SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I12P130
Volume 11 Issue 12, 314-326, December 2024

Original Article

Hybrid Wavelet-LSTM-Transformer Model for Fault Forecasting in Power Grids

Diaa Salman[1], Cem Direkoglu[2], Nabeel Altanneh[3], Abdulaziz Ahmed[4]

[1]*Department of Electrical Engineering, Faculty of Engineering, Jamhuriya University of Science and Technology, Mogadishu, Somalia.*
[2]*Department of Electrical and Electronics Engineering, Middle East Technical University, Northern Cyprus Campus, 99738 Guzelyurt, Northern Cyprus, Mersin 10, Turkey.*
[3]*Department of Electrical Engineering, College of Engineering and Technology, Palestine Technical University-Kadoorie, Yafa Street, Tulkarm P.O. Box 7, Palestine.*
[4]*Department of Electrical Engineering, Faculty of Engineering, Jamhuriya University of Science and Technology, Mogadishu, Somalia.*

[1]*Corresponding Author : D.Salman@just.edu.so*

Received: 17 October 2024 Revised: 17 November 2024 Accepted: 15 December 2024 Published: 31 December 2024

***Abstract** - In power grid management, accurate forecasting of faults is the key to providing uninterrupted electricity supply. Errors can result in significant losses of property as well as severe harm. Better alternatives adhere to the complex power grid data that frequently confuse conventional forecasting techniques and produce less-than-ideal forecasts. This research proposes a novel hybrid Wavelet-LSTM-Transformer model for power grid fault prediction. Long Short-Term Memory (LSTM) network sequential learning capabilities, attention processes of the Transformer model, and time-frequency analysis capabilities of wavelet transform are combined in the suggested model. In order to forecast trends, our approach maintains the notion of long-term memory while capturing short-term variations. This paper demonstrates that the developed model outperforms the most comparable work using numerous experimental trials and comparisons. It offers a workable method to boost the worldwide resilience and dependability of power grid systems. The results stress combining several modeling techniques to tackle difficult forecasting problems in important infrastructure sectors.*

***Keywords** - Fault forecasting, Deep Learning, Hybrid model, Power grid.*

1. Introduction

The basis of modern society is the contemporary power grid, which supplies electrical current to consumers from industrial sites. However, the intricate network of interrelated components of these networks presents several difficulties, including hardware failure, natural calamities, and operational mistakes. Building stable and dependable electrical systems becomes dependent on early problem diagnosis. Fault detection is a crucial technique in power systems to ensure the security and reliability of electric networks [1].

Energy is still more important than ever for economic expansion in the modern world. The many problems their networks face include equipment failure and environmental risks. To avoid probable outages, reduce the likelihood of safety risks, and optimize the potential of the power system, prompt problem identification and resolution is essential in this constantly changing environment. Unlike the technical aspects of power networks, fault identification raises safety issues [2]. Early problem diagnosis is necessary to protect the public and utility staff from dangerous circumstances caused by power outages. Moreover, effective mistake detection capacity has already significant economic consequences [3].

Actually, it helps to extend the life of essential infrastructure, shields expensive electrical equipment from damage, and reduces the cost of individual repairs. Basically, efficiency is the key to fault detection. With prompt interference, operators can boost electricity distribution, curtail losses, and maintain overall efficiency in the entire system.

This article aims to explore the intricate link between fault detection and grid efficiency, in that identifying mistakes actively helps energy networks perform better. Several strategies are utilized to identify errors in time series data, and these approaches can be roughly classified into statistical, machine learning, and hybrid methods. Time series fault forecasting uses several techniques to warn and prevent potential system disturbances [4].

Conventional time series analysis models like Autoregressive Integrated Moving Average (ARIMA) [1] and Exponential Smoothing State Space Models [5], with differing emphases, respectively, capture previous patterns to predict future faults. Moreover, machine learning techniques like decision trees and Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) use data to model their predictions [6].

Ensemble approaches and hybrid models combine different methods for higher accuracy. Adopting adaptive forecasting, feature engineering and online learning methods ensures that models can properly adjust with changing conditions and take full advantage of useful features for the task at hand fault prediction [7]. The selection of methodologies is based on the features of time series data and system needs; frequently, redundant usage of several methodologies is necessary to improve fault prediction accuracy or dependability.

This work presents the Wavelet-LSTM-Transformer, a hybrid forecasting model that integrates three key elements: the transformer, LSTM, and the wavelet transform. Wavelet transform allows the extraction of several resolution characteristics. Analyzing signals with various frequency properties is made easier using it. Temporal relations are particularly well-suited for representation by a class of recurrent neural networks known as long-range dependencies, or LSTM. Meanwhile, the transformer architecture represents remote interactions very well through attentional mechanisms and achieves parallel processing free from sequential order restrictions.

Both components are combined since the model attempts to leverage the feature extraction of Wavelet Transforms and the long-range dependency modeling of LSTM while also considering global relationships as assessed by the transformer. With this integrated strategy, people should be able to project work flexibly and excellently to suit all relevant applications. As a result, the model can more effectively identify complex patterns from time series data in several categories.

An innovative contribution of this work is the development of Wavelet-LSTM-Transformer, a heterogeneous prediction model integrating wavelet transform, LSTM, and transformer. Its all-inclusive time series research frameworks, which combine long-term dependency modeling and global relationship comprehension with many multi-resolution feature extraction techniques, give it its distinctiveness.

The diversity of the model implies that the forecast accuracy can be increased over previous methods. Furthermore, this work pioneers the combination of deep learning approaches. It has established a canon for the next research to tackle difficult analytical problems employing inventive technical combinations by effectively merging several components into one unified and flexible structure. The model results are contrasted with five other models: Wavelet-LSTM, LSTM-Transformer, Wavelet-Transformer, LSTM, and Transformer models. The outcome demonstrates the robustness of the suggested model since it can be trusted in such situations and outperforms the other models in all the evaluation criteria.

2. Literature Review

Electrical distribution networks must be reliable, safe, and efficient; fault predicting is essential. Even hard to foresee failures could lead to widespread outages, equipment damage, or even safety risks. Apart from obvious disruptions, ignoring problems could have serious repercussions. Downtime results in substantial financial losses, as do increased operating costs from emergency repairs and maintenance activities; and, most significantly, the loss of safe supplies brought on by catastrophic catastrophes has a far-reaching economic impact.

Like this, waves outward. Every time something goes wrong, By contrast, proactive fault forecasting can be very beneficial to utilities and operators since it enables them to take preventative measures to lower risks or expenses. It is also a method of increasing grid capacity overall to meet the demands made on it while maximizing grid performance [8].

