Transfer Learning Approach using Simulated Induction Motor Bearing Data: A Comparative Analysis of SE-ResNet and its Hybrid Variants

Lydiah Aywa Sikinyi
Department of Electrical Engineering, Pan African University, Institute for Basic Sciences Technology and Innovation, Nairobi, Kenya | Department of Electrical and Communication Engineering, Masinde Muliro University of Science and Technology, Kakamega, Kenya
lydiah.sikinyi@gmail.com (corresponding author)

Christopher Maina Muriithi
Department of Electrical Engineering, Pan African University, Institute for Basic Sciences Technology and Innovation, Nairobi, Kenya | Department of Electrical and Electronics Engineering, Murang'a University of Technology, Murang'a, Kenya
cmmuriithi@mut.ac.ke

Livingstone Ngoo
Department of Electrical Engineering, Pan African University, Institute for Basic Sciences Technology and Innovation, Nairobi, Kenya | Department of Electrical and Electronics Engineering, Multimedia University of Kenya, Nairobi, Kenya
livingngoo@gmail.com

Duncan Shitubi
DEFEND Organization, Toronto, Canada
dshitubi@outlook.com

Received: 3 March 2025 | Revised: 31 March 2025 | Accepted: 19 April 2025

 DOI: https://doi.org/10.48084/etasr.10758

ABSTRACT

Early detection of incipient bearing faults in induction motors has proven crucial in predictive maintenance, helping avoid machine downtime and costly repairs. The main challenge is collecting sufficient data for deep learning models since faults are a rare occurrence. This paper investigates the efficacy of a transfer learning approach for induction motor bearing fault diagnosis using simulated vibration data. Healthy and faulty bearings of different severities were simulated in MATLAB for various noise magnitudes. A Squeeze and Excitation Residual Network (SE-ResNet), previously trained on a large dataset for bearing faults of a Permanent Magnet Synchronous Motor (PMSM), is used as a feature extractor. By leveraging pre-trained knowledge, the model's weights were fine-tuned using Bayesian Optimization, aiming to mitigate the data scarcity issue while maintaining accurate fault classification. The model's performance was compared against three hybrid architectures incorporating Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM) layers. The study aims to assess the impact of adding recurrent layers to capture temporal dependencies within simulated vibration signals. Contrary to expectations, the hybrid models did not improve the classification accuracy compared to the standalone pre-trained SE-ResNet. The test accuracy remained the same for all the models at 97.297% whereas the computational cost increased for the hybrid models. This paper analyzes these findings, highlighting the challenges of transfer learning with simulated data.

Keywords-Bayesian optimization; bearing fault detection; hybrid architectures; modeling and simulation; squeeze and excitation residual network; transfer learning

I. INTRODUCTION

Induction motors play a pivotal role in industrial applications, making their operational reliability critical. Bearings, which support the rotational movement of the motor, are often susceptible to wear and faults over time [1, 2]. When left undetected, bearing faults can propagate and cause severe damage to machinery, resulting in system breakdowns and costly correctional maintenance. Therefore, early fault detection strategies aim at maximizing motor lifespan and preventing major disruptions in industrial processes. Traditional bearing fault detection often relies on preventive maintenance and manual analysis, which can be insufficient in identifying early-stage failures. The development of advanced machine learning techniques has opened new avenues for improving fault diagnosis [3, 4]. However, it calls for feature engineering [5] which can be time-consuming and requires domain expertise. Deep learning, which automatically learns complex features from raw data, has emerged as a promising solution for bearing fault detection [6].

Various methodologies have been proposed to enhance detection accuracy including the use of vibration analysis [7], which when combined with modern deep learning techniques, provides a comprehensive approach to bearing health assessment [8]. For instance, in [9], the vibration signals collected from sensors were passed through variational mode decomposition and the utilized one-dimensional (1-D) Convolutional Neural Network (CNN) was able to diagnose bearing faults with an accuracy of 92.22%. A CNN model using a two-dimensional (2-D) form of vibration signal instead of 1-D was proposed in [10], achieving an accuracy of 97.74%. Authors in [11] presented a hybrid model combining CNN and Gated Recurrent Units (GRUs), integrated with envelope analysis and adaptive mean filtering techniques. These techniques suppressed noise achieving an accuracy of 99.25% when validated using an open-source dataset.

Despite these advancements, one of the primary challenges in applying deep learning to bearing fault detection is the scarcity of labelled data [6]. Faults are often rare events, making the collection of sufficient training data difficult [12]. Datasets can be generated from experimental set-ups but creating a dataset that encompasses various fault severities, environmental conditions, and noise magnitudes is usually time-consuming and resource-intensive. Emphasis is therefore put on the importance of open-source datasets, which provide a foundation for experimentation and benchmarking since they offer a large amount of data. In [13], the Case Western Reserve University (CWRU) bearing database was used to validate the proposed multi-scale domain adaptive network achieving an accuracy greater than 99%. However, open-source datasets are typically created for specific research projects, which often have a narrow focus. For example, a dataset might be designed to study a particular type of bearing under controlled conditions in a specific machine of a certain power rating. This restricts the applicability of open-source datasets to broader scenarios.

Recognizing the data-intensive nature of deep learning, researchers have explored transfer learning as a viable solution to address the scarcity of fault data. Transfer learning has gained significant attention in fault diagnosis by enabling models trained on one dataset, such as an open-source dataset, to generalize effectively to another dataset. This approach is particularly valuable when dealing with limited target domain data. This technique leverages pre-trained models on related tasks to mitigate the need for extensive labelled data in the target domain. The experimental results in [14] suggest that using mixed data as training samples to build neural network models can enhance the accuracy of transfer learning defect identification by over 10%, whereas in [15] the classification results show that using transfer learning reduces training time while enhancing diagnostic accuracy thus showcasing the viability of transfer learning.

Various studies propose data augmentation [16] and simulation techniques to mitigate the data scarcity issue. A MATLAB-based tutorial for simulating rolling element bearing faults was provided in [17] highlighting the ability to generate synthetic fault signals for diagnostic purposes. Although simulating data is a promising alternative, its effectiveness in capturing real-world complexities remains a topic of investigation.

This paper proposes simulating in MATLAB healthy and faulty induction motor bearings of varying severities and noise magnitudes that mimic real-world faults. The simulated bearing dataset is used to train a pre-trained Squeeze and Excitation Residual Network (SE-ResNet) previously developed by the authors of this paper in [12] using transfer learning strategies. The SE-ResNet model is enhanced with three different Recurrent Neural Network (RNN) variants to capture temporal dependencies in vibration signals. The performance of the SE-ResNet and the three hybrid models is compared on both real-world and simulated datasets, highlighting the challenges and potential of transfer learning. This approach demonstrates the potential of leveraging simulated data and transfer learning to overcome data scarcity challenges, achieving impressive test accuracy on the custom dataset.

