MDPI

Article

Efficient DCNN-LSTM Model for Fault Diagnosis of Raw Vibration Signals: Applications to Variable Speed Rotating Machines and Diverse Fault Depths Datasets

Muhammad Ahsan [1,*] and Mostafa M. Salah [2]

1   Department of Measurements and Control Systems, Silesian University of Technology, 44-100 Gliwice, Poland
2   Electrical Engineering Department, Future University in Egypt, Cairo 11835, Egypt;
    mostafa.abdulkhalek@fue.edu.eg
*   Correspondence: muhammad.ahsan@polsl.pl

Abstract: Bearings are the backbone of industrial machines that can shut down or damage the whole process when a fault occurs in them. Therefore, health diagnosis and fault identification in the bearings are essential to avoid a sudden shutdown. Vibration signals from the rotating bearings are extensively used to diagnose the health of industrial machines as well as to analyze their symmetrical behavior. When a fault occurs in the bearings, deviations from their symmetrical behavior can be indicative of potential faults. However, fault identification is challenging when (1) the vibration signals are recorded from variable speeds compared to the constant speed and (2) the vibration signals have diverse fault depths. In this work, we have proposed a highly accurate Deep Convolution Neural Network (DCNN)–Long Short-Term Memory (LSTM) model with a SoftMax classifier. The proposed model offers an innovative approach to fault diagnosis, as it obviates the need for preprocessing and digital signal processing techniques for feature computation. It demonstrates remarkable efficiency in accurately diagnosing fault conditions across variable speed vibration datasets encompassing diverse fault conditions, including but not limited to outer race fault, inner race fault, ball fault, and mixed faults, as well as constant speed datasets with varying fault depths. The proposed method can extract the features automatically from these vibration signals and, hence, are excellent to enhance the performance and efficiency to diagnose the machine's health. For the experimental study, two different datasets—the constant speed with different fault depths and variable speed rotating machines—are considered to validate the performance of the proposed method. The accuracy achieved for the variable speed rotating machine dataset is 99.40%, while for the diverse fault dataset, the accuracy reaches 99.87%. Furthermore, the experimental results of the proposed method are compared with the existing methods in the literature as well as the artificial neural network (ANN) model.

Keywords: vibration signals; variable-speed rotating machine; artificial neural network; deep convolution neural network; long short-term memory; softmax; deep learning

1. Introduction

Different sorts of rotating machines are used in the industry that are equipped with bearings as a fundamental rotating element. These rotating machines include industrial motors, compressors, fans, turbines, and so on [1–4]. It is crucial for the rotating machines to work smoothly in the industrial environment, without which a sudden switch-off could occur in the entire industrial processing. To avoid this situation, predictive maintenance of the rotating machines is required. Vibration signals from the rotating element bearings possess efficient information about the rotating machine conditions. Vibration signals from rotating machines are evaluated using different digital signal processing (DSP) techniques and machine learning (ML) technologies to detect the condition of the machine. The symmetrical behavior of these signals provides valuable insights into the functioning

and condition of the bearings. When a fault occurs in the bearings, deviations from their symmetrical behavior can be indicative of potential issues or faults. Therefore, by examining the symmetrical characteristics of the vibration signals, it becomes possible to identify and diagnose faults accurately, allowing for timely maintenance and prevention of sudden shutdowns or damages to the entire industrial process.

DSP methods are employed to extract relevant information from vibration signals. The vibration signals acquired from a rotating machine possess background noise and unwanted signals, which cause a low signal-to-noise ratio, and, consequently, it is challenging to diagnose the condition of the rotating machines [5]. To solve this issue, filtering is used to reduce noise or isolate certain frequency ranges of interest. Envelope analysis using power spectrum density (PSD) is also widely employed to determine the faults in rotating machines [6], but the vibration signals must have a high signal-to-noise ratio. Alternatively, kurtosis is an excellent statistical index to determine the transient in the vibration signal. Spectral Kurtosis (SK) [7–9] computes kurtosis over each frequency band in the vibration signal. In the literature, different results are presented for the fault diagnosis in vibration signals using kurtosis and SK [5,9–11]. However, it is challenging to determine the faults and machine conditions because of the non-ideal vibration signals.

ML technologies, on the other hand, are employed to train algorithms on labeled data to categorize the condition of the machine as normal or abnormal based on vibration patterns. Clustering, decision trees, neural networks, and support vector machines are some of the approaches that are used in ML techniques [12–18]. Furthermore, intelligent health diagnosis efficiently detects the faults in the rotating machine and generates results on its own using ML methods [19,20]. These intelligent methods are usually composed of two main steps: the first one is feature learning using a neural network or signal processing, and the second one is classification using pattern reorganization methods [21].

Moreover, deep learning methods have shown better performance in many fields including image classification, object detection, segmentation, and speech precogition [22]. Deep learning methods can handle high-dimensional complex data and learn features because of their multi-layer neural networks [23–25]. Deep learning is a form of ML that includes training multiple-layer artificial neural networks to discover complicated patterns and correlations in data. Deep learning approaches come in a variety of flavors, including [26]:

- Autoencoders: These are neural networks that have been trained to learn a compressed representation of input data. The network is initially trained to encode the input data into a lower-dimensional representation before decoding it back into its original format. Autoencoders are frequently used to extract features and reduce dimensionality.
- Deep belief networks: These are generative models with numerous layers of hidden units. Deep belief networks are trained through unsupervised learning and may be utilized for image and voice recognition.
- Deep Boltzmann machines: These are similar to deep belief networks, but they employ a different form of model known as a Boltzmann machine. Deep Boltzmann machines are also learned via unsupervised learning and may be utilized for tasks such as collaborative filtering and anomaly detection.
- Recurrent neural networks (RNNs) are neural networks that are designed to process sequential input, such as text or time series data. RNNs include loops in their network design that allow them to recall prior inputs and learn dependencies over time. RNNs are frequently used for language modeling and speech recognition.
- Convolutional neural networks (CNNs): These are neural networks that employ convolutional layers to learn spatial patterns in a picture or audio input. CNNs are frequently used for object identification and speech recognition.

Many researchers have combined CNN with other methods to improve the performance of fault diagnosis in rotating machines. For example, CNN is combined with a hierarchical convolution network [27] and hierarchical symbolic analysis [28,29], respectively, for bearing fault diagnosis. In [30], vibration signals are preprocessed using

continuous wavelet transform and then CNN is applied to diagnose the condition. Feature alignment method is used in [31] with multiscale CNN and multivariate encoder information-based CNN is applied in [32] for gearbox fault diagnosis. In [33], normalized deep CNN (DCNN) is applied for imbalanced fault diagnosis, and in [23], DCNN is applied to a noisy environment under different working loads.

Each of these deep learning approaches has advantages and disadvantages, and the methods used are determined by the goals and data at hand. Among these methods, DCNN is one of the most efficient methods for vibration fault diagnosis and prognosis in rotating machines because of two reasons [21,34]. The first reason is that raw vibration data can directly be applied to DCNN without manually extracting features because DCNN has the ability to extract the features automatically from raw data. The second reason for DCNN's popularity is that it has better performance with less training data compared to other neural network architectures [21,35]. However, the performance of DCNN is affected by raw vibration datasets from variable-speed rotating machines and variations in fault depths.

To improve the performance of DCNN, different sorts of classifiers are reported in the literature, including the Bayesian classifier, Artificial Neural Network (ANN), and Support Vector Machine (SVM) [14,36–38]. The Bayesian classifier is a statistical classifier that uses Bayes' theorem to assess the likelihood of a data point belonging to a certain class based on the likelihood of the attributes associated with that class. It is a simple and effective classifier, but it implies that the characteristics are independent of one another and that the feature distribution is known. ANNs are a type of machine learning model that is inspired by the structure and function of the human brain. ANNs are made up of linked nodes or neurons that execute mathematical operations on input data to anticipate output. Since they can learn complicated patterns and correlations in data, ANNs are frequently employed for categorization jobs. SVMs are a sort of supervised learning algorithm that is used for classification and regression analysis. SVMs seek a hyperplane that best separates the distinct classes in the input space. SVMs are frequently used for fault classification because they are good at managing high-dimensional data and can manage non-linear correlations between features.

Other classifiers that are commonly used for fault classification include decision trees, random forests, and k-nearest neighbors (KNN). The choice of the classifier will depend on the specific requirements of the task, such as the size and complexity of the dataset, the number of classes, and the desired level of accuracy and interpretability. The first two classifiers work efficiently if enough training data are available; otherwise, their performance is poor. On the other hand, SVM works efficiently with less training data because of its efficient network generalization capability and high classification accuracy [39]. However, if the data are redundant, the performance of SVM decreases because of its shallow structure [40].

