Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2022 doi:10.20944/preprints202208.0031.v1

Article

Fault Prediction Based on Leakage Current in Contaminated Insulators Using Enhanced Time Series Forecasting Models

Nemesio Fava Sopelsa Neto [1]*, Stefano Frizzo Stefenon [2,3], Luiz Henrique Meyer [1], Raúl García Ovejero [4], and Valderi Reis Quietinho Leithardt [5,6]

1 Department of Electrical Engineering, Regional University of Blumenau. Rua São Paulo 3250, 89030-000 Blumenau, Brazil; meyer@furb.br
2 Fondazione Bruno Kessler. Via Sommarive 18, 38123 Trento, Italy; sfrizzostefenon@fbk.eu
3 Department of Mathematics, Informatics and Physical Sciences, University of Udine. Via delle Scienze 206, 33100 Udine, Italy
4 Expert Systems and Applications Laboratory, E.T.S.I.I. of Béjar, Universidad de Salamanca. 37700 Salamanca, Spain; raulovej@usal.es
5 COPELABS, Lusófona University of Humanities and Technologies. Campo Grande 376, 1749-024 Lisboa, Portugal; valderi@ipportalegre.pt
6 VALORIZA, Research Center for Endogenous Resources Valorization. Instituto Politécnico de Portalegre, 7300-555 Portalegre, Portugal
* Correspondence: nsopelsa@furb.br

Abstract: To improve the monitoring of the electrical power grid it is necessary to evaluate the influence of contamination in relation to leakage current and its progression to a disruptive discharge. In this paper, insulators were tested in a saline chamber to simulate the increase of salt contamination on their surface, and to evaluate the supportability of these components. From the time series forecasting of the leakage current, it is possible to evaluate the development of the fault before a flashover occurs. Choosing which method to use is always a difficult task since some models may have higher computational effort. In this paper, for a complete evaluation, the long short-term memory (LSTM), group method of data handling (GMDH), adaptive neuro-fuzzy inference system (ANFIS), bootstrap aggregation (bagging), sequential learning (boosting), random subspace, and stacked generalization (stacking) ensemble learning models are analyzed. A review and comparison of these well-established methods for time series forecasting is performed. From the results of the best structure of the model, the hyperparameters are evaluated and the Wavelet transform is used to obtain an enhanced model.

Keywords: LSTM; GMDH; ANFIS; Ensemble Learning Models; Wavelet; Time Series Forecasting

1. Introduction

Power distribution insulators are responsible for the electrical insulation and mechanical support of the cables of electrical power transmission and distribution aerial networks [1]. Because they are usually installed outdoor, they are constantly exposed to adverse weather conditions [2]. External agents must be considered when choosing the type of insulator to be used in a network because depending on the environment and the insulator chosen, they can compromise the proper operation of the grid and the life of the insulator itself. Typically the insulators are manufactured using materials such as porcelain, glass, or polymers [3].

Contamination (pollution, salinity, biological agents) can make the surface of the insulator more conductive, increasing the possibility of partial discharges (PDs), dry band arcing, and consequently decreasing the efficiency of the system [4]. Contamination is a problem in places close to industries, agriculture, mining, coastal regions, or unpaved streets, that may accumulate on the surface of the insulators [5]. Therefore, in some critical locations, cleaning or washing of the insulators is required to mitigate insulation failure in the power system [6]. The layer of contamination deposited on the insulator, when dry, is not very conductive, but in the presence of moisture (light rain, fog) its conductivity can be increased by the formation of an electrolyte [7]. As the conductivity increases, PDs become more frequent and with greater intensity, which can damage the surface of the insulator and evolve into a dry band arcing that in turn may develop into complete rupture known as flashover [8].

In the presence of a layer of wet contamination and leakage current on the surface of the insulator, heat dissipation occurs, which can evaporate certain portions of this wet layer, drying out part of this conductive path [9]. This dry, non-conductive region is known as the dry band. These dry regions, which interrupt the conductive path due to evaporation, can concentrate an intense electric field, which can, in turn, lead to disruption of the air over the dry band, causing what is called dry band arcing, may causing damage to the insulation [10].

Nowadays, there are a variety of techniques and equipment specialized in the detection of defective insulators. This analysis can be performed from visual inspection techniques [11] or even taking insulator samples for bench tests. The equipment commonly used for inspection of the network are ultrasound detectors [12], acoustic sensors [13], infrared [14] and ultra-violet cameras. Software with a focus on protection and security has been increasingly used [15], which can also be an alternative for use in monitoring the electrical power system. This maintenance is performed by field technicians that when detecting possible defective insulators, clean or, if necessary, replace the insulator.

The use of machine learning to predict the increase in contamination levels, partial discharges, or/and faulty insulators has been recently growing. Because the failures do not follow a linear pattern their monitoring is a challenging task. Models that use deep layers for time series prediction are becoming popular [16], as well as models that combine simpler models to result in a more robust structure. Among the ensemble learning methods for time series forecasting, the highlights are bootstrap aggregation (bagging) [17], sequential learning (boosting) [18], random subspace [19], random forest [20], and stacked generalization [21].

