



UNIVERSITÀ
DEGLI STUDI
DI UDINE

Università degli studi di Udine

Wavelet LSTM for Fault Forecasting in Electrical Power Grids

Original

Availability:

This version is available <http://hdl.handle.net/11390/1235624> since 2022-10-30T17:54:51Z

Publisher:

Published

DOI:10.3390/s22218323

Terms of use:

The institutional repository of the University of Udine (<http://air.uniud.it>) is provided by ARIC services. The aim is to enable open access to all the world.

Publisher copyright

(Article begins on next page)

Article

Wavelet LSTM for Fault Forecasting in Electrical Power Grids

Nathielle Waldrigues Branco ^{1,*} , Mariana Santos Matos Cavalca ¹ , Stefano Frizzo Stefenon ^{2,3} 
and Valderi Reis Quietinho Leithardt ^{4,5} 

¹ Department of Electrical Engineering, Santa Catarina State University, R. Paulo Malschitzki 200, Joinville 89219-710, Brazil

² Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy

³ Department of Mathematics, Informatics and Physical Sciences, University of Udine, Via delle Scienze 206, 33100 Udine, Italy

⁴ COPELABS, Lusófona University of Humanities and Technologies, Campo Grande 376, 1749-024 Lisboa, Portugal

⁵ VALORIZA, Research Center for Endogenous Resources Valorization, Instituto Politécnico de Portalegre, 7300-555 Portalegre, Portugal

* Correspondence: nathi@uniplaclages.edu.br

Abstract: An electric power distribution utility is responsible for providing energy to consumers in a continuous and stable way. Failures in the electrical power system reduce the reliability indexes of the grid, directly harming its performance. For this reason, there is a need for failure prediction to reestablish power in the shortest possible time. Considering an evaluation of the number of failures over time, this paper proposes performing failure prediction during the first year of the pandemic in Brazil (2020) to verify the feasibility of using time series forecasting models for fault prediction. The long short-term memory (LSTM) model will be evaluated to obtain a forecast result that an electric power utility can use to organize maintenance teams. The wavelet transform has shown itself to be promising in improving the predictive ability of LSTM, making the wavelet LSTM model suitable for the study at hand. The assessments show that the proposed approach has better results regarding the error in prediction and has robustness when statistical analysis is performed.

Keywords: electrical power grids; fault forecasting; long short-term memory; time series forecasting; wavelet transform



Citation: Branco, N.W.; Cavalca, M.S.M.; Stefenon, S.F.; Leithardt, V.R.Q. Wavelet LSTM for Fault Forecasting in Electrical Power Grids. *Sensors* **2022**, *22*, 8323. <https://doi.org/10.3390/s22218323>

Academic Editor: Hossam A. Gabbar

Received: 30 September 2022

Accepted: 26 October 2022

Published: 30 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For electricity to reach consumers in a stable and continuous way, an electrical power grid must work independently of weather conditions [1]. To keep the electrical distribution system running, it is necessary to evaluate the performance of the electrical system's equipment through simulation, and then the disturbance conditions present in the electrical power grid can be identified [2]. Disturbances that occur in an electrical power system can significantly affect the power supply, variations in voltage level, and increased surface conductivity, or contact by the conductors with the ground can result in faults, which affect power quality [3].

Time series forecasting can be used to identify the possibility of a failure occurring, which is a promising way to assist the decision-making process for maintenance teams in an electric power utility [4]. As the increase in failures has a strong relationship with weather conditions, in rainy seasons, there is a greater chance of a failure occurring, so the study of this variation in relation to a time series is an important aspect in this context [5].

The use of wavelet transforms for noise reduction is an approach that is effective when there is high nonlinearity in the time series [6]. When using high-frequency bandwidth filters, there may be a loss of information, considering that a high frequency might be related to the occurrence of a failure. Considering that the wavelet transform evaluates the signal energy, high frequencies are not totally eliminated, thus maintaining the main signal

characteristics [7]. Thus, a hybrid approach that combines a deep learning model with a wavelet transform can be an interesting approach [8].

Long short-term memory (LSTM) is a model applied in deep learning that has been widely used by researchers for time series forecasting [9–11]. Its units solve the vanishing gradient problem partially, since LSTM units allow the gradients to flow unchanged [12]. Based on the advantages of the wavelet transform and the promising capabilities of LSTM [13–16], this work proposes using a combination of those techniques in a method named wavelet LSTM. For this purpose, a study will be conducted using the alarm data obtained from a recloser of a power utility company in the *Serrana* region of Santa Catarina, Brazil.

The main contributions of this research are the following:

- We propose a hybrid wavelet LSTM model which has a higher predictive capacity than the standard LSTM model. The wavelet LSTM shows itself to be a more stable model for time series prediction that can be used in several applications.
- We evaluate a time series regarding the variation in the number of failures in distribution networks with bare cables due to the presence of contamination and contact of foreign materials with the grid, resulting in disruptive discharges in the power grid.
- We present a solution for evaluating failure history based on a time series that can be used in other works, in which it is necessary to evaluate the number of failures over time.

The continuation of this paper is organized as follows. Section 2 presents a review of related works and the used data. Section 3 presents the proposed method. In Section 4, the results are analyzed. Section 5 presents a conclusion and a discussion of possible future works.

2. Related Works

In electrical distribution systems, an electrical fault is defined as an anomaly in a particular piece of equipment causing a forced interruption in the operation of the electrical power grid [17]. There are two classes of faults: transient and permanent. Transient faults are anomalies of short duration that disappear soon after the action of protective devices, having as common causes atmospheric discharges, momentary contacts between the conductors and the ground, the opening of an electric arc, and materials without adequate insulation. Permanent faults are faults that continue to exist until it is possible to replace the defective component or equipment [18].

Through fault diagnosis, it is possible to detect where the fault occurred as well as its size, duration, and impact on the electrical power system [19]. Among the most current fault diagnosis methods are Bayesian networks [20], fuzzy logic [21], the Kalman filter [22], and other mathematical models based on artificial intelligence. The use of artificial intelligence techniques for fault identification has been growing over the years and become a hot topic today, especially for electric power systems [23]. Deep learning models have been increasingly used to improve the ability to identify faults in an electrical grid [24–26]. However, as these models have a large number of layers, they require more computational effort, making the choice of the appropriate model a challenge [27].

From the image processing of failed components, it is possible to identify patterns and thus improve their identification in the field [28]. Several researchers are using object detection and image classification based on convolutional neural network (CNN) models [29–31].

CNNs can be specifically applied to improve the ability to identify faulty components, as shown by Liu et al. [32] and Sadykova et al. [33] using the You Only Look Once (YOLO) approach, Li et al. [34] with an improved Faster R-CNN, and Wen et al. [35] using an Exact R-CNN. As presented by Sadykova et al. [33], the YOLO model is a promising alternative for identifying insulators during power grid inspections and handling large datasets, where data augmentation techniques can also be applied to avoid early overfitting. In this context,

a super-resolution CNN can perform reconstruction of the blurred images to perform the expansion of the dataset [36].

The YOLO model has been updated, and variations in its structure can result in significant performance improvements. According to Liu et al. [32], the YOLOv3-dense model proposed by them reached up to 94.47% for insulator identification using varied image backgrounds. In comparison, for the same dataset, YOLOv3 reached 90.31% accuracy. Previous versions such as YOLOv2 reached a maximum of 83.43%, making it clear that in many situations, the use of non-standard models can be a promising alternative.

Variations of the YOLO models have been proven to be very efficient for locating insulators on transmission lines. According to Liu et al. [37], MTI-YOLO has a higher average precision than YOLO-tiny and YOLO-v2. Liu et al. [38] proposed an improved YOLOv3 model that is better than the YOLOv3 and YOLOv3-dense models. Hu and Zhou [39] showed that YOLOv4 can reach an accuracy of 96.2% for insulator defect detection, and using YOLOv4, Xing and Chen [40] had a precision of 97.78% for insulator identification.

When the distribution power system does not have insulation on the medium voltage conductors, trees might touch the conductors, resulting in discharges to the ground [41]. This type of fault is common in rural power grids that are close to wooded areas. To prevent these faults, electric power utilities perform pruning of trees that are close to the network, thus reducing the chance of discharges to the ground [42].

Insulators installed outdoors are exposed to environmental variations, such as dust accumulation on their surfaces [43]. When contamination accumulates on insulators, their surface conductivity increases, generating leakage current until a discharge occurs [44]. When there is high humidity in the air, the conductivity increases even more, consequently increasing the chance of faults in the grid [45]. One type of contamination that has a significant impact on the conductivity of insulators is salt contamination, which can be measured by the equivalent salt deposit density [46].

Considering all these possible faults [47], time series forecasting comes as an alternative to prepare maintenance teams in advance for an event based on the historical knowledge of data variation over time [48]. A forecast that is many steps ahead is challenging, as each step ahead contains the accumulated forecast error of the previous step [49]. Therefore, time series forecasting needs to take into account how many steps ahead can be considered to obtain acceptable assertiveness [50].

