Convolutional neural network with batch normalisation for fault detection in squirrel cage induction motor

Prashant Kumar | Ananda Shankar Hati

Department of Mining Machinery Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand, 826004, India

Correspondence
Ananda Shankar Hati, Department of Mining Machinery Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand, 826004, India. Email: anandashati@iitism.ac.in

Abstract
Early fault detection in an induction motor is the need of modern industries for minimal downtime and maximum production. A learning technique known as the Convolutional Neural network (CNN) provides automated and reliable feature extraction and selection. Considering these inherent traits of CNN, this study proposes a CNN in combination with batch normalisation (BN)-based fault detection approach for simultaneous detection of bearing fault and broken rotor bars in squirrel cage induction motors (SCIMs). The SCIM vibration signals have different patterns for different defects, and the architecture of CNN is used in this study for fault diagnosis. For an efficient fault feature extraction, the proposed method uses CNN having multiple stacked layers with BN for faster training. In the proposed method, a CNN model with small kernel size is used along with adaptive gradient optimizer and BN to avoid performance degradation and optimum results. For the validation of the proposed technique, a test set-up is used along with different fault conditions. The proposed method is also compared with the existing state-of-the-art methods to illustrate its effectiveness.

1 | INTRODUCTION

Squirrel cage induction motors (SCIMs) are prime mover for modern industries and industrial revolution 4.0. Condition monitoring of SCIM is essential for achieving industrial revolution 4.0. The mechanical faults are significant in SCIMs, and their early detection is essential [1]. Early detection of these faults can help in timely schedule maintenance. Major faults in the induction motor are bearing faults and broken rotor bar (BRB) faults. For monitoring the health of bearing and rotor, vibration monitoring and motor current signature analysis are often used [2–4]. The vibrations induced by bearing fault and BRBs are inherent to the system. Faults are often characterised by characteristic frequency in the vibration spectrum and current spectrum. Each fault has an associated characteristic frequency in vibration and current spectrum for identification. Frequency spectrum analysis is used to detect the faults in induction motors [5]. Fast Fourier transform (FFT) is often used as a viable tool for frequency-domain (FD) based analysis of the time-domain (TD) signal. The amplitude at specific frequencies in the frequency-domain signal can help in detecting anomalies, that is faults. Although FFT analysis is significantly used for fault monitoring and diagnostic in IMs, the algorithm suffers from problems which include overlapping of closely located spectral components, sensitivity to low signal-to-noise ratio, spectral leakage that affects its efficiency [6].

Machine fault diagnosis (MFD) in SCIMs is separated into various stages. It includes data acquisition, pre-processing, feature selection and feature extraction. Finally, Machine Learning (ML) models are executed for fault detection [7,8]. In [9], inter turn fault detection technique for induction motors using feature extraction from current pattern in 2-D space is proposed. Most ML-based MFD requires the features for the input for the ML models. Input data is pre-processed to form an input vector with the help of feature extraction and selection [8]. Sometimes, feature vector having higher dimension is converted to lower dimension by independent component analysis [10] and provides better interpretation of data [11]. The statistical features based on time domain, frequency domain, and time frequency domain are often used for fault analysis [12–15]. ML algorithms like k-nearest neighbour (k-NN) [16], Artificial Neural Network (ANN) [17–19], Support Vector Machines (SVM) [20,21], Decision trees (DT) [22], Bayesian...
Classifier [23], random forest (RF) [24] and deep learning [25] are employed in developing intelligent fault detection system. Feature selection and extraction are important parameter for applying ML methods for fault detection. Features from vibration and current signal can be extracted from the TD signal [26], FD signal [27], and time-frequency domain [28,29]. Statistical features like kurtosis, root mean square, kernel density estimation, crest factor and crest-crest value can be calculated from TD signal [30]. The features from FD signal can be extracted from FFT transformation method [31]. Short-time Fourier transform (STFT) methods [32], wavelet transform (WT) and Dual-tree complex WT can be used to extract features [33]. There are other methods of feature extraction such as Hilbert Huang Transform (HHT) [34], empirical mode decomposition [33] and intrinsic mode function [35]. Authors have proposed a multiple fault detection method in induction motor by extracting features from signal processing tool like Matching Pursuit (MP) and Discrete Wavelet Transform (DWT) and different ML-based classifier [36]. Authors have used high dimensionality reduction technique along with neural network for fault detection in induction motors [37]. Traditionally, ML algorithms have been used for fault detection. ML algorithm classifier accuracy is dependent on feature extraction and selection, and it requires human expertise. The feature extractor has to be redesigned based on the type of faults and their conditions. Conventional ML algorithms have shallow architectures which restrict it from classifying significant nonlinearities and is time-consuming for fault classification [38].

