Vol. 06, No. 1 (2024) 251-260, doi: 10.24874/PES.SI.24.02.008

Proceedings on Engineering Sciences

www.pesjournal.net

FAULT DETECTION AND CATEGORIZATION USING AN ADVANCED MACHINE LEARNING TECHNIQUE FOR INDUSTRIAL ROTATIONAL MACHINERY

Divya Paikaray[1]
Naveen Kumar Rajendran
Vaishali Singh
Pulkit Srivastava

Received 03.11.2023.
Received in revised form 28.12.2024.
Accepted 06.01.2024.
UDC – 004.052.42

Keywords:

Industrial rotating machinery, Fault identification, Categorization, Industrial rotation, Dandelion Optimized CatBoost (DO-CB).

A B S T R A C T

The difficulty of fault identification as well as categorization in industrial rotating machinery is fixed by this study, which introduces a revolutionary Dandelion Optimized CatBoost (DO-CB) technique. The suggested framework makes use of the CB algorithm, which is enhanced by the DO method. The first step in the suggested DO-CB approach is gathering sensor data from rotating gear to record different operational settings. To ensure robustness, the recommended approach is developed on identified data and includes a variety of fault scenarios. Additionally, the Python tool used for identifying faults and classification is the basis for the implementation of the DO-CB approach. The experimental findings show how well the suggested method works to precisely identify and classify problems in industrial rotating gear. In comparison to benchmark defect detection techniques, the suggested DO-CB approach performs better, demonstrating its capacity to manage intricate patterns and fluctuations in the data.

1. INTRODUCTION

A crucial aspect of predictive maintenance and industrial automation involves the recognizing and categorizing defects in rotating gears used in various industries. The maintenance and dependability of rotating machinery become increasingly critical as companies expand and rely more on complex equipment for smooth operations. This field of study addresses challenges in locating and classifying issues in equipment, including generators, turbines and motors integral components of numerous industrial processes (Li, X., et al., 2020). Fault detection is crucial for rotating machinery as it can prevent downtime, avert catastrophic failures and enhance overall productivity. Failures in rotating machinery can have extensive consequences in the complex realm of industrial operations, leading to compromised safety, increased maintenance costs and production delays.

[1] Corresponding author: Divya Paikaray
Email: divya.p@arkajainuniversity.ac.in

Consequently, for industries to achieve operational excellence, it is imperative to develop and implement new defect detection and classification systems (Li, Z., et al., 2019). Early identification of errors before they escalate is a fundamental objective in this field. Addressing issues in their initial stages not only mitigates the financial impact of unplanned breakdowns but also extends the lifespan of machinery (Surendran, R., et al., 2022). The landscape of defect detection has undergone significant transformation with the advent of sensor technology and the Internet of Things (IoT). Live surveillance and collecting of data through sensors integrated into rotating machinery enable continuous evaluation of equipment health, facilitating proactive maintenance techniques (Zhang, W., et al., 2019). Numerous signals and data associated with rotating machinery undergo analysis during the defect detection process.

Classifying defects in rotating equipment is an equally vital undertaking that goes beyond simple identification. The process involves categorizing the characteristics and intensity of defects to enable focused reactions (Senanayaka, J.S.L., et al., 2017). Fault classification empowers maintenance teams to prioritize issues based on their potential impact on machinery performance and overall operational integrity. Organizations can manage resources by distinguishing between minor imbalances and significant structural breakdowns. This enables them to address the most urgent issues promptly while preparing for long-term maintenance and improvements (Chen, S., et al., 2020). Fault detection is the systematic monitoring of machinery to identify any abnormalities or deviations from the expected level of performance. Within the domain of industrial rotating equipment, these anomalies can manifest as vibrations, unusual noises, changes in temperature, or fluctuations in other measurable factors. Prompt identification of these irregularities is crucial to prevent minor issues from evolving into significant malfunctions, which can result in costly downtime and maintenance. Furthermore, proactive defect detection aligns with the primary goal of predictive maintenance, enabling organizations to plan repairs and replacements proactively, thereby minimizing disruptions to production operations.

Identifying and classifying defects is an additional aspect of this discipline that enhances the precision of maintenance procedures. Recognizing that not every mistake has identical ramifications or necessitates identical remedies is crucial. By classifying defects based on their nature, severity and potential impact, industrial operators can prioritize maintenance operations and allocate resources effectively (Xia, M., et al., (2017)). For example, distinguishing between misalignment, imbalance and bearing wear in a rotating machine allows for targeted treatments that optimize the use of time and resources.

The process of integrating fault detection and classification systems into the broader industrial environment is intricate. It involves seamlessly incorporating analytics platforms, data collection systems and sensors with existing supervisory control and data acquisition (SCADA) systems (Choudhary, A., et al., 2021). Ensuring interoperability with enterprise asset management (EAM) systems is crucial in transforming the knowledge derived from problem detection into actionable maintenance plans. This contributes to the development of a comprehensive industrial asset management strategy (Li, Y., et al., 2020). Identifying and classifying problems in industrial rotating gear presents several challenges and constraints. The complexity arises from the diversity of machinery types and operating conditions, as different machines can exhibit distinct defect signs. Fault identification becomes even more intricate due to inherent fluctuations in operational factors, such as speed, load and climatic variables. Limitations arise from the availability and quality of sensor data, as sparse or insufficient data can compromise the accuracy of defect detection systems.

