Performance Analysis of Hybrid CNN–Random Forest Model for Fault Diagnosis in Domestic Generator Sets Using Continuous Wavelet Transform Features

[1]Unwana I. Ibanga, [2]Emmanuel O. Ogungbemi, and [3]Ekerette B. Ibanga
[1,2,3,] Department of Electrical and Electronics Engineering, University of Uyo, Nigeria
unwanaiwok@uniuyo.edu.ng | ogungbemioluropo@uniuyo.edu.ng | ebibanga@gmail.com

Received: 25-MARCH-2026; Reviewed: 26-MAY-2026; Accepted: 10-AUGUST-2026
https://dx.doi.org/10.46792/fuoyejet.v11i3.1553

ORIGINAL RESEARCH

Abstract— Reliable and accurate fault diagnosis of domestic power generators is essential for improving operational reliability and reducing maintenance costs, particularly in resource-constrained environments. This study proposes a hybrid acoustic fault-diagnosis framework that integrates Continuous Wavelet Transform (CWT)-based time–frequency analysis, a Convolutional Neural Network (CNN) for deep feature extraction, and a Random Forest (RF) classifier for fault classification. Acoustic signals acquired from 25 domestic 5 kVA generator sets, totalling 15,023 samples and representing 5 operating conditions (Normal, Carburettor fault, Exhaust fault, Valve fault, and Plug faults), were transformed using CWT into 2-D spectrograms and used to develop and evaluate the framework. The model was trained, tested, and validated using 70%, 20%, and 10% of the data, respectively. Experimental results yielded an overall classification accuracy of 95.95%, with a precision of 95.99%, a recall of 95.95%, and an F1-score of 95.95%. The per-class evaluation showed consistently high diagnostic performance. The results demonstrate the potential of using CWT features with a hybrid CNN-RF model to provide a robust, non-intrusive, and practically deployable diagnostic approach for resource-constrained domestic power-generation settings.

Keywords: Acoustic Signal Analysis, CNN–Random Forest Hybrid, Continuous Wavelet Transform (CWT), Generator Fault Diagnosis, Predictive Maintenance.

1 INTRODUCTION

Unreliable electricity supply remains a persistent challenge in Nigeria, where households and small businesses routinely depend on domestic generator sets for backup power. Grid deficiencies ranging from prolonged outages to voltage instability have made self-generation essential.

Nationally, representative household survey data show that 24.6% of Nigerian households own a generator, with 11.3% relying on one as their primary electricity source and 12.4% as a secondary source (Odeyale, et al., 2023). At the business level, a survey of small and medium enterprises found that fewer than half (49.2%) depend on the grid at all, with the remainder needing self-generated power to cope with supply that is frequently inconsistent (49.3%), inadequate (35.4%), or limited to just 4–5 hours a day (38.6%) (Oshio et al., 2024). This entrenched reliance on self-generated power, and the infrastructural failures driving it, is echoed in Nigeria's broader energy poverty landscape (Obileke, 2026). With the widespread use of domestic generators, they are prone to faults such as spark-plug degradation, carburettor malfunction, exhaust blockage, and valve leakage. If undetected, these faults reduce efficiency, increase fuel consumption, and lead to costly downtime. Current diagnostic practices are largely manual, relying on technicians' experience and visual inspection, which are often slow and error-prone. This limitation underscores the need for automated, reliable, and scalable fault-diagnosis methods, particularly for small-scale systems where existing industrial solutions are not readily applicable.

This work proposes a hybrid Convolutional Neural Network–Random Forest (CNN–RF) framework for non-intrusive fault diagnosis in domestic generator sets using acoustic signals. Generator audio is transformed into time–frequency spectrograms using the Continuous Wavelet Transform (CWT), enabling a CNN to learn discriminative features, while an RF classifier performs the final classification. This combination leverages the complementary strengths of both approaches to improve performance, particularly for subtle or overlapping fault types.

Before this, hybrid CNN–RF and related deep-learning/ensemble studies have focused on industrial, grid-scale, or laboratory-controlled equipment using electrical, vibration, or sensor signals. Small-scale, non-intrusive, acoustically monitored domestic generators remain comparatively unexplored. Relative to this literature, the present work makes the following distinct contributions:

1. a first empirical acoustic-fault dataset for 5 kVA domestic generators, covering 5 operating conditions across 25 physical units, addressing the current scarcity of public domestic-generator acoustic data;
2. a non-intrusive, low-cost diagnostic pipeline that requires no electrical or vibration instrumentation, in contrast to most prior CNN–RF hybrid studies;
3. an adaptation and evaluation of the CWT–CNN–RF hybrid architecture, previously validated mainly on power-electronic and industrial-machinery signals,

Corresponding Author
Section B- ELECTRICAL/COMPUTER ENGINEERING & RELATED SCIENCES
Can be cited as:
Ibanga U. I, Ogungbemi E. O., and Ibanga E. B. (2026). Performance Analysis of Hybrid CNN–Random Forest Model for Fault Diagnosis in Domestic Generator Sets Using Continuous Wavelet Transform Features. FUOYE Journal of Engineering and Technology (FUOYE**JET), 11(3), 384-391.**
https://dx.doi.org/10.46792/fuoyejet.v11i3.1553

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)

to small-scale, engine-driven domestic generators; and

4. a contextualised comparative benchmarking framework that explicitly distinguishes acoustic non-intrusive diagnosis from higher-accuracy but more intrusive or controlled-environment approaches.

This study provides a practical, non-intrusive approach to generator fault diagnosis, with the potential to reduce maintenance costs and improve operational reliability in resource-constrained settings.