In past few years, deep learning (DL) has gained lot of attention from researchers for fault detection. The deep architecture of the DL algorithm learns the various degrees of information that are linked to various levels of abstraction [39]. DL overcomes the disadvantages of ML classifier by implicitly learning numerous multiplex features from the signal. DL algorithms like Convolutional Neural Network (CNN), Recurrent Neural Network, Deep Belief Network (DBN), Deep Boltzmann Machine (DBM) and Stacked Auto Encoder (SAE). DL has promised a lot in pattern recognition problem, natural language processing, image analysis and machine fault diagnosis [40–42]. In [43], authors have investigated bearing fault detection using STFT and DL. Authors proposed a 1-D CNN model for motor fault detection [44]. A deep CNN has been used for fault classification based on the image generated from raw signals by time-frequency representation [45]. Authors [46] have used CNN for bearing fault and broken bar fault detection. Bearing fault detection has been presented using three DL methods namely, DBN, DBM and SAE [47]. In [48], the authors have proposed a CNN model for bearing fault diagnosis using envelope order spectrum from vibration signals. Authors [49] proposed a fault identification technique using CNN and FFT for bearing fault detection in induction motor. In [50], authors have propounded a bearing fault diagnosis of induction motor using CNN and empirical WT. Authors [51] investigated the bearing fault diagnosis using CNN model with the help of FFT analysis of vibration and utilizing root mean square data from the FFT analysis. In [52], authors investigated the fault detection in motors using CNN and STFT of vibration signals. Authors [53] have proposed a technique for mechanical faults classification using CNN-based hidden Markov models. In [54], authors have proposed a hierarchical CNN for identifying different fault states of rolling bearing. Authors [55] have propounded a bearing anomaly detection with the help of hierarchical adaptive CNN model.

The available techniques can achieve a good result in fault diagnosis, but very few techniques can work directly on raw vibration signals. Many available techniques have poor adaptability to new data obtained from different conditions than that of training data of trained fault detection classifier. It lacks domain adaptability [56]. The maximum number of SCIMs are operating under noisy condition; various pre-processing tools are required for data cleaning. Few methods can only detect fault conditions using raw, noisy signals with high accuracy without signal pre-processing. Most of the CNN models implemented in previously mentioned papers have not more than 10-weighted layers and models using more than 10 layers have not been scrutinized and their performance have not been analysed. Also, available methods have used time-frequency imaging technique or other techniques for image conversion from sensor data. These conversion relies on expert knowledge and proper selection of parameters for image generation. The vibration signals are multifaceted signals in SCIMs operates in noisy environment and harsh conditions. The bearing and BRB fault detection require an efficient and deeper feature learning. The deep architectures of CNN are required for learning these complex and deeper features. Moreover, several studies show that with deeper network, the accuracy tends to saturate and consequently starts degrading [57]. In [58], authors have proposed a sparse DL method for improving the deep network and have studied the fault detection model based on this technique for MFD in motors.

By utilizing the advantages of CNN models with higher number of convolution layer (CL) and improvements with the help of batch normalisation (BN), a fault detection model for BRB and bearing fault detection in SCIMs has been proposed. The contribution of the study is enlisted below:

- CNN-based fault detection model is proposed with more than 15 convolution layers for efficient and autonomous feature learning along with batch normalisation for faster training and avoiding the network degradation problem.
- A simple method for converting vibration data to image is developed for image generation which are used as an input to the proposed fault detection model.
- The major advantage of proposed method is its simple CNN architecture with number of filters doubling through every stack of convolution layers.
- The CNN-based fault detection technique facilitates automatic feature extraction and selection.
- The CNN architecture allows the fault detection model to have good domain adaptability for detecting faults from new data and it is one of the major advantages and novelty of the proposed work.
- The simple proposed vibration data to image conversion method abolishes the need of signal processing tools like
STFT, wavelet transform (WT), and HHT for image generation from the sensor data.

- The proposed method also facilitates end-to-end learning.

The hardware set-up scrutinizes the proposed method with a comprehensive performance evaluation of the proposed method. The results were also compared with the existing state-of-the-art methods for fault detection in SCIM. Section 2 describes CNN and its composition. Proposed CNN-based fault detection methodology has been presented in section 3. Section 4 shows details of test set-up and analysis of the proposed work. Results and discussion have been illustrated in section 5. Finally, section 6 concludes the work.

2 | CONVOLUTIONAL NEURAL NETWORK

CNN belongs to a special class of deep neural network. It differs from conventional neural networks in a sense that it uses convolution in the layers while the traditional method uses matrix multiplication. It has topology like ANN with three layers, namely, input layer, hidden layers and output layers. The hidden layers are an important part of CNN which consists of numerous hidden layer and it includes multiple CLs and sub-sampling layers (SLs). CNN generally uses rectified Linear Unit (RELU) activation function and different CLs and SLs stacked together. CNN model inherently facilitates feature extraction and selection. This section presents different layers of CNN along with mathematical interpretation.