Souza, R.M., et al., (2021) investigated unsupervised data, which was straightforward to gather. It employed a three-stage training method involving representation grouping and enhanced supervised learning. The validation was conducted on two datasets related to rotating devices and the results indicated that the proposed approach demonstrated promising diagnostic performance. The utility of the suggested method in addressing unsupervised learning fault diagnostic problems was thus confirmed.

Tang, S., et al., (2019) presented the utilization of a "predictive maintenance model called PdM-CNN ("Convolutional Neural Network"). The model utilized data from a single vibration sensor placed on the motor-drive end bearing, a configuration found in the industry. They demonstrated the model's capability to detect and categorize defects in industrial rotating gear.

Souza, R.M., et al., (2021) proposed intelligent defect detection based on deep learning (DL), which has captured the interest of scholars. DL offered automated feature learning and fault categorization, prompting an examination of DL and DL-based intelligent fault diagnostic systems. The techniques of DL-based fault diagnostics for rotating machinery, particularly bearings, gearboxes and pumps, were outlined and explored. Cross-validation findings indicated that the suggested approach exhibited high diagnostic accuracy. Ding, A., et al., (2019) introduced deep reinforcement learning as an intelligent diagnosis technique aimed at addressing the limitations of discussed diagnostic approaches. The proposed process underwent validation using datasets from two types of rotating machinery: hydraulic pumps and rolling bearings. These datasets consisted of numerous raw vibration signals recorded under various operating conditions and health states. The suggested approach demonstrated its capability to yield outstanding outcomes.

Brito, L.C., et al., (2022) proposed the Shapley Additive Explanations (SHAP) method for diagnosing and detecting problems in spinning equipment. The approach consisted of three components: defect diagnosis, defect detection and feature extraction (Zhang, Y., et al., (2021)). Anomaly detection methods were employed to confirm the presence of faults independently. The models demonstrated superior outcomes in terms of defect diagnosis and detection. Sobie, C., et al., (2018) investigated issues related to race roller bearings defects by producing data for training from superior models of roller bearing movements. The data was employed to train machine learning (ML) algorithms, which were tested against four experimental datasets. Various ML strategies were evaluated, ranging from established statistics-based feature algorithms to CNN. The technique outperformed a statistical feature-based classifier in terms of classification accuracy.

Luwei, K.C., et al., (2018) proposed an artificial neural network (ANN) classification method that could enable autonomous fault classification in rotating machinery. Vibration-based condition monitoring (VCM) techniques were highly advanced, as they could assign unique fault diagnostic signatures to specific rotating machine faults. VCM approaches were widely employed in practice for the detection and categorization of issues in rotating machines. The integrated approach of techniques was suggested; due to its adaptability to new features and flaws. Li, Y., et al., (2020) presented a DL algorithm for detecting faults in rotating equipment. Acknowledging the challenge of obtaining labeled data in real-world sectors, the research presented data augmentation approaches to synthesize more samples suitable for training the model. The suggested technique demonstrated high diagnostic accuracy even with a limited original training dataset, showcasing outstanding performance.

Carino, J.A., et al., (2018) presented ensemble technique for detecting and changing classification. The approach was divided into four major stages: feature evaluation and reduction, recognition of unexpected events, diagnostics with e-Class Evolving Classifiers and training for model improvement. The proposed approach for identifying faults was evaluated and compared to a traditional method for identifying faults. The models' precision was monitored separately.

Wu, C., et al., (2019) discussed the development of a CNN that could learn characteristics directly from the original vibration signals and identify defects. The suggested method's efficacy was proven using data from the Prognostics and Health Management (PHM) gear challenge as well as testing on a planetary gearbox.

Chen, Z., et al., (2019) introduced a “Transferable Convolutional Neural Network (TCNN)” to enhance target task learning. Leveraging extensive source task datasets, a single-dimensional CNN was constructed and trained. The proposed solution capitalized on the learning capabilities of a DL Model but also incorporated past information from the source task. The approach was demonstrated to be stable and resilient, leading to superior outcomes. Luo et al., (2018), proposed a Computer Numerical Control (CNC) technique for early defect identification under time-varying circumstances. A DL model was constructed to select the impulse responses from the vibration that automatically signals over a lengthy 288-day period. The findings demonstrated the method's ability to reflect the machine tool's state of condition accurately.

Li, Z., et al., (2019) presented a DL-driven approach for degradation assessment and fault categorization was presented. With multiple hidden layers, a Deep Neural Network (DNN) learned multiple nonlinear transformations with high complexity compared to traditional data-driven methods, enabling the identification of discriminative information and capturing significant variations from industrial data. Based on the data gleaned from vibration signals in mechanical equipment, the numerical results demonstrated that used approaches were capable of classifying failures.

Dineva, A., et al., (2019) proposed a unique multi-label classification technique called k-nearest neighbors (KNN) intended to diagnose several faults and assess the severity of each problem in the presence of noise. Comparisons were made between the performance of several multi-label classification models and with most of the vibration data labeled, the prediction accuracy was quite good.

Kolar, D., et al., (2020) introduced a model based on CNN for defect diagnostics in rotary equipment using DL techniques. By employing the CNN-trained model, the system could classify data into one of the four classes, enabling it to function and deliver precise diagnostic outcomes.