This research is essential since it presents a scalable and efficient approach to diagnosing bearing faults, which can be applied in environments where the provision of real-world data is limited. This solution offers industries a reliable predictive maintenance tool capable of detecting incipient faults and preventing catastrophic failures in induction motors.

II. METHODOLOGY

A. Mathematical Modeling of Induction Motor in MATLAB Simulink

In this study, a 370 W, 415 V, 50 Hz, 2-pole, three-phase (3-ph) WEG induction motor with a bearing type 6203 was selected with the following parameters: $R_1 = 22.6345\ \Omega$, $R'_2 = 11.1965\ \Omega$, $X_1 = 13.5498\ \Omega$, $X'_2 = 14.7358\ \Omega$, $X_m = 462.9375\ \Omega$, $I_{rated} = 0.84\ \text{A}$, $T_{rated} = 1.26\ \text{Nm}$, $N_{rated} = 2825\ \text{rpm}$, and $J = 0.0004\ \text{kgm}^2$ [18]. The experiments were performed at a high sampling rate of 64 kHz.

A detailed dynamic mathematical model of the induction motor was developed in Simulink [19]. The model incorporated the following components: stator circuit, rotor circuit, and mechanical subsystem. The stator circuit modelled the

electrical characteristics of the stator windings and included voltage equations and stator windings represented by resistances and inductances. The rotor circuit modelled the electrical and mechanical behaviour of the rotor and consisted of rotor windings represented by resistances, inductances, electromagnetic induction, and voltage and current equations. The mechanical subsystem included the rotor inertia, bearing dynamics, and load torque. The electromagnetic torque equation for the induction motor was given by:

$$T_e = \frac{3}{2} \times \frac{p}{2} \times L_m \times (i_{qs} i_{dr} - i_{ds} i_{qr}) \quad (1)$$

where p is the number of poles, L_m is the mutual inductance between the stator and rotor windings, while i_{ds}, i_{qs} are the stator currents and i_{dr}, i_{qr} are the rotor currents in the direct (d) and quadrature (q) axes.

B. Simulation of Healthy and Faulty Bearings

Bearing faults of varying severities were simulated using MATLAB to generate a diverse dataset. The simulation involved introducing artificial defects into the bearing model and generating corresponding vibration signals. The faults were modelled by utilizing the dynamic torque equations of the induction motor, and vibration was derived from the acceleration of the rotor's motion. The torque equation governing the rotor's dynamics is expressed as:

$$T_e(t) = J \frac{d\omega(t)}{dt} + T_{load}(t) \quad (2)$$

where T_e is the electromagnetic torque generated by the motor, J is the rotor's moment of inertia, $\omega(t)$ is the angular velocity of the rotor, and T_{load} is the load torque applied to the system.

The acceleration $\alpha(t)$ of the rotor, which was used as the vibration signal, was derived from the angular velocity:

$$\alpha(t) = \frac{d\omega(t)}{dt} = \frac{T_e(t) - T_{load}(t)}{J} \quad (3)$$

The load torque for healthy bearings was taken as $T_{load} = T_0$. In the presence of bearing faults, the load torque was modified to reflect the impact of the fault. Different severities of the bearing faults were modelled by adjusting the amplitude of the fault-induced load torque oscillations, which introduced disturbances to the rotor's motion [20, 21]:

$$T_{load}(t) = T_0(1 + A_{fault} \cdot k_s \cdot \cos(\omega_{fault} t)) \quad (4)$$

where T_0 is the nominal load torque, A_{fault} is the fault severity amplitude, $k_s = l/1000$ is the severity factor with l being the length of damage, and $\omega_{fault} = 2\pi f_c$ is the angular fault velocity. The characteristic frequency f_c was for faults in the inner raceway, the outer raceway, or a combination of the aforementioned faults. This framework simulated both healthy and faulty bearings across various severity levels based on the length (l) of the damage, i.e. minor: $l \leq 2$ mm, moderate: 2 mm $< l \leq 4.5$ mm, and severe faults: $l > 4.5$ mm [22].

The characteristic frequencies of rolling bearings were calculated by [20]:

$$f_{outer} = \frac{N_b}{2} \cdot f_r \cdot (1 - \frac{D_b}{D_c} \cos\beta) \quad (5)$$

$$f_{inner} = \frac{N_b}{2} \cdot f_r \cdot (1 + \frac{D_b}{D_c} \cos\beta) \quad (6)$$

where N_b is the number of bearing balls, D_c is the bearing separator diameter (mm), D_b is the ball diameter (mm), D_{inner} is the inner raceway ring diameter (mm), D_{outer} is the outer raceway ring diameter (mm), β is the ball contact angle, and f_r is the rotor speed in Hz. The load torque of the combination fault $T_{load.combi}$ was found by adding the load torque of the inner race fault to that of the outer race fault:

$$\begin{aligned} T_{load.combi}(t) = {} & T_0(1 + A_{fault} \cdot k_s \cdot \cos(\omega_{inner} t)) \\ & + T_0(1 + A_{fault} \cdot k_s \cdot \cos(\omega_{outer} t)) \end{aligned} \quad (7)$$

where ω_{inner} and ω_{outer} represent the angular velocity of the inner and outer race fault, respectively.

A Gaussian noise block was introduced into the Simulink model and added to the vibration signals. Its variance was adjusted (0.01, 0.1, 0.2, and 0.3) to create different noise magnitudes, allowing the mimicking of real-world conditions to simulate various levels of environmental disturbances and measurement noise, ranging from low-noise environments with minimal external disturbances and high-quality measurements to highly noisy industrial scenarios with significant external noise and measurement variability. This assisted in the evaluation of the model's robustness against varying noise levels. The noise equation was represented as:

$$\alpha_{noisy}(t) = \alpha(t) + N(t) \quad (8)$$

where $\alpha_{noisy}(t)$ is the noisy vibration signal, $\alpha(t)$ is the clean vibration signal derived from the torque equation, and $N(t)$ is the noise term modelled as Gaussian white noise with varying standard deviations.

A custom dataset was created consisting of labelled images representing the inner race, outer race, combination (inner and outer race), and healthy bearings across different severities and noise levels. The rotational speed and load torque were kept constant as the variance was varied for the three operating conditions shown in Table I. For each of the settings, 4 measurements of 4.5 s each were recorded for the four variances selected, and the first 0.5 s were cut out to leave out the transients. The scope block was used to visualize the data allowing for real-time observation of the signal characteristics. These data were then exported to the MATLAB workspace and processed to generate images, which served as input for the SE-ResNet and hybrid models.