2.1 | Convolution layer

CL consists of numerous kernels having smaller height and width than the input image. Each kernel convolves with the input image to develop an activation map consisting of neurons. Kernels scan at all spatial positions of the image for extracting the locating features and reduces the dimension. The outcomes of convolutional operation are passed through an activation function to generate the output. In the last layer, the activation function is used. The activation function called ReLU has been used as it has non-saturating properties; the gradient is always high (=1) if the neuron activates. The mathematical model of the convolution layer is given below [59]:

\[ x^p_n = f\left( \sum_{m \in T_n} x^{p-1}_{m} \Delta k^p_{mn} + b^p_n \right) \]  

(1)

where \( \Delta \), \( T_n \), \( n \), \( k \) and \( f \) represent the convolution operation, selection of input maps, \( p \)th layer in the network, kernel size with size \( ZXZ \) and non-linear activation function, respectively.

2.2 | Sub-sampling layer

A SL often immediately follows a CL in CNN. Its role is to down-sample the output of a CL along both the spatial dimensions of height and width. It reduces the dimensionality of the input. Sub-sampling gives the representation invariant with a marginal translation. Mathematically, it may be represented by:

\[ x^p_n = f\left( p_{down}(x^{p-1}_n + b^p_n) \right) \]  

(2)

where \( down(.) \) denotes a sub-sampling function. It reduces the dimension of each \( g \)-by-\( g \) block of image and makes its size smaller than the previous layer. Max-sampling or an average-sampling is generally used as a down-sampling function. The max-sampling function segregates the input image from the convolution layer into a set of non-overlapping segments and, for each such segments, the maximum value is given as output. For average-sampling, the output will be the average value.

2.3 | Fully connected layer

Fully connected layers (FLs) perform segregation based on the features of the previous layer. It uses softmax function as an activation function for output. In FL, all neurons are associated to all activation functions of the previous layer. FL gathers all highlights from the previous element map for characterisations. This function assigns decimal probabilities to each class in a multi-class problem. This function is used as a cost function for the output layer to convert the class classification problem. This function converts the output of the preceding layer into the probability of each state when solving the state classification problem. This function is given by:

\[ \sigma(v)_n = \frac{\exp^{v_n}}{\sum_{d=1}^{D} \exp^{v_d}}, \quad n = 1, ..., D \]  

(3)

2.4 | Batch normalisation

The training of deep network is a challenging and complex task as an input of each layer changes during training [60]. It slows training by requiring lower learning rates and precise parameter selection. It makes it difficult to train a model with saturating non-linearities. It leads to internal covariant shift issue and can be resolved by normalising layer. The BN allows the use of high learning rate, reduces internal covariance shift problem, and accelerates the training process. BN is placed right after the convolution layer and before the activation unit. For a BN with \( r \)-dimensional input, \( x = (x_1, x_2, ..., x_r) \), the transformation is given by:
\[ \bar{x}^i = \frac{x^{(i)} - E[x^{(i)}]}{\sqrt{\text{Var}[x^{(i)}]} } \]  

(4)

\[ y^{(i)} = y^{(i)} \bar{x} + \beta^{(i)} \]  

(5)

where \( y^{(i)} \), \( \gamma^{(i)} \), and \( \beta^{(i)} \) denote the output of neuron, scale and shift parameter, respectively. The features are standardized independently in each dimension and accelerate convergence. \( y^{(i)} \) and \( \beta^{(i)} \) ensure that the characteristics of the features remain intact. It is required to calculate the loss function for optimizing the algorithm and for applying the optimization algorithms like stochastic gradient descent, gradient descent with momentum and variable learning rate and Adam stochastic optimization [61]. The loss function is calculated with the help of cross-entropy (CE) between the target data and output data. It is given as:

\[ CE(u, \bar{v}) = -\sum_x u(x) \log \bar{v}(x) \]  

(6)

where CE represents the cross entropy, \( u(x) \) represents target probability distribution and \( \bar{v}(x) \) denotes estimated probability distribution.

3 | CNN-BASED FAULT DETECTION

3.1 | Structure of proposed CNN

This study proposes a CNN-based technique for simultaneous detection of bearing and BRB faults. The fault detection system using CNN consists of a stack of multiple CL with a max-pooling layer in between stacked convolution layer along with BN layer. The data for condition monitoring is taken from accelerometers mounted at different locations on SCIM. The training dataset is used for training and validation dataset is essential for stopping the training before classifier gets over-fitted. The generalised classification property of developed classifier is evaluated with the testing dataset. The developed method has the edge over ML methods as it avoids the use of manual feature extraction and selection. CNN inherently allows the efficient and automatic feature extraction.

