International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

ENHANCING PERFORMANCE OF BEARING FAULT DIAGNOSIS USING MACHINE LEARNING

S. R. Karthiga[1], N. Giridharprasath[2], K. Saloman[3], D. Vignesh[4]

[1]Assistant Professor, Department of Electronics and Communication Engineering, Krishnasamy College of Engineering and Technology, Cuddalore , Tamilnadu, India.

[2,3,4] UG Final Year Student, Department of Electronics and Communication Engineering, Krishnasamy College of Engineering and Technology, Cuddalore, Tamilnadu, India.

ABSTRACT

This study explores the application of machine learning techniques for intelligent machine fault diagnosis using real-time accelerometer sensor data. Leveraging an ensemble of classifiers including Multinomial Logistic Regression, Random Forest and Support Vector Machine, the research aims to enhance fault diagnosis accuracy and reliability. The use of accelerometer sensor data in real-time provides valuable insights into machine health and performance, enabling proactive maintenance and minimizing downtime. By integrating multiple machine learning algorithms, the ensemble approach offers robustness and versatility in handling diverse fault patterns and noise in sensor data. The study investigates the effectiveness of each classifier and their combination in accurately identifying and classifying machine faults in various operating conditions. Through comprehensive experimentation and validation, the research contributes to advancing intelligent fault diagnosis systems, facilitating early fault detection, and optimizing maintenance strategies in industrial and mechanical systems.

Keywords— Taxonomy, Machine learning, RF, SVM Multinomial Logistic Regression.

Introduction

The efficient and reliable operation of rotating machinery is essential in various industrial sectors such as manufacturing, transportation and energy generation. As an integral part of such machines, bearings play an important role in maintaining proper operation. Over time, however, bearings are susceptible to faults due to factors such as wear, lubrication issues, or defects, which can lead to unexpected downtime, increased maintenance costs and if not specified, it could lead to even greater failure. Traditional bearing fault detection methods are usually based on manual probing or specialized signal processing techniques. However, these techniques have limitations in dealing with real-world complexities and may not detect faults early enough to prevent catastrophic damage or downtime In recent years, there has been a growing interest in using deep learning techniques to improve the performance of bearing fault detection systems. Deep learning models, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs), offer promising solutions to automatically detect bearing errors by extracting logical patterns from sensor data unstructured directly by ingesting large amounts of sensor data, such as vibration signals or noise generation Informative- Learn Features This data-driven approach enables the search for subtle anomalies indicative of bearing faults in even at noise and complex work environments.

Literature review

Bearing Fault identification and recognition methods with help of techniques such as Machine learning to make accurate diagnosis. Here discussing the related works as following

[1] LSISMM uses an iterative least squares method to extract fault signals from vibration or thermal signals, enhancing fault detection in rotating machinery Infrared thermal micrographs focus on specific machine areas, providing thermal analysis advanced for faster fault detection and identification through high-resolution heat maps.

[2] Combines HHT and energy entropy to identify and classify multiple faults occurring simultaneously in a system, particularly effective for complex fault patterns in hoist spindle devices. The drawbacks of employing these techniques include their dependence on high-quality data, complexity, and challenges in parameter optimization and result interpretation.

[3] uses WPT (decomposes a signal into wavelet packets) and LS-SVM (regression tasks) enables the extraction of relevant features from vibration signals and the construction of an accurate wear degree prediction model.

[4] Digital twin technology uses AI techniques such as machine learning to create virtual mode of physical assets, enabling real-time simulation and optimization. Data integration and seamless collaboration facilitate the collection and analysis of disparate data for informed decision making in intelligent manufacturing

[5] CIPCABPNN method combines interval data compression with a constructed adaptive binary probabilistic neural network. It integrates compressed interval data with a neural network-based classifier for failure prediction, leveraging both interval data characteristics and neural network capabilities.

