Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Proposed Damage Detection and Isolation from Limited Experimental Data Based on a Deep Transfer Learning and an Ensemble Learning Classifier

Chuanyuan Tan[1], Fanru Gao[2], Chaoda Song[3], Meicai Xu[4], Yizhou Li[5], Haowei Ma[6*]
[1]*College of Information Technology and Engineering department, Marshall University, West Virginia, USA, 77449*
chuanyuantan396@gmail.com
[2]*Department of Mechanical and Aerospace Enginerring, Case Western Reserve University, Cleveland, OH, 44106, USA*
fxg149@case.edu
[3]*Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA*
cxs965@case.edu
[4]*Department of Biosystems & Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA*
xumeica2@msu.edu
[5]*Department of Electrical, Computer, and Systems Engineering, Case Western Reserve University, Cleveland, 44106, USA*
yxl3527@case.edu
[6*]*Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH 44106, USA*
hxm502@case.edu

ARTICLE INFO

Received: 18 Mar 2025

Revised: 25 May 2025

Accepted: 20 Jun 2025

ABSTRACT

A re-enactment model can give knowledge into the trademark ways of behaving of various wellbeing conditions of a genuine framework; be that as it may, such a recreation can't represent all intricacies in the framework. The work proposes a Deep ensemble learning procedure that utilizes straightforward programmatic experiences for fault diagnosis in a genuine framework. A straightforward shaft-plate framework was utilized to create a significant arrangement of source information for three wellbeing conditions of a rotor framework, and that information is utilized to prepare, approve, and test a redid profound neural organization. The learning model is pretrained on reproduction information based on RT-WT and was utilized as a space and class invariant summed up with highlight feature extractor i.e. EXPSO-STFA and the extricated highlights was handled with Deep ensemble learning. A fault diagnosis strategy in view of Transfer learning i.e. Dense Net and convolutional neural organization (CNN) permits different modern members to cooperatively prepare a global fault-diagnosis approach without exchanging local data. Model preparation is privately executed inside each modern member, and the cloud server refreshes the worldwide model by totalling the nearby models of the members. In particular, a versatile technique is intended to change the model conglomeration span as per the criticism data of the modern members to diminish the correspondence cost while guaranteeing model exactness. The proposed Transfer learning and ensemble learning classifier technique were additionally approved by contrasting its exhibition and the prior profound learning models of DBN, CNN, LSTM and GRU in terms of computational cost, generalizability, feature extraction, , network size and boundaries.

Keywords: Deep convolution neural networks (DCNN), Retinex based Wavelet Transform (RT-WT), Explorated Particle Swarm Optimization Embedded on Sea Turtle Foraging Algorithm (EXPSO-STFA), Fault diagnosis, transfer learning (TL), signal processing, Pre-processing, Image Net, Dense Net and image classification.

I. Introduction

Somewhat recently, with the inexorably mind boggling construction of mechanical hardware design and quick advancement of sensor innovation, the obtaining of vibration signal has become simple and has carried new points of view and difficulties to the customary keen fault diagnosis of pivoting machinery [1][2]. Albeit these customary insightful fault diagnosis strategies referenced above can accomplish great outcomes, they are undeniably founded on the accompanying two suspicions: (1) countless marked fault data tests are accessible and (2) the preparation and testing tests are imparted to a similar likelihood negligible conveyance. In any case, in genuine designing, it is an

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

extravagance to gather enormous marked fault data tests, and the gathered information from obscure activity condition are not drawn from a similar likelihood minor dispersion [3][4].

The most commonly used methods of fault diagnosis include model-based methods and data-driven methods [5]. In model-based methods, the physics underlying the system's behavior are modelled and used for fault diagnosis. It is difficult or even impossible to precisely model the behaviour of complex systems, owing to the wide range of structural complexities and environmental uncertainties that affect such systems [6]. Data-driven methods use data obtained from sensors in the system to carry out fault diagnosis; these methods do not require much knowledge about the underlying kinematics and physics of the failure of the system [7][8][9][10]. In traditional data-driven fault diagnosis techniques using machine learning, the signals from sensors are usually subjected to pre-processing (e.g., noise removal, domain transformation (time to frequency), signal decomposition (empirical mode decomposition)), extraction of discriminative features (e.g., time and frequency domain statistical features), selection of features that are more sensitive to damage (e.g., feature ranking), and processing of the selected features with supervised or unsupervised machine learning algorithms [11]. The machine learning algorithms for fault diagnosis performance is heavily dependent on the set of discriminative features that is selected [12]. A set of statistical features may work well for one problem and may fail completely for another problem in the same domain but on a different scale [13]. In general, there is no optimized set of processing steps for fault diagnosis using handcrafted statistical features from sensor data and machine learning algorithms. For instance, a data-driven diagnostic strategy that uses simulation data may not be generalized to a dataset from an experimental setup of the same problem without a complex process of model updating [14][15][16]. Moreover, the extraction of damage sensitive features is labour-intensive and requires considerable diagnostic skills and domain expertise [17][18].

In this manner, an enormous degree exists for working on the exhibition of profound learning so it can advance successfully from little preparation informational collection [19]. In this article, an original fault location and confinement approach for resolving the issue of restricted trial information is proposed. The proposed framework joins Deep Transfer Learning with Ensemble Learning Classifier to recognize the harm in machinery and detach them precisely.

The article's remaining portions are: Section 2 goes with the associated work and problem description. Section 3 presents the recommended methodology. Section 4 demonstrates the findings and discussions, while Section 5 ends the research.

2. Review of Literature

In 2020, Guo *et al.* [20] introduced a profound learning-based technique that separated the harm highlights from mode shapes without using any hand-designed element or earlier information. To meet different necessities of the harm situations, convolutional neural organization (CNN) calculation was used and planned another organization engineering: a multi-scale module, which helped in extricating highlights at different scales that can decrease the impedance of debased information; stacked lingering learning modules, which helped in speeding up the organization combination; and a worldwide normal pooling layer, which helped in diminishing the utilization of figuring assets and getting a relapse execution. The outcomes showed that the technique had exactness upgrades of something like 10% over other organization designs.

In 2020, Chen *et al.* [21] introduced another information transfer network with an inadequate auto encoder and a profound convolutional neural organization. In engineering, input prompt turning speed (IRS), as functional condition data, was taken care of into the scanty in the objective area with the goal that the functional data was incorporated into the model preparation as opposed to utilize the halfway vibration dataset as it were. Then, at that point, profound convolutional neural organization (DCNN) was used to remove highlights from crude vibrations. The model empowered the fault design limit acknowledgment under fluctuating functional circumstances for moving bearing fault discovery.

In 2021, Khan *et al.* [22] presented a supervised learning framework of transfer learning. A source simulation domain with three health states was employed to detect, isolate, and quantify five health states in the target experimental domain that minimized the gap in the response characteristics of the two domains. The source domain was comprised of the simulation model of a few representative health states of the target domain, and simulation models were not required for all prospective health states of the actual target system. Synthetic augmentation of the experimental data

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

enhanced the performance of this approach in terms of training/validation accuracy (from 88.8% to 99.5%), test accuracy (90% to 97.14%), and ROC area (from 97% to 100%).

In 2020, Shao *et al.* [23] showed a transfer diagnosis of rotor-bearing framework faults under various working circumstances in light of cutting edge observing measures. Stochastic pooling and LReLU were joined to work on the exhibition of the fundamental CNN. Boundary transfer was utilized to give great beginning boundaries to empower the planned CNN to take on new spaces with restricted accessible preparation information. The technique was applied to dissect the warm pictures of rotor-bearing framework gathered under various working circumstances. The examination results affirmed the unrivaled execution of the technique contrasted and the current strategies in fault diagnosis of rotor-bearing framework.