The architecture of a CNN model for fault detection is given in Figure 1. Convolutional networks (ConvNets) have achieved success in image classification problems. The structure of proposed model is motivated by VGG19 model [62], which proved its efficiency on image classification. It has 19 layers (16 CL’s, 3 FL’s). The kernel size of 3 × 3 is used throughout the CNN structure having stride of 1. The max-pooling layers are used having stride of 2. Small kernel size in CL helps in reducing the number of parameters despite having depth in the structure. The hyper-parameters of convolution layer and max-pooling layer are tabulated in Table 1 and Table 2, respectively. During the training, input to ConvNets is a fixed size. The information is passed through the stack of CL with a small kernel size. BN layer is also placed after the convolution layer. In one of the layers, 1×1 filter is used as a linear transformation of the input channels. Max-pooling is used in between layers for reducing the dimension of the representative features. Fully connected layers follow the stack of convolution layer with the combination of max-pooling layer and BN layer: 4096 channels are there in two layers and five channels in last layer (five for five states of SCIM). The final layer comprises softmax function. The layout of fault detection scheme is shown in Figure 2. As shown in the Figure 2, the vibration data is acquired from the accelerometers mounted on the body of SCIM. The acquired vibration data is converted to multiple images with the help of proposed 1-D signal to image conversion technique. The dataset of images is divided into the training, test and validation set. The proposed CNN model is trained on the

![Figure 1](image-url)

**Figure 1** Architecture of proposed Convolutional Neural network model with batch normalisation.
images of the training set. The details of dataset are elaborated in the next section.

The hidden layers in VGG19 use ReLU and it reduce the training time as compared to the established model. The three ReLU units make the decision function more discriminative. The kernel of smaller size allows VGG to have a more depth in the network and, more depth that is more weighted layers improve the performance. Also, the architectures are composed of multiple levels of non-linear operation. It has been also found that shallow arrangement of significance 2 requires exponential width to execute a limit that a deep architecture of polynomial width could realize. The flowed and small portrayal learned by deep structures are more efficient than that learned by shallow architecture [63,64]. It is critical to use deep structures other than simply shallow models to get the hang of convincing representation of data. The depth of the networks allows better feature learning [65,66] and BN facilitates the faster convergence.

3.2 1-D vibration signal to image conversion method

The conventional system requires significant pre-processing before analysis as data-driven methods cannot handle the raw data signals. The data pre-processing includes feature extraction and feature selection from raw signals which require human expertise. Selecting appropriate features is an exhaustive work, and it drives the whole fault detection process. The vibration signal is a 1-D signal and is transformed into a suitable form for input to the proposed CNN fault classifier. For developing the images from the vibration data, the vibration data is normalised in the range (0–1). To generate images of size $224 \times 224 \times 3$, the vibration data is divided into multiple segments with 256 data points. Each segment is converted to a 2-D array. This 2-D array is used for developing the images with the help of Python Imaging Library package of the python. Each image requires 256 vibration data points.

4 TEST SET-UP AND ANALYSIS

Simultaneous fault detection is always a challenging task, and it requires precise data for analysis. The proposed CNN model with BN needs sufficient data for training and testing purpose.

4.1 Test set-up

The test set-up consists of 5 kW SCIM with three accelerometers mounted on a different location on its yoke and is
shown in Figure 3. The set-up also includes National Instruments based USB X-series data acquisition (DAQ) system with LabVIEW interface. The acquired data is prepared in workstation having Intel Xeon E3 processor, 64 GB RAM and NVIDIA Quadro P400 graphics processing unit (GPU). The proposed CNN model is developed in Python with the help of Tensorflow and Keras packages. Also, GPU is used for faster computation. Table 3 enlists the details of the test set-up.

Figures 4–6 show the healthy rotor, rotor with one damaged bar, and rotor with two damaged bars, respectively. Figures 7–10 show the healthy bearing and bearing with ball defect (BBDF), bearing with inner race fault (BIRF), and bearing with outer race fault (BORF), respectively. 1-D vibration signals are acquired with the help of three accelerometers placed on the body of SCIM in the drive end using NI DAQ and interfaced with the workstation by LabVIEW software.
The experimental set-up, as shown in Figure 3, has been used for collecting the data. In the experiment, five states of SCIM were considered, namely, healthy (H), BIRF, BORF, BBDF and BRB. For generating images from the vibration data, vibration signals are obtained from the three accelerometers mounted on the drive end of the SCIM at the sampling rate of 10 kHz, and it includes the tri-axial vibration acquired from each accelerometer. Vibration signals are acquired for all the five states under no-load, 25%, 50%, 75% and full-load. Vibration data for each state is randomly fused and used for generating images. In total, 15,000 images were generated with the help of
selected hyper-parameter. There is no standardization available for a selection of hyper-parameter. Accuracy depends on the kernel size selection, number of stacked layers, padding and pooling. For all CLs, kernel size is $3 \times 3$. After passing through stacked CLs and max-pooling layer, the size of the output data is reduced. The final fully connected layer has a softmax function.