Many issues and interruptions brought about by power grid faults may impact the effectiveness and dependability of electrical distribution networks. The system's stability is important since faults can lead to oscillations, frequency changes, and even widespread failures in an unstable network. Unexpected load redistributions, overloads of transmission lines, and disturbances to supply-demand equilibrium are among the more complicated consequences of changes in power flow dynamics. Concurrently, the most typical consequence is equipment loss [9].

This happens when technology problems cause important infrastructure parts, such as transformers, circuit breakers, and safety relays, to fail or deteriorate. These problems are exacerbated by the following power flow transfer phenomena, which makes flexible control techniques necessary to manage and change network structure on the go. Voltage variation is another frequent occurrence and results from issues that cause voltage levels to spike, drop, or vary.

These modifications directly impact the quality of the electricity supplied, which increases the possibility of sensitive equipment breaking down when needed most. It is clear from weighing these effects that efficient defect detection, forecasting, and management methods are necessary. This highlights the need to take proactive measures to improve the power grid and lessen the detrimental effects while offering secure supervision of power distribution networks [10, 11]. Figure 1 summarizes the main effects of the grid faults. As power grid fault forecasting is closely related to the overall efficiency, safety and reliability of electric distribution systems, it has long been a hotly discussed issue. Traditional forecasting methods have provided valuable information but often fail to deal with power grid data's complex, nonlinear and dynamic nature.

That is why statistics and determinism mainly determine these models. On the other hand, bringing deep learning heralded a sea change; it opened up an unprecedented realm of possibilities for making more accurate and dependable predictions. Table 1 discusses a variety of different forecasting techniques and their characteristics.

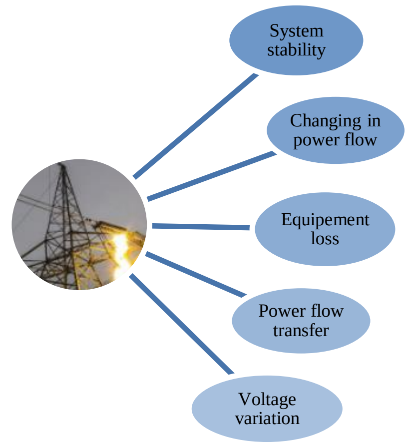

Fig. 1 The effects of power grid faults

Table 1. A comparison of forecasting methods and deep learning architectures

Ref.	Method	Description/ Characteristics
[12]	Time-Series Analysis	Identifies patterns and trends by analyzing historical data.
[13]	Seasonal Decomposition	Sorts time series data by trend, seasonality, and residuals.
[14]	Autoregressive Models	Uses historical data and variables with a lag time to predict future values.
[15]	Exponential Smoothing	Prioritizes up-to-date data for prediction by using weighted averaging.
[16]	Auto Regressive Integrated Moving Average (ARIMA)	Uses a combination of moving average and autoregressive components to make predictions.
[17]	RNN	Sequential data modeling involves capturing temporal dependencies.
[18]	LSTM	A specialized RNN with a memory cell is designed to address the vanishing gradient problem.
[19]	Transformer	Architecture that utilizes attention mechanisms and parallel processing.
[20]	Convolutional Neural Network (CNN)	Employs convolutional layers to extract features hierarchically.
[21]	Generative Adversarial Network (GAN)	Comprises generator and discriminator networks for data generation.
[22]	LSTM-ARIMA	Combines LSTM and ARIMA to capture both prolonged relationships and time-based patterns effectively.
[23]	CNN-LSTM	The approach utilizes CNN for extracting features and LSTM for modeling sequential data.
[24]	Wavelet-LSTM	The approach uses Wavelet Transform for feature extraction and LSTM for capturing temporal relationships.

3. Algorithms Background and Preliminaries

3.1. Wavelet Transform

The Wavelet Transform is a mathematical tool that, like the Fourier Transform, is used to decompose functions into frequency components. On the other hand, the Fourier Transform uses sinusoids, but the Wavelet Transform employs wavelets, which are brief waves with a limited period. One of the most essential aspects of wavelets is their ability to record frequency and position information simultaneously. Examining non-stationary signals, such as time series data with fluctuating frequency components across time, is particularly useful for wavelet analysis. This is so because multi-resolution analysis is possible with wavelets [25]. From the viewpoint of time-series analysis, wavelet transformation makes it feasible to extract significants properties at multiple scales, making it possible to depict complex temporal patterns in a more complete and adaptive manner [25]. A function f(t) which may be decomposed into its wavelet coefficients W(a,b) making use of the Wavelet Transform, which is:

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt \tag{1}$$

The wavelet function that has been scaled by a and translated by b is represented as $\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$. The wavelet coefficient captures both temporal and frequency localizations and is suitable for assessing nonstationary signals in time-series data.

The first step in ensuring consistency when analyzing power grid failure data over time is to compile and standardize the raw data. After that, this data is divided into several parts using the Wavelet Transform, like in a tale separating the main and subordinate topics. Important properties are extracted from these parts to identify and understand various power system failures. Specialized algorithms are applied after noise reduction techniques to uncover patterns and potential issues in the grid. By early identification and resolution of problems, this approach contributes to the stability of the electrical system.

3.2. Long Short Term Memory

A subset of Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks are designed to assist machines in learning how to depend on information over the short and long terms [26]. LSTMs employ input, forget, and output gates with specialized memory cells to regulate the information flow and allay worries about vanishing and increasing gradient issues that plague regular RNNs when processing long sequences. That will then enable them to absorb lengthier sequences more successfully. In sequential datasets, LSTMS may identify complex patterns and connections. They are, therefore, beneficial for jobs requiring a deep grasp of context and temporal links, such as sequence modeling, natural language processing, and time-series forecasting. Gating mechanisms are particular components used to build LSTM networks. These systems, which act similarly to filters, govern the flow of information inside the network's memory cells [26]. These gating mechanisms are extremely important in defining how information is to be stored, accessed, or deleted over time, which is represented as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$
$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{5}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{6}$$
$$h_t = o_t \odot \tanh(c_t) \tag{7}$$

Where c_t act like a memory bank that holds onto crucial details at each step while h_t provide a condensed summary of those stored details. As new data x_t arrives, the forgotten gate f_t determines which old information to retain or discard, and the input gate i_t identifies which new insights should be incorporated. A candidate $\tilde{c}_t$ proposes potential updates to the memory and the output gate o_t assist in determining the final summary h_t. Through the coordinated actions of these gates and specific functions like sigmoid and tanh, the LSTM effectively processes sequence, ensuring it captures and retains the essential information for accurate predictions or interpretations.

3.3. Transformer

The Transformer design represents a paradigm jump in sequence modelling since it includes a unique attention mechanism that captures global data linkages. Transformers cannot run sequentially like RNNs and LSTMs. Rather, they evaluate each component, employing self-attention processes to determine its relative relevance [27]. Transformers can preserve complex linkages and dependencies over the input sequence thanks to the simultaneous processing and attention approach. Transformers allow this to be possible." Transformers are thus especially useful for jobs requiring interdependence over great distances and contextual awareness. Transformers have shown remarkable performance in many fields, including natural language processing, photo identification, and time-series forecasting. This makes transformers quite flexible and helpful for managing sequential data. [27].

