NEAT: A Resilient Deep Representational Learning for Fault Detection Using Acoustic Signals in IIoT Environment

JARWAR, Aslam <http://orcid.org/0000-0002-5332-1698>, KHOWAJA, Sunder Ali <http://orcid.org/0000-0002-4586-4131>, DEV, Kapal <http://orcid.org/0000-0003-1262-8594>, ADHIKARI, Mainak and HAKAK, Saqib <http://orcid.org/0000-0002-6670-1032>

NEAT: A Resilient Deep Representational Learning for Fault Detection using Acoustic Signals in IIoT Environment

Muhammad Aslam Jarwar*, *Member IEEE*, Sunder Ali Khowaja, Kapal Dev*†, *Member IEEE*, Mainak Adhikari, and Saqib Hakak

***Abstract*—Fault diagnostics involving the Internet of Things (IoT) sensors and edge devices is a challenging task due to their limited energy and computational capabilities. Another challenge concerning IoT sensors or devices is the incursion of noise when used in an industrial environment. The noisy samples affect the decision support system that could lead to financial and operational losses. This paper proposes a Noisy Encoder using Artificial Intelligence of Things (NEAT) architecture for fault diagnosis in IoT edge devices. NEAT combines autoencoders and Inception module to co-train the clean and noisy samples for solving the said problem. Experimental results on benchmark datasets reveal that the NEAT architecture is noise resilient in comparison to the existing works. Furthermore, we also show that the NEAT architecture has lightweight characteristics as it yields a lower number of parameters, weight storage, training, and testing times that support its real-life applicability in an Industrial IoT environment.**

***Index Terms*—Deep learning, industrial internet of things, representational learning, intelligence of things, noisy encoders, fault diagnosis and maintenance.**

I. INTRODUCTION

FAULT detection is one of the important tasks that are widely used in industrial maintenance at a larger scale. Conventionally, fault detection is performed in three stages: (1) data acquisition from Internet of Things (IoT) sensors; (2) feature extraction to represent the raw data into meaningful patterns; and (3) fault classification using Artificial Intelligence (AI) techniques. Some studies consider one more step which is the pre-processing in case if the data needs to be made compliant with the underlying system [1], [2]. For instance, transmission of data through edge devices, broadcasting, sending signals to the administrator remotely via 5G communication systems, and so forth. The four step model (including pre-processing) can be categorized as Artificial Intelligence of Things (AIoT) for industrial processes. The aforementioned workflow is mostly considered to be feasible for laboratory experiments rather than for daily process operation due to the fact that very few samples regarding faults are available. Furthermore, acquisition of data from IoT sensors or transmission of the data through 5G communication systems may introduce noise in the data samples that can affect the end decision obtained using the opted AI method. Nevertheless, extensive studies have been conducted on fault detection for proactive maintenance which can decrease the operation and maintenance cost, while maximizing productivity and profit [3], [4].

Existing studies representing the hypothetical AI models for fault detection mostly take into account two levels [5]. The first level is concerned with the data collection and analytics aspect that requires the acquisition of symptotic data or extracting relevant features for fault detection. The second level deals with data processing, analyzing, and applying the machine learning models on the data acquired from IoT. In the literature, features of gearbox faults are acquired through vibration signals, temperature level, lubrication condition, and acoustic signals [6], [7], [8], [9]. Authors [10], have suggested that the acoustic signals are more effective as compared to the vibration signals for early diagnosis of fault. However, results of current studies [5], [7], shows that the vibration signals are more effective for the prediction of fault when the signals are stationary, while the acoustic signals are more suitable in the non-stationary loads.

Initially, machine learning techniques such as support vector machines and Artificial Neural Networks (ANN) are used to detect and diagnose the fault from the data acquired from IoT sensors [11], [12]. But later the deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) become popular due to automatic feature extraction and better optimization support [13], [14]. However, the CNNs characterized by serial architectures such as VGG16 and VGG19, are computationally complex [15] which is not compliant with the industrial environment as it requires low computationally complex systems.

As mentioned earlier that one of the hindrances for early prediction of fault in Industrial Internet of Things (IIoT) systems is the background noise which is common in the industrial workspace. An example of representational learning in an IIoT environment is shown in Figure 1. Many studies have shown that such noise can affect the data acquired

*Corresponding Author
† Senior Author
Muhammad Aslam Jarwar with University of Manchester, UK. (email: aslam.jarwar@manchester.ac.uk)
Sunder Ali Khowaja with Faculty of Engineering and Technology, University of Sindh, Jamshoro, Pakistan. (email: sandar.ali@usindh.edu.pk)
Kapal Dev is with the Department of institute of intelligent systems, University of Johannesburg, South Africa.(e-mail:kapal.dev@ieee.org)
Mainak Adhikari with Department of Computer Science, Indian Institute of Information Technology Lucknow, India. (e-mail:mainak@iiitl.ac.in)
Saqib Hakak with University of New Brunswick, Canada (e-mail:saqib.hakak@unb.ca)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3109668, IEEE Internet of Things Journal