Among the algorithms for time series forecasting, ensemble learning models in general have high performance and lower computational effort [51], and they may be promising approaches for failure prediction. Various ways of combining the weak learners can be used to create a model that has a greater capacity, such as bagging, boosting, and stacking [52]. Further optimized models, such as the Bayesian optimization-based dynamic ensemble proposed by Du et al. [53], can be used and are even applied with nonlinear data [54].

Many variations of ensemble models for time series forecasting can be found, such as efficient bootstrap stacking, presented by Ribeiro et al. [55], or extreme gradient boosting, proposed by Sauer et al. [56]. Especially for power system failure prediction, the wavelet transform combined with ensemble models becomes a superior approach to well-established models, such as the adaptive neuro-fuzzy inference system [57]. Therefore, ensemble models are successful approaches for multi-step forward prediction [58], which is equivalent to what is being evaluated in this paper.

Due to the existing features in the structure of LSTM, it is one of the most qualified models for handling chaotic time series, since it has the ability to remember distant values and interpret the order of dependencies, which are essential characteristics for prediction models. Abbasimehr and Paki [59] used the attention mechanism to attain an enhanced LSTM model. Related to the power system using LSTM, Guo et al. [60] and Ko et al. [61] presented research about wind power forecasting. Specifically for fault prediction, Guo et al. [62] proposed a modified LSTM version to improve the safe and reliable operation of mechanical equipment.

Faults

Most faults that occur in an electrical power distribution system with naked cable are caused by direct contact with the network. This occurs mainly when the weather conditions are bad (rain and intense wind), increasing the likelihood of contact from trees with the power grid. Another failure that can occur frequently is when insulators lose their insulating capacity due to contamination or when the insulators are damaged [63].

To perform time series evaluation, all failures that occur on the same day are added up to obtain a daily failure rate over time and thus evaluate the influence of the change in season in relation to the increase in failures in an electrical power grid. These failures are evaluated in relation to the alarms registered by the electric power utility company during the evaluated period. Some examples of alarms are presented in Table 1.

Table 1. Example of alarms that have been registered in the considered period.

| Day | Time | Failure Record |
|-------------------|----------|----------------------------|
| 6 January 2020 | 11:09:41 | Current Phase B |
| 6 January 2020 | 11:09:50 | Current Phase A |
| 6 January 2020 | 17:10:32 | Current Phase C |
| 26 January 2020 | 13:57:37 | Recloser Communic. Failure |
| 06 April 2020 | 10:04:23 | Relay 50/51 (Neutral) |
| 01 June 2020 | 17:24:51 | Current Phase A |
| 30 June 2020 | 14:11:56 | Phase Voltage C |
| 27 August 2020 | 10:00:54 | Neutral Protection |
| 27 August 2020 | 11:58:48 | Current Phase C |
| 11 September 2020 | 03:06:12 | Current Phase A |
| 29 December 2020 | 13:56:32 | Relay 50/51 (Phase A) |

The alarms listed in Table 1 contain the day, time, and reason for the failure. The original dataset, which presents all recorded alarms, is available at <https://github.com/SFStefenon/FailuresPowerGrid2020> (accessed on 21 October 2021).

Since failures generally occur in a nonlinear pattern, this evaluation was based on statistical analysis, and it was not possible 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 highest number of failures occurring.

In this paper, the evaluation of the history of recorded faults is in relation to the year 2020 (from 1 January to 31 December), and this history corresponds to the sum of all the faults of the distribution branches in the Lages region (Brazil) based on data provided by *Centrais Elétricas de Santa Catarina* (CELESC). In total, there were 366 days recorded, considering that the year 2020 was a leap year. Figure 1 presents the sum of the alarms regarding faults per day in this period.

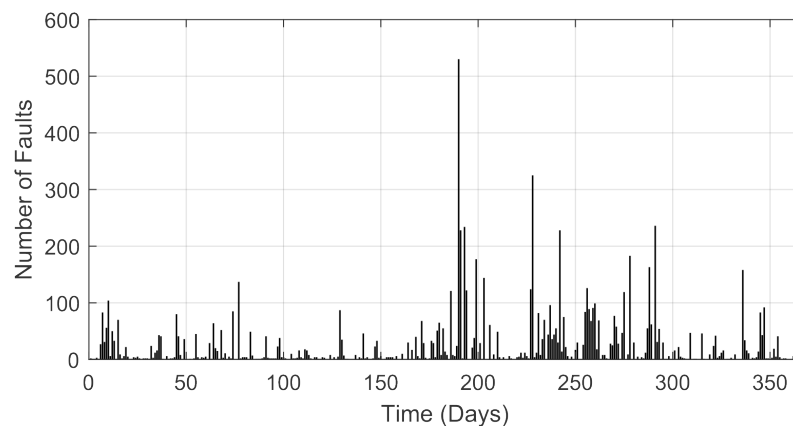


Figure 1. Failures registered in the power grid in 2020 (Lages region).

3. Wavelet LSTM

The wavelet LSTM method is a combination of the wavelet transform and long short-term memory. This approach has been widely used for fault diagnosis, as presented in the work of Sabir et al. [64] and Jalayer, Orsenigo, and Vercellis [65] for electrical machines, and especially for rolling bearings in the work of Tan et al. [66]. In scenarios that involve time series forecasting, its application can be extended to the Internet of Things [67–69], industry applications [70–72], and sustainability [73].

To apply the wavelet LSTM model here, initially, the time series passes through the wavelet filter to reduce the noise and nonlinearities. After the signal is decomposed and reconstructed, the LSTM receives the filtered signal and performs the prediction. The complete structure of this approach is presented in Figure 2 and will be explained in this section.

The structure of the proposed method (presented in Figure 2) can be divided into steps as follows. In the first step (A), the original input signal (shown in Figure 1) is loaded. In the next step, the wavelet transform is applied, which is divided into two parts, namely the signal decomposition (B) and its reconstruction (C) into the time series. In the next step, the denoising signal is normalized (D), where the variation in the number of failures is evaluated in relation to all recorded faults in the considered period. In the last step, the LSTM model is used to perform the time series prediction one step ahead (E).

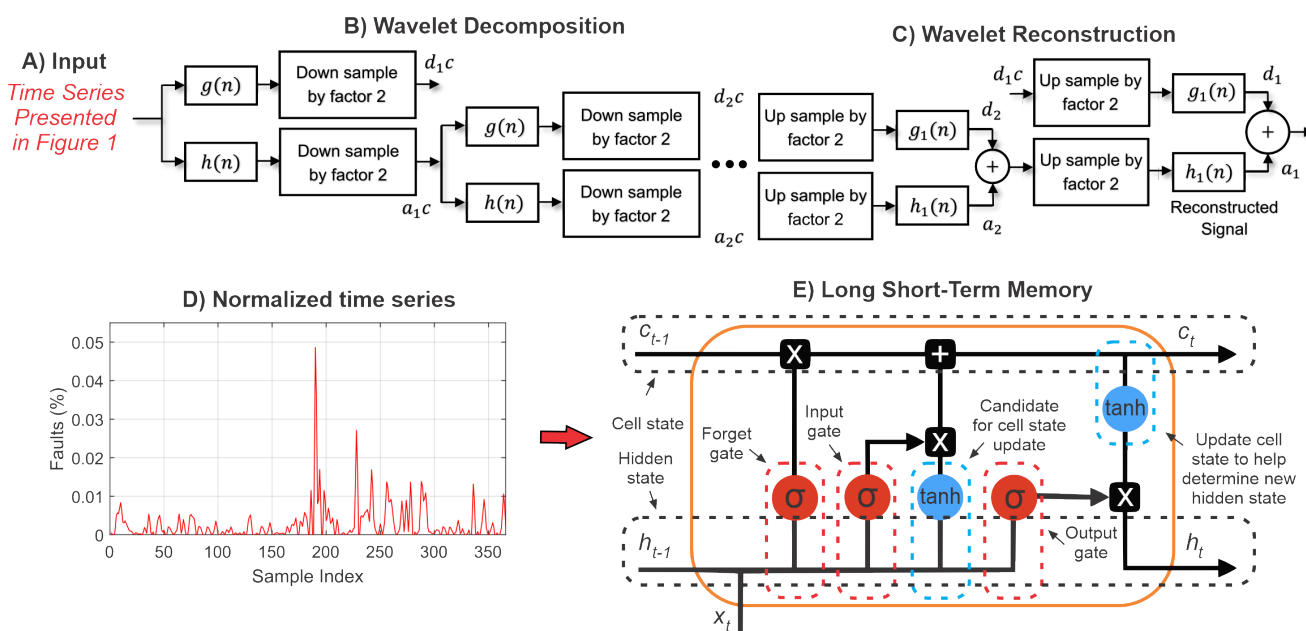


Figure 2. Structure of the wavelet long short-term memory model.