[6] In NLP, RNNs process sequential data, while CNNs excel in local feature extraction. BERT and other transformer models use the focus for context understanding, and achieve higher performance. Sequential examples aid in interpretation and summary, and are enhanced by focus. Transfer learning with pre-trained embedding models is standard for use with large datasets.

[7] uses supervised algorithms like SVM and Random Forests to classify labeled data, while unsupervised methods like clustering and anomaly detection detect subnormal behavior Deep learning models like CNN and RNN capture complex patterns in sensor data, while feature engineering extracts context. Ensemble learning enables multiple models to be combined for accuracy.

[8] uses Convolutional neural networks (CNNs) form the foundation for object recognition based on deep learning, enabling the sequential representation of visual information. Region-based convolutional neural networks (R-CNN) pioneered the approach of proposing regions of interest and applying CNNs to each. Single shot multibox detectors (SSDs) transform object recognition by predicting bounding box and class probabilities simultaneously, facilitating the development of fast-paced You Only Look Once (YOLO) models and speed Feature Pyramid Network (FPN) improves detection by connecting features of different scales.

Proposed work

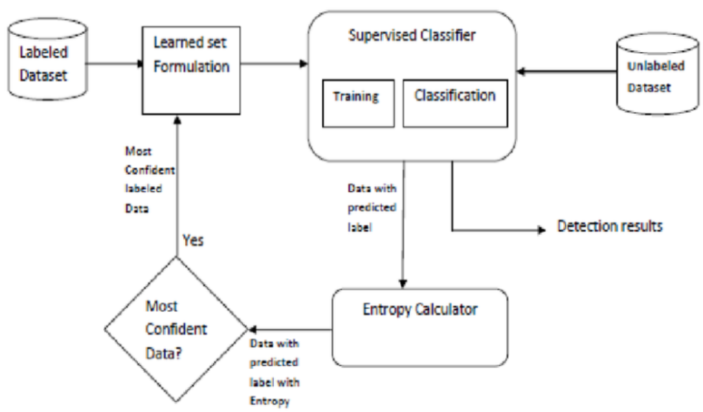

Fig. 1. Systematic Block diagram of Proposed work

The proposed system for intelligent machine fault diagnosis integrates machine learning algorithms, including Support Vector Machines (SVM), Multinomial Logistic Regression and Random Forests (RF) to enhance diagnostic accuracy and efficiency. This system aims to address the challenges of timely fault detection and classification in complex machinery by leveraging the unique strengths of each algorithm. SVM offers robust classification by identifying optimal hyper planes in high-dimensional feature spaces, enabling the accurate categorization of faults based on input data. Multinomial Logistic Regression, known for its simplicity and effectiveness in pattern recognition, supplements SVM by providing intuitive and adaptable fault detection capabilities. Additionally, Random Forests contribute to the system's accuracy through ensemble learning, aggregating predictions from multiple decision trees to improve fault diagnosis performance. By combining these machine learning techniques, the proposed system empowers industries to achieve proactive maintenance strategies, minimize downtime, and optimize operational efficiency by swiftly identifying and addressing machine faults before they escalate.

A. *Training and Testing phases*

1) *Data Collection*

This is the phase to upload various datasets as input about the faults in the bearing with its characteristics such as leaf dimensions which helps to train the model for the classification and labelled. We collected and processed the data and further examined and crosschecked the dataset against several available online source. It is ensured that the dataset consists of a comprehensive and diversified range of dataset to train the model and further divided the datasets let into training and testing.

2) *Pre-Processing*

It is the step to create a raw data to the mode. The median filter is a non-linear digital filtering techniques, frequently used to suppress noise from an image or signal.

Data Cleaning: A primary step in getting a bearing's datasets ready for machine learning model training is data cleaning. In this process, redundant datasets were eliminated, annotations were reviewed and corrected, data augmentation was used, and the dataset was split. By ensuring the enhanced quality of the data used for training, we can develop more accurate and reliable systems for detecting and classifying the characteristics of bearing faults.