Recently, researchers have been working on new techniques in order to analyze the rotating machine conditions; however, the accuracy and sensitivity of vibration analysis techniques, including digital signal processing and machine learning, can be affected by changes in working conditions or by constraints such as limited sensor data, high levels of noise, and complex vibration patterns. Furthermore, the above-mentioned literature review relies on huge training data, which is challenging for rotating machines in real industrial fault diagnosis environments. Moreover, there is little literature on condition diagnosis for variable speed vibration data using CNN, and even less has been researched about intelligent fault diagnosis for variable speed vibration data using DCNN.

The Long Short-Term Memory (LSTM) architecture excels at capturing and modeling long-term dependencies, which is crucial for fault classification in variable-speed rotating machines. The following are how LSTM addresses specific requirements in this context:

- Variable-speed rotating machines generate sequential data where each time step is influenced by the preceding ones. LSTM's recurrent connections allow for the capture of temporal dependencies in the data. By retaining information from previous time steps in its memory cell, LSTM can learn and exploit the patterns and relationships that exist across different speed regimes and time periods.

- Variable-speed machines produce data sequences of varying lengths depending on the duration of operation or the occurrence of faults. LSTM is designed to handle variable-length sequences as it processes data in a step-by-step manner, adapting to the varying time lengths. This flexibility makes LSTM well-suited for accommodating the dynamic nature of variable-speed rotating machines.
- LSTM's ability to recognize and learn speed-dependent features is crucial for fault classification in variable-speed rotating machines. By training on historical data that include speed information, LSTM can capture the relationships between speed and fault characteristics. It can then leverage these learned associations to make accurate fault predictions and classifications when new data are presented, considering the specific speed regime of the machine.
- Faults in rotating machines can exhibit complex patterns that may be difficult to detect using traditional techniques. LSTM's architecture allows it to learn and model complex relationships within the data. It can automatically extract relevant features, recognize subtle fault patterns, and capture the interactions between speed variations and fault signatures. This enables LSTM to provide accurate fault classifications even in challenging scenarios.

By utilizing LSTM's architecture, operators can leverage its ability to handle temporal dependencies, accommodate variable-length sequences, recognize speed-dependent features, and learn complex fault patterns. This makes LSTM a suitable choice for fault classification in variable-speed rotating machines. It can enhance the accuracy and reliability of fault diagnosis, enabling timely maintenance actions and improving the overall performance and longevity of the machines.

Motivated by the aforementioned literature, this paper proposes an efficient model consisting of DCNN-LSTM with SoftMax classifier to diagnose the faults in raw vibration signals. The proposed model is efficient for multiple fault diagnoses such as outer race fault, inner race fault, ball fault, and mixed faults in vibration signals acquired from variable speed rotating machines and diverse fault depths datasets. Different from the existing DCNN models, the proposed model consists of DCNN and LSTM with SoftMax classifier for multiple faults diagnosis and is efficient for both (1) variable speed rotating machines where the vibration data were recorded when the speed of the rotating machine was first increased and then decreased, and (2) a diverse fault depths dataset that consists of healthy bearings, inner race faults, outer race faults, and ball faults with different fault depths. The experimental results driven using the proposed model are compared with the existing models in the literature, and it is concluded that, for both datasets, the proposed model is more efficient in terms of accuracy to diagnose the different fault conditions.

The rest of the paper is arranged as follows. Section 2 includes the materials and methods that describe the proposed model and experimental setups used to record the vibration datasets. This section includes two different datasets utilized to validate the proposed model. In Section 3, the simulation results are presented for both datasets. This section also compares the results concluded for the proposed DCNN-LSTM model with the existing models in the literature and the ANN model with SoftMax classifier to illustrate the superiority of the proposed model over the existing models in the literature. Finally, the conclusion of this research work is presented at the end.

2. Materials and Methods

2.1. Structure of DCNN Model

This section elaborates on the DCNN model with SoftMax classifier as depicted in Figure 1. DCNN is the sub-branch of deep neural networks (DNNs). Recently, DCNNs are extensively used in image processing, object detection, natural language processing, and speech recognition. DCNNs analyze spatial correlations between nearby pixels and represent input data better than autoencoders and multilayer perceptrons. DCNN has achieved great advances in image recognition research as a standard approach for collecting data features in deep learning models. We have applied the DCNN-LSTM with the SoftMax

classifier for fault classification in one-dimensional data from rotating bearings having variable speed settings in this study. DCNN illustrated in Figure 1 has three fundamental layers including the (i)convolutional layer, (ii) fully-connected layer, and (iii) output layer.
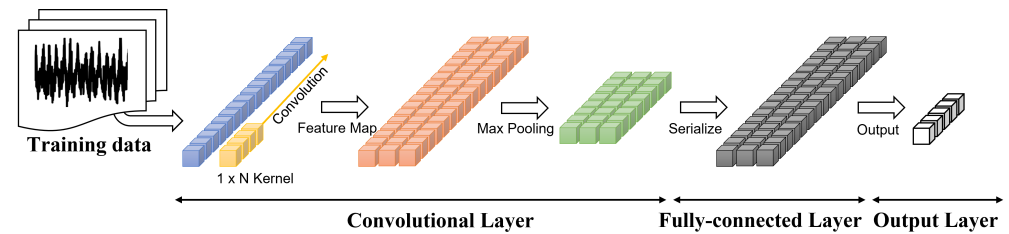

Figure 1. Block diagram of the DCNN model for fault classification in rotating machines.

The convolution layer is the essential layer in the DCNN that is equipped with a kernel. The input vibration signals fed to the convolution layer are filtered to extract the feature maps. To learn the multiple features, a convolutional kernel is used to perform the convolutions with input signals. Convolution procedures produce a powerful feature map in the convolution layer. By conducting local feature extraction on input data, the number of network parameters and model complexity is minimized. The following is the convolution formula:

$$h(t) = (x * w)(t) = \sum_{\tau=-m}^{m} x(t-\tau).w(\tau); \tag{1}$$

where $x \in \mathbb{R}^n$ is the input vibration signal and $w \in \mathbb{R}^{2m-1}$ is the kernel.

The constructed feature map is then forwarded to the pooling layer, which is the subsection of the convolutional layer as shown in Figure 1. The pooling layer reduces the parameters and dimensions of the network by applying a down-sampling tool. A down-sampling tool combines comparable features. Average pooling and maximum pooling are two commonly used pooling techniques. In average pooling, the average value of the patch is computed on the activation map, whereas in maximum pooling, the maximum value of the patch is computed on the activation map. To compute the pooling, the following expression is used:

$$x_j^n = f\left(\beta_j^n down(x_j^{n-1}) + b_j^n\right); \tag{2}$$

where x_j^n is the output, x_j^{n-1} is the input to the layer n, and b_j^n is the network bias. The operator $down(.)$ represents the down-sampling tool, $f(.)$ is the activation function, and β represents the network's weight.

An activation function is a function that interacts with a neural network's neurons and is in charge of translating inputs to outputs. In the literature, different activation functions are frequently used including the sigmoid, *tanh*, rectified linear unit (ReLU), and its derivatives. The sigmoid and *tanh* functions are significant to the gradient vanishing problem because they are saturated activation functions. ReLU, an unsaturated activation function, on the other hand, partially eliminates gradient vanishing and accelerates convergence. ReLU returns positive numbers precisely as they are, while instantly changing negative values to zero. Following convolution, the ReLU function is computed. Negative input causes ReLU activation to be zero and is defined as follows:

$$ReLU = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \leq 0 \end{cases}; \tag{3}$$

A dying ReLU problem arises when the ReLU is used as an activation unit; if the input is 0 or negative, the slope of the function is zero, and the network cannot be backpropagated

or learned. To resolve this problem, the Leaky ReLU activation function can be used as given below:

$$Leaky\ ReLU = \left\{ \begin{array}{ll} x & \text{if } x > 0 \\ 0.1x & \text{if } x \leq 0 \end{array} \right. ; \tag{4}$$

The pooling layer sends the data features to the fully connected layer where a linear transformation is applied to the input data $x^{(k-1)}$ through a weighted matrix ω^k. Where k is the network of hidden layers. The outputs of the fully connected layer are influenced by each input and can be represented as follows:

$$y^k = f(\omega^k x^{k-1} + b^k); \tag{5}$$

where b^k represents the network offset.