In 2021, Yan *et al.* [24] introduced a CNN mine derrick slowing mechanism fault diagnosis strategy. An information encoding strategy was introduced to change over the one layered signal gathered by the sensor into a two-layered picture. The first informational collection tests addressed by the two-layered exhibit were planned to the preparation set and the test set addressed by the three-layered cluster, which expanded the effectiveness and unwavering quality of data transmission. The better CNN for fault diagnosis has accomplished 99.375% exactness.

In 2022, Dong *et al.* [25] introduced another diagnosis structure in view of to tackle the little example issue in genuine diagnosis situation of moving component heading race faults. In this structure, reproduction information gathered from dynamic model of moving component course were used to help the genuine situation diagnosis that lacked information in view of CNN and boundary transfer techniques, working straightforwardly on crude vibration signals without manual element extraction. The outcomes showed that the recreation information helper and four positive transfer procedures, this diagnosis system was attractive to taking care of the little example issue in genuine diagnosis situation where the gained information of the bearing to be analyzed was lacking to prepare a viable diagnosis model.

In 2020, He *et al.* [26] introduced a naturally kinds of turning equipment cross working circumstances, ensemble transfer CNNs. They were utilized to change the essential CNN, and afterward progressions of adjusted CNNs were developed with multi-channel indications. With the assistance of boundary transfer, source CNNs was applied to manage target tests. Another choice combination procedure was intended to intertwine every individual outcome deftly.

In 2020, Han *et al.* [27] introduced a space transformation approach, JDA based DTN to exploit a pre-prepared network from the source space and get the model transferred with unlabeled information from the objective space. Through broad analyses on three datasets, the outcomes showed that the DTN with JDA beats the cutting edge draws near. Contrasted and the shallow strategies, i.e., SVM, RF, EMD, TJM, TCA and JDA, DTN with JDA accomplished 25.2%, 24.3%, 31.4% 15.4%, 16.7%, 15.8% enhancements for the typical exactness in ten diagnosis assignments.

In 2021, Wang *et al.* [28] introduced a profound transfer multi-wavelet auto-encoder. To begin with, multi-wavelet initiation work was utilized to configuration new-type profound multi-wavelet auto-encoder. Then, comparability measure was utilized to choose excellent helper tests for well pre-preparing a source model containing comparative qualities with the objective space. At last, boundary information obtained from the source model was transferred to target model utilizing not very many objective preparation tests.

In 2020, Zhang *et al.* [29] introduced a profound learning-based fault diagnosis strategy to address the imbalanced information issue utilizing generative ill-disposed networks. Various age modules were taken on for information increase of the minority classes. Through antagonistic preparation between the generator and discriminator, the planning between the appropriations of clamor and genuine information was laid out to create extra phony examples to adjust and additionally grow the preparation dataset. In view of the trial approvals on two turning machinery datasets, the information driven fault analytic model fundamentally profited from the created counterfeit examples.

3. Proposed Methodology

Fault location and separation can be characterized as a sort of example acknowledgment basically. The previous ten years has seen the ascent and improvement of information driven fault diagnosis explores, because of their high information handling effectiveness and benefits as far as computerization and insight[30]. The customary investigates about information driven fault diagnosis will generally gain an order model from a lot of named fault information to

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

have adequate speculation limit. In order to conquer the existing issue the work has developed transfer learning based CNN and Ensemble learning to detect damage as shown in figure 1.

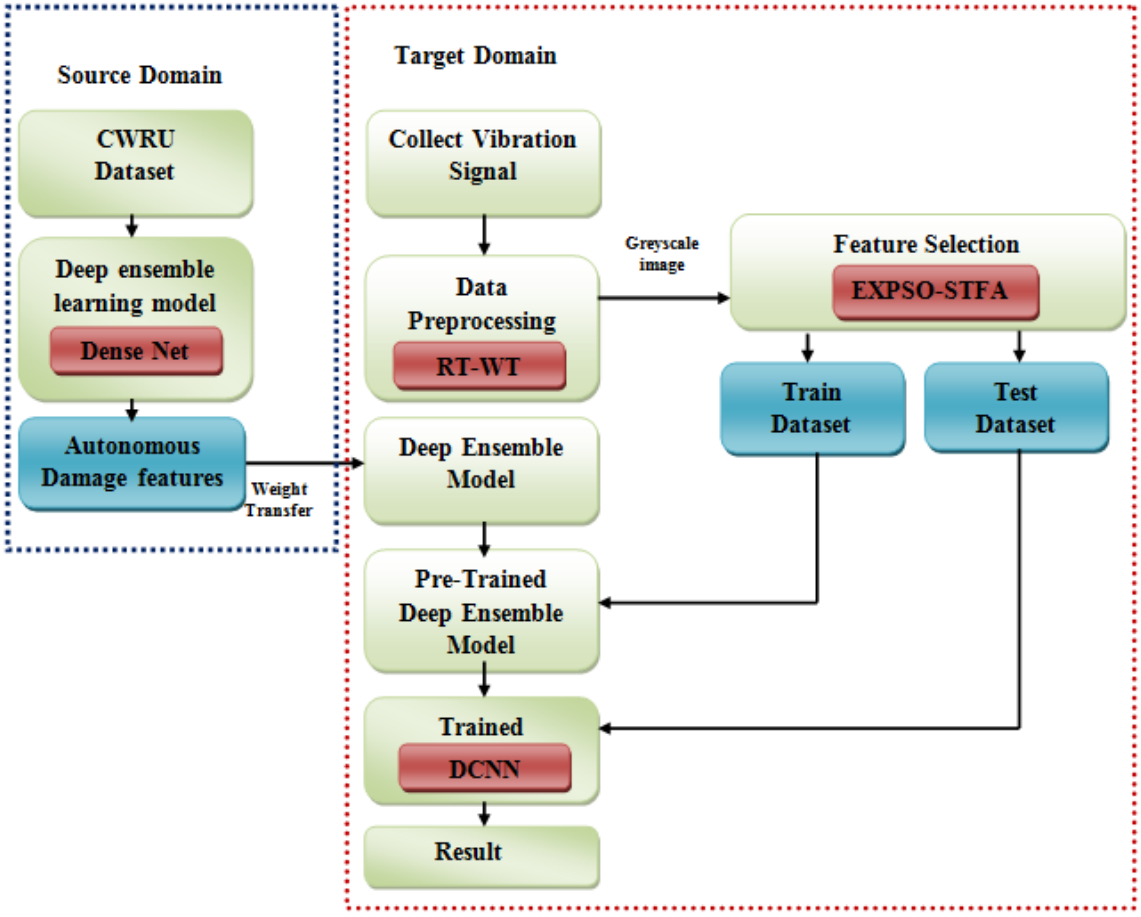


Figure 1: Proposed framework

3.1. Source Domain

The work is carried out based on CWRU Datasets .CWRU dataset comprises of the test bench for motor performance assessment consists of:

- Motor with 2 HP power
- Torque transducer
- Dynamometer
- Control electronics

3.1.1. Deep Ensemble Learning Model

The residual neural network can solve gradient descent and disappearance better than the convolutional neural network. Because each layer has distinct weights, the number of residual neural network parameters rises with depth [31, 32]. Deep Ensemble learning employs Dense Net and CNN as training and testing networks to overcome this challenge. A dense block, transition layer, and bottleneck layer make up Dense Net. Figure 2 shows the dense block's multilayer network and composite function.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