2 RELATED WORKS

Recent advances in data-driven fault diagnosis show that integrating machine learning with signal-processing techniques can significantly enhance detection accuracy. Ensemble methods, such as RF, provide robustness through aggregated decision-making (Aldrich & Auret, 2010), (Fezai, et al., 2020), while CNNs enable effective feature extraction from structured inputs. Time–frequency transformation techniques, particularly the CWT, have been widely adopted to convert one-dimensional signals into informative spectrograms suitable for deep-learning models (Sun et al., 2022). Although very high classification accuracies have been reported in controlled environments (Lupea & Lupea, 2025), model performance often degrades under real-world operating conditions.

To address this limitation, recent studies have increasingly explored hybrid frameworks that combine deep learning, feature engineering, and ensemble learning. Such approaches have demonstrated improved robustness in bearing-fault diagnosis (You et al., 2022), (Najdi et al., 2024), sensor-fault detection (Sun, et al., 2020), and electrical-machine monitoring (Paraskevopoulos et al., 2023). These methods have leveraged the complementary strengths of deep models for feature learning and classical models for stable decision-making to achieve better generalisation.

Power-system applications offer some of the most explicit evidence for what these hybrids can do. Ahmad et al. (2023), for instance, paired CNN-based feature extraction with RF classification for real-time transient stability assessment from phasor measurement unit data — an application where fault conditions vary considerably and a model needs to hold up across that variation, not just perform well on average. Anwar et al. (2025) took a different combination, folding RF, K-Nearest Neighbours (KNN), and Long Short Term Memory(LSTM) networks into a single ensemble for transmission-line fault detection. The resulting 99.96% accuracy is more notable for what it suggests about ensembling as a general-purpose reliability strategy than for the number itself.

A recent parallel study showcased the combination of CNN-LSTM rather than CNN-RF. Alhanaf et al. (2024) built a CNN–LSTM framework for smart-grid fault detection that captures spatial patterns through the CNN and temporal dependencies through the LSTM component — a division of labour that turns out to matter more than raw model capacity, since faults in power systems are inherently both spatial and time-dependent events. Alsumaidaee et al. (2023) applied broadly the same logic to medium-voltage switchgear using a 1D-CNN–LSTM architecture, reporting near-perfect classification across several fault categories. Together, these studies make a reasonably strong case that sequence-aware architectures are well matched to electrical fault signals specifically.

This hybrid strategy has been successfully adopted in many other areas. Sun et al. (2020) reached 100% accuracy on hydrogen-sensor fault diagnosis with a CNN–RF model, even under noisy input conditions — a result that is easier to credit given how the RF stage was explicitly used to absorb some of the CNN's sensitivity to noise rather than relying on the CNN alone. Ahsan et al (2024) worked in a different domain again, rotating machinery, where their CWT–LeNet-5–LSTM framework held up consistently well across a range of load conditions. What these two studies share, despite the very different equipment involved, is the same underlying speculation that a fixed feature extractor followed by a more adaptable classifier generalises better than either component would alone.

Overall, the literature reveals three key trends: (1) the growing use of time–frequency representations, such as the CWT, for feature extraction; (2) the effectiveness of CNNs in learning discriminative features from transformed signals; and (3) the increasing adoption of hybrid and ensemble models to improve robustness and generalisation. However, most existing studies focus on large-scale industrial systems or controlled experimental settings, and the application of these techniques to small-scale, non-intrusive environments — such as domestic generator sets remains limited. This gap motivates the present study, which extends existing approaches by combining CWT-based feature representation with a CNN–RF hybrid model for acoustic-based fault diagnosis in domestic generator systems.

3 METHODOLOGY

The proposed hybrid model combines a CNN for feature extraction with an RF classifier, while CWT is used to transform the input audio into 2-D spectrograms. This RF–CNN–CWT framework is well suited to noisy and variable operating conditions, as reported by Ahmad *et al*. (2023). This robustness is particularly relevant because the audio recordings from the generator sets were acquired under non-uniform recording conditions.

The development pipeline begins with feature extraction, where the CWT transforms the generator's acoustic fault signals into time–frequency spectrograms that serve as input to the CNN. The CNN then automatically learns discriminative patterns from the spectrograms and generates class-probability outputs, eliminating the need for manual feature engineering. These probability

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)

http://dx.doi.org/10.46792/fuoyejet.v11i3.1553 engineering.fuoye.edu.ng/journal

features are subsequently passed to the RF classifier, which performs the final fault classification. This hybrid approach is particularly effective for distinguishing fault classes with similar acoustic characteristics, such as Plug, Carburettor, and Valve faults. Model development was implemented in Python 3.11 using TensorFlow/Keras for image preprocessing, scikit-learn for the RF classifier and performance evaluation, NumPy for numerical computations, and pickle for loading the pre-trained CNN model. Figure 1(a) illustrates the overall methodology, while Figure 1(b) shows that the dataset was split into training, validation, and testing sets before any feature extraction, windowing, or segmentation, ensuring no information leaked between data partitions.

3.1 Data Acquisition

This stage involved data recording, data collation, labelling, and organisation. Audio was recorded from 25 5kVA generator sets under normal (healthy) and various fault conditions, including exhaust, plug, valve, and carburettor faults. Recording used a consistent sampling rate of 44.1 kHz, a consistent microphone distance of 30–50 cm, with minimal background noise. Multiple sessions per fault class were recorded to ensure variability and generalisation, and a smartphone with a high-quality microphone was used to simplify setup.

