collected sensor data are first preprocessed and used as input features. An ANN model is then employed to extract high-level representations from the original data. These learned features are further fed into multiple machine learning models for classification. Finally, the best-performing models are selected and combined to produce the final prediction. This approach aims to take advantage of both deep feature extraction and the robustness of traditional ML models.

The rest of this paper is organized as follows. Section 2 reviews related work in machine learning-based fault diagnosis approaches. Section 3 introduces the proposed methodology in detail. Section 4 describes the experimental setup and dataset. Section 5 presents and discusses the experimental results. Finally, Section 6 concludes the paper and outlines future work.

2. Related Works

Recent studies on fault diagnosis of electromechanical equipment have increasingly adopted machine learning techniques to improve diagnostic accuracy and reliability. Compared with traditional rule-based or purely signal-threshold methods [11, 12], machine learning approaches can better capture nonlinear relationships hidden in sensor data such as vibration, temperature, speed, and torque [13, 14]. A recent review also noted that machine learning-based fault diagnosis for rotating machinery usually follows a data-driven pipeline of signal acquisition, feature extraction, and fault classification, and has become an important direction in intelligent maintenance research.

Among classical machine learning methods, random forest and support vector machine have been widely used in machinery fault diagnosis. For instance, Tang et al. proposed a rolling bearing fault diagnosis method based on improved fast spectral correlation and optimized random forest, showing that the combination of signal processing and ensemble learning can improve diagnosis performance [15]. Zhao et al. proposed a wind turbine bearing fault diagnosis method based on stochastic subspace identification and multi-kernel Support Vector Machine (SVM), in which vibration features were first extracted and then classified by a hybrid SVM structure [16].

With the development of deep learning, more researchers have shifted from handcrafted features to automatic feature learning. Liu et al. proposed a denoising convolutional autoencoder and 1-D Convolutional Neural Network (CNN) framework for rotating machinery fault diagnosis under noisy conditions, where the autoencoder was used for denoising and the CNN was used for diagnosis [17]. Zhang et al. introduced an ensemble deep autoencoder model for rotating machinery and reported that it could automatically learn representative features and outperform traditional methods that depend on handcrafted features [18]. In another study, a 1DCNN-LSTM model was developed for bearing fault diagnosis, combining convolutional layers for local feature extraction and LSTM units for sequence modeling [19]. These works indicate that deep learning can improve feature representation, especially for complex and noisy sensor signals.

Although these studies have achieved encouraging results, limitations still remain. While deep learning models can extract high-level features automatically, single-model approaches may still fail to fully combine the representation ability of neural networks with the stability and interpretability of classical machine learning classifiers. These limitations motivate the development of hybrid frameworks that integrate deep feature extraction with multiple downstream classifiers for more reliable fault diagnosis.

3. Deep Feature-based Ensemble Learning for Electromechanical Fault Diagnosis

3.1 The Introduction of Ensemble Learning

Ensemble learning is a machine learning technique that improves prediction performance by combining multiple models instead of relying on a single classifier [20, 21]. The basic idea is that different models may capture different patterns in the data, and integrating their predictions can lead to more accurate and stable results. Common ensemble methods include bagging, boosting, and voting. Bagging methods, such as random forest, reduce variance by training multiple models on different subsets of data. Boosting methods focus on improving weak learners by giving more attention to misclassified samples. Voting-based methods combine predictions from multiple models through majority or weighted voting.

Ensemble learning has been widely applied in various classification tasks, especially when the data are complex or noisy. By using the diversity among models, ensemble methods can enhance robustness and reduce the risk of overfitting compared with single-model approaches.

3.2 Overview of the Developed Framework

The overall framework of the proposed method is illustrated in Figure 1. First, operational data are collected from electromechanical equipment through IoT-based sensing systems. These data include key variables such as air temperature, process temperature, rotational speed, and torque. After preprocessing, the selected features are used as inputs to an ANN for representation learning.

The ANN is employed to automatically extract high-level features from the original data, aiming to capture complex nonlinear relationships among different variables.

Compared with raw features, these learned representations are more informative and suitable for classification tasks. The extracted features are then fed into multiple machine learning models, including random forest, decision tree, and k-nearest neighbors, to perform fault classification.

After model training, the best-performing classifiers are selected based on evaluation metrics. Their outputs are further combined through a weighted fusion strategy to obtain the final prediction result. This framework integrates deep feature extraction and ensemble learning, aiming to improve both accuracy and robustness in fault diagnosis tasks.