One of the most effective way to assess surface insulation degradation is by monitoring the leakage current [22]. According to Ghunem *et al.* [23] the leakage current is the main cause of fires on poles, which may result wildfires. When an insulator is contaminated there may be an increase in leakage current until there is a disruptive failure [24]. The contamination accumulates over time and becomes embedded in the surface of the insulator [25]. The evaluation of the increase in leakage current can be an indication that a disruptive failure will occur [26], so this paper proposes to predict this current increase using the several approaches.

The major difficulties in using methods for time series forecasting is the definition of which model to use. Many authors support proposals for hybrid models, in which several techniques are combined, however, a complete comparison to determine if the approach is the most adequate is usually not performed. In this paper, the models long short-term memory (LSTM) [27], group method of data handling (GMDH) [28], adaptive neuro fuzzy inference system (ANFIS) [29], bagging [30], boosting [31], random subspace [32], and stacking ensemble learning models [33] are compared. After defining the structure of each model that has the best performance for the used data set, the hyperparameters of the models and the use of Wavelet transform are evaluated, which has wide application for time series [34].

The remainder of this paper is structured as follows: Section 2 discusses the problem of contamination in insulators and the laboratory procedure used to perform the simulation of contamination accumulation. Section 3 presents the methods used for time series forecasting. Section 4 analyzes and discusses the results. Section 5 presents the conclusion.

2. Insulators Contamination

Since the power distribution insulators are usually installed outdoors, they are exposed to external agents such as sunlight, rain, and wind [35]. As time passes, it is natural for small amounts of particles such as dust or salt to be deposited on the insulator. This layer of dirt on the surface of the insulator is called contamination [36]. Contamination is one of the reasons of failure in power grid insulators [37]. This happens mainly in coastal areas due to the salinity present in the sea air, on unpaved streets, and in industrial regions with mining activities or chemical industries that generate suspended dirt. This contamination is distributed on the surface of the insulator in a non-uniform way [38].

The presence of contamination on the insulator's surface does not mean that it needs to be replaced and normally has no harmful effect as long as moisture is not present [39]. However, in the presence of moisture a conductive path can be generated which decreases the insulation between the high-voltage phases and ground [40]. When cracks and fissures occur in the insulator, the rate of contamination deposition may accelerate and consequently increases its leakage current, making it more susceptible to partial discharge events and flashovers [41].

For polymeric insulators Maraaba *et al.* [42] showed that insulators with up to two years of use have a better hydrophobic level compared to equivalent equipment with fifteen years of operation. According to their study, there is a gradual reduction in hydrophobicity over time, which impairs the performance of the insulator. It is even recommended to replace the insulators after a field period longer than 19.2 years due to changes in their electrical and mechanical characteristics.

In the presence of a contaminated and wet filament, dry band arcing occur, these dissipate energy through the joule effect, which can generate dry bands since the evaporation capacity is greater than the capacity to fill the affected region with water [43]. These dry bands bring the high voltage point closer to the earth, concentrating intense electric fields, which in turn can give rise to a series of dry band arcing [44]. The discharges can damage the surface of the polymeric insulator giving rise to cracks that diminish its hydrophobic property and, in the long term, can lead to a complete rupture of the insulation [45].

Among the techniques for assessing contamination, the equivalent salt deposit density (ESDD) evaluates the amount of salt (NaCl) dissolved in a certain area measured in mg/cm^2, this value is defined by washing the insulator with a specific amount of water and then measuring the conductivity of the water [46]. From the ESDD measurement, it is possible to measure the current state of a specific insulator that can be extrapolated to several insulators in the same region with equivalent exposure time [47].

The major disadvantage of the ESDD method is the need to remove the insulator from the transmission line to perform an accurate measurement [48]. An alternative is to estimate the ESDD through the information of the leakage current [49], a consequence of this salt deposition, thus not requiring the removal of the insulator.

The accumulation of contamination and consequently PDs can cause the equipment to be at risk and lead to outages, which makes the monitoring of these insulators essential for electrical utility [50,51]. If maintenance is not done and the insulation fails, a technician should perform corrective maintenance, searching and replacing the insulator in the field in an emergency manner [52].

2.1. Laboratory Setup

To evaluate the contamination in the laboratory, tests are performed in saline chambers, where controlled contamination situations are simulated on the surface of the insulator [53]. The experiments can be performed in two ways: The first method consists of using salty water to generate saline spray. In the second method, the contamination is applied directly to the surface of the insulator. In both methods the salt levels on the surface of the insulator can be increased until the dielectric breakdown.

In the case of the experiment conducted in this paper with salt spray, the first method is applied. To simulate salt contamination that accumulates over time on the surface of the insulators, six insulators were mounted in a saline chamber. A 8.66 kV RMS 60Hz was applied to the insulators (same phase) and the salt concentration was increased gradually. This voltage level is defined by the NBR 10621 (similar to IEC 60507), standard used by the electrical power utility, specifically this standard deals with the determination of the characteristics of supportability under artificial pollution for insulators in electric power grids for the 15 kV class. The arrangement of this experiment is shown in Figure 1.