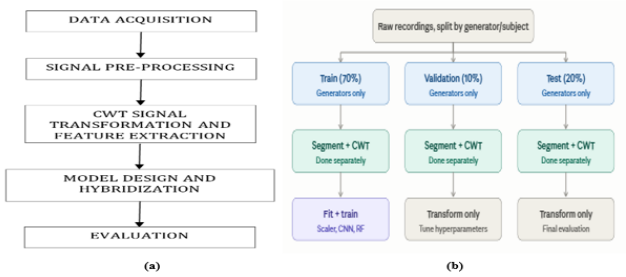

Fig. 1. Methodology pipeline

The different fault audio files were stored in .wav format, and the number of segments obtained per class was: Carburettor (3,108), Exhaust (3,108), Valve (5,699), Plug (2,849), and Normal (259), giving a total of 15,023 segments after the 100ms segmentation with 30% window overlap. The collected generator faults are internal, relating to corrosion, wear and tear, and/or blockages arising from the engine itself rather than from external parts or electrical-component failure.

It should be noted that this class distribution is markedly imbalanced: the Normal class (259 segments) is under-represented relative to the Valve class (5,699 segments), an imbalance ratio of approximately 1:22. This imbalance is a known characteristic of real-world fault datasets, where healthy operation is recorded far less frequently than fault conditions, and it has direct implications for the interpretation of aggregate performance metrics and for the comparatively lower recall obtained for the Normal class (Section 4.2). This limitation, and how it should be mitigated in future iterations of this work (e.g., class weighting, SMOTE-based oversampling, or targeted additional data collection for the Normal class), is discussed further in Section 5.

3.2 SIGNAL PRE-PROCESSING

The real-world data collected came from multiple sources (25 generator sets), so noise, missing values, or irrelevant features can bias the model, mislead it, or reduce prediction accuracy. This pre-processing pipeline was designed to create a robust foundation for reliable fault diagnosis by producing clean, structured data from the raw recordings. All audio data passed to the wavelet transformation and classification stages were made uniform, time-aligned, amplitude-normalised, and matched to the same sampling rate before feature extraction and model training. The pre-processing steps were:

- Audio recordings captured at different times were resampled to a uniform sampling rate of 44.1 kHz to ensure consistency across all files, regardless of their original sampling rates.
- LibROSA's *load* and *fix_length* functions were used to standardise the audio data. *fix_length* trims each clip to a fixed duration of 5 seconds, ensuring uniform input length, while *load* normalises the signal amplitude to the range [−1, 1], reducing variation due to differences in recording volume and supporting better model convergence.

Following pre-processing, the dataset was cleaned by removing duplicates, handling missing values, and correcting inconsistencies. The data was then split into training (70%), validation (10%), and test (20%) sets. We performed this split at the individual audio-segment level before sliding-window segmentation. Because segments from the same underlying recording session can appear across multiple partitions, and to avoid potential optimistic bias in the reported metrics, we also performed a generator-level (subject-independent) split in which no segment from a given generator unit appears in both the training and test sets.

To enhance data quality and capture temporal features for each audio recording, we segmented the data into 100ms segments with 30% overlap using a sliding-window approach in Python's LibROSA library to preserve transient information and ensure continuous signal coverage. This segmentation makes pseudo-stationary signals, enabling more effective analysis of time-varying features. It also better captures short-duration faults, increases the number of training samples, localises feature extraction, reduces overfitting, lowers computational cost, and improves generalisation to unseen data.

3.3 CWT FEATURE TRANSFORMATION

To enhance the detection of fault-related transients and improve sensitivity to localised anomalies, the CWT (Morlet wavelet) was applied to the 1-D raw acoustic signals as input, and the CWT was computed with PyWavelets (pywt.cwt) using scales from 1 to 127. This approach produced feature-rich time–frequency representations for the CNN-based analysis stage to capture both high- and low-frequency fault

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)

http://dx.doi.org/10.46792/fuoyejet.v11i3.1553 engineering.fuoye.edu.ng/journal

characteristics. We then computed the squared magnitudes of the CWT coefficients to obtain the power scalogram, highlighting the energy distribution across time and frequency. CWT-spectrogram images derived from the organised training and validation directories of the generator fault audio signals were loaded using Keras' ImageDataGenerator, which rescaled pixel values to [0, 1] and prepared scalogram batches of 224×224 images using OpenCV to match CNN input requirements. They were saved in .jpeg format using Matplotlib for efficient loading and batch processing during training.

3.4 MODEL DEVELOPMENT AND HYBRIDIZATION

The proposed model adopts a hybrid architecture combining a CNN-based feature extractor with an RF classifier. The CNN passes learned features to the RF, leveraging the CNN's ability to automatically extract patterns and the RF's robustness and interpretability for classification, to improve performance under real-world conditions. The overall architecture of the RF–CNN model with CWT input, shown in Figure 2, comprises three convolutional blocks that hierarchically extract spatial features. Each block includes a convolutional layer with ReLU activation to identify localised patterns, with filter counts progressively increasing from 32 (first block) to 64 (second block) to 128 (final block), enabling the model to capture increasingly intricate hierarchical patterns.