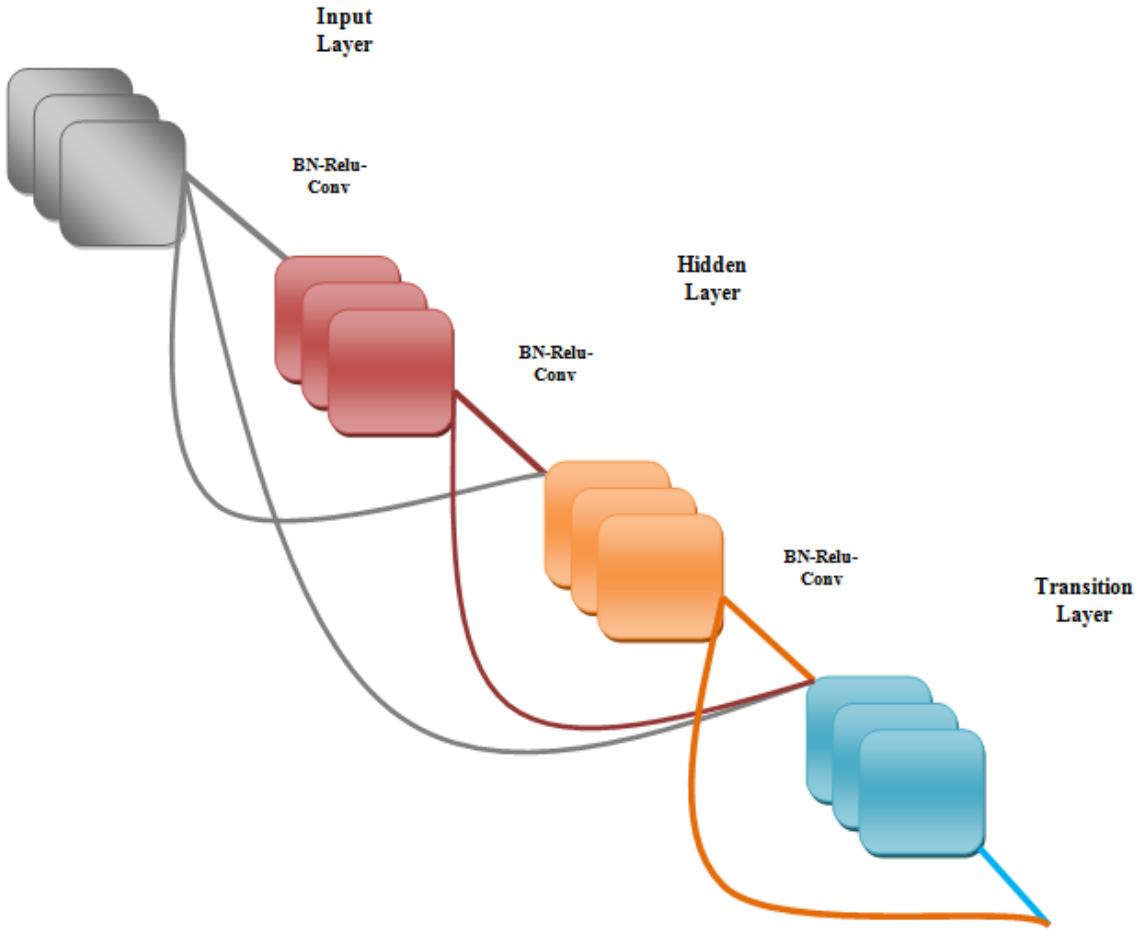


Figure 2: Dense Net Model

Normalization, linear rectification, and convolution compensate the composite function [33,34]. The l-layer of its l^{th} layer network gets the previous layer's $l-1$ feature map outputs. Construction formula:

$$I_l = \mathrm{M}\left(\left[I_0, I_1, \ldots\ldots I_{l-1}\right]\right) \tag{1}$$

$$\Phi = \Gamma\left(I_i\{w_i\}\right) + I \tag{2}$$

$$\Gamma = w_2\beta\left(w_1 I + B_1\right) + B_2 \tag{3}$$

$$\beta(I) = \max(0, I) \tag{4}$$

Where $\forall$ and Φ implies input and output vectors of the layers, w implies weight in the weight matrix, and Γ implies residual mapping to be learned, β signifies the Rectified Linear Unit activation function, and w_1, w_2 and B_1, B_2 implies weights and biases of the first layer and the second layer, respectively. Finally the autonomous features are extracted according to the weight and transferred to target domain.

3.2. Target Domain

Target Domain comprises of Pre-processing and validation of pre-processed data for fault diagnosis.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.2.1. Pre-processing

Initial vibrations were mostly ambient vibrations and sensor-white-noise. Training datasets for neural networks should only include valid vibration segments and labels. Raw signal data was cleaned to reduce noise and obtain valid segments. Short-term energy characterizes each frame in a vibration signal [35, 36]. Data cleaning included a short-term energy analysis since valid signals vary from noise in acceleration amplitude (short-term energy).

Data pre-processing is imperative for conventional strategies since it is challenging to straightforwardly utilize crude data. One of the pre-processing strategies is the fake of elements, which is generally tedious and needs master information. Training and testing of data should be clean in order to avoid probability of error while detecting fault. The pre-processed signal data is given by the below equation:

$$\chi^{\pm} = \chi^{pre}\left[\Theta_{Coll}\right] \tag{5}$$

$$\Omega = \left[\Theta_{Coll}\right] \tag{6}$$

Where, Ω indicates the pre-processed target data.

CNN can analyze 2-D or 3-D data. CNN's one-dimensional diagnostic performance is poor. Therefore, 1-D vibration signals must be converted to 2-D for CNN [37, 38, 39]. As a time–frequency domain conversion technique, the signal image is first processed by wavelet transform and then by the Retinex algorithm. The following steps are followed:

Step1: The image is wavelet-transformed. Wavelet transform yields horizontal and vertical low-frequency (LL), horizontal high-frequency and vertical low-frequency (HL), horizontal low-frequency and vertical high-frequency (LH), and horizontal and vertical (HH) components.

Step2: Retinex now processes horizontal and vertical LL. Gaussian low pass filtering is used to prevent losing high-frequency components.

Step3: Retinex logarithm of LL ($\log \xi_{LL}$) is:

$$\log \xi_{LL} = \log \Omega_{LL} - \log I_{LL} \tag{7}$$

Where, Ω_{LL} denotes the fused image and I_{LL} is the illumination for each pixel.

Step4: The LL (ξ_{LL}) reflectance is then calculated.

$$\xi_{LL} = \exp\left(\log \Omega_{LL} - \log I_{LL}\right) \tag{8}$$

Step5: Retinex algorithm-processed as ξ_{LL} and inverse wavelet transform-processed HL, LH, HH. The image is then improved.

3.2.2. Feature Selection

Feature selection helps to choose the optimal features for a classification model. Feature selection enhances model accuracy and reduces calculation time. Various hybrid optimization techniques have been devised to choose important features, but exploration-exploitation of the search space remains a hurdle. Exploration selects new locations from a large search area, whereas exploitation refines previously visited sites to improve solution consistency. Exploration reduces process performance but gives fresh search sites, whereas exploitation refines the existing solution but drives the process to locally optimum solutions. Exploration and exploitation problems may lead to poor management of ambiguous data, reducing damage detection efficiency. To solve this challenge, researchers created an Explorated Particle Swarm Optimization Embedded on Sea Turtle Foraging Algorithm (EXPSO-STFA). The created approach balances Exploration with Exploitation and efficiently selects features. The developed EXPSO-STFA is fully designed in order to balance exploration and exploitation. The EXPSO contributes towards improving the exploration by avoiding the premature convergence of the particle near the optimal point whilst the optimal point is at the local optimum. Normally, the PSO algorithm doesn't achieve a better exploration rate, in order to obtain a better exploration, the PSO algorithm is combined along with a simulated annealing optimization technique. On the other

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

hand, to provide better exploitation for PSO, the work has inhabited STFA working mechanism in the PSO algorithm. STFA, which is fully based on local search, is embedded in EXPSO to exploit the promising area founded by EXPSO.