5 | RESULTS AND DISCUSSION

5.1 | Results

The adjustment of hyper-parameter is essential for developing an efficient classifier. After the adjustment of hyper-parameters and layers, input images from the training set are fed to the first layer and then processed to subsequent layers as per the algorithm. The 16 weighted layers CNN with BN-based fault classifier is trained with the training dataset. The developed technique performance is evaluated with the help of test dataset. The training and validation accuracy with respect to epoch is shown in Figure 11. The confusion matrix (CM) for the proposed CNN model is given in Figure 12. CM shows the performance of the proposed CNN-based classifier. The overall accuracy of the classifier is 99.6%. The performance evaluation indices like the precision ratio $p$, recall ratio $r$, and $F_1$ score can estimate the performance of the proposed classifier and are given by:

$$p = \frac{TP}{TP + FP}$$

(7)
TABLE 4 Performance indices of different fault labels of proposed CNN and batch normalisation-based model

| Label | BBDF (%) | BIRF (%) | BORF (%) | BRB (%) | H (%) |
|-------|----------|----------|----------|---------|-------|
| \(p\) | 100      | 100      | 98.04    | 100     | 100   |
| \(R\) | 100      | 100      | 99       | 99      | 99    |
| F1    | 100      | 100      | 99.01    | 99.49   | 99.49 |

Abbreviations: BBDF, bearing with ball defect; BIRF, bearing with inner race fault; BORF, bearing with outer race fault; BRB, broken rotor bar; CNN, Convolutional Neural network; H, healthy.

TABLE 5 Comparison of performance indices of different fault labels using SVM, k-NN, DT and Shallow CNN

| Label       | BBDF (%) | BIRF (%) | BORF (%) | BRB (%) | H (%) |
|-------------|----------|----------|----------|---------|-------|
| SVM         |          |          |          |         |       |
| \(\rho\)    | 74.46    | 81.05    | 73.34    | 69.48   | 78.43 |
| \(r\)       | 70.18    | 74.25    | 70.08    | 75.18   |       |
| F1          | 76.77    | 73.79    | 69.83    | 76.77   |       |
| k-NN        |          |          |          |         |       |
| \(\rho\)    | 69.51    | 62.59    | 69.23    | 69.68   | 80.48 |
| \(r\)       | 71.33    | 69.55    | 71.08    | 74.18   |       |
| F1          | 77.20    | 69.39    | 70.51    | 77.20   |       |
| DT          |          |          |          |         |       |
| \(\rho\)    | 71.76    | 64.33    | 70.73    | 72.75   | 76.57 |
| \(r\)       | 68.25    | 65.25    | 72.14    | 73.15   | 77.25 |
| F1          | 69.98    | 64.78    | 71.43    | 72.95   | 76.90 |
| Shallow CNN |          |          |          |         |       |
| \(\rho\)    | 92.25    | 93.50    | 98.13    | 93.51   | 92.32 |
| \(r\)       | 92.67    | 92.67    | 93.56    | 92.88   | 93.56 |
| F1          | 92.46    | 93.08    | 95.79    | 93.19   | 92.93 |

Abbreviations: BBDF, bearing with ball defect; BIRF, bearing with inner race fault; BORF, bearing with outer race fault; BRB, broken rotor bar; CNN, Convolutional Neural network; DT, decision trees; H, healthy; k-NN, k- nearest neighbour; SVM, support vector machine.

\[ r = \frac{TP}{TP + FN} \quad (8) \]
\[ F1 = \frac{2\rho \times r}{\rho + r} \quad (9) \]

where TP, FP, and FN denote the true positive samples, false-positive samples, and false-negative samples, respectively. TP represents the positive samples which are classified as positive. FP represents the negative samples that are classified as positive. FN represents the positive samples that are classified as negative. The positive samples mean the sample which belongs to current fault type and negative samples mean the sample which do not belong to current fault type. The performance indices of each fault label are given in Table 4. As it can be seen from Table 4, the value of \(\rho\), \(r\) and \(F1\) score is well over 95%, which demonstrates that probability of misclassification of two faults is well below 0.01%. Even if any fault is misclassified, overall accuracy will still improve with consequent fault classification. The proposed method abolishes the use of hand-crafted features. The 16 weighted layer along with BN achieves the high accuracy in the fault classification with minimal effort. The advantage of the proposed method is also demonstrated by comparing it with other AI tools like SVM, k-NN and DT.