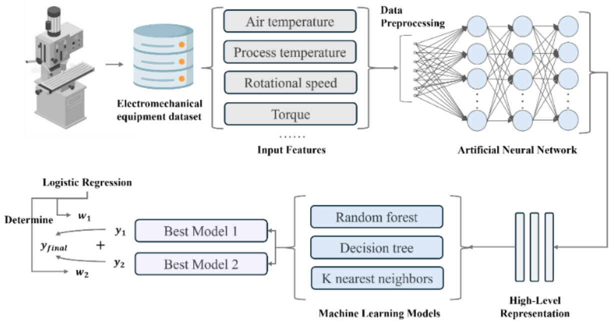

Figure 1: Overview of the proposed deep feature-based ensemble learning framework for electromechanical equipment fault diagnosis

3.3 ANN-based Deep Feature Extraction

The artificial neural network is used to learn high-level representations from the original input features [22, 23]. Let the input sample be denoted as $\mathrm{x} \in \mathbb{R}^d$, where d is the number of input features.

In this study, a fully connected ANN with four layers shown in Figure 2 is constructed. The network consists of an input layer, three hidden layers with 64, 32, and 16 neurons, respectively, and an output layer with a sigmoid activation function. The rectified linear unit (ReLU) is used as the activation function in all hidden layers to introduce nonlinearity. The transformation of the l-th layer is defined as:

$$\mathrm{h}^{(l)} = \sigma\left(\mathrm{W}^{(l)}\mathrm{h}^{(l-1)} + \mathrm{b}^{(l)}\right) \tag{1}$$

where $\mathrm{h}^{(l)}$ denotes the output of the l-th layer, $\mathrm{W}^{(l)}$ and $\mathrm{b}^{(l)}$ are the weight matrix and bias vector, $\sigma(\cdot)$ is the activation function, and $\mathrm{h}^{(0)} = \mathrm{x}$.

The third hidden layer with 16 neurons is selected as the feature representation layer. The extracted feature vector is defined as:

$$\mathrm{z} = \mathrm{h}^{(L-1)} \in \mathbb{R}^k \tag{2}$$

where z is the extracted feature vector and k is the dimension of the learned feature space.

The ANN is trained using the binary cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \tag{3}$$

where y_i and $\hat{y}_i$ denote the true label and predicted probability of the i-th sample, respectively, and N is the total number of samples.

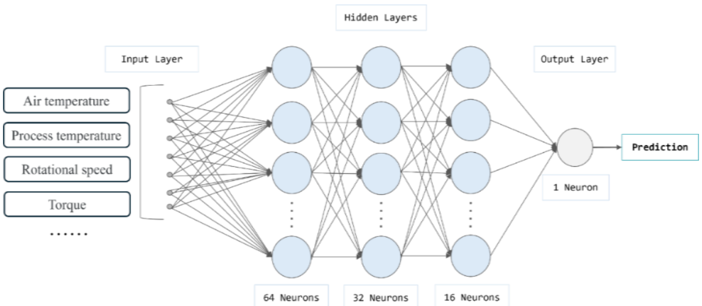

Figure 2: The architecture of the developed artificial neural network

3.4 Machine Learning-based Classification Models

The extracted feature representation $\mathbf{z}$ is used as the input for multiple machine learning classifiers including random forest, decision tree and k-nearest neighbors [24-26]. The introduction of these algorithms is as follows: 1) Random forest is an ensemble-based method that builds multiple decision trees using random subsets of data and features, and combines their outputs to improve robustness and reduce overfitting. 2) Decision tree is a simple and interpretable model that recursively partitions the feature space based on feature thresholds, making it suitable for capturing nonlinear decision boundaries. 3) K-nearest neighbors is a distance-based method that assigns labels according to the majority class of the closest samples in the feature space, which is effective for modeling local data structures. By combining these models, different perspectives of the learned feature space can be exploited for more reliable fault classification.

For random forest (RF), the final prediction is obtained by aggregating multiple decision trees:

$$\hat{y} = \mathrm{mode}(T_1(\mathrm{z}), T_2(\mathrm{z}), \dots, T_M(\mathrm{z})) \quad (4)$$

where $T_m(\cdot)$ denotes the m-th decision tree and M is the total number of trees.

For decision tree (DT), the model recursively splits the feature space based on an impurity criterion such as the Gini index:

$$G = 1 - \sum_{c=1}^{C} p_c^2 \quad (5)$$

where p_c represents the probability of class c at a given node and C is the number of classes.

For k-nearest neighbors (KNN), the prediction is determined by the majority label of the nearest samples:

$$\hat{y} = \mathrm{mode}\left(y_{i_1}, y_{i_2}, \dots, y_{i_K}\right) \quad (6)$$

where $i_1, \dots, i_K$ are the indices of the K nearest samples under a distance metric such as Euclidean distance $\| \mathrm{z}_i - \mathrm{z}_j \|_2$.

3.5 Adaptive Weight Fusion via Logistic Regression

To further improve prediction performance, the outputs of the best-performing classifiers are combined using a logistic regression-based fusion strategy. Logistic regression is a linear classification model that estimates the probability of a binary outcome through a weighted combination of input variables followed by a sigmoid function [27, 28]. It is widely used due to its simplicity, interpretability, and ability to learn optimal weights from data. In this work, logistic regression is used as a meta-classifier to adaptively combine the prediction probabilities of the selected models, allowing the framework to assign higher importance to more reliable classifiers.