The EXPSO-STFA algorithm is presented below.

Step1: Initial particles are generated randomly.

Step2. Calculate each particle's fitness

Step3: To provide a variety of search, every particle is transformed to a new position using the Simulated Annealing (SA) algorithm according to the Expression

$$\mathrm{I}_1^*(t+1) = \mathrm{I}_1^*(t) + r_1 - r_1.2.randn() \tag{9}$$

Where parameter r_1 decreases as generation increases, $randn()$ provides a standard normal distribution sample with 0 mean and 1 standard deviation.

Step 4: If particles' fitness is better than personal best κ_{PBEST} , update personal best κ_{PBEST} with fitness.

Step 5: If a particle's fitness is better than its global best position κ_{GBEST} , update (κ_{GBEST}) with its fitness.

Step 6: Calculate each particle's velocity I_{vel}^* and position I_{Pos}^* using,

$$\mathrm{I}_{Pos}^*(t) = \ell[\mathrm{I}_1^*(t) - \mathrm{I}_i^*(t)] \tag{10}$$

$$\mathrm{I}_{vel}^*(t+1) = \mathrm{I}_i^*(t) + \mathrm{I}_{Pos}^*(t) + \left[\frac{f(\mathrm{I}_i^*(t)) - f(\mathrm{I}_i^*(t-1))}{f(\mathrm{I}_i^*(t-1)}\right]\left[\mathrm{I}_i^*(t) - \mathrm{I}_i^*(t-1)\right] \tag{11}$$

Step 7: Calculate κ_{GBEST} fitness from particle's food position(i). Perceived κ_{GBEST} and computed velocity affect the particle's future location. Comparing $\hbar_{ij}(t)$ particle fitness to food source fitness. If particle fitness is greater than the food source's, κ_{GBEST} set to zero. If particle fitness is lower than the food source's, then it is.

$$\hbar_{ij}(t) = \frac{f(\lambda_i)}{\sum_{q=1}^{p} f(\lambda_q)} e^{-\left[\frac{D_{ij}^2}{2\sigma^2(t)}\right]} \tag{12}$$

Where $f(\lambda_i)$ represents the food source i fitness, D_{ij} is the distance between particle j and the food source i , time lowers $\sigma(t)$'s convergence rate.

$$\sigma(t) = \sigma_\circ e^{-\left[\frac{1}{t}\right]} \tag{13}$$

Step 8: Identify particle's j top food source $\hbar_{ij}(t)$ utilizing,

$$d = \arg\max[\hbar_{ij}(t)] \tag{14}$$

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Step 9: Update particle position

$$U_i^*(t+1) = U_i^*(t) + \chi \mathrm{I}_{Pos}^*(t) + \hbar_{Ij}(t)[\lambda_j - U_i^*(t)] \tag{15}$$

Where, $\hbar_{Ij}$ implies κ_{GBEST} fitness of the food source I .

Step 10: Fitness sorts the combined Features list. Next, the top N features are chosen.

Step11: Steps 3-10 are repeated till end.

Thus, the feature selection approach gives the most relevant damage-detection features in a data frame.

$$\Omega_{selection}^{IDS} = \left[\mathrm{I}_1^T, \mathrm{I}_2^T, \mathrm{I}_3^T, \mathrm{I}_4^T, \mathrm{I}_5^T, \mathrm{I}_6^T, \mathrm{I}_7^T\right] \tag{16}$$

Figure 2 illustrates the outline procedure for performing feature selection based on EXPSO-STFA in the form of a pseudo code.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Input: Extracted Features
Output: Selection of Top N Features

Begin
Initialize the particles randomly
Evaluate the fitness of each particle
$t = 1$
While $(t < \max_{iter})$
If $\kappa_p \rightarrow local\ \min ima$
Perform transformation of the particle for variety search using,
$$\Gamma_i^*(t+1) = \Gamma_i^*(t) + r_1 - r_1.2.randn()$$
End if
If $(\kappa_p > \kappa_{pbest})$
Return $\kappa_{pbest} = \kappa_p$
End if
If $(\kappa_{pbest} < \kappa_{gbest})$
Return $\kappa_{gbest} = \kappa_{pbest}$
End if
Calculate velocity and position of particle using,
$$\Gamma_{Pos}^*(t) = \ell[\Gamma_1^*(t) - \Gamma_i^*(t)]$$
$$\Gamma_{vel}^*(t+1) = \Gamma_i^*(t) + \Gamma_{Pos}^*(t) + \left[\frac{f(\Gamma_i^*(t)) - f(\Gamma_i^*(t-1))}{f(\Gamma_i^*(t-1)}\right][\Gamma_i^*(t) - \Gamma_i^*(t-1)]$$
Evaluate the fitness of food source that is perceived by particle using,
If $(\kappa_{gbest} > \hbar_{ij}(t))$
Set $\kappa_{gbest} = 0$
Else
Control the fitness value using,
$$\hbar_{ij}(t) = \frac{f(\lambda_i)}{\sum_{q=1}^{p} f(\lambda_q)} e^{-\left[\frac{D_{ij}^2}{2\sigma^2(t)}\right]}$$
End If
Identify the particle with highest food source using,
$$d = \arg\max[\hbar_{ij}(t)]$$
Update the position of particle using,
$$U_i^*(t+1) = U_i^*(t) + \chi\Gamma_{Pos}^*(t) + \hbar_{Ij}(t)[\lambda_j - U_i^*(t)]$$
Sort top N features
End While
End Begin

Figure 3: Pseudo Code for EXPSO-STFA

3.2.2. Training and Testing

The new damage detection approach enhances accuracy, converges quicker, and aggregates visual information at a deeper neural network level.

The training and testing is done based on fused pre-processed data i.e.

$$F = \left[\Theta_{Coll} + I_{coll}\right]$$

Each sub-band enters the convolutional layer, which conducts two-dimensional convolutions between the extracted input features F_n^G, where G and n implies the sub-band levels and feature indices. The convolutional layer is computed as

$$\hbar_n^G = \sum_{i=0}^{n}\sum_{j=0}^{n} \omega_{i,j}^G F_{i,j}^G + \zeta_{i,j}^G \quad (17)$$

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$\hbar_n^G = Z\left(\sum_{i=0}^{N}\sum_{j=0}^{N}\omega_{i,j}^G F_{i,j}^G + \zeta_{i,j}^G\right) \tag{18}$$

Where, i, j denotes rows and columns that represents the features, $\omega_{i,j}^G$ denotes the G subband weight , $\zeta_{i,j}^G$ implies bias, Z implies relu activation function.