SoftMax is also an activation function that is usually used at the output layer. The output layer classifies the different sorts of predicted outputs. In this paper, DCNN also possesses a SoftMax classifier at the output layer to perform this task. A classifier applies a probability distribution to the input data to convert it into a vector form whose total sum is one. Compared to the other classifiers, the SoftMax classifier computes quickly, is simple to implement, and performs better. It can efficiently determine the probability $P(y^{(i)=j|x^{(i)}}$ of $x^{(i)}$ for each label j, where $i = 1, 2, 3, ..., K$, and $x^{(i)}$ and $y^{(i)}$ are the training set and its corresponding labels, respectively; where $i \in M$, and M is the total number of training samples. Figure 2 shows the standard form of the SoftMax classifier.
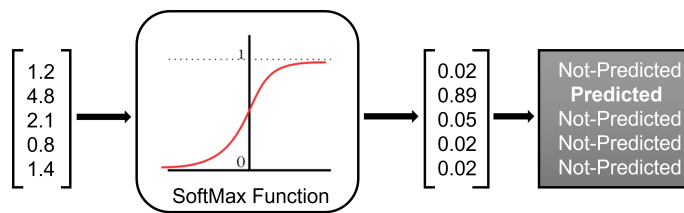

Figure 2. SoftMax classifier.

The following hypothesis function estimates the probability of each label's input data:

$$f\theta^{x^{(i)}} = \begin{bmatrix} P(y^{(i)} = 1 | x^{(i)}; \theta) \\ P(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ P(y^{(i)} = K | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum\limits_{k=1}^{K} e^{\theta_k^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_K^T x^{(i)}} \end{bmatrix}; \tag{6}$$

where $[\theta_1, \theta_2, ..., \theta_K]^T$ are the parameters of the SoftMax regression model and $\theta \sum\limits_{i=1}^{K} e^{\theta_k^T x^{(i)}}$ is the normalized distribution such that the sum of this normalized distribution is equal to one. In order to represent the probabilities for each class, this classifier function ensures that the outputs are positive values between 0 and 1.

2.2. Structure of LSTM

LSTM is a type of recurrent neural network (RNN) architecture that is specifically designed to handle and model sequential data. Figure 3 illustrates the structure of a typical LSTM block. LSTM is particularly effective in capturing and learning long-term dependencies and patterns in time series data, making it well-suited for time series vibration data.
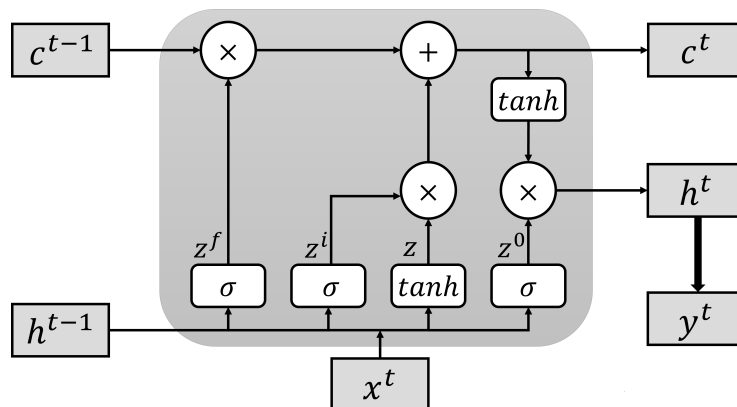
Figure 3. Structure of a typical LSTM block.

Traditional RNNs suffer from the vanishing gradient problem, which hinders their ability to capture long-term dependencies. LSTM addresses this issue by introducing a memory cell and a gating mechanism. The memory cell retains information over long sequences, allowing the network to selectively remember or forget information based on the input data.

The key components of an LSTM unit are the following:

- Cell State (c^t): The cell state serves as the memory of the LSTM. It carries information across time steps, allowing the network to maintain long-term dependencies.
- Input Gate (z^i): The input gate controls the amount of new information that is added to the cell state at each time step. It decides which parts of the input are relevant and should be stored in the cell state.
- Forget Gate (z^f): The forget gate determines which parts of the cell state should be forgotten or discarded. It selectively removes information that is no longer relevant, preventing the cell state from being cluttered with unnecessary information.
- Output Gate (z^0): The output gate controls the amount of information that is output from the cell state to the next layer or as the final prediction. It determines which parts of the cell state are relevant for the current time step.

The following are the mathematical formulas of LSTM units:

$$z = \tanh(w[x^t, h^{t-1}] + b) \tag{7}$$

$$z^i = \sigma(w^i[x^t, h^{t-1}] + b^i) \tag{8}$$

$$z^f = \sigma(w^f[x^t, h^{t-1}] + b^f) \tag{9}$$

$$z^0 = \sigma(w^0[x^t, h^{t-1}] + b^0) \tag{10}$$

The cell state c^t is given by the sum of the Hadamard product $(*)$ of forget gate and the previous cell state and the Hadamard product of the input gate and cell update z. Following is the formula of the cell state c^t:

$$c^t = z^f * c^{t-1} + z^i * z; \tag{11}$$

Similarly, the new hidden state h^t is given by:

$$h^t = z^0 * \tanh(c^t); \tag{12}$$

and the current output y^t is given by:

$$y^t = \sigma(w'h^t). \tag{13}$$

The LSTM architecture enables the network to effectively capture and retain long-term dependencies in sequential data. By selectively storing and forgetting information at each

time step, the network can learn to recognize and leverage relevant patterns in the data, even when they occur over long time spans.

In our research, we incorporate LSTM layers into our DCNN-LSTM model to capture temporal dependencies in the vibration data. The LSTM layers act as a crucial component for understanding the sequential nature of the signals, enabling our model to make accurate fault classifications in variable-speed rotating machines and diverse fault depths.

2.3. Proposed Models

In the initial part of this section, the data segmentation process is examined in order to create separate training and testing datasets. Subsequently, the development of the proposed model takes place. This section further demonstrates the ANN model with the SoftMax classifier, which is employed to compare the performance of the proposed DCNN-LSTM model with the SoftMax classifier.

2.3.1. Data Segmentation

All vibration signals were divided into small sequences known as segments using a sliding window. The length of the sliding window was 1000 which was moved left with a stride of 200 to make the segments of the vibration signal as shown in Figure 4. Each segment of the vibration signals was assigned the corresponding fault name as a tag. The segmented data was then shuffled and divided into training and testing datasets with 70% and 30% of the total data segments, respectively.
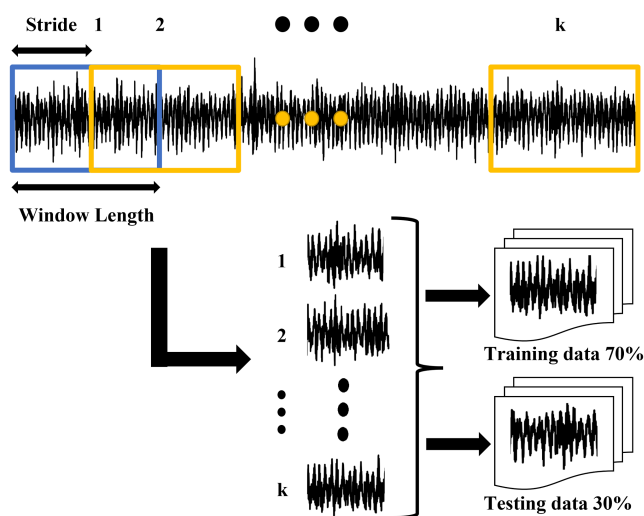

Figure 4. Data segmentation and training and testing dataset construction.

Shuffling is important before training the proposed model to improve the performance, reduce bias, enhance generalization, optimize gradient descent, and mitigate overfitting. These are detailed as follows:

- Reduce bias: When the data are ordered in a certain way, such as being sorted by class labels, it can introduce bias during training. Shuffling the data helps to ensure that the model sees a diverse range of samples from different classes, reducing the potential bias.
- Enhance generalization: If the data are not shuffled and there is a particular order or pattern in the dataset, the model may learn to rely on that pattern instead of learning the underlying relationships between features and labels. Shuffling the data helps to break any sequential patterns and encourages the model to learn more generalized representations.
- Improve gradient descent optimization: Optimization algorithms like stochastic gradient descent (SGD) work by updating the model's parameters based on mini-batches

of data. Shuffling the data ensures that each mini-batch contains a random sample of data, leading to more effective updates and faster convergence.

- Mitigate overfitting: Shuffling the data helps to prevent overfitting by introducing randomness in the training process. Overfitting occurs when a model becomes too specialized in the training data and fails to generalize well to new, unseen data. Shuffling the data helps to make the model more robust and less prone to overfitting.

