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

Research Article

Highly Reliable CI-JSO based Densely Connected Convolutional Networks using Transfer Learning for Fault Diagnosis

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: 20 Mar 2025

Revised: 28 May 2025

Accepted: 22 Jun 2025

ABSTRACT

Fault diagnosis is basic in modern frameworks since early location of issues could save important time at any point as well as decrease support costs. The component extraction interaction of conventional fault diagnosis is tedious and relentless work. As of late, with the fast improvement of the deep learning (DL) technique, it has shown its prevalence with an end-over end process and has been applied to classification and different fields. Somewhat, it settles the impediments of guide component abstraction in the conventional fault diagnosis strategy. Notwithstanding, the accessible preparation information is in many cases restricted, and it will corrupt the exhibition of DL strategies. In order to conquer the existing issues the work has integrated DCNN and TL based on CI-JSO selection strategy for fault diagnosis to deal with various fault types. A signal dispensation strategy that changes over one-layered is first and foremost applied, and it can wipe out the impact of high quality elements. To avoid this type of signal dispensation the work uses MMS-CWT to handle the signal decomposition. Thereafter, to improve the feature training, the selection of feature is carried out using CI-JSO. Furthermore, an ideal DCNN is planned and prepared with the Image Net datasets, which can extricate the undeniable level highlights of monstrous images. At last, TL is additionally evolved to smear the information erudite in the basic information appropriation to the objective information conveyance, which incredibly diminishes the reliance on preparing information and further develops the speculation execution of DCNN. Some famous classification strategies are likewise added to the examination. Results demonstrate the way that the proposed technique can definitively distinguish different fault types and have the most elevated classification precision among different strategies.

Keywords: Deep convolution neural networks (DCNN), MinMax Scalar based continuous wavelet transform (MMS-CWT),Confidence interval based jellyfish search optimization (CI-JSO), Fault diagnosis, transfer learning (TL), signal processing, Pre-processing, Image Net, DenseNet and image classification.

1. Introduction

The fast developing of industrialisation, electrical machines are dependably present in businesses and zapped transportation frameworks. Frequently, these machines work under cruel conditions and complex circumstances, for example, high surrounding temperature, high dampness and over-burden, which can ultimately bring about engine breakdowns that lead to high support costs, serious monetary misfortunes and security dangers [1, 2]. Pivoting hardware is a critical part in modern frameworks which is broadly utilized in mechanical gear. It as a rule works in an unforgiving climate for quite a while. Assuming that a fault happens in pivoting hardware parts, for example, bearing and stuff, it is probably going to influence the typical activity of the whole mechanical gear, or even undermine the wellbeing of labourers [3, 4]. In this manner, to guarantee the functional unwavering quality of turning apparatus and

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

Research Article

keep away from shocking mishaps in modern frameworks, it is important to distinguish deformities and disappointments of the pivoting hardware parts as soon as conceivable precisely [5,6].

Because of the requirement for long haul ceaseless checking and the presence of huge information, AI (ML) - based fault diagnosis techniques stand out. Since ML-based diagnosis techniques can process and use helpful data from sufficient verifiable information, they are viewed as an extremely encouraging and incredible asset [7, 8]. Fault diagnosis can be characterized as a sort of example acknowledgment basically. It comprises of three stages to recognize the mal function types: signal information signal assortment, include extraction, and mal function classification [9, 10]. The last two are the most basic advances which influence the exhibition of mal function operation strategies. The current mal function operation techniques can be for the most part separated into two sorts: the customary information driven-based mal function operation strategy and the deep learning (DL) - based mal function operation technique [11]. The conventional information driven-based mal function operation techniques need to physically remove and select elements.

Conversely, DL-based mal function operation techniques can satisfy the entire course of component extraction, include choice and variations consequently. Signal information data contains an abundance of data, subsequently various sorts of mal functions can be analyzed precisely through the signal information data [12, 13]. The customary information driven-based mal function operation strategy comprises of four stages: gather the signal information data, remove the component from the signal information data, sift through the element which can mirror the mal function state, and info the element into the variations the variations results [15]. A research input the trademark boundaries preparing and testing, joined with the water-greased up harsh machine test to confirm that the technique has high exactness in distinguishing various kinds of mal functions. The author introduced a thought, which joins influence time-recurrence word reference, present moment coordinating, and SVM. The proposed technique can extricate great highlights under a very low data-to-commotion proportion, screen the running status of machine continuously, and successfully identify early breakdown of machine [16, 17]. The researcher thought of a fluffy entropy technique to separate the non-direct elements concealed in the signal information data and utilized the group SVM to arrange numerous kinds of mal functions. The investigation results showed that various kinds of machinemal functions were successfully distinguished [18, 19]. The commitment of this work is that an original CWT-DCNN-TL method is suggested to perform precise mal function recognition for moving direction.

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

2. Literature Review

In 2021, Zhu et al. [20] introduced an appropriate transfer learning strategy for bearing mal function operation under factor working circumstances. L1 regularization was acquainted with stifle over fitting and further develops the speculation capacity of the method, to accomplish the motivation behind transfer method boundaries. As per the application prerequisites of mal function operation under factor load working circumstances, a mal function operation strategy in view of L1 regularization transfer learning was introduced. The method showed that the technique had high exactness and great speculation capacity for mal function operation under factor load working circumstances.

In 2021, Zhang et al. [21] introduced a DL-based open set space variation strategy for apparatus fault diagnosis issues. Ill-disposed learning plan was acquainted with get area invariant highlights from source and target areas. To recognize the objective exception classes, an example level biased component was used, which mirrored the likenesses of the objective occurrences with the foundation modules. It Investigated two commonsense pivoting machine datasets were completed for approvals. It was seen that high cross-area fault judgment difficult precision was accomplished by this technique, and the objective exception classes were successfully perceived.

In 2020, Wang et al. [22] introduced a semi-regulated fault diagnosis approach called DADAN that depended on transfer learning to learn area invariant and discriminative elements. DADAN utilized the occasion based strategy to advance additional discriminative elements from crude signals. Besides, the technique utilized a space ill-disposed ways to deal with adjust the conveyance of the extricated highlights across the source and the objective areas. The security of the preparation cycle was ensured in light of the fact that it utilized the Wasserstein distance, which is a proportion of the dispersion between two spaces.