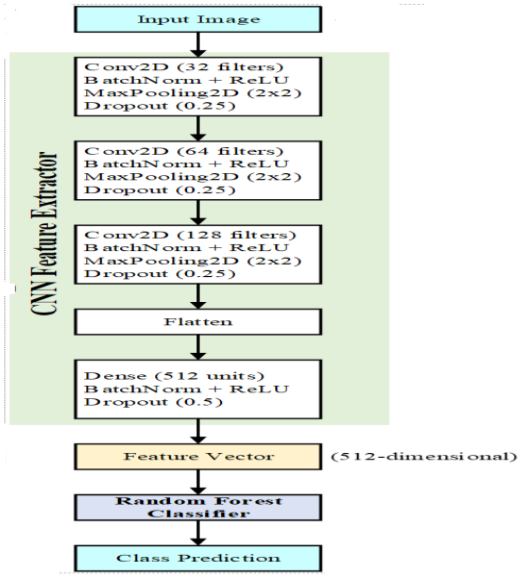

Fig. 2. Architecture for the RF-CNN with CWT

The final block output is flattened into a one-dimensional feature vector and passed through a dense layer with 512 units and ReLU activation, followed by batch normalisation (to stabilise training by standardising layer outputs) and max pooling (to reduce spatial dimensionality while preserving critical features). Dropout layers mitigate overfitting by stochastically deactivating neurons during training; batch normalisation and dropout are reapplied to improve generalisation, producing a refined 512-dimensional feature vector passed to the RF classifier.

The authors pre-trained the CNN in an earlier study using the same underlying CWT-spectrogram dataset (Ibanga et al., 2025), and reused its class-probability output here as the fixed feature extractor for the RF classifier. In this work, we loaded the pre-trained model (cnn_cwt_model.pkl) via Python's pickle module rather than retraining it. Ibanga et al. (2025) report the full CNN architecture and training details, including the loss function, optimiser, learning rate, and number of training epochs. The CNN feature-extractor architecture was developed using a custom Keras/TensorFlow Sequential CNN. It was designed for the five fault classes with an input 224 × 224 × 3 image, while the hybrid CWT CNN for this study was loaded from a pickle file:

model_cwt = pickle.load(open('cnn_cwt_model.pkl', 'rb')).

The architecture configuration is presented in Table 1.

Table 1. CNN development specifications

Layer	Configuration
Conv2D	32 filters, 3 x 3 Kernel ReLU
Batch Normalization	Applied
Max Pooling	2x2
Dropout	0.25
Conv2D	64 filters, 3 x 3 Kernel ReLU
Batch Normalization	Applied
Max Pooling	2x2
Dropout	0.25
Conv2D	128 filters, 3 x 3 Kernel ReLU
Batch Normalization	Applied
Max Pooling	2x2
Dropout	0.25
Flatten	--
Dense	512 Neurons, ReLU
Batch Normalization	Applied
Dropout	0.5
Output Dense	5 Neurons, Softmax

In the five-class categorical classification problem of the previous work (Ibanga et al 2025), CNN was trained using categorical cross-entropy with a softmax output layer:

loss = categorical_crossentropy;

The loss function is explicitly specified in the CNN compilation step mathematically as shown in Equation (1), the loss is:

$$L = -\sum_{c=1}^{5} y_c \log(\hat{y}_c) \quad (1)$$

where (y_c) is the true one-hot encoded class and $(\hat{y}_c)$ is the CNN's predicted probability for class (c). The Adam optimiser was used with a learning rate of $(\eta = 0.0001 = 10^{-4})$

A snippet of the training call was:

model.fit(train_generator, steps_per_epoch=len(train_generator), epochs=20,
validation_data=validation_generator,
validation_steps=len(validation_generator))

The CNN training parameters are summarized in Table 2.

Table 2. Training procedure

Training Parameters	Value
Epochs	20
Batch Size	32
Optimizer	Adam
Learning Rate	10^{-4}
Loss	Categorical Cross-entropy
Input	224 x 224 x 3
No of Classes	5
Image normalization	Pixel values x 1/255
Validation	Separate Validation Generator
Steps per epoch	len(train_generator)
Validation steps	len(validation_generator)

To achieve the CNN–RF hybrid, the CNN was used as a probabilistic feature extractor such that for every CWT

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)

http://dx.doi.org/10.46792/fuoyejet.v11i3.1553 engineering.fuoye.edu.ng/journal

image, the CNN generated a vector of class probabilities using: model.predict(x);

The Continuous Wavelet Transform (CWT) transformed the acoustic signals into 2D time–frequency representations over scales 1–127. The spectrogram images were then resized to 224 × 224 × 3, normalised by rescaling pixel intensities by 1/255 and supplied in batches of 32 for CNN processing. The resulting probability vectors became the input features for the Random Forest.

The contribution of the present study lies specifically in the design, tuning, and evaluation of the RF classification stage built on top of this fixed, previously validated CNN feature extractor, and in the associated per-class and comparative performance analysis (Section 4). The CNN predicted class probabilities for each image (soft predictions), which served as deep-learned features. To prevent statistical data leakage from fitted transforms on the full dataset from influencing the model before splitting, the CNN probabilities were standardised using StandardScaler before RF training; the scaler was fit only on X_train_prob, and transform () was applied to the validation and test data.

The Random Forest classifier was implemented using the RandomForestClassifier class in scikit-learn and trained on the CNN-derived probability features (X_train_prob) with the true class labels (y_train), allowing the RF to exploit the nonlinear decision boundaries present in the CNN output space to improve classification performance and robustness. For each of the five possible fault classes (Normal, Carburettor, Exhaust, Valve, Plug), the algorithm builds an ensemble of decision trees, where for each tree it - (1) draws a bootstrap sample of the CNN-derived probability features; (2) at each node, randomly selects a subset of these features and the gini index - RandomForestClassifier(criterion='gini') was used as the impurity measure for node splitting; and (3) repeats this process until the stopping conditions below is reached. The final predicted class is obtained by majority vote across all trees.