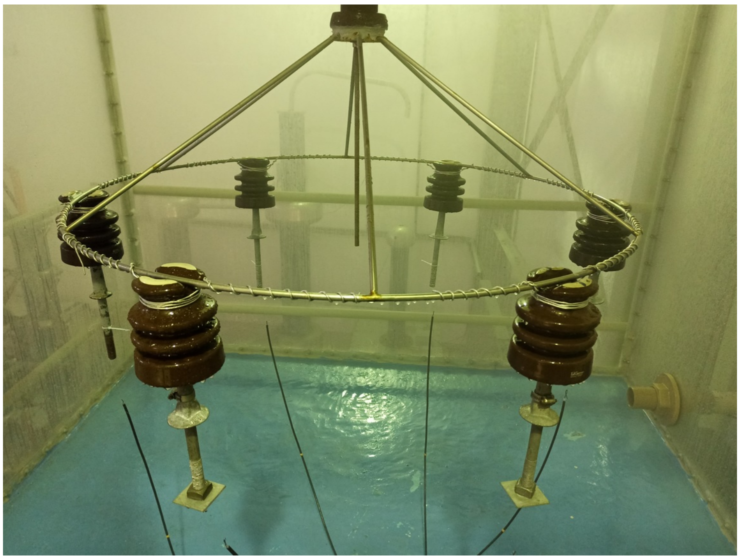

Figure 1. Insulators in a saline chamber experiment.

To monitor and record the applied voltage and resulting leakage current, an interface was developed in LabVIEW software. Each of the insulators was individually connected to the ground to measure the leakage current through a shunt resistor. Figure 2 presents the measured values of the leakage current during the experiment.

From the six analyzed insulators, two did not have flashed over and the respective leakage current values were measured until the end of the experiment. The other insulators were monitored only until the surface breakdown occurred. Using the signal recorded during the laboratory experiment, time series forecasting models are applied to evaluate the capability of predicting the development of a failure in relation to the increase in leakage current.

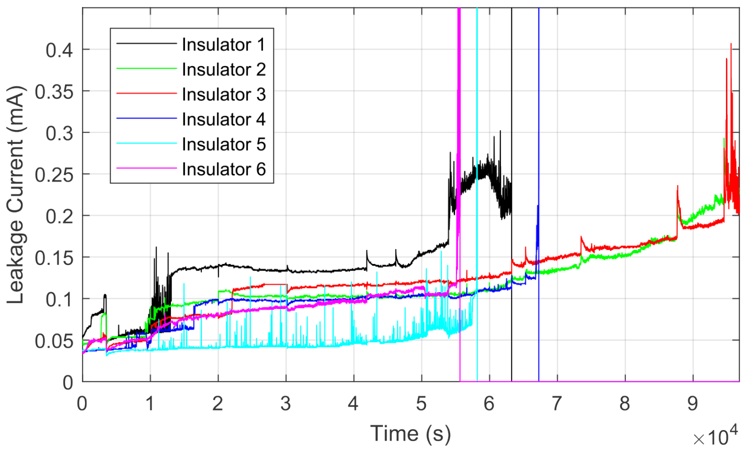

Figure 2. Leakage current of the contaminated insulators.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2022 doi:10.20944/preprints202208.0031.v1

3. Time Series Forecasting

The values from the time series are used up to time t, this way it is possible to predict the value in the future, $t + P$ [54]. Thus, a mapping is created from the sample points n, sampled in each unit Δt in time,

$$\hat{x} = x(t - (n-1)\Delta t), \ldots,\ x\ (t - \Delta t),\ x(t), \quad (1)$$

to a predicted value,

$$y(t + P) = f(x). \quad (2)$$

To predict the values of the next steps forward, the answers of the training sequence are changed on a time interval. Using the time series forecasting step ahead approach, each input sequence of the time step, learn to predict the value of the next time step [55]. To obtain the expected values of future time steps, the output of the training sequences is shifted by a single time step [56]. From the time series evaluation it is feasible to predict the development of a flashover voltage considering contamination conditions on electrical power insulators [57].

Several models can be used to time series forecasting, making the choice of the appropriate model a difficult task. LSTM has been applied in deep learning by several authors due to its promising features in dealing with non-linear data [58]. GMDH has performance advantages because it is an adaptive model that disregards neurons that do not help in the training process [59].

ANFIS has the advantages of fuzzy logic for time series forecasting [60]. The combination of simpler models makes ensemble learning approach a promising alternative for forecasting [61], like the bagging, boosting, random subspace, and stacking. These approaches will be explained and compared in this paper. The structure of these models is presented in Figure 3 and will be explained in this section.

3.1. LSTM

LSTM is a recurrent neural network used in deep learning that has become increasingly popular [62]. The major advantage of using the LSTM is that it is able to learn long-term dependencies, being able to handle nonlinear variations of the system, which is an important feature for time series forecasting [63].

A LSTM unit consists of a cell with an input port, an output port, and a forgetting port [64]. The unit remembers values at arbitrary time intervals and the three gates control the flow of information in and out of the cell [65]. The LSTM can be calculated according to the equations:

$$\begin{aligned} i_t &= \sigma_g(W_i x_t + R_i h_{t-1} + b_i), \\ f_t &= \sigma_g(W_f x_t + R_f h_{t-1} + b_f), \\ o_t &= \sigma_g(W_o x_t + R_o h_{t-1} + b_o), \end{aligned} \quad (3)$$

where σ_g is the gate activation function, W and R are earning matrices and b is the polarization matrix, these values are assigned to the network training [66].

In a LSTM recurrent unit h_{t-1} is the hidden state at the previous timestep $t - 1$ (short-term memory), c_{t-1} is the cell state at previous timestep $t - 1$ (long-term memory), x_t is the input vector at current timestep t, h_t is the hidden state at current timestep t, and c_t is the cell state at current timestep t [67].