Given a sequence $X = \{x_1, x_2, \ldots, x_n\}$, the self-attention mechanism computes the attention scores Attention (Q, K, V) as follows:

Query, Key, and Value Projections

$$Q = X \cdot W_Q \tag{8}$$
$$K = X \cdot W_K \tag{9}$$
$$V = X \cdot W_V \tag{10}$$

Where, W_Q, W_K, W_V are weight matrices.

3.3.1. Scaled Dot-Product Attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

Where d_k is the dimension of the key vector, and softmax is applied along the rows to obtain the attention scores.

4. Proposed Wavelet LSTM Transformer Model

A novel architectural approach is offered via the Wavelet-LSTM-Transformer paradigm. Integration of transformer-based sequence modelling, LSTM-based temporal modelling, and wavelet transformation in this architecture provides synergistic effects. Using this integration, the model may effectively include and exploit temporal relationships and multi-resolution features in the input data. The method is thus particularly suitable for intricate time-series applications and those involving power grid forecasting.

Three primary activities comprise the prediction model: input preprocessing and representation, sequence encoding using Wavelet-LSTM layers, and sequence modeling using Transformer layers. The model structure is enumerated in Figure 2 as a flowchart.

4.1. Input Representation and Preprocessing

Preprocessing of the incoming time series data for power grids involves standardization and perhaps denoising. This ensures data consistency and reduces the number of errors brought on by noise sources. After this, a wavelet modification of the data makes it possible to extract multi-resolution features that can record regional and worldwide patterns.

In the long term, this makes the model more resilient and flexible. Following the conversion, an embedding layer prepares the continuous wavelet coefficients for further analysis by the LSTM and Transformer layers.

Let X represent the input time-series data for the shape of (T, d), where T is the sequence length and d is the feature dimension.

Then, Wavelet Transformation

$$W = \text{WaveletTransform}\ (X) \tag{12}$$

Where, W represents the wavelet-transformed data, capturing multi-resolution features.

$$E = \text{Embedding}(W) \tag{13}$$

Where E represents the embedded data suitable for subsequent processing.

4.2. Sequence Encoding Using Wavelet-LSTM Layers

The encoded data then passes via a sequence of LSTM layers. As detectives, these layers identify the temporal trends and connections in the order. Because of their better gating mechanisms, these Long Short Term Memory (LSTM) layers can comprehend short- and long-term connections. The method next goes on to the Wavelet Integration phase. Here, the LSTM layers are boosted using wavelet-transformed data. Consider it as providing the model with a specialized lens to comprehend complicated patterns at various stages in the sequence fully. The merging of wavelet-transformed features with the LSTM's innate capabilities gives a formidable model capable of decoding intricate, multi-resolution patterns in the data.

Given LSTM parameters $\Theta_{\mathrm{LSTM},}$, the LSTM encoder is defined as,

$$\mathrm{H} = \mathrm{LSTM}(\mathrm{E}; \Theta_{\mathrm{LSTM}}) \tag{14}$$

Where H contains the LSTM-encoded sequence representations.

Then, the Wavelet integration can be represented by,

$$\mathrm{H}_{\mathrm{wavelet}} = [\mathrm{H}, \mathrm{W}] \tag{15}$$

$\mathrm{H}_{\mathrm{wavelet}}$ integrates the LSTM-encoded features with the wavelet-transformed data.

5. Sequence Modeling using Transformer Layers

Following the sequence's encoding by the LSTM layers, the data passes via a sequence of Transformers blocks. Position-wise feed-forward networks and multi-head self-attention processes distinguish each of these blocks. These architectural configurations are extremely useful when understanding complex interactions and dependencies over several time intervals and feature dimensions.After completing Transformer processing, the outputs are carefully mixed with the previously encoded LSTM characteristic. We can combine the input sequence's temporal dynamics and multi-resolution data into a comprehensive picture by integrating several fusion procedures or concatenative methodologies. Finally, the collected information is fed via a prediction layer designed to generate final predictions or classifications relevant to the intended tasks, such as fault forecasting or anomaly identification. This concludes the modelling procedure. Given Transformer parameters $\Theta_{\mathrm{Transformer}}$, then the Transformer blocks are defined as

$$\mathrm{O} = \mathrm{Transformer}\left(\mathrm{H}_{\mathrm{wavelet}}; \Theta_{\mathrm{Transformer}}\right) \tag{16}$$

Where, O contains the Transformer-encoded sequence representations.

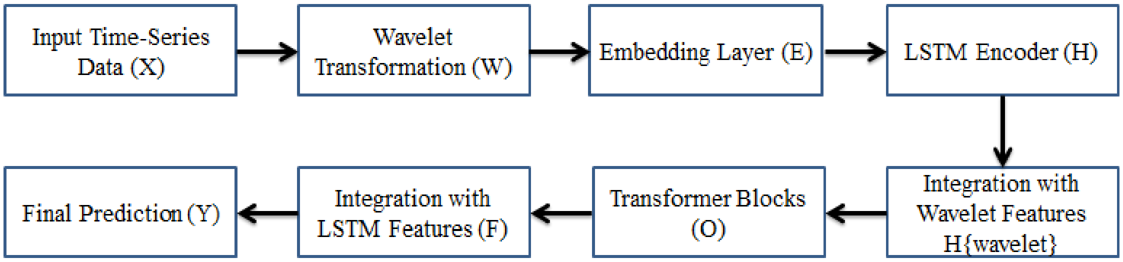

Fig. 2 Summary of the model structure

Then, the integration and fusion can be represented by

$$F = [H_{wavelet}, O] \tag{17}$$

Where, F integrates the LSTM-encoded and Transformer-encoded features.

Then, the final prediction layer can be represented by:

$$Y = \text{Dense}(F; \Theta_{Dense}) \tag{18}$$

Where, Y represents the final predictions or classifications, and Θ_{Dense} are the parameters of the dense prediction layer.

To sum up, Figure 2 shows the suggested model structure as the Wavelet-LSTM-Transformer model uses an embedding layer (E) to represent the data in a lower dimension after a wavelet transformation (W) to extract temporal features from the input time-series data (X). The LSTM encoder (H) extracts sequential patterns from the embedded data. Wavelet characteristics are also integrated into the LSTM encoder (H) to improve temporal representation. The LSTM features (F) are combined with Transformer blocks (0) to capture long-range relationships and spatial patterns. Finally, the model generates a final prediction (Y) using the combined knowledge from the LSTM and Transformer layers, resulting in a comprehensive approach to understanding and forecasting time-series data.