Hyperparameter tuning for the Random Forest stage was performed using GridSearchCV with cross-validation, searching over the number of estimators, maximum tree depth, and minimum samples per leaf/split. The configuration achieving the highest cross-validated accuracy on the training data was retained for final evaluation. The final selected hyperparameter values tuned using GridSearchCV with 10-fold cross-validation over the grid — n_estimators, max_depth, min_samples_leaf — returned by GridSearchCV, for full reproducibility, are shown in the code snippet below:

```
# RF Hyperparameter Grid
param_grid ={ 'n_estimators':[100, 150],'max_depth':[10,
20],'min_samples_split': [5, 10],'min_samples_leaf': [2, 4]}

# GridSearchCV Usage
grid_search = GridSearchCV(estimator=classifier, param_grid=param_grid,
cv=10, n_jobs=-1, verbose=3 )

# Best Parameter Extraction
best_params = grid_search.best_params_
```

3.5 MODEL EVALUATION

This hybrid architecture leverages the CNN's strength in learning rich feature representations from CWT images and the RF's robustness in handling high-dimensional feature inputs while reducing overfitting. Cross-validation (via sklearn.model_selection's train_test_split, cross_val_score, and GridSearchCV) was applied to the RF classifier trained on the CNN-extracted features to assess robustness and generalisation.

The trained model was used to predict fault classes on the validation and test data, and standard evaluation metrics — accuracy, precision, recall, and F1-score — were computed, together with a full confusion matrix and per-class breakdown (Section 4.2), to assess the hybrid model's performance in a manner appropriate for an imbalanced multi-class dataset.

4 RESULTS AND DISCUSSION

4.1 WAVEFORMS OF GENERATOR AUDIO SIGNALS AND SPECTROGRAMS

Figures 3 and 4 present the waveforms and spectrograms of the recorded generator audio signals, respectively. The waveforms in Figure 3 visualise the recorded generator fault audio signals, showing the temporal characteristics, amplitude variations, and frequency patterns associated with different fault conditions. Figure 4 shows the corresponding CWT spectrograms, capturing frequency–time domain features of the fault audio signals and providing compact representations suitable for training deep-learning models. Visual inspection of the CWT spectrograms reveals distinguishing spectral-temporal signatures across fault types that plausibly explain the CNN's ability to learn discriminative features. Valve-fault spectrograms tend to show short, high-energy transients concentrated in higher-frequency bands, consistent with the metallic impact character of valve seating/leakage events, whereas Exhaust-fault spectrograms tend to show

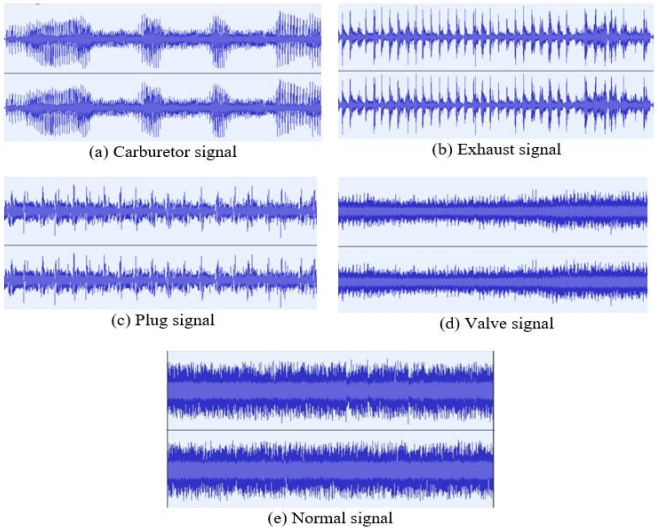

Fig. 3. Waveforms of fault audio signals recorded

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)

http://dx.doi.org/10.46792/fuoyejet.v11i3.1553 engineering.fuoye.edu.ng/journal

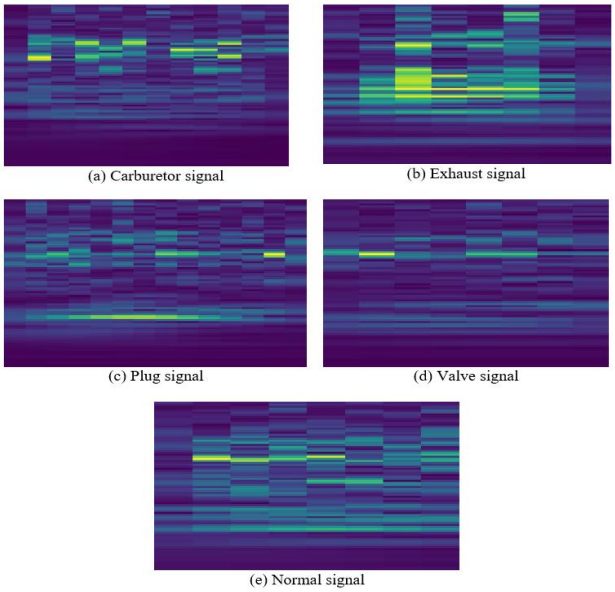

Fig. 4. CWT Spectrogram for CNN+RF Model

more sustained energy concentrated in lower-frequency bands, consistent with continuous gas-leakage and turbulent-flow noise. Plug- and Carburettor-related faults show intermediate, more diffuse energy patterns, which is also consistent with the greater confusion observed between these classes and Valve in the confusion matrix discussed below (Section 4.2).

4.2 PERFORMANCE EVALUATION OF RF-CNN HYBRID MODEL WITH CWT SPECTROGRAMS

The integrated framework — CNN feature extraction from CWT time–frequency representations, followed by RF classification — was assessed on the training, validation, and test datasets. The CNN's convolutional layers extracted discriminative spatial-spectral attributes from the transformed input signals, and the RF classifier performed final classification by aggregating predictions from multiple decision trees. Aggregate performance metrics are reported in Table 3.