2.3.2. DCNN-LSTM Model with SoftMax Classifier

The block diagram of the proposed DCNN-LSTM model is illustrated in Figure 5. We designed a deep learning model consisting of Convolutional 1D (Conv1D) and LSTM layers to effectively extract temporal and spatial features from the vibration data. The workings of each layer in a general DCNN model and LSTM block are also elaborated on in the previous section and given in Figures 1 and 3, respectively.

The proposed DCNN-LSTM model is more efficient compared to the existing models in the literature as well as ANN in terms of accuracy of fault diagnosis using raw vibration datasets from (1) variable speed rotating machines and (2) diverse fault depths datasets. The proposed model used a kernel at the input layer that helps in the convolution and construction of a feature map from the input signals.
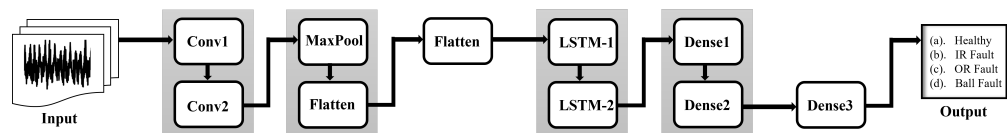

Figure 5. Block diagram of the proposed DCNN–LSTM model with SoftMax classifier.

The given architecture of the proposed model is given in Table 1. The CNN architecture applies convolutions and pooling operations to extract features from the input data and then uses dense layers for classification. ReLU activation functions are applied to introduce non-linearity, while the SoftMax function produces a probability distribution over the output classes.

Table 1. Proposed DCNN-LSTM model with SoftMax classifier.

Layer (Type)	Activation Function	Output Shape	No. of Parameters
Layer 1 (Convolution)	ReLU	(None, 901, 64)	6464
Layer 2 (Convolution)	ReLU	(None, 852, 32)	102,432
Layer 3 (MaxPooling)	-	(None, 213, 32)	0
Layer 4 (Flatten)	-	(None, 6816)	0
Layer 5 (Reshape)	-	(None, 213, 32)	0
Layer 6 (LSTM)	Sigmoid	(None, 213, 64)	24,832
Layer 7 (LSTM)	Sigmoid	(None, 16)	5184
Layer 8 (Dense)	ReLU	(None, 100)	1700
Layer 9 (Dense)	ReLU	(None, 50)	5050
Layer 10 (Dense)	SoftMax	(None, 5)	255

The model starts with two Conv1D layers. The first Conv1D layer applies 64 filters with a kernel size of 100 to capture important patterns in the input vibration signals. The second Conv1D layer follows with 32 filters and a kernel size of 50 to further enhance the learned features. The MaxPooling1D layer reduces the spatial dimensions of the output, facilitating efficient information processing.

To bridge the gap between the CNN and LSTM layers, we used a Reshape layer to adjust the output shape from the previous layers. This reshaping operation ensures compatibility with the input requirements of the subsequent LSTM layer. In our model, we reshaped the output to have a fixed number of time steps and 32 features.

The LSTM layers are essential for capturing temporal dependencies in the vibration data. We employed two LSTM layers to effectively model the sequential nature of the signals. The first LSTM layer consists of 64 memory units and is configured to return sequences, providing outputs for each time step. The second LSTM layer has 16 units and does not return sequences, condensing the temporal information into a fixed-length representation.

To further process the extracted features, we incorporated additional dense layers. These fully connected layers transform the information from the previous layers into a more abstract representation. The first dense layer has 100 units with ReLU activation, which introduces non-linearity to the model. The second dense layer follows with 50 units, aiding in the hierarchical abstraction of features.

For the final classification, we employed a dense output layer with a number of units equal to the unique fault classes in the dataset. The SoftMax activation function was applied to produce class probabilities, enabling multi-class fault classification.

Our model architecture was implemented using TensorFlow's Keras API. The summary of the model reveals the layer-wise configuration, including the number of parameters and the shapes of each layer's output. Furthermore, the training process utilized a batch size of 500 samples per iteration, and the model was trained for a total of 50 epochs. The optimization algorithm employed was Adam, with a learning rate set to 0.005. The proposed DCNN-LSTM model demonstrates the promising potential for fault classification in variable-speed rotating machines by effectively leveraging both temporal and spatial information from vibration datasets.

2.3.3. ANN Model with SoftMax Classifier

The block diagram of the ANN model is depicted in Figure 6. Each segment from the training dataset was fed to the input layer of the ANN to train the model.

The developed ANN architecture is given in Table 2 and consists of five dense layers with ReLU activation functions and a final dense layer with SoftMax activation function. Here is a breakdown of each layer:

- Layer 1 (Dense): This layer has 1024 neurons with ReLU activation function. The output shape of this layer is (None, 1024), where "None" indicates that the batch size can be variable. The number of parameters in this layer is 1,025,024, which is calculated as (input shape × number of neurons) + number of biases (1).
- Layer 2 (Dense): This layer has 512 neurons with ReLU activation function. The output shape of this layer is (None, 512), and it has 524,800 parameters.
- Layer 3 (Dense): This layer has 256 neurons with ReLU activation function. The output shape of this layer is (None, 256), and it has 131,328 parameters.
- Layer 4 (Dense): This layer has 128 neurons with ReLU activation function. The output shape of this layer is (None, 128), and it has 32,896 parameters.
- Layer 5 (Dense): This layer has 5 neurons with SoftMax activation function. The output shape of this layer is (None, 5), which corresponds to the number of classes in the classification task. The SoftMax function is used to convert the output of the previous layer into a probability distribution over the classes. This layer has 645 parameters.
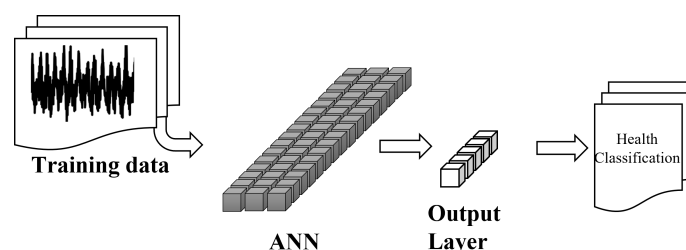

Figure 6. Block diagram of the ANN model for fault classification in rotating machines.

Table 2. ANN model with SoftMax classifier.

Layer (Type)	Activation Function	Output Shape	No. of Parameters
Layer 1 (Dense)	ReLU	(None, 1024)	1,025,024
Layer 2 (Dense)	ReLU	(None, 512)	524,800
Layer 3 (Dense)	ReLU	(None, 256)	131,328
Layer 4 (Dense)	ReLU	(None, 128)	32,896
Layer 5 (Dense)	SoftMax	(None, 5)	645

The ANN architecture is designed for fault diagnosis and is suitable for a classification task with 5 classes. The ReLU activation functions are commonly used in deep learning because they allow the network to learn nonlinear relationships between features and labels, while the SoftMax function is used for multi-class classification problems to produce a probability distribution over the possible classes.

2.4. Experimental Setups and Datasets

2.4.1. Variable Speed Dataset

The prototype model is presented in Figure 7. Experiments were carried out to acquire vibration data from bearings running at time-varying speeds in order to validate the efficacy of the proposed technique for bearing failure diagnosis under unknown time-changing speed settings [41,42]. The acquired dataset is different from the existing datasets in the literature that are recorded under constant speed settings. The datasets were collected with healthy bearings, inner race fault, outer race fault, ball fault, and combination fault. The shaft was moved by a variable-speed motor operated by the AC controller. Two ER16K bearings were linked to the shaft; the one on the right side is healthy, and the one on the left side is replaced with three distinct bearings for independent testing. An accelerometer (ICP accelerometer, Model 623C01) was installed on the left side of the bearing housing to collect vibration data. The data were acquired by the NI data acquisition board (NIUSB-6212BNC) and sampled by LABVIEW. The sampling frequency is 200 kHz with a 10 s sampling duration. The inner race and outer race fault frequencies are $5.43 f_r$ and $3.57 f_r$, respectively, where f_r is the shaft rotational frequency. The parameters of the bearing are listed in Table 3 and the datasets are available at http://dx.doi.org/10.17632/v43hmbwxpm.1, (accessed on 1 January 2022).
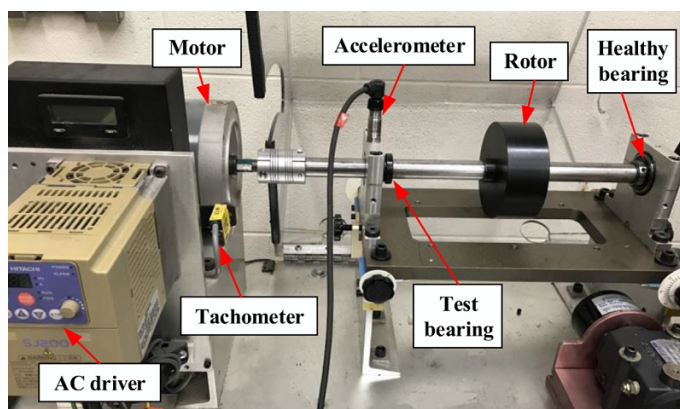
Figure 7. Experimental setup [42].