González-Muñiz, A., et al., (2020) introduced a method for monitoring the state and detecting faults in rotating machinery using a one-dimensional deep convolutional neural network (1D DCNN). The developed system was intended to use on a rotating apparatus with a total of seven potential operational conditions. The findings suggested that their Convolutional Neural Network (CNN) model had a level of accuracy that was comparable to that of traditional classifiers.

Gong, W., et al., (2019) introduced a new approach called the enhanced “convolutional neural network-support vector machine (CNN-SVM)” technique. The approach improved the conventional CNN architecture by incorporating global average pooling and SVM. The findings validated the superiority of the suggested technique over other established intelligent methods, including SVM, KNN, “Back-propagation artificial neural network (BPNN)”, “Deep back-propagation network (DBPN)” and classic CNN.

Khan, M.A., et al., (2022) proposed fault identification and diagnosis (FDD) methods for anticipating a variety of bearing problems that occurs in rotating machinery. Traditional procedures, statistical techniques and artificial intelligence-based architectures, such as ML and DL, were explored for the purpose of identifying defects in rotating electrical machines. Furthermore, external resources like DL algorithms and variable frequency devices (VFD) were suggested to obtain precise outcomes. Li, C., et al., (2016) introduced deep statistical feature learning (DSFL) as a method for analyzing vibration data in rotating machinery. The proposed methodology was utilized as a sophisticated statistical technique for acquiring profound knowledge of features in the gearbox and bearing systems. The suggested strategy is evaluated in comparison to conventional approaches such as support vector machines and Gaussian-Bernoulli restricted Boltzmann machines (GRBM). The results indicated that deep statistical feature learning had higher classification accuracy for fault patterns compared to other models.

Mehta, A., et al., (2021) discussed the application of Infrared Thermography (IRT) in diagnosing bearing faults. The thermal picture underwent a “two-dimensional discrete wavelet transform (2D-DWT)” for breakdown. “Principal Component Analysis (PCA)” was used to decrease the dimensionality of characteristics that are extracted. The appropriate selection of these parameters led to improved outcomes.

Wu, C., et al., (2021) presented a hybrid classification auto encoder (HCAE) model that employed a softmax classifier to diagnose the health state using the encoded characteristics obtained from the autoencoder. The validity of the suggested approach was confirmed by applying it to a dataset of motor bearings and an industrial hydro turbine dataset. The empirical findings demonstrated that their system could achieve significantly high diagnostic accuracies with a minimal proportion of labeled data.

In this paper, we introduce a novel, Dandelion Optimized CatBoost to improve the reliability as well as efficiency of fault detection devices, contributing to the overall operational stability and maintenance optimization of industrial rotational machinery.

1.1 Contribution

- The collected CWRU dataset is a great asset for scholars and professionals engaged in the development and evaluation of algorithms for the early identification of malfunctions in rotating machinery.
- The Dandelion Optimisation approach improves the process of discovering as well as classifying faults, enabling prompt and targeted maintenance actions.
- The CatBoost algorithm is designed to handle category data and achieve excellent performance in machine learning tasks. It enables the effective analysis of intricate data patterns in machinery operations.
- The DO-CB model identifies and categorizes defects in rotating machinery, providing a more dependable and effective solution.

The remaining part of this article is categorised into the subsequent sections: Part 2, Methodology; Part 3, Result and Part 4, Conclusion.

2. RESEARCH METHODS

In this paper, we collected a CWRU dataset that is used to identify and classify different types of defects in industrial spinning equipment. The Dandelion Optimized CatBoost (DO-CB) strategy is presented as an innovative and potentially more effective technique. This approach enhances the predictive maintenance capabilities of rotational machinery in industrial settings, thereby minimizing downtime, reducing maintenance costs and improving overall operational efficiency.

2.1 Dataset

The CWRU dataset is a comprehensive and typical collection of bearing problems. Its purpose is to verify and enhance motor state assessment methods. An experimental setup was conducted using a 2-horsepower Reliance motor. The facility was operated at four different loads throughout the experiments: 0 hp, 1 hp, 2 hp and 3 hp. The sampling frequency utilized was 12 kHz. The CWRU dataset's enduring fault class comprises “inner race defects, ball defects and outer race defects." Every fault class is characterized by two distinct diameters, namely 8 mils and 16 mils. Hence, the CWRU dataset is categorized into seven classes, consisting of one normal class and six fault classes, each representing two different fault diameters for a total of three fault classes. Without rearranging the information, we utilized 51,200 points of sampling to represent distinct bearing faults. These points were divided into 100 samples, with each sample including 512 sampling points for training purposes. The rest of the 12,800 points of sampling were used for testing, divided into 25 pieces from each load. There were 400 different samples allocated 100 samples for training, allocated for testing in each fault class. Table 1 contains specific information about the dataset (Wei, H., et al., (2021)).

Table 1. The CWRU dataset fault classifications.

Class	*payload*	*Defect Diameter*	*Error Situation*	*Training*	*Testing*
0	0	Not Applicable	Operating Normally	400	100
1	1	10	Misaligned shaft	400	100
2	2	20	inequality	400	100
3	3	30	Equipment Erosion	400	100
4	1	15	alignment deviation	400	100
5	2	25	Disparity	400	100
6	3	35	Equipment Erosion	400	100

2.2 Synergistic Fault Identification

Dandelion Optimization, stimulated by the dispersal pattern of dandelion seeds, is employed in tandem with the CB algorithm for efficient fault identification and classification in industrial rotating machinery. This synergistic approach combines the robust gradient-boosting capabilities of CB with the stochastic search methodology of Dandelion Optimization.