6. Empirical Study

6.1. Dataset Description

The history of recorded faults is evaluated in this paper in relation to the year 2020 (beginning on January 1 and ending on December 31), and this history corresponds to the total number of faults that have occurred in distribution branches in the Lages region of Brazil. The Centrais Elétricas de Santa Catarina (CELESC) provided the data for this evaluation. Considering that 2020 was a leap year, there were a total of 366 days for which records were kept. The total number of alarms regarding defects that occurred during this period is displayed in Figure 3.

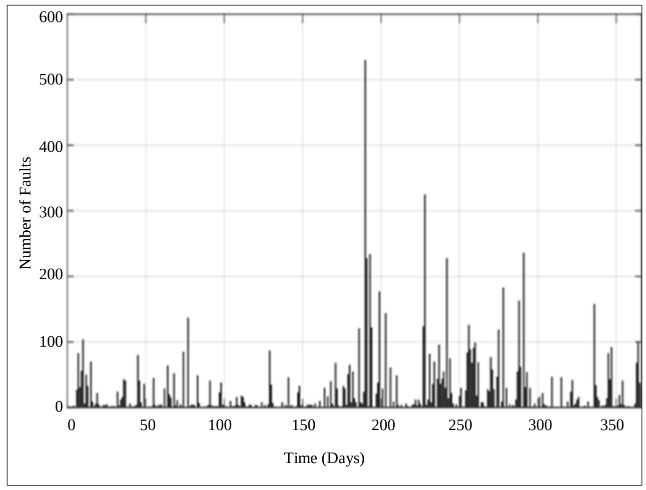

Fig. 3 Power grid failures documented in the Lages Region in 2020

Since failures generally occur in a nonlinear pattern, this evaluation was based on statistical analysis, and it was impossible to determine exactly when a failure would occur. However, it was possible to evaluate in which period of the year there was a greater chance of the most failures occurring. Performing a time series evaluation involves adding up all of the failures on the same day. This allows for the calculation of a daily failure rate over some time, which can then be used to assess the impact of the change in season on the rise in the number of failures in an electrical power grid. Regarding the alarms filed by the electric power utility business during the period being examined, these failures are evaluated in connection to themselves. Table 2 contains a few examples of alarms, which are shown here.

Table 2. An illustration of recorded alarms during the specified time frame

Date	Time	Fault Record
07/01/2020	00:24:02	Current Phase C
06/04/2020	10:04:32	Current Phase A
06/04/2020	10:05:53	Current Phase B
09/06/2020	15:21:21	Recloser Failure
30/06/2020	14:11:05	Relay 50/51 (Phase C)
10/11/2020	17:06:48	Neutral Protection
31/12/2020	10:30:47	Current Phase A

6.2. Baseline Model

Baseline models are crucial to power grid fault predictions, revealing the efficacy and improvements of new methods. Branco et al. [24] applied hybrid Wavelet-LSTM for fault forecasting in electrical power grids, which examines predictive modelling during the 2020 Brazilian pandemic, which is a good example. In this study, the authors used the Wavelet-LSTM model to improve LSTM network prognostication. The ability of the wavelet transform to enhance temporal data extraction and provide the LSTM model with better precision and robustness led to the selection of this model design.

This work constructed a hybrid wavelet-LSTM-transformer model that combines transformer, LSTM, and wavelet transform elements in a synergistic framework. After a knowledge of the benefits and drawbacks of numerous models, a composite framework that incorporates their best aspects was created. Wavelet-enhanced feature extractions, LSTM temporal modeling, and Transformer attention methods are claimed to generate a model architecture with improved prediction fidelity and durability.

These analytical frameworks combined create a scholarly discussion with significant implications for electrical power system failure prediction paradigms. This report carefully assesses the pros and disadvantages of each model architecture to determine its overall benefits. Robustness profiles, computational efficiency, and prediction accuracy measure analysis expose every model's special value offers and development prospects. Comparative evaluations go beyond numerical assessments, including qualitative examinations of every modelling paradigm's epistemological, practical, and theoretical foundations. This multifaceted analysis contributes deep theoretical knowledge and real-world application to academic debate.

6.3. Working Structure

This power fault prediction system uses a transformer, Long Short-Term Memory, and Wavelet Transform topologies. Time series power grid measurements guarantee homogeneity, and preprocessing eliminates noise. A more precise study of local and global patterns is made possible by the Wavelet Transform by retrieving multi-resolution attributes from preprocessed data.

LSTM layers record temporal dependencies and patterns; transformer layers record complex interactions across time steps and feature dimensions. Transformers and LSTMs are used in the hybrid design to increase the model's lifetime and predictability.

RMSE, MAE, and R^2 assess performance, and the model uses SGDM, Adam, and RMSprop optimizers. Advanced optimization techniques and neural network architectures are combined in this technology to produce reliable and accurate power grid failure forecasts. The inquiry workflow is shown in Figure 4. The results of the proposed model are then compared with five different models, Wavelet-LSTM, LSTM-Transformer, Wavelet-Transformer, LSTM, and Transformer models, in terms of the mentioned evaluation metrics.