TABLE I. OPERATING CONDITIONS FOR SIMULATED DATA

No.	Rotational speed [rpm]	Load torque [Nm]	Name of setting
0	1500	0.7	N15_M07
1	900	0.7	N09_M07
2	1500	0.1	N15_M01

C. Data Pre-processing and Model Training

The original [656×875×3] images were scaled and cropped to 224×224×3 size to prevent overfitting and improve the model's ability to recognize the target objects. The simulated dataset had only 245 images, i.e. 12 images from the healthy class, 32 images from the combination faults class, 95 images representing the outer race faults class, and 106 representing images from the inner race faults class. Some of the images are presented in Figure 1.

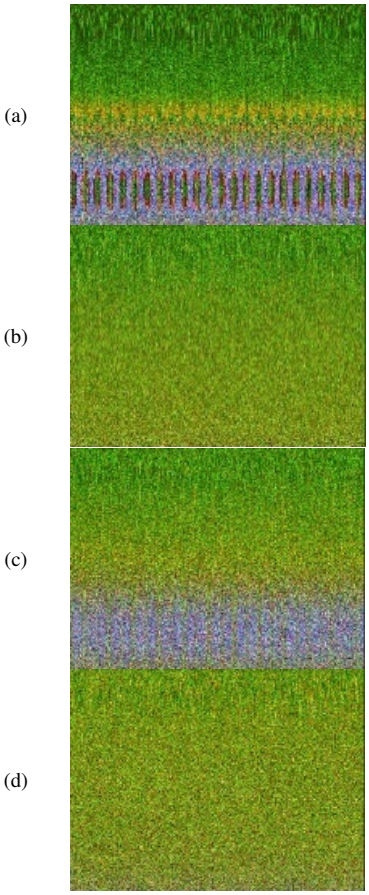

Fig. 1. Simulated bearing data images for: (a) Combination bearing fault (N15_M01_FV_KB02_1), (b) healthy bearing (N15_M01_FV_K001_1), (c) inner raceway bearing fault (N15_M01_FV_KI02_2), and (d) outer raceway bearing fault (N15_M01_FV_KA01_3).

These images were then used to test the SE-ResNet model from [12]. The SE-ResNet was originally trained on bearing fault data from a Permanent Magnet Synchronous Motor (PMSM) and training it on induction motor faults would prove its ability to generalize to a related but unseen task. However, this model did not perform well on classifying the simulated test dataset. All the combination faults were misclassified as either inner race fault or outer race fault. The combination fault images were therefore divided into either outer raceway or inner raceway fault, depending on which fault had more severity. The final layers of the SE-ResNet were changed to capture the new conditions, therefore the classification layer had three classes instead of four.

The simulated images were divided using a 70:15:15 split resulting in 171 for training, 37 for validation, and 37 for testing. The SE-ResNet model was retrained using the custom dataset through transfer learning and Bayesian Optimization was leveraged to optimize the model's performance. Bayesian Optimization is a powerful technique that systematically explores the model's hyperparameter space, including parameters such as learning rate, batch size, and number of epochs, to identify the optimal configuration.

D. Training of SE-ResNet Model using Transfer Learning and Bayesian Optimization

The SE-ResNet had performed well on training, validating and testing on an open-source bearing dataset, therefore transfer learning was used to improve the model's generalization to a new but related task i.e. the pre-trained model was used to classify induction motor bearing faults. Instead of starting the learning process from scratch, knowledge learnt from the previous task, which was trained with a lot of labelled data, was leveraged. The initial convolutional layers responsible for low-level feature extraction of the SE-ResNet were frozen to retain the learned representations from the open-source bearing dataset, while the higher layers responsible for classification were retrained using the simulated induction motor dataset. This fine-tuning allowed the model to adapt to the specific characteristics of the new dataset. The initial hyperparameters chosen were a learning rate of 0.0119, momentum of 0.9122, minibatch-size of 64, L2 regularization of 3.52×10^{-5} and 20 epochs. It became evident that the number of epochs was not enough for the model to learn all the necessary features and that the chosen learning rate was overfitting. An exhaustive sweep was done to help in choosing a suitable range for optimization. Bayesian Optimization strategy was then performed on MATLAB's experiment manager for 30 trials to automatically search for the optimal hyperparameter configurations. The following ranges were chosen: learning rate [0.001-0.09], categorical mini batch-size [16, 32, 64, 128], momentum [0.8–0.98], number of epochs [50-150], and dropout rate [$1e^{-5}$ – $1e^{-2}$]. This approach efficiently identified the best parameters while enhancing model performance without extensive manual tuning.

E. Training of Hybrid Models using Transfer Learning and Bayesian Optimization

RNNs are known to have the vanishing gradient problem. For this study, variants of RNNs i.e. LSTM, BiLSTM, and GRU were used since they are effective in learning long-term dependencies. Each architecture combined a pre-trained SE-ResNet initialized with weights to leverage spatial feature extraction capabilities, with RNN layers to model temporal dynamics. Only one layer of each variant was added in place of

the RNN Layer block shown in Figure 2. This led to a total of three hybrid models i.e. SE-ResNet-LSTM, SE-ResNet-BiLSTM, and SE-ResNet-GRU.

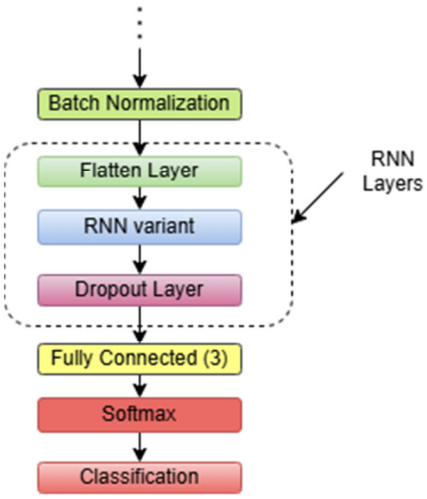

Fig. 2. Layers of the hybrid models.

Transfer learning was applied by fine-tuning the SE-ResNet backbone while training the recurrent layers from scratch, balancing the retention of generalized visual patterns with adaptation to task-specific temporal sequences. Bayesian optimization was then employed to efficiently navigate the hyperparameter space, optimizing critical parameters such as learning rate and dropout rates [0.1 -0.5]. The range for the other parameters remained the same as for the SE-ResNet model. This approach minimized manual tuning and accelerated convergence by iteratively updating probabilistic models to prioritize hyperparameters that improve validation accuracy.