Table 3. Aggregate Performance Metrics for the RF–CNN Hybrid Model with CWT Spectrograms (cf. Ibanga *et al.*, 2025)

Model	Accuracy	Precision	Recall	F1-Score
RF-CNN	95.95%	95.99%	95.95%	95.95%

The RF–CNN framework achieved an overall classification accuracy of 95.95%. A precision of 95.99% reflects a low false-positive rate, while a recall of 95.95% shows that most fault events were correctly identified with minimal false negatives. The F1-score of 95.95%, closely aligned with both precision and recall, confirms a well-balanced model. Given the class imbalance described in Section 3.1, however, these aggregate metrics can mask class-specific weaknesses. A confusion matrix and per-class breakdown are therefore presented below.

Figure 5 presents the confusion matrix obtained for the RF–CNN hybrid model on the test set, and Table 4 reports the corresponding per-class precision, recall, and F1-score, computed directly from the confusion matrix (support = number of true test samples per class).

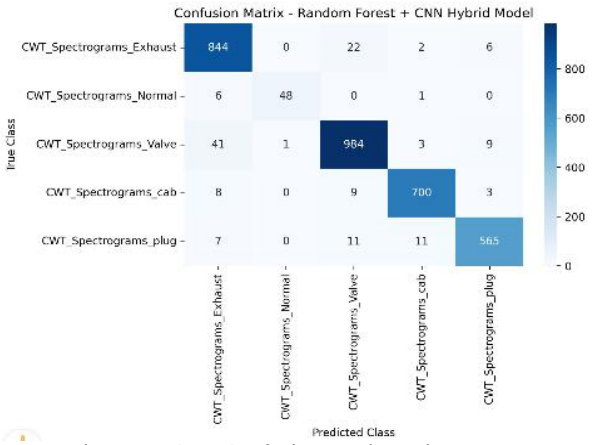

Fig. 5. RF-CNN Confusion matrix on the test set

Table 4. RF-CNN Hybrid Model Per-Class Performance

Fault Class	Support	TP	Precision	Recall	F1-Score
Exhaust	874	844	0.9316	0.9657	0.9483
Normal	55	48	0.9796	0.8727	0.9230
Valve	1038	984	0.9591	0.9480	0.9535
Carburettor (Cab)	720	700	0.9763	0.9722	0.9744
Plug	594	565	0.9691	0.9512	0.9602
Macro average	3281	3141	0.9631	0.9420	0.9519
Weighted average	3281	3141	0.9578	0.9573	0.9575

The per-class analysis presented here provides the more granular evidence needed to substantiate the claim that the hybrid model copes well with subtle, overlapping fault types rather than just asserting it. On that test, the Carburettor class comes out best (F1 = 0.9744), which fits with its spectrograms being visually the most distinct of the five. Valve and Plug are not far behind (F1 = 0.9535 and 0.9602). Normal is the outlier: recall drops to 0.8727, meaning roughly one in eight genuine Normal segments was misread, almost always as Exhaust. It would be tempting to read this as the model struggling to tell healthy operation from a specific fault signature, but given that Normal makes up only 259 of the 15,023 segments in the dataset, the simpler explanation is exposure — the classifier has simply seen comparatively little of what "normal" sounds like. Its precision remained high (0.9796), so at least the model is not carelessly guessing "Normal"; when it does predict the class, it is usually right. Where the model does slip, it slips in a fairly consistent direction: Exhaust and Valve are confused with each other more than any other pair (22 Exhaust segments called Valve, 41 Valve segments called Exhaust). That kind of two-way confusion usually points to something structural rather than random noise. In this case, probably overlapping vibration frequencies or a shared mechanical coupling path within the generator housing that makes the two fault signatures genuinely harder to tell apart acoustically. Carburettor and Plug show smaller, similarly mutual mix-ups with Valve. The one result worth flagging as encouraging is that only a single true fault sample, out of the entire Valve class, was ever

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)

http://dx.doi.org/10.46792/fuoyejet.v11i3.1553 engineering.fuoye.edu.ng/journal

mistaken for Normal. In a predictive-maintenance setting, that is the error you most want to avoid, since it is the one that leads to a missed fault rather than a false alarm. The Normal-class recall here is a fairly direct illustration of the consequence of imbalance in practice, and it is the class most in need of the mitigation steps proposed in Section 5.

4.3 PERFORMANCE COMPARATIVE ANALYSIS

A comparative analysis of this work with related hybrid models, based on feature type, application domain, and reported accuracy, is presented in Table 5. This comparison is intended to be contextual rather than a strict, like-for-like performance benchmark: the compared studies differ substantially in signal modality (acoustic vs. electrical vs. vibration), application scale (domestic vs. industrial/grid), and dataset characteristics, and several were evaluated on more tightly controlled or class-balanced datasets. Table 5 should therefore be read as situating this work within the broader hybrid CNN–RF/ensemble literature rather than as evidence of superiority or inferiority in an absolute sense.

Table 5. Comparative Analysis of the CNN_RF Hybrid Model and Related Works.

