IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 October 2023, accepted 5 November 2023, date of publication 13 November 2023,
date of current version 17 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3332243

RESEARCH ARTICLE

A Scalo Gram-Based CNN Ensemble Method With Density-Aware SMOTE Oversampling for Improving Bearing Fault Diagnosis

MUHAMMAD IRFAN[1], ZOHAIB MUSHTAQ[2], NABEEL AHMED KHAN[3], SALIM NASAR FARAJ MURSAL[1], SAIFUR RAHMAN[1], MUAWIA ABDELKAFI MAGZOUB[4,5], (Member, IEEE), MUHAMMAD ARMGHAN LATIF[6], FAISAL ALTHOBIANI[7], IMRAN KHAN[2], AND GHULAM ABBAS[8], (Senior Member, IEEE)

[1]Electrical Engineering Department, College of Engineering, Najran University, Najran 61441, Saudi Arabia
[2]Department of Electrical, Electronics and Computer Systems, College of Engineering and Technology, University of Sargodha, Sargodha 40100, Pakistan
[3]Department of Electrical Engineering, Riphah International University, Islamabad 46000, Pakistan
[4]Electrical and Electronics Engineering Department, Sudan Technological University, Omdurman 13315, Sudan
[5]Electrical and Electronics Engineering Department, National University, Khartoum 11111, Sudan
[6]Department of Computer and Information System, Cleveland State University, Cleveland, OH 44115, USA
[7]Faculty of Maritime Studies, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[8]Department of Electrical Engineering, The University of Lahore, Lahore 54000, Pakistan

Corresponding author: Muawia Abdelkafi Magzoub (muawia.magzoub@nu.edu.sd)

This work was supported by the Deanship of Scientific Research, Najran University, Saudi Arabia, through the Distinguished Research Groups, under Grant NU/DRP/SERC/12/8.

ABSTRACT Machine learning (ML) based bearing fault detection is an emerging application of Artificial Intelligence (AI) that has proven its utility in effectively classifying various faults for timely measures. There are myriad studies dedicated to the effective classification of bearing faults under different conditions and experimental settings. In this study, we proposed a weighted voting ensemble (WVE) of three low-computation custom-designed convolutional neural networks (CNNs) to classify bearing faults at 48 KHz. Some of the recent studies have exploited 1-d time-series signals and time-frequency based 2-d transformations for bearing fault classification. However, 1-d signals lack contextual information and higher-dimensional interpretations whereas time-frequency based transformations provide a more appropriate, visually perceivable and explainable representation of the time and frequency changes. Therefore in this study, a scalogram based representation of the signals is leveraged for classification using the CNN. Furthermore, the class imbalance is a significant challenge that affects the modelling behavior and possibly create biases. This study provides a novel density and distance hybrid over-sampling approach namely Density-Aware SMOTE(DA-SMOTE) built upon the SMOTE methodology for a more refined representation of synthetic samples within the minority class distribution. The experimentation procedures were carried out before and after the oversampling and it was observed that the balanced dataset acquired much better accuracy then the imbalanced dataset. This is evident by the fact that the highest validation accuracy for the proposed ensemble method (WVCNN) reached at 0-HP and 1-HP reached 99.28% and 99.13% while for the over-sampled dataset the accuracy soared to 99.71% and 99.87% for 0 and 1-HP respectively. The performance was evaluated for other metrics apart from the accuracy to assess the model's performance in terms of chance occurrences and the class wise performance.

INDEX TERMS Bearing fault detection, deep convolutional neural network, transfer learning, fine tuning, time series, data pre-processing.

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Tedesco.

I. INTRODUCTION

In order to ensure reliable performance of mechanical systems, condition monitoring of bearings is paramount. This

process is heavily reliance upon various data sources and techniques, each of which plays a critical role in detecting potential failures. The key data sources are inclusive of vibration analysis, temperature sensing, lubrication analysis, and acoustic sensing [1]. The vibration analysis is quite instrumental in highlighting changes within the respective vibrational patterns that serve as early fault indicators. Temperature sensors track the heat signatures during bearing operations, further revealing the underlying issues like friction or other lubrication problems. The acoustic sensors extract sound signals produced by the bearings and analyze them for deviations indicating potential fault types [2], [3].

An accurate classification of the detected bearing faults is important to effectively monitor the condition and apprising of the severity of entailing consequences. In such regards, the deep learning techniques, especially the convolutional neural networks (CNN), have time after time demonstrated their efficacy in accurate pin-pointing the fault types for enhanced monitoring [4]. These CNN models excel at processing varying data representations like time-frequency based vibration spectrograms. As deep learning models struggle to find patterns and suffer from over-fitting when the data set is insufficient or monotonous, hence pre-trained models provide a substantial edge by cutting the training costs and the accumulation of large data-sets. Transfer learning, a strategy that fine-tunes pre-trained neural network models, expedites model training, particularly when dealing with limited labeled data. Non-conventional approaches, such as Siamese networks and graph neural networks, provide unique perspectives on bearing fault classification. Siamese networks facilitate one-shot learning, while graph neural networks capture intricate dataset relationships [5], [6].

However, the bearing faults classification comes with some challenges, the primary one being the class-imbalance which arises due to varying distribution of faults in real-world systems. Some of the faults, like wear and misalignment are marked as common, while the others like corrosion or brinelling, are much rarer. Additionally, the data collection biases have a tendency to further complicate this class imbalance, as certain faults classes are more receptive during inspections, ultimately leading to over-representation in the dataset [1].

Hence an effectively addressed class imbalance remains pivotal for achieving accurate fault classification results. There are several strategies that are employed to mitigate this imbalance issue and consequently improve the fault classification accuracy [4]. The strategies that are commonly employed include oversampling, which increases the minority class instances, while under-sampling reduces the majority class samples to match the minority class. Oversampling usually involve synthetic data generation techniques including traditional methods like SMOTE, ADASYN sampling and other interpolations while the AI based techniques i.e. Generated Adversarial Networks (GAN) employ learned patterns for artificial instances generation for minority classes to diversify the dataset. Additionally in cost-sensitive learning, classes are assigned different misclassification costs, ultimately prioritizing minority class accuracy during the training. Techniques like these contribute collectively to overcome challenges posed by imbalance classes in bearing fault classification [7], [12].

With the fast-paced emergence of effective deep learning algorithms being refined each data, developments in convolutional and Time Series algorithms have pushed the research in condition monitoring to new heights [20]. Techniques employing novel combinations [21], [22] have surfaced, enabling effective diagnoses and prognosis of bearing failures. Similarly, the constraints of large data-sets have been entertained by novel architectures and transfer learning models being used interchangeably with certain architectural modifications for required results, as suggested in [23]. Some non-conventional CNN techniques such as Siamese networks [24] and the graph neural networks have effectively been integrated into the classification pipeline for effective results. Overall these techniques have proven to be successful in effectively distinguishing various faults types in bearings.

Being a newly tapped domain there are many challenging gaps and unaddressed facets that form a plethora of possibilities for improvement in bearing fault classification. One such aspect is the evaluation criteria. Accuracy has been taken as a primary measure or the sole assessing criterion in most of the studies, which deprives the model of thorough evaluation [24], [25], [26]. Owing to the imbalanced lineament of the data [27], the biases and class-wise agreements for proposed techniques should be properly weighed across various performance matrices for reliance [28], [29]. Additionally, the computational costs of classification and fault diagnosis should be considered for deployable scenarios. This study contributes by leveraging three distinct convolutional neural networks (CNN) to create a weighted voting ensemble for a robust classification of signals at 0 and 1 HP Load. Three custom built CNN architectures were proposed for the ensemble model-ling to evaluate performance, assess computational costs, robustness, and effectively com-pare against existing techniques. Moreover, the class imbalance of the dataset was dealt with using a novel density and distance weighted hybrid SMOTE methodology allowing for a more refined oversampling of minority class instances and an improved classification. Furthermore, the time-frequency based representation of the fault signals was implemented using wavelet transform for efficaciously utilizing the abilities of an adaptive spatial feature oriented classification approach. This study contributes as follows:

1) A weighted voting ensemble of three streamlined and resource efficient convolutional neural networks is proposed for robust classification of bearing faults.
2) A density-distance weighted hybrid SMOTE methodology is implemented by refining the SMOTE generated samples through a ranking of density values for a uniform weighted distance and density aware oversampling.

TABLE 1. Comparison of existing studies on ball bearing fault diagnosis.

Reference	Year	Dataset	Methodology	Classes	Results	Limitations
[23]	2021	SKF	Novel method using WGAN and CNN for imbalanced datasets.	9	99.6%	Cost Inefficient Augmentation
[26]	2023	CWRU	Novel Architecture SE-ResNet152 derived from ResNet	4	96.42%	Classification performance, Under-Sampled Dataset
[30]	2017	CWRU	Proposed a custom CNN architecture.	4	99.5%	Imbalanced Classes
[25]	2023	CWRU	Proposed a graph neural network based approach	4	99.1%	Complex pre-processing, Under-represented classes
[24]	2022	Private	Proposed a Siamese Based Network to evaluate similarity and classify faults.	5	98.1%	No Augmentation
[22]	2022	CWRU	Residual Network based on Markov transition field.	10	98.52%	Imbalanced Classes
[21]	2019	CWRU	CNNEPDNN model using raw time domain signals.	4	98.10%	Computationally expensive

3) The time-frequency based scalogram representation of bearing fault signals is used for a more transient spatial features extraction.

The remaining part of the paper is organized as follows: Section II, the Literature Review is given. In Section III, the methodologies used are described. The Results and Discussions are given in Section IV respectively. While the Section V concludes the paper.

II. LITERATURE REVIEW

The literature study pertaining to bearings reveal a scintillating insight into the re-search and development focused on fault detection and diagnosis. Various studies have been sieved through to reveal the state-of-the-art methodologies being employed in regard to the effective diagnosis of bearings for possible abnormalities while highlighting their significance.

One such study by Tang et al. [23] provided a novel take on the process of bearing fault diagnosis by employing Wasserstein Generative Networks (WGAN) for combating the imbalanced classes. The vibrational signals were converted to gray-scale images and then fed to a generative network to create synthetic samples, representatives of the minority class. Moreover, a novel convolutional neural network (CNN) named SECNN was then used for fault identification with an achieved accuracy of 98.2% on SKF dataset for 9 classes. Another study by Wu et al. [26] proposed a novel architecture name SE-ResNet by modifying an existing residual connections-based network called ResNet. The author explicitly opted for automated feature extraction by converting raw signals to image-based representations using continuous wavelet transform (CWT). In this study conducted on the CWRU dataset for 4 classes, the signals were converted to Spectrogram prior to being subjected to SE-ResNet for classification. The model achieved an accuracy of 96.42%.