2.2.1 Dandelion Optimization

Dandelion Optimization can enhance the efficiency of feature selection and model parameter tuning, enabling the identification of abnormalities in rotational machinery. To investigate the behavior of dandelion seeds, a unique swarm intelligence system was developed and suggested as a solution for ongoing optimization issues. There are two subpopulations of dandelions in Dandelion Algorithm (DA) that can be seeded and those cannot. Distinct seeding techniques are employed for each subgroup. Meanwhile, conducting a subpopulation suitable for sowing is another method of seeding to prevent slipping into the local optimum. A dandelion scatters its seeds indiscriminately. One way to conceptualize dandelion seeding is as a search for the optimal in acertain vicinity around a point. The three steps involved in this process are as follows:

When seeds are in the rising stage, they can locally float in communities or ascend spirally, influenced by eddies from above, depending on the weather. In space, as they descend during the falling stage, flying seeds alter their course. Seeds are planted in selected spots to facilitate sprouting during the landing stage as shown in Equation (1).

$$C_j = rand \times (V - K) + K \tag{1}$$

Here j represents an integer between 1 and the size of the population. dim Represents the dimension of the dandelion vector, $rand$ denotes a pseudo-random value in the range of 0 to 1 and K and V are defined as follows in Equation (2-4):

$$K = [1_1, \dots, 1_{dim}], V = [v_1, \dots, u_{dim}] \tag{2}$$

$$f_{best} = \left(f(C_j)\right) \tag{3}$$

$$C_{elite} = C_j \tag{4}$$

During the dispersion stage, dandelion seeds must attain a specific height to detach from their parent plant. The wind's direction and the air's humidity influence the height they reach. Climate and wind speed are the primary factors affecting dandelion seed distribution. Wind speed determines the range a seed can travel. Weather conditions impact dandelion growth in nearby and distant areas, determining whether seeds can disperse. Two types of weather are relevant here:

Category 1: wind speed distributions are considered as lognormal distribution $Y \sim N(u, \sigma^2)$. Due to the even distribution of random values along the Y-axis, there's a higher chance for dandelion seeds to germinate and spread to distant places. In this case, exploration is prioritized by DO. Dandelion seeds disperse in various ways around the search area via the wind. Wind speed influences the height a dandelion seed attains. Stronger winds at higher altitudes result in seeds that are spread over a greater distance. The vortexes above the dandelion seeds continually shift due to wind speed, causing the seeds to ascend in a spiral pattern, as shown as follows in Equation (5-9):

$$C_{s+1} = C_s + \alpha * u_w * u_z * ln\, ln\, Y * (C_t - C_s) \tag{5}$$

$$C_t = rand\ (1, Dim) * (V - K) + K \tag{6}$$

$$ln\, ln\, Y = \{\frac{1}{z\sqrt{2\pi}} exp\, exp\left[-\frac{1}{2\sigma^2}(lnz)^2\right] \quad z \geq 0\ \ 0 \qquad z < 0 \tag{7}$$

$$\alpha = rand\ () * \left(\frac{1}{S^2}s^2 - \frac{2}{S}s + 1\right) \tag{8}$$

$q = \frac{1}{f^{\theta}};\ \ u_w = q * cos\ cos\ \theta; u_z = q * sin\ sin\ \theta$
(9)

Where, θ falls within the range of [−π, π]. Category 2: On a wet day, dandelion seeds cannot rise in alignment with the wind because of things like humidity and air resistance. In the specific context where dandelion seeds are utilized, the corresponding mathematical Equations (10-11),

$$C_{s+1} = C_s * l \tag{10}$$

$$l = 1 - rand\ (\,) * r \tag{11}$$

$r = rand\ (\,) * \left(\frac{1}{S^2-2S+1}s^2 - \frac{1}{S^2-2S+1}s + 1 + \frac{1}{S^2-2S+1}\right)$
(12)

The mathematical equation for the ascent of dandelion seeds is as follows in Equation (13)

$$W_{s+1} = \{C_s + \alpha * u_w * u_z * ln\ ln\ \Upsilon\ * (C_t - C_s)\ \ if\ rand\ n < 1.5\ C_s * l\ \ else \tag{13}$$

Following the ascending stage, the average position data depicts the descent of the dandelion, promoting population growth and the emergence of potential communities. The corresponding mathematical formula is presented below in Equation (14)

$C_{s+1} = C_s - \alpha * \beta_s * (C_{mean-s} - \alpha * \beta_s * C_s)$
(14)

The mathematical formulation for the average location of the population in the jth iteration represented byC_{mean-s}, is as follows in Equation (15)

$$C_{mean-s} = \frac{1}{pop}\sum_{j=1}^{pop} C_j \tag{15}$$

This part of the DO technique focuses on exploitation. The dandelion seed selects its landing location randomly, drawing from its experiences in the preceding two phases. As the number of iterations increases, the algorithm aims to identify the best option. The optimal placement is the approximate location where dandelion seeds are most likely to germinate successfully. To converge precisely to the global optimum, search agents in their local communities leverage the specialist knowledge of the existing elite. The evolution of the population will eventually unveil the ideal global solution. This behavior manifests in the following ways in Equation (16)