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

Research Article

In 2021, Ruan et al. [23] showed the inadequacies experienced by the conventional fault diagnosis draws near and introduced a connection based semi-directed technique, which precisely perceived different gearbox fault conditions under restricted named tests. In this technique, the example connection finding and pseudo mark learning system was coordinated into an ever-evolving structure to perform named test increase, and the regulated transfer preparing methodology was acquainted with direct fault classification process. The adequacy and prevalence of this technique were approved in light of two gearbox fault diagnosis tests. Exploratory outcomes showed that this strategy acquired a magnificent and stable fault diagnosis execution when contrasted with conventional fault diagnosis approaches.

In 2020, Pang et al. [24] introduced a smart diagnosis approach in view of DCNN and source signal dataset to work on the productivity and precision of mal function diagnostics of planetary gearboxes. The dataset was valued as the contribution for the CNN models (indicated as dataset-CNN) on the grounds that the dataset permitted nonlinear component improvement and sound decrease. Also, TL (TL) was used with to address the difficulties of CNN challenges. A dataset network was created to analyze the mal functions in a muddled powerful reaction. Since dataset upgraded the symptomatic highlights, this approach significantly worked on the preparation and execution of CNN and TL processes.

In 2020, Mao et al. [25] introduced another transfer fault diagnosis strategy for moving bearing in light of an organized DANN model. This technique presented discriminative data and organized data of different results into DANN. The discriminative capacity and mathematical security of area transformation was be really improved by learning organized data among various medical issue. Subsequently, this technique gave another answer for taking care of the fault diagnosis issue with an inadequate measure of observing information.

In 2021, Lv et al. [26] introduced another wise fault diagnosis strategy in view. The likely connection between various however related mechanical parts was information was created. In this strategy, a deep transfer network (DTN) was utilized to separate transferable highlights from the source and target areas at the same time. Furthermore, the multi-part powerful appropriation variation (MDDA) strategy built a weighted blended bit work, which joined the upsides of various bit capacities to plan the transferable highlights to a bound together element space. The outcomes showed that when unlabeled information in the objective area was obtained, the DTN with MDDA technique distinguished the well being conditions of genuine breeze turbine heading and keep up with great power under boisterous information.

In 2020, Huang et al. [27] introduced a clever DACN to manage the compound mal function analysis of IE near multi-space speculation task, which decoupled the complexmal function into different single mal functions in light of a lot of marked single-mal function information in the cause area and enormous measure of information in the MT areas. The DACN can be summed up from the preparation working circumstances to a new or outrageous. In particular, the assignment of decoupling was accomplished the case layers to build, and the area speculation was carried out through acquainting the antagonistic preparation with adjust the element dissemination across numerous areas.

In 2020, Zheng et al. [28] introduced a diagnosis plot to move bearing under a difficult space speculation situation, in which more expected disparities among different source areas ought to be killed and just ordinary examples of the objective space were accessible during the preparation stage. To accomplish adequate speculation execution, a diagnosis conspire joined some priori diagnosis information and a deep space speculation network for fault diagnosis (DDGFD) was expounded. Through the signal preprocessing steps directed by the deduced diagnosis information, the contributions of DDGFD with an essential steady significance across areas were developed from the vibration signal.

3. Proposed Methodology

Here the diagnosis of fault has been discussed in and detailed manner. The work has integrated DCNN and TL -based approach to predict Fault. The network is trained using training dataset with known labels and then tested it on public datasets. The process flow is illustrated in figure 1.

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

Research Article

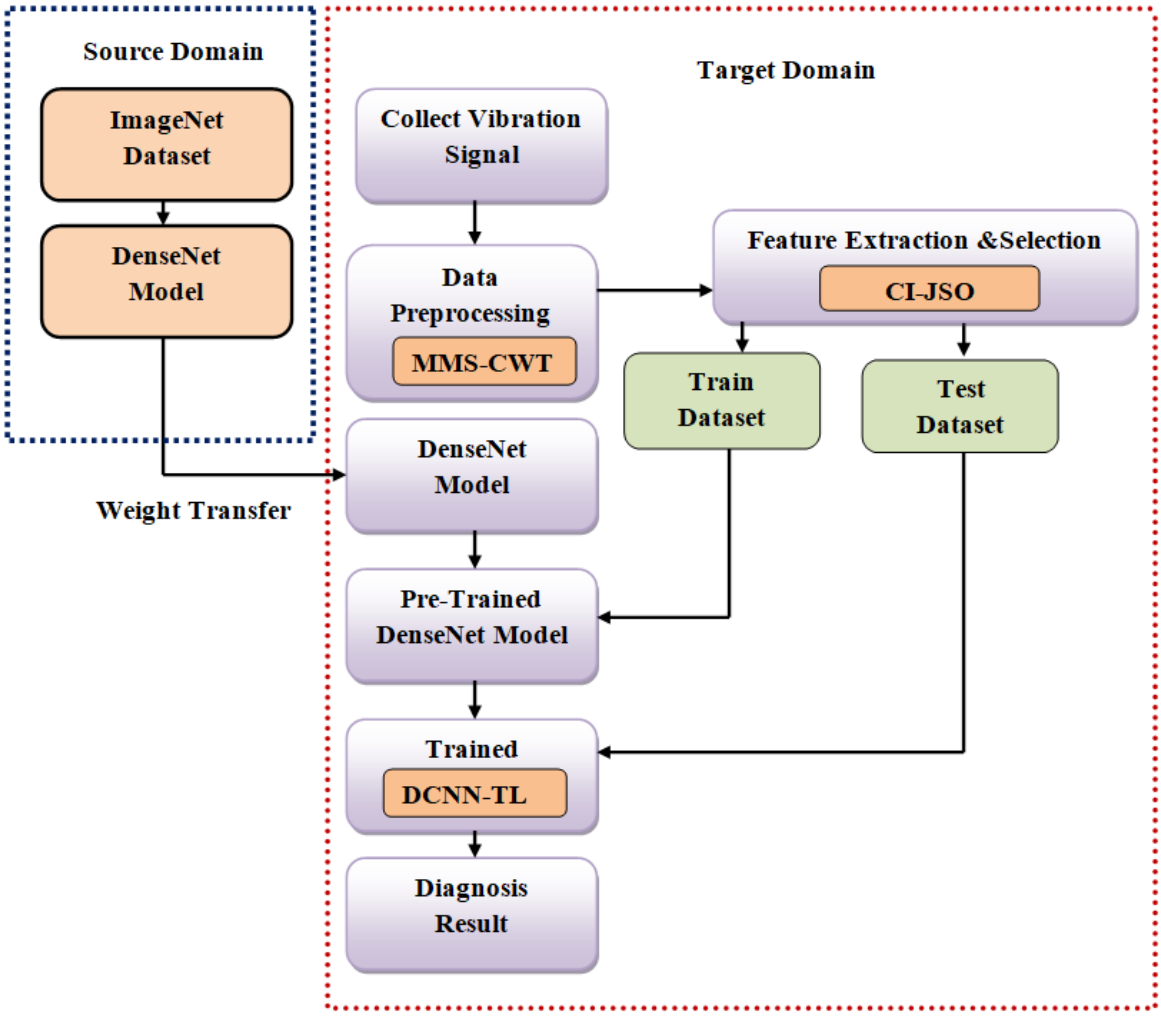


Figure 1: Proposed Methodology

3.1. Source Domain

The hierarchical framework created by Word Net is the basis on which the Image Net is created. Image Net's main objective is to have 50 million high-quality, properly labelled images (500-1000 per synset).