CWRU has been used recurrently for its diversity and reliability in the domain of bearing faults classification, along with its open-sourcing enabling researchers to reproduce the results of the experimentations and build upon the experiment. Hence another study on this dataset by Verstraete et al. [30] demonstrated the results of time-frequency analysis of the fault waveform by employing STFT, Wavelet Transform and Hilbert-Huang transforms respectively. These transforms curbed the need for manual- labor-intensive feature extraction and automated the process. These respective transformations were then fed to a convolutional neural network for fault identifications. The highest accuracy achieved for the following dataset in terms of scalogram and Spectrogram inputs approached 99.5% for 4 class classifications.

Some studies drifted away from the conventional classification of CNN networks to shift the perspective onto other networks in order to enhance performance. One such study by Xiao et al. perfectly emulated this idea by harnessing the power of graph neural net-works for effective classification [25]. The graphical approach took into account the correlation between weak and strong fault signals. A graph of similarity between samples was created first and then forward to a Graph Neural Network for effective detection. The validation results achieved on the similar CWRU dataset for 4 classes were 99.1%. Another study by Zhao et al. [24] proposed a Siamese Network-based approach to not only classify the faults but to get insights into the similarity relation between samples. The authors utilized the network to train the model on a small dataset. The authors modified the standard Siamese network in a way that a classification head was added while the Euclidean distance measurement was replaced with a network measurement. Different tests at varying input samples were conducted to evaluate the model's performance. The maximum achieved validation accuracy on private datasets using the custom-defined ISNN method was around 98.1% for 5 classes respectively.

Some more studies effectively exploiting the potential of CNN were seen during the literature analysis. One such study by Yan et al. proposed MTF-ResNet an architecture inspired by ResNet and Markov Transition Field for effective identification of fault patterns in the data [22]. The authors used augmentation techniques along with a time-frequency conversion of vibrational data for the classification. The problems pertaining to vanishing gradients and raw signal-to-image conversion have been addressed using the Markov Transition Field (MTF) based conversion of signals into 2-dimensional representations. The achieved validation accuracy for 10 classes was around 98.52% on the CWRU dataset.

Another feature fusion-based ensemble Convolutional Neural Network was pro-posed by Li et al. [21] the network named CNNEPDNN was used to classify varying vibration sensor signals in an ensemble to get highly accurate outputs. Moreover, the proposed method by the respective authors effectively entertained the loss of significant signal in-formation during the pooling operations. The validation accuracy achieved for the 4 classes on the CWRU dataset was 98.10%.

Being a aspect of machine learning models, data imbalance poses a significant problem. Several studies have proposed innovative techniques and algorithms to address the challenges associated with accurate fault diagnosis under imbalance conditions. These techniques have utilized both the traditional and deep learning based approaches for effective up-sampling. In regards to the tradition up-sampling, Hang et al. [13] proposed an improvement to Synthetic Minority Over-sampling Technique (SMOTE), by incorporating Principal Component Analysis (PCA) in order to enhance the performance in high-dimensional imbalanced fault diagnosis data for rolling bearings. The respective work the authors demonstrated significantly improved classification performance, which is an imperative part of accurate detection of faults. In a similar vein, Wei et al. proposed TMD-SMOTE algorithm, further improving SMOTE's up-sampling capabilities. The algorithmic approach put forth by the authors not only considers the problem of distribution marginalization but also addresses complexity of algorithm itself [15].

Similarly, Li et al. proposed the Minority-class Sensitive Fault Diagnosis approach (MSFD), aiming to reduce imbalance and enhance diagnostic model's sensitivity to minority samples. This approach addresses a crucial aspect of fault diagnosis in imbalanced datasets [16]. Duan et al. introduced the MeanRadius-SMOTE algorithm, an enhancement of the traditional SMOTE oversampling technique. This algorithm effectively avoids the generation of useless and noisy samples, leading to improved prediction accuracy in unbalanced datasets. Addressing the challenge of generating meaningful samples is vital in fault diagnosis [17].

The above literature study provides insight into the ongoing and previous research with the domain of bearing fault analysis and identification under balanced and imbalanced states. The research elucidates certain aspects of the research and the techniques employed in effectively harnessing the potential of current deep learning and traditional tools. Moreover, being a recently tapped domain, there is still a need for effective research in fault diagnostics and explainability. The related literature on the detection and classification of ball bearing fault has been summarized in Table.1.

III. METHODOLOGY

A. DATASET DESCRIPTION

An open-source and widely used benchmark dataset for the bearing fault diagnosis CWRU (Case Western Reserve University) [27] is used herein for the purpose of evaluating the proposed method against bearing fault diagnosis.

The apparatuses used in extracting the data consist of a 2 horsepower motor, a torque transducer and a dynamometer. As shown in Figure 1, varying range of fault conditions were deliberately imprinted on individual bearings, including inner race, outer race, and ball faults, as well as their combination. Moreover, the dataset was extracted for two frequencies i.e., 12Khz and 48Khz at 0,1,2,3 Horse Power(HP). [31].

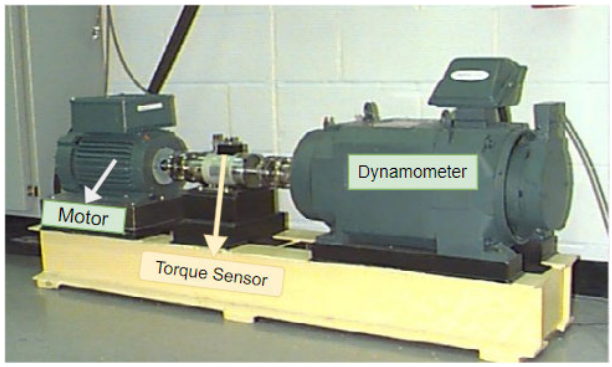

FIGURE 1. Experimental setup to acquire faulty bearing signals.

In this study, the fault dataset for drive-end 48K samples at no-load condition (NLC) and single load condition (SLC) was used due to the presence of class-imbalance serving as an optimal application for our proposed approach. Within the 48Khz, the fault diameter of 0.007, 0.014 and 0.021 inches were also taken at the mentioned load conditions for classification. Additionally, the normal baseline data was merged with the prior mentioned data to be classified alongside. The empirical procedures to evaluate the proposed model's efficiency were based on 14 classes respectively.

B. DATA PREPROCESSING

1) PRE-PROCESSING OF BEARING FAULT VIBRATIONS

The pre-processing stage plays a significant part in preparing the acquired data for further analysis and training. In this study, we experimented on the vibration signals acquired at 48,000 Hz (48KHz), under 0 and 1 HP load conditions. The data points at 48KHz were subjected to specific pre-processing steps to augment the effectiveness of the subsequent analysis and cope with any unbalanced classes that may skew the model's classification behavior. In this

regard we implemented a novel synthetic oversampling method by using a weighted-distance and density hybrid approach involving the SMOTE generated samples and refining them.

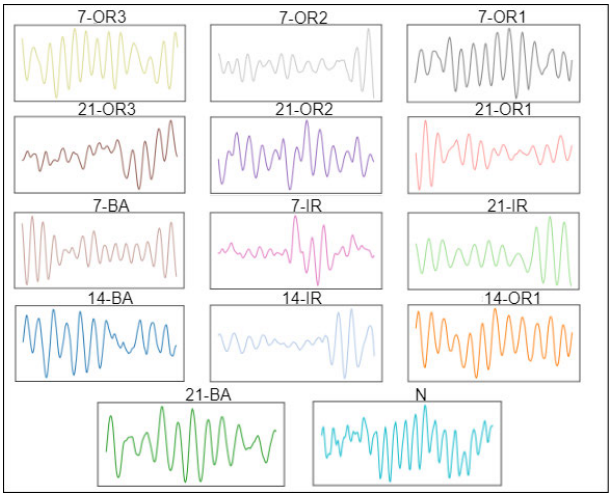

FIGURE 2. Visualization of fault waveforms for 0-HP.

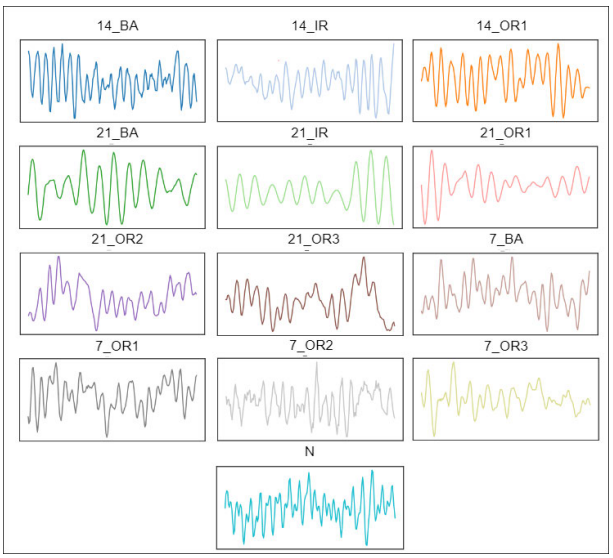

FIGURE 3. Visualization of fault waveforms for 1-HP.

For NLC and SLC at 1797 rpm, the dataset for 48 KHz was extracted for 3 fault diameters, having a total of 14 classes. The class distribution included normal baseline class values, 3 classes each for Inner Race Faults(IRF) and Ball Faults(BF), 7 classes for Outer Race Faults(ORF) at varying orientations and fault diameters respectively, and given in Figure 4 and Figure 5. These classes were segregated and processed using Matlab. The visual illustration of the raw waveform at aforementioned specifications is given in Figure 2 and Figure 3 for O and 1 HP based load conditions respectively.

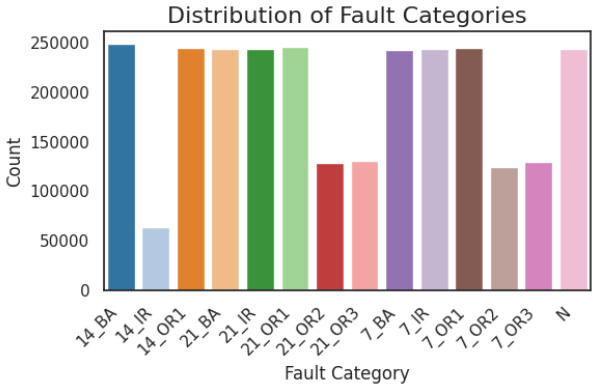

FIGURE 4. Class distribution for 0-HP 48 KHz samples.