Thereafter, the convolutional image is followed by the spatial-temporal pooling layer that avoids the fixed-length representation of the image in terms of size/scale and also provides object deformation. It helps to avoid the loss of image. The STP layer is added to the last convolutional layer which consists of 3 pyramid levels (6x6, 3x3, 1x1) totally 46bins and generates a fixed-length output, which is fed into the fully connected layer.

$$\hbar_n^G = \mathrm{P}_{layers}^{STP}\left(\left|F_{i,j}^G\right|\right) \tag{19}$$

Where, $\mathrm{P}_{layers}^{STP}$ denotes an STP pooling layer. Finally, the convolutional layer and the STP pooling layer form the convolutional neural network i^{th} layer together, that is capable of capturing the low-level information and reducing the complexity of the image. So, convolutional layer output is compressed and delayed.

$$\hbar_n^G(F_{Flatten}) = \left[F_1, F_2, F_3, F_4, \ldots F_n\right] \tag{20}$$

It creates an N-dimensional vector, where N implies number of compressed feature classes. A softmax activation function (act_σ) is utilized to label input images.

$$\hbar_m = act_{sm}\left(\sum_{i=1}^{n}\omega_{i,j}F_{i,j} + \zeta_{i,j}\right) \tag{21}$$

The softmax provides the probability-based detection of the image. The threshold for the probability is evaluated using the ROC curve. For the developed approach, there is a need for a high true positive rate and less false positive rate. The threshold value selected is (0.88 TPR, 0.12FPR). The points lying inside the curve denote the damage image and outside the curve denotes normal image.

Vector calculus and chain rule are used for CNN learning.

Let β be scalar (i.e., $\beta \in R$) and $\Omega \in R^H$ be vector. So, if β is a function of Ω, then the β partial derivative with respect to Ω be vector, expressed as:

$$\left(\frac{\partial\beta}{\partial\Omega}\right)_i = \left(\frac{\partial\beta}{\partial\Omega}\right)_i \tag{22}$$

Its i-th element is $\left(\frac{\partial\beta}{\partial\Omega}\right)_i$: Notably,

$$\left(\frac{\partial\beta}{\partial\Omega^T}\right) = \left(\frac{\partial\beta}{\partial\Omega}\right)^T \tag{23}$$

Also, suppose Ω is a function of κ. The Ω partial derivative is:

$$\left(\frac{\partial\Omega}{\partial\kappa^T}\right)_{ij} = \frac{\partial\Omega_i}{\partial\kappa_i} \tag{24}$$

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

This fractional derivative is a H W matrix whose ith-row-jth-column entry is $\frac{\partial \Omega_i}{\partial \kappa_i}$. In a chain-like argument, it's clear that translates κ to Ω. Chain rule calculates:

$$\left(\frac{\partial \beta}{\partial \kappa^T}\right), as \left(\frac{\partial \beta}{\partial \kappa^T}\right)=\left(\frac{\partial \beta}{\partial \Omega^T}\right)\left(\frac{\partial \Omega}{\partial \kappa^T}\right) \tag{25}$$

Using a loss function, one may analyze the difference between a CNN's prediction κ^L and the target $t,\kappa^1 \to w^1,\ t,\kappa^2 \to .....,\ t,\kappa^L \to w^L=\beta,,$. Complex functions are often used. Prediction output is $\arg\max_i \kappa_i^l$. Convolution process:

$$\Omega i^{l+1}, j^{l+1}, d=\sum_{i=0}^{h}\sum_{j=0}^{w}\sum_{k=0}^{d} F_{i,j,k}\times \kappa_{i^{i+1}}^{L}+, ij^{l+1}+j,k \tag{26}$$

The filter F has $\left(h\times w\times d^l\right)$ size, therefore convolution will be $\left(h^l-h+1\right)\times\left(w^l-w+1\right)$ with D slices, which means

$$\Omega\left(\kappa^{l+1}\right)\text{in } \mathrm{P}^{h^{l+1}\times w^{l+1}\times d^{l+1}},\ \ h^{l+1}=h^l-h+1, w^{l+1}=w^l-w+1, d^{l+1}=d \tag{27}$$

In Inception V3, each training example's label $\eta \in \{1,\ldots n\}$ probability is calculated as:

$$P(\eta \mid n)=\frac{\exp(z_k)}{\sum_{i}^{k}\exp(z_k)}, \tag{28}$$

where β implies non-normalised log probability. Normalizing $q(\eta \mid n)$ ground truth distribution across labels $\sum_k q(\eta \mid n)=1$. Cross-entropy gives this model's loss.

$$l=\sum_{k=1}^{K}\log(P(\eta))q(\eta) \tag{29}$$

Cross-entropy loss is distinguishable with respect to logits, and may be employed for gradient training of deep models, where the simple gradients are:

$$\frac{\partial l}{\partial z_k}=P(\eta)-q(\eta) \tag{30}$$

The above eqn is bounded between −1 and 1. Usually, when minimising the cross entropy, it increased log-likelihood of right label. Inception V3 analyzes a label distribution independent of training examples $u(\eta)$ with a smooth parameter $\in$, where given a training example, the label distribution $q(\eta \mid n)=\vartheta_{\eta|,\Omega}$ is updated by:

$$q^{l(\eta|n)}=(1-\in)\vartheta_{\eta,n}+\in u(\eta), \tag{31}$$

Which combines initial $q(\eta \mid n)$ and fixed $u(\eta)$ weighted distributions. A label-smoothing regularization with uniform distribution $u(\eta)=1/K$ yields:

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$q^{l(\eta|n)} = (1-\in)\vartheta_{\eta,n} + \frac{\in}{K}, \tag{32}$$

Cross-entropy is defined as follows.

$$h(q', P) = -\sum_{k=1}^{K} \log(P(\eta)) q^{l(k)} = (1-\in)h(q', P) + \in h(u, P) \tag{33}$$

Label-smoothing regularization is comparable to using a single cross entropy loss $h(q, P)$ with a pair of losses $h(q, P)$ and $h(u, P)$. Second loss hinders label distribution P divergence from previous with weight $\frac{\in}{\in +1}$, which is similar to determining the Kullback–Leibler divergence. Now based on the above perceptions the TL model from source domain gets added up with the deep convolutional neural network as shown in figure 4.

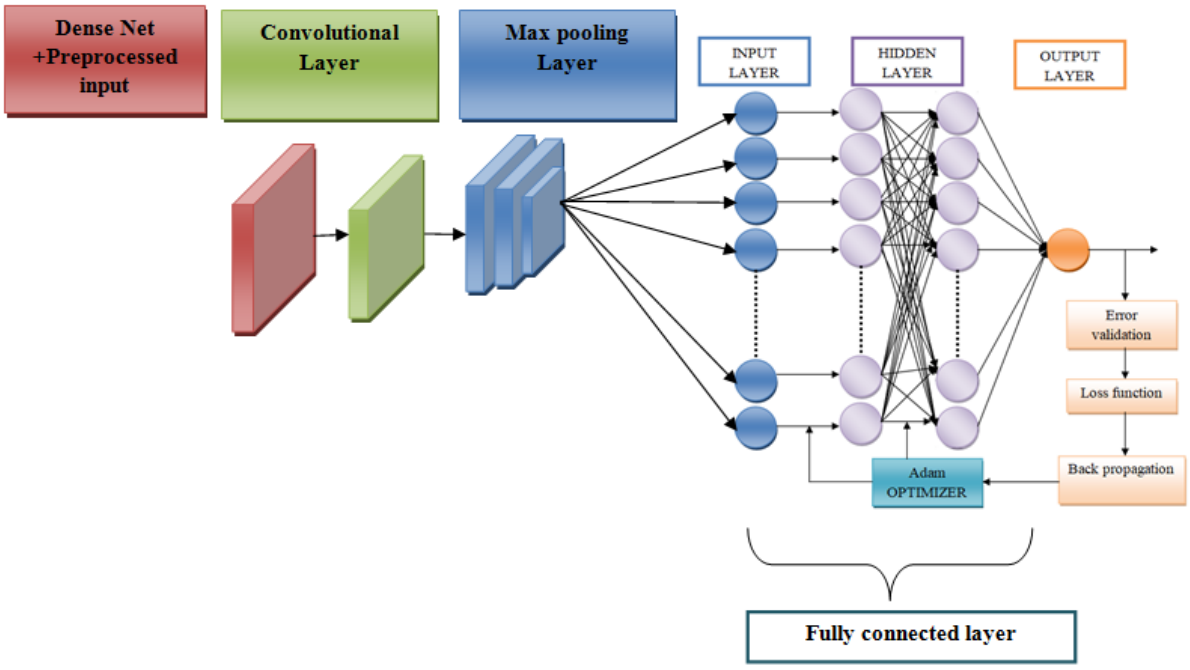


Figure 4: Proposed Deep Ensemble learning

4. Result and discussions

The presentation of various Deep Learning Neural Networks along with proposed Methodology is analysed through various performance metrics. The work is performed in the working platform of Python based DWRU datasets. The detail evaluation is presented in upcoming sections.