$C_{s+1} = C_{elite} + levy\ (\delta) * \alpha * (D_{elite} - C_s * \partial)$
(16)

In the fifth iteration, the dandelion seed was planted in the optimal spot, symbolized by D_{elite}. Levy(δ), representing the Levy flight function, is computed as follows in Equation (17-19)

$$levy\ (\delta) = t * \frac{x*\sigma}{s^{\frac{1}{\beta}}} \tag{17}$$

A random number between 0 and 2 is denoted by β and in this work, β is set to 1.25. The variables x and s are random numbers within the range [0, 1], Whereas s is a constant value at 0.01. The mathematical equation for σ is,

$$\sigma = \left(\frac{\Gamma(1+\beta)*sinsin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right)*\beta*2^{((\beta-12))}}\right) \tag{18}$$

$$\partial = \frac{2s}{S} \tag{19}$$

The algorithm specifies the upper limit for the amount of repetitions to terminate the optimization procedure.

2.2.1 2.2.2 CatBoost algorithm

CatBoost is capable of processing intricate data from sensors on machines, making it possible to see minute anomalies that might be signs of possible problems. A novel ensemble technique utilizing gradient boosting with decision trees is called CB. CatBoost's standard model utilizes a symmetric tree, an application of increasing. The sequential iteration of a set of classifiers produces a robust classifier. In comparison to conventional boosting approaches, CB employs an enhanced algorithm for computing leaf values and it utilizes a unique method to handle categorical information. CB leverages integrated category features, enabling it to exploit feature linkages and enhance the feature dimension. To reduce the present iteration's function of loss, CatBoost'sl iteration seeks to identify g_l that can be computed using equation (20).

$$g_l - arg\ arg\ \frac{1}{n}\ \sum_{l=1}^{n}\ \left(-e_l(w_l, z_l) - g(w_l)\right)^2 \tag{20}$$

Where $e_l(w,z) = \frac{\partial K(z, E_{l-1}(w))}{\partial E_{l-1}(w)}$is the estimate of the gradient.

Where $K\big(z, E_{l-1}(w)\big)$is the loss function and the current learner created by the finished k-1 step iteration$E_{l-1}(w)$, the process involves sorting the samples from the initial collection of pieces based on an unsystematic arrangement, denoted as $\sigma = \{\sigma_0, \sigma_1, \dots, \sigma_m\}$. Then, m distinct support models, denoted as M_1, M_2 ... M_n is initialized to achieve an impartial estimation of the gradient using the ordered boosting approach. For each sample $x_i$obtained from a training set without replacement, a unique model needs to be trained. The gradient estimate concerning sample x_i is obtained using M_i and the basic learner is trained using this gradient estimate. Equation (21) illustrates the final result, representing the strong learner of this iteration.

$$E_l(w) = E_{l-1}(w) + g_l \tag{21}$$

Comparatively speaking, the objective of the l^{th} iteration is to determineg_l, following which the aim is to minimize and decrease the training data's objective function. This entails reducing the model's inaccuracy of prediction in the training set to obtain the CB system.

2.2.3 Dandelion Optimized CatBoost (DO-CB)

An innovative method for defect identification and classification in industrial rotating machinery involves the combination of Dandelion Optimization and CatBoost. Taking inspiration from the natural foraging behaviour of dandelions, Dandelion Optimization optimizes model parameters by improving the algorithm's exploration-exploitation balance. This hybrid model is further enhanced by CB, a robust gradient boosting method and performs well in handling intricate connections and categorical characteristics in data related to industrial machinery. The integration of these methods improves fault detection efficiency along with accuracy, enabling the swift identification and classification of issues in rotating machinery. This creative combination provides a robust approach for preserving the reliability and performance of industrial systems. Algorithm 1 shows the pseudo code for DO-CB.

Algorithm 1: Pseudocode for Dandelion Optimized CatBoost

```
Initialize Dandelion Optimization parameters
Initialize CatBoost hyperparameters
while not termination_condition:
    Update Dandelion positions based on fitness
    Apply CatBoost training on Dandelion positions
    Explore neighborhood and exchange information
    Apply local search and update positions
    Update Dandelion wind parameters
    Evaluate CatBoost performance and fitness
    Maintain diversity in Dandelion population
end while
```

3. RESULTS AND DISCUSSION

The recommended task was executed with "four RTX 2080Ti graphics cards and PyTorch 1.5.1" was employed for software. These datasets served as verification to assess the models' functionality.

The performance evaluation of the proposed approach involves assessing it in terms of Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) as well as Mean Absolute Error (MAE) and conducting a comparative analysis with other existing methods, including Deep Belief Network (DBN) (Shao, H., et al., 2018) and SVR (Shao,H., et al., 2018).

The loss metric is used to quantify prediction errors during the training and validation phases of fault detection and categorization for industrial rotational machinery. The goal is to minimize the disparity between predicted and actual values. The training graph monitors these metrics across epochs, aiming for high accuracy and low loss. Figure 1 depicts the loss validation of the proposed method.