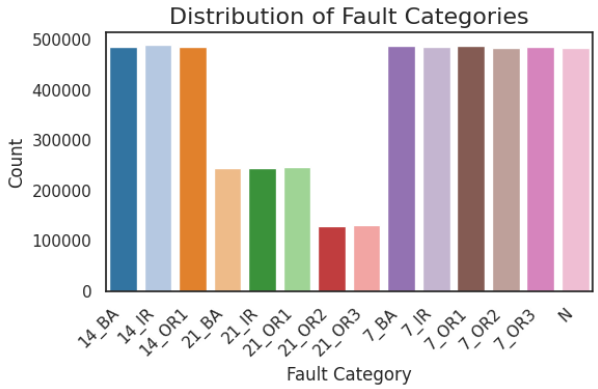

FIGURE 5. Class distribution for 1-HP 48 KHz samples.

The Figure 4 and Figure 5 illustrates the distribution for normal and faulty classes at 48KHz under NLC and single load. The y-axis represents the sample size whilst the x-axis represents the class name. Each class name mapped on the x-axis for each colored bar represents the fault type and the diameter of the respective fault. Except for the ORF, where the orientation is also mentioned along with type and diameter. The acronym representation for each class in x-axis is given such that BA represents the BF, OR represents the ORF, IR represents the IRF and the N represents the normal baseline. The number before the acronyms separated by an underscore represents the diameter of the fault in mills (where 1 inch = 1000mills). The number after the ORF acronym represents three orientations depicted by 1,2 and 3 at 0', 90' and 180' respectively. In case of NLC, the number of samples for the majority classes lies between 243K to 249K respectively, while the minority classes have a varying sample size ranging from the lowest tier class 14_IR having a sample size of 63788 and the rest ranging from 124K to 130K respectively. Similarly for SLC, the majority samples lie between 482K to 489K samples, while the minority samples lie between 128K to 248K samples.

2) OVER-SAMPLING

The class distribution for the faulty samples is uneven, hence rendering the classification susceptible to biases. In order

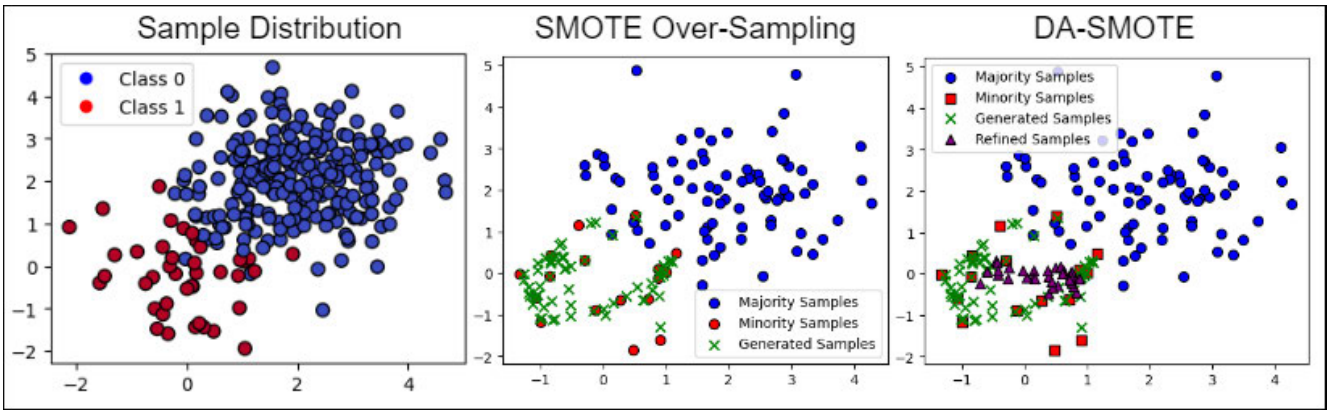

FIGURE 6. Sample distribution for SMOTE and DA-SMOTE.

to prevent this potential shortcoming within the dataset we proposed a novel over-sampling algorithm derived from SMOTE. Synthetic over-sampling via SMOTE is open to a number of possible complications that may influence the classification behavior of the models. The generated samples from SMOTE more often then not display signs of potential over-fitting, information loss, noise-sensitivity and decision boundary influence. These significant drawbacks can affect the models accuracy and other distinguishing traits. Herein we propose a Density-Aware oversampling methodology (DA-SMOTE) to effectively refine the generated samples acquired through SMOTE algorithm and applying a density based sample selection for a more appropriate adjustment within the bounds of minority class distribution as shown in Figure 4.

For the over-sampled dataset D_s Let X_{min} be the matrix of minority class features while X_{gen} be the SMOTE generated feature matrix of minority class and k be the number of nearest neighbors. The distance d_i of each synthetic sample with its k-neighbors under the bounds of the minority region is given by:

$$d_i = distance(X_{(min(i))}, X_{(min(neig[i]))}) \tag{1}$$

where,

neigh[i] = Indices of neighbors surrounding the minority sample

Next we calculate the average distance d_{avg} for each synthetically generated sample as:

$$d_{(avg[i])} = \frac{1}{k}\sum_{m=1}^{k} d_{i[m]} \tag{2}$$

While the distribution density D_ρ of minority class around a synthetic samples is given such that:

$$D_{\rho[i]} = \frac{1}{d_{avg[i]}} \tag{3}$$

After calculating the density of each synthetic sample generated via SMOTE, we will rank the samples based on their density values in a descending order. With the higher density samples at the top and the lower density values at the bottom. We can define the density based ranking R as

$$R = sort[D_\rho, Descending] \tag{4}$$

Out of the sorted densities, top N densities are chosen to be further refining. For each one of the selected synthetic samples, the indices of the k-nearest neighbors are computed with lying within X_{min}. Furthermore, the average features values are computed for the nearest neighbors such that,

$$N_{avg[i]} = \frac{1}{k}\sum_{j=1}^{k} X_{min[j]} \tag{5}$$

where,

$N_{avg[i]}$ = Average feature values of the nearest neighbors

Finally the average values of synthetic samples are adjusted towards the Average value $N_{avg[i]}$ of the nearest neighbors using:

$$X_{ref[i]} = X_{gen[i]} + \alpha(N_{avg[i]} - X_{gen[i]}) \tag{6}$$

here,

$X_{ref[i]}$ = Refined Sample

α = Adjustment parameter

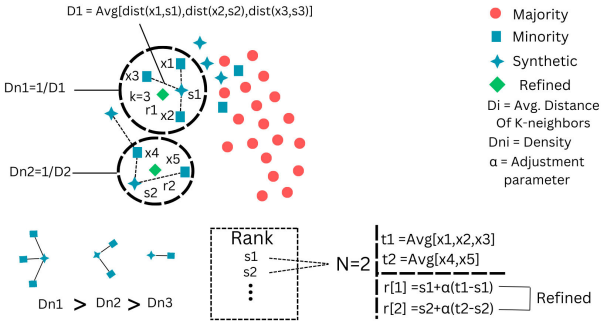

FIGURE 7. Density-Aware (DA) SMOTE oversampling.

C. WINDOW SEGMENTATION

For the extraction of meaningful information from the raw over-sampled vibration signals, they were divided into windows or smaller segments. The segmentation helps to extract localized features along with any recurrent patterns associated with fault types. Continuous vibrational waveforms were converted in fixed-length segregation also known as time-windows or frames. The size of stride determines the extent of overlapping segments and often used to capture temporal dynamics. The experimentations herein were performed on a window size of 1024 with a strides of 800. The mathematical expression for window segmentation can be given as

$$W_i = S[i * W_s : i * W_s + W_l] \tag{7}$$

where,

S = Time series signal
W_i = Segmented Window
i = Index
W_s = Stride size
W_l = Window length

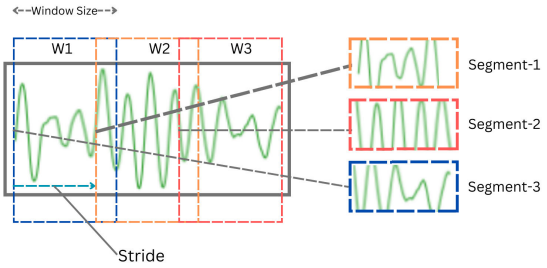

FIGURE 8. Raw signal segmentation.

1) TIME-FREQUENCY REPRESENTATION

Once segmented over-sampled signals are generated, they are transformed into scalogram for the representation of signals in the time-frequency domain. The respective transformation holds viable potential to enhance the representation of both frequency and time components of the signals, providing a visual representation of the signal's energy distribution across both time and frequency dimensions This allows for the identification and extraction of key features and patterns that are crucial and relevant for fault classification tasks.

Let $S'[t]$ represent the segmented signal at time index t from the dataset D'. The scalogram W[a,b] of S' at a specific scale a and translation b is calculated as:

$$W[a,b] = \int_{-\infty}^{\infty} S'[t] * \phi^*(\frac{t-b}{a})dt \tag{8}$$

where,

a = scale parameter to controls width of the wavelet.
b = translation parameter to shift the wavelet along the time axis.
t = time index.
ϕ^* = complex conjugate of analyzing wave-function.

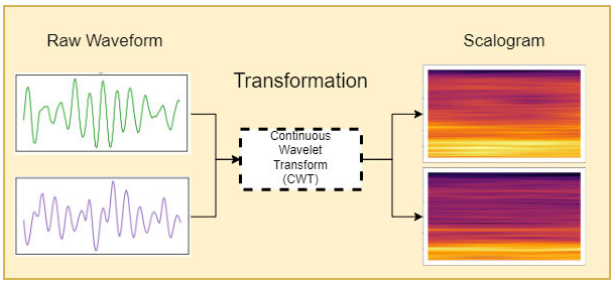

FIGURE 9. Raw over-sampled signals to Scalogram.

D. CONVOLUTIONAL NEURAL NETWORKS(CNN)

In this context, several approaches toward CNN-based classification have been explored. Deep learning-based automated feature extraction methods that make use of hierarchical structure to capture significant high and low-level details at varying levels. CNN have repeatedly demonstrated itself effective in classification of bearing faults, owing to its automated extraction of intricate details from both hierarchical and local patterns.