Vibration signals were recorded using different faulty bearings with variable speeds. In the undertaken datasets, the speed was first increased and then decreased gradually. The vibration signals in the time domain can be visualized in Figure 8. These signals consist of a vibration signal from a healthy bearing, a vibration signal with a ball fault, a vibration signal with an inner race fault, a vibration signal with an outer race fault, and a vibration signal with combined faults composed of inner race fault, outer race fault, and ball

fault. Table 4 illustrates the experimental dataset from the variable-speed rotating machine. The fault signals in the variable-speed rotating machines experienced a change in speed, with the average rotating speed increasing from 14.5 Hz to 24 Hz and then decreasing to 18 Hz. This variability in speed introduces an additional level of complexity when it comes to fault classification. In total, 49,975 segments were built from all vibration signals.

Table 3. ER16K bearing specification.

Description	Variable	Value
Number of balls	n	9
Ball diameter	d	7.94 mm
Pitch diameter	D	38.52 mm
Bearing contact angle	α	0

Table 4. Variable-speed dataset.

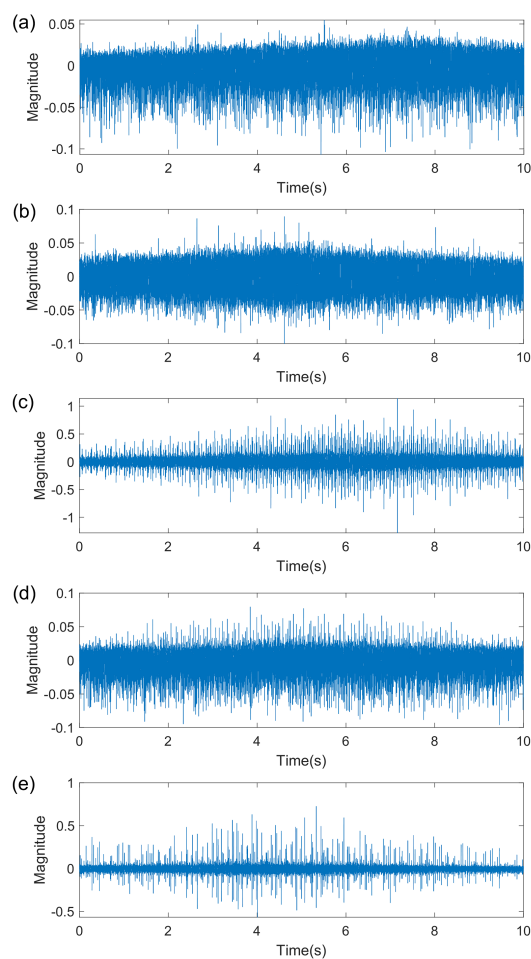| Fault Class	Rotating Speed
Healthy bearing	First increased then decreased
Ball fault	First increased then decreased
Inner race fault	First increased then decreased
Outer race fault	First increased then decreased
Combined fault	First increased then decreased

Figure 8. Raw vibration signals with variable speed conditions: (**a**) vibration signal from healthy bearing, (**b**) vibration signal with ball fault, (**c**) vibration signal with inner race fault, (**d**) vibration signal with outer race fault, and (**e**) vibration signal with combined faults.

The dimensionality increases when all the vibration signals are treated together for fault classification. However, to visualize all vibration signals together in a 2D graph, the t-Distributed Stochastic Neighbor Embedding (t-SNE) tool was used. t-SNE reduces the dimensionality to visualize high-dimensional datasets. The resultant 2D graph can be seen in Figure 9. From Figure 9, it can be seen that all the faults are mixed with each other and difficult to distinguish.
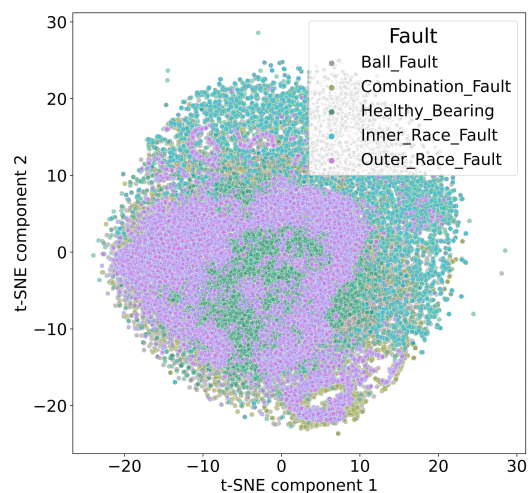

Figure 9. Visualization of all fault types for variable speed rotating machine using t-SNE.

2.4.2. Diverse Fault Depths Dataset

The dataset used in this study is derived from the Case Western Reserve University (CWRU). It comprises vibration signal data obtained from an experimental setup featuring a 2-horsepower induction motor equipped with a deep ball bearing of type 6205-2RS JEM SKM, as depicted in Figure 10. The vibration data were collected using accelerometers with a sampling frequency of 12 kHz, while the motor operated under a load.
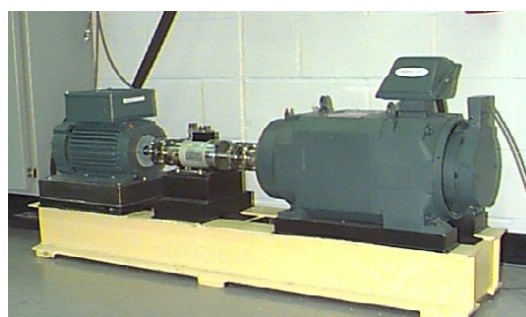

Figure 10. CWRU experimental setup.

This dataset encompasses various fault conditions, including healthy bearings, ball faults, inner race faults, and outer race faults. Each fault category consists of multiple sub-datasets, which were collected at different fault severities. These severity levels represent different degrees of machinery damage.

The CWRU vibration datasets have gained significant recognition in the research community and are widely employed for the development and evaluation of fault detection and diagnosis algorithms. Researchers utilize these datasets to test and refine their methods.

To access the datasets used in this study, you can visit the CWRU Bearing Data Center website at https://csegroups.case.edu/bearingdatacenter, (accessed on 1 January 2022). For reference, Table 5 provides an overview of the dataset specifically utilized in this research.

Table 5. CWRU dataset with different fault depths.

Fault Class	Symbol	Fault Depth
Healthy bearing	N	-
Ball fault	007_BA	0.007 inch
Ball fault	014_BA	0.014 inch
Ball fault	021_BA	0.021 inch
Ball fault	028_BA	0.028 inch
Inner race fault	007_IR	0.007 inch
Inner race fault	014_IR	0.014 inch
Inner race fault	021_IR	0.021 inch
Inner race fault	028_IR	0.028 inch
Outer race fault	007_OR1	0.007 inch
Outer race fault	007_OR2	0.007 inch
Outer race fault	007_OR3	0.007 inch
Outer race fault	014_OR1	0.014 inch
Outer race fault	021_OR1	0.021 inch
Outer race fault	021_OR2	0.021 inch
Outer race fault	021_OR3	0.021 inch

Figure 11 represents the 2D plot of all the fault types for the CWRU dataset using t-SNE. It is a popular dimensionality reduction technique used for visualizing high-dimensional data in a lower-dimensional space.
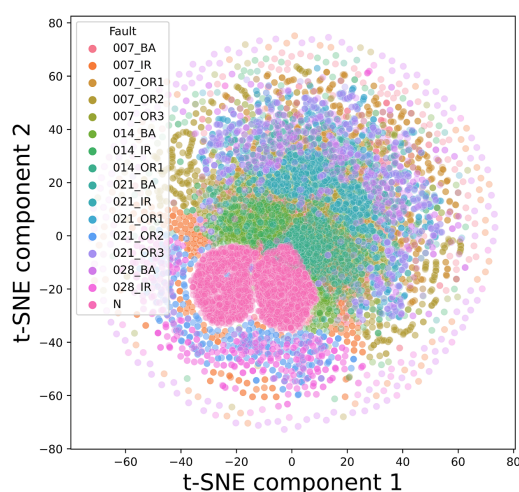
Figure 11. Visualization of all fault types of the CWRU dataset using t-SNE.