3.1.1. DenseNet Model

The residual neural network, compared to convolutional neural networks and other deep learning approaches, could, to certain level, solve the problem of gradient descent and disappearance. Since each layer has its own weight, the number of parameters in a residual neural network rises as the number of layers increases [29–31]. Fortunately, densely connected convolutional networks (DenseNet) will efficiently resolve this problem. A dense block, a transition layer, and a bottleneck layer compensate DenseNet. An -layer network and a composite function are depicted in figure 2 for the dense block.

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

Research Article

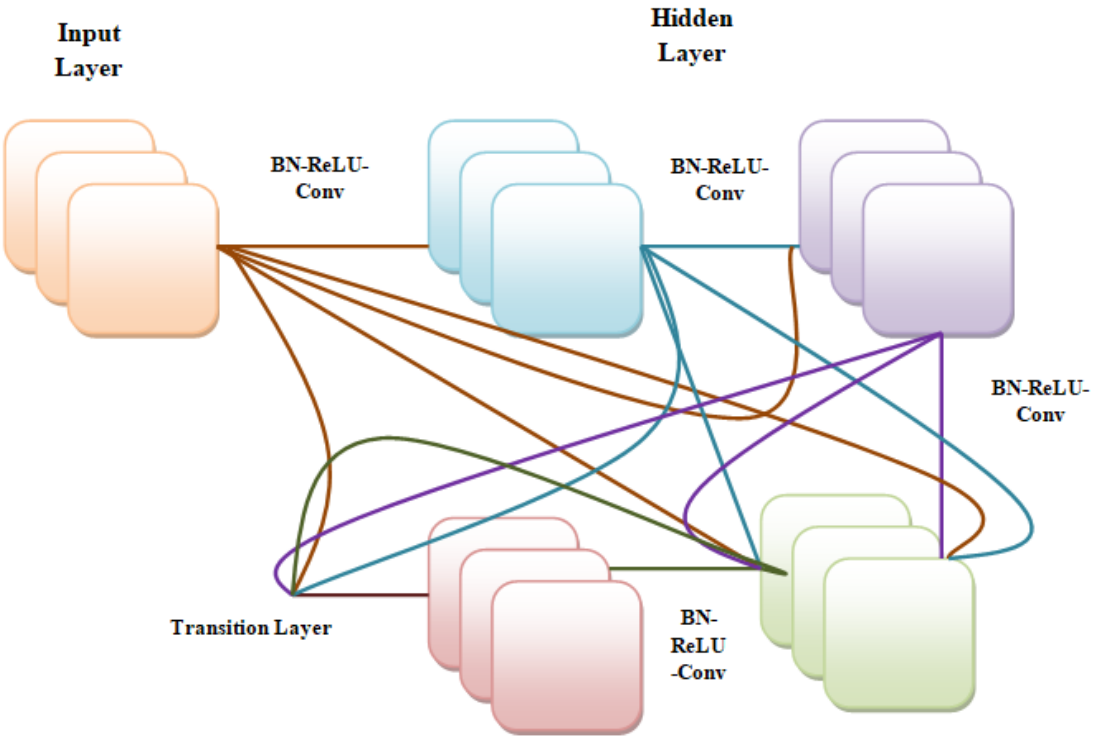


Figure 2: DenseNet Model

Normalization, linear rectification, and convolution constitute the composite function. The l-layer of their l^{th} layer network receives the previous layer's $l-1$ feature map outputs. Its construction formula is as follows:

$$\forall_l = \mathrm{M}\left(\left[\forall_0, \forall_1, \ldots\ldots \forall_{l-1}\right]\right) \tag{1}$$

$$\Phi = \Gamma\left(\forall_i \{w_i\}\right) + \forall \tag{2}$$

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

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

Where $\forall$ and Φ implies layers' input and output vectors, w implies weight in the weight matrix, and Γ implies the 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.

3.2. Target Domain

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

3.2.1. Pre-processing

Initial vibrations have mostly been environmental vibrations and sensor-white-noise. Training datasets for neural networks should only include valid vibration segments and labels. Raw signal data is cleaned to reduce noise and obtain valid segments. Short-term energy identifies each frame in a vibration signal [32, 33]. 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

Research Article

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:

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

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

Where, Ω indicates the fused source data and pre-processed target data.

CNN can analyse 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 [34, 35]. CWT can convert 1-D vibration signals into 2-D time–frequency spectrums that convolution layers may use directly. The scaling coefficient for extending and shrinking the $\wp(t)$ to $\wp(t)\ to\ \wp(t/a)$ leads to massive calculation time and delayed convergence, therefore the work combines MinMax scaling with CWT to avoid the defects for a 1-D vibration signal sequence $\Omega(t)$. This signal's CWT is

$$\Re_{CWT}(S,T) = \frac{1}{\sqrt{|S|}}\int_{-\infty}^{\infty}\Omega(t)\wp\left(\frac{t-T}{S}\right)dt = \left\langle \Omega, \wp_{S,T}\right\rangle \tag{7}$$

Where $\wp(t)$ implies real or complex wavelet-generating function. T implies time-shift coefficient for analysis. $\Omega(t)$ implies scale coefficient for expanding and decreasing of the $\wp(t)$ to $\wp(t)\ to\ \wp(t/a)$ that is expressed as :

.

$$\wp(t) = \frac{\wp(t) - \wp(t)_{\min}}{\left[\wp(t)_{\max} - \wp(t)_{\min}\right]} \tag{8}$$

The continuous wavelet transform in equation (7) is vibration signal inner product $\Omega(t)$ and a set of WBFs $\wp_{S,T}(t)$.

3.2.2. Feature extraction and Selection

As the conventional CNN has the capability to extract features and select best feature , but high resolution image local features are ignored which might comprise of high informative data. This elimination leads to high error rate and also leads to high computation time. In order to avoid that the work extract features and selects the best relevant features in order to strengthen the fault diagnosis [36, 37]. The work has proposed a confidence interval-based Jellyfish (jf) search optimization that extracts and selects the best features to train the Classification model.