1) CONVOLUTIONAL LAYERS

The Convolutional Layers make use of a set of filters or kernels for convolving across an input image in order to extract features. For an image input $I \in D$, where D is the dataset of converted 2d images. The filter F is convolved with Ito produce an output feature map.The mathematical representation of convolutional operation is given as

$$m(i,j) = f[\sum(F * I)] \tag{9}$$

where,

m(i,j) = feature map obtained after convolving.
f = activation function

2) ACTIVATION LAYER

Non-linearity is added to the convolved output element-wise, by means of an activation function. More often ReLU is used as an activation function and represented as follows:

$$ReLU[t] = max(0,t) \tag{10}$$

where,

t = Input to activation function

3) POOLING LAYER

The feature maps obtained through convolution undergo pooling operations to reduce their spatial dimensions. This ensures the extraction of dominant features and aids in achieving translational invariance. The most commonly used pooling operation is max pooling, which selects the maximum value within each pooling region. Mathematically, the output of the pooling layer can be represented as

$$P_i = g(m_i) \tag{11}$$

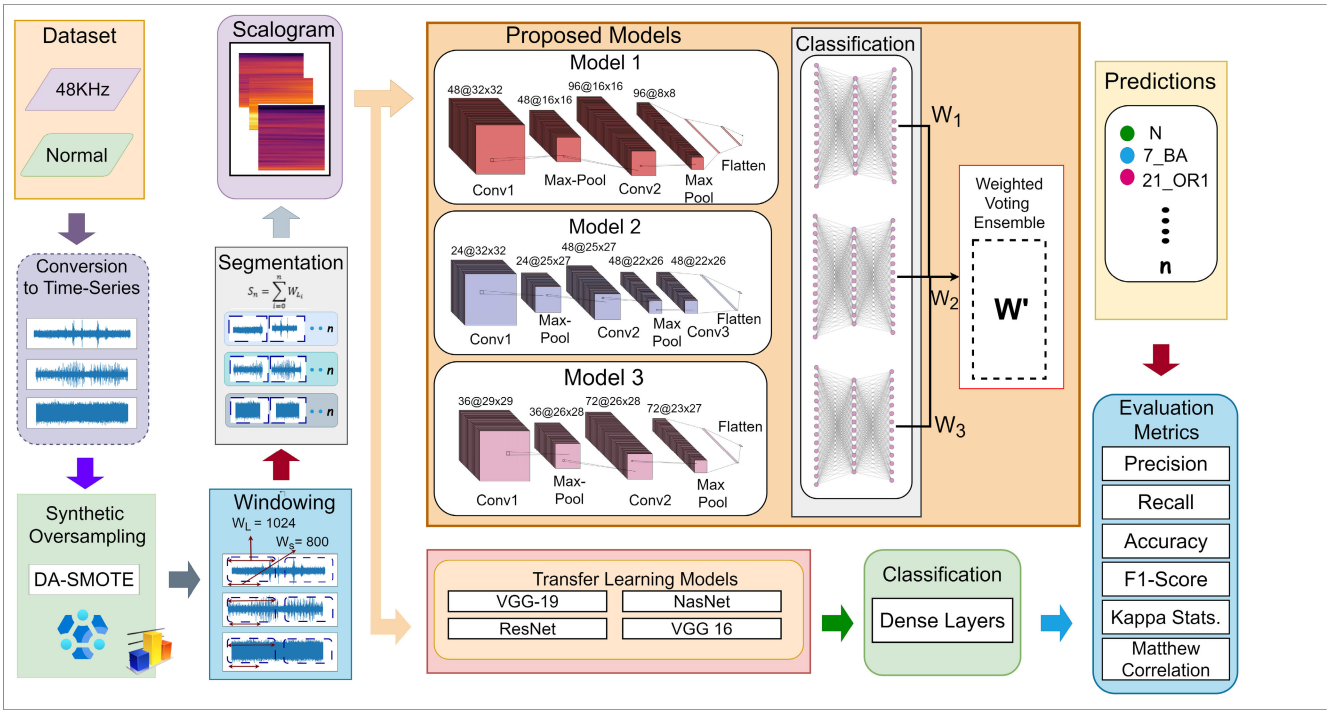

FIGURE 10. Block diagram of proposed - methodologies.

where,

P_i = pooled feature map
m_i = input feature map
g = pooling function

4) FULLY CONNECTED LAYER

The fully connected layer is responsible for classifying the features extracted by the convolutional and pooling layers. It consists of neurons that are connected to all the neurons in the previous layer. Mathematically, the output of the fully connected layer O can be represented as:

$$O = \sigma(\sum(W * F) + b) \tag{12}$$

where,

O = Output vector
F = Input feature vector
W = Weight matrix
b = Bias term
σ = Activation function (e.g., softmax for classification) that introduces non-linearity.

E. TRANSFER LEARNING AND PRE-TRAINED NETWORKS

Transfer learning is an efficacious remedy to a small dataset or under -performing models. Transfer learning includes the training of a model on a large dataset usually on millions of samples to learn generic feature representations of objects. These feature representations are then used by freezing the initial layers and using the knowledge of the previous layers to navigate the training of custom data. Some of the most effective transfer learning models that are also employed in this study include VGG-19 [32], NasNet [33], ResNet50 [34] and VGG16 [35].

F. PROPOSED METHODOLOGY

In this study, a robust ensemble approach was employed, utilizing a weighted voting ensemble strategy that combines the outputs of three individual CNN models. Each of these CNN models was meticulously crafted with varying parameters to explore a diverse range of computational complexities and performance characteristics. Through comprehensive evaluation, the ensemble method was bench-marked against cutting-edge transfer learning models as well as the proposed methodologies. The interplay between these models and their performance nuances are elucidated within the schematic diagram presented in Figure 9, showcasing the effectiveness of the ensemble's weighted voting mechanism in harnessing the collective predictive capabilities of the three CNN models.

1) WEIGHTED VOTING ENSEMBLE(WVE)

In the following paper, we implemented an effective strategy to enhance the precision and reliability of classifying scalogram representations of bearing fault signals through a Weighted Voting Ensemble of three distinct Convolutional Neural Network (CNN) models namely WVCNN. The respective ensemble technique is meticulously tailored to address the complexities of bearing fault classification by harnessing the collective predictive abilities of individual

TABLE 2. Architectural description for proposed CNN models.

Model	Layers	Kernel	Stride	Size	Activation	Maps	Parameters
SCNN1	Conv2D	4,2	1,1	32,32,48	ReLU	48	1200
	Max-Pool	2,2	2,2	16,16,48	-	48	-
	Conv2D	4,2	1,1	16,16,96	ReLU	96	36960
	Max-Pool	2,2	2,2	8,8,96	-	96	-
	Flatten	-	-	1,6144	-	-	-
	Dense1	-	-	144	ReLU	-	884880
	Dense2	-	-	16/14	Softmax	-	925,360
	Total Layers	Conv.	Dense	Dropout	Flatten		Trainable / Non-Trainable
	5	2	2	-	1		925,360 / -
SCNN2	Conv2D	5,5	1,1	28,28,24	ReLU	24	1824
	Max-Pool	4,2	1,1	25,27,24	-	24	-
	Conv2D	4,4	1,1	25,27,48	ReLU	48	18480
	Max-Pool	4,2	1,1	22,26,48	-	48	-
	Conv2D	3,3	1,1	22,26,48	ReLU	48	
	Flatten	-	-	1,27456	-	-	-
	Dropout				Rate = 0.1		
	Dense	-	-	16/14	Softmax	-	480,400
	Total Layers	Conv.	Dense	Dropout	Flatten		Trainable / Non-Trainable
	5	2	1	1	1		480,400 / -
SCNN3	Conv2D	4,4	1,1	29,29,36	ReLU	36	1764
	Max-Pool	4,2	1,1	26,28,36	-	36	-
	Conv2D	3,3	1,1	26,28,72	ReLU	72	23400
	Max-Pool	4,2	1,1	23,27,72	-	72	-
	Flatten	-	-	1,44712	-	-	-
	Dropout				Rate = 0.1		
	Dense	-	-	16/14	Softmax	-	740572
	Total Layers	Conv.	Dense	Dropout	Flatten		Trainable / Non-Trainable
	5	2	1	1	1		740572 / -

CNN models. Each of the corresponding CNN network proposed for this ensemble is configured with distinct architectural variations and hyper-parameters to capture diverse characteristics of the scalogram data. The weighted voting ensemble operation entails averaging the predictions of the individual models while manually assigning specific weights based on their respective validation performance. The weight coefficients are computed to maximize the influence of models with superior validation accuracy on the final prediction. The models designed for this ensemble are explicitly defined in the forthcoming sections. Whereas the ensemble method is mathematically formulated as:

$$P_{ensemble} = \frac{\sum_{i=1}^{3} w_i P_i}{\sum_{i=1}^{3} w} \tag{13}$$

where,

w_i = Weights assigned to CNN model i

P_i = Weighted predictions

2) PROPOSED ENSEMBLE MODEL 1 (SCNN1)

The first model SCNN1 in Figure 10 consists of five layers. First of which is a convolutional layer for initial feature extraction. The number of feature maps obtained from the first layer(L1) is 48 with kernel size K_{size} of (4,2) and default stride S of 1. The activation function $\rightarrow A_{fun}$ used in successive layers is ReLU. The output feature maps from the L1 are max pooled (P_{max}) to reduce the dimension via a K_{size} (2,2). In the second layer(L2), 96 feature maps $\rightarrow f_{map}$ were extracted using a K_{size} of (4,2) followed by a P_{max} operation on a receptive field of (2,2). The output from L2 is flattened to form a 1-d feature vector, which is then fed to a dense layer having 144 neurons and for the final classification a dense layer of 14 neurons is used, depicting the number of classes from 48KHz input data. The probability distribution for each class is obtained via softmax classifier. The optimizer used during the back propagation is Adam, set at a learning rate of 0.0001. Categorical Cross Entropy (C.C.E) was used as a loss function to estimate the discrepancy between the ground truth and predicted labels. The training was performed on a batch size of 64 and iterated for 50 epochs.

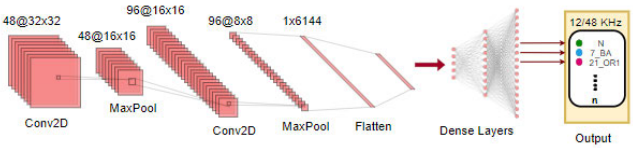

FIGURE 11. Architecture for SCNN1.

3) PROPOSED ENSEMBLE MODEL 2 (SCNN2)

The second model namely SCNN2 shown in Figure 11 consists of 5 layers. L1 is a convolutional layer with f_{map} of 24 with the K_{size} kept at (5,5). The convolution operation is swiftly followed by P_{max} operation having a K_{size} (4,2) with S at (1,1). For L2, 48 f_{map} were obtained by convolving a Kernel of size (4,4), shortly followed by a P_{max} operation with K_{size} of (4,2). The third layer also involved a convolution operation at a K_{size} of (3,3) acquiring a f_{map} of 48. The outputs from the respective f_{map} were flattened using a flatten layer to form a latent vector and strictly followed by a dropout layer set

a 0.1 to mitigate the risk of overfitting. The output from the dropout layer is fed to a dense network of 14 neurons for fault classification. The optimizer used in this setting was Adam with a learning rate of 0.0001. The loss function used was C.C.E and the number of epochs was set to 50 with an input batch size of 64.