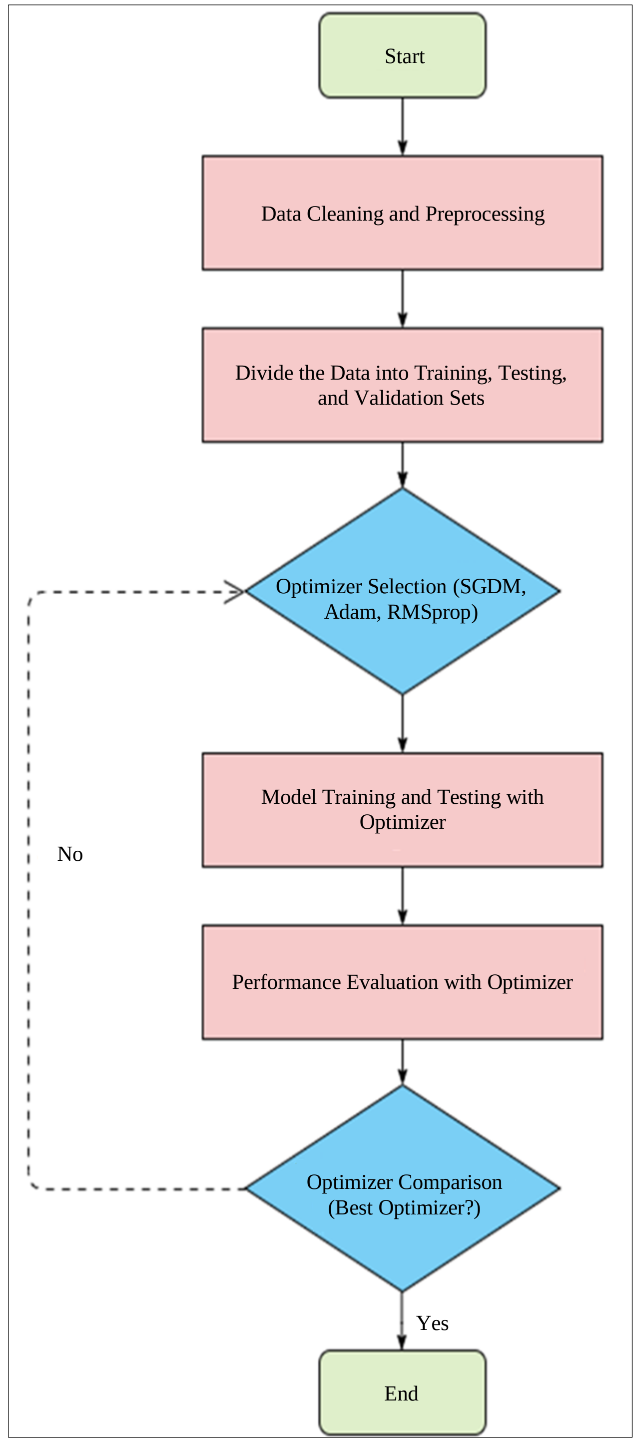

Fig. 4 Workflow and optimizer selection flowchart

7. Results and Discussions

7.1. Model Setup

The proposed Wavelet-LSTM-Transformer model captures input data's complicated temporal dependencies and feature interactions. Multiple levels in the model serve different roles in the hierarchical data processing pipeline. The model initially receives input sequences with dimensions of (None, 100, 10), where None allows for flexible batch sizes, 100 represents sequence length, and 10 represents feature dimensionality at each time step. Data transformations begin with this basic input layer, which channels sequences into the LSTM layer. Long-range relationships and temporal dynamics in data sequences are captured well by the LSTM component with 71,168 parameters. A dropout layer post-LSTM selectively deactivates neurons with a 20% dropout rate during training to strengthen the model's overfitting resistance and generalization ability.

A Multi-Head Attention mechanism with 527,488 parameters is added to the architecture to identify and highlight data patterns and relationships. The attention-based layer improves the model's ability to capture local and global data sequence dependencies. A tightly linked neural network segment with 16,512 parameters refines extracted features after the attention layer, enabling an intricate change that prepares the input for downstream processing.

A key concatenation layer smoothly integrates attention and dense layer features into a consolidated feature map of dimensions (None, 100, 256). A thick layer with 16,448 parameters feeds a singular-neuron output layer that generates model predictions to complete the architectural ensemble. The model has 631,681 trainable parameters, demonstrating its ability to identify complicated patterns and derive complex correlations from data sequences. Figure 5 illustrates the summary of the proposed model.

7.2. Numerical Results

In machine learning and predictive analytics, models are routinely compared to existing methods to determine their efficacy and practicality. This study compared our "Wavelet-LSTM-Transformer" model to Branco et al.'s [24] and LSTM-Transformer, Wavelet-Transformer, LSTM, and Transformer models. The proposed model's predictive capability and generalizability in power grid failure forecasting were tested. Table 3 illustrates the proposed model's numerical analysis and compares it with Branco et al. (Wavelet-LSTM model) reached by 200 epochs, and the wavelet transform with one node is selected.

The analysis began with RMSE and MAE measures. Lower numbers indicate better model predicting accuracy. Here, the proposed model performed well. In all three optimizers, SGDM, ADAM, and RMSprop, the proposed model had lower RMSE and MAE than Branco et al..

The model can explain more data variability when the R2 value is close to 1. Our Model and Branco et al. [1] revealed low R2 values across all evaluated optimizers, indicating that while providing accurate predictions, these models may not capture all data patterns and nuances.

Optimizers are crucial to deep learning model training and performance. Their main role is adjusting model characteristics like weights and learning rates to decrease loss function. We tested several optimizers to see which improved model performance. The performance measures, especially R2, reveal the optimizer's ability to explain data variance. Our trials showed that the ADAM optimizer, known for its adjustable learning rates and momentum, performed well.

These optimizers can handle complex loss landscapes, ensuring faster convergence and reliable generalization. However, optimizers like RMSprop have advantages, especially for non-stationary targets. Their adaptive mechanisms protect them against unpredictable gradients, making model training stable and efficient. Choosing an optimizer is a strategic decision that can affect the model's predictive power and generalizability. Our study shows the effectiveness of several optimizers and emphasizes the importance of optimizer selection in deep learning research and implementation. A comparison has been made on six different models: Wavelet-LSTM using Branco et al. results, Wavelet- LSTM- Transformer, LSTM-Transformer, wavelet-transformer, LSTM, and Transformer Models as shown in Table 3.

Comparative examination shows the proposed model's potential to challenge fault forecasting methods. While the models performed well across optimizers, further training epochs or ensemble approaches may improve the proposed model. Future studies could also investigate the patterns behind the disparities, improving power grid failure forecasting tools.

According to the results of our empirical research, RMSprop consistently outperformed SGDM and ADAM across all of the criteria studied. These results highlight the relevance of optimizer selection, which also highlights the potential superiority of RMSprop for the particular model and dataset being considered.

Figure 6 illustrates the proposed model performance regarding training and validation losses for the three optimizers. Moreover, Figure 7 shows the actual and predicted values of the power grid faults using the proposed model, and Figure 8 illustrates the actual and predicted values of the power grid faults using the Branco et al. model.

```
____________________________________________________________________________________________________
 Layer (type)                   Output Shape              Param #    Connected to
====================================================================================================
 input_layer (InputLayer)       [(None, 100, 10)]         0          []

 lstm_1 (LSTM)                  (None, 100, 128)          71168      ['input_layer[0][0]']

 dropout (Dropout)              (None, 100, 128)          0          ['lstm_1[0][0]']

 multi_head_attention (Mult     (None, 100, 128)          527488     ['dropout[0][0]',
 iHeadAttention)                                                      'dropout[0][0]']

 dense_2 (Dense)                (None, 100, 128)          16512      ['multi_head_attention[0][0]']

 dropout_1 (Dropout)            (None, 100, 128)          0          ['dense_2[0][0]']

 concatenate (Concatenate)      (None, 100, 256)          0          ['dropout[0][0]',
                                                                      'dropout_1[0][0]']

 dense_3 (Dense)                (None, 100, 64)           16448      ['concatenate[0][0]']

 output_layer (Dense)           (None, 100, 1)            65         ['dense_3[0][0]']

====================================================================================================
Total params: 631681 (2.41 MB)
Trainable params: 631681 (2.41 MB)
Non-trainable params: 0 (0.00 Byte)
```

Fig. 5 The proposed model summary

Table 3. The numerical results of the proposed model compared with five different models for time series fault forecasting