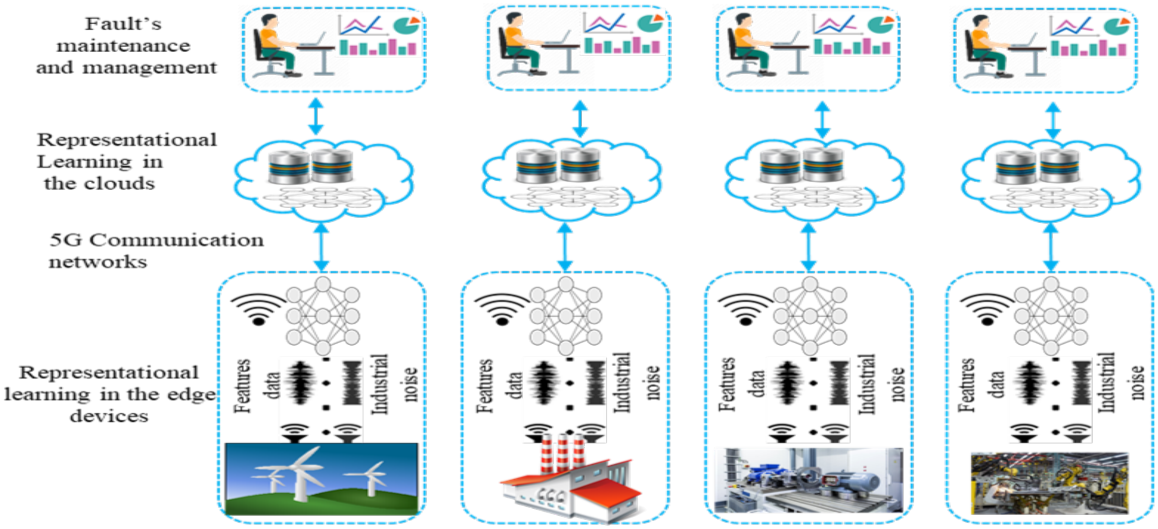

Fig. 1. Example of representational learning in the IIoT

using the IoT sensors, which in turn degrades the diagnostic performance, accordingly. Some of the traditional methods that deal with the filtering of such noise include filtering through frequency responses, wavelet transform, symplectic geometry mode decomposition, singular spectrum analysis, empirical mode decomposition, singular value decomposition, power spectral density, and more [16]. The problem with these methods is that they require additional optimization of parameters (number of parameters depend on the opted method), and the in dependency of the filtering method. The explicit use of filtering makes it difficult to generalize the performance with varying noise levels. The said issue compels the filtering method to be trained explicitly for each noise level [17] that limits the practical applicability of such methods. Furthermore, to the best of our knowledge, none of the methods have integrated the noise removal and detection tasks for fault detection in the gearbox maintenance within a single learning network. A single learning network refers to an end-to-end learning process without indulging extra processes or inference rules to deal with secondary tasks, such as noise removal.

In order to implicitly remove the noise and naturally augmenting the data for the prediction of faults, we propose a Noisy Encoders using Artificial Intelligence of Things (NEAT) architecture that uses end-to-end learning strategy for analyzing, transforming, and training of a deep learning model on the limited data acquired from the IoT sensors. We used three publicly available datasets to evaluate the proposed work for the fault analytics and predictive maintenance. The raw acoustic data was arranged in a multi-dimensional time and frequency dimension matrix with multiple channels to train with the deep learning model. The NEAT architecture co-trains the clean and noisy samples to make the detection method noise resilient as well as generalizing a consistent performance on multiple datasets with limited samples. The contributions of this work are summarized as follows:

- The transformation and representation of raw acoustic signals to multi-dimensional matrices.
- Injection of random noise to the samples for simulating industrial workspace environment.
- We proposed NEAT architecture that integrates the noise modeling within the training process.
- The state-of-the-art results are reported on three publicly available datasets for fault analytics.

The paper is organized as follows: Section II describes the literature review; Section III presents the proposed methodology; Section IV discusses the experimental results on three publicly available and benchmark datasets for fault detection; Section V concludes the article.

II. Related Work

The wind turbine gearboxes are a bunch of mechanical components integrated into a system where these components work together in the environment of lubrication, sound, vibration, and electrical pulses [9]. In order to observe the condition of gearboxes automatically and to detect the faults, various methodologies have been proposed. Authors in [9], mentioned that the fault in the gearbox can be detected by analyzing the quality of lubrication, acquisition, and analysis of sound emissions, by sensing the vibration among the components in the gearbox and electrical pluses analysis. The traditional methods such as onsite fault diagnosis and maintenance or analyzing faults automatically at the central locations needs resources and computing power. Additionally, the collected data from remote devices might be corrupt and lost due to wireless communication channels [2]. Our article aims to provide a learning model to detect faults with high accuracy even though the received data are corrupt while yielding a lower number of parameters and fast testing times.

To transform and represent the acoustic or vibration-based features in the frequency domain and to diagnose the faults. The Fourier Transform (FT) and Short Time Fourier Transform (STFT) signal processing techniques have been widely used in the literature for transformation, representation, and diagnosis

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3109668, IEEE Internet of Things Journal

of acoustic or vibration-based faults [16]. Chuan *et al.* [18] used the acoustic emission sensors and accelerometers to capture the gearbox fault features from sound and vibration signals, respectively. The study used a random forest algorithm to classify the fault patterns.