3. Experimental Results

3.1. Case I: Variable-Speed Vibration Dataset

In this section, simulation results are presented for both the ANN and DCNN-LSTM models and are compared with each other to differentiate the efficiencies of both models. The ANN and DCNN-LSTM model with SoftMax classifier were applied to the vibration data with variable speed of the rotating machine to classify the different sorts of faults. The constructed high-dimensional dataset was divided into training and testing datasets with 70% and 30% ratios, respectively.

The training and validation accuracy using the ANN model described in Table 2 was 99.40% and 95.08%, respectively. Figure 12 shows the accuracy graph of the developed ANN model. The green line shows the training accuracy, while the red line shows the validation accuracy.
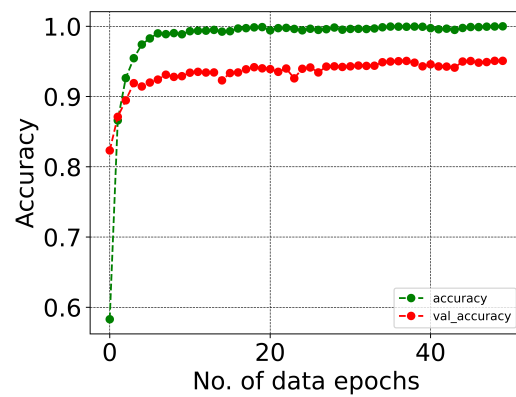
Figure 12. Accuracy of ANN model.

The accuracy was improved to 99.40% when the DCNN-LSTM model with SoftMax classifier was trained using the vibration data from variable-speed rotating machines. Figure 13 shows the accuracy graph of the model.
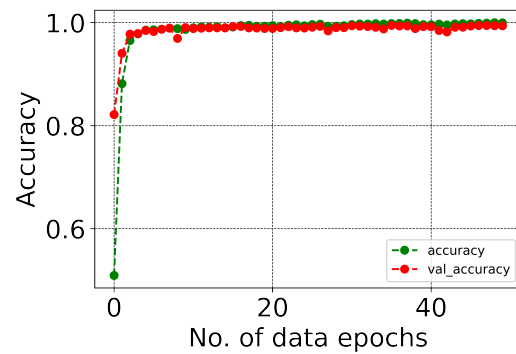

Figure 13. Accuracy of the proposed DCNN-LSTM model.

Furthermore, the confusion matrix illustrates the accuracy of each fault type. Figures 14 and 15 show the confusion matrices concluded from both the ANN and DCNN-LSTM models.
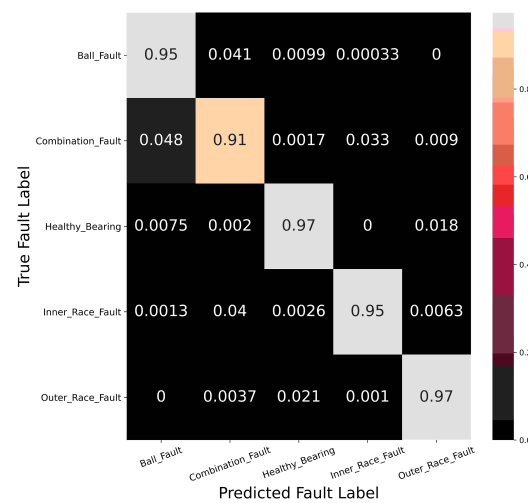

Figure 14. Confusion matrix of the ANN model for variable speed rotating machine.
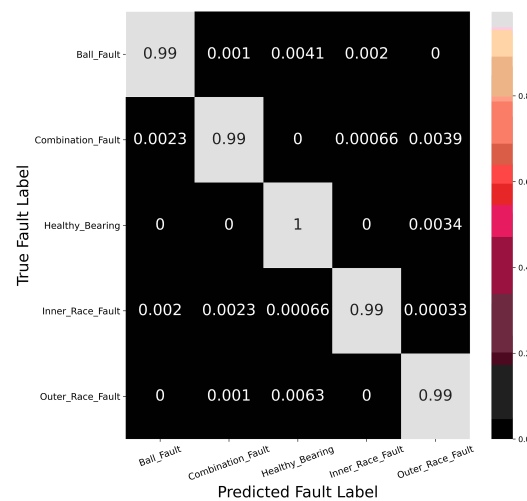
Figure 15. Confusion matrix of the DCNN-LSTM model for variable speed rotating machine.

From the confusion matrices of both models, it can be concluded that the DCNN-LSTM is more efficient compared to the ANN model. From the diagonal entries of the confusion matrix of the DCNN-LSTM model, it can be seen that the ball faults, combined faults, and healthy signals are 100% classified, while the inner race faults and outer race faults are 99% classified. Moreover, the t-SNE graph shown in Figure 16 illustrates that each fault can be classified and visualized in a 2D plot.
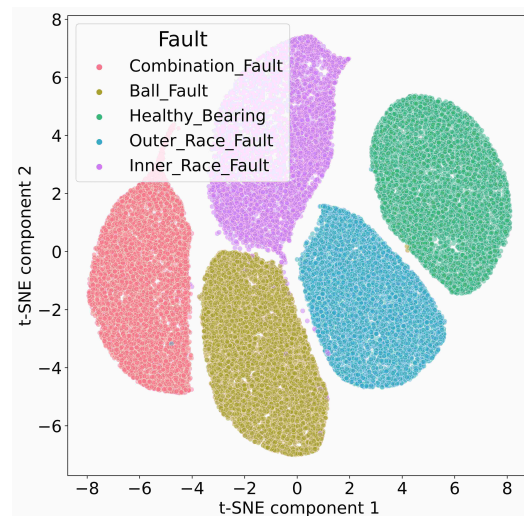

Figure 16. t-SNE graph of the validation dataset with variable speed rotating machine.

Table 6 displays a comparison between the findings of the proposed model and the existing literature. In Reference [43], four methods (EEMD, VMD, BMD, and GBMD) were proposed for analyzing the variable speed vibration dataset and their accuracy rates were 73.33%, 80.00%, 86.67%, and 96.67%, respectively. However, this study solely focused on the normal, inner race fault, and outer race fault classes. In contrast, our investigation considered five categories of classes, which include healthy bearing, ball fault, inner race fault, outer race fault, and combination fault.

Reference [44] introduced five methods (FEE-SVM, DFF-MLP, DFF-1DCNN, DFF-TICNN, and DFF-Lightweight 1DCNN) for diagnosing fault conditions in variable speed data. The accuracy rates for these techniques were reported as 45.60%, 55.16%, 90.08%, 93.06%, and 96.26%, respectively. The experimental results were also derived using the ANN model with SoftMax classifier that had an accuracy of 95.08%. In comparison to the above results, our proposed approach, which utilizes a DCNN-LSTM model with SoftMax

classifier, offers a more efficient means of fault diagnosis for the variable speed rotating machinery, achieving an accuracy of 99.40%.

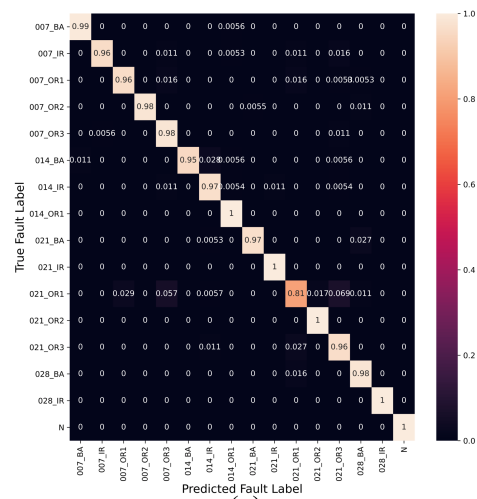**Table 6.** Comparison with existing literature for variable speed dataset.

| References	Models	Accuracy
[43]	EEMD	73.33%
	VMD	80.00%
	BMD	86.67%
	GBMD	96.67%
[44]	DFF-SVM	45.60%
	DFF-MLP	55.16%
	DFF-1DCNN	90.08%
	DFF-TICNN	93.06%
	DFF-Lightweight 1DCNN	96.26%
Proposed method	**ANN with SoftMax**	**95.08%**
	DCNN-LSTM with SoftMax	**99.40%**

3.2. Case II: Diverse Fault Depths Vibration Dataset

The proposed model of the DCNN-LSTM model with the SoftMax classifier was evaluated to diagnose the fault conditions using a diverse fault depths vibration dataset having different fault depths and a load. The CWRU dataset consists of one healthy and fifteen faulty conditions. The faulty conditions are further classified into three categories: ball fault, inner race fault, and outer race fault with different fault depths as described in Table 5. The confusion matrices for the DCNN-LSTM model with SoftMax classifier and the ANN model with SoftMax classifier were generated using the test dataset and shown in Figure 17.

