Dual-Branch Variational Autoencoder with Ensemble Learning for Fault Classification in Inverter-Based Resource-Dominated Power Systems

Gabrielly Rocha
Department of Electrical Engineering
Federal University of Vale do São Francisco
Juazeiro, Brazil
gabrielly.rocha@discente.univasf.edu.br

Shipra Tiwari
Department of Electrical and Computer Engineering
Michigan Technological University
Michigan, USA
shiprat@mtu.edu

Rodrigo P. Ramos
Department of Electrical Engineering
Federal University of Vale do São Francisco
Juazeiro, Brazil
rodrigo.ramos@univasf.edu.br

Flavio B. Costa
Department of Electrical and Computer Engineering
Michigan Technological University
Michigan, USA
fbcosta@mtu.edu

***Abstract*—The increasing integration of inverter-based resources (IBRs) introduces nonlinear dynamics that complicate fault diagnosis in modern power systems. Classical machine learning methods relying on handcrafted features struggle with complex temporal and spectral fault patterns. This paper presents a hybrid deep learning approach using dual-branch Variational Autoencoders (VAEs) to conjointly process time-domain and frequency-domain representations of voltage and current signals. Latent features from both branches are classified by a bidirectional long short-term memory (BiLSTM) recurrent neural network, and their outputs are fused using a stacking ensemble with AdaBoost as the meta-classifier. A comprehensive dataset was generated from a simulated 14-bus grid with diverse inverter controls and fault scenarios. Experimental results show that the proposed framework achieved an accuracy of 94.14% and F1-score of 93.74, indicating consistent performance and ability to capture nonlinear fault characteristics. These findings suggest enhanced reliability for fault classification in renewable-rich power grids.**

***Index Terms*—Fault diagnosis, inverter-based resources, deep learning, variational autoencoder, time–frequency analysis.**

I. Introduction

The proliferation of inverter-based resources (IBRs) in modern electrical grids has fundamentally altered the dynamic characteristics of fault events. Unlike conventional synchronous machines, power electronic interfaces exhibit nonlinear control behaviors that significantly affect current and voltage waveforms during disturbances [1]. As a result, fault classification in IBR-dominated systems demands advanced methods capable of learning complex, nonstationary patterns from multichannel measurements [2]–[4].

Traditional machine learning approaches, such as support vector machines, decision trees, and random forests [5]–[8], typically rely on handcrafted statistical or spectral features. Despite these methods can achieve reasonable accuracy under restricted operating conditions, their performance generally deteriorates when underlying system dynamics change or when inverter controls introduce nonlinear distortions. To address these challenges, deep learning models have emerged as a promising alternative, owing to their ability to automatically extract discriminative features directly from raw data.

Among deep models, the Variational Autoencoder (VAE) [9] has shown strong capability for learning compact latent representations that preserve essential characteristics of complex signals. However, relying on a single input domain may limit discriminative power, since certain fault signatures are more pronounced either in the waveform transient or in its spectral content. Prior studies indicate that combining time and frequency representations improves fault separability, particularly under conditions where inverter controls introduce nonlinear distortions [10]–[12]. In addition, previous works show that combining time- and frequency-domain features improves the discrimination of transient fault patterns, especially in scenarios with limited or imbalanced data.

To overcome these limitations, this study introduces a hybrid diagnostic framework that combines two complementary VAEs, one operating on time-frequency representations derived from continuous wavelet transforms (CWT) [12], and another trained on time-domain waveforms, together with an ensemble learning strategy. The latent features extracted by each VAE are used to train individual Bidirectional LSTM (BiLSTM) classifiers. Their probabilistic outputs are subsequently fused through a stacking meta-classifier using the AdaBoost approach [13]. This hybrid configuration leverages the representational power of deep generative models and the stability of ensemble methods to improve classification accuracy and robustness in inverter-dominated grids.

By incorporating both time-domain and CWT-based

979-8-3315-8138-1

frequency-domain features, the proposed approach captures a richer description of fault transients and demonstrates improved generalization compared to single-modality and conventional machine learning baselines. Quantitative results on a simulated fourteen-bus dataset with multiple fault types confirm the effectiveness of the method.