Let the predicted probabilities of two selected models be: $p_1(\mathrm{z}), p_2(\mathrm{z})$

The final prediction is defined as:

$$\hat{y} = \sigma(w_1 p_1(\mathrm{z}) + w_2 p_2(\mathrm{z}) + b) \quad (7)$$

where w_1 and w_2 are learnable weights, b is the bias term, and $\sigma(\cdot)$ is the sigmoid function.

The parameters are optimized by minimizing the binary cross-entropy loss:

$$\mathcal{L}_{fusion} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (8)$$

where y_i and $\hat{y}_i$ denote the true and predicted labels of the i-th sample, respectively, and N is the number of training samples.

4. Experimental Setup

4.1Dataset Collection and Preprocessing

- **Dataset Description**

The dataset used in this study is the Predictive Maintenance Dataset (AI4I 2020), which is publicly available on Kaggle and originally proposed by Matzka [29]. It is a synthetic dataset modeled after a real milling machine and contains 10,000 samples with 14 features.

Each sample represents the operating condition of an electromechanical system and includes several process-related variables. The main features consist of air temperature, process temperature, rotational speed, torque, and tool wear, which reflect the physical state of the equipment during operation. In addition, product-related information such as product type is included, representing different quality levels (low, medium, and high). The dataset also contains a binary label named “machine failure,” indicating whether a failure occurs in a given sample.

The failure label is determined by five independent failure modes, including tool wear failure, heat dissipation failure, power failure, overstrain failure, and random failure. A sample is labeled as failure if any of these conditions are satisfied. It should be noted that these failure modes are not directly used as input features in this study to avoid data leakage.

- **Dataset Preprocessing**

To ensure the quality and reliability of the input data, several preprocessing and feature engineering steps are performed before model training. First, irrelevant features are removed from the dataset. Identifier-related variables such as UDI and Product ID are excluded since they do not contribute to the prediction task. In addition, failure mode indicators (TWF, HDF, PWF, OSF, RNF) are removed to prevent data leakage, as the target variable "machine failure" is directly derived from these conditions. After this step, the remaining variables are used as input features X, and the corresponding label is denoted as $y \in \{0,1\}$.

The dataset is then randomly divided into training and test sets with a ratio of 80:20 using stratified sampling to preserve the class distribution. This ensures that both subsets maintain similar proportions of normal and failure samples.

Furthermore, the dataset suffers from a significant class imbalance problem, where normal samples greatly outnumber failure samples. Such imbalance may bias the model toward the majority class and reduce its ability to detect failures. To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training data [30, 31].

SMOTE generates synthetic samples for the minority class by interpolating between existing samples. For a given minority sample x_i, a new sample is constructed as:

$$x_{new} = x_i + \lambda(x_{nn} - x_i) \tag{9}$$

where x_{nn} is a randomly selected nearest neighbor of x_i within the minority class, and $\lambda \in [0,1]$ is a random coefficient. This approach increases the diversity of minority samples and helps the model learn a more balanced decision boundary.

SMOTE is applied only to the training set, while the test set remains unchanged to ensure a fair evaluation of the model performance. After this process, the training data become more balanced, which improves the robustness and generalization ability of the model in fault diagnosis tasks.

4.2 Training Configuration

The proposed framework is implemented using the TensorFlow deep learning framework and Scikit-learn framework with the Python programming language. The ANN is trained using the Adam optimizer, which is widely used for efficient stochastic gradient optimization. The binary cross-entropy loss function is adopted to handle the binary classification task of machine failure prediction.

During training, the model is trained for 30 epochs with a batch size of 64. A validation split of 0.2 is used to monitor the training process and prevent overfitting.

The input data are first processed using SMOTE to balance the class distribution, and the balanced training set is used for model learning.

After training, the ANN is used as a feature extractor, and the learned representations are fed into multiple machine learning models, including random forest, decision tree, and k-nearest neighbors. These models are trained using their default settings, with the number of estimators in the random forest set to 100. The main hyperparameters used in this study are summarized in Table 1.

Table 1. Training hyperparameters of the proposed framework

Parameter	Value
Programming language	Python
Deep learning framework	TensorFlow
Optimizer	Adam
Loss function	Binary cross-entropy
Batch size	64
Epochs	30
Validation split	0.2
Random forest estimators	100

5. Empirical Results on Electromechanical Fault Diagnosis

5.1 Data Distribution and Feature Analysis

To better understand the characteristics of the dataset, an initial analysis of feature distributions and correlations is conducted. As shown in Figure 3, the distributions of key input features, including air temperature, process temperature, rotational speed, torque, and tool wear, exhibit different patterns. The temperature-related variables follow relatively smooth and approximately normal distributions, indicating stable operating conditions. In contrast, rotational speed shows a slightly skewed distribution with a long tail, suggesting the presence of varying operating regimes. Torque also follows a near-normal distribution centered around a moderate range, while tool wear appears to be more uniformly distributed across its range. These observations indicate that the dataset contains both stable and variable features, which may contribute differently to fault prediction.