To use the wavelet transform, the signal was first decomposed using the wavelet packet transform (WPT) method to obtain the energy coefficient of the signal. This procedure considers both sides of the spectrum (high and low frequencies). The decomposition may be denoted by

$$W_{\Psi,x}(A, B) = \frac{1}{\sqrt{A}} \int_{-\infty}^{+\infty} x(t) \Psi * \left(\frac{t - B}{A} \right) dt, \quad A \neq 0 \tag{1}$$

where $x(t)$ is the signal to be decomposed, Ψ is the time-based function (mother wavelet), and A and B are the scale and displacement parameters, respectively [74]. Given a discretization, the high-pass filter $g(n)$ is

$$g(n) = h(2N - 1 - n). \tag{2}$$

where $h(n)$ is the low-pass filter. Thereby, the mother wavelet and the scaling function (Φ) are given by

$$\Psi(n) = \sum_{i=0}^{N-1} g(i)\Phi(2n - i), \tag{3}$$

$$\Phi(n) = \sum_{i=0}^{N-1} h(i)\Phi(2n - i). \tag{4}$$

The WPT performs a new decomposition on each interaction using the coefficients from the previous iterations and therefore indicates that the total number of coefficients is determined by the number of iterations. Each wavelet packet coefficient can be determined based on its frequency level. The WPT decomposes the all elements of the frequencies, and thus its use results in both low- and high-frequency components. By using the tree structure created by the approximation decomposition coefficients, an optimal binary value is obtained. An example of the tree structure for wavelet decomposition is presented in Figure 3.

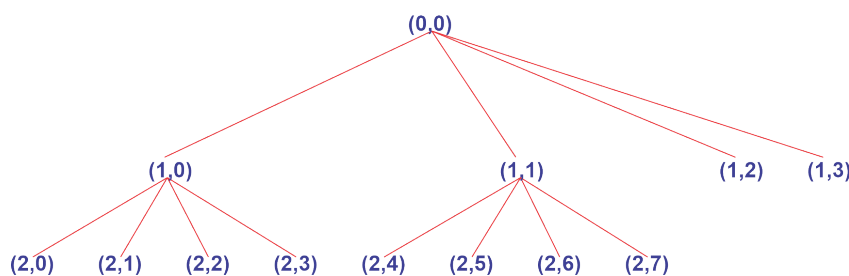


Figure 3. Tree decomposition.

As can be seen, paths 1, 2 and 1, 3 are not used after optimizing the structure, resulting in an optimized decomposition. After wavelet packet decomposition based on the optimum binary wavelet packet tree, the signal is reconstructed while considering the number of defined nodes. With the reconstructed filtered signal, a time series is obtained that is used for LSTM forecast evaluation.

LSTM is a recurrent neural network that has feedback, allowing the model to remember distant values. For the time series forecasting starting from D samples, we use

$$x(t - (D - 1)\Delta), \dots, x(t - \Delta), x(t) \tag{5}$$

to predict the future value

$$x(t + P), \tag{6}$$

where P represents the steps forward and Δ is the period of the samples. In this paper, Δ is equal to 1 day, where all faults of the same day were summed. In this paper, one-step-ahead prediction was used ($P = 1$).

LSTM is capable of understanding order dependence in problems that require sequence prediction, making it promising for time series forecasting [75]. In an LSTM algorithm, each cell is divided into three gates: the input (l_t), output (o_t), and forgetting (f_t) gates [76], where f_t controls how much information will be forgotten and how much will be remembered, and the useful information for the states is added through the l_t and o_t to determine how much of the current state must be assigned to the output [77]. The LSTM can be defined by the following equations:

$$\begin{aligned} l_t &= \sigma_g(W_l x_t + R_l h_{t-1} + b_l), \\ f_t &= \sigma_g(W_f x_t + R_f h_{t-1} + b_f), \\ o_t &= \sigma_g(W_o x_t + R_o h_{t-1} + b_o). \end{aligned} \tag{7}$$

where b is the polarization matrix, R and W are earnings matrices, and σ_g is the activation function [78]. LSTM has an input activation function G and output activation function H , which are used to update the cell and the hidden state as given in the following equations:

$$\begin{aligned}\tilde{c}_t &= G(W_c x_t + R_c h_{t-1} + b_c), \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t, \\ h_t &= o_t \circ H(c_t).\end{aligned}\quad (8)$$

To find the predicted values of future time steps, the training response sequences are shifted by a time step. Thus, at each time step in the input sequence, the net learns to forecast the following time-step value. To feed the compared models, normalization is performed using the values as percentages in relation to the total number of failures recorded in the evaluated period.

To improve the efficiency of the proposed model, three optimizers were evaluated. Stochastic gradient descent (SGD) updates the parameters of the neural net to minimize the loss function by taking small steps in each iteration (i) in the direction of the negative loss gradient:

$$\theta_{i+1} = \theta_i - \alpha \nabla E(\theta_i). \quad (9)$$

SGD with momentum (SGDM) helps accelerate the gradient vectors in the correct directions, resulting in faster convergence [79]. Here, α is the learning rate, θ is the vector of parameters, and $E(\theta_i)$ is the loss function to be optimized.

Root mean squared propagation (RMSProp) employs learning rates that differ per parameter and can adapt automatically to the loss function that is optimized. In this way, the algorithm keeps a moving average of the squares of the parameter, which is calculated as follows:

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2) [\nabla f(x_i)]^2. \quad (10)$$

where β_2 is the decay rate of the moving average. The algorithm takes the moving average to normalize the parameter updates individually [80]:

$$x_{i+1} = x_i - \frac{\alpha \nabla f(x_i)}{\sqrt{v_i} + \varepsilon}. \quad (11)$$

The adaptive moment estimation (ADAM) method computes the learning rates for each parameter. The downward averages of the last m_i and the squared gradients of the last v_i are calculated as follows:

$$m_i = \beta_1 m_{i-1} + (1 - \beta_1) \nabla f(x_i), \quad (12)$$

$$v_i = \beta_2 m_{i-1} + (1 - \beta_2) [\nabla f(x_i)]^2. \quad (13)$$

ADAM works by using moving averages to update the parameters of the network in the following way:

$$x_{i+1} = x_i - \frac{\alpha m_i}{\sqrt{v_i} + \varepsilon}. \quad (14)$$

Considered Measures

In this paper, the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were considered, given by

$$\text{RMSE} = \sqrt{\frac{1}{\eta} \sum_{i=1}^{\eta} (y_i - \hat{y}_i)^2}, \quad (15)$$

$$\text{MAE} = \frac{1}{\eta} \sum_{i=1}^{\eta} |y_i - \hat{y}_i|, \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (17)$$

where y_i is the observed value, \hat{y}_i is the predicted output, and \bar{y}_i is the average of the observed value [81].

The statistical evaluation was performed with 50 runs using the same parameter configuration, where the mean, median, and standard deviation were evaluated. The simulations were computed using an Intel Core I5-7400 with 20 GB of RAM using MATLAB software. For a comparative assessment, the adaptive neuro-fuzzy inference system (ANFIS) [82,83], group method of data handling (GMDH) [84], bagging [85], random subspace [86], and stacking [87] ensemble learning methods were evaluated.

4. Analysis of Results

The first evaluation regarded the analysis of the time series forecast performed in relation to the percentage of data used for training and testing of the neural network. The evaluation of this parameter is important because it can be used to define the minimum amount of data needed for training the model. The best results presented in this section are highlighted in bold.

The evaluation results are shown in Table 2, considering a training ratio from 50 to 90 percent. The test used the difference equivalent percentage to complete the dataset, in which the validation stage was not considered.

Table 2. Evaluating the influence of the training and testing relationship.

| Train/Test | RMSE | MAE | R ² | Time (s) |
|------------|---|---|----------------|--------------|
| 50/50 | 8.02×10^{-3} | 3.03×10^{-3} | 0.1516 | 17.21 |
| 60/40 | 6.24×10^{-3} | 7.47×10^{-4} | 0.1681 | 18.53 |
| 70/30 | 4.73×10^{-3} | 3.59×10^{-4} | 0.0325 | 18.64 |
| 80/20 | 3.60×10^{-3} | 1.18×10^{-3} | 0.2779 | 20.47 |
| 90/10 | 4.29×10^{-3} | 5.49×10^{-5} | 0.0884 | 19.68 |

Using 80% of the data for training and 20% of the data for testing gave the best RMSE and R² values, and hence this ratio was used in further analysis. As can be seen in Figure 4, there was major difficulty in predicting the data due to the nonlinearities in the time series, considering that in some cases, there were several failures in a short period of time.

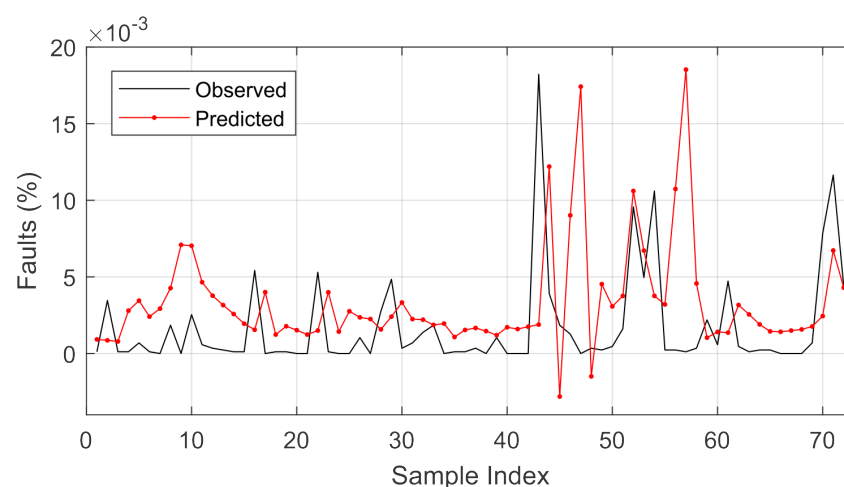


Figure 4. Preliminary analysis of fault prediction capability.