II. Data Acquisition and Preprocessing

The dataset was obtained from a detailed simulation of a 14-bus electrical grid containing multiple IBRs, developed in [14], including type IV wind turbine generators [15], grid-following inverters, and grid-forming converters, as well as synchronous generators [16], as illustrated in Fig. 1. The system topology represents a medium-voltage distribution network operating at $f = 50$ Hz. The signals were sampled with a sampling rate $f_s = 10$ kHz. High-fidelity models were employed to capture the complex dynamics and accuracy of fault signatures, including realistic transmission line parameters, inverter control strategies, and operational limits consistent with standard grid codes [15], [16].

Ten distinct types of electrical faults were simulated under various operating scenarios to capture the diversity of possible fault signatures in mixed-generation systems: three single line-to-ground faults (AG, BG, CG), three phase-to-phase faults (AB, BC, CA), three double line-to-ground faults (ABG, BCG, CAG), and one balanced three-phase fault (ABC). Each fault was applied with resistance values of 0.01 Ω, 1 Ω, 10 Ω, 20 Ω, and 50 Ω, and with fault inception angles ranging from 0° to 360° in 10° steps. The three-phase voltages $V_a(k)$, $V_b(k)$, $V_c(k)$ and currents $I_a(k)$, $I_b(k)$, $I_c(k)$ were acquired at bus 4 and used as input for feature extraction. In total, 1850 labeled fault samples were generated (185 simulations per fault type), ensuring a balanced dataset. The fault locations are shown as curved arrows in Fig. 1.

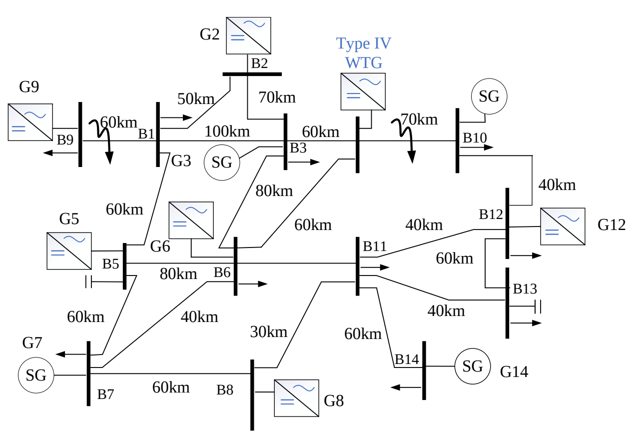

Fig. 1. IBR-dominated power grid.

To capture both transient and spectral characteristics of the measured electrical quantities, a one-cycle time segment of the signal following each fault occurrence was selected to extract relevant features, and a dual-domain feature extraction approach was implemented. The fault detection is beyond the scope of this study, and it can be addressed using any advanced techniques, such as the approach developed in [12]. Therefore, only fault cases were simulated and the beginnig of the fault is known.

a) Time-Domain Representation: The raw sequence of the six electrical signals, normalized to a range of [-1, 1], preserves temporal dependencies among all phases and allows the one-dimensional convolutional encoder to learn transient dynamics directly from the time series.

b) Time-Frequency Representation: Each channel was decomposed using a continuous Morse wavelet transform with 12 voices per octave and a sampling frequency of 10 kHz. The corresponding spectrogram was converted into a 224 $\times$ 224 grayscale, histogram-equalized image to enhance local contrast. The six resulting grayscale maps were concatenated to form a tensor of size 224×224×6, suitable for the two-dimensional encoder input of the VAE model.

III. Proposed Hybrid VAE and Stacking Architecture

The proposed multi-stage methodology is implemented in two distinct training phases: (A) a dual-branch base-classifier stage, where two independent VAE-BiLSTM branches are jointly trained to learn optimal representations and classify faults from the domain and frequency domains, and (B) a meta-learning stage, where an ensemble meta-model is trained to fuse the probabilistic outputs from both base-classifiers, producing a final, robust prediction. The complete architecture is shown in Fig. 2.