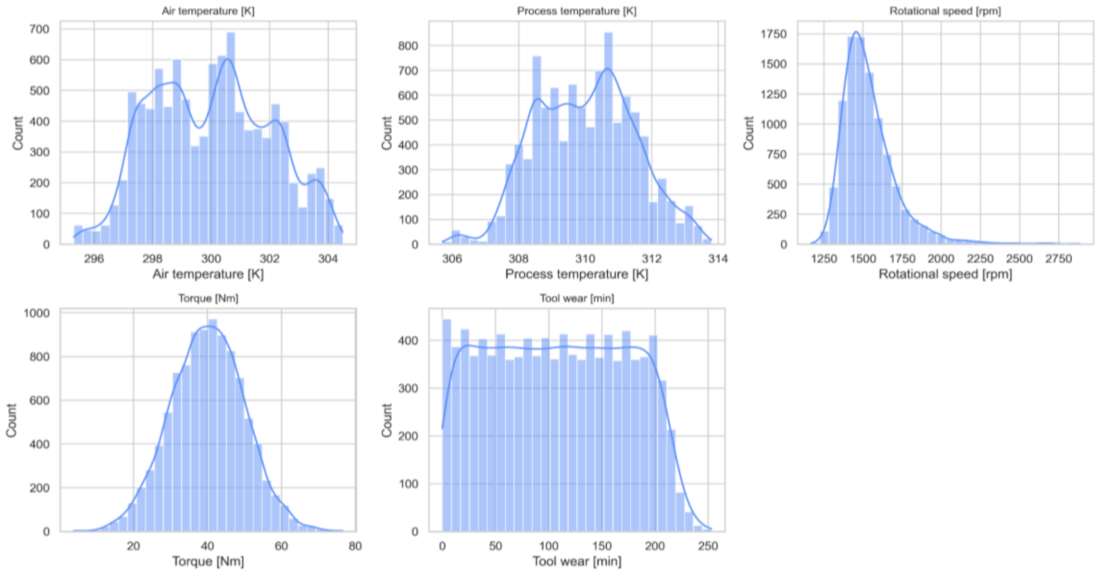

Figure 3: Distribution of key input features

Figure 4 presents the correlation matrix among the input features and the target variable. It can be observed that air temperature and process temperature are strongly positively correlated, which is expected since process temperature is partially derived from air temperature. Other features, such as rotational speed and torque, show relatively weak correlations with each other.

The correlation between individual features and the target variable (machine failure) is generally low to moderate, suggesting that fault occurrence is influenced by a combination of multiple factors rather than a single dominant variable.

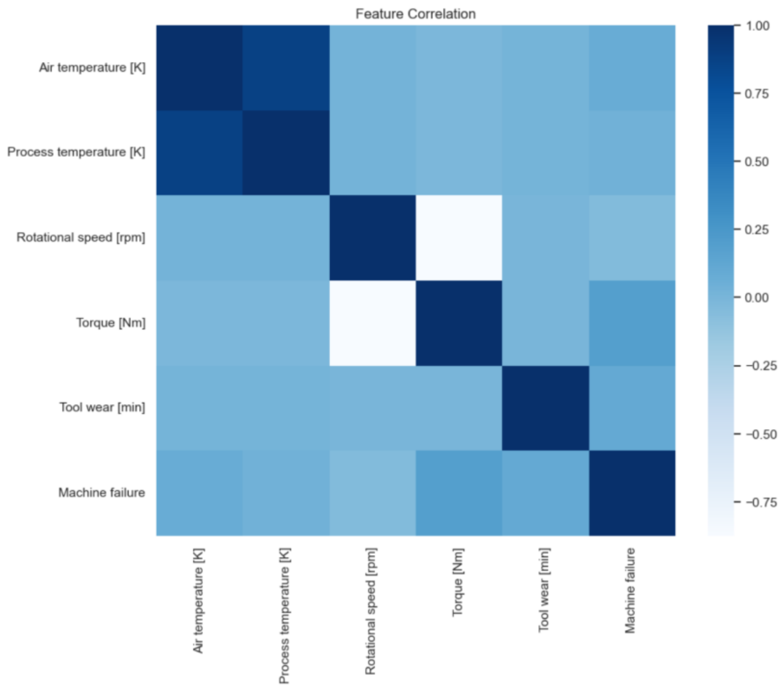

Figure 4: Correlation matrix of input features and target variable

As shown in Figure 5, the original dataset shows a severe class imbalance, where the number of normal samples significantly exceeds that of failure samples. Such imbalance may lead to biased model training, causing the classifier to favor the majority class and overlook minority failure cases.