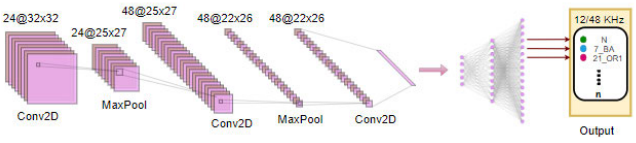

FIGURE 12. Architecture for SCNN2.

4) PROPOSED ENSEMBLE MODEL 3 (SCNN3)

L1 of the third model SCNN3 shown in Figure 12 is set at a K_{size} of (4,4) to acquire the f_{map} of 36. P_{max} is applied on f_{map} obtained through L1 with the K_{size} of (4,2) in order to reduce the dimensionality. L2 follows along with a f_{map} of 72 having a receptive field set at (3,3), followed by a P_{max} operation with K_{size} of (4,2). The outputs are then fed to a flatten layer to obtain a 1-d vector of features. An intermediate dropout layer set at a 0.1 is placed before the last dense layer of 14 neurons for reducing the risks of over-fitting. The final dense layer consists of 14 neurons for obtaining the class probability score through the softmax classifier. Adam is used an optimizer with a learning rate of 0.0001 with a loss function C.C.E. Accuracy, MSE and MAE are specified as evaluation metrics. The batch size and the number of epochs are set at 64 and 50 respectively.

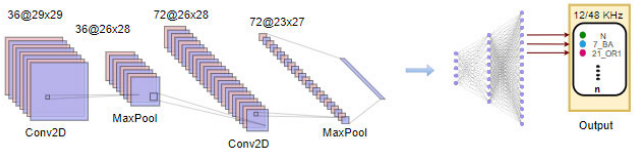

FIGURE 13. Architecture for SCNN3.

G. EVALUATION METRICS

1) PRECISION

Precision represents accurately predicted instances, specifically the True Positives, in regards to all predictions labeled as Positive, inclusive of both True Positives and False Positives. Mathematically, it can be represented as:

$$Precision_{(PR)} = \frac{T_P}{T_P + F_P} 100\% \tag{14}$$

2) ACCURACY

Accuracy calculates the ratio of correctly classified instances to the total number of instances i.e., all true and false instances. It can be expressed mathematically as:

$$Accuracy_{(AC)} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{15}$$

3) RECALL

Recall assesses the model ability in reducing the False Negatives. It is quantified by the ratio of True Positives to the sum of False Positives and True Positives. Mathematically, Recall is expressed as:

$$Recall_{(RC)} = \frac{T_P}{T_P + F_P} 100\% \Rightarrow 1 - Rate_{False_{Neg}} \tag{16}$$

4) F1-SCORE

The F1-Score provides a numerical representation of equilibrium between precision and recall. It can essentially be defined as the harmonic mean of precision and recall. F1-Score is particularly valuable when addressing the class-imbalanced data. Its calculation is expressed as:

$$F1Score_{(FS)} = \frac{2 \times T_P}{2 \times T_P + F_P + F_N} \tag{17}$$

5) KAPPA STATISTICS

The Kappa Statistics, also known as Cohen's Kappa Coefficient, is a statistical metric that represents the agreement between actual and predicted class labels. It evaluates the possibility of chance or random agreement. The mathematical representation of the Kappa Statistics is given by:

$$Kappa_{stats(KS)} = \frac{p_o - p_e}{1 - p_e} \tag{18}$$

IV. RESULTS AND DISCUSSIONS

The respective section provides description in regards to the evaluation of the proposed ensemble technique along with a comparison with some other state-of-the-art methods. The experimental procedures undertaken in this study involve three custom Convolutional Networks carefully designed and integrated into a weighted voting ensemble for a robust and reliable classification on faults samples. The experimentation was performed on an imbalanced and an over-sampled dataset in order to evaluate the impact of class balancing on the model's performance. Furthermore, the performance of our proposed methods was evaluated with pre-trained models as comprehensively defined in the coming discussion.

A. CLASSIFICATION WITHOUT OVERSAMPLING

The acquired validation accuracy(VA) from the training of scalogram representations given in the Figure 14 and Figure 15 using the proposed models were 99.08%, 99.16% and 99.22% for the individual models and 99.28% for the weighted ensemble in case of NLC. While for the SLC, the VA attained by training the individual models on imbalanced dataset is 98.43%, 98.66 % and 99.10 % for the three CNN models and 99.13 % for the ensemble. These results were encouraging and were further assessed for reliability by weighing them across multiple matrices including F1-score, recall, MCC and KS. These matrices provided insight into the individual models performance for each class and were indicative of any possible biases towards a class. The results were further compared with state-of-the-art transfer learning

models including ResNet50, VGG-19, NasNet and VGG-16 to analyze the ability of these models in addressing the classification challenge. The results however alluded to an under-performance of these models except for the VGG-16 and VGG-19 which managed to perform well by obtaining a VA of 92.63% and 93.47% for NLC and 95.02 % and 96.51 % for the SLC. The rest of the models achieved VA accuracy of 82.32% and 73.73% for NLC and 84.93 % and 72.01 % for SLC respectively.

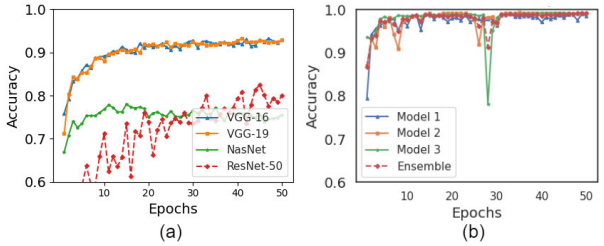

FIGURE 14. (a) Val accuracy of pre-trained models on 0-HP imbalanced dataset (b) Val accuracy of proposed methods On 0-HP imbalanced dataset.

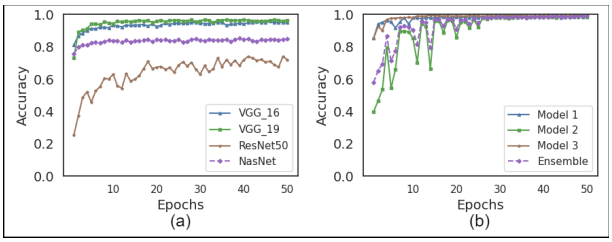

FIGURE 15. (a) Val accuracy of pre-trained models on 1-HP imbalanced dataset (b) Val accuracy of proposed methods On 1-HP imbalanced dataset.

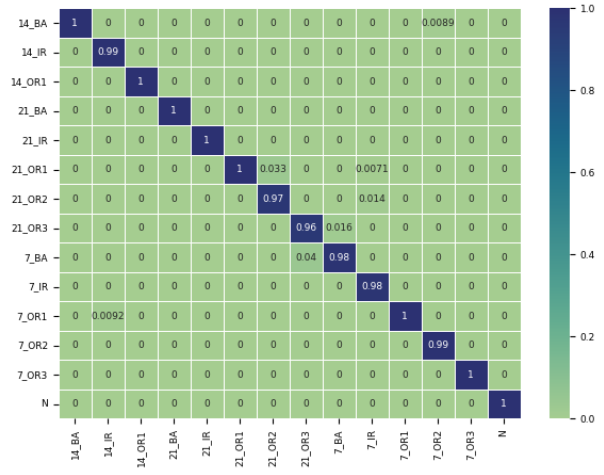

FIGURE 16. Confusion matrix of voting ensemble model (WVCNN) for imbalanced dataset.

The above evaluation results from each of the proposed methodologies and the transfer learning models provide the thorough insight into the results obtained from the imbalanced dataset. It can be seen that the proposed methodologies achieved high training and validation accuracy while from the transfer learning models VGG-19 and VGG-16 were on par with the results of our proposed methods. The rest of the pre-trained models under-performed in this scenario. The confusion matrix of WVCNN revealed the class wise classification score for 14 classes of 48KHz data under-imbalance data for 0-HP. The left diagonals depict the true predictions while on the horizontal and vertical axis lay the false predictions. The count for each class is normalized for proportional representation of class wise predictions.

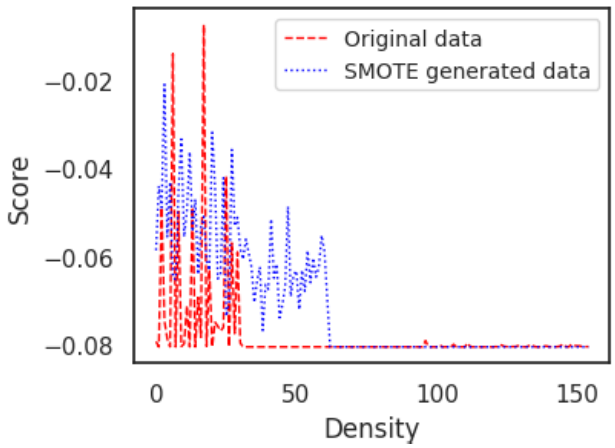

FIGURE 17. Kernel density estimation of both synthetic and original samples.

B. CLASSIFICATION WITH PROPOSED OVERSAMPLING TECHNIQUE (DA)

1) EVALUATION OF PROPOSED DENSITY-AWARE (DA) SMOTE OVER-SAMPLING

The proposed Density Aware Weighted Hybrid SMOTE approach generates synthetic samples by addressing both the distance and density in order to maintain a more balanced sampling approach. By emphasizing regions with high density and appropriate distances the generated samples are refined. In order to evaluate the sample generation for the under-represented classes of bearing faults signals, quantitative and qualitative assessment methods are considered. The first of which is the Kernel Density Estimation of the DA-SMOTE refined and the original samples as shown in Figure 17. The figure above illustrates a kernel density estimation of both the original and synthetic datasets. In the figure above, the synthetic data is depicted as a smooth blue line, while the original data is represented by a red line characterized by substantial spikes. The stark contrast between the blue line of the synthetic data and the red line of the original data can be attributed to the way the data was generated and its implications:

1) The synthetic data, generated using DA-SMOTE, emphasizes creating new data points in a manner that fosters even distribution across the feature space. This results in a smooth, continuous distribution in

the kernel density estimation plot. The synthetic data, by design, aims to bridge gaps between minority class instances and, in doing so, mitigates the presence of concentrated spikes.

2) The original data exhibits substantial spikes in its kernel density estimation. This un-even distribution could stem from inherent characteristics of the dataset, such as clustering and class imbalance. The presence of pronounced spikes suggests that there are regions of high density where data points are clustered together, contributing to the unevenness in the distribution.