During the training phase it is possible to define several optimizers, the most popular of them are stochastic gradient descent with momentum (SGDM) [68], adaptive moment estimation (ADAM) [69] and RMS propagation (RMSProp) [70]. In this paper the preliminary evaluation will be performed using the SGDM with 50 hidden units and after defining the best structure all the mentioned optimizers will be evaluated.

 doi:10.20944/preprints202208.0031.v1

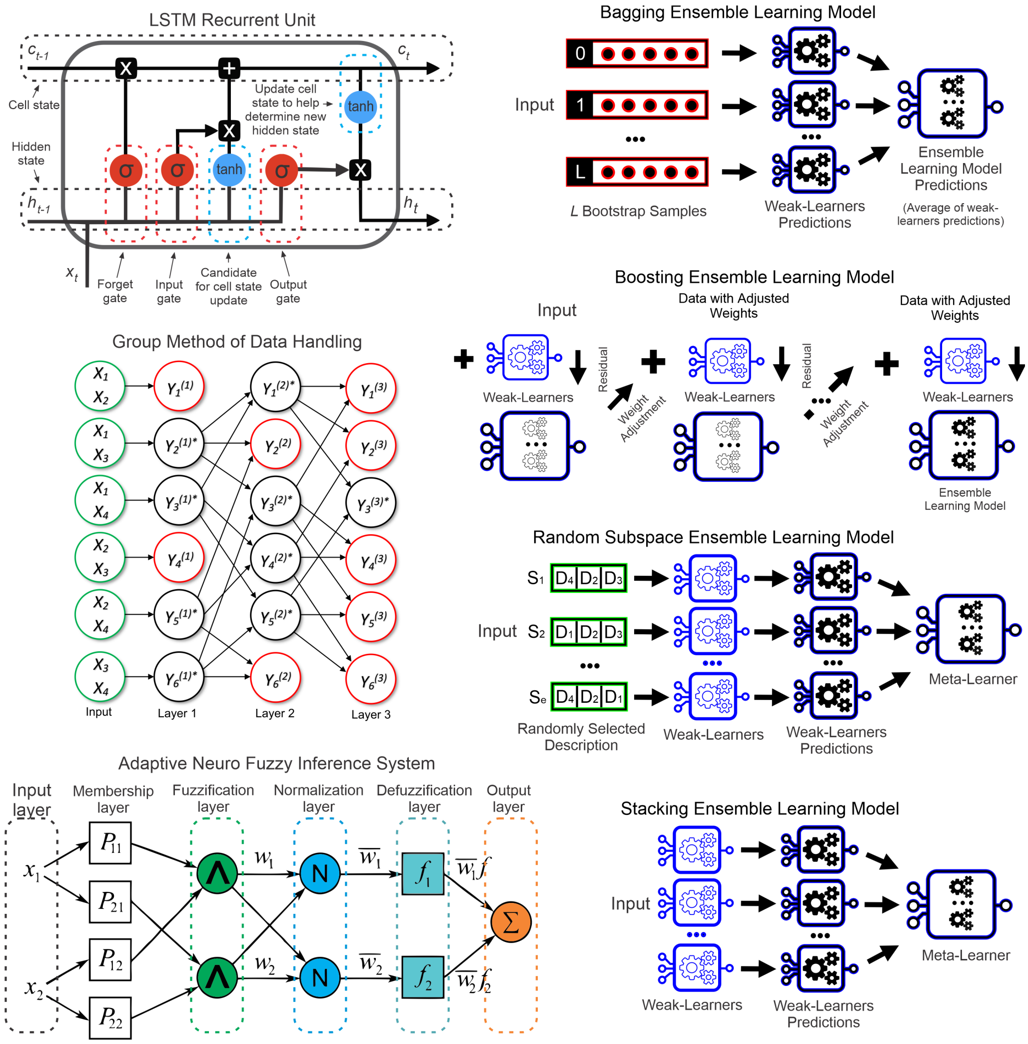

Figure 3. Structure of the considered models.

3.2. GMDH

Group method of data handling is distinguished by been an inductive approach that performs the ranking of gradually complicated polynomial models and selects the best possible solution using an external criterion [71]. The external criterion is one major feature of GMDH, as it describes the requirements of the model. In this model the number of hidden layers and the number of their neurons are determined automatically [72].

The GMDH selects the best structure that results in better performance to obtain an optimized network, when the minimum is no longer reduced with the previous layer, the network prediction error stops [73]. For a comparison of the models, a maximum of 50 neurons were initially used. The coefficients of GMDH are solved with regression methods for each pair of input variables x_i and x_j, where:

$$\hat{y}(x_1, \ldots, x_n) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \ldots \quad (4)$$

$$G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a x_i x_j. \quad (5)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2022 doi:10.20944/preprints202208.0031.v1

In this paper, the coefficients are estimated by the least-squares error (LSE) function, according:

$$\mathrm{LSE} = \begin{cases} \hat{y}(x_1, \ldots, x_n) = G(x_i,\ x_j) \\ e = \sum_{n=1}^{N} (y - \hat{y})^2 \\ \frac{de}{da_k} = 0,\ \ k = 1,\ 2,\ 3,\ 4,\ 5. \end{cases} \tag{6}$$