III. RESULTS AND DISCUSSION

A. *Performance of the SE-ResNet Model on Simulated Dataset*

The acceleration signals from the simulation exhibited distinct patterns corresponding to healthy and faulty conditions. As fault severity escalated, the amplitude and frequency variations also increased. The addition of Gaussian white noise with varying variance levels effectively simulated different operational environments. The 245 images created were first used as a test dataset on the pre-trained SE-ResNet. The classification accuracy of the model was very poor at 40.408%. This was because the model was trained on a different machine type (PMSM) with a distinct set of bearing fault characteristics. Since the induction motor's mechanical properties differ from those of PMSM, the fault patterns in the vibration data were different. The operational conditions of the test dataset did not also fully match the conditions of the training dataset. The pre-trained model was therefore retrained with the new split data leading to an improved classification performance of 86.531%. However, all the combination faults were misclassified as seen in Figure 3.

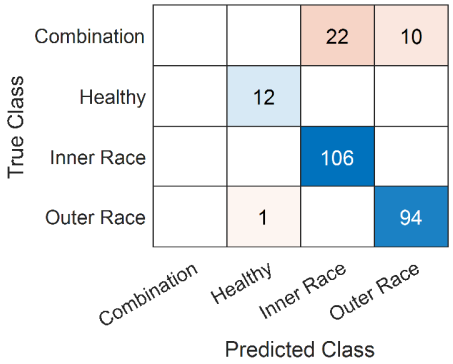

Fig. 3. Confusion matrix of the SE-ResNet model.

Since the primary focus was the early detection of the faults, the combination fault class was eliminated from the dataset and its images were distributed between inner raceway and outer raceway faults depending on which fault was more severe. The class reduction led to an increase in data for the inner raceway and outer raceway faults, allowing the model to learn unique characteristics. This simplified the classification task and improved the model's performance, leading to a more accurate and reliable prediction. Bayesian Optimization was used to train the models using a set range of hyperparameters. The best trial, which took 12 min, 54 s, achieved a training and validation accuracy of 100% as shown in Figure 4, and had the following hyperparameter values: learning rate: 0.0055, momentum: 0.8903, L2 regularization: 0.0046, epochs: 74, and mini-batch size: 16. This model performed well when tested using the remaining 15% of the custom dataset, achieving a test accuracy of 97.297%, demonstrating its efficacy in classifying bearing faults based on simulated vibration data. The confusion matrix in Figure 5 provides a detailed view of the classification performance across different fault categories, while Figures 6 and 7 show the accuracy and loss graphs of the SE-ResNet after transfer learning.

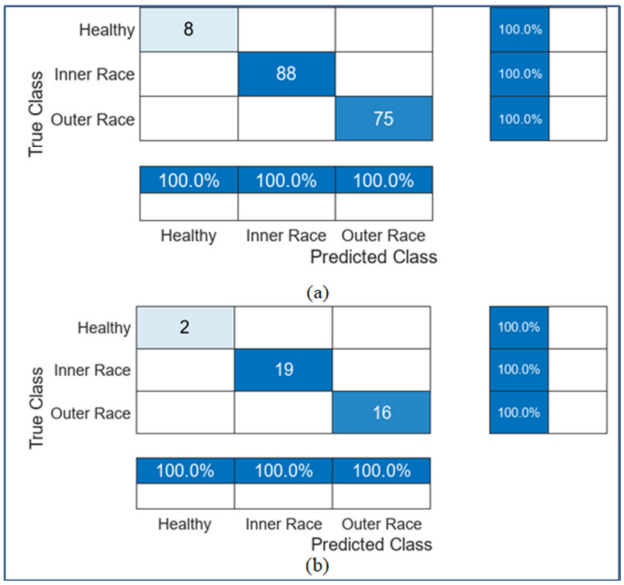

Fig. 4. Confusion matrices of (a) training and (b) validation data of the retrained SE-ResNet model.

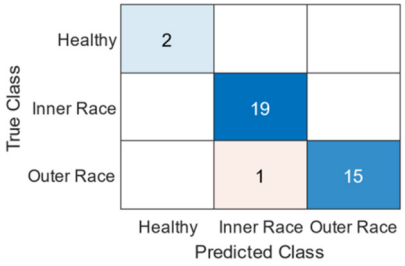

Fig. 5. Confusion matrix for the test dataset of a retrained SE-ResNet model.

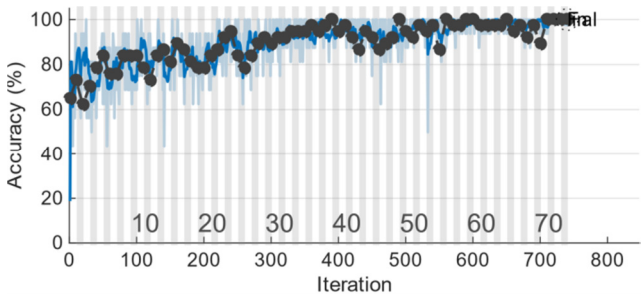

Fig. 6. Training (blue) and validation (black) accuracy graph for the SE-ResNet after transfer learning.

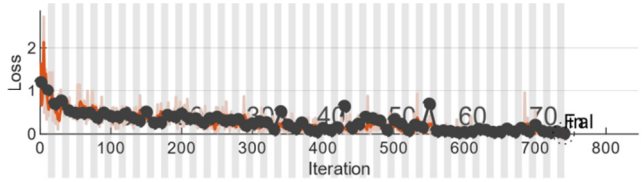

Fig. 7. Training (blue) and validation (black) accuracy graph for the SE-ResNet after transfer learning.

B. Performance of the SE-ResNet-LSTM Model on the Simulated Dataset

The SE-ResNet-LSTM model combines the deep feature extraction capabilities of the SE-ResNet with the temporal modeling power of the LSTM network. The motivation behind this architecture is to leverage SE-ResNet for spatial feature extraction from vibration scalograms while utilizing LSTM layers to capture temporal dependencies present in the data. With a learning rate of 0.0011, momentum of 0.9703, L2-regularization of 0.0027, 138 epochs, minibatch-size of 128, and a dropout probability of 0.1, the model achieved a training accuracy of 99.2188% and a validation accuracy of 94.5946%. This took a total of 18 min and 22 s. Figures 8 and 9 show the accuracy and loss graphs.