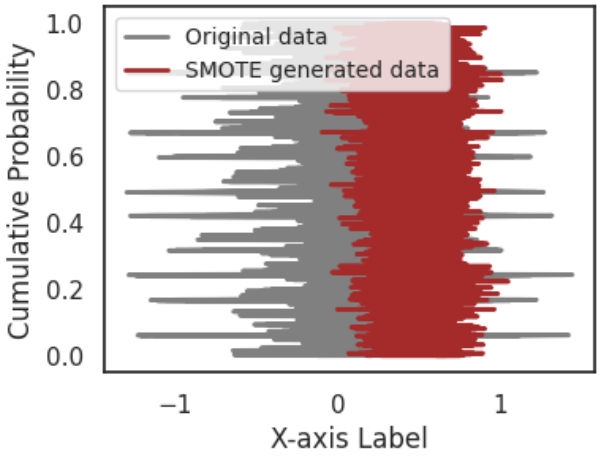

FIGURE 18. Cumulative distribution function (CDF) of SMOTE.

The even distribution of synthetic data (blue line) supports the learning process of deep learning models. By populating previously underrepresented regions, it allows models to better capture the underlying data distribution and patterns. This, in turn, enhances their ability to generalize and make accurate predictions on unseen data. The Cumulative Distribution Function (CDF) graph in Figure 18, visually illustrates the cumulative probability of both the original dataset and the dataset generated using the Synthetic Minority Over-sampling Technique (SMOTE).

The CDF of a distribution serves as a mapping function that associates each value within the distribution with its cumulative probability. In simpler terms, the CDF informs us about the likelihood that a randomly chosen value from the distribution will be equal to or less than a specific value.

In the graphical representation, the x-axis signifies the values present in the distribution, while the y-axis corresponds to the cumulative probability. The gray solid line illustrates the CDF of the original data, whereas the maroon line portrays the CDF of the DA-SMOTE generated data.

The rightward shift of the generated data's CDF can be seen in comparison to the CDF of the original data. This shift indicates that the generated dataset has a higher likelihood of containing larger values in comparison to the original data. In other words, the generated data encompasses a greater proportion of values that are larger or more extreme than what's seen in the original dataset.

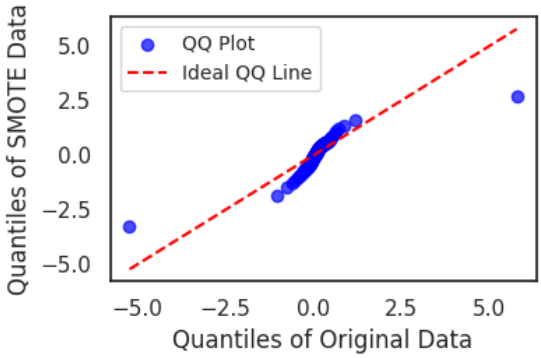

FIGURE 19. QQ-Plot Of DA-SMOTE and original data.

The reason for this shift can be attributed to the way synthetic samples are generated. DA-SMOTE introduces synthetic samples by interpolating between existing minority class samples on the basis of the densities and distances of the neighbors. As a result, the new synthetic samples can exhibit more diverse characteristics, leading to an expansion of the dataset's range. This shift in the CDF signifies that the over-sampling technique, by design, influences the distribution to include a broader spectrum of values, potentially contributing to improved model generalization and predictive accuracy. The QQ-Plot, a graphical tool for distribution comparison, displays the quantiles of two datasets, the synthetic and the original, against each other. Ideally, the points should follow a straight line, indicating similar distributions. A slightly curved plot, as illustrated in Figure 19, indicates that the generated data adheres to the distribution of the original data. The points deviating from the line — the one in the top-right and the other in the bottom-left imply outliers and variations. The top-right outlier suggests larger values in the SMOTE-generated data than expected, possibly due to synthetic sample interpolation. The bottom-left dot implies smaller values in the original data, indicating inherent variability. These deviations, though minor, highlight disparities and demand consideration within the broader analysis and model development to ensure their impact is appropriately managed.

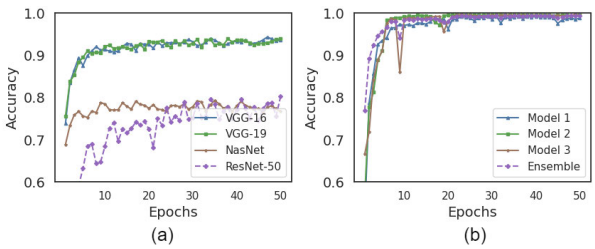

FIGURE 20. (a) val accuracy of pre-trained models on over-sampled dataset for 0 - HP (b) Val accuracy of proposed methods on over-sampled dataset for 0 - HP.

TABLE 3. Performance comparison of proposed methods with pre-trained models at NLC.

Model	Precision	Recall	F1-Score	Accuracy	Matthew Corr.	Kappa Stats.	Pred. Time
			Evaluation Metrics				
			Before Over-Sampling				
SCNN1	0.9908	99.12	0.9911	0.9908	0.9906	0.9906	0.383s
SCNN2	0.9916	0.9925	0.9926	0.9916	0.9919	0.9919	0.35s
SCNN3	0.9923	0.9917	0.9917	0.9922	0.9925	0.9925	0.217s
WVCNN	0.9925	0.9929	0.9929	0.9928	0.9932	0.9932	0.851s
ResNet	0.8804	0.8720	0.8641	0.8232	0.8648	0.8628	0.354s
NasNet	0.7409	0.7212	0.7209	0.7373	0.7585	0.7584	0.7s
VGG-19	0.9297	0.9351	0.9344	0.9263	0.9230	0.9225	0.319s
VGG-16	0.9358	0.9425	0.9425	0.9347	0.9415	0.9412	0.649s
			After Over-Sampling				
SCNN1	0.9930	0.9929	0.9929	0.9913	0.9923	0.9923	0.413s
SCNN2	0.9941	0.9940	0.9940	0.9957	0.9935	0.9935	0.377s
SCNN3	0.9909	0.9908	0.9907	0.9963	0.9900	0.9900	0.319s
WVCNN	0.9975	0.9971	0.9971	0.9971	0.9973	0.9975	0.850s
ResNet50	0.8213	0.8194	0.8128	0.8279	0.8062	0.8091	0.99s
NasNet	0.7864	0.7871	0.7871	0.7772	0.7656	0.7659	0.86s
VGG-19	0.9323	0.9236	0.9238	0.9345	0.9169	0.9176	0.693s
VGG-16	0.9468	0.9568	0.9552	0.9487	0.9540	0.9539	0.669s

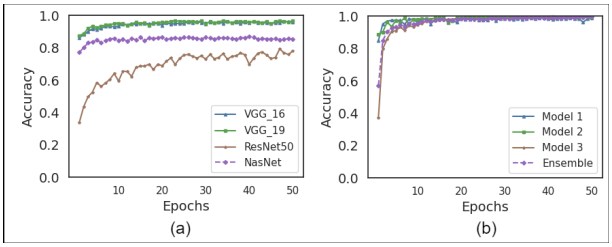

FIGURE 21. (a)Val accuracy of pre-trained models on over-sampled dataset for 1 - HP (b) Val accuracy of proposed methods on over-sampled dataset for 1 - HP.

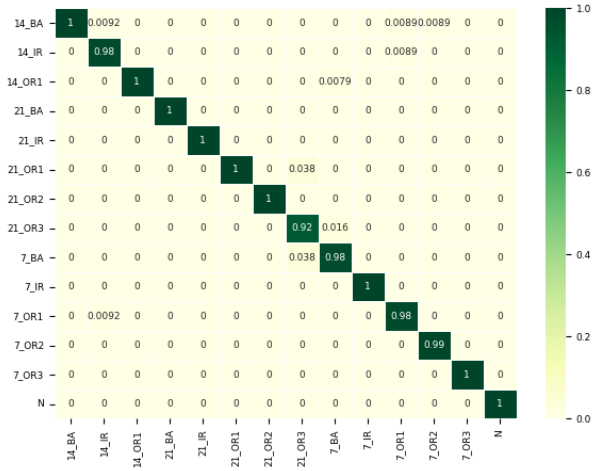

FIGURE 22. Confusion Matrix of Voting Ensemble Model (WVCNN) for over-sampled dataset.

2) EVALUATION OF WEIGHTED VOTING ENSEMBLE ON PROPOSED OVER-SAMPLING METHOD

The validation accuracies (VA) obtained through training the proposed models on scalo-gram representations yielded impressive results of 99.13%, 99.57%, and 99.63% for individual models, reaching a notably enhanced 99.71% accuracy after employing the weighted ensemble technique following DA-SMOTE augmentation for NLC. Similarly, the oversampled results for SLC yeilded 99.04%, 99.42% and 99.82% for individual CNN models and 99.87% for the respective ensemble. These outcomes provided a promising foundation, further weighed by comprehensive assessment across multiple evaluation metrics encompassing F1-score, recall, Matthews Correlation Coefficient (MCC), and Kappa statistic(KS). These metrics offered insights into the performance of individual models across different classes, shedding light on potential class-specific biases. The results were subsequently assessed against state-of-the-art transfer learning models such as ResNet50, VGG-19, NasNet, and VGG-16. Notably, the outcomes indicated an increase in performance of most transfer learning models as compared to the under-sampled counter-parts, with VGG-16 and VGG-19 achieving commendable validation accuracies of 93.45% and 94.87%, respectively. The remaining models achieved VA accuracies of 82.79% and 77.72% for NLC. For the SLC the validation accuracies of the pre-trained models are 95.97%, 96.61% for VGG_16 and VGG_19, while 78.08% and 85.51 % for ResNet50 and NasNet. These interpretations underscore the significant impact of the DA-SMOTE technique in enhancing the accuracy of the classification process and highlight the relative advantages of the ensemble approach in comparison to individual models and existing transfer learning methodologies.