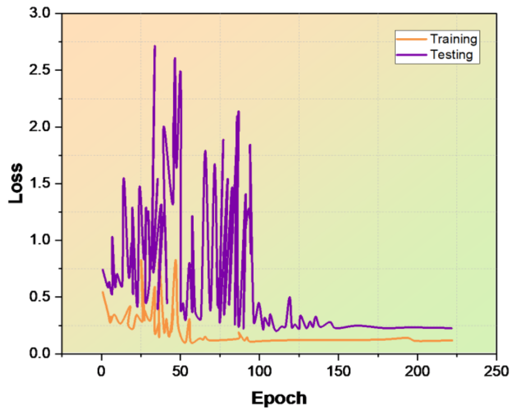

Figure 1. DO-CB loss validation on the CWRU dataset during training

RMSE quantifies the average amount of discrepancies between expected and actual data. The calculation involves finding the square root of the average of the squared differences between the expected and actual values. This metric is valuable for assessing the accuracy of a model in detecting various faults in rotational machinery. Figure 2 and Table 2 depict the comparative evaluation of RMSE in suggested and traditional methods. When compared to existing methods such as DBN and SVR, which have RMSE values of 0.015 as well as 0.016, respectively, the suggested DO-CB achieves an RMSE value of 0.011. Our suggested approach produced better results for robustness identification and incorporates a range of errors. The RMSE equation (22) is as follows,

$$RMSE = \sqrt{\frac{1}{m}\sum_{j=1}^{m} \left(z_j - \hat{z}_j\right)^2} \tag{22}$$

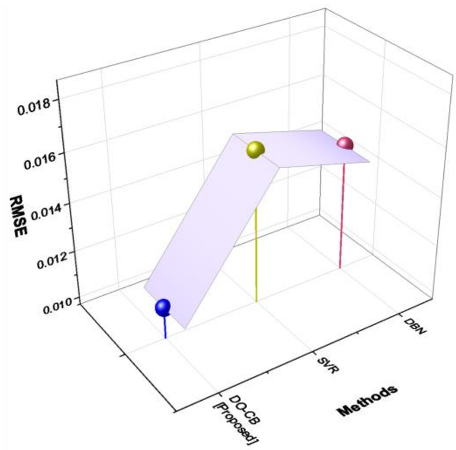

Figure 2. Result of RMSE

Table 2. Result of RMSE

Methods	RMSE
DBN	0.015
SVR	0.016
DO-CB [Proposed]	0.011

MAPE is a statistic expressed as a percentage that calculates the average absolute percentage difference between expected and actual values. It offers a comparative assessment of the precision of a model in identifying different malfunctions in rotating machinery. The comparative evaluation of MAPE is shown in Figure 3 and Table 3. When compared to the currently existing methods, such as DBN along with SVR, which have MAPE values of 0.9283 and 1.0323, respectively, the suggested DO-CB has a MAPE value of 0.7546. The proposed methodology demonstrates superiority over the existing methods for identifying data to ensure robustness and includes a variety of faults. The MAPE equation (23) is as follows,

$$MAPE = \frac{1}{m}\sum_{j=1}^{m} \left|\frac{z_j - \hat{z}_j}{z_j}\right| \times 100\ \% \tag{23}$$

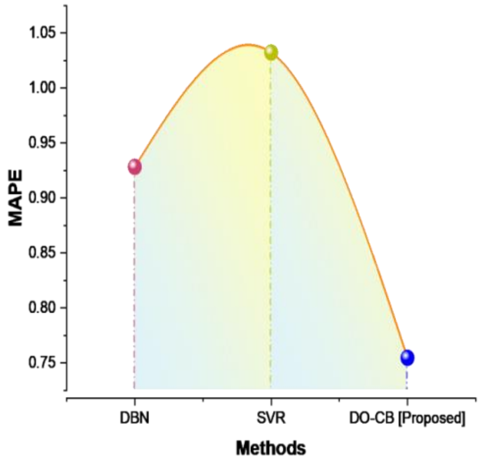

Figure 3. Result of MAPE

Table 3. Result of MAPE

Methods	MAPE
DBN	0.9283
SVR	1.0323
DO-CB [Proposed]	0.7546

MAE is a measure of the average absolute difference between expected and actual values. It assesses the total quantity of mistakes by assigning equal weight to errors without regard for their direction. The comparative evaluation of MAE is shown in Figure 4 and Table 4. When compared to the currently existing methods, such as DBN and SVR, which have MAE values of 0.0098 and 0.0108, respectively, the suggested DO-CB has a MAE value of 0.0074. The suggested technique outperforms existing methods for identifying data to assure resilience and incorporates a range of faults. The MAE equation (24) is as follows,

$$MAE = \frac{1}{m}\sum_{j=1}^{m} \left|z_j - \hat{z}_j\right| \tag{24}$$

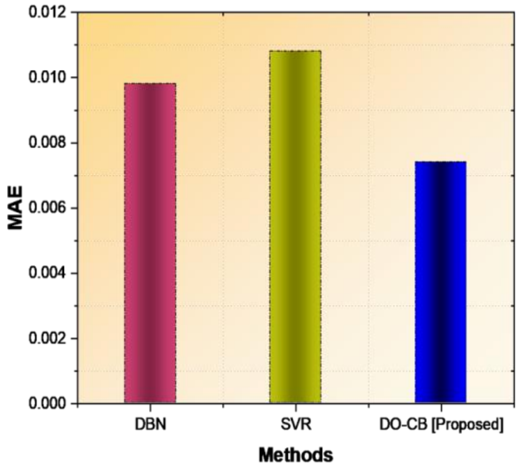

Figure 4. Result of MAE

Table 4. Result of MAE.