After applying the SMOTE technique, the class distribution becomes more balanced, with the number of failure samples significantly increased. This allows the model to better learn the characteristics of the minority class and improves its ability to detect faults. By reducing the bias toward the majority class, SMOTE helps enhance the overall robustness and effectiveness of the fault diagnosis model.

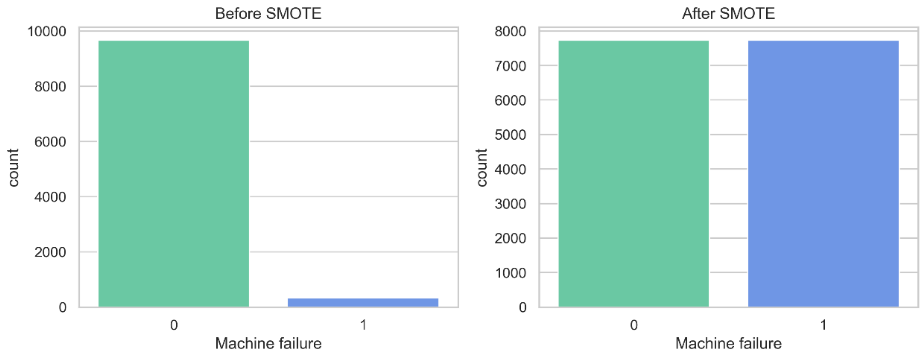

Figure 5: Class distribution before and after applying SMOTE

5.2 Performance of Developed Framework and Feature Representation Analysis

To evaluate the effectiveness of the ANN, both the training process and the quality of the learned feature representations are analyzed. As shown in Figure 6, the training and validation loss decrease steadily as the number of epochs increases, indicating that the ANN model is able to learn meaningful patterns from the data. The gap between the training and validation curves remains relatively small throughout the training process, suggesting that overfitting is well controlled. Although slight fluctuations can be observed in the validation loss, the overall trend remains stable, which demonstrates the robustness of the model.

To further investigate the quality of the learned feature representations, t-SNE is applied to visualize the extracted features, as shown in Figure 6. It can be observed that the ANN-transformed features exhibit a clear separation between normal and failure samples. The two classes form distinct clusters with limited overlap, indicating that the learned feature space has strong discriminative ability. Compared with the original feature space, the separation between classes is significantly improved, which facilitates more accurate classification by subsequent machine learning models.

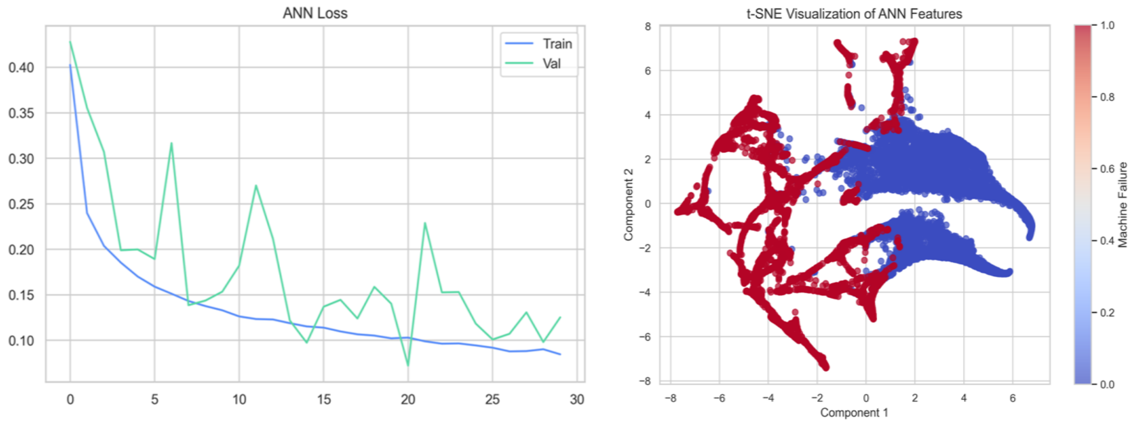

Figure 6: Training and validation loss of the ANN model and t-SNE visualization of ANN-extracted features

To further evaluate the effectiveness of the learned feature representations, the extracted features from the ANN are fed into different machine learning models, including RF, DT, and KNN. The performance of these models is compared using multiple evaluation metrics, including accuracy, precision, recall, and F1-score, as shown in Figure 7.

It can be observed that all models achieve relatively high accuracy, indicating that the ANN-extracted features provide a strong basis for classification. Among the models, the random forest and KNN classifiers generally achieve better performance compared with the decision tree, particularly in terms of precision and F1-score. This suggests that ensemble-based and distance-based methods are more effective in capturing the structure of the learned feature space.