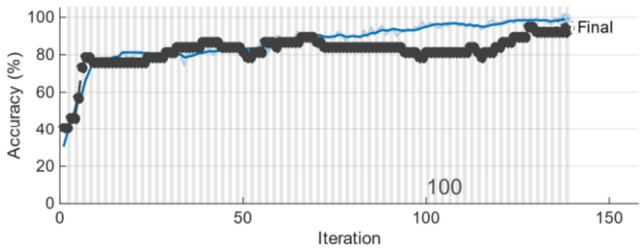

Fig. 8. Training (blue) and validation (black) accuracy graph for the SE-ResNet-LSTM after transfer learning.

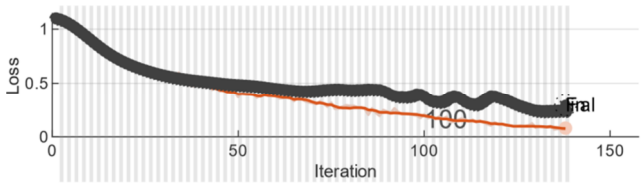

Fig. 9. Training (blue) and validation (black) accuracy graph for the SE-ResNet-LSTM after transfer learning.

The training accuracy, as seen in the confusion matrix in Figure 10(a), dropped from 100% for SE-ResNet to 99.22% for SE-ResNet-LSTM, while the validation accuracy decreased more significantly to 94.59% as seen in Figure 10(b). This suggests that the LSTM layer might not have effectively captured useful temporal patterns from the simulated dataset, potentially indicating a lack of strong temporal dependencies in the simulated vibration signals. Additionally, the training time increased due to the recurrent nature of LSTM, leading to higher computational costs. The test accuracy remained the same as the base SE-ResNet model at 97.297%, as shown by the confusion matrix in Figure 10(c).

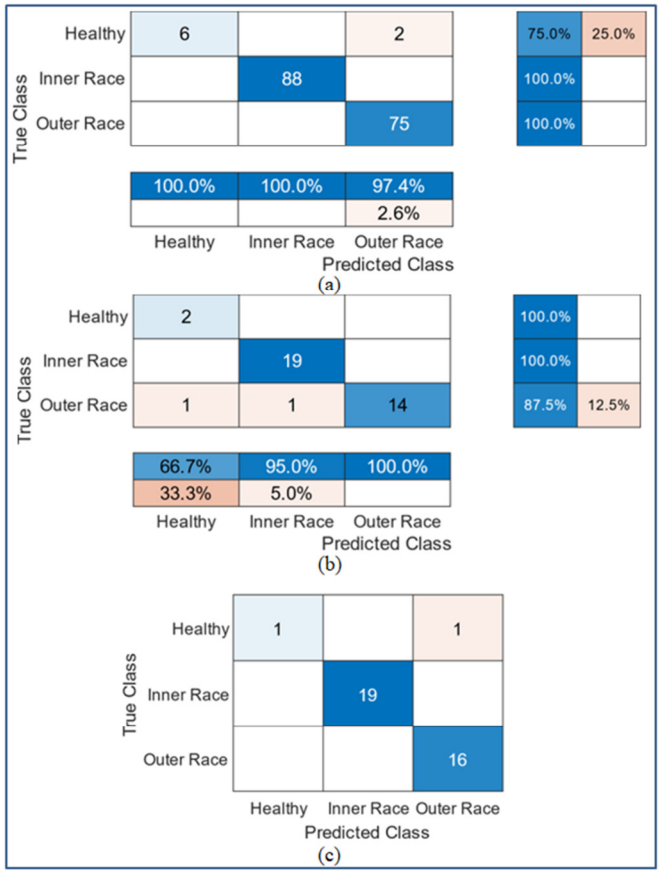

Fig. 10. Confusion matrices for (a) training, (b) validation, and (c) test datasets of the retrained SE-ResNet-LSTM model.

C. Performance of the SE-ResNet-GRU Model on the Simulated Dataset

The SE-ResNet-GRU model demonstrated a slightly improved training accuracy of 99.61% over the SE-ResNet-LSTM (99.22%), with increased memory requirements and training time of 37 min 33 s, because the minibatch-size of the best performing experiment was small at only 16 as compared to the LSTM's 128. The other parameters were: learning rate 0.0559, momentum 0.9439, L2-regularization 0.0003, 100 epochs, and a dropout probability of 0.5. The model achieved a validation accuracy of 94.5946%, the same as that of SE-ResNet-LSTM.

Similarly, the test accuracy on the simulated dataset was the same to the baseline SE-ResNet model's at 97.297%, further supporting the hypothesis that the simulated dataset lacks strong temporal dependencies that would justify the inclusion of recurrent layers. Figures 11 and 12 show the training and validation accuracy and loss graphs, respectively.

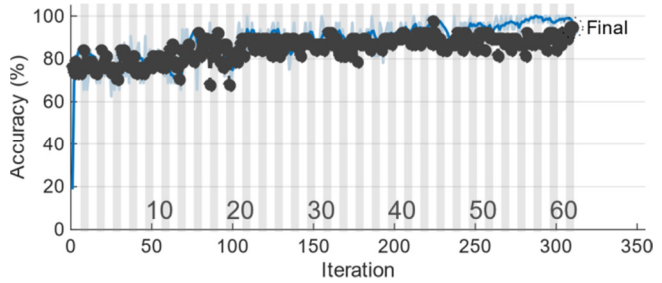

Fig. 11. Training (blue) and validation (black) accuracy graph for the SE-ResNet-GRU after transfer learning.

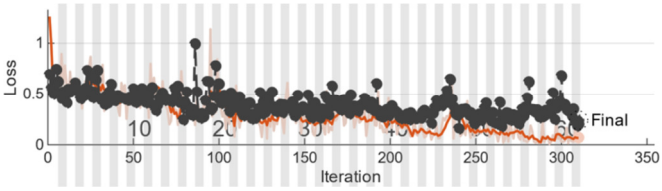

Fig. 12. Training (blue) and validation (black) accuracy graph for the SE-ResNet-GRU after transfer learning.

Despite the lower accuracy compared to pure SE-ResNet, the SE-ResNet-GRU model exhibited slightly better performance than the SE-ResNet-LSTM, suggesting that GRU's simplified gating mechanism helped mitigate some of the overfitting or ineffective learning observed with LSTM. However, the overall performance difference remained marginal. Figure 13 shows the confusion matrices of the training, validation, and test datasets.