A. Phase 1: Dual-branch Base-Classifier

The first stage of the model focuses on learning a comprehensive representation of the input signals with two parallel branches that serve as the base-learners for the ensemble. Unlike conventional pipelines where feature extraction and classification are decoupled, the encoder of the VAE and the BiLSTM classifier are optimized together under a single composite loss function. The encoder maps the input signal to a latent vector $\mathbf{z}$ parameterized by a mean and variance, while the decoder reconstructs the original signal to preserve its informative structure. The BiLSTM consumes $\mathbf{z}$ directly to predict the fault class, allowing the latent space to be shaped for discrimination rather than reconstruction alone.

a) Encoder: The encoder transforms the input tensor $\mathbf{X}$ into the parameters of a diagonal Gaussian distribution, defined by the two vectors output from the final layer of the encoder network: a mean vector μ and a log-variance vector $\log \sigma^2$. For each input data point, $\mathbf{z}$ is sampled from a diagonal Gaussian distribution.

The encoder is composed of convolutional layers (2D for frequency-domain data and 1D for time-domain sequences) to extract hierarchical spatial or temporal features prior to projection into the latent distribution.

979-8-3315-8138-1

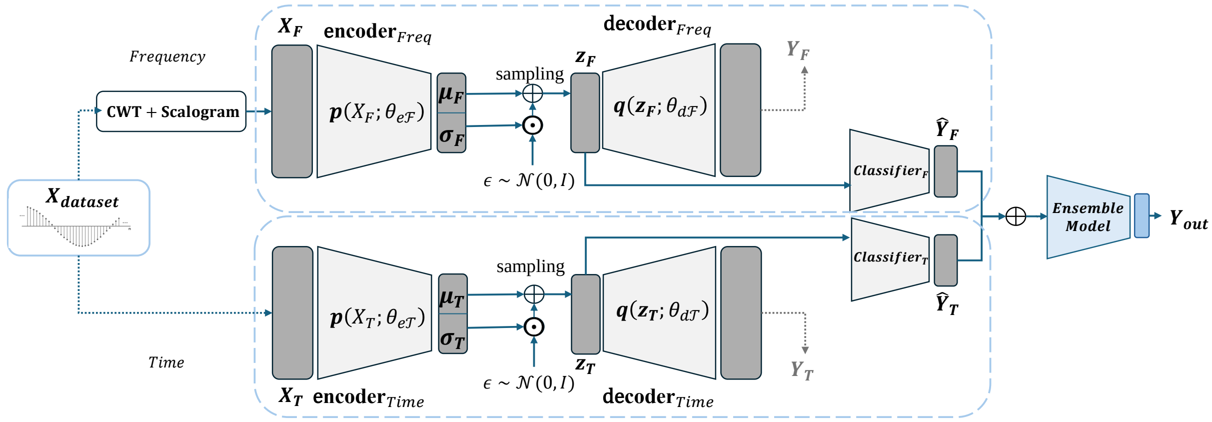

Fig. 2. Proposed VAE and meta-classifier architecture.

b) Latent Space and Reparameterization: To allow stochastic sampling while preserving differentiability, the latent representation is obtained using the "reparameterization trick" $z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$ [9]. This formulation allows gradients to flow through μ and σ during training while allowing the model to learn uncertainty-aware latent representations. The sampled latent vector $\mathbf{z}$ is then passed directly to the BiLSTM classifier.

c) Decoder: The decoder reconstructs an estimated tensor $\hat{\mathbf{X}}$ from the latent vector $\mathbf{z}$, forming a generative path that enforces latent consistency. It mirrors the encoder architecture using transposed convolutional layers to upsample back to the original domain, with a sigmoid activation function at the output layer to ensure valid data range.