In addition, the radar chart provides a more intuitive comparison of model performance across different metrics. While the differences in accuracy are relatively small, variations can be observed in precision, recall, and F1-score, highlighting the importance of considering multiple evaluation criteria, especially under imbalanced data conditions. Based on these results, the best-performing models including ensemble ANN+RF and ensemble ANN+KNN are selected for further fusion.

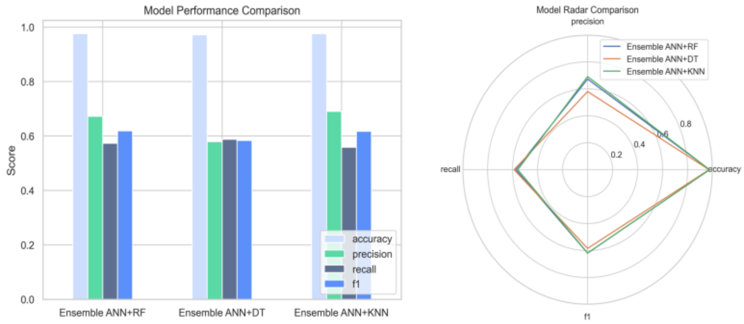

Figure 7: Performance comparison of machine learning models based on ANN-extracted features

Based on the previous analysis, the two best-performing models, namely ANN+RF and ANN+KNN, are selected for further fusion. Their prediction probabilities are combined using a logistic regression-based adaptive weighting strategy to obtain the final ensemble model. As shown in Figure 8, the confusion matrix indicates that the proposed model achieves high accuracy in identifying normal samples, with most samples correctly classified. At the same time, the model is also able to detect a considerable number of failure cases, although a small number of misclassifications still exist.

This is mainly due to the inherent difficulty of distinguishing failure samples under imbalanced conditions.

In terms of evaluation metrics, the final model achieves high accuracy, while keeping balanced precision, recall, and F1-score. Compared with individual models, the fusion approach shows improved overall performance, especially in terms of F1-score, which reflects a better trade-off between precision and recall. This demonstrates that the proposed ensemble strategy effectively integrates the strengths of different classifiers.

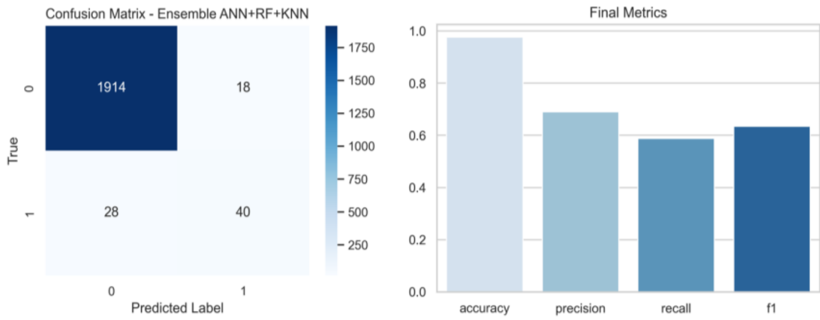

Figure 8: Confusion matrix and evaluation metrics of the final ensemble model (ANN+RF+KNN)

To further evaluate the performance of different models, confusion matrices of the ensemble models are presented in Figure 9.

It can be observed that all models achieve high accuracy in identifying normal samples, as indicated by the large number of true negatives.

However, differences can be observed in their ability to detect failure cases. Among the models, the ensemble ANN+RF+KNN and ANN+RF achieve a better balance between correctly identifying failure samples and reducing false alarms, while ANN+DT shows relatively weaker performance with more misclassifications.

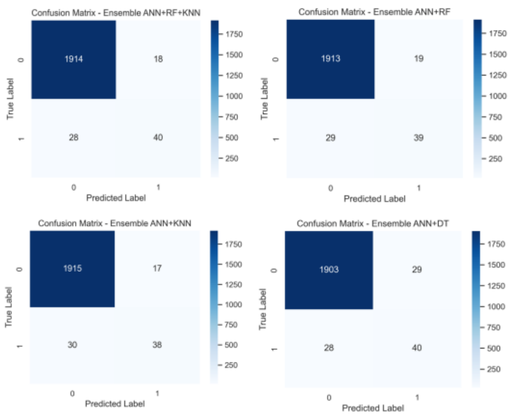

Figure 9: Confusion matrices of different ensemble models based on ANN-extracted features

Table 2 provides a detailed comparison of all models using accuracy, precision, recall, and F1-score. It can be seen that the proposed ensemble model (ANN+RF+KNN) achieves the highest overall performance, with an accuracy of 0.9765 and an F1-score of 0.6349. Compared with single models such as ANN, RF, KNN, and DT, the ensemble models consistently achieve better results, demonstrating the effectiveness of combining multiple classifiers. Table 2 provides a detailed comparison of all models using accuracy, precision, recall, and F1-score. It can be seen that the proposed ensemble model (ANN+RF+KNN) achieves the highest overall performance, with an accuracy of 0.9765 and an F1-score of 0.6349. Compared with single models such as ANN, RF, KNN, and DT, the ensemble models consistently achieve better results, demonstrating the effectiveness of combining multiple classifiers.