The failures that occurred after the middle of the year were due to the rainy season, which starts after winter in the Southern Hemisphere. The greater presence of bad weather

conditions favors the development of faults in electrical power distribution systems. In the following analysis, the optimizer and the number of hidden units are evaluated (see Table 3).

Table 3. Assessment of the number of hidden units (HUs) using different optimizers.

| Optimizer | HU | RMSE | MAE | R ² | Time (s) |
|-----------|------|---|---|----------------|--------------|
| SGDM | 50 | 3.49×10^{-3} | 9.12×10^{-4} | 0.1957 | 17.74 |
| | 100 | 3.50×10^{-3} | 9.12×10^{-4} | 0.2035 | 18.29 |
| | 200 | 3.51×10^{-3} | 1.17×10^{-3} | 0.2089 | 19.58 |
| | 500 | 3.44×10^{-3} | 9.18×10^{-4} | 0.1624 | 25.06 |
| | 1000 | 3.44×10^{-3} | 9.37×10^{-4} | 0.1623 | 36.66 |
| ADAM | 50 | 4.79×10^{-3} | 1.77×10^{-3} | - | 17.94 |
| | 100 | 7.69×10^{-3} | 4.58×10^{-3} | - | 19.94 |
| | 200 | 3.98×10^{-3} | 6.51×10^{-4} | 0.5554 | 19.21 |
| | 500 | 3.96×10^{-3} | 8.29×10^{-4} | 0.5469 | 25.55 |
| | 1000 | 3.94×10^{-3} | 1.52×10^{-3} | 0.5242 | 35.22 |
| RMSprop | 50 | 6.22×10^{-3} | 3.01×10^{-3} | - | 21.58 |
| | 100 | 5.65×10^{-3} | 1.22×10^{-3} | - | 19.14 |
| | 200 | 3.93×10^{-3} | 6.64×10^{-4} | 0.5212 | 21.64 |
| | 500 | 3.35×10^{-3} | 1.00×10^{-3} | 0.1074 | 27.51 |
| | 1000 | 3.16×10^{-3} | 2.96×10^{-5} | 0.0190 | 36.50 |

In this evaluation, the SGDM optimizer had more stable results for the coefficient of determination, presenting a smaller variance in relation to the change in hidden units. When comparing all the models, the best results occurred using 200 hidden units, considering the coefficient of determination. In some cases, it was not possible to measure the coefficient of determination due to the high intensity of the variation in the prediction using the ADAM and RMSprop optimizers. Considering that there was a large variation in the values, statistical analysis was performed (and will be presented) using 200 hidden units.

A wavelet transform for noise reduction was added for the following analysis. This transform should be used with caution, as it can result in a loss of features for the signal. Figure 5 shows the results of the wavelet transform relative to the original signal using one node, and Figure 6 shows this comparison using two nodes.

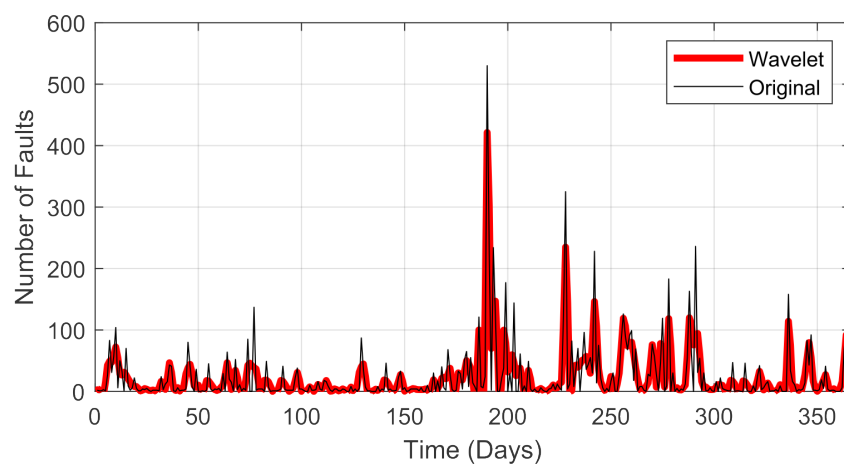


Figure 5. Evaluation of the wavelet transform with one node.

The use of two nodes considerably altered the response of the transform, hindering practical application. When three nodes or more are used, the signal loses its characteristics, and this was not considered in this paper.

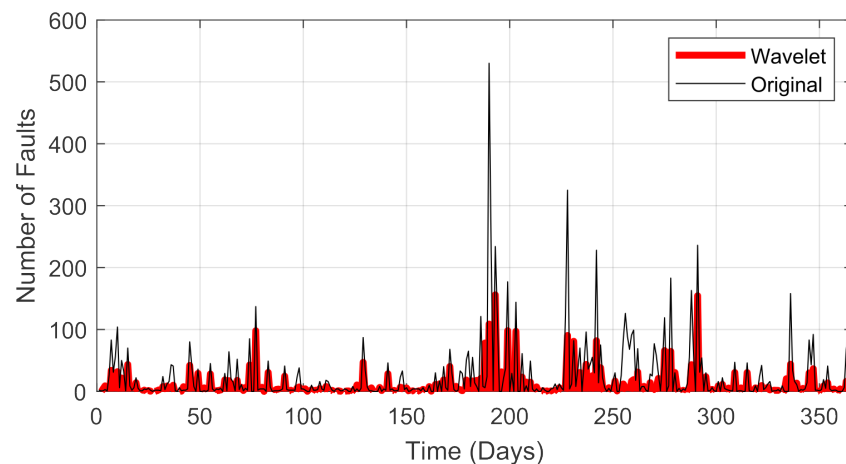


Figure 6. Evaluation of the wavelet transform with two nodes.

The complete analysis of the wavelet transform depth variation is presented in Table 4. Considering that there was an error value that made the prediction not suitable for analysis, the use of two nodes was disregarded after this evaluation. As mentioned earlier this can also be observed when the wavelet transform was compared in this configuration in relation to the original signal (see Figure 6).

Table 4. Assessment of depth using different one and two nodes.

| Nodes | Depth | RMSE | MAE | R ² | Time (s) |
|-------|-------|---|---|----------------|--------------|
| 1 | 1 | 2.19×10^{-3} | 4.54×10^{-4} | 0.3375 | 23.14 |
| | 2 | 2.16×10^{-3} | 4.69×10^{-4} | 0.3547 | 22.13 |
| | 3 | 2.22×10^{-3} | 5.28×10^{-4} | 0.3132 | 23.61 |
| | 4 | 2.16×10^{-3} | 3.53×10^{-4} | 0.3509 | 22.76 |
| | 5 | 2.17×10^{-3} | 3.70×10^{-4} | 0.3493 | 21.71 |
| 2 | 1 | 2.22×10^{10} | 1.24×10^9 | 0.6888 | 21.34 |
| | 2 | 2.13×10^{10} | 1.24×10^9 | 0.7155 | 22.75 |
| | 3 | 2.45×10^{10} | 6.79×10^9 | 0.6214 | 18.05 |
| | 4 | 2.32×10^{10} | 1.90×10^9 | 0.6601 | 18.74 |
| | 5 | 2.44×10^{10} | 6.58×10^9 | 0.6244 | 18.69 |

The best coefficient of determination was reached using a depth equal to two in the wavelet transform, getting close to the best MAE value, which happened using a depth equal to four. Considering these results, a depth equal to two was used for statistical analysis, which is presented in Table 5.

Table 5. Statistical evaluation.

| Model | Optimizer | Mean | Median | Std Dev. |
|--------------|-----------|---|---|---|
| LSTM | SGDM | 3.50×10^{-3} | 3.49×10^{-3} | 2.62×10^{-5} |
| | ADAM | 7.87×10^{-3} | 8.04×10^{-3} | 2.63×10^{-3} |
| | RMSprop | 4.90×10^{-3} | 4.95×10^{-3} | 6.41×10^{-4} |
| Wavelet LSTM | SGDM | 2.19×10^{-3} | 2.19×10^{-3} | 2.76×10^{-5} |
| | ADAM | 1.79×10^{-3} | 1.27×10^{-3} | 2.27×10^{-3} |
| | RMSprop | 1.67×10^{-3} | 1.54×10^{-3} | 7.15×10^{-4} |

The wavelet LSTM model was superior in all comparative analyses to the LSTM model with respect to the RMSE. Even when varying the optimizer, the wavelet LSTM model

showed promise for the analysis in question. The best average RMSE result was obtained using the RMSprop optimizer in the wavelet LSTM model. The comparison between the prediction result and the original signal is presented in Figure 7.