d) Joint Loss Optimization: The overall training loss function for each branch is given by:

$$\mathcal{L}_{\text{Freq}} = \underbrace{\left(\mathcal{L}^{f}_{\text{rec}} + D^{f}_{\text{KL}}\right)}_{\text{VAE}} + \underbrace{\lambda_f \mathcal{L}_{\text{CE}}(\hat{\mathbf{Y}}_{\mathbf{F}}, \mathbf{Y})}_{\text{Classifier}} \tag{1}$$

$$\mathcal{L}_{\text{Time}} = \underbrace{\left(\mathcal{L}^{t}_{\text{rec}} + D^{t}_{\text{KL}}\right)}_{\text{VAE}} + \underbrace{\lambda_t \mathcal{L}_{\text{CE}}(\hat{\mathbf{Y}}_{\mathbf{T}}, \mathbf{Y})}_{\text{Classifier}} \tag{2}$$

where $\mathcal{L}^{f}_{\text{rec}}$ and $\mathcal{L}^{t}_{\text{rec}}$ denote the mean squared error (MSE) reconstruction losses for the frequency and time branches, respectively, $\mathcal{L}_{\text{CE}}$ stands for the well-known cross-entropy loss function, and λ_t and λ_f are regularization parameters for the time and frequency domains, respectively. The terms D^{f}_{KL} and D^{t}_{KL} correspond to the Kullback–Leibler (KL) divergence regularizer [9], forming the standard evidence lower bound (ELBO), which enforces the approximate posterior distributions to remain close to a standard Gaussian prior. For a latent vector of dimension d, the KL divergence is expressed as:

$$D_{\text{KL}} = \frac{1}{2}\sum_{i=1}^{d}\left(\mu_i^2 + \sigma_i^2 - \log(\sigma_i^2) - 1\right). \tag{3}$$

As a result, each branch outputs a confidence score for each fault class, $\hat{\mathbf{Y}}_{\mathbf{T}}$ and $\hat{\mathbf{Y}}_{\mathbf{F}}$, corresponding to the time-domain and frequency-domain branches, respectively.

B. Phase 2: Meta-Learning via Stacking Ensemble

The probabilistic outputs from the time-domain and frequency-domain branches ($\hat{\mathbf{Y}}_{\mathbf{T}}$ and $\hat{\mathbf{Y}}_{\mathbf{F}}$) were concatenated to form the meta-feature matrix $\mathbf{X}_{\text{meta}} = [\hat{\mathbf{Y}}_{\mathbf{T}}, \hat{\mathbf{Y}}_{\mathbf{F}}]$. This matrix represents the confidence distribution of both base-models across all fault classes. Importantly, the meta-classifier is trained solely on these prediction-level features, without accessing the original input data, consistent with the stacking ensemble paradigm.

The meta-classifier was then trained using $\mathbf{X}_{\text{meta}}$ on MATLAB with the AdaBoost multiclass extension method, proposed by Freund and Schapire [13], designed to emphasize ambiguous class boundaries. Instead of reweighting samples alone, the extension method maintains a distribution $D_t(i, y)$ over example–label pairs (x_i, y) with $y \neq y_i$, allowing the boosting process to focus on distinguishing the correct label from the most confusable incorrect alternatives. At iteration t, the weak learner h_t is trained and evaluated using the pseudo-loss:

$$\varepsilon_t = \frac{1}{2}\sum_{(i,y)\in B} D_t(i, y)\left[1 - h_t(x_i, y_i) + h_t(x_i, y)\right], \tag{4}$$

where $B = \{(i, y) : y \neq y_i\}$ is the set of all mislabel pairs, and $h_t(x_i, y)$ represents the degree to which the weak classifier deems label y to be plausible for input x_i. The learner is then weighted by $\alpha_t = \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$, and the distribution D_t is updated to increase the influence of the most challenging instance–label pairs in subsequent rounds. After M iterations, the final ensemble prediction is given by:

$$\hat{Y}_{\text{final}} = \arg\max_{y\in\mathcal{Y}} \sum_{t=1}^{M} \alpha_t \, h_t(x, y). \tag{5}$$

In this framework, the weak learner h_t is a shallow decision tree. This stacking meta-classifier then acts as a second-level learner that integrates the complementary information captured by both time and frequency-domain branches, resulting in improved fault classification accuracy and robustness.