It is also worth noting that although KNN achieves the highest recall (0.8088), its precision is relatively low, indicating a higher false positive rate. In contrast, the proposed ensemble model maintains a better balance between precision and recall, resulting in a higher F1-score. This confirms that the adaptive fusion strategy can effectively integrate the strengths of different models and improve overall fault diagnosis performance.

Table 2. Performance comparison of different models in terms of accuracy, precision, recall, and F1-score

Model	Accuracy	Precision	Recall	F1-score
Ensemble ANN+RF+KNN	0.9765	0.6896	0.5882	0.6349
Ensemble ANN+RF	0.9760	0.6724	0.5735	0.6190
Ensemble ANN+KNN	0.9765	0.6909	0.5588	0.6178
Ensemble ANN+DT	0.9715	0.5797	0.5882	0.5839
ANN	0.9625	0.4684	0.7647	0.5810
RF	0.9630	0.4727	0.7647	0.5842
KNN	0.9250	0.2879	0.8088	0.4247
DT	0.9545	0.4080	0.7500	0.5284

6. Discussion

The experimental results demonstrate that the proposed framework effectively improves fault diagnosis performance for electromechanical equipment. By combining ANN-based feature extraction with multiple machine learning models, the framework is able to capture complex patterns in the data and enhance classification accuracy. As shown in the results, the ANN successfully transforms the original features into a more separable representation space, which significantly benefits the downstream classifiers.

Compared with individual models, the ensemble approaches consistently achieve better performance, especially in terms of F1-score. This indicates that combining different models can provide complementary information and improve the balance between precision and recall. In particular, the final fusion model (ANN+RF+KNN) achieves the best overall performance, confirming the effectiveness of the proposed adaptive weighting strategy. The results also show that relying on a single model may lead to biased predictions, especially under imbalanced data conditions.

However, it is important to note that this study has some limitations. Firstly, the dataset used in this research is synthetic and may not fully reflect the complexity of the real industrial environment. Due to the relatively rare occurrence of machine failures and the restrictions imposed by confidentiality and production management requirements, obtaining real industrial fault data is often very difficult. Although the AI4I 2020 dataset has been widely used in predictive maintenance research, further verification on real industrial data is still needed. Secondly, although SMOTE helps alleviate the problem of class imbalance, it may introduce some synthetic samples that do not perfectly represent the real fault patterns. Thirdly, the current framework is mainly data-driven and does not explicitly integrate physical laws such as mechanics or thermodynamics, which may lead to the learned statistical correlations not fully reflecting the actual physical mechanisms. Moreover, the current framework focuses on static data and does not consider the time dependence during the operation of the equipment.

Future work can explore the use of real-world datasets and integrate time-series models such as recurrent neural networks or Transformers to further enhance performance. Additionally, integrating domain knowledge, physical constraints, and improving the interpretability of the model are crucial for practical deployment in industrial applications. Although current research provides useful theoretical foundations for intelligent fault diagnosis, additional verification under real industrial operating conditions is still necessary before large-scale practical deployment.

7. Conclusions

This paper proposes a hybrid fault diagnosis framework that integrates ANN-based feature extraction with ensemble machine learning models for electromechanical equipment. The ANN is first used to transform the original data into a more informative feature space, and multiple classifiers are then applied to improve prediction performance. A logistic regression-based fusion strategy is further introduced to combine the strengths of the best-performing models.

Experimental results show that the proposed method achieves better performance than individual models, particularly in terms of F1-score, which reflects a more balanced trade-off between precision and recall under imbalanced data conditions. The results also demonstrate that the learned feature representations play an important role in improving classification performance.

Overall, the proposed framework provides a simple yet effective solution for fault diagnosis tasks. Future work will focus on applying the method to real industrial datasets, incorporating temporal information, and improving model interpretability to support practical deployment.

References

[1] Kim T, Behdinan K. Data–driven prediction and uncertainty quantification of process–structure–property relationships for additively manufactured continuous carbon fiber reinforced polymer composites. Journal of Intelligent Manufacturing. 2026 Mar 26:1-20.

[2] Qiu Y, Hui Y, Zhao P, Wang M, Guo S, Dai B, Dou J, Bhattacharya S, Yu J. The employment of domain adaptation strategy for improving the applicability of neural network-based coke quality prediction for smart cokemaking process. Fuel. 2024 Sep 15;372:132162.