Method	Optimizer	RMSE	MAE	R^2
Branco et al. [24]	RMSprop	3.93 × 10−3	6.64 × 10−4	0.5212
	SGDM	3.51 × 10−3	1.17 × 10−3	0.2089
	ADAM	3.98 × 10-3	6.51 × 10−4	0.5554
The Proposed Model	RMSprop	**1.42× 10−3**	**3.38× 10−4**	**0.6576**
	SGDM	**2.51 × 10−3**	**3.25× 10−4**	**0.6593**
	ADAM	2.14× 10−3	5.97× 10−4	0.6571
LSTM-Transformer	RMSprop	0.00438	0.00695	0.1591
	SGDM	0.00392	0.00139	0.1254
	ADAM	0.00495	0.00731	0.1985
Wavelet-Transformer	RMSprop	7.25× 10−3	9.87× 10−4	0.5321
	SGDM	6.85× 10−3	8.31× 10−3	0.2354
	ADAM	6.98× 10−3	8.65× 10−4	0.5425
LSTM	RMSprop	0.0495	0.0698	0.1595
	SGDM	0.535	0.775	0.1065
	ADAM	0.0532	0.0783	0.1142
Transformer	RMSprop	0.656	0.8211	0.1052
	SGDM	0.0621	0.0863	0.1432
	ADAM	0.0736	0.0935	0.1596

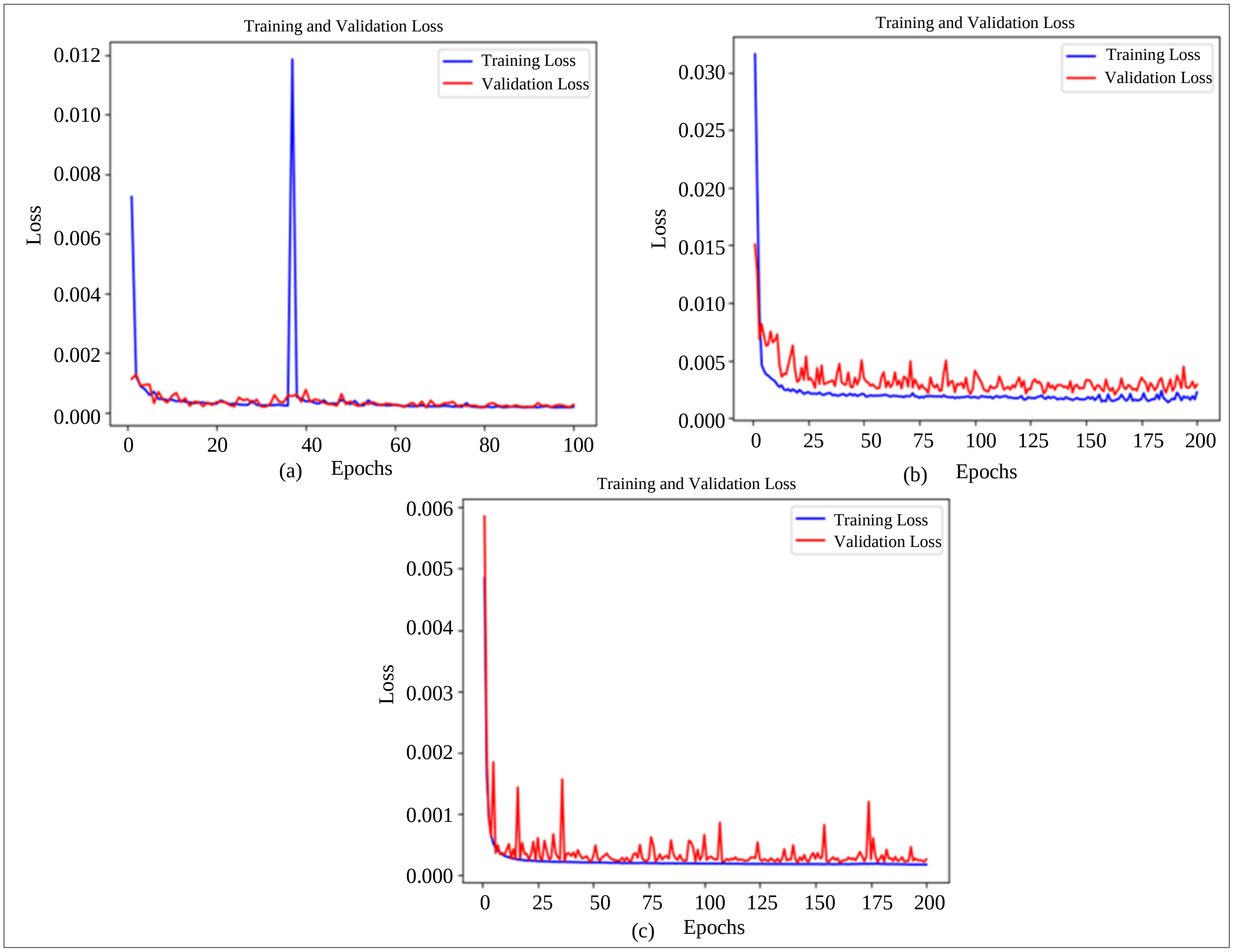

Fig. 6 the training and validation loss plot for the model of wavelet-LSTM-transformer in the case of (a) RMSprop optimizer, (b) SGDM optimizer, and (c) ADAM optimizer.

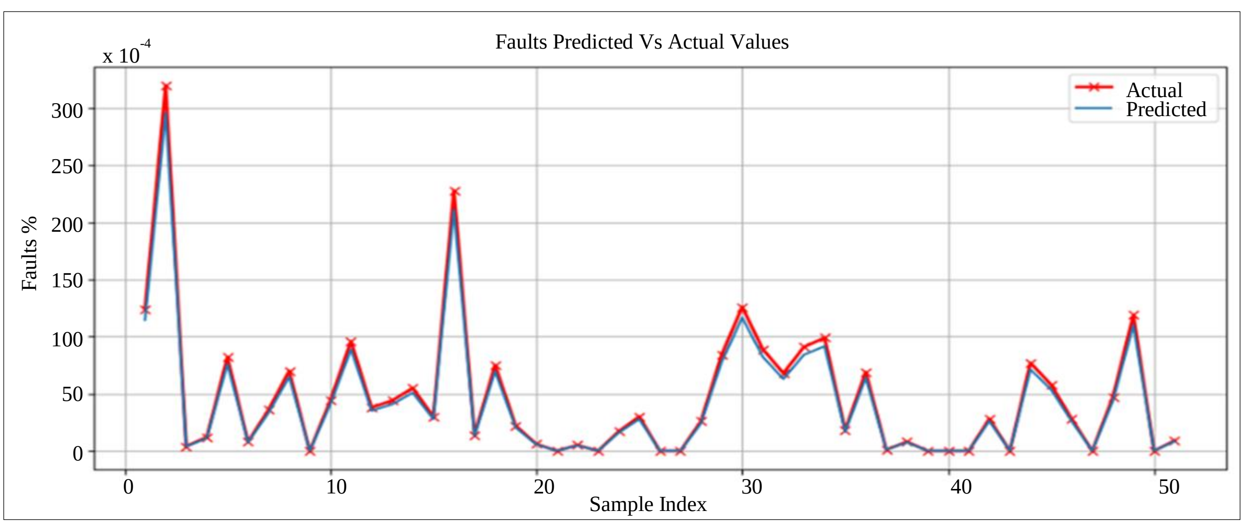

Fig. 7 Actual vs Predicted faults using wavelet-LSTM-transformer model with RMSprop optimizer