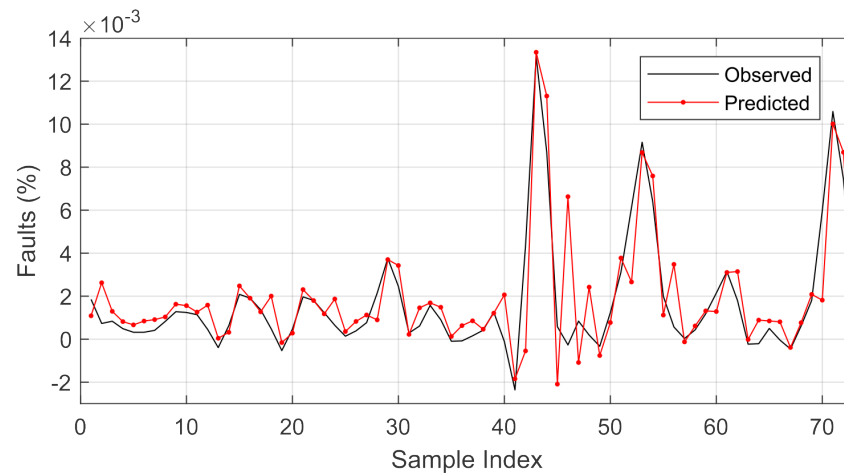


Figure 7. Comparison of the prediction using the wavelet LSTM model for the observed values.

The results of the comparison of the predicted and observed values (shown in Figure 7) mean that it would be possible to estimate the next day's failures based on the recorded history (considering that the prediction is one step ahead, which corresponds to 1 day). As the forecast is accomplished one step ahead, the history of the data recorded until the next expected forecast is used to predict the next one. Considering that the forecast is performed in relation to the sum of the failure records that occurred over time, it would be possible to estimate the number of failures for the next day.

Whereas failures are related to weather conditions, there is a tendency for them to increase depending on the time of year, which is the focus of this research. After the highest accumulated value of the number of failures, there was an oscillation in the prediction, something that was expected due to this abrupt variation in the time series.

This statement is supported based on the error calculated by the difference between the predicted and observed values, as presented in Figure 8.

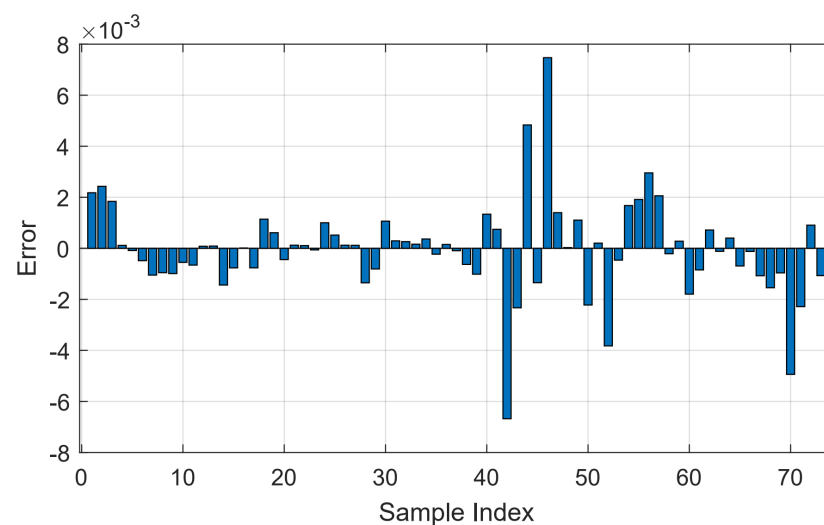


Figure 8. Error given by the difference between predicted and observed values.

Once the best parameter configurations for the LSTM wavelet model were defined, benchmarking, presented in Table 6, was performed, aiming to compare the proposed model with well-established models.

Table 6. Benchmarking.

| Method | Structure | RMSE | MAE | R ² | Time (s) |
|--------------|-------------------|---|---|----------------|-------------|
| Ensemble | Bagging | 3.33×10^{-3} | 8.55×10^{-4} | 0.0830 | 6.16 |
| | Random Subspace | 3.38×10^{-3} | 1.03×10^{-3} | 0.1149 | 2.18 |
| | Stacking | 4.24×10^{-3} | 2.48×10^{-3} | 0.7536 | 1.17 |
| GMDH | 3 Max Layers | 3.60×10^{-3} | 5.24×10^{-4} | 0.2621 | 0.60 |
| | 5 Max Layers | 3.62×10^{-3} | 7.87×10^{-4} | 0.2796 | 1.23 |
| | 10 Max Layers | 3.52×10^{-3} | 7.65×10^{-4} | 0.2088 | 2.52 |
| ANFIS | FCM Clustering | 3.68×10^{-3} | 5.90×10^{-4} | 0.3226 | 4.86 |
| | Grid Partitioning | 3.91×10^{-3} | 4.23×10^{-4} | 0.4866 | 12.39 |
| | Influence Radius | - | - | - | - |
| Wavelet LSTM | RMSprop Opt. | 1.45×10^{-3} | 4.11×10^{-5} | 0.7064 | 21.032 |

The LSTM wavelet model presented the best results regarding the RMSE and MAE, despite requiring the longest time to converge. The ensemble stacking model presented the best coefficient of determination with the shortest time required for convergence of the ensemble models, with this being an indication that this model may be promising in this evaluation. The other models compared presented similar RMSE results, not being superior to the proposed model. The ANFIS model using the radius of influence did not converge because the time series used had insufficient data for this structure.

5. Conclusions

The prediction of faults in an electrical distribution system is necessary to ensure the operation of the power grid. By analyzing the variation of a time series, it is possible to verify the presence of a higher number of failures during a certain time and thus define a more effective correction strategy. Based on time series forecasting, an electric utility company may know when there are higher chances to have faults before they happen and thus have more defined strategies to deal with them.

It is noticeable that there was difficulty in this prediction due to the large variation in the number of failures in some seasons of the year, which was mainly related to the rainy season. Using traditional models, the forecast results were ineffective, so it was necessary to combine algorithms and create a hybrid model to meet the needs of the problem.

The wavelet LSTM model showed better results in all analyses compared with the standard LSTM model, including better results in the statistical analysis, being an appropriate model for the evaluation presented in this paper. Using this model, it is possible to have failure prediction indicators that can help the organization of maintenance teams, thus reducing the response time when a disruptive failure occurs.

The LSTM model has a high predictive ability for chaotic time series, which was the case in the need for solving this task. Using the LSTM model without the inclusion of additional filters made it impossible to perform an acceptable prediction, considering that the abrupt variation in the time series of the number of failures made it necessary to perform a smoothing of the series through the inclusion of a filter. The results showed that the wavelet LSTM model was acceptable for the analysis in question, being superior to the ensemble learning methods, GMDH, and ANFIS.

Future work can be conducted regarding the type of failure. The failures can vary, for instance, because of direct contact with the grid or leakage current. Specific analysis on which type of failure occurs more frequently and how to avoid this type of failure is promising work to be carried out in the future.

Author Contributions: Writing—original draft and methodology, N.W.B.; supervision and writing—review and editing, M.S.M.C.; writing—review and editing, software, and formal analysis, S.F.S.; project administration and supervision, V.R.Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by national funds through the Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) by the project “VALORIZA—Research Centre for Endogenous Resource Valorization” under Grant UIDB/05064/2020 and Grant UIDB/04111/2020 and in part by the Instituto Lusófono de Investigação e Desenvolvimento (ILIND) under Project COFAC/ILIND/COPELABS/3/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this paper were provided by *Centrais Elétricas de Santa Catarina* 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 21 October 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations and Symbols

The following abbreviations and symbols are used in this manuscript:

| Abbreviation | Meaning |
|--------------|---|
| ADAM | Adaptive moment estimation |
| ANFIS | Adaptive neuro-fuzzy inference system |
| CNN | Convolutional neural network |
| GMDH | Group method of data handling |
| LSTM | Long short-term memory |
| MAE | Mean absolute error |
| RMSE | Root mean square error |
| RMSProp | Root mean squared propagation |
| SGD | Stochastic gradient descent |
| SGDM | Stochastic gradient descent with momentum |
| WPT | Wavelet packet transform |
| YOLO | You Only Look Once |
| Symbol | Meaning |
| b | Polarization matrix |
| f | Forgetting gate |
| g | High-pass filter |
| i | Iteration |
| h | Low-pass filter |
| m | Elementary moving average |
| n | Number of samples |
| o | Output gate |
| t | Time step |
| v | Moving average of the squares of the elements |
| x | Original signal |
| y | Observed value |
| \hat{y} | Predicted output |
| \bar{y} | Average of the observed value |
| A | Scale parameter |
| B | Displacement parameter |
| E | Loss function |
| G | Input activation function |
| H | Output activation function |
| P | Steps forward |
| R and W | Earning matrices |
| R^2 | Coefficient of determination |

| | |
|----------|-----------------------|
| α | Learning rate |
| β | Decay rate |
| η | Number of predictions |
| σ | Activation function |
| ι | Input gate |
| θ | Vector of parameters |
| Δ | Period of the samples |
| Ψ | Mother wavelet |