Recently, deep learning has been adapted extensively for the diagnosis of fault from acoustic modality. Xueyi *et al.* [13] used the CNN and Gated Recurrent Unit (GRU) models to analyze the acoustic and vibration signals for the classification of various conditions of pitting. The study proposed to train a one-dimensional CNN model with acoustic features followed by the stacking of GRU to model the temporal dependencies of vibration data. Yong *et al.* [19], used an end-to-end CNN model for diagnosing the fault in the gearbox from acoustic signals. The authors extracted the features from the load and non-load situation using the gearbox and then trained the CNN model by integrating the time-domain and frequency-domain data. Cao *et al.* [20] adopted a transfer learning approach to diagnosing the gear fault. The motivation behind their work is to use small scale datasets while generalizing the accuracy with the help of pre-trained network architectures. Yu *et al.* [21] considered the fusion of acoustic and vibrational modalities for training a CNN to improve the fault diagnosis performance. The fusion of the aforementioned data modalities was based on improved evidence theory to attain reliable fault diagnostics. Zhuang *et al.* [5] worked on filtering the noise and extracting clean features from the noised vibration signals. The filtered data is used with 1-D CNN without crafting features. However, to improve the model efficacy and accuracy they developed multi-scale kernels (an attentive kernel residual network) that extracts multi-scale features. In contrast to their approach, we added the noise in the data and co-train the model in order to make our model resilient in an IIoT environment. Wang *et al.* [22] focused on the characteristics of the fault diagnostic system within the context of intelligent manufacturing rather than improving the performance of fault diagnosis. Their study highlights the topology structure, network environment, and operational framework of fault diagnosis using the IoT approach. Wang *et al.* [23] considered the case study of the IIoT environment that needs to be less computationally complex in order to have real-life applicability. In this regard, the study proposed a lightweight CNN by introducing depthwise separable convolutional layers for reducing the training and inference time, accordingly. Chen *et al.* [24] suggested the integration of semantics in the fault diagnosis approach while omitting the use of data preprocessing. The approach transforms the data into temporal logic formulas that can be interpreted by humans as well. Furthermore, the temporal logic is then modeled using Markov decision processes and reinforcement learning to detect the faults. Over the years, many studies have been proposed for fault diagnosis, in general, using CNNs. For instance, Zhang *et al.* [25] used deep learning algorithm to diagnose the fault patterns from turbofan dataset. Weimer *et al.* [26] employed CNN to identify the defects through an industrial inspection. Ince *et al.* [27] employed 1-D CNN for the monitoring of motor condition. Abdeljaber *et al.* [28] also used CNN to detect the damage in real-time and achieved satisfactory results.

Considering the above-mentioned studies, the sole focus is to either transform the data/feature space or to alter the pre-trained network architectures for improving the fault diagnosis process. It should be noted that all the works are centered towards the IIoT environment one way or another but they have not considered the problem of noise integration that is very common in the given context. Cheng *et al.* [16] recently, proposed a noise reduction method for the gear fault diagnosis using the adaptive weighted symplectic geometry decomposition method. However, their method urges that denoising should be performed explicitly and independently of the detection process. Considering the real-time operations, such methods could likely improve the performance while adding computational overhead.

In contrast, this study proposes a network architecture that practically yields lower computational complexity (due to inception layers) and incorporates the modeling of noise parameters within the training process to improve the fault diagnosis. Provided that the noise reduction method is integrated within the detection framework, we assume that it yields lower computational overhead, and thus, is suitable for the IIoT environment.

III. PROPOSED METHODOLOGY

The methodology section for the fault diagnosis comprises of four parts as shown in Fig. 2. The first part acquires the acoustic signals in time domain and extracts the frequency information to form a sample matrix. The sample matrices are then stacked to form a multi-dimensional matrix representing the acoustic signals with multiple channels. The second stage injects a random Gaussian noise to the multi-dimensional matrix for corrupting the signals. The third stage considers both the clean and noisy signals to be trained with the proposed network architecture. In the fourth stage, we perform quantitative experiments to evaluate the proposed work for fault diagnosis in terms of accuracy, processing times, number of parameters, and weight storage, accordingly.

A. Data Preprocessing and Transformation

This subsection deals with the acquisition of time and frequency information and its transformation to multi-dimensional matrix, accordingly. Let us consider the time domain information for each sample is represented by a vector τ such that $\tau = [td_1, \ldots, td_k]$ where k is the sequential parameter. However, the vector τ represents a single sample. The stacked acoustic samples can be represented as a sample matrix $M = [\tau_1, \ldots, \tau_J]$ where J represents the number of samples, respectively. In order to obtain frequency information, Fourier transform is applied to the sample matrix M. We concatenate the time and frequency domain information in a single matrix T as shown in equation 1.

$$T = \begin{bmatrix} td_{11} & \ldots & td_{1k} & \ldots & fd_{11} & \ldots & fd_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ td_{j1} & \ldots & td_{jk} & \ldots & fd_{j1} & \ldots & fd_{jp} \end{bmatrix} \quad (1)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3109668, IEEE Internet of Things Journal

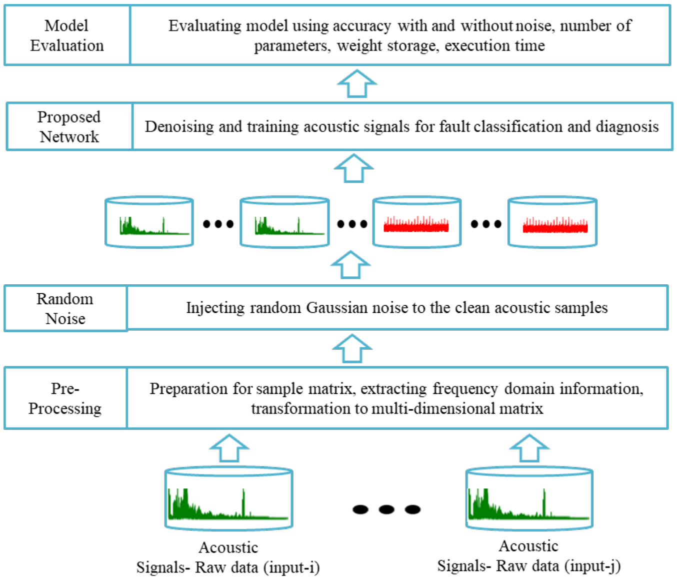

Fig. 2. Proposed workflow for fault analytics using AIoT

where $T \in \mathbb{R}^{j\times(k+p)}$ and is referred to as a single channel time and frequency information matrix. In order to construct a multi-dimensional matrix, the sample matrices T are stacked together to form Ω such that $\Omega = [T_1, \ldots, T_q]$ where T_q refers to the q-th channel of $j \times (k + p)$ acoustic signal sample matrix. The transformed multi-dimensional matrix is analogous to the Red Green Blue (RGB) image characteristics that can directly be trained deep network architectures.