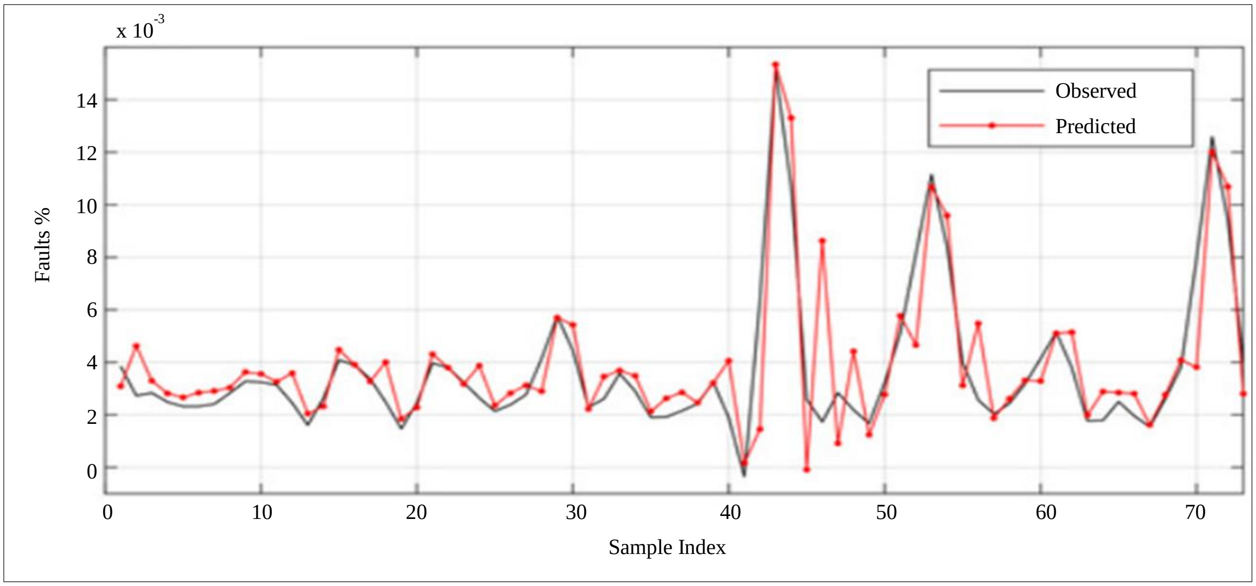

Fig. 8 Actual vs Predicted faults using wavelet-LSTM model

This research aims to conduct an exhaustive evaluation of six different models specifically designed for fault forecasting in power networks. Among these models were variants of the Wavelet-LSTM, LSTM-Transformer, Wavelet-Transformer, LSTM, and Transformer architectures. These models were trained and assessed using three distinct optimizers: RMSprop, SGDM, and ADAM. By use of criteria such as RMSE, MAE, and reliability R squared, we were able to determine the suitability of each model for fault prediction tasks and get knowledge about its relative performance.

It is demonstrated that the hybrid wavelet-LSTM-transformer model is a very effective model across all optimizer configurations. Remarkably, the Wavelet-LSTM-Transformer model constantly obtained the lowest RMSE values when trained with RMSprop, SGDM, and ADAM optimizers. These findings show that the proposed model is resilient and applicable to capture the intricate temporal correlations found in fault data. Moreover, the Wavelet-LSTM-Transformer model frequently outperformed the baseline Branco et al. model as well as other variations, including Transformer, LSTM-Transformer, Wavelet-Transformer, and LSTM, under all optimizer parameters, proving its superiority in fault forecasting tasks. This was shown by the model frequently beating the baseline model.

Moreover, the optimizer selection impacted the model performance because each optimizer demonstrated varying success across various model designs. The fact that the ADAM optimizer delivered competitive solutions for most models repeatedly shows how well it can traverse complex optimization environments. Conversely, the architecture of the underlying model affected the uneven performance of RMSprop and SGDM. This emphasizes the need to choose optimizers with great thought while training models.The Transformer model performed far worse than others, particularly when trained using SGDM and RMSprop optimizers. This study shows that the Transformer design may not be as good at capturing long-range connections in failure data.

These findings significantly impact fault forecasting in power grids and highlight the significance of selecting appropriate model architectures and optimizer settings to obtain the best possible predictive accuracy. The created WLST model mainly exhibits promise for practical use. It makes proactive maintenance methods in electrical power systems easier to apply and offers better performance and dependability in problem prediction.

Taking into consideration that the forecast is one step forward, which corresponds to one day, the results of the comparison of the anticipated and observed values, which are displayed in Figure 7, indicate that it would be possible to estimate the failures that will occur on the following day based on the recorded history. As each stage of the forecast is completed, the history of the data that has been captured up until the next projected forecast is looked at to make predictions about the subsequent forecast.

Because the forecast concerns the aggregate of the failure records that have occurred with time, it could estimate the number of failures that would occur the following day. Failures are related to weather conditions; nevertheless, they tend to rise depending on the time of year, which is the topic of this research. Failures primarily occur during the winter months. Following the point at which the total number of failures reached its highest accumulated value, there was a fluctuation in the prediction, which was anticipated due to the sudden change in the time series.

8. Conclusion

The power grid environment needs forecasting systems to provide accurate and reliable data to prevent future breakdowns and guarantee continuous service. We examine six different designs that were trained using three distinct optimizers. With consistent performance above baseline models and different changes across all optimizer settings, it became evident that the Hybrid Wavelet-LSTM-Transformer model was the most promising method. These results show the need for optimizer selection for the model training process and how well the proposed model captures the intricate temporal dynamics of fault data.

We have actual proof from our experiments that the suggested model may achieve the goals. Concerning power system data, the model demonstrated a remarkable capacity to capture complex patterns, including long-term dependencies and short-term oscillations. This is a significant breakthrough because the proposed model regularly outperformed the Wavelet-LSTM model alone. Among these metrics were more significant R squared values and reduced RMSE and MAE values.