The proposed selection method is inspired on jellyfish behaviour in the water. However, due to stochastic location updation, the base chimp optimization technique achieves a high convergence rate. The random update produces insufficient numbers that take a long time to convergence. To prevent this, the work incorporates confidence interval-based Jellyfish (jf) search optimization [38, 39].

Because the ocean provides a lot of nutrients, JF are drawn to it. The path of the water currents $\vec{\forall}_{trend}$ is established by summing all the vectors out of each JF in the marine to the JF in the most excellent place at the time. The ocean current is simulated by Equation (9).

$$\vec{\forall}_{trend} = \frac{1}{n_{pop}}\sum\vec{\forall}_{trend_i} = \frac{1}{n_{pop}}\sum\left(\Gamma^{*} - E_C\Gamma_i\right) = \Gamma^{*} - E_C\frac{\sum\Gamma_i}{n_{pop}} = \Gamma^{*} - E_C\mu \tag{9}$$

$$df = E_C\mu \tag{10}$$

$$\vec{\forall}_{trend} = \Gamma^{*} - df \tag{11}$$

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

Research Article

$$set\ \beta = rand^{\sigma}(0,1) \times \mu \tag{12}$$

Where $n_{pop}$is the amount of JF; $\Gamma^{*}$is the tentacles that presently has the finest locality in the flock; E_C is the factor that influences the attractiveness; μ is the mean locality of all JF; df is the ratio between both the jellyfish's current perfect place and the mean location of all jellyfish. Built on the hypothesis that JF have a typical spatial dispersion in all aspects , $\pm \alpha\beta$ a distance of likewise around mean position incorporates a given probability of all JF, where β is the distribution's standard deviation (SD) [40]. Therefore,

$$df = \alpha \times \beta \times rand^{f}(0,1) \tag{13}$$

$$set\ \beta = rand^{\sigma}(0,1) \times \mu \tag{14}$$

Thereafter, the finest location of the JF is given by

$$\Gamma_i(t+1) = \Gamma_i(t) + CI \times \left(\Gamma^{*} - \alpha \times CI \times \mu\right) \tag{15}$$

$$CI = \left[\overline{\Gamma} + z\frac{s}{\sqrt{n}}, \overline{\Gamma} - z\frac{s}{\sqrt{n}}\right] \tag{16}$$

Where, $\overline{\Gamma}$ denotes the sample mean of the features, z denotes confidence level (CI) value, s indicates SD and n denotes the sample size.

Whenever the swarms are first established, many jellyfish display type A motility. They are progressively exhibiting type B movements throughout time. The motion of jellyfish around their own places is referred to as type A movement, and the associated revised position of each JF is provided by.

$$\Gamma_i(t+1) = \Gamma_i(t) + \vartheta \times rand(0,1) \times \left(u_b - l_b\right) \tag{17}$$

Where U_b and l_b are indicated to be maximum and minimum bound search spaces, respectively; $\vartheta > 0$ is a action coefficient, associated to the length of suggestion around jellyfish's location.

To imitate type B motion, a JF (j) different one than of attention is chosen at random, and the direction of the current is determined by a vector from the JF of interest (i) to the picked JF (j). Whenever the amount of food available to the selected JF (j) exceeds that available to the JF of interest, the latter moves toward the former; if the quantity of food available to the selected JF (j) is less than that available to the JF of interest (i), the latter moves directly away from it (i). As a result, in a swarm, each JF goes toward the improved way to locate food. Eqs 16-18. Replicate the speed and direction and the modified location of a JF. This movement is seen as an efficient use of the information retrieval space.

$$\vec{S} = \Gamma_i(t+1) - \Gamma_i(t) \tag{18}$$

$Where, \vec{S} = rand(0,1) \times \vec{d}$

$$\vec{d} = \begin{cases} \Gamma_j(t) - \Gamma_i(t)\ if\ f(\Gamma_i) \geq f(\Gamma_j) \\ \Gamma_i(t) - \Gamma_j(t)\ if\ f(\Gamma_i) < f(\Gamma_j) \end{cases} \tag{19}$$

Where f is an objective function of location Γ hence,

$$\Gamma_i(t+1) = \Gamma_i(t) + \vec{S} \tag{20}$$

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

Research Article

A period control instrument is utilized to distinguish the sort of movement over the course of time. It manages not even sort An and type B swarm ways of behaving, yet additionally the movements of jellyfish towards another approaching tide. The obligation cycle component is known in full in the accompanying subsection.

Jellyfish populations are typically seeded at random. As a result, this map is used in this investigation

$$\Gamma_{i+1} = \kappa\Gamma_i(1-\Gamma_i), \quad 0 \le \Gamma_0 \le 1 \tag{21}$$

Γ_i denotes location's logistically chaotic value of the i^{th} JF; Γ_0 is used for generating initial inhabitants of jellyfish, $\Gamma_0 \in (0,1)$, and parameter κ is set to 4.0. 3.4. Circumstances of confinement Oceans can be found all over the earth. Because the earth is spherically symmetric, a jellyfish that wanders outside of the confined search region will revert to the reverse bound.

$$\begin{cases} \Gamma'_{i,d} = (\Gamma_{i,d} - U_{b,d}) + l_b(d) \ if \ \Gamma_{i,d} > U_{b,d} \\ \Gamma'_{i,d} = (\Gamma_{i,d} - U_{b,d}) + l_b(d) \ if \ \Gamma_{i,d} < U_{b,d} \end{cases} \tag{22}$$

$\Gamma_{i,d}$ Indicates the location of the i th JF in d^{th} dimension; $\Gamma_{i,d}$ states the rationalized position after examination border line constraint. $U_{b,d}$ and $l_{b,d}$ are d^{th} dimension maximal and minimal in search spaces, respectively.

Finally the outcome from the three FSM is validated and based on intersection baseline features are selected.

3.2.2. Image Classification Using Transfer Learning

Image classification is one of the managed AI issues which intend to arrange the images of a dataset into their particular classifications or marks. Classification of images of different canine varieties is an exemplary image classification issue. The conventional CNN leads to high computation time to get trained for large dataset and leads to high error rate .In order to avoid that the work has developed a Deep Convolutional Neural Network-TL.

The completely associated layer of Deep Convolutional Neural Network is supplanted by GAP to build the pre-prepared Deep Convolutional Neural Network. At long last, the gathered vibration signals are changed over into greyscale images and contribution to the pre-prepared Deep Convolutional Neural Network to acquire the diagnosis results.