The results suggest that the DCNN-LSTM model with SoftMax classifier outperformed the ANN model with SoftMax classifier in terms of accuracy, with an accuracy of 99.87% compared to 97.88%. The difference in accuracy could be due to the superior ability of the DCNN-LSTM model in learning complex patterns in the dataset, which is crucial for accurate classification. Furthermore, the experimental results demonstrate that the proposed model is effective in classifying faulty conditions and provide insights into the superior performance of the DCNN-LSTM model with SoftMax classifier over the ANN model with SoftMax classifier in this specific classification problem. Figure 18 shows the fault classification and representation in a 2D graph using t-SNE. From Figure 18, it can be concluded that the fault conditions can be distinguished for the test dataset.

Table 7 shows how the accuracy of the proposed DCNN-LSTM model with SoftMax classifier compares to other existing models in the literature. As we can see from Table 7, the model referenced in [45] achieves an accuracy of 99.17%. However, it is important to note that the authors of [45] only considered normal bearings, inner race faults, and two types of outer race faults in their experiments. In contrast, this paper considers 16 different classes of bearing conditions, as shown in Table 5. In [46], five different methods (CNN, MAML, Reptile, Reptile with GC, and EML) were evaluated, with accuracies of 90.46%, 92.51%, 92.63%, 93.48%, and 98.78%, respectively. In this paper, only two fault depths (0.007 inches and 0.021 inches) were considered. In [47], six methods (RF, GRU, CNN, CNN-GRU, MSCNN, and MDRMA-MSCM) were evaluated for the CWRU dataset and the accuracy of each method was 87.28%, 97.33%, 97.54%, 98.22%, 99.06%, and 99.71%, respectively. However, the 0.028-inch fault depth was not considered for the experimental results. In contrast to the above-mentioned literature, the proposed model in this paper was developed using the DCNN-LSTM model with SoftMax classifier, and all available fault depths (0.007 inches, 0.014 inches, 0.021 inches, and 0.028 inches) were considered for the experimental work to evaluate the performance and superiority of the proposed model. The results illustrated in Table 7 show the superiority of the proposed model.

(a)
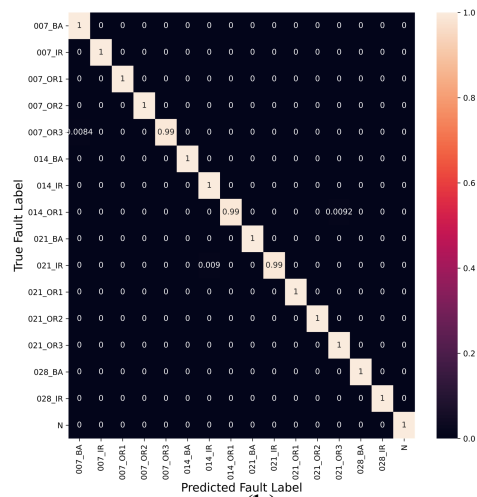
(b)

Figure 17. Confusion matrix for diverse fault depths vibration dataset: (**a**) ANN model and (**b**) DCNN-LSTM model.
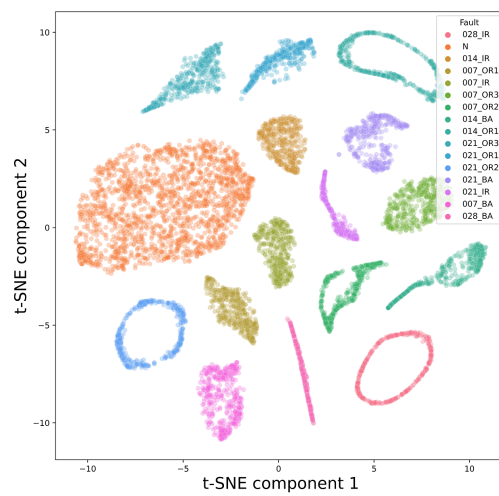

Figure 18. t–SNE graph of the validation dataset for diverse fault depths vibration dataset.

Table 7. Comparison with existing literature for the CWRU dataset.

References	Models	Accuracy
[45]	-	99.17%
[46]	CNN	90.46%
	MAML	92.51%
	Reptile	92.63%
	Reptile with GC	93.48%
	EML	98.78%
[47]	RF	87.28%
	GRU	97.33%
	CNN	97.54%
	CNN-GRU	98.22%
	MSCNN	99.06%
	MDRMA-MSCM	99.71%
Proposed method	**ANN with SoftMax**	**97.88%**
	DCNN-LSTM with SoftMax	**99.87%**

4. Conclusions

This paper proposes an efficient model of the DCNN-LSTM model with SoftMax classifier for diagnosing faults in rotating machines, capable of diagnosing faults in both variable speed and diverse fault depths datasets. In contrast to the current DCNN-LSTM models, the novel model presented in this study exhibits enhanced efficacy in two distinct scenarios. Firstly, it demonstrates efficiency in analyzing vibration data obtained from variable-speed rotating machines. The recorded data encompass instances where the rotational speed of the machine undergoes a sequential increase and subsequent decrease. This characteristic sets it apart from prior DCNN-LSTM models that primarily focus on constant-speed operating conditions. The proposed DCNN-LSTM model achieved an impressive overall condition prediction accuracy of 99.40% for variable-speed rotating machines. In addition, when compared to the existing literature, it was found that the proposed model outperformed the other approaches in terms of prediction accuracy.

Secondly, the proposed model showcases its effectiveness in handling a diverse fault depths dataset. This dataset comprises various fault categories, including healthy bearings, inner race faults, outer race faults, and ball faults. Notably, each fault category incorporates fault instances with different levels or depths of damage to the machinery. This aspect distinguishes the proposed model from existing approaches, which often concentrate on specific fault types or do not explicitly consider varying fault depths. The DCNN-LSTM model proposed in this study demonstrated an outstanding overall condition prediction accuracy of 99.87% when applied to a diverse fault depths dataset. Furthermore, when compared to existing literature, it is evident that the proposed model surpasses other approaches in terms of prediction accuracy.

Moreover, the proposed model eliminates the need for manual feature extraction from the vibration datasets. This examination of symmetrical characteristics enables accurate identification and diagnosis of faults, leading to timely maintenance and the prevention of sudden shutdowns or damages to the entire industrial process. It streamlines the analysis process by directly feeding the raw vibration signals into the model. This feature circumvents the laborious task of manually identifying and extracting relevant features, thus enhancing the efficiency and simplicity of the diagnostic process.

The performance of the proposed model was compared with an ANN model for both datasets, and it was found that the proposed DCNN-LSTM model with the SoftMax classifier is more efficient in diagnosing fault conditions. Additionally, the results of the proposed model were compared with existing models in the literature, demonstrating its ability to effectively diagnose faults in both variable speed and diverse fault depths

datasets with a load. Overall, it was concluded that the proposed DCNN-LSTM model with SoftMax classifier is a promising tool for fault diagnosis in rotating machines with raw vibration datasets.