979-8-3315-8138-1

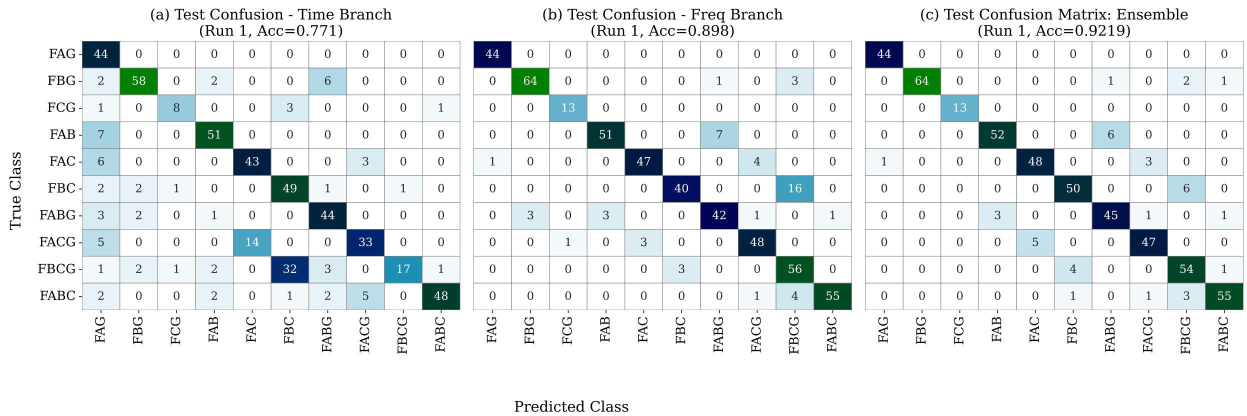

Fig. 3. Confusion matrices: (a) Time branch, (b) Frequency branch, and (c) VAE-stacking (AdaBoost) ensemble.

IV. Performance Assessment and Results

To comprehensively evaluate the proposed framework, VAE-Stacking performance was compared against frequency-domain and time-domain only architectures, as well as using Bagging as a meta-model instead of Adaboost.

Furthermore, the performance was benchmarked against three classical machine learning techniques: Support Vector Machine (SVM), Random Forest (RF) with bagging and Random Forest (RF) with boosting [13]. To this end, a two-level discrete wavelet decomposition was applied using the Daubechies wavelet (db8) as the mother wavelet. From the detail coefficients at each decomposition level i, represented by D_{ij} for $j = 1, 2, \ldots, N_d$ (where N_d denotes the number of detail coefficient samples), eight statistical metrics were extracted, as detailed in Table I [17]. This process produced a total of 16 wavelet-based statistical features for each voltage and current signal.

The classifiers' performance was evaluated using confusion matrices to calculate accuracy. A confusion matrix quantifies the classifier's predictions, capturing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), from which the accuracy is computed.

For the classical models, to enhance the robustness of the results, a bootstrap approach was employed. Data was split randomly, with two-thirds allocated to the training subset and one-third to the testing subset. This process was repeated ten times to reduce variance and increase reliability, and the final accuracy for each model was obtained by averaging results across iterations.

Table II shows the results for the classical and one-branch VAE-Stacking models. As it can be noticed, the one-branch approach does not improve or even decreases performance when compared to classical models.

The quantitative results of the new proposed model, averaged over independent runs, are presented in Table III. The stacking ensemble, which fuses latent features from both domains via AdaBoost meta-classification, achieved the highest test accuracy (0.9414 ± 0.0192) and mean F1-score (0.9374 ± 0.0206). In contrast, the frequency-only and time-only models yielded lower accuracy and F1-scores, indicating reduced discriminative power when relying on a single modality, as well as the classical machine learning methods, demonstrating that the complex, jointly-trained latent features from the dual-branch VAEs provided a significant advantage in discriminating complex fault patterns that classical methods could not fully capture.