CNN uses vector calculus and chain rule.

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

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

Thus, $\left(\frac{\partial\beta}{\partial\Omega}\right)$ has the same size as Ω and its i-th element is $\left(\frac{\partial\beta}{\partial\Omega}\right)_i$: Notably,

$$\left(\frac{\partial\beta}{\partial\Omega^T}\right) = \left(\frac{\partial\beta}{\partial\Omega}\right)^T \tag{24}$$

Additionally, assume is a vector and Ω is a function of κ. Ω's partial derivative with respect to κ is then:

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

Research Article

$$\left(\frac{\partial \Omega}{\partial \kappa^T}\right)_{ij} = \frac{\partial \Omega_i}{\partial \kappa_i} \tag{25}$$

This fractional derivative is a H W matrix whose i-th row/j-th column juncture is $\frac{\partial \Omega_i}{\partial \kappa_i}$. In a chain-like argument, it's clear that a function maps κ to Ω, and another function maps Ω to β. Chain rule determines:

$$\left(\frac{\partial \beta}{\partial \kappa^T}\right), as \left(\frac{\partial \beta}{\partial \kappa^T}\right) = \left(\frac{\partial \beta}{\partial \Omega^T}\right)\left(\frac{\partial \Omega}{\partial \kappa^T}\right) \tag{26}$$

Using a loss function, one could evaluate 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. $\arg\max_i \kappa_i^l$ implies prediction output.

Convolution method:

$$\Omega i^{l+1}, j^{l+1}, d = \sum_{i=0}^{h} \sum_{j=0}^{w} \sum_{k=0}^{d} F_{i,j,k} \times \kappa_{i^{l+1}}^{L} +, ij^{l+1} + j, k \tag{27}$$

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

$$\Omega(\kappa^{l+1}) \text{in } \mathrm{P}^{h^{l+1} \times w^{l+1} \times d^{l+1}}, \quad h^{l+1} = h^l - h + 1, w^{l+1} = w^l - w + 1, d^{l+1} = d \tag{28}$$

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

$$P(\eta \mid n) = \frac{\exp(z_k)}{\sum_{i}^{k} \exp(z_k)}, \tag{29}$$

Normalizing $q(\eta \mid n)$ ground truth distribution across labels: $\sum_k q(\eta \mid n) = 1$ Cross-entropy gives this model's loss:

$$l = \sum_{k=1}^{K} \log(P(\eta)) q(\eta) \tag{30}$$

Cross-entropy loss is differentiable with respect to logits z_k, hence it can be used in deep models' gradient training. The gradients have the simple form:

$$\frac{\partial l}{\partial z_k} = P(\eta) - q(\eta) \tag{31}$$

The above eqn is bounded between −1 and 1. Usually, when minimising the cross entropy, it maximises the accurate label's log-probability. Inception V3 uses a label distribution independent of training samples $u(\eta)$ to avoid overfitting. with a smooth parameter $\in$, where for a training example; the label distribution $q(\eta \mid n) = \vartheta_{\eta|,\Omega}$ is simply replaced by:

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

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

Research Article

It's a combination of $q(\eta \mid n)$ with weights $1-\in$ and the fixed distribution $u(\eta)$ with weights. It becomes after label-smoothing regularisation with uniform distribution $u(\eta)=1/K$ as:

$$q^{l(\eta|n)} = (1-\in)\vartheta_{\eta,n} + \frac{\in}{K}, \tag{33}$$

This is also called cross-entropy.

$$h(q',P) = -\sum_{k=1}^{K} \log(P(\eta))q^{l(k)} = (1-\in)h(q',P) + \in h(u,P) \tag{34}$$

The label-smoothing regularisation is equivalent to applying a single cross entropy loss $h(q,P)$ with a pair of losses $h(q,P)$ and $h(u,P)$, with the second loss penalising the deviation of the predicted label distribution P from the prior u with weight $\frac{\in}{\in+1}$, which is equal to calculating 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.

3.2.2.1. Transfer learning

Transfer learning is a well known deep learning strategy that follows the methodology of utilizing the information that was learned in an undertaking and applying it to tackle the issue of the connected objective errand. In this way, rather than making a neural network without any preparation we "transfer" the learned elements which are essentially the "loads" of the network. To carry out the idea of transfer learning, we utilize "pre-prepared models". Figure 3 shows the flowchart of image classification using Transfer Learning.

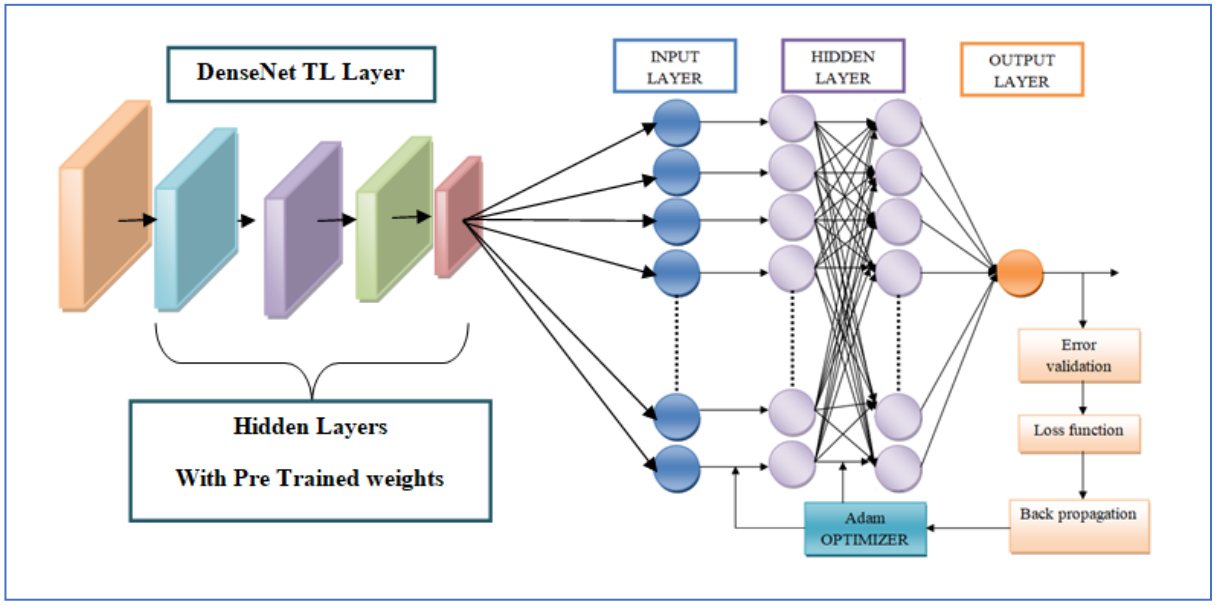


Figure 3: Flowchart of Image Classification using Transfer Learning

4. Result and Discussions

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

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

Research Article

4.1. Comparison with Other Methods

The examination results are displayed in Table 1. The boundary segment in the table alludes that the quantity of loads remembered for each DCNN, while the size section means the necessary space to except these data. It very well may be seen from the outcomes that the model and the DenseNet121 model accomplish the most noteworthy test exactness contrasted and different models. The DBMobileNetV1 model necessitates just preparing after TL. The running season of the CNN-1536-128 model is 60s, which is quicker than the wide range of various models. Be that as it may, this model accomplishes the least classification precision. The DenseNet-TL model can accomplish an exceptionally high arrangement precision of 99.85% with a moderately low preparation season of 96.73s.