To make the analysis easier, the results can be expressed in matrix form as:

$$a = (X^T X)^{-1} X^T y \tag{7}$$

where,

$$X = \begin{Bmatrix} 1 & x_{i1} & x_{j1} & x_{i1}x_{j1} & x_{i1}^2 & x_{j1}^2 \\ 1 & x_{i2} & x_{j2} & x_{i2}x_{j2} & x_{i2}^2 & x_{j2}^2 \\ 1 & x_{i3} & x_{j3} & x_{i1}x_{j2} & x_{i3}^2 & x_{j3}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in} & x_{jn} & x_{in}x_{jn} & x_{in}^2 & x_{jn}^2 \end{Bmatrix}. \tag{8}$$

3.3. ANFIS

An ANFIS is a neural network based on the Takagi-Sugeno-Kang inference model. This method unites both the benefits of neural networks and fuzzy systems in the same structure. By using both characteristics of these methods, this approach can deal with systems involving imprecise and non-linear data [74].

The behavior of ANFIS can be understood by observing variables related to membership functions, the relationship from inputs to outputs, and fuzzy rules. Given these features, the ANFIS model might be adopted for chaotic time series forecasting [75]. The model optimization was evaluated in two ways, being the backpropagation using a gradient descent to calculate all parameters, and the hybrid method that performs combination of backpropagation to calculate the input membership parameters, and least-squares estimation to calculate the output membership parameters.

3.4. Ensemble learning models

Ensemble learning modeling is based on the divide-and-conquer challenge dedicated to enhancing the accuracy of models. Several weak learners individually perform a particular task, and when their results are combined, a more efficient model regarding the accuracy is achieved [76].

The ensemble learning approach has better results because every base model learns distinct features of the data, and then when the outputs are combined, the entire pattern of the data is learned by aggregating the weak learners (base models) [77]. Because of the suitability of the ensemble learning method to handle different types of data, applications can be found in various fields such as energy [78], security [79], public health [80], industry [81], and environment [82].

The weak learners used in this paper are support vector regression (SVR), these base models were used considering that they are an efficient approach for ensemble learning models [83]. The form of the SVR is defined by a convex optimization problem with linear constraints, denoted by:

$$\min_{\mathbf{w},b,\sigma} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{m} \sigma_i \tag{9a}$$

$$\text{s.t.}:\ y_i - (\mathbf{w} \cdot \mathbf{x}_i + b) \leq \varepsilon + \sigma_i \tag{9b}$$

$$\text{s.t.}:\ (\mathbf{w} \cdot \mathbf{x}_i + b) - y_i \leq \varepsilon + \sigma_i \tag{9c}$$

$$\sigma_i \geq 0, \forall i = 1, \ldots, m \tag{9d}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2022 doi:10.20944/preprints202208.0031.v1

wherein $\mathbf{w}$ and b are the normal vector and the bias, respectively of a training dataset ($\mathbf{x}_i$, y_i), and ε is a margin of tolerance. The σ_i is employed to transform (9b) into a soft constraint (penalized by C), letting the optimization problem satisfy the constraint, even in ambiguous cases. To accomplish the relationship between input and output data,

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \tag{10}$$

where the forecasting values are $f(\mathbf{x})$ and the mapping of the input vector $\mathbf{x}$ is ϕ; b and $\mathbf{w}$ are the coefficients calculated by the minimization of the risk function (R):

$$R(f) = \frac{1}{N} \sum_{i=1}^{N} L_\varepsilon\Big(\mathbf{y}_i, \mathbf{w}^T \phi(\mathbf{x}_i) + b\Big) \tag{11}$$

and the loss function (L_ε) is used to penalize the training errors, evaluated by:

$$L_\varepsilon(\mathbf{y}, f(\mathbf{x})) = \begin{cases} 0 & if |\mathbf{y} - f(\mathbf{x})| \leq \varepsilon \\ |\mathbf{y} - f(\mathbf{x})| - \varepsilon & otherwise. \end{cases} \tag{12}$$

Using the L_ε in the regularized function the task becomes a quadratic programming problem. The minimization of this function can be rewritten as an equivalent optimization problem, referred to as the *primal* problem:

$$\begin{gathered} \min_{(w,b,\xi,\xi^*)} \; R(w,b,\xi,\xi^*) = \\ C\sum_{i=1}^{n}(\psi_s(\xi_i) + \psi_s(\xi_i^*)) + \tfrac{1}{2}\|w\|^2 \end{gathered} \tag{13}$$

subject to

$$\begin{cases} y_i - \langle w \cdot \phi(x_i) \rangle - b \leq (1-\beta)\epsilon + \xi_i; \\ \langle w \cdot \phi(x_i) \rangle + b - y_i \leq (1-\beta)\epsilon + \xi_i^*; \\ \xi_i^* \geq 0; \quad \xi_i^* \geq 0 \quad \forall_i \end{cases} \tag{14}$$

wherein

$$\psi_s(\xi) = \begin{cases} \frac{\xi^2}{4\beta\epsilon} & if \quad \xi \in [0, 2\beta\epsilon]; \\ \xi - \beta\epsilon & if \quad \xi \in [2\beta\epsilon, +\infty]. \end{cases} \tag{15}$$