4.1. Performances of Different Classifiers

The test accuracy on the augmented data increased from 85% to 90% in comparison with the performance on the measured data, which would not have been possible in the case of over fitting due to synthesized signals. To further explore the proposed method' robustness and its ability to bridge the gap between simple computer simulations and actual experiments, the deep learning model trained on simulation data was compared with pre-existing deep learning models of DBN, CNN, LSTM and GRU in terms of feature extraction. The pre-existing deep learning models are trained and optimized on natural values and have fixed network architectures. The vibration signal dataset that is commonly employed to train the existing pretrained networks is usually a subset of the Image Net database. To verify the performance of the customized deep learning model for autonomous feature extraction from a limited amount of

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

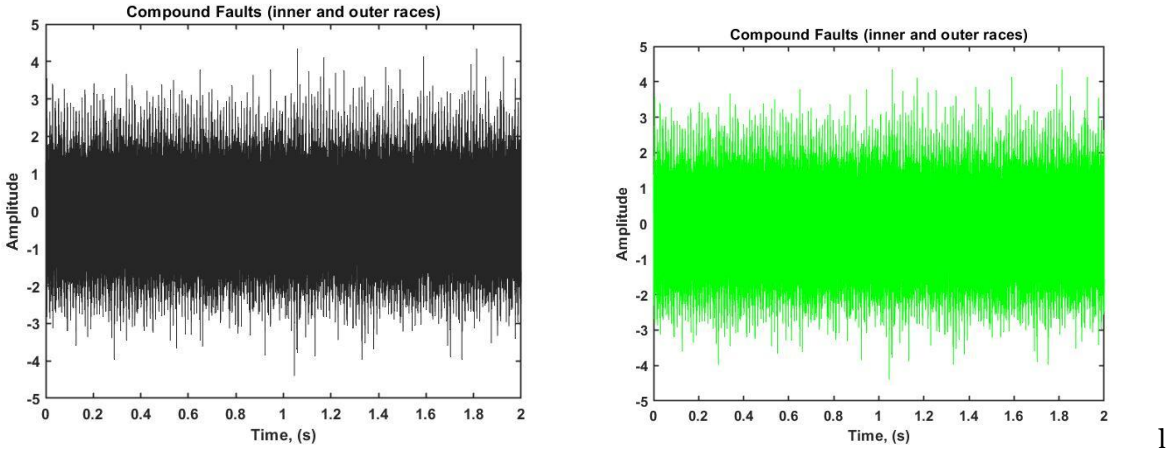
experimental data, Table 1 depicts the performance of the cubic SVM on autonomously extracted features from the simulation model was compared with the performance of DBN, CNN, LSTM and GRU.

Table 1: Comparison of accuracy variation of deep learning methods

Techniques	Diagnostic Accuracy	Mean square error
Proposed	90	0.05
DBN	89	0.06
CNN	88	0.07
LSTM	87	0.08
GRU	85	0.09

The proposed method used equivalent arrangement, for instance, various convolutional blocks and same two fc-layers tending to picture components and one fc-layer for course of action use. Regardless, differentiated and past association designing, Transfer learning goes much further with more CNN layers, which meant the possibility of "outstandingly significant organization" (Simonyan and Zisserman, 2014) giving an orientation to design and capable utilization of CNN. The most indisputable responsibility of the Transfer Learning configuration is to show that by applying little channel size (or piece size) like 1 × 1, 3 × 3 and growing significance of association can really deal with the model execution. Meanwhile, pretrained TL Model on Image Net has brilliant distortion on another instructive assortment, which gives an environment to TL study.

4.1.1.Compound Fault Diagnosis



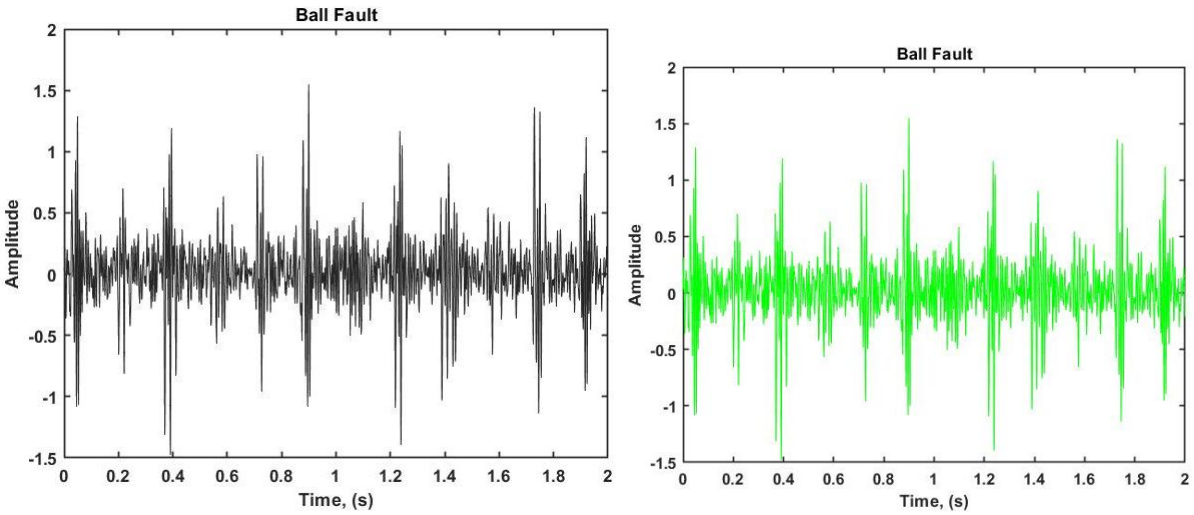
(a)

(b)

Figure 5: Compound Faults Diagnosis Graphical representation

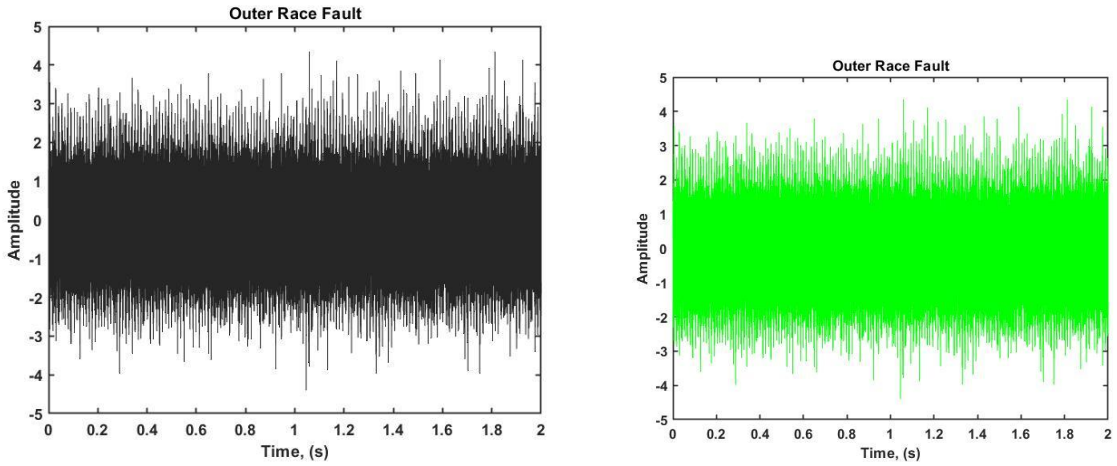
Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