Methods	Accuracy (%)	Size (MB)	Time (s)
Proposed DenseNet-TL	99.90	35	33
DenseNet121	97.89	40	40
DBMobileNetV1	96.77	45	50
CNN-1536-128	96.07	60	60

Table 1: Comparison of other Deep Learning Methods with the Proposed Method

4.1.1. Fault Diagnosis Result from Proposed Method

The proposed DCNN-TL method was implemented to Dataset to recognize the machine faults. Since the DL as a rule presents a certain statistical method, we played out the discovery for multiple times for each source signals and accepted the normal as the fault identification result. To comprehend the instrument of the DCNN includes extraction. The Confusion Matrix of the dataset is given in Figure 4.

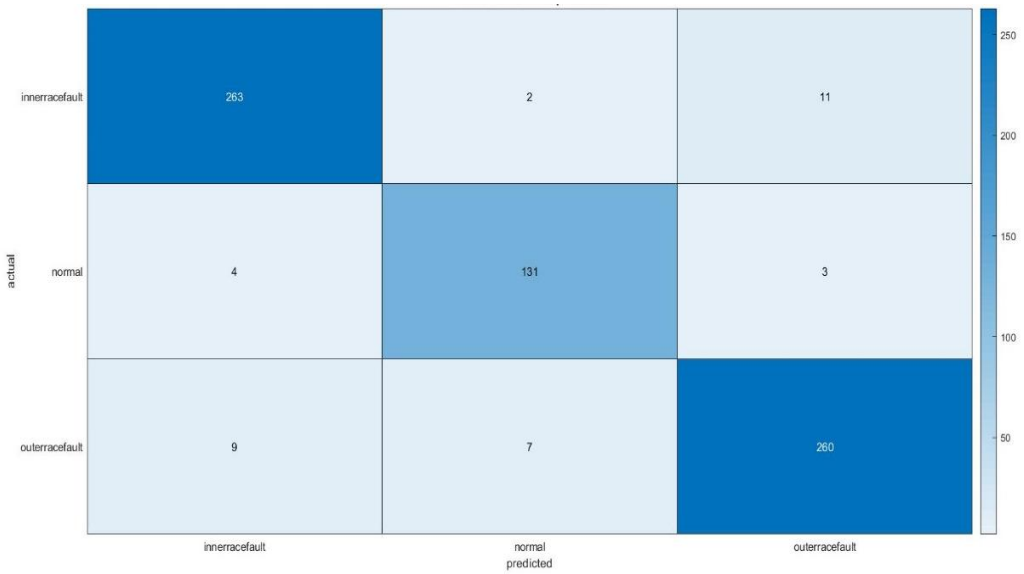
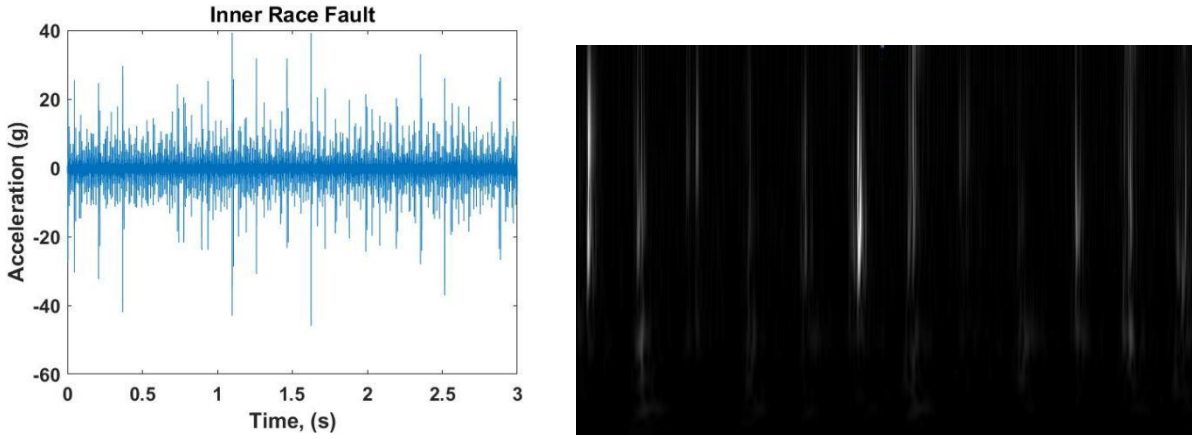


Figure 4: Confusion Matrix of Count of Actual vs. Predicted Layers

It additionally notices void may be seen from the figure that after adequate preparation ages, the preparation misfortune decreases to 0 and the exactness ranges to close to 100%, which shows that the planned DCNN model is sensible. Then, the DCNN highlights were taken care of into the TL model. To imagine the fault distinguishing proof execution, the t-circulated statisticalneighbor installing (t-SNE) [46] was utilized to depict the two-layered guide of the TL yields. DCNN can remove helpful fault elements to further develop the checking execution of the TL. All the ten activity conditions can be obviously recognized after the DCNN-TL processing. In this way, the CWT DCNN-TL model can accomplish great fault detection accuracy.