D. Performance of the SE-ResNet-BiLSTM Model on the Simulated Dataset

The SE-ResNet-BiLSTM was also trained, validated, and tested on the simulated data. A training accuracy of 100% and validation accuracy of 94.5946% were achieved under the following hyperparameters: learning rate 0.0439, momentum 0.8449, L2 regularization 0.0045, 97 epochs, dropout probability 0.5, and a mini-batch size of 16. This performance was achieved within 36 min 42 s. Figure 16 shows the confusion matrix of the split dataset, whereas Figures 14 and 15 show the accuracy and loss graphs of the SE-ResNet-BiLSTM model. The test accuracy mirrored the other models at 97.297%.

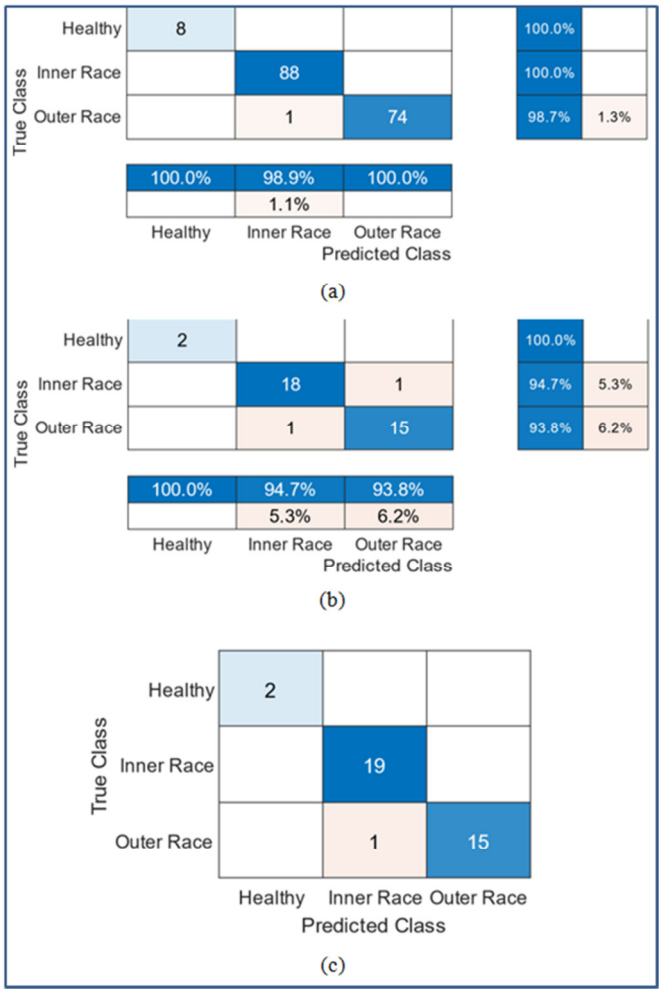

Fig. 13. Confusion matrices for (a) training, (b) validation, and (c) test datasets of the retrained SE-ResNet-GRU model.

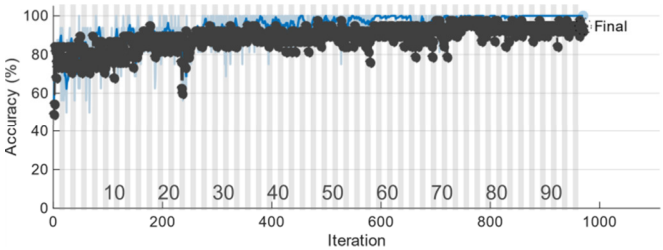

Fig. 14. Training (blue) and validation (black) accuracy graph for the SE-ResNet- BiLSTM after transfer learning.

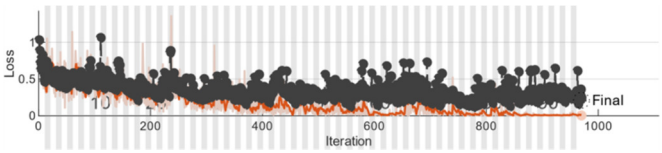

Fig. 15. Training (blue) and validation (black) accuracy graph for the SE-ResNet- BiLSTM after transfer learning.

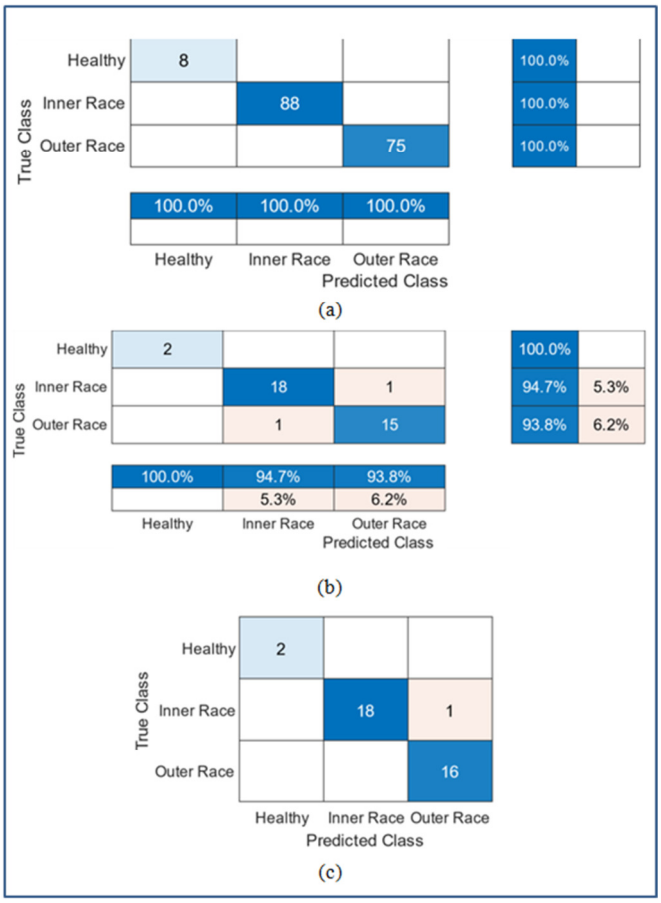

Fig. 16. Confusion matrices for (a) training, (b) validation, and (c) test datasets of the retrained SE-ResNet- BiLSTM model.

E. Performance of the SE-ResNet and Hybrid Modesl on the Open Source Dataset

The SE-ResNet and Hybrid Models had earlier been trained, validated, and tested using an open-source dataset [12]. This dataset was made up of healthy and faulty bearings, of which the faulty bearings were from both artificial damages and accelerated lifetime tests. Results showed that adding the RNN variants led to an improvement in performance. Table II shows the results obtained from training, validating, and testing the various models using the open-source dataset. However, when using simulated data, the addition of RNNs did not improve the performance of the SE-ResNet model and only the computational cost increased due to the increased complexity. The lack of improvement in accuracy suggests that the temporal features may not have been sufficiently different or relevant for the simulated data. It is evident that the additional complexity is not justified in the case of the simulated dataset.