3) COMPARISON WITH EXISTING STUDIES

There has been ongoing research in bearing fault detection domain, and as such many new algorithms and techniques have been proposed in the literature to effectively diagnose the bearing faults. Existing techniques have employed innovative approaches toward bearing classification with some including generative models to counter class imbalance and biasing whilst others opted for a time-frequency based approach [36], [37]. These novel takes at one end effectively

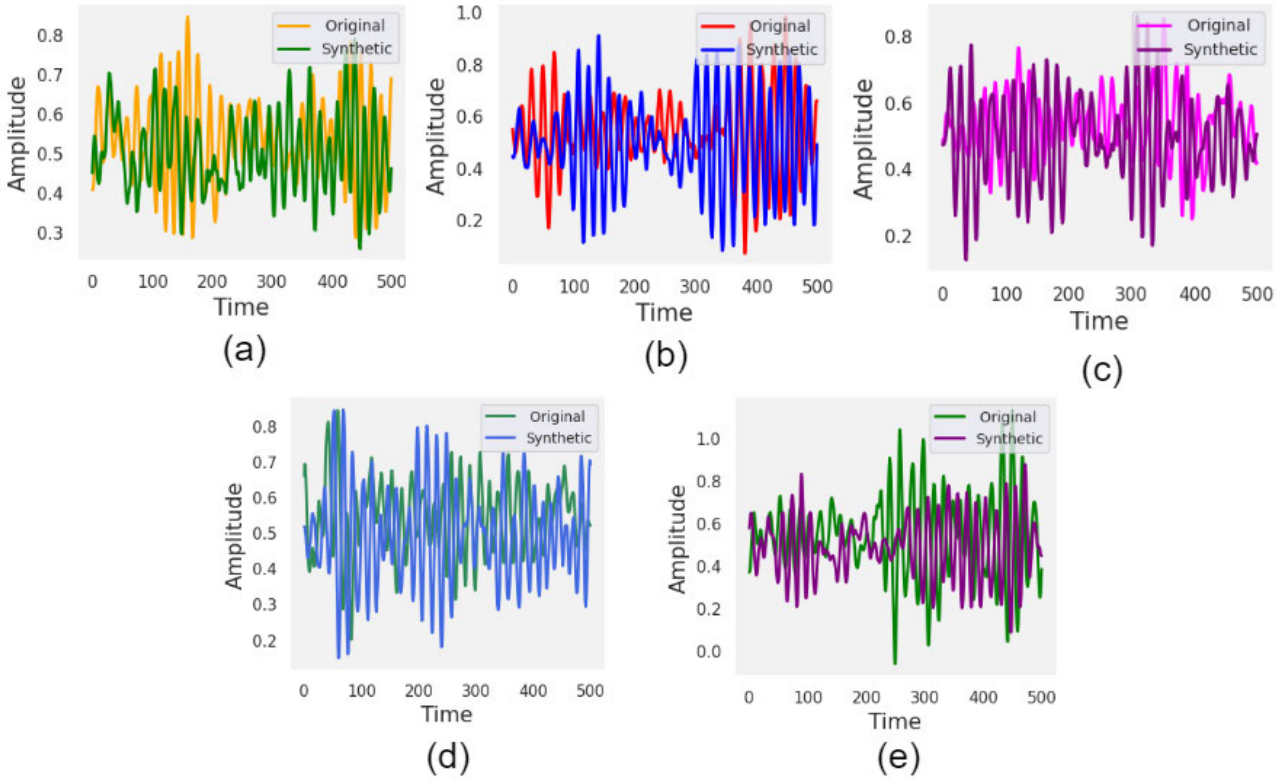

FIGURE 23. Visualization of original and synthetic wave-forms for imbalanced classes (a) 21_{OR2} (b) 21_{OR3} (c) 7_{OR2} (d) 7_{OR3} (e) 14_{IR}.

TABLE 4. Performance comparison of proposed methods with pre-trained models at SLC.

	Evaluation Metrics						
Model	Precision	Recall	F1-Score	Accuracy	Matthew Corr.	Kappa Stat.	Pred. Time
				Before Over-Sampling			
SCNN1	0.9816	0.9798	0.9807	0.9843	0.9763	0.9763	0.450s
SCNN2	0.9832	0.9732	0.9805	0.9866	0.9788	0.9751	0.465s
SCNN3	0.9894	0.9806	0.9813	0.9910	0.98.87	0.9875	0.477s
WVCNN	0.9897	0.9817	0.9811	0.9913	0.9905	0.9905	0.484s
NasNet	0.8423	0.8521	0.8445	0.8493	0.7025	0.6555	0.522s
ResNet	0.7208	0.7215	0.7236	0.7201	0.5032	0.4581	0.603s
VGG_16	0.9513	0.9415	0.9417	0.9502	0.9012	0.8723	0.499s
VGG_19	0.9607	0.9532	0.9518	0.9651	0.9255	0.8936	0.507s
				After Over-Sampling			
SCNN1	0.9826	0.9732	0.9815	0.9904	0.9878	0.9876	0.423s
SCNN2	0.9914	0.9853	0.9825	0.9942	0.9905	0.9912	0.437s
SCNN3	0.9945	0.9892	0.9917	0.9982	0.9962	0.9979	0.445s
WVCNN	0.9951	0.9905	0.9931	0.9987	0.9978	0.9980	0.451s
ResNet	0.7802	0.8015	0.7907	0.7808	0.7224	0.7017	0.505s
NasNet	0.8542	0.8736	0.8627	0.8551	0.7907	0.8013	0.524s
VGG_16	0.9601	0.9578	0.9566	0.9597	0.9504	0.9502	0.498s
VGG_19	0.9645	0.9587	0.9583	0.9651	0.9546	0.9555	0.502s

countered the associated problems but still had a large margin for improvement due to their complex lineament, inadequate evaluation criteria and lower accuracy. Additionally, most of the studies didn't address the class balancing while those which did were computationally expensive and complex in this regard. Whereas the proposed methods in this study effectively provide a straight forward take on data augmentation and class balancing by proposing a novel distance and density weighted hybrid take on SMOTE oversampling of under-represented bearing fault classes. Furthermore, the scalogram based representation was procured for a transient representation of time-frequency components with a weighted ensemble for robust classification. Our proposed ensemble method achieved state-of-the-art results on both imbalanced and balanced data, complemented by a variety of quantitative and qualitative evaluation criterion to ascertain the reliability of the models and the augmentations. These evaluations further reveal the relevance and performance

TABLE 5. Comparative analysis of this study with recent works.

Comparison With Existing Studies					
Reference	Year	Dataset	Methodology Used	Classes	Accuracy
[26]	2023	CWRU	Custom CNN architecture derived from ResNet-152.	4	96.42
[30]	2017	CWRU	Custom CNN architecture	4	99.5%
[25]	2023	CWRU	Graph Neural Network	4	99.1%
[22]	2022	CWRU	Residual Network based on Markov Transition Field	10	98.52%
[21]	2019	CWRU	Custom CNN architecture	4	98.10%
This Study	2023	CWRU	WVE original data (Imbalance) WVE with Novel Density Distance based SMOTE (balance data)	14	99.28% / 99.71%

of our proposed methodologies in contrast to the recently published works.

C. LIMITATIONS AND CHALLENGES

While our study proposed innovative take on bearing fault classification under imbalanced settings, there are some limitations that can further be worked upon. The proposed CNN algorithms although posses lower individual parametric count as given in Table 2, when used in ensemble the computational costs are somewhat increased. This can be improved in the future by working on more lightweight models or optimizing the proposed methods to suit the use-case. Furthermore, the generalizability of our proposed ensemble technique to diverse data-sets and industrial settings warrants further investigation. For instance, emphasizing the model's explainability while addressing the practical considerations, including computational resource requirements and real-world deployment challenges. The proposed DA-SMOTE, albeit performing efficiently in this scenario may need further evaluation on diverse data for a more credible insight into its potential and limitations.

V. CONCLUSION AND FUTURE WORK

The present study delves into the endeavor of precisely classifying bearing faults for an effective condition monitoring. These strives were directed towards classification across varying fault diameters and types which is a pivotal facet impacting machinery performance in both industrial and domestic contexts. The faulty conditions have far-reaching economic and safety implications, hence a focused exploration is undertaken at a sampling frequency of 48KHz to refine the study's scope.

In this context, an effective ensemble approach is used by combining the predictions of the three efficient individual CNN models in a weighted voting, while harnessing the collective potency of its constituents in an attempt to elevate the classification accuracy of fault conditions. This proposed approach leverages time series signals that are trans-formed into the respective scalogram representations—to fully exploit CNN's capacity for spatial analysis and feature extraction. Furthermore, a significant addition to this endeavor is the incorporation of the DA-SMOTE technique, which substantially enhanced the dataset quality by incorporating the interpolated samples within the under-represented classes while amplifying the robustness of the proposed models against skews and biases. The proposed ensemble model showed adaptability in handling the classification of scalogram images, while being robust to varying computational complexities through parameter adjustments. Within the 48KHz framework, the classification extended to 14 classes and yielded notably promising outcomes. In order to evaluate the adaptability of our methodologies across diverse conditions, transfer learning techniques are used for the effective comparison of our model's outcome with the state-of-the-art methods. In this regard, rigorous evaluation, employing metrics such as Accuracy, Recall, Precision, F1-Score, and Kappa Statistics, provides insights into model performance for each class, unearthing potential biases.

In summation, our weighted voting ensemble models outshine established transfer learning techniques and models featured in analogous studies concerning the same dataset. The acquired efficiency of our model shows considerable promise for real-time classification applications. While the study is anchored to specific fault diameters, frequency and classes, the work can be extended to the application of adept bearing fault detection and classification, driving improvements in maintenance practices, curbing downtime, and enhancing safety in industrial and domestic spheres. Future investigations will include further refining of the proposed methodologies and testing them on other datasets with requisite modifications within the oversampling methodology and the data-refining procedures.

ACKNOWLEDGMENT

The authors acknowledge the support from the Deanship of Scientific Research, Najran University. Kingdom of Saudi Arabia, for funding this work under the Distinguished Research funding program grant code number (NU/DRP/SERC/12/8).

REFERENCES

[1] X. Peng, B. Zhang, and D. Gao, "Research on fault diagnosis method of rolling bearing based on 2DCNN," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Aug. 2020, pp. 693–697.