References

- Araya, J.; Montaña, J.; Schurch, R. Electric Field Distribution and Leakage Currents in Glass Insulator Under Different Altitudes and Pollutions Conditions using FEM Simulations. *IEEE Lat. Am. Trans.* **2021**, *19*, 1278–1285. <https://doi.org/10.1109/TLA.2021.9475858>.
- Sun, J.; Yang, Q.; Cui, H.; Ran, J.; Liu, H. Distribution Line Fault Location With Unknown Fault Impedance Based on Electromagnetic Time Reversal. *IEEE Trans. Electromagn. Compat.* **2021**, *63*, 1921–1929. <https://doi.org/10.1109/TEMC.2021.3097105>.
- Liu, Z.; Chen, H.; Hu, Z.; Li, Y.; Wu, X.; Peng, H. Fault Detection System for 500 kV AC Fault Current Limiter Based on High-Coupled Split Reactor. *IEEE Trans. Appl. Supercond.* **2021**, *31*, 1–7. <https://doi.org/10.1109/TASC.2021.3101745>.
- Stefenon, S.F.; Bruns, R.; Sartori, A.; Meyer, L.H.; Ovejero, R.G.; Leithardt, V.R.Q. Analysis of the Ultrasonic Signal in Polymeric Contaminated Insulators through Ensemble Learning Methods. *IEEE Access* **2022**, *10*, 33980–33991. <https://doi.org/10.1109/ACCESS.2022.3161506>.
- Medeiros, A.; Sartori, A.; Stefenon, S.F.; Meyer, L.H.; Nied, A. Comparison of artificial intelligence techniques to failure prediction in contaminated insulators based on leakage current. *J. Intell. Fuzzy Syst.* **2021**, *42*, 3285–3298. <https://doi.org/10.3233/JIFS-211126>.
- Rhif, M.; Ben Abbès, A.; Farah, I.R.; Martínez, B.; Sang, Y. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Appl. Sci.* **2019**, *9*, 1345. <https://doi.org/10.3390/app9071345>.
- Ameid, T.; Menacer, A.; Talhaoui, H.; Azzoug, Y. Discrete wavelet transform and energy eigen value for rotor bars fault detection in variable speed field-oriented control of induction motor drive. *ISA Trans.* **2018**, *79*, 217–231. <https://doi.org/10.1016/j.isatra.2018.04.019>.
- Stefenon, S.F.; Kasburg, C.; Nied, A.; Klaar, A.C.R.; Ferreira, F.C.S.; Branco, N.W. Hybrid deep learning for power generation forecasting in active solar trackers. *IET Gener. Transm. Distrib.* **2020**, *14*, 5667–5674. <https://doi.org/10.1049/iet-gtd.2020.0814>.
- Chandra, R.; Goyal, S.; Gupta, R. Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction. *IEEE Access* **2021**, *9*, 83105–83123. <https://doi.org/10.1109/ACCESS.2021.3085085>.
- Hu, Y.; Sun, X.; Nie, X.; Li, Y.; Liu, L. An Enhanced LSTM for Trend Following of Time Series. *IEEE Access* **2019**, *7*, 34020–34030. <https://doi.org/10.1109/ACCESS.2019.2896621>.
- Ma, C.; Dai, G.; Zhou, J. Short-Term Traffic Flow Prediction for Urban Road Sections Based on Time Series Analysis and LSTM BILSTM Method. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 5615–5624. <https://doi.org/10.1109/TITS.2021.3055258>.
- Zhang, S.; Wang, Y.; Liu, M.; Bao, Z. Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM. *IEEE Access* **2018**, *6*, 7675–7686. <https://doi.org/10.1109/ACCESS.2017.2785763>.
- Kim, W.H.; Kim, J.Y.; Chae, W.K.; Kim, G.; Lee, C.K. LSTM-Based Fault Direction Estimation and Protection Coordination for Networked Distribution System. *IEEE Access* **2022**, *10*, 40348–40357. <https://doi.org/10.1109/ACCESS.2022.3166836>.
- Qiao, M.; Yan, S.; Tang, X.; Xu, C. Deep Convolutional and LSTM Recurrent Neural Networks for Rolling Bearing Fault Diagnosis under Strong Noises and Variable Loads. *IEEE Access* **2020**, *8*, 66257–66269. <https://doi.org/10.1109/ACCESS.2020.2985617>.
- Stefenon, S.F.; Freire, R.Z.; Meyer, L.H.; Corso, M.P.; Sartori, A.; Nied, A.; Klaar, A.C.R.; Yow, K.C. Fault detection in insulators based on ultrasonic signal processing using a hybrid deep learning technique. *IET Sci. Meas. Technol.* **2020**, *14*, 953–961. <https://doi.org/10.1049/iet-smt.2020.0083>.
- Ma, Y.; Oslebo, D.; Maqsood, A.; Corzine, K. DC Fault Detection and Pulsed Load Monitoring Using Wavelet Transform-Fed LSTM Autoencoders. *IEEE J. Emerg. Sel. Top. Power Electron.* **2021**, *9*, 7078–7087. <https://doi.org/10.1109/JESTPE.2020.3019382>.
- Furse, C.M.; Kafal, M.; Razzaghi, R.; Shin, Y.J. Fault Diagnosis for Electrical Systems and Power Networks: A Review. *IEEE Sens. J.* **2021**, *21*, 888–906. <https://doi.org/10.1109/JSEN.2020.2987321>.
- Li, B.; Cui, H.; Li, B.; Wen, W.; Dai, D. A permanent fault identification method for single-pole grounding fault of overhead transmission lines in VSC-HVDC grid based on fault line voltage. *Int. J. Electr. Power Energy Syst.* **2020**, *117*, 105603. <https://doi.org/10.1016/j.ijepes.2019.105603>.
- Wadi, M.; Elmasry, W. An Anomaly-based Technique for Fault Detection in Power System Networks. In Proceedings of the 2021 International Conference on Electric Power Engineering, Gaza, 23–24 March 2021; ICEPE: Palestine, 2021; pp. 1–6. <https://doi.org/10.1109/ICEPE-P51568.2021.9423479>.
- Wu, J.; Zhang, L.; Bai, Y.; Reniers, G. A safety investment optimization model for power grid enterprises based on System Dynamics and Bayesian network theory. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108331. <https://doi.org/10.1016/j.res.2022.108331>.
- Sadi, M.A.H.; AbuHussein, A.; Shoeb, M.A. Transient Performance Improvement of Power Systems Using Fuzzy Logic Controlled Capacitive-Bridge Type Fault Current Limiter. *IEEE Trans. Power Syst.* **2021**, *36*, 323–335. <https://doi.org/10.1109/TPWRS.2020.3003294>.
- Rigatos, G.; Serpanos, D.; Zervos, N. Detection of Attacks Against Power Grid Sensors Using Kalman Filter and Statistical Decision Making. *IEEE Sens. J.* **2017**, *17*, 7641–7648. <https://doi.org/10.1109/JSEN.2017.2661247>.