B. Random noise injection

Existing studies have proved that the cleansing of noisy data is one of the fundamental steps for IIoT sensor data analytics [16]. However, most of the methods perform this step in an explicit manner that increases the computational overhead. In this work, we intend to design a network architecture that could handle the noisy data while improving the fault detection performance, simultaneously. We add noise and handle it on our own way in order to develop a noise resilient network architecture for noisy IIoT environment that can detect industrial faults with improved precision and lower computing overhead.

In this regard, we employ the random Gaussian noise injection method to the multi-dimensional matrix T, respectively. The reason for choosing the Gaussian noise is two-fold. The first is the wide usage and acceptability of Gaussian noise in the existing studies [16], the second is the ease of generating the said type of noise that will be helpful in reproducing the given results, accordingly. The formulation for adding random Gaussian noise is given in equation 2.

$$\hat{T} = T + \gamma \cdot \delta; \delta \sim \mathcal{N}(0, \sigma^2), \tag{2}$$

where $\hat{T}$ corresponds to the noisy multi-dimensional matrix and δ refers to the random noise term that follows the Gaussian distribution with zero mean and σ^2 variance. The scale of the noise term is regulated through the parameter γ that can be varied to analyze the noise effect on the fault diagnosis. The noise is added to the samples after they are transformed via multi-dimensional matrix. The selection of the variance and the regulation term for adding noise is performed on an empirical basis suggesting that noise is varied in conjunction with the detection accuracy, i.e. [40% - 100%]. The noise levels that affect the detection accuracy to fall lower than 40% are not considered in this study. We employed a library for recognition and organization of speech and audio (librosa) to add random noise with varying scales.

C. Network Architecture

The proposed network architecture is shown in Fig. 3. As it can be noticed that the clean and noisy samples undergo a denoising autoencoder which serves not only the purpose of denoising the acoustic signals but also extracts feature encodings that can be further used to train the samples for fault detection. Our main focus is to diagnose the fault from acoustic signals given that the samples can be corrupted by noise. However, one can obtain the denoised samples using the decoding part of the autoencoder as shown in Fig. 3. The main task of denoising autoencoders is to train such a kind of architecture that can reconstruct/clean the noisy samples. In process of learning reconstruction, the autoencoders generate embeddings that could be used as learnable patterns or features. A study concluded that the denoising autoencoders imbued with random Gaussian noise can approximate the gradient of the input's density function as shown in equation 3.

$$\frac{\partial log\rho(w)}{\partial w} \approx (\mathcal{A}(w) - w)) \cdot \frac{1}{\gamma \cdot \delta}, \tag{3}$$

where $\rho(w)$ is the density function's energy, $\mathcal{A}(.)$ refers to the autoencoder, and w refers to the optimizable weights, respectively. The training process compels the autoencoder to learn the vector field in the direction of input data. In our study, we are interested in the denoising of input samples but also the embeddings that can be used as learnable patterns or vector fields. The vector fields can support the extraction of learning features and handling of noisy samples, simultaneously [16]. In this regard, the denoising autoencoder in the proposed architecture learns to approximate

$$\frac{\partial log\rho(w|X)}{\partial w} \approx (\mathcal{A}(w) - w)) \cdot \frac{1}{\gamma \cdot \delta}, \tag{4}$$

In equation 4. $X = T, \hat{T}$ and $\rho(w|X)$ is the conditional distribution of the weights given the set of clean and noisy samples. We want to make the inputs similar to the clean samples, thus, use a gradient ascent approach iteratively for approximating the conditional distribution $\rho(w|X)$ as shown in equation 5.

$$w \leftarrow eps \cdot \frac{\partial log\rho(w|X)}{\partial w} + w = eps \cdot (\mathcal{A}(w) - w) + w \tag{5}$$

The eps is the epsilon value representing the step size of the gradient ascent. Once we obtain the embeddings from the encoding layers, the network architecture performs feature aggregation to combine the patterns from both the clean and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3109668, IEEE Internet of Things Journal

noisy samples. In this regard, we use the relation-net based aggregation function as proposed in [29]. The original study performs the aggregation function as shown in equation 6.

$$X^t = \sum w_{ij} \cdot \gamma^l(X_i^l, X_j^l) \tag{6}$$

In equation 6, $\gamma^l(X_i^l, X_j^l)$ refers to the computation for combining the information from two neighboring nodes in conjunction with their corresponding weights. We modify the computation for combining the two embeddings rather than the message vectors. The $\gamma^l(\cdot)$ adds the output by forwarding the embeddings through a fully connected linear layer. The output then undergoes BatchNorm, Dropout, and ReLU layers. Note that the aforementioned arrangement of layers for feature aggregation is similar to its original configuration [29].