Data Balancing: Developing reliable models requires balanced data sets. The algorithm is trained on a representative data set with the same samples from each class. Before training the learning model, we balance the data set. Class imbalances were addressed through undersampling and oversampling. Oversampling recreates samples from minority classes randomly to improve data set representation. Undersampling tests for majority group control by removing samples. Balanced datasets are effective in detection and classification by accurately characterizing defects.

Data Augmentation & Resize: Data augmentation uses available data to extend the training data set. This method is good for small or biased. The accuracy and robustness of our classification system increased due to the large, much more diverse data set. After data enhancement, dataset becomes free from outliers and non e values.

Annotation & Labelling: We utilised the Labelling tool to annotate, creating a bounding box to the affected part and label it, to assign the corresponding fault type.

B. *Methodology*

This project is carried over by two machine learning classifiers as following

1) Support Vector Machine

The goal of support vector machine (SVM) optimization is to find an ideal hyper plane that efficiently separates classes and maximizes the difference between them. This process is achieved by solving a quadratic programming problem with the goal of minimizing the cost function under certain constraints. The objective function deducts a penalty for misclassification and encourages the selection of hyper planes that maximize margins, which is the distance between the hyper plane and the closest data points from each class. The SVM identifies a small, known set of training data points as support vectors, which are important elements defining sophisticated hyper planes. These support vectors are close to the decision boundary and play an important role in determining the optimal hyper plane. Using support vectors, SVM can effectively classify new data points based on their proximity to the decision boundary, making it a powerful tool for a variety of classification tasks, including image recognition, text segmentation and error detection in the bearing.

Mathematical Insights into Support Vector Machines

Consider a binary classification problem with two classes labeled +1 and -1. We have a training data set of input feature vectors X and their corresponding class labels Y.

The equation for the linear hyperplane can be written as:

$$w^T x + b = 0$$

The vector W represents the normal vector to the hyperplane. that is, the direction parallel to the hyperplane. The parameter b in the equation represents the offset or distance between the hyperplane and the origin and the normal vector w.

The distance between a data point x_i and the decision boundary can be calculated as:

$$d_i = \frac{w^T + x_i + b}{\|w\|}$$

where ||w|| represents the Euclidean norm of the weight vector w. Euclidean norm of the normal vector W

For Linear SVM classifier :

$$\hat{y} = \{ 1 \quad : \quad w^T x + b \ \geq 0 \ 0 \quad : \quad w^T x + b \ < 0$$

Optimization:

- **For Hard margin linear SVM classifier:**

$$minimize\ \frac{1}{2} w^T w = minimize\ \frac{1}{2}\|w\|^2$$

$$subject\ to\ y_i(w^T\ x + b) \geq 1\ for\ i = 1,2\ 3, \ldots, m$$

The target variable or label of the ith training sample is identified by the symbol header in this case. Here, ti = -1 for negative cases (when yi = 0) and ti = 1 for positive cases (when yi = 1), respectively. We need a decision boundary that satisfies the constraint:

- **For Soft margin linear SVM classifier:**

$$minimize\ \frac{1}{2} w^T w + C \sum_{i=1}^{m} \zeta_i$$

$$subject\ to\ y_i(w^T\ x + b) \geq 1 - \zeta_i\ and\ \zeta_i \geq 0$$

$$for\ i = 1,2\ 3, \ldots, m$$

Dual Problem: A dual-problem optimization problem that requires finding the Lagrange multipliers associated with the support vector can be used to solve the SVM. The optimal Lagrange factor α(i) that maximizes the following two objective functions.

$$minimize: \frac{1}{2}\sum_{i\to m}\sum_{j\to m}\alpha_i\alpha_j t_i t_j K(x_i.x_j) - \sum_{i\to m}\alpha_i$$

where,

- αi is the Lagrange multiplier associated with the ith training sample.
- K(xi, xj) is a kernel function that computes the similarity between observations of xi and xj. It enables SVM to solve nonlinear classification problems by using high-dimensional models.
- The term ∑αi represents the sum of all Lagrange coefficients.