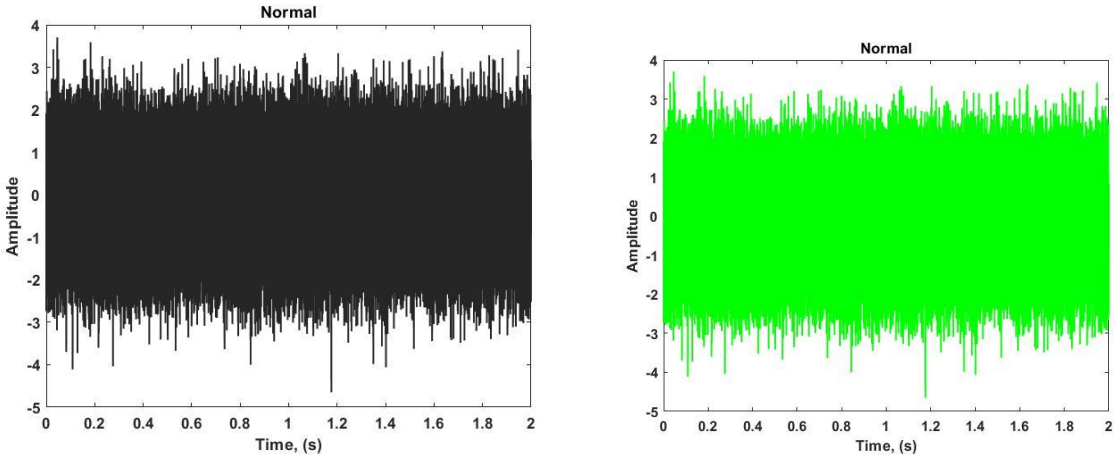
(a) (b)

Figure 6: Balls fault detection graphical representation



(a) (b)

Figure 7: Outer race fault diagnosis graphical representation

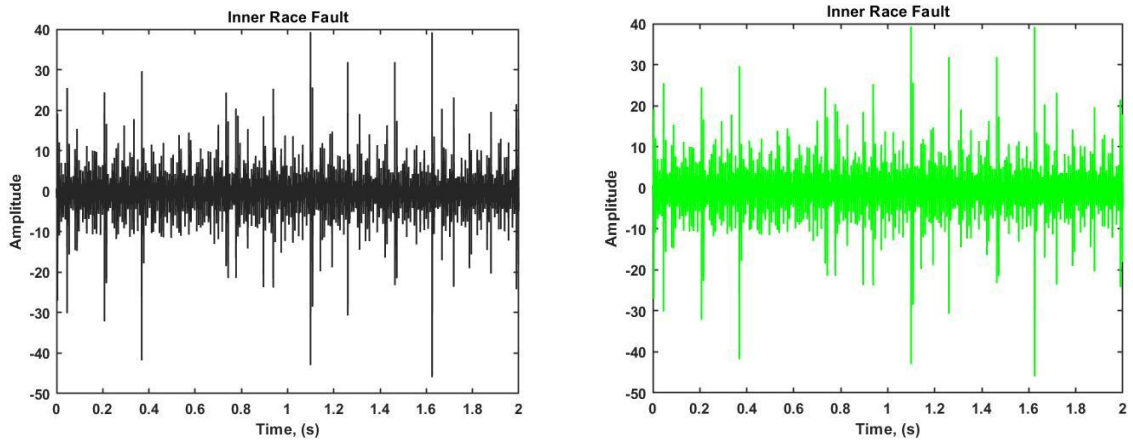


(a) (b)

Figure 8: Normal state fault detection graphical representation

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



(a) (b)

Figure 9: Inner race fault detection graphical representation

The external and inward ring fault highlights after the deconvolution activity are displayed in above Figures. It very well may be seen from the time space chart that the first shock qualities can be decreased partially.

The reverberation band of the external race fault is found. Bandpass sifting is performed on this recurrence band, and afterward the envelope range is investigated. In the envelope range, the trademark recurrence of the external race of the bearing is found. Also, the internal ring fault signal is broken down similarly. The internal ring fault reverberation band is found in the proposed technique and the trademark recurrence of the inward ring fault is tracked down in the envelope range.

In given figure, some high-frequency contents are observed in the scalograms of the normal state, implying either the presence of noise or some other small unavoidable defects alongside the small unbalance in the system. Additionally, comparing the scalograms from the experimental data with their simulated counterparts shows that the scalograms of the experimental data demonstrate more complex behaviour in terms of time-frequency content, which confirms that extremely simple mathematical models cannot replicate the exact dynamic response behaviour of an actual system with and without defects. Furthermore, the health states of normal and oil whirl were not considered in the source simulation data. In the next section, the CNN model pretrained on simulation data is used to automatically extract discriminative features from the scalograms of experimental data.

5. Conclusion

This work proposed a realm and session invariant comprehensive feature extractor using a supervised learning framework of transfer learning. A source simulation domain with three health states was employed to detect, isolate, and quantify five health states in the target experimental domain without minimizing the gap in the response characteristics of the two domains. The source domain was comprised of the simulation model of a few representative health states of the target domain, and simulation models were not required for all prospective health states of the actual target system. The proposed methodology relies on transfer learning; where a customized deep learning model is trained, validated, and tested on a substantial set of simulation data, and then the pretrained model is employed to autonomously extract discriminative features from a small experimental target dataset. This work also discussed the synthetic augmentation of the limited experimental data using virtual sensors, where the output from the virtual sensors was defined in terms of the actual sensors using the concept of coordinate transformation. Synthetic augmentation of the experimental data enhanced the proposed technique' performance in terms of training/validation accuracy. The proposed approach' efficiency is authenticated by likening its results with the pre-remaining profound learning models of DBN, CNN, LSTM and GRU in terms of training, testing, generalization, size of the network, parameters of the network, and computational time. The current approach was found to perform relatively better in terms of generalizability and computation cost with more flexibility for a given engineering problem. The proposed approach autonomously extracts discriminative features from the vibration-based scalograms of a limited experimental dataset and eliminates the need for labour-intensive hand-crafted statistical features. The generalized autonomous discriminative features are robust to variations in the operating conditions, severity levels of different

Research Article

health states, and source-and-target-domain scale. This work could be extended to assess faults in laminated composites, gearboxes, industrial robots, civil infrastructures, etc.

References

[1] Tang, S., Zhu, Y. and Yuan, S., 2021. An improved convolutional neural network with an adaptable learning rate towards multi-signal fault diagnosis of hydraulic piston pump. Advanced Engineering Informatics, 50, p.101406.

[2] Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., Yan, R. and Gryllias, K., 2022. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. Mechanical Systems and Signal Processing, 167, p.108487.

[3] Zhang, X., Zhang, M., Wan, S., He, Y. and Wang, X., 2021. A bearing fault diagnosis method based on multiscale dispersion entropy and GG clustering. Measurement, 185, p.110023.