5.1.1 | Comparison with the ML-based models and shallow CNN for fault detection

For comparing with the ML algorithm-based fault detection, the statistical features in the time and frequency domain were used for fault diagnosis using SVM, kNN and DT. The time-domain features like mean, variance, crest, kurtosis, skewness, root mean square and shape factor and frequency domain features, namely, crest, kurtosis, variance and mean energy were used. The value of \(\rho\), \(r\) and \(F1\) is tabulated in Table 5. The performance indices in Table 5 show the performance of SVM, k-NN, and DT-based classifier. The accuracy of SVM, k-NN, and DT based classifier are 75.2%, 69.96%, and 71.21%, respectively. Also, the values of \(\rho\), \(r\) and \(F1\) are below 75%, which reflects that chances of misclassification of two states are on the higher side. The conventional ML approach requires feature extraction and selection, which is a challenging task and requires past information and expertise. A shallow CNN network is also considered for showing the efficacy of the proposed method. Same dataset which is utilized for the proposed CNN model is utilized for assessing the shallow CNN model. The shallow CNN model is comprising following layers:

- \(2 \times (C + \text{ReLU})\)
- Pooling layer
- \(3 \times (C + \text{ReLU})\)
- Pooling layer
- Fully Connected layer with softmax function

This model can be considered shallow with respect to the proposed CNN model. The prediction accuracy of shallow CNN model is 93.06% which is lower than proposed CNN model. The values of \(\rho\), \(r\) and \(F1\) (Table 5) for shallow CNN model are below 95%, which depict that probability of misclassification of two states is higher. The proposed method overcomes the pitfalls of traditional learning methods and simple CNN model, and proposed CNN model allows efficient learning along with the acceleration of process by BN. The combination of CNN with multiple CLs and pooling layers along with BN ensures the precise and fast fault classification in SCIM.

5.1.2 | Performance evaluation with bearings dataset

The bearing dataset from the Case Western Reserve University (CWRU) Bearing Data Centre [67] was analysed for
the application of CWRU bearing dataset is given in Figure 13. CM affirms the performance of the proposed CNN put together classifier with respect to CWRU bearing dataset. The overall accuracy of the proposed classifier is 99.6%. The performance indices of each fault label is given in Table 6. It is visible from the CM in Figure 13 that 35% samples out of 3600 samples are classified rightly and over all accuracy is 99.61% using the proposed CNN method. Also, it can be seen from Table 6, the values of $p$, $r$ and F1 score are well over 98%, which demonstrate that probability of misclassification of two states is below 0.01%. Even if any fault is misclassified, overall accuracy will still improve with consequent fault classification.

5.2 | Discussions

The superiority of proposed CNN model with batch normalisation is also illustrated by the fact that within 50 epochs as shown in Figure 11, the model achieves a good accuracy with minimal loss. Achieving good efficiency at fewer epochs ensures that the developed technique is optimised and is effective for fault detection in SCIM. Also, the ability to apply this technique to raw signals makes it more reliable and less vulnerable as compared to other techniques. As data is in huge amount, the training takes time; however, it is worth investing significant time in training for an accurate model. The proposed CNN model is implemented on the test set-up data as well as Case Western Reserve University Bearing Data Centre dataset, and the efficiency in both cases is more than 99.50%. It shows the robustness and adaptability of the proposed CNN model. Also, the proposed CNN model is compared with conventional ML models and shallow CNN model for performance evaluation and comparison.

6 | CONCLUSION

This study proposes a novel fault detection technique for detecting the bearing faults and broken bar fault using CNN model with batch normalisation. The application of DL improved the performance and accuracy of fault diagnosis. The CNN model with higher number of CL allowed it to adaptively mine the data for identifying dominant fault features and classify the healthy and different fault states of SCIM with reasonable accuracy. The main contributions of the developed method are enlisted below:

- Autonomous fault feature learning
- Better feature learning owing to higher number of convolution layers
- Independence from signal processing tools like STFT, WT, and HHT for image generation from the sensor data
- End-to-end learning capabilities and better domain adaptability

The comparative investigation with the ML methods and shallow CNN model testify the performance of proposed...
method in learning feature without human expertise and prior knowledge of any signal processing technique. Also, the proposed method is tested on the CWRU bearing dataset and the accuracy is good. It makes the proposed method a convenient tool for fault diagnosis. The stacked CI with small kernel size allowed the signal to be analysed with precision and accuracy. The model can also be extended to other faults detection.