The SVM decision boundary can be defined in terms of these Lagrange optimal parameters and support vectors. Once the two problems are solved and the Lagrange optimal parameters are found, training samples with i > 0 are support vectors, whereas the decision boundary is given by:

$$w = \sum_{i\to m}\alpha_i t_i\, K(x_i.x) + b$$

$$t_i(w^T x + b) = 1 \leftrightarrow b = w^T x - t_i$$

2) Multinomial Logistic Regression

Logistic regression is a distributional model. This is designed for data sets with numeric input variables and categorical target variables with two values or two classes. Such problems are called binary classification problems. Logistic regression is designed for binomial problems; the objective is modeled using a binomial probability distribution function. Learning scores map to 1 for a positive class or outcome and 0 for a negative class or outcome. The fit model predicts the probability that Example Class 1. By default, logistic regression cannot be used for classification tasks with more than two class labels, the so-called multiclass classification.Instead, multi-classification problems.

One of the popular methods for optimizing logistic regression in multiclass classification problems is to divide the multiclass classification problem into multiple binary classification problems and fit a standard logistic regression model for each subproblem. Another approach is to modify the logistic regression model to directly support the prediction of multiclass labels. In particular, to state the probability that the input instance belongs to any known class label.

A probability distribution describing multiclass probabilities is called a multinomial probability distribution. A logistic regression model designed to detect and predict multinomial probability distributions is called multinomial logistic regression. Similarly, the default or standard logistic regression can be called a binomial logistic regression.

Binomial logistic regression: Standard logistic regression that predicts the binomial probability (i.e., for both classes) for each input instance.
Multinomial logistic regression: A modified form of logistic regression that predicts a multinomial probability (i.e., more than two subjects) for each input instance.

3) Random Forest

Random forest is a popular machine learning algorithm that includes supervised learning methods. It can be used for classification and regression problems in ML. It is based on the concept of cluster learning, the process of combining multiple classifiers to solve a complex problem and improve modeling performance, as the name implies. "A random forest is a classification of a number of classified decision tree. It takes predictions from each tree and determines the final result based on the majority of the predictions' votes. The larger the number of trees in the forest, the greater the accuracy, and the difficulty of constraint excess is eliminated
.Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.
Steps in Random Forest method,
Step-1: Select random K data points from the training set.
Step-2:,Create decision trees associated with selected data points (subsets).
Step-3: Select the number N for the decision trees you want to create.
Step-4: Repeat Steps 1 and 2 again.
Step-5: Find the prediction of each decision tree for the new data points, win the majority, and assign the new data points to that category.

By integrating these model outputs, the proposed system aims to improve fault detection and reduce false alarms, thereby speeding up maintenance and reducing execution time under industrial conditions.

Result &Analysis

In this chapter, real time datasets, content extraction and classification methods were used in the process. which can help in evaluating the performance using accuracy metrics. The accuracy metric is tested as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

The proposed algorithm provide improved accuracy rate than the other machine learning algorithms.