Hybrid Models	Feature type	Fault/Non-fault System	Accuracy	Source
CNN-RF	CWT spectrograms from acoustic signals	Domestic Power generator sets	95.95%	This work
CNN-RF	Electrical waveforms (voltage and current)	Monitoring of transient stability assessment in electric power systems	99.95%	(Ahmad, *et al*, 2023)
CNN-RF	2-D gray matrix images transformed from one-dimensional time-domain data of fault signals	Hydrogen gas sensor	100%	(Sun, *et al*., 2020)
RF-LSTM Tuned KNN	Using voltage and current patterns across transmission line phases	Fault detection and classification in power transmission lines	99.96%	(Anwar, *et al*., 2025)
CWT-LeNet-5-LSTM technique	CWT from Vibration signal recordings from a 2-horsepower (HP) induction motor fitted with a 6205-2RS JEM SKM deep groove ball bearing.	The vibration data from the Case Western Reserve University (CWRU), including vibration signal recordings from an experimental setup involving-In a 2-horsepower (HP) induction motor fitted with a 6205-2RS JEM SKM deep groove ball bearing,	92.5% - 99.6% for different loads	(Ahsan *et al*, 2024)
CNN-LSTM	1D current and voltage inputs across all lines within the test network	Fault Detection and Classification in Ring Power System in IEEE 9-bus	99.95%	(Alhanaf,*et al*, 2024)
CNN-LSTM.	1D- audio data in both the time domain (TD) and frequency domain (FD).	Medium voltage (MV) switchgear in power systems	100%	(Alsumaidaee *et al*., 2023)
Discrete wavelet–CNN	3-phase currents	Various stator winding fault conditions in induction motors.		(Paraskevopoulos *et al*., 2023),

4.3.1 FEATURE TYPE COMPARISON

The clearest dividing line in Table 5 is between studies built on electrical measurements and the handful, including this one, built on physical signals such as vibration or sound. Ahmad et al. (2023), Anwar et al. (2025), Alhanaf et al. (2024), and Alsumaidaee et al. (2023) all draw their features from voltage and current waveforms — a sensible choice for grid-connected equipment already wired for monitoring, but one that presupposes exactly the kind of electrical instrumentation a domestic generator owner is unlikely to have. Ahsan et al. (2024) and Paraskevopoulos et al. (2023) instead work from vibration signals, which brings them slightly closer to the non-intrusive spirit of this study, though vibration sensors still require physical attachment to the machine. Sun et al. (2020) is the nearest methodological relative here, also converting 1-D signals into 2-D representations for CNN input, but for hydrogen sensors rather than acoustic engine faults. Set against this backdrop, the acoustic-CWT approach used in the present work is the only one that requires no contact with the machine at all — a distinction that matters more for domestic deployment than for raw classification accuracy.

4.3.2 APPLICATION/FAULT SYSTEM COMPARISON

This work focused on domestic generator sets, a relatively underexplored area in the literature. Other authors' work covered electric power transmission and distribution systems (Ahmad et al., 2023; Anwar et al., 2025; Alhanaf et al., 2024), rotating machinery and motors (Ahsan et al., 2024; Paraskevopoulos et al., 2023), hydrogen gas sensors (Sun et al., 2020), and medium-voltage switchgear (Alsumaidaee et al., 2023). Most prior work focuses on industrial-scale or grid-related equipment, whereas this work uniquely addresses small-scale, residential-use generators, filling a practical gap.

4.3.3 PERFORMANCE (ACCURACY) COMPARISON

Table 6. Accuracy Comparison (Contextual)

Model	Accuracy
This work (CNN_RF)	95.95%
CNN-RF (Ahmad *et al*., 2023)	99.95%
CNN-RF (Sun *et al*., 2020)	100%
RF-LSTM Tuned KNN	99.96%
CWT-LeNet-5-LSTM	92.5–99.6%
CNN-LSTM (Alhanaf *et al*., 2024)	99.95%
CNN-LSTM (Alsumaidaee *et al*., 2023)	100%

Although this work does not outperform all others in raw accuracy, 95.95% is a strong result given the non-intrusive nature of acoustic sensing and the severe class imbalance present in the underlying dataset (Section 3.1). Models reporting 100% accuracy in Table 6 were generally developed and evaluated under highly controlled, class-balanced, or grid-scale conditions, which are not directly comparable to the real-world, imbalanced, domestic-generator setting studied here.

5 CONCLUSION AND FUTURE WORK

This study developed and evaluated a hybrid fault-diagnosis framework for domestic power generators that integrates CWT-based time–frequency analysis with a CNN feature extractor and RF classifier. By combining the automated feature-learning capability of deep learning with the robust decision-making characteristics of ensemble machine learning, the proposed framework provides an effective approach for acoustic-based generator fault diagnosis without requiring handcrafted feature engineering. The approach addresses the growing need for accurate, reliable, and computationally efficient fault-monitoring solutions suitable for resource-constrained environments where uninterrupted power supply remains a critical challenge.

Experimental evaluation demonstrated that the proposed CNN–RF architecture achieved an overall classification accuracy of 95.95%, with corresponding precision, recall, and F1-score of 95.99%, 95.95%, and 95.95%, respectively.

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)

http://dx.doi.org/10.46792/fuoyejet.v11i3.1553 engineering.fuoye.edu.ng/journal

Per-class analysis further showed consistently high diagnostic performance across most fault categories, with the remaining misclassifications occurring primarily between the acoustically similar Exhaust and Valve faults. The comparatively lower recall for the Normal operating condition was largely attributable to the significant class imbalance in the dataset rather than limitations of the hybrid architecture itself. Collectively, these results demonstrate that the proposed framework is capable of accurately distinguishing generator operating conditions using only acoustic signatures, thereby providing a practical and non-invasive alternative to conventional sensor-based monitoring systems. Despite these encouraging results, several factors constrain the generalisability of the present study. The dataset exhibits substantial class imbalance, particularly for the Normal operating condition, which may influence performance on minority classes. In addition, the CNN feature extractor was adopted from a previously validated model without retraining, allowing the present work to focus specifically on evaluating the effectiveness of the RF classification stage. Furthermore, the experimental dataset comprises only 5 kVA domestic generators and a limited set of mechanical fault categories, while comparisons with related studies involve datasets of differing characteristics, preventing direct one-to-one benchmarking.