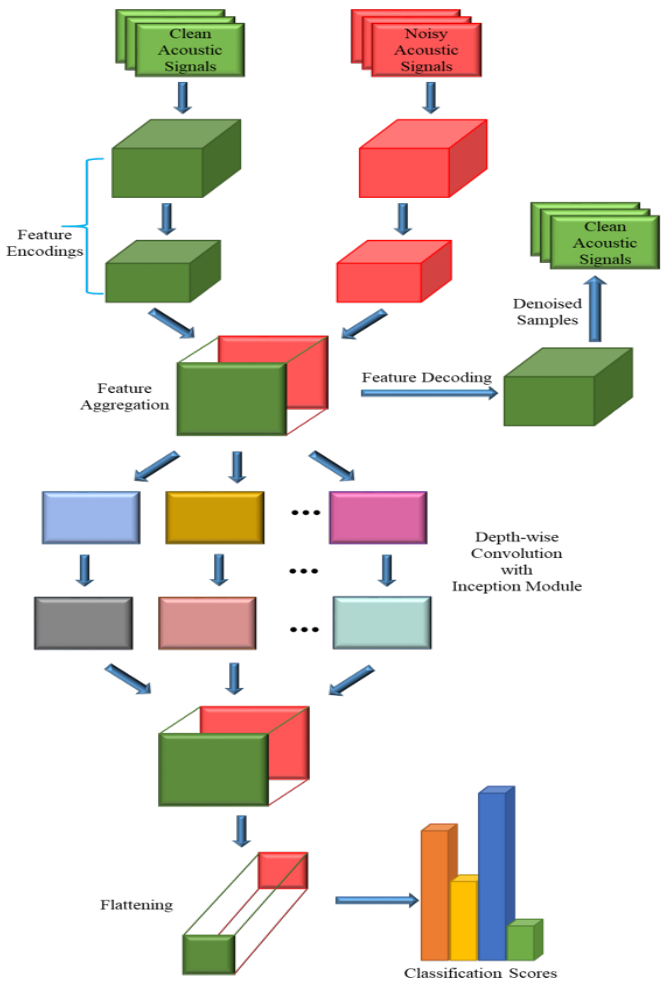

Fig. 3. Noisy encoders using artificial intelligence of things (NEAT) architecture for fault diagnosis

In order to classify and predict the fault in the gearbox, the Inception module of the Inceptionv3 model has been applied (refer to Fig. 3). The rationale for choosing the aforementioned model is two-fold: The first refers to the computational complexity incurred by the Inception module as it yields lower computation time than Serial convolutional and residual network architectures. The second is the banks of bandpass and low-pass filter representations in their weight distributions [30]. The selection of the Inception module in the proposed architecture naturally provides an edge to deal with the random noise problem by applying the bandpass and low-pass filters while modeling noise parameters explicitly within the training process. It should be noted that we used Inception block A for fine-tuning in Inceptionv3 architecture and froze the remaining layers, accordingly. The depth-wise convolution using the inception module first performs the dimensionality reduction of the feature maps and then uses standard 1x1 convolution to increase the dimensions. The depthwise convolution is most suitable for high dimensional inputs that can increase the expressive capability of the network to model the task at hand. In order to train the proposed network, we used a hierarchical search space [23] that uses the decomposed blocks and slowly increases the kernel size while reducing the number of filters. This not only reduces the computational overhead in the training process but also ensures the performance boost for the detection of gear faults. In the proposed network architecture, the dropout ratio is set to 0.35, the δ is set to 0.5 while the γ value varies from 0.1 to 0.9. The γ value yielding the lowest accuracy was selected to report the results in this study. We used the ADAM optimizer with default settings. The learning rate is set to 0.001 with a decay rate of 0.05 after every 30 epochs. The weight decay and the momentum was set to be 3e-4 and 0.73, respectively. The eps value was set to be 0.6, accordingly. We use the cross-category entropy loss to train the inception module for classifying the nature of gear fault based on the employed dataset. The Inceptionv3 part of the NEAT architecture has been trained for 150 epochs. The network architecture details are consistent for all the employed datasets to prove the generalization of NEAT architecture, accordingly.

IV. Experimental Results

In this section, we provide the experimental results for fault diagnosis on multiple datasets with and without the addition of noise. We also compare our method with several pre-trained network architectures to validate its efficacy. All the experiments were carried out using python libraries for deep learning on a machine corei5 clocked at 3.4 GHz, RTX 1080Ti GPU, and 32GB of RAM. We deployed scalable data and representational learning server as Infrastructure-as-a-Service (IaaS) in a private cloud environment installed with the OpenStack platform. We evaluated our method in terms of accuracy, computation time, and number of parameters accordingly. The reasons for using accuracy instead of other metrics are two-fold. The first is the wide acceptance of accuracy metric in the field of AI and the second is the fair comparison with existing studies. Existing studies report their methods' performance in terms of accuracy, therefore, using the same metric allow us to asses the performance and compare it in a fair manner.

A. *Case Western Reserve University Dataset*

The aforementioned dataset [31] is widely used for mechanical fault diagnosis and is considered to be a good dataset to verify the deep learning approach [32]. The Case Western Reserve university dataset testing board comprises

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3109668, IEEE Internet of Things Journal

control electronics, a dynamometer, a torque sensor, and a 2-horsepower motor. More details on the dataset can be obtained from [31]. We follow the train-test protocol as adopted by some of the recent studies.

We used 70% and 10% of the dataset comprising of 10 fault types for training and validation, respectively. The remaining 20% was used for testing and evaluating the proposed approach. Besides, we evaluate some existing approaches including Support Vector Machine (SVM), AlexNet, LeNet, GoogleNet, ShuffleNet, DarkNet, Xception, and ResNet for the diagnosis. We also evaluated the same methods and architectures while adding noisy samples. The γ value was set to be 0.8 for the experiments on the aforementioned dataset. The results are reported in Fig. 4. Considering the literature review, some studies suggest that the LSTMs are better at dealing with time-series data due to their ability of handling temporal streams. However, addition of noise heavily distorts the temporal stream characteristics which in turn degrades the performance of such models. In order to prove the above-mentioned assumption, we compare the results with LSTM networks as well. For the sake of generality, we used grid search method for all of the above-mention pre-trained networks and shallow learning techniques to select the optimal parameters, accordingly. It can be observed from the obtained results that the proposed method outperforms the existing works even when the samples are corrupted by noise with a large margin. Furthermore, it was observed during training that AlexNet, GoogleNet, and LeNet suffer from vanishing gradient problem that eventually resulted in decreased performance. However, unlike the study [23] we achieved better results for the aforementioned pre-trained architectures by tweaking some parameters. The existing works also achieve good results when adding the noise but the performance gets decreased by almost 24% and 27%, respectively. It should be noted that the results from [19] and [23] are reproduced based on our understanding of their implementation. ShuffleNet yields the lowest number of parameters and weight storage, i.e. 0.97 million and 4.385 MB, respectively, whereas the proposed approach yields 2.73 million parameters and 10.425 MB of storage, putting it to be the lowest in comparison to the other works. Furthermore, the proposed work yield a training time of 26.54 minutes and a testing time of 29.32 seconds, accordingly.