Algorithm	Accuracy (%)
SVM	94
MULTINOMIAL LOGISTIC REGRESSION	94
RF	96

Table 1: Accuracy Percentage level of different Algorithms

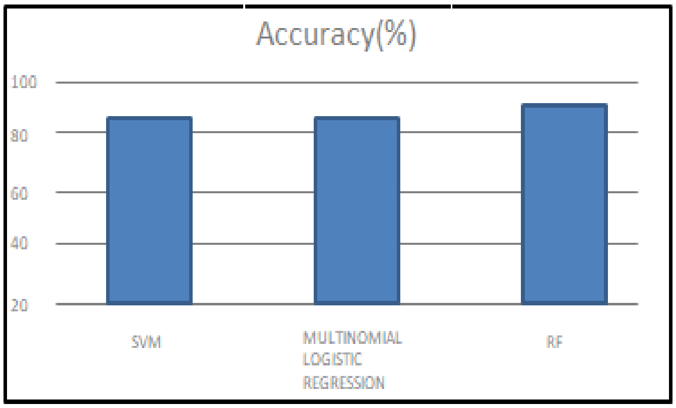

Fig 2. Performance report

Conclusion & Future work

In conclusion, the utilization of an ensemble of machine learning models, including Random Forest, SVM and Multinomial Logistic Regression represents a significant advancement in intelligent machine fault diagnosis. Through this integrated approach, we have demonstrated the capability to accurately detect and diagnose faults in industrial machinery, enabling proactive maintenance and minimizing downtime. The ensemble of machine learning models offers several key advantages. By combining these models, we enhance fault diagnosis accuracy and robustness, particularly in the presence of diverse fault patterns and noisy sensor data. Furthermore, the real-time nature of the system enables timely detection and response to emerging faults, preventing costly equipment failures and optimizing operational efficiency. Looking ahead, further research and development in this field could focus on refining the ensemble learning approach, optimizing model parameters, and exploring the integration of additional sensor modalities for enhanced fault diagnosis capabilities. Additionally, efforts to deploy and validate these systems in real-world industrial environments will be critical to demonstrating their effectiveness and scalability.

REFERENCES:

1. X. Li, H. Shao, S. Lu, J. Xiang, And B. Cai, ``Highly Efficient Fault Diagnosis of Rotating Machinery Under Time-Varying Speeds Using Lstm and Small Infrared Thermal Images," IEEE Trans. Syst., Man, Cybern. Syst., Early Access, May. 14, 2022, Doi: 10.1109/Tsmc.2022.3151185.
2. J. Gu, Y. Peng, H. Lu, B. Cao, And G. Chen, ``Compound Fault Diagnosis and Identification of Hoist Spindle Device Based on Hilbert Huang and Energy Entropy," J. Mech. Sci. Technol., Vol. 35, No. 10, Pp. 42814290,Oct. 2021.
3. H. Wang And L. Chan, ``Wavelet Packet Transform-Assisted Least Squares support Vector Machine for Gear Wear Degree Diagnosis," Math. Problems eng., Vol. 2021, Pp. 19, Sep. 2021.
4. Z. Huang, Y. Shen, J. Li, M. Fey, And C. Brecher, ``A Survey On AI-Driven digital Twins in Industry 4.0: Smart Manufacturing and Advanced Robotics, "Sensors, Vol. 21, No. 19, P. 6340, Sep. 2021.
5. L. Yang, G. Jia, F. Wei, W. Chang, C. Li, And S. Zhou, ``The CIPCA-BPNN Failure Prediction Method Based on Interval Data Compression and dimension Reduction," Appl. Sci., Vol. 11, No. 8, P. 3448, Apr. 2021.
6. D. W. Otter, J. R. Medina, And J. K. Kalita, ``A Survey of the Usages of deep Learning for Natural Language Processing," IEEE Trans. Neural Netw.Learn. Syst., Vol. 32, No. 2, Pp. 604624, Feb. 2021.

7. N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, And X. Ding,``Embracing Imperfect Datasets: A Review of Deep Learning Solutions for medical Image Segmentation,'' Med. Image Anal., Vol. 63, Jul. 2020,Art. No. 101693.
8. Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, And A. K. Nandi, ``Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap, ''Mech. Syst. Signal Process., Vol. 138, Apr. 2020, Art. No. 106587.
9. L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, And. Martikainen, ``Deep Learning For Generic Object Detection: A Survey, ''Int. J. Comput. Vis., Vol. 128, No. 2, Pp. 2613 18, Jan. 2020.
10. P. Li, D. Wang, L. Wang, And H. Lu, ``Deep Visual Tracking: Review and Experimental Comparison,'' Pattern Recognit., Vol. 76, Pp. 323338,Apr. 2018.