[4] Zhang, X., Han, B., Wang, J., Zhang, Z. and Yan, Z., 2021. A novel transfer-learning method based on selective normalization for fault diagnosis with limited labeled data. Measurement Science and Technology, 32(10), p.105116.

[5] Jian, C., Yang, K. and Ao, Y., 2021. Industrial fault diagnosis based on active learning and semi-supervised learning using small training set. Engineering Applications of Artificial Intelligence, 104, p.104365.

[6] Hong, D., Bang, S. and Kim, B., 2021. Unsupervised Condition Diagnosis of Linear Motion Guide Using Generative Model Based on Images. IEEE Access, 9, pp.80491-80499.

[7] Ruan, H., Wang, Y., Qin, Y. and Tang, B., 2021, October. An Enhanced Intelligent Fault Diagnosis Method to Combat Label Noise. In 2021 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD) (pp. 1-6). IEEE.

[8] Li, X., Jiang, X., Wang, Q., Yang, L., Wang, Z., Shen, C. and Zhu, Z., 2022. Multi-perspective deep transfer learning model: A promising tool for bearing intelligent fault diagnosis under varying working conditions. Knowledge-Based Systems, p.108443.

[9] Chen, X., Yang, R., Wen, H. and Guan, S., 2021, December. Transfer Learning with Unsupervised Domain Adaptation Method for Bearing Fault Diagnosis. In 2021 CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes (SAFEPROCESS) (pp. 1-6). IEEE.

[10] Han, T., Zhou, T., Xiang, Y. and Jiang, D., 2022. Cross-machine intelligent fault diagnosis of gearbox based on deep learning and parameter transfer. Structural Control and Health Monitoring, 29(3), p.e2898.

[11] Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R. and Chen, X., 2020. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. ISA transactions, 107, pp.224-255.

[12] Li, X., Li, X. and Ma, H., 2020. Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. Mechanical Systems and Signal Processing, 143, p.106825.

[13] Peng, P., Zhang, W., Zhang, Y., Xu, Y., Wang, H. and Zhang, H., 2020. Cost sensitive active learning using bidirectional gated recurrent neural networks for imbalanced fault diagnosis. Neurocomputing, 407, pp.232-245.

[14] Azamfar, M., Li, X. and Lee, J., 2020. Intelligent ball screw fault diagnosis using a deep domain adaptation methodology. Mechanism and Machine Theory, 151, p.103932.

[15] Ma, X., Lin, Y., Nie, Z. and Ma, H., 2020. Structural damage identification based on unsupervised feature-extraction via Variational Auto-encoder. Measurement, 160, p.107811.

[16] LeCun, Y., Haffner, P., Bottou, L. & Bengio, Y. (1999), Object recognition with gradient-based learning, in Proceedings of the Shape, Contour & Grouping in Computer Vision, 823.

[17] Li, B. & Mosalam, K. M. (2013), Seismic performance of reinforced-concrete stairways during the 2008 Wenchuan earthquake, ASCE Journal of Performance of Constructed Facilities, 27(6), 721–30.

[18] Moehle, J. (2014), Seismic Design of Reinforced Concrete Buildings, McGraw Hill Professional, New York, NY.

[19] Morgenthal, G. & Hallermann, N. (2014), Quality assessment of Unmanned Aerial Vehicle (UAV) based visual inspection of structures, Advances in Structural Engineering, 17(3), 289–302.

[20] Guo, T., Wu, L., Wang, C. and Xu, Z., 2020. Damage detection in a novel deep-learning framework: a robust method for feature extraction. Structural Health Monitoring, 19(2), pp.424-442. DOI: 10.1177/1475921719846051

Research Article

[21] Chen, P., Li, Y., Wang, K. and Zuo, M.J., 2020. A novel knowledge transfer network with fluctuating operational condition adaptation for bearing fault pattern recognition. Measurement, 158, p.107739.
[22] Khan, A., Kim, J.S. and Kim, H.S., 2021. Damage Detection and Isolation from Limited Experimental Data Using Simple Simulations and Knowledge Transfer. Mathematics, 10(1), p.80.
[23] Shao, H., Xia, M., Han, G., Zhang, Y. and Wan, J., 2020. Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer convolutional neural network and thermal images. IEEE Transactions on Industrial Informatics, 17(5), pp.3488-3496.
[24] Yan, F., Li, J., Miao, D. and Cao, Q., 2021. Fault diagnosis method of mine hoist braking system based on deep convolution neural network. DOI: 10.21203/rs.3.rs-877486/v1
[25] Dong, Y., Li, Y., Zheng, H., Wang, R. and Xu, M., 2022. A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: Solving the small sample problem. ISA transactions, 121, pp.327-348.
[26] He, Z., Shao, H., Zhong, X. and Zhao, X., 2020. Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. Knowledge-Based Systems, 207, p.106396.
[27] Han, T., Liu, C., Yang, W. and Jiang, D., 2020. Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application. ISA transactions, 97, pp.269-281.
[28] Wang, Z., Liu, Q., Chen, H. and Chu, X., 2021. A deformable CNN-DLSTM based transfer learning method for fault diagnosis of rolling bearing under multiple working conditions. International Journal of Production Research, 59(16), pp.4811-4825.
[29] Zhang, W., Li, X., Jia, X.D., Ma, H., Luo, Z. and Li, X., 2020. Machinery fault diagnosis with imbalanced data using deep generative adversarial networks. Measurement, 152, p.107377.
[30] Mosalam, K. M., Takhirov, S. M. & Park, S. (2014), Applications of laser scanning to structures in laboratory tests and field surveys, Structural Control and Health Monitoring, 21(1), 115–34.
[31] Nagi, J., Ducatelle, F., Di Caro, G. A., Cires¸an, D., Meier, U., Giusti, A., Nagi, F., Schmidhuber, J. & Gambardella, L. M. (2011), Max-pooling convolutional neural networks for vision-based hand gesture recognition, in Proceedings of the IEEE International Conference on Signal & Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 342– 47.
[32] Nair, V., & Hinton, G. E. (2010), Rectified linear units improve restricted Boltzmann machines, in Proceedings of the 27th International Conference Machine Learning (ICML10), Haifa, Israel, 807–14.
[33] Oquab, M., Bottou, L., Laptev, I. & Sivic, J. (2014), Learning and transferring mid-level image representations using convolutional neural networks, in Proceedings of the IEEE International Conference Computer Vision & Pattern Recognition (CVPR), 1717–24.
[34] Pan, S. J. & Yang, Q. (2010), A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–59.
[35] Ribeiro, M. T., Singh, S. & Guestrin, C. (2016), "Why should I trust you?" Explaining the predictions of any classifier, in Proceedings of the International Conference Knowledge Discovery & Data Mining, San Francisco, CA, 1135– 44.
[36] Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1988), Learning representations by back-propagating errors, Cognitive Modeling, 5(3), 696–99.
[37] Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D. & Batra, D. (2016), Grad-cam: why did you say that? Visual explanations from deep networks via gradient-based localization, arXiv:1611.01646. Deep transfer learning for image-based structural damage recognition 21
[38] Sezen, H., Whittaker, A. S., Elwood, K. J. & Mosalam, K. M. (2003), Performance of reinforced concrete buildings during the August 17, 1999 Kocaeli, Turkey earthquake, and seismic design and construction practice in Turkey, Engineering Structures, 25(1), 103–14.
[39] Li, C., Zheng, H., Sun, Y., Wang, C., Yu, L., Chang, C., ... & Liu, B. (2024). Enhancing Multi-Hop Knowledge Graph Reasoning through Reward Shaping Techniques. arXiv preprint arXiv:2403.05801.