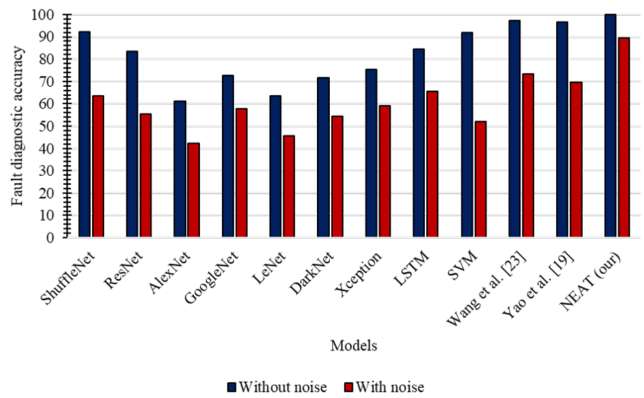

Fig. 4. Fault diagnostic result on Case Western Reserve University dataset and comparison with existing approaches

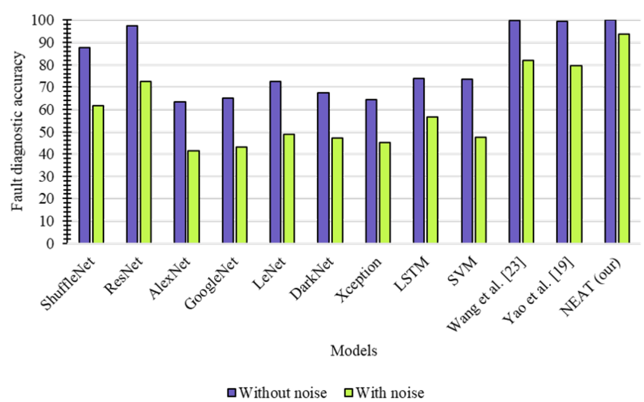

Fig. 5. Result on MFPT dataset and comparison with existing approaches

B. MFPT Bearing Fault Dataset

The Mechanical Failure Prevention Technology (MFPT) dataset was prepared by the NRG system's chief engineer [33]. The dataset includes inner race failure, outer race failure, and baseline bearing data under various loads. We follow the data preparation protocol as performed in [23]. We use the same pre-trained architectures, existing models, and works for the evaluation on the MFPT dataset. The ratio of training, validation, and test set has been kept the same, i.e. 70%, 10%, and 20%, respectively. The γ value was set to be 0.5 for the experiments on the MFPT dataset. The results are reported in Fig. 5. The proposed method outperforms the existing architectures and the methods with and without noise, significantly. It can be noticed from results that AlexNet, GoogleNet, and LeNet, underperform in general. One of the assumptions for such results is the less volume of training data that can hinder the performance. We also assume that the training strategy including both noisy and clean samples is a natural way of data augmentation that helps the proposed method in achieving the best performance with limited data availability. Similar to the previous dataset, the ShuffleNet yields the lowest number of parameters and weight storage, i.e. 0.97 million and 4.37 MB, followed by the proposed architecture 3.49 million and 14.36 MB, respectively. The proposed method takes almost 7 minutes to train the network while 18.3 seconds to test.

C. Gearbox Acoustic Dataset

To predict the faults in the gearbox of a motor, we used the publicly available gearbox acoustic dataset [19]. The dataset was acquired with four types of acoustic sensors at the 16000 Hz sampling frequency. During the acquisition of data, the speed of the motor was set at 1800 revolutions per minute. In the data acquiring experiments, multiple types of audio sensors were used and installed at various suitable locations around the motor with the gearbox in order to capture the high-quality acoustic samples for better prediction and classification accuracy.

The dataset contains two types of samples: the first set of samples contain four types of features captured during the no-load on the motor. Similarly, the second set of samples also contains four types of features captured when the gearbox was used with the load. The manual programmable brake was used to simulate the non-load and load states of the gearbox. The four features in each type of samples include normal, pitting, tooth fracture, and wear categories, respectively. In each feature such as normal, pitting, etc., there are four types of input samples; and each input type contains 40 acoustic files with 60 seconds length. We follow the same protocol and use the ratio of 70%, 10%, and 20% data for training, validation, and testing, respectively. We evaluate the same architectures and existing methods for evaluation and comparison, accordingly. The γ value was set to be 0.7 for this experiment. The experimental results on the gearbox acoustic dataset are reported in Fig. 6. It is revealed that the proposed work achieves better performance in general, in comparison to the existing works and the pre-trained network architectures with and without the addition of noise. Furthermore, the proposed network architecture consistently yields a low number of parameters, weight storage, training, and testing times in comparison to the existing works and architectures, except for ShuffleNet. For the gearbox acoustic dataset, the number of parameters and weight storage yielded by the proposed work is 3.86 million and 15.9 MB, respectively. The training and test time on the gearbox acoustic dataset was noted to be 5.9 minutes and 16.43 seconds.

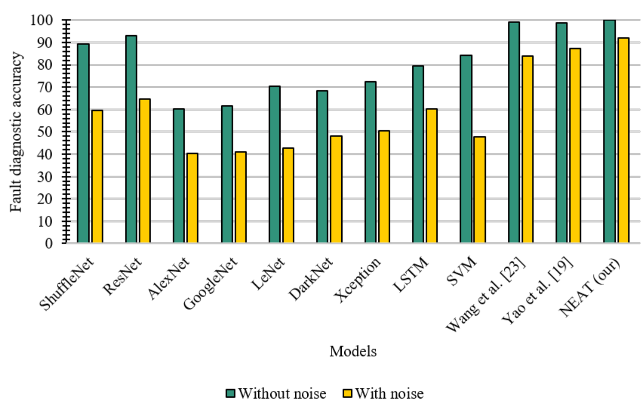

Fig. 6. Result on gearbox acoustic dataset and comparison with existing models

V. Conclusion

In this article, we proposed a novel architecture, i.e. Noisy Encoders using Artificial Intelligence of Things (NEAT) for the classification and prediction of faults from Internet of Things (IoT) sensors. To make the study compliant with the Industrial Internet of Things (IIoT) environment and real-life applicability we proposed NEAT architecture that not only performs the detection well enough but also able to handle the noise that is considered to be unavoidable. The proposed workflow based on Artificial Intelligence of Things (AIoT) includes four stages, i.e. data preprocessing, data transformation, random noise injection, and training of the multi-dimensional matrix. The NEAT architecture considers both noisy and clean samples and passes them through the autoencoders. The representation from autoencoders for both clean and noisy samples are aggregated and trained using the Inceptionv3 network (Inception module A) for fault diagnosis.