Author Contributions: Conceptualization, M.A. and M.M.S.; Methodology, M.A.; Software, M.A.; Validation, M.A.; Formal analysis, M.M.S.; Investigation, M.A.; Data curation, M.A.; Writing—original draft, M.A. and M.M.S.; Writing—review& editing, M.M.S.; Supervision, M.M.S.; Project administration, M.M.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The datasets were obtained from Mendeley Data by the University of Ottawa and Bearing Data Center by the Case Western Reserve University. These datasets can be accessed from the following URLs: (1) http://dx.doi.org/10.17632/v43hmbwxpm.1 and (2) https://csegroups.case.edu/bearingdatacenter, respectively.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, L.; Cai, G.; Wang, J.; Jiang, X.; Zhu, Z. Dual-Enhanced Sparse Decomposition for Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 450–461. [CrossRef]
2. Jiang, G.; He, H.; Xie, P.; Tang, Y. Stacked Multilevel-Denoising Autoencoders: A New Representation Learning Approach for Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 2391–2402. [CrossRef]
3. He, Q.; Zhao, J.; Jiang, G.; Xie, P. An Unsupervised Multiview Sparse Filtering Approach for Current-Based Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 5569–5578. [CrossRef]
4. Yu, X.; Tang, B.; Zhang, K. Fault Diagnosis of Wind Turbine Gearbox Using a Novel Method of Fast Deep Graph Convolutional Networks. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 6502714. [CrossRef]
5. Ahsan, M.; Bismor, D. Early-Stage Fault Diagnosis for Rotating Element Bearing Using Improved Harmony Search Algorithm with Different Fitness Functions. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3519309. [CrossRef]
6. Abboud, D.; Antoni, J.; Sieg-Zieba, S.; Eltabach, M. Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. *Mech. Syst. Signal Process.* **2017**, *84*, 200–226. [CrossRef]
7. Udmale, S.S.; Singh, S.K. Application of Spectral Kurtosis and Improved Extreme Learning Machine for Bearing Fault Classification. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 4222–4233. [CrossRef]
8. Hu, Y.; Bao, W.; Tu, X.; Li, F.; Li, K. An Adaptive Spectral Kurtosis Method and its Application to Fault Detection of Rolling Element Bearings. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 739–750. [CrossRef]
9. Jérôme, A. The spectral kurtosis: A useful tool for characterising non-stationary signals. *Mech. Syst. Signal Process.* **2006**, *20*, 282–307. [CrossRef]
10. Dong, W.; Peter, W.T.; Kwok, L.T. An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2013**, *35*, 176–199. [CrossRef]
11. Jérôme, A. Fast computation of the kurtogram for the detection of transient faults. *Mech. Syst. Signal Process.* **2007**, *21*, 108–124. [CrossRef]
12. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [CrossRef]
13. Robert, B.R.; Jérôme, A. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [CrossRef]
14. Soualhi, A.; Medjaher, K.; Zerhouni, N. Bearing Health Monitoring Based on Hilbert–Huang Transform, Support Vector Machine, and Regression. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 52–62. [CrossRef]
15. Huang, W.; Gao, G.; Li, N.; Jiang, X.; Zhu, Z. Time-Frequency Squeezing and Generalized Demodulation Combined for Variable Speed Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 2819–2829. [CrossRef]
16. Yan, R.; Gao, R.X. Hilbert–Huang Transform-Based Vibration Signal Analysis for Machine Health Monitoring. *IEEE Trans. Instrum. Meas.* **2006**, *55*, 2320–2329. [CrossRef]
17. He, Q.; Song, H.; Ding, X. Sparse Signal Reconstruction Based on Time-Frequency Manifold for Rolling Element Bearing Fault Signature Enhancement. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 482–491. [CrossRef]
18. Attoui, I.; Boutasseta, N.; Fergani, N. Novel Machinery Monitoring Strategy Based on Time–Frequency Domain Similarity Measurement with Limited Labeled Data. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3500708. [CrossRef]
19. Qian, W.; Li, S.; Jiang, X. Deep transfer network for rotating machine fault analysis. *Pattern Recognit.* **2019**, *96*, 106993. [CrossRef]

20. Li, W.; Huang, R.; Li, J.; Liao, Y.; Chen, Z.; He, G.; Yan, R.; Gryllias, K. A prospective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mech. Syst. Signal Process.* **2022**, *167*, 108487. [CrossRef]

21. Mohammed, H.; Abdoulhdi, A.B.O.; Ali, N.A.; Muhannad, A.W.; Abdallah, A. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Eng. J.* **2023**, *14*, 101945. [CrossRef]

22. Samir, K.; Takehisa, Y. A review on the application of deep learning in system health management. *Mech. Syst. Signal Process.* **2018**, *107*, 241–265. [CrossRef]

23. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [CrossRef]

24. Li, X.; Zhang, W.; Xu, N.X.; Ding, Q. Deep Learning-Based Machinery Fault Diagnostics with Domain Adaptation Across Sensors at Different Places. *IEEE Trans. Ind. Electron.* **2020**, *67*, 6785–6794. [CrossRef]

25. Zhou, Y.; Zhi, G.; Chen, W.; Qian, Q.; He, D.; Sun, B.; Sun, W. A new tool wear condition monitoring method based on deep learning under small samples. *Measurement* **2022**, *189*, 110622. [CrossRef]

26. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]

27. Lu, C.; Wang, Z.; Zhou, B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network-based health state classification. *Adv. Eng. Inform.* **2017**, *32*, 139–151. [CrossRef]

28. Saravanakumar, R.; Krishnaraj, N.; Venkatraman, S.; Sivakumar, B.; Prasanna, S.; Shankar, K. Hierarchical symbolic analysis and particle swarm optimization based fault diagnosis model for rotating machineries with deep neural networks. *Measurement* **2021**, *171*, 108771. [CrossRef]

29. Yang, Y.; Zheng, H.; Li, Y.; Xu, M.; Chen, Y. A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network. *ISA Trans.* **2019**, *91*, 235–252. [CrossRef]

30. Guo, J.; Liu, X.; Li, S.; Wang, Z. Bearing Intelligent Fault Diagnosis Based on Wavelet Transform and Convolutional Neural Network. *Shock Vib.* **2020**, *2020*, 6380486. [CrossRef]

31. Chen, J.; Huang, R.; Zhao, K.; Wang, W.; Liu, L.; Li, W. Multiscale Convolutional Neural Network with Feature Alignment for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3517010. [CrossRef]

32. Jiao, J.; Zhao, M.; Lin, J.; Zhao, J. A multivariate encoder information-based convolutional neural network for intelligent fault diagnosis of planetary gearboxes. *Knowl.-Based Syst.* **2018**, *160*, 237–250. [CrossRef]

33. Feng, J.; Yaguo, L.; Na, L.; Saibo, X. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mech. Syst. Signal Process.* **2018**, *110*, 349–367. [CrossRef]

34. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [CrossRef]

35. Wang, J.; Mo, Z.; Zhang, H.; Miao, Q. A Deep Learning Method for Bearing Fault Diagnosis Based on Time-Frequency Image. *IEEE Access* **2019**, *7*, 42373–42383. [CrossRef]

36. Nassim, L.; Nida, S.O.; Sami, O. Support Vector Machines for Fault Detection in Wind Turbines. *IFAC Proc. Vol.* **2011**, *44*, 7067–7072. [CrossRef]

37. Jürgen, S. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]

38. Muralidharan, V.; Sugumaran, V. A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Appl. Soft Comput.* **2012**, *12*, 2023–2029. [CrossRef]

39. Liu, R.; Yang, B.; Zhang, X.; Wang, S.; Chen, X. Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *75*, 345–370. [CrossRef]

40. Savita, A.; Amit, C. Hybrid CNN-SVM Classifier for Handwritten Digit Recognition. *Procedia Comput. Sci.* **2020**, *167*, 2554–2560. [CrossRef]

41. Huan, H.; Natalie, B.; Ming, L. Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction. *J. Sound Vib.* **2018**, *414*, 43–60. [CrossRef]

42. Huan, H.; Natalie, B. Bearing vibration data collected under time-varying rotational speed conditions. *Data Brief* **2018**, *21*, 1745–1749. [CrossRef] [PubMed]

43. Geng, H.; Peng, Y.; Ye, L.; Guo, Y. Fault identification of rolling bearing with variable speed based on generalized broadband mode decomposition and distance evaluation technique. *Digit. Signal Process.* **2022**, *129*, 103662. [CrossRef]

44. Lu, F.; Tong, Q.; Feng, Z.; Wan, Q.; Li, Y.; Wang, M.; Cao, J.; Guo, T. Explainable 1DCNN with demodulated frequency features method for fault diagnosis of rolling bearing under time-varying speed conditions. *Meas. Sci. Technol.* **2022**, *33*, 095022. [CrossRef]

45. Jin, Z.; Xiao, Y.; He, D.; Wei, Z.; Sun, Y.; Yang, W. Fault diagnosis of bearing based on refined piecewise composite multivariate multiscale fuzzy entropy. *Digit. Signal Process.* **2023**, *133*, 103884. [CrossRef]

46.   Che, C.; Wang, H.; Xiong, M.; Ni, X. Few-shot fault diagnosis of rolling bearing under variable working conditions based on ensemble meta-learning. *Digit. Signal Process.* **2022**, *131*, 103777. [CrossRef]

47.   Chu, C.; Ge, Y.; Qian, Q.; Hua, B.; Guo, J. A novel multi-scale convolution model based on multi-dilation rates and multi-attention mechanism for mechanical fault diagnosis. *Digit. Signal Process.* **2022**, *122*, 103355. [CrossRef]