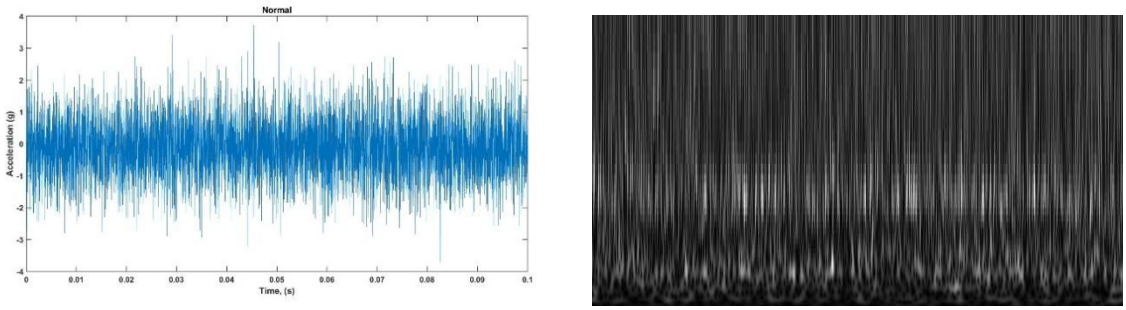
Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



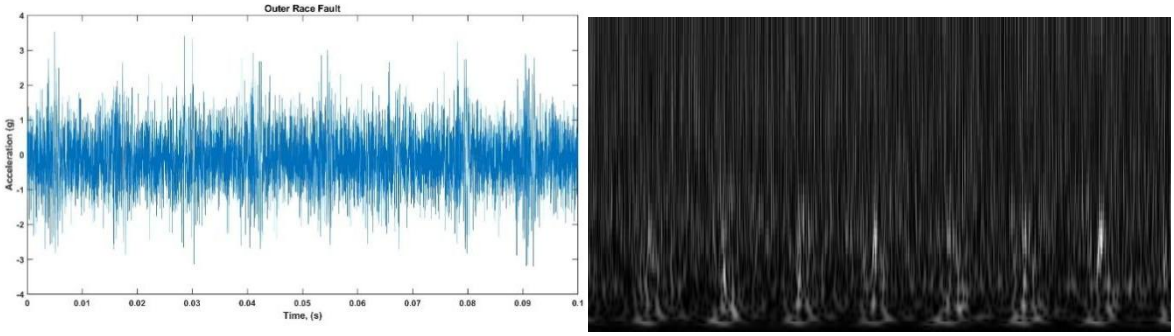
(a) (b)

Figure 5: Inner race fault Pre-processed signal vs. Inner Greyscale image



(a) (b)

Figure 6: Normal race fault Pre-processed signal vs. Normal Greyscale image



(a) (b)

Figure 7: Outer race fault Pre-processed signal vs. Outer Greyscale image

The Figure 5-7 manifest the fault diagnosis after-effect of the datasets in one computation of the ten rehashed times utilizing the disarray grid. As should be visible in the figure, the found accuracy percentage of the ten fault types from the taken signal vibrations is agreeable. The typical found higher percentage of the taken signal vibrations. It ought to underline that for the little size Taken signal vibrations, the found higher percentage is similar to that of the enormous size Dataset IV. Consequently, the suggested CWT-DCNN-TL model is powerful for various sizes of the taken signal vibrations. To feature the adequacy of the suggested technique, correlations with DCNN and TL were done. In this review, either DCNN or TL took on similar design and boundaries as in the DCNN-TL model. The examination results, where TL produces higher found higher percentage than DCNN for little size taken signal vibrations however lower higher percentage for enormous size taken signal vibrations.

Research Article

This is on the grounds that TL is intended for little size data examination while DCNN is brought into the world for large data investigation. For every one of the four gatherings of taken signal vibrations, the proposed DCNN-TL strategy creates the best accuracy percentage against the other two techniques. It records the internet-based acknowledgment season of the three strategies in the model testing stage. One can take note of that the TL is quicker than the DCNN; and the DCNN processes the testing taken signal vibrations quicker than the proposed strategy. By and by, the acknowledgment speed of the relative multitude of three techniques is in a similar size level. The benefit of the CWT DCNN-TL model lies in that it can give more precise fault found to various sizes of taken signal vibrations than the DCNN and TL techniques; this component is particularly significant in the utilizations of the aeronautic trade and the breeze turbines, where the accuracy percentage is substantially more vital than the computational time on the grounds that a little enhancement for the accuracy percentage might save a large number of upkeep cost.

5. Conclusion

DCNN-TL diagnosis strategy consolidating DCNN and TL are proposed for mechanical faults detection, which is able to do significant level component extracted mode and variation. The ideal DCNN built by GAP and deep transfer learning is utilized as an element extractor to upgrade the component learning capacity. Three source vibration signals, specifically the CWRU source vibration signal, SPCP source vibration signal, and UPB source vibration signal, are used for fault diagnosis tests. It very well may be seen from the outcomes that the recommended DCNN-TL strategy can accomplish 100 percent higher percentage on the three source vibration signals. The suggested strategy can build the demonstrative higher percentage of fault diagnosis and lessening the estimation cost. In future work, we will integrate other TL strategies and classifiers into the suggested variation system to work on its strength. In addition, the vibration signal in genuine applications is normally non-standard, and we will confirm it in our future work.

Reference

[1] 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.

[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] Liao, M., Liu, C., Wang, C. and Yang, J., 2021. Research on a rolling bearing fault detection method with wavelet convolution deep transfer learning. IEEE Access, 9, pp.45175-45188.

[4] Zhang, Z., Xu, X., Gong, W., Chen, Y. and Gao, H., 2021. Efficient federated convolutional neural network with information fusion for rolling bearing fault diagnosis. Control Engineering Practice, 116, p.104913.

[5] Zhou, X., Mao, S. and Li, M., 2021. A Novel Anti-Noise Fault Diagnosis Approach for Rolling Bearings Based on Convolutional Neural Network Fusing Frequency Domain Feature Matching Algorithm. Sensors, 21(16), p.5532.

[6] Wei, M., Liu, Y., Zhang, T., Wang, Z. and Zhu, J., 2022. Fault Diagnosis of Rotating Machinery Based on Improved Self-Supervised Learning Method and Very Few Labeled Samples. Sensors, 22(1), p.192.

[7] 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.

[8] Kim, T. and Chai, J., 2021. Pre-Processing Method to Improve Cross-Domain Fault Diagnosis for Bearing. Sensors, 21(15), p.4970.

[9] Nie, X. and Xie, G., 2021. A novel framework using gated recurrent unit for fault diagnosis of rotary machinery with noisy labels. Measurement Science and Technology, 32(5), p.055107.

[10] Shen, Y., Chen, B., Guo, F., Meng, W. and Yu, L., 2021. A Modified Deep Convolutional Subdomain Adaptive Network Method for Fault Diagnosis of Wind Turbine Systems. IEEE Transactions on Instrumentation and Measurement.