23. Stefenon, S.F.; Yow, K.C.; Nied, A.; Meyer, L.H. Classification of distribution power grid structures using inception v3 deep neural network. *Electr. Eng.* **2022**, 1–13. <https://doi.org/10.1007/s00202-022-01641-1>.
24. Haj, Y.E.; El-Hag, A.H.; Ghunem, R.A. Application of Deep-Learning via Transfer Learning to Evaluate Silicone Rubber Material Surface Erosion. *IEEE Trans. Dielectr. Electr. Insul.* **2021**, *28*, 1465–1467. <https://doi.org/10.1109/TDEI.2021.009617>.
25. Stefenon, S.F.; Singh, G.; Yow, K.C.; Cimatti, A. Semi-ProtoPNet Deep Neural Network for the Classification of Defective Power Grid Distribution Structures. *Sensors* **2022**, *22*, 4859. <https://doi.org/10.3390/s22134859>.
26. Zhao, M.; Barati, M. A Real-Time Fault Localization in Power Distribution Grid for Wildfire Detection through Deep Convolutional Neural Networks. *IEEE Trans. Ind. Appl.* **2021**, *57*, 4316–4326. <https://doi.org/10.1109/TIA.2021.3083645>.
27. Mantach, S.; Lutfi, A.; Moradi Tavasani, H.; Ashraf, A.; El-Hag, A.; Kordi, B. Deep Learning in High Voltage Engineering: A Literature Review. *Energies* **2022**, *15*, 5005. <https://doi.org/10.3390/en15145005>.
28. Corso, M.P.; Perez, F.L.; Stefenon, S.F.; Yow, K.C.; García Ovejero, R.; Leithardt, V.R.Q. Classification of Contaminated Insulators Using k-Nearest Neighbors Based on Computer Vision. *Computers* **2021**, *10*, 112. <https://doi.org/10.3390/computers10090112>.
29. Wu, H.; Hu, Y.; Wang, W.; Mei, X.; Xian, J. Ship Fire Detection Based on an Improved YOLO Algorithm with a Lightweight Convolutional Neural Network Model. *Sensors* **2022**, *22*, 7420. <https://doi.org/10.3390/s22197420>.
30. Vieira, J.C.; Sartori, A.; Stefenon, S.F.; Perez, F.L.; de Jesus, G.S.; Leithardt, V.R.Q. Low-Cost CNN for Automatic Violence Recognition on Embedded System. *IEEE Access* **2022**, *10*, 25190–25202. <https://doi.org/10.1109/ACCESS.2022.3155123>.
31. Hou, L.; Chen, C.; Wang, S.; Wu, Y.; Chen, X. Multi-Object Detection Method in Construction Machinery Swarm Operations Based on the Improved YOLOv4 Model. *Sensors* **2022**, *22*, 7294. <https://doi.org/10.3390/s22197294>.
32. Liu, C.; Wu, Y.; Liu, J.; Sun, Z. Improved YOLOv3 Network for Insulator Detection in Aerial Images with Diverse Background Interference. *Electronics* **2021**, *10*, 771. <https://doi.org/10.3390/electronics10070771>.
33. Sadykova, D.; Pernebayeva, D.; Bagheri, M.; James, A. IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV Imaging. *IEEE Trans. Power Deliv.* **2020**, *35*, 1599–1601. <https://doi.org/10.1109/TPWRD.2019.2944741>.
34. Li, X.; Su, H.; Liu, G. Insulator Defect Recognition Based on Global Detection and Local Segmentation. *IEEE Access* **2020**, *8*, 59934–59946. <https://doi.org/10.1109/ACCESS.2020.2982288>.
35. Wen, Q.; Luo, Z.; Chen, R.; Yang, Y.; Li, G. Deep Learning Approaches on Defect Detection in High Resolution Aerial Images of Insulators. *Sensors* **2021**, *21*, 1033. <https://doi.org/10.3390/s21041033>.
36. Chen, H.; He, Z.; Shi, B.; Zhong, T. Research on Recognition Method of Electrical Components Based on YOLO V3. *IEEE Access* **2019**, *7*, 157818–157829. <https://doi.org/10.1109/ACCESS.2019.2950053>.
37. Liu, C.; Wu, Y.; Liu, J.; Han, J. MTI-YOLO: A Light-Weight and Real-Time Deep Neural Network for Insulator Detection in Complex Aerial Images. *Energies* **2021**, *14*, 1426. <https://doi.org/10.3390/en14051426>.
38. Liu, J.; Liu, C.; Wu, Y.; Xu, H.; Sun, Z. An Improved Method Based on Deep Learning for Insulator Fault Detection in Diverse Aerial Images. *Energies* **2021**, *14*, 4365. <https://doi.org/10.3390/en14144365>.
39. Hu, X.; Zhou, Y. Insulator defect detection in power inspection image using focal loss based on YOLO v4. In Proceedings of the International Conference on Artificial Intelligence, Virtual Reality, and Visualization (AIVRV 2021), Sanya, China, 16 December 2021; SPIE: Sanya, China, 2021; Volume 12153, pp. 90–95. <https://doi.org/10.1117/12.2626694>.
40. Xing, Z.; Chen, X. Lightweight algorithm of insulator identification applicable to electric power engineering. *Energy Rep.* **2022**, *8*, 353–362. <https://doi.org/10.1016/j.egyrs.2022.01.209>.
41. Stefenon, S.F.; Seman, L.O.; Pavan, B.A.; Ovejero, R.G.; Leithardt, V.R.Q. Optimal design of electrical power distribution grid spacers using finite element method. *IET Gener. Transm. Distrib.* **2022**, *16*, 1865–1876. <https://doi.org/10.1049/gtd2.12425>.
42. Yang, C.; Chen, T.; Yang, B.; Zhang, X.; Fan, S. Experimental study of tree ground fault discharge characteristics of 35 kV transmission lines. In Proceedings of the 2021 IEEE Sustainable Power and Energy Conference (iSPEC), Nanjing, China, 23–25 December 2021; Volume 1, pp. 2883–2891. <https://doi.org/10.1109/iSPEC53008.2021.9735502>.
43. Stefenon, S.F.; Neto, C.S.F.; Coelho, T.S.; Nied, A.; Yamaguchi, C.K.; Yow, K.C. Particle swarm optimization for design of insulators of distribution power system based on finite element method. *Electr. Eng.* **2022**, *104*, 615–622. <https://doi.org/10.1007/s00202-021-01332-3>.
44. Salem, A.A.; Abd-Rahman, R.; Al-Gailani, S.A.; Kamarudin, M.S.; Ahmad, H.; Salam, Z. The Leakage Current Components as a Diagnostic Tool to Estimate Contamination Level on High Voltage Insulators. *IEEE Access* **2020**, *8*, 92514–92528. <https://doi.org/10.1109/ACCESS.2020.2993630>.
45. Stefenon, S.F.; Seman, L.O.; Sopelsa Neto, N.F.; Meyer, L.H.; Nied, A.; Yow, K.C. Echo state network applied for classification of medium voltage insulators. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107336. <https://doi.org/10.1016/j.ijepes.2021.107336>.
46. Cao, B.; Wang, L.; Yin, F. A Low-Cost Evaluation and Correction Method for the Soluble Salt Components of the Insulator Contamination Layer. *IEEE Sens. J.* **2019**, *19*, 5266–5273. <https://doi.org/10.1109/JSEN.2019.2902192>.
47. Stefenon, S.F.; Corso, M.P.; Nied, A.; Perez, F.L.; Yow, K.C.; Gonzalez, G.V.; Leithardt, V.R.Q. Classification of insulators using neural network based on computer vision. *IET Gener. Transm. Distrib.* **2021**, *16*, 1096–1107. <https://doi.org/10.1049/gtd2.12353>.
48. Hou, C.; Wu, J.; Cao, B.; Fan, J. A deep-learning prediction model for imbalanced time series data forecasting. *Big Data Min. Anal.* **2021**, *4*, 266–278. <https://doi.org/10.26599/BDMA.2021.9020011>.
49. Taieb, S.B.; Atiya, A.F. A Bias and Variance Analysis for Multistep-Ahead Time Series Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 62–76. <https://doi.org/10.1109/TNNLS.2015.2411629>.