TABLE II. PERFORMANCE METRICS FOR SE-ResNet ON THE OPEN SOURCE DATASET

Model	Training accuracy (%)	Validation accuracy (%)	Testing accuracy (%)
SE-ResNet	98.44	99.48	99.48
SE-ResNet-LSTM	100	99.74	99.74
SE-ResNet-GRU	100	100	99.74
SE-ResNet-BiLSTM	100	100	99.47

F. Performance Metrics of SE-ResNet and Hybrid Models

Precision, recall, and F1-score were also computed to evaluate the performance of the models. The results are shown in Tables III - VI. The high precision values indicate that the models have a strong ability to minimize false positives, accurately predicting true faults with minimal error. The SE-ResNet and SE-ResNet-GRU had the highest precision at 0.9833. This is shown in Tables II and IV. This value is very important in bearing fault diagnosis since the desired model should minimize false positives hence false alarms.

The recall values demonstrate the models' effectiveness in capturing almost all actual bearing faults thus minimizing false negatives. The SE-ResNet-BiLSTM had the highest recall value of 0.9825 as seen in Table V. The high F1-scores obtained in this study highlights the models' strong performance in classifying bearing faults, especially the SE-ResNet-GRU at 0.9809. The SE-ResNet-LSTM had the worst performance across all metrics compared to the other models.

Since the hybrid models increased the computational cost, the marginal improvement in certain metrics is not merited. The SE-ResNet is therefore considered to be the best option both in terms of accuracy and computational efficiency. The results demonstrate that the SE-ResNet model not only classifies faults with high accuracy but also maintains robust detection across different fault categories, ensuring that both precision and recall are well-balanced for practical predictive maintenance applications.

TABLE III. PERFORMANCE METRICS FOR SE-ResNet

Class	Precision	Recall	F1-score
Healthy	1.0000	1.0000	1.0000
Inner Race	0.9500	1.0000	0.9744
Outer Race	1.0000	0.9375	0.9677
Average	0.9833	0.9792	0.9807

TABLE IV. PERFORMANCE METRICS FOR SE-ResNet-LSTM

Class	Precision	Recall	F1-score
Healthy	1.0000	0.5000	0.6667
Inner Race	1.0000	1.0000	1.0000
Outer Race	0.9412	1.0000	0.9697
Average	0.9804	0.8333	0.8788

TABLE V. PERFORMANCE METRICS FOR SE-ResNet-GRU

Class	Precision	Recall	F1-score
Healthy	1.0000	1.0000	1.0000
Inner Race	0.9500	1.0000	0.9744
Outer Race	1.0000	0.9375	0.9677
Average	0.9833	0.9792	0.9807

TABLE VI. PERFORMANCE METRICS FOR SE-ResNet-BiLSTM

Class	Precision	Recall	F1-score
Healthy	1.0000	1.0000	1.0000
Inner Race	1.0000	0.9474	0.9730
Outer Race	0.9412	1.0000	0.9697
Average	0.9804	0.9825	0.9809

G. Performance Analysis

The pre-trained SE-ResNet achieved a test accuracy of 97.297% on the simulated data, demonstrating its ability to generalize to a different dataset. Similarly, the hybrid models, SE-ResNet-LSTM, SE-ResNet-GRU, and SE-ResNet-BiLSTM, maintained a test accuracy of 97.297% and did not show significant improvements on the other performance metrics. However, when using the open-source dataset, the hybrid models performed better than the SE-ResNet achieving a higher accuracy. Some of the reasons for this performance are:

- Domain Shift: The primary reason for the lack of improvement is likely the domain shift between the open-source dataset (real-world) and the simulated data. While the pre-trained SE-ResNet learnt robust features from real-world data, these features may not be optimally aligned with the characteristics of simulated signals. Simulated data often lacks the noise, variability, and complexity of real-world measurements.
- Simplicity of Simulated Data: Simulated data, while useful for controlled experiments, may lack the complex temporal dependencies present in real-world signals. The temporal patterns in simulated data might be simpler or more deterministic, making the RNN layers redundant or ineffective.
- Feature Saturation: The pre-trained SE-ResNet may have already extracted the most salient features from the simulated data. In this case, the RNN layers might not be able to learn additional relevant temporal patterns, leading to overfitting or negligible improvements.
- Lack of Real-World Noise: Real-world noise and variations present within the open-source dataset may have created a more robust feature extractor. The simulated data may be too clean, and the RNN portions of the model may be trying to find patterns that do not exist.

IV. CONCLUSION

This study presents a comprehensive approach to bearing fault detection in induction motors by leveraging simulated data, transfer learning, and Bayesian Optimization. By modeling bearing faults of varying severities using torque-based equations in MATLAB Simulink, and deriving acceleration as the vibration signal, a robust dataset was generated under different noise conditions. This method of modelling replicated real-world vibration physics more accurately, unlike prior works that relied on simplified frequency-domain simulations [21]. The transformation of these vibration signals into time-frequency domain images facilitated effective feature extraction by the SE-ResNet and the hybrid models.

Transfer learning, utilizing a pre-trained SE-ResNet originally developed for PMSMs, was employed to adapt the models to the simulated induction motor dataset. This cross-domain transfer learning demonstrated the versatility of the designed pre-trained model. Bayesian Optimization further enhanced the model by fine-tuning hyperparameters, resulting in a high fault classification accuracy of 97.297% with limited data. This result was higher than the 94% test accuracy achieved in [22], where the authors trained a one-dimensional CNN multiscale attention model using simulated bearing vibration signals.

The inclusion of mathematical modeling and detailed simulation procedures adds depth to the methodology, demonstrating the viability and effectiveness of simulation-based approaches in addressing data scarcity challenges in fault diagnosis. While a pre-trained SE-ResNet effectively generalized to the simulated induction motor bearing data, the integration of RNN layers did not yield significant improvements. The domain shift between real-world and simulated data, the simplicity of simulated signals, and potential overfitting were likely contributing factors. This research brings out the importance of model selection and its implications for efficient deployment in resource-constrained industrial environments.

The proposed method demonstrates high accuracy in simulated environments, offering a scalable and reliable solution for predictive maintenance, particularly in industrial environments where real-world fault data is limited. However, several avenues for future work remain, such as expanding the range of fault types, exploring real-time implementation strategies, incorporating more complex noise models, such as non-Gaussian noise to better simulate real-world conditions, and hybrid modeling approaches such as combining simulation-based data with limited real-world data.