Rewriting the *dual* problem:

$$\begin{gathered} \min_{(\alpha,\alpha^*)} \; R(\alpha,\alpha^*) = \\ \tfrac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_i^*)K(x_i,x_j) \\ -\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i + \sum_{i=1}^{n}(\alpha_i + \alpha_i^*)(1-\beta)\epsilon \\ +\tfrac{\beta\epsilon}{C}\sum_{i=1}^{n}(\alpha_i^2 + \alpha_i^{*2}) \end{gathered} \tag{16}$$

subject to

$$\begin{gathered} 0 \leq \alpha_i \leq C, \forall_i, 0 \leq \alpha_i^* \leq C, \forall_i \\ and \\ \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0. \end{gathered} \tag{17}$$

The Kernel functions (K) used in the SVR for this paper are linear (19), radial basis function (RBF) (18), and polynomial (20). For an overall evaluation, the quadratic programming (L1QP) [84], iterative single data algorithm optimization (ISDA) [85], and sequential minimal optimization (SMO) [86] optimizers were used. For the initial comparison of ensemble models, the linear Kernel function and the L1QP optimizer were adopted. After the best-fit model was defined, all the presented kernel functions and optimizers were evaluated.

$$K(x_j, x_k) = \exp(-\|x_j - x_k\|^2). \tag{18}$$

$$K(x_j, x_k) = x_j\prime\, x_k. \tag{19}$$

$$K(x_j, x_k) = (1 + x_j\prime\, x_k)^q. \tag{20}$$

Several models are employed in the field of ensemble learning, in which bootstrap aggregation (bagging) [87], sequential learning (boosting) [88], random subspace [89], and stacked generalization [90] can be highlighted. This grouping is intended to bring together the weakest multiple models to reduce their general susceptibility to the bias-variance, therefore making the prediction more robust.

3.4.1. Bagging

The bagging ensemble method is a type of parallel method, aimed to generate a more robust set of models than individual models that compose it (week learners). Bagging is focused on reducing the variance of the resulting model, in this method independent learners are considered independent of each other, then it is possible to train them simultaneously [91]. The bagging method can also be called bootstrap aggregation, since the bootstrap sample is created initially for each model, and afterwards the model is aggregated, combined by mean rule (in regression cases) [92].

3.4.2. Boosting

The boosting ensemble learning is a process characterized as a sequential learning approach, weak learners are not independently trained, focusing on the reduction of the bias of the individual models [93]. Indeed, for the regression tasks, the effectiveness of the boosting approach is due to the fact that, after the result of the first weak model, the following model try to improve accuracy by fitting models to the residual of the previous models [94].

Using the regularization parameter, overfitting is avoided. In fact, the boosting paradigm trains new models iteratively, concentrating on observations that the prior models had more difficulty in predicting, which therefore makes the predictive model less impartial [95]. Since the goal is to reduce the bias of simpler predictors, it is proper to use a simpler model with higher bias and lower variance [96].

3.4.3. Random Subspace

Random subspace ensemble learning model is a popular random sampling method, that was introduced by Ho [97] to improve the performance of weak classifiers and to improve the classification accuracy of individual classifiers [98]. According to Pham *et al.* [99], random subspace is an ensemble approach where the original high-dimensional feature vector is randomly sampled to create the low-dimensional subspaces, and multiple classifiers are then combined on these random subspaces for the final decision.

3.4.4. Stacked Generalization

Stacking ensemble learning combines several different predictive models into a single model, working in layers or levels. This concept introduces the meta-learning, which represents an asymptotically optimum learning system, and is intended to minimize generalization errors by reducing the bias of its generalizers [100]. In fact, a stacked model is created from the predictions of weak learners, which are used as features. The characteristics allow the resulting model to cluster the initial models, so the model disregards the results that performed poorly [101].

3.5. Wavelet

For the purpose of comparing the proposed model, the signal will be filtered using the Wavelet transform (WT) to assess whether the use of filters is promising for the application. In the WT, information is extracted from each signal segment and treated [102]. Initially, Wavelet energy coefficient is obtained after the signal decomposition by the Wavelet packets transform (WPT), considering that the information on both sides of the spectrum is considered in this procedure [103].

The WPT performs a new decomposition in each interaction, based on the coefficients of the previous iterations, and thus indicates that the final number of coefficients depends on the number of iterations [104]. The orthogonal Wavelet is decomposed into Wavelet packages (WP) and thus a vector tree structure is created. The structure is divided into two parts, the first being an approximation coefficient vector, and the second a detailed result vector [105]. The WP function can be obtained by:

$$W^n_{j,\ k}(t) = 2^{j/2} W^n(2^j t - k) \tag{21}$$

where k represents the translation operator, j is a scalable parameter, and n is the oscillation parameter. The first two WP functions for $n = 0$ and $n = 1$ are:

$$W^0_{0,0}(t) = \phi(t),\ W^1_{0,0}(t) = \psi(t). \tag{22}$$

Equation (21) represents the scale function, and equation (22) represents the main function. The equations for $n = 2, 3, \ldots, N$, can be defined according to the following relations:

$$W^{2n}_{0,0}(t) = \sqrt{2} \sum_k \delta(k) W^n_{1,k}(2t - k), \tag{23}$$

$$W^{2n+1}_{0,0}(t) = \sqrt{2} \sum_k \zeta(k) W^n_{1,k}(2t - k), \tag{24}$$

where $\zeta(k)$ is a high-pass filter and $\delta(k)$ is a low-pass filter [106]. The coefficients $\Omega^n_j(k)$ can be calculated by the product of $x(t)$ by $W^n_{j,k}$, expressed by:

$$\Omega^n_j(k) = \int_{-\infty}^{\infty} x(t) W^n_{j,\ k} dt. \tag{25}$$

Each WP coefficient can be determined according to its frequency level. While the Wavelet decomposes the elements of low frequency, the WPT decomposes the elements of all frequencies, so its use results in components of low and high frequencies [107]. Using the tree structure, generated by the decomposition coefficients of approximation, an optimal binary value is obtained. The resulting subtree can be much smaller than the original and thus its can make the algorithm more efficient [108].