Future research should therefore focus on expanding the dataset to include a wider range of generator capacities, manufacturers, operating conditions, and fault types while improving class balance through additional data acquisition and advanced imbalance-learning strategies. Incorporating complementary sensing modalities such as vibration and temperature, investigating alternative time–frequency representations including Mel-Frequency Cepstral Coefficients (MFCC) and Short-Time Fourier Transform (STFT), and exploring feature-selection and dimensionality-reduction techniques may further enhance diagnostic performance and model interpretability. Equally important is the implementation and validation of the proposed framework on embedded and edge-computing platforms to evaluate computational latency, energy efficiency, and real-time operational capability under practical deployment conditions.Overall, the findings demonstrate that the proposed CNN–RF hybrid framework constitutes a robust, accurate, and computationally efficient solution for acoustic-based fault diagnosis of domestic power generators. Beyond its strong predictive performance, the framework provides a promising foundation for intelligent condition-monitoring systems that support predictive maintenance, reducing unplanned equipment downtime and improving the operational reliability of small-scale power generation systems in both developed and resource-constrained environments.

REFERENCES

Ahmad, H., Muhammad, Y., Maliha, S., & Najeeb, U. (2023). Monitoring transient stability in electric power systems using a hybrid model of convolutional neural network and random forest. *Electric Power Components and Systems, 52*(6), 946–958.

Ahsan, M., Muhammad, W. H., Jose, R., & Mohamed, A. (2024). Enhanced fault diagnosis in rotating machinery using a hybrid CWT-LeNet-5-LSTM model: Performance across various load conditions. *IEEE Access, 13*, 1026–1045.

Akindele, O. A., & Adejumobi, D. O. (2017). Domestic electric power generator usage and residents' liveability milieu in Ogbomoso, Nigeria. *Environmental Management and Sustainable Development, 6*(1), 91–104.

Aldrich, C., & Auret, L. (2010). Fault detection and diagnosis with random forest feature extraction and variable importance methods. *IFAC Proceedings Volumes, 43*(9), 79–86.

Alhanaf, A. S., Murtaza, F., & Hasan, H. B. (2024). Fault detection and classification in a ring power system with DG penetration using a hybrid CNN-LSTM. *IEEE Access, 12*, 59953–59975.

Alsumaidaee, Y. A., Johnny, K. S., Chong, T. Y., Sieh, K. T., Chai, P. C., Talal, Y., & Ahmed, N. A. (2023). Fault detection for medium voltage switchgear using a deep learning hybrid 1D-CNN-LSTM model. *IEEE Access, 11*, 97574–97589.

Anwar, T., Chaoxu, M., Muhammad, Z. Y., Wajid, K., Saqib, K., Ahmad, O. H., & Ievgen, Z. (2025). Robust fault detection and classification in power transmission lines via ensemble machine learning models. *Scientific Reports, 15*(1), 2549.

Fezai, R., Dhibi, K., Mansouri, M., Trabelsi, M., Hajji, M., Bouzrara, K., & Nounou, M. (2020). Effective random forest-based fault detection and diagnosis for wind energy conversion systems. *IEEE Sensors Journal, 21*(5), 6914–6921.

Ibanga, E., Udofia, K., Udofia, K., Iwok, U., & Ogungbemi, E. (2025). Machine learning-based fault diagnosis and characterization system for power-generating sets using customized fault audio dataset. *American Journal of Signal Processing, 13*(1), 1–16.

Lupea, I., & Lupea, M. (2025). Continuous wavelet transform and CNN for fault detection in a helical gearbox. *Applied Sciences, 15*(2), 950.

Najdi, B., Mohammed, B., & Mohammed, N. K. (2024). Bearing fault diagnosis with a hybrid CWT-ResNet-LSTM model. In *Proceedings of the International Conference on Digital Technologies and Applications* (pp. 454–463). Springer Nature Switzerland.

Obileke, K. (2026). Energy Poverty in Nigeria: A Review. *Sustainability*, *18*(5), 2387.

Odeyale, O. C., Oke, O. O., Falana, A. R., Adeoye, A. S., Ogunsola, J. O., & Marizu, J. T. (2023). Environmental pollution as health depreciator: The case of household generator use in Nigeria. *Global Journal of Pure and Applied Sciences, 29*(1), 11–17.

Oshio, L. E., Owenvbiugue, R. O., & Chukwuemeka, H. E. (2024). Assessment of the Impacts of Energy Poverty on Small and Medium Businesses in Benin City, Nigeria Journal of Economics and Business, 7(3) 10.31014/aior.1992.07.03.604

Paraskevopoulos, D., Spandonidis, C., & Giannopoulos, F. (2023). Hybrid wavelet-CNN fault diagnosis method for ships' power system. *Signal, 4*, 150–166.

Sun, Q., Yu, X., Li, H., Peng, F., & Sun, G. (2022). Fault detection for power electronic converters based on continuous wavelet transform and convolution neural network. *Journal of Intelligent & Fuzzy Systems, 42*(4), 3537–3549.

Sun, Y., Hongquan, Z., Tingting, Z., Zhihui, Z., Bin, S., & Lixin, Y. (2020). A new convolutional neural network with a random forest method for hydrogen sensor fault diagnosis. *IEEE Access, 8*, 85421–85430.

You, K., Qiu, G., & Gu, Y. (2022). Rolling bearing fault diagnosis using hybrid neural network with principal component analysis. *Sensors, 22*(21), 8161

© 2026 The Author(s). Published by Faculty of Engineering, Federal University Oye-Ekiti.
This is an open access article under the CC BY NC license. (https://creativecommons.org/licenses/by-nc/4.0/)