[3] Qiu Y, Hui Y, Zhao P, Dou J, Bhattacharya S, Dai B, Yu J. The interpretable deep learning-based coke quality prediction strategy aligned with domain knowledge for sustainable smart cokemaking applications. Fuel. 2025 Oct 15;398:135533.

[4] Marques G, Roque Ferreira C, Pitarma R. A system based on the internet of things for real-time particle monitoring in buildings. International journal of environmental research and public health. 2018 Apr;15(4):821.

[5] Paudel N, Neupane RC. A general architecture for a real-time monitoring system based on the internet of things. In Proceedings of the 2019 3rd international symposium on computer science and intelligent control 2019 Sep 25 (pp. 1-12).

[6] Uddin MS, Alam JB, Banu S. Real time patient monitoring system based on Internet of Things. In2017 4th International conference on Advances in Electrical Engineering (ICAEE) 2017 Sep 28 (pp. 516-521). IEEE.

[7] Yadav DK, Kaushik A, Yadav N. Predicting machine failures using machine learning and deep learning algorithms. Sustainable manufacturing and service economics. 2024 Jan 1;3:100029.

[8] Das MK, Rangarajan K. Performance monitoring and failure prediction of industrial equipments using artificial intelligence and machine learning methods: A survey. In2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) 2020 Mar 11 (pp. 595-602). IEEE.

[9] Aggarwal CC. Neural networks and deep learning. Cham: Springer; 2018 Sep.

[10] Qamar R, Zardari BA. Artificial neural networks: An overview. Mesopotamian Journal of Computer Science. 2023 Aug 2;2023:124-33.

[11] Oniésko A, Lucas P, Druzdzel MJ. Comparison of rule-based and Bayesian network approaches in medical diagnostic systems. In Conference on Artificial Intelligence in Medicine in Europe 2001 Jun 28 (pp. 283-292). Berlin, Heidelberg: Springer Berlin Heidelberg.

[12] Hossain MS, Khalid MS, Akter S, Dey S. A belief rule-based expert system to diagnose influenza. In2014 9Th international forum on strategic technology (IFOST) 2014 Oct 21 (pp. 113-116). IEEE.

[13] Gao Z, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. IEEE transactions on industrial electronics. 2015 Mar 26;62(6):3757-67.

[14] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical systems and signal processing. 2020 Apr 1;138:106587.

[15] Tang G, Pang B, Tian T, Zhou C. Fault diagnosis of rolling bearings based on improved fast spectral correlation and optimized random forest. Applied sciences. 2018 Oct 10;8(10):1859.

[16] Zhao H, Gao Y, Liu H, Li L. Fault diagnosis of wind turbine bearing based on stochastic subspace identification and multi-kernel support vector machine. Journal of Modern Power Systems and Clean Energy. 2019 Mar;7(2):350-6.

[17] Liu X, Zhou Q, Zhao J, Shen H, Xiong X. Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network. Sensors. 2019 Feb 25;19(4):972.

[18] Zhang Y, Li X, Gao L, Chen W, Li P. Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method. Measurement. 2020 Feb 1;151:107232.

[19] Sun H, Zhao S. Fault diagnosis for bearing based on 1DCNN and LSTM. Shock and Vibration. 2021;2021(1):1221462.
[20] Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. Frontiers of Computer Science. 2020 Apr;14(2):241-58.
[21] Sagi O, Rokach L. Ensemble learning: A survey. Wiley interdisciplinary reviews: data mining and knowledge discovery. 2018 Jul;8(4):e1249.
[22] Wu YC, Feng JW. Development and application of artificial neural network. Wireless Personal Communications. 2018 Sep;102(2):1645-56.
[23] Zou J, Han Y, So SS. Overview of artificial neural networks. Artificial neural networks: methods and applications. 2009 Jan:14-22.
[24] Rigatti SJ. Random forest. Journal of insurance medicine. 2017 Jan 1;47(1):31-9.
[25] De Ville B. Decision trees. Wiley Interdisciplinary Reviews: Computational Statistics. 2013 Nov;5(6):448-55.
[26] Kramer O. K-nearest neighbors. In Dimensionality reduction with unsupervised nearest neighbors 2013 Dec 27 (pp. 13-23). Berlin, Heidelberg: Springer Berlin Heidelberg.
[27] LaValley MP. Logistic regression. Circulation. 2008 May 6;117(18):2395-9.
[28] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. John Wiley & Sons; 2013 Feb 26.
[29] Matzka S. Explainable artificial intelligence for predictive maintenance applications. In2020 third international conference on artificial intelligence for industries (ai4i) 2020 Sep 21 (pp. 69-74). IEEE.
[30] Bunkhumpornpat C, Sinapiromsaran K, Lursinsap C. DBSMOTE: density-based synthetic minority over-sampling technique. Applied Intelligence. 2012 Apr;36(3):664-84.
[31] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002 Jun 1;16:321-57.