Methods	MAE
DBN	0.0098
SVR	0.0108
DO-CB [Proposed]	0.0074

4. CONCLUSION

In this study, we introduced a novel approach, DO-CB, developed on identified data and include a variety of fault scenarios. Synergistic Dandelion Optimization enhances the model's feature space search, optimizing the CatBoost method. This novel technique improves the recognition and classification of industrial failures in rotating machinery. The CWRU dataset was gathered and experimental results showed RMSE (0.011), MAPE (0.7546), as well as MAE (0.0074). The suggested method's outcomes were contrasted to other utilized algorithms and the outcomes of the evaluations showed that the suggested strategy was more effective for detecting various faults of rotational machinery. The efficiency of machine learning models is based on the data's quality. If the input information is noisy, complete and accurate, it can lead to false detections as well as misclassifications. In future research, continued advancements in deep learning techniques may lead to more sophisticated models capable of learning intricate patterns in machinery data, thereby improving fault detection accuracy.

References:

Brito, L.C., Susto, G.A., Brito, J.N., & Duarte, M.A. (2022).An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery.*Mechanical Systems and Signal Processing, 163*,https://doi.org/10.1016/j.ymssp.2021.108105

Carino, J.A., Delgado-Prieto, M., Iglesias, J.A., Sanchis, A., Zurita, D., Millan, M., Redondo, J.A.O., & Romero-Troncoso, R. (2018).Fault detection and identification methodology under an incremental learning framework applied to industrial machinery:*IEEE Access, 6,* 49755-49766. https://doi.org/10.1109/ACCESS.2018.2868430

Chen, S., Meng, Y., Tang, H., Tian, Y., He, N., & Shao, C. (2020).*Robust deep learning-based diagnosis of mixed faults in rotating machinery:IEEE/ASME Transactions on Mechatronics, 25(5),* 2167-2176.https://doi.org/10.1109/TMECH.2020.3007441

Chen, Z., Gryllias, K., & Li, W. (2019).Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network.*IEEE Transactions on Industrial Informatics, 16(1),* 339-349. https://doi.org/10.1109/TII.2019.2917233

Choudhary, A., Mian, T., & Fatima, S. (2021).*Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images:Measurement, 176,* 109196.https://doi.org/10.1016/j.measurement.2021.109196

Dineva, A., Mosavi, A., Gyimesi, M., Vajda, I., Nabipour, N., &Rabczuk, T. (2019). Fault diagnosis of rotating electrical machines using multi-label classification. *Applied Sciences, 9(23),* 5086. https://doi.org/10.3390/app9235086

Ding, Y., Ma, L., Ma, J., Suo, M., Tao, L., Cheng, Y., & Lu, C. (2019).Intelligent fault diagnosis for rotating machinery using deep Q-network based health state classification: A deep reinforcement learning approach.*AdvancedEngineering Informatics, 42,* 100977.https://doi.org/10.1016/j.aei.2019.100977

Gong, W., Chen, H., Zhang, Z., Zhang, M., Wang, R., Guan, C., & Wang, Q. (2019). A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. *Sensors, 19(7),* 1693.https://doi.org/10.3390/s19071693

González-Muñiz, A., Díaz, I., & Cuadrado, A.A. (2020).DCNN for condition monitoring and fault detection in rotating machines and its contribution to the understanding of machine nature.*Heliyon, 6(2).* https://doi.org/10.1016/j.heliyon.2020.e03395

Gupta, K., & Muzakkir, S. M. (2023b). A Model for Prediction of Outer Race Defects of Rolling Contact Bearing based on Vibration Data Using Machine Learning Algorithms. *Tribology in Industry*, *45*(4), 676–685. https://doi.org/10.24874/ti.1540.09.23.11

Khan, M.A., Asad, B., Kudelina, K., Vaimann, T., &Kallaste, A. (2022). The Bearing Faults Detection Methods for Electrical Machines—The State of the Art. *Energies, 16(1),* 296.https://doi.org/10.3390/en16010296

Kolar, D., Lisjak, D., Pająk, M., &Pavković, D. (2020). Fault diagnosis of rotary machines using deep convolutional neural network with wide three-axis vibration signal input. *Sensors, 20(14),* 4017. https://doi.org/10.3390/s20144017

Kumar, V., Tewari, R. P., Pandey, R., & Rawat, A. (2023b). Triboinformatic modeling of wear in total knee replacement implants using machine learning algorithms. *Journal of Materials and Engineering*, *1*(3), 97–105. https://doi.org/10.61552/jme.2023.03.001

Li, C., Sánchez, R.V., Zurita, G., Cerrada, M., & Cabrera, D. (2016). Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. *Sensors, 16(6), 895.*https://doi.org/10.3390/s16060895

Li, X., Li, X., & Ma, H. (2020).Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery:*Mechanical Systems and Signal Processing, 143,* 106825.https://doi.org/10.1016/j.ymssp.2020.106825

Li, X., Zhang, W., Ding, Q., & Sun, J.Q. (2020).Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation:*Journal of Intelligent Manufacturing, 31,* 433-452.https://doi.org/10.1007/s10845-018-1456-1