[2] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[3] J. Li, Y. Liu, and Q. Li, "Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method," *Measurement*, vol. 189, Feb. 2022, Art. no. 110500.
[4] C. Li, J. V. de Oliveira, M. Cerrada, D. Cabrera, R. V. Sánchez, and G. Zurita, "A systematic review of fuzzy formalisms for bearing fault diagnosis," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 7, pp. 1362–1382, Jul. 2019, doi: 10.1109/TFUZZ.2018.2878200.
[5] A. Yunusa-Kaltungo, J. K. Sinha, and A. D. Nembhard, "A novel fault diagnosis technique for enhancing maintenance and reliability of rotating machines," *Struct. Health Monitor.*, vol. 14, no. 6, pp. 604–621, Nov. 2015.
[6] Y. Li, X. Wang, S. Si, and S. Huang, "Entropy based fault classification using the case western reserve university data: A benchmark study," *IEEE Trans. Rel.*, vol. 69, no. 2, pp. 754–767, Jun. 2020.
[7] A. Fernández, S. García, J. Luengo, E. Bernadó-Mansilla, and F. Herrera, "Genetics-based machine learning for rule induction: State of the art, taxonomy, and comparative study," *IEEE Trans. Evol. Comput.*, vol. 14, no. 6, pp. 913–941, Dec. 2010, doi: 10.1109/TEVC.2009.2039140.
[8] E. F. Swana, W. Doorsamy, and P. Bokoro, "Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset," *Sensors*, vol. 22, no. 9, p. 3246, Apr. 2022, doi: 10.3390/s22093246.
[9] H. Zhiyi, S. Haidong, J. Lin, C. Junsheng, and Y. Yu, "Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder," *Measurement*, vol. 152, Feb. 2020, Art. no. 107393.
[10] T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, and E. Xu, "Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions," *ISA Trans.*, vol. 119, pp. 152–171, Jan. 2022.
[11] J. Jiao, M. Zhao, J. Lin, and K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, vol. 417, pp. 36–63, Dec. 2020.
[12] S. Tang, S. Yuan, and Y. Zhu, "Deep learning-based intelligent fault diagnosis methods toward rotating machinery," *IEEE Access*, vol. 8, pp. 9335–9346, 2020, doi: 10.1109/ACCESS.2019.2963092.
[13] Q. Hang, J. Yang, and L. Xing, "Diagnosis of rolling bearing based on classification for high dimensional unbalanced data," *IEEE Access*, vol. 7, pp. 79159–79172, 2019, doi: 10.1109/ACCESS.2019.2919406.
[14] M. A. Ali, A. A. Bingamil, A. Jarndal, and I. Alsyouf, "The influence of handling imbalance classes on the classification of mechanical faults using neural networks," in *Proc. 8th Int. Conf. Model. Simul. Appl. Optim. (ICMSAO)*, Manama, Bahrain, Apr. 2019, pp. 1–5, doi: 10.1109/ICMSAO.2019.8880437.
[15] W. Sun, C. Cheng, and G. Yu, "Research on classification of imbalanced data set based on TMDSMOTE algorithm," *Frontiers Soc., Sci. Technol.*, vol. 2, no. 8, pp. 5–12, 2020.
[16] H. Li, R. Fan, and Q. Shi, "Generative oversampling and deep forest based minority-class sensitive fault diagnosis approach," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Toronto, ON, Canada, Oct. 2020, pp. 3629–3636, doi: 10.1109/SMC42975.2020.9283190.
[17] F. Duan, S. Zhang, Y. Yan, and Z. Cai, "An oversampling method of unbalanced data for mechanical fault diagnosis based on MeanRadius-SMOTE," *Sensors*, vol. 22, no. 14, p. 5166, Jul. 2022, doi: 10.3390/s22145166.
[18] D. S. Chaudhari and V. Kumar, "Review paper of wear on journal bearing," *Int. J. for Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 5, pp. 4826–4828, May 2023, doi: 10.22214/ijraset.2023.52800.
[19] E. Yadav and V. K. Chawla, "An explicit literature review on bearing materials and their defect detection techniques," *Mater. Today, Proc.*, vol. 50, pp. 1637–1643, Jan. 2022, doi: 10.1016/j.matpr.2021.09.132.
[20] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019, doi: 10.1109/ACCESS.2019.2912200.
[21] Li, Huang, and Ji, "Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network," *Sensors*, vol. 19, no. 9, p. 2034, Apr. 2019, doi: 10.3390/s19092034.
[22] J. Yan, J. Kan, and H. Luo, "Rolling bearing fault diagnosis based on Markov transition field and residual network," *Sensors*, vol. 22, no. 10, p. 3936, May 2022, doi: 10.3390/s22103936.
[23] H. Tang, S. Gao, L. Wang, X. Li, B. Li, and S. Pang, "A novel intelligent fault diagnosis method for rolling bearings based on Wasserstein generative adversarial network and convolutional neural network under unbalanced dataset," *Sensors*, vol. 21, no. 20, p. 6754, Oct. 2021, doi: 10.3390/s21206754.
[24] X. Zhao, M. Ma, and F. Shao, "Bearing fault diagnosis method based on improved Siamese neural network with small sample," *J. Cloud Comput.*, vol. 11, no. 1, p. 79, Nov. 2022, doi: 10.1186/s13677-022-00350-1.
[25] L. Xiao, X. Yang, and X. Yang, "A graph neural network-based bearing fault detection method," *Sci. Rep.*, vol. 13, no. 1, p. 5286, Mar. 2023, doi: 10.1038/s41598-023-32369-y.
[26] G. Wu, X. Ji, G. Yang, Y. Jia, and C. Cao, "Signal-to-image: Rolling bearing fault diagnosis using ResNet family deep-learning models," *Processes*, vol. 11, no. 5, p. 1527, May 2023, doi: 10.3390/pr11051527.
[27] *Download a Data File | Case School of Engineering | Case Western Reserve University*. Accessed: Jul. 8, 2023. [Online]. Available: https://engineering.case.edu/bearingdatacenter/download-data-file
[28] D. Verstraete, A. Ferrada, E. L Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017, pp. 1–17, 2017, Art. no. 5067651, doi: 10.1155/2017/5067651.
[29] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, p. 52, Mar. 2023, doi: 10.3390/computation11030052.
[30] F. Example, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017, pp. 1–17, Oct. 2017, doi: 10.1155/2017/.5067651.
[31] *Apparatus & Procedures | Case School of Engineering | Case Western Reserve University*. Accessed: Jul. 8, 2023 [Online]. Available: https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures
[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*. Computational and Biological Learning Society, 2015, pp. 1–14.
[33] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[36] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
[37] S. P. Lee, "Bearing life evaluation of automotive wheel bearing considering operation loading and rotation speed," *Trans. Korean Soc. Mech. Eng. A*, vol. 40, no. 6, pp. 595–602, Jun. 2016, doi: 10.3795/ksme-a.2016.40.6.595.

MUHAMMAD IRFAN received the Ph.D. degree in electrical and electronic engineering from Universiti Teknologi PETRONAS, Malaysia, in 2016. He has two years of industry experience (October 2009–October 2011) and six years of academic experience (since January 2017) in teaching and research. Currently, he is an Associate Professor with the Electrical Engineering Department, Najran University, Saudi Arabia. He has authored more than 250 research articles in reputed journals, books, and conference proceedings (Google Scholar citations of 2700 and H-index of 23). His current research interests include automation and process control, energy efficiency, condition monitoring, vibration analysis, artificial intelligence, the Internet of Things (IoT), big data analytics, smart cities, and smart healthcare.

ZOHAIB MUSHTAQ received the B.Sc. degree from Islamia University, the M.S. degree from Government College University, Lahore, and the Ph.D. degree in electrical engineering from the National Taiwan University of Science and Technology, in 2020. He was an Assistant Professor with Riphah International University. He is currently an Assistant Professor of electrical engineering with the University of Sargodha, Sargodha, Pakistan. He has published research in IEEE and other reputable journals. His current interests include neural networks, machine learning, deep learning, computer vision, and data science.

NABEEL AHMED KHAN received the Electrical Engineering degree from Riphah International University, Islamabad. His current research interests include deep learning, signal processing, computer vision, machine learning, and data science.

SALIM NASAR FARAJ MURSAL received the Ph.D. degree in electronic and communication engineering. Currently, he is an Assistant Professor with the College of Engineering, Najran University, Saudi Arabia. He has more than ten years of academic experience in teaching and research. He has authored research articles in several journals, books, and conference proceedings.

SAIFUR RAHMAN received the Ph.D. degree in electrical and electronic engineering. Currently, he is an Associate Professor with the College of Engineering, Najran University, Saudi Arabia. He has more than 12 years of academic experience in teaching and research. He has authored more than 70 research articles in reputed journals, books, and conference proceedings. His current research interests include artificial intelligence, the Internet of Things (IoT), big data analytics, smart cities, and smart healthcare.

MUAWIA ABDELKAFI MAGZOUB (Member, IEEE) received the B.Sc. degree in electrical engineering from Al-Zaim Al-Azhari University, Khartoum, Sudan, in 2005, the master's degree in electronics engineering from the Sudan University of Science and Technology, Khartoum, in 2008, and the Ph.D. degree in electrical and electronics engineering from Universiti Teknologi PETRONAS, Malaysia, in 2017. He was a Lecturer with the Al-Jraif Sharq Technical College, Khartoum. He was a Senior Engineer with the Department of Tooling and Automation, Finisar II-VI (M) Sdn Bhd, Malaysia, for two years and eight months. He was a Postdoctoral Researcher with the Department of Electrical and Electronics Engineering, Universiti Teknologi PETRONAS, for one year. He is currently an Assistant Professor with the Department of Electrical and Electronics Engineering, Sudan Technological University, and National University, Sudan. His current research interests include automation and process control, power electronics, machine control and machine applications, artificial intelligence, smart grids, renewable energy and energy systems, machine learning, and data analysis.

MUHAMMAD ARMGHAN LATIF received the B.Sc. degree from the University of Management and Technology, Sialkot, Pakistan, and the M.S. degree from Cleveland State University, Cleveland, Ohio, USA. His current research interests include neural networks, machine learning, deep learning, computer vision, and data science.

FAISAL ALTHOBIANI is currently an Associate Professor with the Marine Engineering Department, Faculty of Maritime, King Abdulaziz University, Jeddah, Saudi Arabia. He has more than six years of academic experience in teaching and research. He has authored several research articles in reputed journals, books, and conference proceedings. His current research interests include marine sciences, marine engineering, condition monitoring, and fault diagnosis systems.

GHULAM ABBAS (Senior Member, IEEE) received the B.E. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2004, and the M.E. and Ph.D. degrees in electrical engineering from Institut National des Sciences Appliquées de Lyon (INSA Lyon), France, in 2008 and 2012, respectively. He is currently a Professor with the Department of Electrical Engineering, The University of Lahore, Lahore, Pakistan. He has published several papers in various international and national IEEE conferences and journals of repute, such as IEEE, Elsevier, and MDPI. His current research interests include analog and digital controller designs for power switching converters, soft computing, and power system optimization. He is a Reviewer of several premier conferences and journals, including IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *Asian Journal of Control*, *Engineering Optimization*, and IEEE ACCESS. He has served for several electrical and computer engineering journals as a guest editor.

IMRAN KHAN YOUSUFZAI received the B.Sc. degree in electrical engineering from the Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan, in 2010, the M.Sc. degree in control engineering from Mohammad Ali Jinnah University, Islamabad, in 2012, and the Ph.D. degree in control engineering from the Capital University of Science and Technology, Islamabad, in 2016. He was a Visiting Scholar with The Ohio State University, Columbus, USA. He is currently an Assistant Professor with the Department of Electrical Engineering, College of Engineering and Technology, University of Sargodha, Pakistan. He has more than eight years of industrial and teaching experience. His current research interests include the analysis and design of nonlinear control systems, especially sliding mode control, for various industrial applications.