3.6. Considered measures

Aiming to forecast the leakage current of the contaminated insulators, it is promising to evaluate the evolution of the failure through time series analysis, thus efficiently estimating the moment when the component is vulnerable to suffering a disruptive discharge [109].

The most commonly used measures of performance in relation to forecast error are root-mean-square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) [110], calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{26}$$

$$\mathrm{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \tag{27}$$

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{28}$$

where the error is calculated by the difference in the observed value y_i to the predicted output $\hat{y}_i$ [111].

Typically for classification tasks, the metrics are based on the confusion matrix [112], however for prediction the metrics are based on the error [113]. The coefficient of determination (R^2) measures the adjustment of a statistical model to the observed values of a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2022 doi:10.20944/preprints202208.0031.v1

random variable, this is another widely used metric for evaluating regressions, calculated by:

$$R^2 = 1 - \frac{RSS}{TTS}, \tag{29}$$

where RSS is the relationship between the residual sum of squares and TSS is the total sum of squares, given by:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{30}$$

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2, \tag{31}$$

where $\bar{y}_i$ is the average of the observed value [114]. When the best model configurations were found, 100 runs were performed to evaluate the mean (Eq. 32), median (Eq. 33), standard deviation (Eq. 34), and variance (Eq. 35). The simulations were evaluated using an Intel Core I5-7400, 20 GB of random-access memory, with MATLAB software, version R2019.

$$\bar{x}_i = \frac{1}{m} \sum_{i=1}^{m} x_i. \tag{32}$$

$$\text{Median} = \begin{cases} x_{(m+1)/2}, & \text{if } m \text{ is odd,} \\ \frac{x_{(m/2)} + x_{(m/2)+1}}{2}, & \text{if } m \text{ is even.} \end{cases} \tag{33}$$

$$\text{Std Dev.} = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x}_i)^2}. \tag{34}$$

$$\text{Variance} = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x}_i)^2, \tag{35}$$

where m is the number of performed runs (100 in this paper), x_i is the results of the RMSE of each run (i), and $\bar{x}_i$ is the mean result of all the simulations.

4. Analysis of Results

The leakage current of the insulators was recorded with a time interval of one second between each record, during the experiment 100,000 records were made, corresponding to approximately 27 hours and 46 minutes of evaluation. As the time series is long-term, the down sample method of order 5 is used to reduce the time series length. During the experiment the water salinity went from 142.3 to 133,600.0 (uS), the pressure from 3 to 5 (BAR) and the water flow from 10.4 to 20.8 (ml/s), there was no significant change in temperature, humidity, and applied voltage.

The dielectric breakdown occurred in four out of the six tested insulators. Only the insulators that resulted in a disruptive discharge were considered for the analysis since this is the condition that should be predicted when an increase in insulator contamination occurs. One of the insulators at the end of the experimental analysis is shown in Figure 4.

After recording the variation of leakage current over time due to the increased accumulation of contamination on the insulator surface, time series forecasting models were applied to evaluate the prediction capacity of the failure development. The best results for each structure are highlighted in bold in this section.

4.1. Time Series Forecasting Analysis

All the models were compared for a preliminary evaluation of the models' performance using standard parameter configurations, as explained in the previous section. The comparison between the models is performed according to variations of their structure, in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2022 doi:10.20944/preprints202208.0031.v1

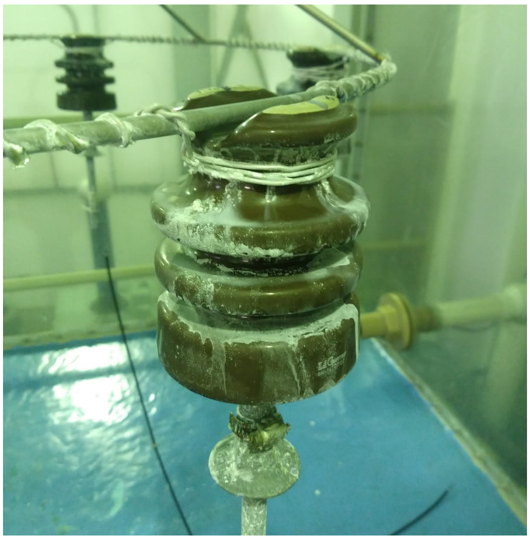

Figure 4. Insulator with salt contamination accumulated on its surface at the end of the experiment.

the LSTM and GMDH models this variation is relayed by increasing the size of the network through the inclusion of more layers, the results are presented in Table 1.

Table 1. Overall comparison of the models.