TABLE I
List of Abbreviations

NEAT	Noisy Encoders using Artificial Intelligence of Things
IIoT	Industrial Internet of Things
IoT	Internet of Things
AIoT	Artificial Intelligence based Internet of Things
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
LSTM	Long Short-Term Memory Networks
STFT	Short Time Fourier Transform
VGG	Visual Geometry Group
GRU	Gated Recurrent Unit
BatchNorm	Batch Normalization
ReLU	Rectified Linear Units
ADAM	Adaptive Moment Estimation
MFPT	Mechanical Failure Prevention Technology

We performed extensive experiments on three publicly available and benchmark datasets to demonstrate the efficacy of the proposed approach. It is shown that the NEAT architecture not only is better at detection performance but also yields the lowest number of parameters, weight storage, training, and testing times, after the ShuffleNet. It is also assumed that the inclusion of noisy samples in the training process is a natural way of augmentation that not only makes the NEAT architecture noise resilient but also helps the network to cope with low data volume. The experimental outcome obtained by the proposed approach proves to be suitable for its applicability in the IIoT environment. We believe that the deep learning approach in this study for classification and prediction of gearboxes' faults from their acoustic emission will facilitate in the unmanned monitoring management of proactive maintenance scheduling and increase the production of machines.

This work integrates the noisy samples to train the NEAT architecture in an end-to-end learning approach. Although, we propose an extended decoding part to denoise the samples but we did not conducted experiments to show its efficacy which can be considered as one of the limitations to this approach. As a future work, we intend to propose the parameter selection strategy that can achieve the generalization across the NEAT architecture without fine-tuning each of the network blocks individually. Furthermore, we also intend to develop a self-supervising model that enables multi-agents to diagnosis new types of faults and recovery of devices using AIoT approach.

References

[1] L. Feng and C. Zhao, "Fault Description Based Attribute Transfer for Zero-Sample Industrial Fault Diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1852–1862, 2020.

[2] Y. Peng, W. Qiao, and L. Qu, "Compressive sensing-based missing-data-tolerant fault detection for remote condition monitoring of wind turbines," *IEEE Transactions on Industrial Electronics*, 2021.

[3] S. A. Khowaja and P. Khuwaja, "Q-learning and LSTM based deep active learning strategy for malware defense in industrial IoT applications," *Multimedia Tools and Applications*, pp. 1–27, 2021.

[4] K. Dev, R. K. Poluru, L. K. R, P. K. R. Maddikunta, and S. A. Khowaja, "Optimal Radius for Enhanced Lifetime in IoT using Hybridization of Rider and Grey Wolf Optimization," *IEEE Transactions on Green Communications and Networking*, pp. 1–8, 2021.

[5] Z. Ye and J. Yu, "Akrnet: A novel convolutional neural network with attentive kernel residual learning for feature learning of gearbox vibration signals," *Neurocomputing*, 2021.

[6] D. Cevasco, S. Koukoura, and A. Kolios, "Reliability, availability, maintainability data review for the identification of trends in offshore wind energy applications," *Renewable and Sustainable Energy Reviews*, vol. 136, p. 110414, 2021.

[7] I. Vamsi, G. Sabareesh, and P. Penumakala, "Comparison of condition monitoring techniques in assessing fault severity for a wind turbine gearbox under non-stationary loading," *Mechanical Systems and Signal Processing*, vol. 124, pp. 1–20, jun 2019.

[8] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3757–3767, jun 2015.

[9] V. Inturi, N. Shreyas, K. Chetti, and G. Sabareesh, "Comprehensive fault diagnostics of wind turbine gearbox through adaptive condition monitoring scheme," *Applied Acoustics*, vol. 174, p. 107738, 2021.

[10] J. P. Salameh, S. Cauet, E. Etien, A. Sakout, and L. Rambault, "Gearbox condition monitoring in wind turbines: A review," *Mechanical Systems and Signal Processing*, vol. 111, pp. 251–264, oct 2018.

[11] W. Liu, B. Tang, J. Han, X. Lu, N. Hu, and Z. He, "The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 466–472, apr 2015.

[12] W. Li, S. Zhang, and G. He, "Semisupervised Distance-Preserving Self-Organizing Map for Machine-Defect Detection and Classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 869–879, may 2013.

[13] X. Li, J. Li, Y. Qu, D. He, X. Li, J. Li, Y. Qu, and D. He, "Gear Pitting Fault Diagnosis Using Integrated CNN and GRU Network with Both Vibration and Acoustic Emission Signals," *Applied Sciences*, vol. 9, no. 4, p. 768, feb 2019.

[14] M. A. Jarwar and I. Chong, "Web Objects Based Contextual Data Quality Assessment Model for Semantic Data Application," *Applied Sciences*, vol. 10, no. 6, p. 2181, mar 2020.

[15] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of Residual Connections on Learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, pp. 4278–4284.

[16] J. Cheng, Y. Yang, N. Hu, Z. Cheng, and J. Cheng, "A noise reduction method based on adaptive weighted symplectic geometry decomposition and its application in early gear fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 149, p. 107351, feb 2021.