As shown in Table III, the VAE-Stacking models significantly outperformed all standalone models from Table II. This confirms the core hypothesis that the time and frequency domains are complementary; while the time-domain is weak alone, it provides critical information that resolves ambiguities for the frequency-domain model, resulting in a more accurate final classification. The selection of AdaBoostM2 as the final meta-classifier was based on this comparative analysis. While the bagging ensemble already provided a strong result (93.10% accuracy), AdaBoostM2 achieved a superior and more stable performance (94.14% $\pm$ 0.02). This is attributed to AdaBoost's boosting mechanism, which iteratively focuses on the most difficult-to-classify samples. In this framework, it effectively learns to place more weight on the predictions from the more confident branch (frequency domain) while strategically using the time branch's outputs to correct misclassifications in cases of ambiguity, leading to a more refined and robust final decision boundary.

To provide a deeper insight into the model's behavior, Figure 3 presents the confusion matrices from a single, representative test run for each component. The matrices visually confirm that the time branch (a) is a weak learner with widespread confusion, while the frequency branch (b) is a strong classifier that still exhibits specific errors (e.g., 'FBC' as 'FBG'). The final VAE-Stacking model (c) achieves the highest accuracy by leveraging the ensemble and it was observed in every run. Its visibly cleaner confusion matrix diagonal demonstrates that the AdaBoost meta-learner successfully used the information from the weak time branch to resolve ambiguities and correct the errors of the stronger frequency branch, thus validating the

979-8-3315-8138-1

TABLE I
STATISTICAL FEATURES

Mean	$\mu_i = \frac{1}{N_d}\sum_{j=1}^{N_d} D_{ij}$	Energy	$E_i = \sum_{j=1}^{N_d} \lvert D_{ij}\rvert^2$
Standard variation	$\sigma_i = \left(\frac{1}{N_d}\sum_{j=1}^{N_d}(D_{ij}-\mu_i)^2\right)^{1/2}$	Log energy entropy	$LOE_i = \sum_{j=1}^{N_d} \log\left(D_{ij}\right)^2$
Skewness	$SK_i = \sqrt{\frac{1}{6N_d}}\sum_{j=1}^{N_d}\left(\frac{D_{ij}-\mu_i}{\sigma_i}\right)^3$	Shannon Entropy	$Ent_i = -\sum_{j=1}^{N_d} D_{ij}^2 \log\left(D_{ij}\right)^2$
Kurtosis	$K_i = \sqrt{\frac{N_d}{24}}\left(\frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{D_{ij}-\mu_i}{\sigma_i}\right)^4 - 3\right)$	Root Mean Square	$rms_i = \sqrt{\frac{1}{N_d}\sum_{j=1}^{N_d} D_{ij}^2}$

TABLE II
COMPARATIVE MODEL PERFORMANCES: CLASSICAL AND INDIVIUAL TIME AND FREQUENCY BRANCHES

Model Description	**Performance Results**	
	Model	***Accuracy*** *(%)*
CLassical	SVM	91.30 ± 0.78
Classical	RF Bagging	89.77 ± 2.16
Classical	RF AdaBoost	75.78 ± 1.94
VAE-Branch	VAE-BiLSTM$_F$	91.08 ± 0.04
VAE-Branch	VAE-BiLSTM$_T$	76.73 ± 0.03

TABLE III
COMPARATIVE MODEL PERFORMANCE

Group Head	**Performance Results**		
	Model	***Accuracy*** *(%)*	***F1-Score*** *(%)*
VAE-Stacking	**AdaBoostM2**	**94.14 ± 0.02**	**93.74**
VAE-Stacking	Bagging	93.1 ± 0.034	93.16

quantitative superiority reported in Table III.

V. CONCLUSION

The results demonstrate that the proposed hybrid deep learning approach offers substantial improvements in fault diagnosis accuracy for inverter-based power systems, with the stacking ensemble model achieving the highest performance among the evaluated methods. These findings confirm the effectiveness of combining time and frequency-domain features, and highlight the model's robustness in handling the nonlinear dynamics characteristic of modern renewable-rich grids.