[11] Chai, Z., Zhao, C. and Huang, B., 2021. Multisource-refined transfer network for industrial fault diagnosis under domain and category inconsistencies. IEEE Transactions on Cybernetics. DOI: 10.1109/TCYB.2021.3067786

Research Article

[12] Han, T., Li, Y.F. and Qian, M., 2021. A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-11.
[13] Ghorvei, M., Kavianpour, M., Beheshti, M.T. and Ramezani, A., 2021. Spatial Graph Convolutional Neural Network via Structured Subdomain Adaptation and Domain Adversarial Learning for Bearing Fault Diagnosis. arXiv preprint arXiv:2112.06033.
[14] Sun, Z., Yuan, X., Fu, X., Zhou, F. and Zhang, C., 2021. Multi-Scale Capsule Attention Network and Joint Distributed Optimal Transport for Bearing Fault Diagnosis under Different Working Loads. Sensors, 21(19), p.6696.
[15] Shi, Z., Chen, J., Zi, Y. and Zhou, Z., 2021. A novel multitask adversarial network via redundant lifting for multicomponent intelligent fault detection under sharp speed variation. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-10.
[16] de Jesús Rangel-Magdaleno, J., 2021. Induction Machines Fault Detection: An Overview. IEEE Instrumentation & Measurement Magazine, 24(7), pp.63-71.
[17] C. Zhang, Z. Peng, S. Chen, Z. Li, J. Wang, A gearbox fault diagnosis method based on frequency-modulated empirical mode decomposition and support vector machine, Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 232 (2) (2018) 369–380.
[18] A. Maru, A. Dutta, K.V. Kumar, D.P. Mohapatra, Effective Software Fault Localization Using a Back Propagation Neural Network Computational, in: Intelligence in Data Mining, Springer, Singapore, 2020, pp. 513–526.
[19] J. Yu, Y. Xu, K. Liu, Planetary gear fault diagnosis using stacked denoising autoencoder and gated recurrent unit neural network under noisy environment and time-varying rotational speed conditions, Meas. Sci. Technol. (2019).
[20] Zhu, D., Song, X., Yang, J., Cong, Y. and Wang, L., 2021, March. A Bearing Fault Diagnosis Method Based on L1 Regularization Transfer Learning and LSTM Deep Learning. In 2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE) (pp. 308-312). IEEE. DOI: 10.1109/ICICSE52190.2021.9404081
[21] Zhang, W., Li, X., Ma, H., Luo, Z. and Li, X., 2021. Open-set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning. IEEE Transactions on Industrial Informatics, 17(11), pp.7445-7455. DOI: 10.1109/TII.2021.3054651
[22] Wang, Y., Sun, X., Li, J. and Yang, Y., 2020. Intelligent fault diagnosis with deep adversarial domain adaptation. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-9. DOI: 10.1109/TIM.2020.3035385
[23] Ruan, H., Wang, Y., Li, X., Qin, Y., Tang, B. and Wang, P., 2021. A relation-based semisupervised method for gearbox fault diagnosis with limited labeled samples. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-13. DOI: 10.1109/TIM.2021.3052010
[24] Pang, X., Xue, X., Jiang, W. and Lu, K., 2020. An investigation into fault diagnosis of planetary gearboxes using a bispectrum convolutional neural network. IEEE/ASME Transactions on Mechatronics, 26(4), pp.2027-2037. DOI: 10.1109/TMECH.2020.3029058
[25] Mao, W., Liu, Y., Ding, L., Safian, A. and Liang, X., 2020. A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-13. DOI: 10.1109/TIM.2020.3038596
[26] Lv, M., Liu, S., Su, X. and Chen, C., 2021. Deep transfer network with multi-kernel dynamic distribution adaptation for cross-machine fault diagnosis. IEEE Access, 9, pp.16392-16409. DOI: 10.1109/ACCESS.2021.3053075
[27] Huang, R., Li, J., Liao, Y., Chen, J., Wang, Z. and Li, W., 2020. Deep adversarial capsule network for compound fault diagnosis of machinery toward multidomain generalization task. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-11. DOI: 10.1109/TIM.2020.3042300
[28] Zheng, H., Yang, Y., Yin, J., Li, Y., Wang, R. and Xu, M., 2020. Deep domain generalization combining a priori diagnosis knowledge toward cross-domain fault diagnosis of rolling bearing. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-11. DOI: 10.1109/TIM.2020.3016068
[29] Chai, Z., Zhao, C. and Huang, B., 2021. Multisource-refined transfer network for industrial fault diagnosis under domain and category inconsistencies. IEEE Transactions on Cybernetics. DOI: 10.1109/TCYB.2021.3067786

Research Article

[30] Han, T., Li, Y.F. and Qian, M., 2021. A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-11.
[31] 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.
[32] Ghorvei, M., Kavianpour, M., Beheshti, M.T. and Ramezani, A., 2021. Spatial Graph Convolutional Neural Network via Structured Subdomain Adaptation and Domain Adversarial Learning for Bearing Fault Diagnosis. arXiv preprint arXiv:2112.06033.
[33] Sun, Z., Yuan, X., Fu, X., Zhou, F. and Zhang, C., 2021. Multi-Scale Capsule Attention Network and Joint Distributed Optimal Transport for Bearing Fault Diagnosis under Different Working Loads. Sensors, 21(19), p.6696.
[34] Shi, Z., Chen, J., Zi, Y. and Zhou, Z., 2021. A novel multitask adversarial network via redundant lifting for multicomponent intelligent fault detection under sharp speed variation. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-10.
[35] de Jesús Rangel-Magdaleno, J., 2021. Induction Machines Fault Detection: An Overview. IEEE Instrumentation & Measurement Magazine, 24(7), pp.63-71.
[36] C. Zhang, Z. Peng, S. Chen, Z. Li, J. Wang, A gearbox fault diagnosis method based on frequency-modulated empirical mode decomposition and support vector machine, Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 232 (2) (2018) 369–380.
[37] A. Maru, A. Dutta, K.V. Kumar, D.P. Mohapatra, Effective Software Fault Localization Using a Back Propagation Neural Network Computational, in: Intelligence in Data Mining, Springer, Singapore, 2020, pp. 513–526.
[38] J. Yu, Y. Xu, K. Liu, Planetary gear fault diagnosis using stacked denoising autoencoder and gated recurrent unit neural network under noisy environment and time-varying rotational speed conditions, Meas. Sci. Technol. (2019).
[39] L. Eren, T. Ince, S. Kiranyaz, A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier, J. Signal Process. Syst. 91 (2) (2019) 179–189.
[40] X. Liu, X. Yin, M. Wang, Y. Cai, G. Qi, Emotion recognition based on multicomposition deep forest and transferred convolutional neural network, J. Adv. Comput. Intelligence Intelligent Informat. 23 (5) (2019) 883–890.