**REFERENCES**

1. Tavner, P., et al.: Condition monitoring of rotating electrical machines, vol. 56. IET (2008)
2. Delgado Arredondo, P.A., et al.: Comparative study of time-frequency decomposition techniques for fault detection in induction motors using vibration analysis during startup transient. Shock Vib. 2015 (2015)
3. Picazo Roedenas, M.J., et al.: Combination of noninvasive approaches for general assessment of induction motors. IEEE Trans. Ind. Appl. 51(3), 2172–2180 (2014)
4. Thomson, W.T., Fenger, M.: Current signature analysis to detect induction motor faults. IEEE Ind. Appl. Mag. 7(4), 26–34 (2001)
5. Lacey, SJ.: An overview of bearing vibration analysis. Mainten. Asset. Manag. 23(6), 32–42 (2008)
6. Madisetti, V.: ‘The digital signal processing handbook, -3 volume set’. CRC Press (2009)
7. Jardine, A.K., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. 20(7), 1483–1510 (2006)
8. Yang, H., Mathew, J., Ma, L.: Fault diagnosis of rolling element bearings using basis pursuit. Mech. Syst. Signal Process. 19(2), 341–356 (2005)
9. Eftekhari, M., et al.: Online detection of induction motor’s stator winding short-circuit faults. IEEE Syst. J. 8(4), 1272–1282 (2013)
10. Cao, L., et al.: A comparison of pca, lpc, and ica for dimensionality reduction in support vector machine. Neurocomputing. 55(1-2), 321–336 (2003)
11. LeCun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE. 86(11), 2278–2324 (1998)
12. Lin, S.W., et al.: Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Syst. Appl. 35(4), 1817–1824 (2008)
13. Hsu, H., et al.: Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. Mech Syst. Signal Process. 18(3), 625–644 (2004)
14. Siedlecki, W., Sklansky, J.: A note on genetic algorithms for large-scale feature selection. In: Chen, C.H., Pau, L.F., Wang, P.S.P. (eds.) Handbook of Pattern Recognition and Computer Vision, pp. 88–107. World Scientific (1993)
15. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. 40(1), 16–28 (2014)
16. Casimir, R., et al.: The use of features selection and nearest neighbors rule for faults diagnostic in induction motors. Eng. Appl. Artif. Intell. 19(2), 169–177 (2006)
17. Su, H., Chong, K.T.: Induction machine condition monitoring using neural network modeling. IEEE Trans. Ind. Electron. 54(1), 241–249 (2007)
18. Han, T., et al.: Fault diagnosis system of induction motors based on neural network and genetic algorithm using stator current signals. Int. J. Rotating Mach. 2006, 13 (2006)
19. Sun, W., et al.: A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement. 89, 171–178 (2016)
20. Martínez Morales, J.D., Palacios Hernández, E.R., Campos Delgado, D.: Multiple-fault diagnosis in induction motors through support vector machine classification at variable operating conditions. Electr. Eng. 100(1), 59–73 (2018)
21. Liu, R., et al.: Artificial intelligence for fault diagnosis of rotating machinery: a review. Mech. Syst. Signal Process. 108, 34–47 (2018)
22. Li, Y., et al.: Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine. J. Sound Vib. 428, 72–86 (2018)
23. Baraldi, P., et al.: Comparing the treatment of uncertainty in bayesian networks and fuzzy expert systems used for a human reliability analysis application. Reliab. Eng. Syst. Saf. 138, 170–193 (2015)
24. Quiroz, J.C., et al.: Fault detection of broken rotor bar in ls-pmsm using random forests. Measurement. 116, 273–280 (2018)
25. Janssens, O., et al.: Convolutional neural network based fault detection for rotating machinery. J. Sound Vib. 377, 331–345 (2016)
26. Samanta, B., Al Balushi, K.: Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mech. Syst. Signal Process. 17(2), 317–328 (2003)
27. Malhi, A., Gao, R.X.: Pca-based feature selection scheme for machine defect classification. IEEE Trans. Instrum. Meas. 53(6), 1517–1525 (2004)
28. Lou, X., Loparo, K.A.: Bearing fault diagnosis based on wavelet transform and fuzzy inference. Mech. Syst. Signal Process. 18(5), 1077–1095 (2004)
29. Yen, G.G., Lin, K.C.: Wavelet packet feature extraction for vibration monitoring. IEEE Trans. Ind. Electron. 47(3), 650–667 (2000)
30. Prieto, M.D., et al.: Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. IEEE Trans. Ind. Electron. 60(8), 3398–3407 (2012)
31. Lin, H.C., et al.: Bearing vibration detection and analysis using enhanced fast fourier transform algorithm. Adv. Mech. Eng. 8(10), 1687814016675080 (2016)
32. Li, C., et al.: Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement. ISA Trans. 60, 274–284 (2016)
33. Van, M., Kang, H.J.: Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition. Proc. IME C J. Meeh. Eng. Sci. 230(2), 291–302 (2016)
34. Osman, S., Wang, W.: A morphological hilbert-huang transform technique for bearing fault detection. IEEE Trans. Instrum. Meas. 65(11), 2646–2656 (2016)