Moreover, the model is flexible and reliable, as seen by its adaptability to several optimization techniques, such as ADAM, RMSprop, and SGDM. Prediction accuracy was observed to rise with the synergistic complementing of some optimizers, namely RMSprop, to the Wavelet-LSTM-Transformer design.The data's quality and level of detail define the effectiveness of the Hybrid Wavelet-LSTM-Transformer model. While our results were positive, improving data preparation methods should be the top priority for future studies to provide a more comprehensive picture of grid behavior. Further worries regarding the model's capacity to manage processing efficiency and huge-scale grid networks persist. The structure of the suggested model has to be optimized for higher efficiency while maintaining accuracy to increase its applicability. Constant validation in many real-world grid scenarios is necessary to guarantee the system's dependability and durability.

Funding Statement

This study was performed with Jamhuriya University of Science and Technology, Mogadishu, Somalia.

Acknowledgments

The authors would like to thank the engineering faculty of Jamhuriya University of Science and Technology for their invaluable assistance and support throughout this research. Special thanks to the Department of Electrical Engineering for their guidance and contributions. The authors also appreciate the cooperation and encouragement from their colleagues and peers.

Availability of Data and Materials

Centrais Elétricas de Santa Catarina provided the data used in this paper regarding the alarms of the power distribution grids in the Lages region in Brazil from 1 January to 31 December 2020. These records are available at https://github.com/SFStefenon/ FailuresPowerGrid2020 (accessed on 10 August 2023).

References

[1] Hamid Reza Baghaee et al., "Support Vector Machine-Based Islanding and Grid Fault Detection in Active Distribution Networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 3, pp. 2385-2403, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Muhammad Sarwar et al., "High Impedance Fault Detection and Isolation in Power Distribution Networks Using Support Vector Machines," *Journal of King Saud University - Engineering Sciences*, vol. 32, no. 8, pp. 524-535, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Oktoviano Gandhi et al., "Review of Power System Impacts at High PV Penetration Part I: Factors Limiting PV Penetration," *Solar Energy*, vol. 210, pp. 181-201, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Ali Raza et al., "A Review of Fault Diagnosing Methods in Power Transmission Systems," *Applied Sciences*, vol. 10, no. 4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Yanlin Shi, Sixian Tang, and Jackie Li, "A Two-Population Extension of the Exponential Smoothing State Space Model with a Smoothing Penalisation Scheme," *Risks*, vol. 8, no. 3, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Senlin Zhang et al., "Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ling Zheng et al., "A Fault Prediction of Equipment based on CNN-LSTM Network," *2019 IEEE International Conference on Energy Internet (ICEI)*, Nanjing, China, pp. 537-541, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Stéfano Frizzo Stefenon et al., "Wavelet Group Method of Data Handling for Fault Prediction in Electrical Power Insulators," *International Journal of Electrical Power & Energy Systems*, vol. 123, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Asiye K. Ozcanli, Fatma Yaprakdal, and Mustafa Baysal, "Deep Learning Methods and Applications for Electrical Power Systems: A Comprehensive Review," *International Journal of Energy Research*, vol. 44, no. 9, pp. 7136-7157, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] A.A.P. Bíscaro et al., "Integrated Fault Location and Power-Quality Analysis in Electric Power Distribution Systems," *IEEE Transactions on Power Delivery*, vol. 31, no. 2, pp. 428–436, 2016. [CrossRef] [Google Scholar] [Publisher Link]
[11] Cynthia M. Furse et al., "Fault Diagnosis for Electrical Systems and Power Networks: A Review," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 888-906, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[12] Chen Haomin et al., "Fault Prediction for Power System based on Multidimensional Time Series Correlation Analysis," *2014 China International Conference on Electricity Distribution (CICED)*, Shenzhen, pp. 1294-1299, 2014. [CrossRef] [Google Scholar] [Publisher Link]
[13] Charith Silva, and Mohamad Saraee, "Electricity Distribution Network: Seasonality and the Dynamics of Equipment Failures Related Network Faults," *2020 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[14] Hendrik J. Groenewald, Marius Kleingeld, and Gerrit J. Cloete, "An Autoregressive Fault Model for Condition Monitoring of Electrical Machines in Deep-Level Mines," *2018 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, Cape Town, South Africa, 2018. [Google Scholar] [Publisher Link]
[15] A.V. Koteswara Rao et al., "An Exponential Smoothing based Power Swing Detection Technique for Distance Protection," *2018 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Delhi, India, pp. 182-187, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[16] Danilo Pinto Moreira de Souza, Eliane Da Silva Christo, and Aryfrance Rocha Almeida, "Location of Faults in Power Transmission Lines Using the ARIMA Method," *Energies*, vol. 10, no. 10, 2017. [CrossRef] [Google Scholar] [Publisher Link]
[17] Moath Alrifaey, Wei Hong Lim, and Chun Kit Ang, "A Novel Deep Learning Framework Based RNN-SAE for Fault Detection of Electrical Gas Generator," *IEEE Access*, vol. 9, pp. 21433-21442, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[18] Veerapandiyan Veerasamy et al., "LSTM Recurrent Neural Network Classifier for High Impedance Fault Detection in Solar PV Integrated Power System," *IEEE Access*, vol. 9, pp. 32672-32687, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[19] Ailing Zeng et al., "Are Transformers Effective for Time Series Forecasting?," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, pp. 11121-11128, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[20] Jibin B. Thomas et al., "CNN-Based Transformer Model for Fault Detection in Power System Networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[21] Bin Liao, You Du, and Xiangyun Yin, "Fusion of Infrared-Visible Images in UE-IoT for Fault Point Detection Based on GAN," *IEEE Access*, vol. 8, pp. 79754-79763, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[22] Vikram Bali, Ajay Kumar, and Satyam Gangwar, "A Novel Approach for Wind Speed Forecasting Using LSTM-ARIMA Deep Learning Models," *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, vol. 11, no. 3, pp. 13-30, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[23] Arash Moradzadeh et al., "Hybrid CNN-LSTM Approaches for Identification of Type and Locations of Transmission Line Faults," *International Journal of Electrical Power & Energy Systems*, vol. 135, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[24] Nathielle Waldrigues Branco et al., "Wavelet LSTM for Fault Forecasting in Electrical Power Grids," *Sensors*, vol. 22, no. 21, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[25] M. Sifuzzaman, M.R. Islam, and M.Z. Ali, "Application of Wavelet Transform and its Advantages Compared to Fourier Transform," *Journal of Physical Sciences*, vol. 13, pp. 121-134, 2009. [Google Scholar] [Publisher Link]
[26] Xuan-Hien Le et al., "Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting," *Water*, vol. 11, no. 7, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[27] André Veltman, Duco W.J. Pulle, and R.W. De Doncker, *Fundamentalsof Electrical Drives*, 2nd ed., Power Systems, Springer, pp. 47-82, 2016. [Google Scholar] [Publisher Link]