This research contributes to the advancement of fault diagnosis techniques, providing a foundation for more resilient and efficient predictive maintenance systems in industrial applications.

ACKNOWLEDGMENT

This research was funded by the Pan African University Institute for Basic Sciences, Technology, and Innovation

REFERENCES

[1] H. A. Toliyat, S. Nandi, S. Choi, and H. Meshgin-Kelk, *Electric Machines: Modeling, Condition Monitoring, and Fault Diagnosis*, 1st ed. CRC Press, 2012.

[2] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, May 2023, Art. no. 91, https://doi.org/10.3390/computers12050091.

[3] P. Ilius, M. Almuhaini, M. Javaid, and M. Abido, "A Machine Learning–Based Approach for Fault Detection in Power Systems," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11216–11221, Aug. 2023, https://doi.org/10.48084/etasr.5995.

[4] M. Hakim, A. A. B. Omran, A. N. Ahmed, M. Al-Waily, and A. Abdellatif, "A systematic review of rolling bearing fault diagnoses based

on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations," *Ain Shams Engineering Journal*, vol. 14, no. 4, Apr. 2023, Art. no. 101945, https://doi.org/10.1016/j.asej.2022.101945.

[5] E. A. Burda, G. V. Zusman, I. S. Kudryavtseva, and A. P. Naumenko, "An Overview of Vibration Analysis Techniques for the Fault Diagnostics of Rolling Bearings in Machinery," *Shock and Vibration*, vol. 2022, no. 1, 2022, Art. no. 6136231, https://doi.org/10.1155/2022/6136231.

[6] Y. Zhang, Z. Ren, and S. Zhou, "A New Deep Convolutional Domain Adaptation Network for Bearing Fault Diagnosis under Different Working Conditions," *Shock and Vibration*, vol. 2020, no. 1, 2020, Art. no. 8850976, https://doi.org/10.1155/2020/8850976.

[7] H. Habbouche, Y. Amirat, T. Benkedjouh, and M. Benbouzid, "Bearing Fault Event-Triggered Diagnosis Using a Variational Mode Decomposition-Based Machine Learning Approach," *IEEE Transactions on Energy Conversion*, vol. 37, no. 1, pp. 466–474, Mar. 2022, https://doi.org/10.1109/TEC.2021.3085909.

[8] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, Jan. 2019, https://doi.org/10.1016/j.cogsys.2018.03.002.

[9] H. Zhu, Z. Sui, J. Xu, and Y. Lan, "Fault Diagnosis of Mechanical Rolling Bearings Using a Convolutional Neural Network–Gated Recurrent Unit Method with Envelope Analysis and Adaptive Mean Filtering," *Processes*, vol. 12, no. 12, Dec. 2024, Art. no. 2845, https://doi.org/10.3390/pr12122845.

[10] G. G. V. Díaz, J. C. C. Roque, and R. R. S. Torres, "Improving Healthcare Accessibility in Arequipa, Peru: Designing a Low-Cost IoT Vital Signs Monitoring System," *International Journal of Electrical and Electronics Engineering*, vol. 11, Jul. 2024, https://doi.org/10.14445/23488379/IJEEE-V11I7P104.

[11] K. Wang, W. Zhao, A. Xu, P. Zeng, and S. Yang, "One-Dimensional Multi-Scale Domain Adaptive Network for Bearing-Fault Diagnosis under Varying Working Conditions," *Sensors*, vol. 20, no. 21, Jan. 2020, Art. no. 6039, https://doi.org/10.3390/s20216039.

[12] X. Gong, K. Feng, W. Du, B. Li, and H. Fei, "An imbalance multi-faults data transfer learning diagnosis method based on finite element simulation optimization model of rolling bearing," *Proceedings of the Institution of Mechanical Engineers, Part C*, vol. 238, no. 17, pp. 8924–8940, Sep. 2024, https://doi.org/10.1177/09544062241245826.

[13] S. Djaballah, K. Meftah, K. Khelil, and M. Sayadi, "Deep Transfer Learning for Bearing Fault Diagnosis using CWT Time–Frequency Images and Convolutional Neural Networks," *Journal of Failure Analysis and Prevention*, vol. 23, no. 3, pp. 1046–1058, Jun. 2023, https://doi.org/10.1007/s11668-023-01645-4.

[14] K. Yu, T. R. Lin, H. Ma, X. Li, and X. Li, "A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning," *Mechanical Systems and Signal Processing*, vol. 146, Jan. 2021, Art. no. 107043, https://doi.org/10.1016/j.ymssp.2020.107043.

[15] S. Kim, D. An, and J.-H. Choi, "Diagnostics 101: A Tutorial for Fault Diagnostics of Rolling Element Bearing Using Envelope Analysis in MATLAB," *Applied Sciences*, vol. 10, no. 20, Jan. 2020, Art. no. 7302, https://doi.org/10.3390/app10207302.

[16] "WEG Electric Motors," *Quantum Controls*. https://www.quantum-controls.co.uk/electric-motors/weg-motors/.

[17] P. F. Le Roux and M. K. Ngwenyama, "Static and Dynamic Simulation of an Induction Motor Using Matlab/Simulink," *Energies*, vol. 15, no. 10, Jan. 2022, Art. no. 3564, https://doi.org/10.3390/en15103564.

[18] S. Salnikov, E. Solodkiy, and D. Vishnyakov, "Simulation of three-phase induction motor different bearing faults in Matlab Simulink environment," *Journal of Physics: Conference Series*, vol. 1886, no. 1, Dec. 2021, Art. no. 012009, https://doi.org/10.1088/1742-6596/1886/1/012009.

[19] A. El Idrissi *et al.*, "Bearing Fault Diagnosis for an Induction Motor Controlled by an Artificial Neural Network—Direct Torque Control Using the Hilbert Transform," *Mathematics*, vol. 10, no. 22, Jan. 2022, Art. no. 4258, https://doi.org/10.3390/math10224258.

[20] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification," in *European Conference of the Prognostics and Health Management Society 2016*, 2016.

[21] S. Shen *et al.*, "A physics-informed deep learning approach for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol. 103, Aug. 2021, Art. no. 104295, https://doi.org/10.1016/j.engappai.2021.104295.

[22] Y. Chen, J. Shi, J. Hu, C. Shen, W. Huang, and Z. Zhu, "Simulation data driven time–frequency fusion 1D convolutional neural network with multiscale attention for bearing fault diagnosis," *Measurement Science and Technology*, vol. 36, no. 3, Oct. 2025, Art. no. 035109, https://doi.org/10.1088/1361-6501/adb329.