[17] S. Yin, C. Liu, Z. Zhang, Y. Lin, D. Wang, J. Tejedor, T. F. Zheng, and Y. Li, "Noisy training for deep neural networks in speech recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, p. 2, dec 2015.

[18] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems and Signal Processing*, vol. 76-77, pp. 283–293, aug 2016.

[19] Y. Yao, H. Wang, S. Li, Z. Liu, G. Gui, Y. Dan, J. Hu, Y. Yao, H. Wang, S. Li, Z. Liu, G. Gui, Y. Dan, and J. Hu, "End-To-End Convolutional Neural Network Model for Gear Fault Diagnosis Based on Sound Signals," *Applied Sciences*, vol. 8, no. 9, p. 1584, sep 2018.

[20] P. Cao, S. Zhang, and J. Tang, "Preprocessing-Free Gear Fault Diagnosis Using Small Datasets With Deep Convolutional Neural Network-Based Transfer Learning," *IEEE Access*, vol. 6, pp. 26 241–26 253, 2018.

[21] L. Yu, X. Yao, J. Yang, and C. Li, "Gear Fault Diagnosis through Vibration and Acoustic Signal Combination Based on Convolutional Neural Network," *Information*, vol. 11, no. 5, p. 266, may 2020.

[22] M. Wang, Z. Zhang, K. Li, Z. Zhang, Y. Sheng, and S. Liu, "Research on key technologies of fault diagnosis and early warning for high-end equipment based on intelligent manufacturing and Internet of Things," *The International Journal of Advanced Manufacturing Technology*, vol. 107, no. 3-4, pp. 1039–1048, mar 2020.

[23] Y. Wang, J. Yan, Q. Sun, Q. Jiang, and Y. Zhou, "Bearing Intelligent Fault Diagnosis in the Industrial Internet of Things Context: A Lightweight Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 87 329–87 340, 2020.

[24] G. Chen, M. Liu, and Z. Kong, "Temporal-logic-based Semantic Fault Diagnosis with Time-series Data from Industrial Internet of Things," *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2020.

[25] C. Zhang, J. H. Sun, and K. C. Tan, "Deep Belief Networks Ensemble with Multi-objective Optimization for Failure Diagnosis," in *IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, oct 2015, pp. 32–37.

[26] D. Weimer, B. Scholz-Reiter, and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection," *CIRP Annals*, vol. 65, no. 1, pp. 417–420, 2016.

[27] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, nov 2016.

[28] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *Journal of Sound and Vibration*, vol. 388, pp. 154–170, feb 2017.

[29] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-Learning with Memory-Augmented Neural Networks," in *Proceedings of The 33rd International Conference on Machine Learning*, 2016, pp. 1842–1850.

[30] S. A. Khowaja and S.-L. Lee, "Semantic Image Networks for Human Action Recognition," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 393–419, feb 2020.

[31] Case Western Reserve University, "Case Western Reserve University Bearing Data Center," 2012. [Online]. Available: https://shorturl.at/gptzV

[32] D. Neupane and J. Seok, "Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset With Deep Learning Approaches: A Review," *IEEE Access*, vol. 8, pp. 93 155–93 178, 2021.

[33] E. Bechhoefer, "Condition Based Maintenance Fault Database for Testing of Diagnostic and Prognostics Algorithms," 2012. [Online]. Available: https://shorturl.at/hsBJ8

Muhammad Aslam Jarwar is Research Associate at University of Manchester, United Kingdom. He received the Ph.D. in Information and Communications Engineering from Hankuk University of Foreign Studies, South Korea. He is the author of several peer-reviewed articles and ITU-T technical reports. He is the editorial board member of Progress in Human Computer Interaction Journal and the regular reviewer of IEEE Transactions, Elsevier and Springer Q1 journals. His research interests include Internet of Things (IoT), Data Analytics,Data Provenance, Cyber Security, Knowledge Representation, and AI.

Sunder Ali Khowaja received the Ph.D. degree in Industrial Information Systems Engineering from Hankuk University of Foreign Studies, South Korea. He is currently an Assistant Professor at Department of Telecommunication Engineering, University of Sindh, Pakistan. He is also a regular reviewer of notable journals including IEEE Transactions, IET, Elsevier, and Springer Journals. He has also served as a TPC member for workshops in A* conferences such as Globecom and Mobicom. His research interests include Data Analytics & Machine Learning for Computer Vision applications.

Kapal Dev is Senior Research Associate at university of Johannesburg South Africa . Previously, he was a Postdoctoral Research Fellow with CONNECT Centre, Trinity College Dublin. He worked as 5G Junior Consultant & Engineer at Altran Italia S.p.A, Milan, Italy on 5G use cases. He is very active in leading Erasmus + International Credit Mobility Capacity Building for Higher Education H2020 Co-Fund projects. He is serving as Associate Editor in Wireless Networks, IET Quantum Communication, and IET Networks. He is also serving as Guest Editor in COMCOM, COMNET, and Lead chair in CCNC 2021 workshops.

Mainak Adhikari is currently working as an Assistant Professor at IIIT Lucknow. He has completed his Ph.D in Cloud Computing from IIT(ISM) Dhanbad, India in 2019 and postdoctoral fellowship from University of Tartu, Estonia in 2021. He obtained his M.Tech. from Kalyani University in the year 2013. He earned his B. Tech. Degree from West Bengal University of Technology in the year of 2011. He is the Associate Editor of Cluster Computing and Physical Communication journals.

Dr. Hakak is an assistant professor at the Canadian Institute for Cybersecurity (CIC), Faculty of Computer Science, University of New Brunswick (UNB). Having more than 5+ years of industrial and academic experience, he has received the number of Gold/Silver awards in international innovation competitions and is serving as the technical committee member/reviewer of several reputed conference/journal venues.