35. Pandya, D., Upadhyay, S., Harsha, S.P.: Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using apf-knn. Expert Syst. Appl. 40(10), 4137–4145 (2013)
36. Ali, M.Z., et al.: Machine learning-based fault diagnosis for single-and multi-faults in induction motors using measured stator currents and vibration signals. IEEE Trans. Ind. Appl. 55(3), 2378–2391 (2019)
37. Saucedo Dorantes, J.J., et al.: Multifault diagnosis method applied to an electric machine based on high-dimensional feature reduction. IEEE Trans. Ind. Appl. 53(3), 3086–3097 (2016)
38. Jia, F., et al.: Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech. Syst. Signal Process. 72, 303–315 (2016)
39. Deng, L.: A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans. Signal and Inform. Process. 3, 29 (2014)
40. Szegedy, C., et al.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826 (2016)
41. Greenspan, H., Van Ginneken, B., Summers, R.M.: Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans. Med. Imag. 35(5), 1153–1159 (2016)
42. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature. 521(7553), 436 (2015)
43. He, K., et al.: Deep learning based approach for bearing fault diagnosis. IEEE Trans. Instrum. Meas. 65(1), 7067–7075 (2016)
45. Verstraete, D., et al.: Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. Shock Vib. 2017, 17 (2017)
46. Lee, J.H., Pack, J.H., Lee, I.S.: Fault diagnosis of induction motor using convolutional neural network. Appl. Sci. 9(15), 2950 (2019)
47. Chen, Z., et al.: Deep neural networks-based rolling bearing fault diagnosis. Microelectron. Reliab. 75, 327–333 (2017)
48. Zhu, D., Zhang, Y., Zhao, L.: Fault diagnosis method for rolling element bearing with variable rotating speed using envelope order spectrum and convolutional neural network. J. Intell. Fuzzy Syst. 37(2), 3027–3040 (2019)
49. Pandarakone, S.E., et al.: Deep neural network based bearing fault diagnosis of induction motor using fast fourier transform analysis. In: 2018 IEEE energy conversion congress and exposition (ECCE), pp. 3214–3221. IEEE (2018)
50. Hsueh, Y.M., et al.: Fault diagnosis system for induction motors by cnn using empirical wavelet transform. Symmetry. 11(10), 1212 (2019)
51. Li, S., et al.: An ensemble deep convolutional neural network model with improved ds evidence fusion for bearing fault diagnosis. Sensors. 17(8), 1729 (2017)
52. Wang, L.H., et al.: Motor fault diagnosis based on short-time fourier transform and convolutional neural network. Chin. J. Mech. Eng. 30(6), 1357–1368 (2017)
53. Wang, S., et al.: Convolutional neural network-based hidden markov models for rolling element bearing fault identification. Knowl. Base. Syst. 144, 65–76 (2018)
54. Lu, C., Wang, Z., Zhou, B.: Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Adv. Eng. Inf. 32, 139–151 (2017)
55. Guo, X., Chen, L., Shen, C.: Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement. 93, 490–502 (2016)
56. Zhang, W., et al.: A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors. 17(2), 425 (2017)
57. Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks (2015). arXiv preprint arXiv:150500387
58. Sun, C., et al.: Sparse deep stacking network for fault diagnosis of motor. IEEE Trans. Indust. Inform. 14(7), 3261–3270 (2018)
59. Hoang, D.T., Kang, H.J.: Rolling element bearing fault diagnosis using convolutional neural network and vibration image. Cognit. Syst. Res. 53, 42–50 (2019)
60. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift (2015). arXiv preprint arXiv:150203167
61. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014). arXiv preprint arXiv:14126980
62. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint arXiv:14091556
63. Bengio, Y.: Learning deep architectures for AI. Now Publishers Inc (2009)
64. Zhong, G., Ling, X., Wang, L.N.: From shallow feature learning to deep learning: benefits from the width and depth of deep architectures. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 9(1), e1255 (2019)
65. Bengio, Y., Delalleau, O.: On the expressive power of deep architectures. In: International conference on algorithmic learning theory, pp. 18–36. Springer (2011)
66. Cohen, N., Sharir, O., Shashua, A.: On the expressive power of deep learning: a tensor analysis. In: Conference on learning theory, pp. 698–728 (2016)
67. Loparo, K.: ‘Case western reserve university bearing data center’, Bearings Vibration Data Sets, pp. 22–28. Case Western Reserve University (2012). https://csegroupscaseedu/bearingdatacenter/pages/download-data-file

How to cite this article: Kumar P, Shankar Hati A. Convolutional neural network with batch normalisation for fault detection in squirrel cage induction motor. IET Electr. Power Appl. 2021;15:39–50. https://doi.org/10.1049/elp2.12005