The implications of these results extend to both academic research and industry practice, suggesting new directions for the development of intelligent grid monitoring solutions. Future studies could explore the scalability of this approach in larger and more heterogeneous power networks, as well as its application to real-world operational data. Overall, this work contributes valuable insights to the field, supporting the advancement of automated fault classification techniques and promoting greater reliability in evolving electrical grids. A complete analysis and evaluation for a real-world power grid or a larger scale system will be considered in a future work with more cases of grid-forming and grid-following IBR mixes.

ACKNOWLEDGEMENT

The authors are grateful for the Fulbright Award PS00366046 and NSF grant 2501861.

REFERENCES

[1] R. Vaish, U. Dwivedi, S. Tewari, and S. M. Tripathi, "Machine learning applications in power system fault diagnosis: Research advancements and perspectives," *Engineering Applications of Artificial Intelligence*, vol. 106, p. 104504, 2021.

[2] K. V. Kkuni, S. Mohan, G. Yang, and W. Xu, "Comparative assessment of typical controlrealizations of grid forming converters based ontheir voltage source behaviour," *arXiv preprint arXiv:2106.10048*, 2021.

[3] K. M. Cheema, N. I. Chaudhary, M. F. Tahir, K. Mehmood, M. Mudassir, M. Kamran, A. H. Milyani, and Z. S. Elbarbary, "Virtual synchronous generator: Modifications, stability assessment and future applications," *Energy Reports*, vol. 8, pp. 1704–1717, 2022.

[4] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," *Energy strategy reviews*, vol. 24, pp. 38–50, 2019.

[5] N. Qu, J. Zuo, J. Chen, and Z. Li, "Series arc fault detection of indoor power distribution system based on LVQ-NN and PSO-SVM," *IEEE Access*, vol. 7, pp. 184 020–184 028, 2019.

[6] G. A. Fonseca, D. D. Ferreira, F. B. Costa, and A. R. Almeida, "Fault classification in transmission lines using Random Forest and notch filter," *Journal of Control, Automation and Electrical Systems*, vol. 33, p. 598–609, 2021.

[7] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and Random Forests," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5581–5588, 2017.

[8] A. A. Majd, H. Samet, and T. Ghanbari, "k-NN based fault detection and classification methods for power transmission systems," *Protection and Control of Modern Power Systems*, vol. 2, no. 4, pp. 1–11, 2017.

[9] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations (ICLR)*, 2014.

[10] J. Lin, R. Wang, W. Li, and Y. Liu, "Enhanced rolling bearing fault diagnosis using multimodal deep learning and singular spectrum analysis," *Applied Sciences*, vol. 15, no. 9, p. 4828, 2025.

[11] N. Ciliz, M. Yildirim, and E. Genc, "A deep-learning-based multi-modal sensor fusion approach for detection of equipment faults," *IEEE Access*, vol. 10, pp. 124 438–124 448, 2022.

[12] F. B. Costa, "Fault-induced transient detection based on real-time analysis of the wavelet coefficient energy," *IEEE transactions on power delivery*, vol. 29, no. 1, pp. 140–153, 2013.

[13] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[14] S. Tiwari, F. B. Costa, V. A. Lacerda, O. G. Bellmunt, and C.-W. Ten, "IEEE 2800-2022 standard compliance of mixed-model energy systems with full integration of IBRs," in *2024 IEEE Power and Energy Society General Meeting (PESGM)*, 2024, pp. 1–5.

[15] Y. Wang, M. Wen, and Y. Chen, "A simplified model of type-4 wind turbine for short-circuit currents simulation analysis," *IET Generation, Transmission & Distribution*, vol. 16, no. 15, pp. 3036–3049, 2022.

[16] V. A. Lacerda, E. P. Araujo, M. Chea-Mañe, and O. Gomis-Bellmunt, "Phasor and EMT models of grid-following and grid-forming converters for short-circuit simulations," *Electric Power Systems Research*, vol. 223, p. 109662, 2023.

[17] H. Erişti, Ö. Yıldırım, B. Erişti, and Y. Demir, "Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, vol. 49, pp. 95–103, 2013.