Li, Y., Du, X., Wan, F., Wang, X., & Yu, H. (2020):Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging.*Chinese Journal of Aeronautics, 33(2),* 427-438.https://doi.org/10.1016/j.cja.2019.08.014

Li, Z., Wang, Y., & Wang, K. (2019).A deep learning driven method for fault classification and degradation assessment in mechanical equipment.*Computers in Industry, 104,* 1-10. https://doi.org/10.1016/j.compind.2018.07.002

Li, Z., Wang, Y., & Wang, K. (2019): A deep learning driven method for fault classification and degradation assessment in mechanical equipment.*Computers in Industry, 104,* 1-10. https://doi.org/10.1016/j.compind.2018.07.002

Luo, B., Wang, H., Liu, H., Li, B., & Peng, F. (2018). Early fault detection of machine tools based on deep learning and dynamic identification. *IEEE Transactions on Industrial Electronics, 66(1),* 509-518. https://doi.org/10.1109/TIE.2018.2807414

Luwei, K.C., Yunusa-Kaltungo, A., &Sha'aban, Y.A. (2018).Integrated fault detection framework for classifying rotating machine faults using frequency domain data fusion and artificial neural networks:*Machines, 6(4),* 59.https://doi.org/10.3390/machines6040059

Mehta, A., Goyal, D., Choudhary, A., Pabla, B.S., &Belghith, S. (2021). Machine learning-based fault diagnosis of self-aligning bearings for rotating machinery using infrared thermography.*Mathematical Problems in Engineering, 2021,* 1-15. https://doi.org/10.1155/2021/9947300

Senanayaka, J.S.L., Kandukuri, S.T., Van Khang, H., &Robbersmyr, K.G. (2017, April).Early detection and classification of bearing faults using support vector machine algorithm:In 2017 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD) (pp. 250-255). *IEEE.* https://doi.org/*10.1109/WEMDCD.2017.7947755*

Shao, H., Jiang, H., Li, X., & Liang, T. (2018).Rolling bearing fault detection using continuous deep belief network with locally linear embedding:*Computers in Industry, 96,* 27-39.https://doi.org/10.1016/j.compind.2018.01.005

Sobie, C., Freitas, C., & Nicolai, M. (2018).Simulation-driven machine learning: Bearing fault classification.*Mechanical Systems and Signal Processing, 99,* 403-419. https://doi.org/10.1016/j.ymssp.2017.06.025

Souza, R.M., Nascimento, E.G., Miranda, U.A., Silva, W.J., & Lepikson, H.A. (2021).Deep learning for diagnosis and classification of faults in industrial rotating machinery. *Computers & Industrial Engineering, 153,* 107060.https://doi.org/10.1016/j.cie.2020.107060

Surendran, R., Khalaf, O.I., & Tavera Romero, C.A. (2022).Deep Learning Based Intelligent Industrial Fault Diagnosis Model:*Computers, Materials & Continua, 70(3).*https://doi.org/10.1016/j.isatra.2018.12.025

Tang, S., Yuan, S., & Zhu, Y. (2019).Deep learning-based intelligent fault diagnosis methods toward rotating machinery*:IEEE Access, 8,* 9335-9346. https://doi.org/10.1109/ACCESS.2019.2963092

Wei, H., Zhang, Q., Shang, M., & Gu, Y. (2021).Extreme learning machine-based classifier for fault diagnosis of rotating machinery using a residual network and continuous wavelet transform:*Measurement, 183,* 109864.https://doi.org/10.1016/j.measurement.2021.109864

Wu, C., Jiang, P., Ding, C., Feng, F., & Chen, T. (2019).Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network.*Computers in Industry, 108,* 53-61.https://doi.org/10.1016/j.compind.2018.12.001

Wu, X., Zhang, Y., Cheng, C., & Peng, Z. (2021).A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery.*Mechanical Systems and Signal Processing, 149,* 107327.https://doi.org/10.1016/j.ymssp.2020.107327

Xia, M., Li, T., Xu, L., Liu, L., & De Silva, C.W. (2017).Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks:*IEEE/ASME Transactions on Mechatronics, 23(1),* 101-110.https://doi.org/10.1109/TMECH.2017.2728371

Zhang, W., Li, X., & Ding, Q. (2019).Deep residual learning-based fault diagnosis method for rotating machinery:*ISATransactions, 95,* 295-305. https://doi.org/10.1016/j.isatra.2018.12.025

Zhang, Y., Zhou, T., Huang, X., Cao, L., & Zhou, Q. (2021).Fault diagnosis of rotating machinery based on recurrent neural networks:*Measurement, 171,*108774. https://doi.org/10.1016/j.measurement.2020.108774

DivyaPaikaray
Arka Jain University, Jamshedpur, Jharkhand, India
divya.p@arkajainuniversity.ac.in
ORCID 0000-0001-7886-1538

Naveen Kumar Rajendran
JAIN (Deemed-to-be University), Ramanagara District, Karnataka, India
r.naveenkumar@jainuniversity.ac.in
ORCID 0000-0001-7313-1712

Vaishali Singh
Maharishi University of Information Technology, Uttar Pradesh, India
singh.vaishali05@gmail.com
ORCID 0000-0001-8304-8947

Pulkit Srivastava
Noida Institute of Engineering & Technology,
Greater Noida, Uttar Pradesh, India
p.srivastava@niet.co.in
ORCID 0000-0003-3393-1066