50. Duan, J.; Kashima, H. Learning to Rank for Multi-Step Ahead Time-Series Forecasting. *IEEE Access* **2021**, *9*, 49372–49386. <https://doi.org/10.1109/ACCESS.2021.3068895>.
51. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Yow, K.C.; Mariani, V.C.; Coelho, L.S.; Seman, L.O. Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam. *Electr. Power Syst. Res.* **2022**, *202*, 107584. <https://doi.org/10.1016/j.epsr.2021.107584>.
52. Ribeiro, M.H.D.M.; Coelho, L.S. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl. Soft Comput.* **2020**, *86*, 105837. <https://doi.org/10.1016/j.asoc.2019.105837>.
53. Du, L.; Gao, R.; Suganthan, P.N.; Wang, D.Z. Bayesian optimization based dynamic ensemble for time series forecasting. *Inf. Sci.* **2022**, *591*, 155–175. <https://doi.org/10.1016/j.ins.2022.01.010>.
54. Kim, D.; Kim, C. Forecasting time series with genetic fuzzy predictor ensemble. *IEEE Trans. Fuzzy Syst.* **1997**, *5*, 523–535. <https://doi.org/10.1109/91.649903>.
55. Ribeiro, M.H.D.M.; da Silva, R.G.; Moreno, S.R.; Mariani, V.C.; Coelho, L.S. Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *136*, 107712. <https://doi.org/10.1016/j.ijepes.2021.107712>.
56. Sauer, J.; Mariani, V.C.; Coelho, L.S.; Ribeiro, M.H.D.M.; Rampazzo, M. Extreme gradient boosting model based on improved Jaya optimizer applied to forecasting energy consumption in residential buildings. *Evol. Syst.* **2021**, *13*, 577–588.
57. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Coelho, L.S.; Leithardt, V.R.Q.; Silva, L.A.; Seman, L.O. Hybrid Wavelet Stacking Ensemble Model for Insulators Contamination Forecasting. *IEEE Access* **2021**, *9*, 66387–66397. <https://doi.org/10.1109/ACCESS.2021.3076410>.
58. da Silva, R.G.; Ribeiro, M.H.D.M.; Moreno, S.R.; Mariani, V.C.; Coelho, L.S. A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting. *Energy* **2021**, *216*, 119174. <https://doi.org/10.1016/j.energy.2020.119174>.
59. Abbasimehr, H.; Paki, R. Improving time series forecasting using LSTM and attention models. *J. Ambient Intell. Humaniz. Comput.* **2022**, *13*, 673–691.
60. Yu, R.; Gao, J.; Yu, M.; Lu, W.; Xu, T.; Zhao, M.; Zhang, J.; Zhang, R.; Zhang, Z. LSTM-EFG for wind power forecasting based on sequential correlation features. *Future Gener. Comput. Syst.* **2019**, *93*, 33–42. <https://doi.org/10.1016/j.future.2018.09.054>.
61. Ko, M.S.; Lee, K.; Kim, J.K.; Hong, C.W.; Dong, Z.Y.; Hur, K. Deep Concatenated Residual Network with Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting. *IEEE Trans. Sustain. Energy* **2021**, *12*, 1321–1335. <https://doi.org/10.1109/TSTE.2020.3043884>.
62. Guo, J.; Lao, Z.; Hou, M.; Li, C.; Zhang, S. Mechanical fault time series prediction by using EFMSAE-LSTM neural network. *Measurement* **2021**, *173*, 108566. <https://doi.org/10.1016/j.measurement.2020.108566>.
63. Stefenon, S.F.; Oliveira, J.R.; Coelho, A.S.; Meyer, L.H. Diagnostic of Insulators of Conventional Grid Through LabVIEW Analysis of FFT Signal Generated from Ultrasound Detector. *IEEE Lat. Am. Trans.* **2017**, *15*, 884–889. <https://doi.org/10.1109/TLA.2017.7910202>.
64. Sabir, R.; Rosato, D.; Hartmann, S.; Guehmann, C. LSTM Based Bearing Fault Diagnosis of Electrical Machines using Motor Current Signal. In Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications, ICMLA, Boca Raton, FL, USA, 16–19 December 2019; pp. 613–618. <https://doi.org/10.1109/ICMLA.2019.00113>.
65. Jalayer, M.; Orsenigo, C.; Vercellis, C. Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, Fast Fourier and continuous wavelet transforms. *Comput. Ind.* **2021**, *125*, 103378. <https://doi.org/10.1016/j.compind.2020.103378>.
66. Tan, W.; Sun, Y.; Qiu, D.; An, Y.; Ren, P. Rolling Bearing Fault Diagnosis Based on Single Gated Unite Recurrent Neural Networks. *J. Phys. Conf. Ser.* **2020**, *1601*, 042017. <https://doi.org/10.1088/1742-6596/1601/4/042017>.
67. Leithardt, V.; Santos, D.; Silva, L.; Viel, F.; Zeferino, C.; Silva, J. A Solution for Dynamic Management of User Profiles in IoT Environments. *IEEE Lat. Am. Trans.* **2020**, *18*, 1193–1199. <https://doi.org/10.1109/TLA.2020.9099759>.
68. Viel, F.; Silva, L.A.; Valderi Leithardt, R.Q.; Zeferino, C.A. Internet of Things: Concepts, Architectures and Technologies. In Proceedings of the 2018 13th IEEE International Conference on Industry Applications (INDUSCON), São Paulo, Brazil, 11–14 November 2018; pp. 909–916. <https://doi.org/10.1109/INDUSCON.2018.8627298>.
69. Mendes, A.S.; Silva, L.A.; Blas, H.S.S.; Jiménez Bravo, D.M.; Leithardt, V.R.O.; González, G.V. WCIoT: A Smart Sensors Orchestration for Public Bathrooms using LoRaWAN. In Proceedings of the 2021 Telecoms Conference (ConfTELE), Leiria, Portugal, 11–12 February 2021; pp. 1–5. <https://doi.org/10.1109/ConfTELE50222.2021.9435574>.
70. Itajiba, J.A.; Varnier, C.A.C.; Cabral, S.H.L.; Stefenon, S.F.; Leithardt, V.R.Q.; Ovejero, R.G.; Nied, A.; Yow, K.C. Experimental Comparison of Preferential vs. Common Delta Connections for the Star-Delta Starting of Induction Motors. *Energies* **2021**, *14*, 1318. <https://doi.org/10.3390/en14051318>.
71. Leithardt, V.R.Q. Classifying garments from fashion-MNIST dataset through CNNs. *Adv. Sci. Technol. Eng. Syst. J.* **2021**, *6*, 989–994.
72. Stefenon, S.F.; Seman, L.O.; Schutel Furtado Neto, C.; Nied, A.; Seganfredo, D.M.; da Luz, F.G.; Sabino, P.H.; Torreblanca González, J.; Quietinho Leithardt, V.R. Electric Field Evaluation Using the Finite Element Method and Proxy Models for the Design of Stator Slots in a Permanent Magnet Synchronous Motor. *Electronics* **2020**, *9*, 1975. <https://doi.org/10.3390/electronics9111975>.
73. Muniz, R.N.; Stefenon, S.F.; Buratto, W.G.; Nied, A.; Meyer, L.H.; Finardi, E.C.; Kühl, R.M.; Sá, J.A.S.d.; da Rocha, B.R.P. Tools for Measuring Energy Sustainability: A Comparative Review. *Energies* **2020**, *13*, 2366. <https://doi.org/10.3390/en13092366>.
74. Stefenon, S.F.; Kasburg, C.; Freire, R.Z.; Silva Ferreira, F.C.; Bertol, D.W.; Nied, A. Photovoltaic power forecasting using wavelet Neuro-Fuzzy for active solar trackers. *J. Intell. Fuzzy Syst.* **2021**, *40*, 1083–1096. <https://doi.org/10.3233/JIFS-201279>.

75. Casado-Vara, R.; del Rey, A.M.; Pérez-Palau, D.; de-la Fuente-Valentín, L.; Corchado, J.M. Web Traffic Time Series Forecasting Using LSTM Neural Networks with Distributed Asynchronous Training. *Mathematics* **2021**, *9*, 421. <https://doi.org/10.3390/math9040421>.
76. Fernandes, F.; Stefenon, S.F.; Seman, L.O.; Nied, A.; Ferreira, F.C.S.; Subtil, M.C.M.; Klaar, A.C.R.; Leithardt, V.R.Q. Long short-term memory stacking model to predict the number of cases and deaths caused by COVID-19. *J. Intell. Fuzzy Syst.* **2022**, *6*, 6221–6234. <https://doi.org/10.3233/JIFS-212788>.
77. Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2019**, *323*, 203–213. <https://doi.org/10.1016/j.neucom.2018.09.082>.
78. Liu, Y.; Guan, L.; Hou, C.; Han, H.; Liu, Z.; Sun, Y.; Zheng, M. Wind Power Short-Term Prediction Based on LSTM and Discrete Wavelet Transform. *Appl. Sci.* **2019**, *9*, 1108. <https://doi.org/10.3390/app9061108>.
79. Qian, N. On the momentum term in gradient descent learning algorithms. *Neural Netw.* **1999**, *12*, 145–151. [https://doi.org/10.1016/S0893-6080\(98\)00116-6](https://doi.org/10.1016/S0893-6080(98)00116-6).
80. Kasburg, C.; Stefenon, S.F. Deep Learning for Photovoltaic Generation Forecast in Active Solar Trackers. *IEEE Lat. Am. Trans.* **2019**, *17*, 2013–2019. <https://doi.org/10.1109/TLA.2019.9011546>.
81. Sopelsa Neto, N.F.; Stefenon, S.F.; Meyer, L.H.; Ovejero, R.G.; Leithardt, V.R.Q. Fault Prediction Based on Leakage Current in Contaminated Insulators Using Enhanced Time Series Forecasting Models. *Sensors* **2022**, *22*, 6121. <https://doi.org/10.3390/s22166121>.
82. Stefenon, S.F.; Freire, R.Z.; Coelho, L.S.; Meyer, L.H.; Grebogi, R.B.; Buratto, W.G.; Nied, A. Electrical Insulator Fault Forecasting Based on a Wavelet Neuro-Fuzzy System. *Energies* **2020**, *13*, 484. <https://doi.org/10.3390/en13020484>.
83. Wei, L.Y. A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. *Appl. Soft Comput.* **2016**, *42*, 368–376. <https://doi.org/10.1016/j.asoc.2016.01.027>.
84. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Coelho, L.S.; Rocha, D.F.M.; Grebogi, R.B.; Ruano, A.E.B. Wavelet group method of data handling for fault prediction in electrical power insulators. *Int. J. Electr. Power Energy Syst.* **2020**, *123*, 106269. <https://doi.org/10.1016/j.ijepes.2020.106269>.
85. Ribeiro, M.D.M.; Moreno, S.; Silva, R.G.d.; Larcher, J.H.K.; Canton, C.; Mariani, V.; Coelho, L. Wind power forecasting based on bagging extreme learning machine ensemble model. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Brugge, Belgium, 5–7 October 2022; pp. 345–350. <https://doi.org/10.14428/esann/2022.ES2022-117>.
86. Zhang, Y.; Mo, C.; Ma, J.; Zhao, L. Random Subspace Ensembles of Fully Convolutional Network for Time Series Classification. *Appl. Sci.* **2021**, *11*, 10957. <https://doi.org/10.3390/app112210957>.
87. da Silva, R.G.; Moreno, S.R.; Ribeiro, M.H.D.M.; Larcher, J.H.K.; Mariani, V.C.; dos Santos Coelho, L. Multi-step short-term wind speed forecasting based on multi-stage decomposition coupled with stacking-ensemble learning approach. *Int. J. Electr. Power Energy Syst.* **2022**, *143*, 108504. <https://doi.org/10.1016/j.ijepes.2022.108504>.