Model	Structure	RMSE	MAPE	MAE	R^2	Time (s)
LSTM	1 Deeper Layer	$\mathbf{2.35}\times10^{-2}$	$\mathbf{6.60}\times10^{-2}$	$\mathbf{1.56}\times10^{-3}$	**0.7104**	**75.86**
	2 Deeper Layers	3.01×10^{-2}	8.07×10^{-2}	1.95×10^{-2}	0.5269	107.22
	3 Deeper Layers	4.49×10^{-2}	1.37×10^{-1}	3.17×10^{-2}	0.0506	155.01
	4 Deeper Layers	5.22×10^{-2}	1.63×10^{-1}	3.75×10^{-2}	0.4208	182.07
GMDH	1 Max. Layer	4.79×10^{-3}	3.32×10^{-3}	7.18×10^{-4}	0.9880	2.21
	2 Max. Layers	$\mathbf{4.35}\times10^{-3}$	$\mathbf{1.90}\times10^{-3}$	$\mathbf{4.70}\times10^{-4}$	**0.9901**	2.90
	3 Max. Layers	5.10×10^{-3}	9.06×10^{-3}	2.10×10^{-3}	0.9864	**1.92**
	4 Max. Layers	6.00×10^{-2}	3.11×10^{-2}	7.30×10^{-3}	0.8819	4.68
ANFIS	FCM	1.15×10^{-2}	2.38×10^{-2}	4.85×10^{-3}	0.9304	**25.08**
	Grid Partitioning	5.23×10^{-3}	5.68×10^{-3}	1.45×10^{-3}	0.9857	92.53
	Subt. Clustering	$\mathbf{4.48}\times10^{-3}$	$\mathbf{3.29}\times10^{-5}$	$\mathbf{3.43}\times10^{-5}$	**0.9895**	73.75
Ensemble	Bagging	$\mathbf{4.19}\times10^{-3}$	$\mathbf{1.88}\times10^{-3}$	$\mathbf{3.16}\times10^{-4}$	**0.9909**	4,443.30
	Boosting	3.40×10^{-2}	1.27×10^{-1}	2.73×10^{-2}	0.3971	31,874.70
	Random Subsp.	4.94×10^{-3}	1.09×10^{-2}	2.21×10^{-3}	0.9872	25,155.34
	Stacking	5.04×10^{-2}	1.56×10^{-1}	3.59×10^{-2}	0.3255	**1,295.48**

All ensemble models had a higher computational effort than the other models until convergence, resulting in a considerably higher processing time to be computed. The GMDH was the fastest model for the required processing, this occurred because it uses the number of layers and nodes according to the needs of the task, being an efficient adaptive model. From a greater number of maximum layers used, the GMDH needs more time to converge, similar to what occurs with the LSTM model when deeper layers are used. In this evaluation, the GMDH model was superior to the LSTM model in terms of considered error metrics, coefficient of determination and time to convergence.

The coefficient of determination of the ensemble bagging, ANFIS using subtractive clustering, and GMDH with with maximum use of 2 layers were higher than the other models and their different structures. In the LSTM the use of a deeper network did not result in a progressive increase in model performance with respect to error evaluation, this shows that using a model based on deep learning may not always be the best alternative.

Regarding the MAPE and MAE, the model that had the lowest error result was the ANFIS subtractive clustering followed by the ensemble bagging and GMDH (max

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 August 2022 doi:10.20944/preprints202208.0031.v1

of 2 layers). The RMSE of these models was also lower compared to other structure configurations. These structures had the best results in this comparison, for this reason their hyperparameters will be modified for a more complete evaluation of their capabilities. A comparison of the original (observed) signal and the predicted signal is presented in Figure 5.

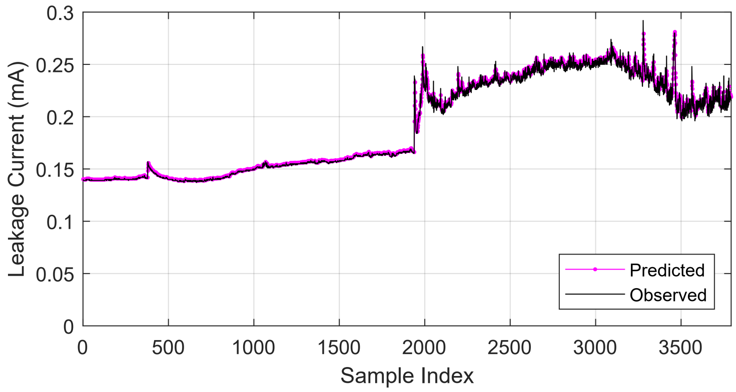

Figure 5. Predicted signal compared to observed signal.

4.2. Hyperparameter Optimization

To obtain models with better performance, the variation in the main configuration parameters of the structures that had superior results of each model was evaluated. Table 2 presents the results of the LSTM model using the SGDM, ADAM and RMSprop optimizers, varying the number of hidden units.

Table 2. Evaluation of LSTM hyperparameters.

Optimizer	Hidden Units	RMSE	MAPE	MAE	R^2	Time (s)
SGDM	10	5.40×10^{-2}	1.67×10^{-1}	3.85×10^{-2}	0.5226	81.53
	20	3.73×10^{-2}	1.12×10^{-1}	2.60×10^{-2}	0.2737	78.02
	30	2.32×10^{-2}	7.09×10^{-2}	1.63×10^{-2}	0.7180	71.09
	40	2.77×10^{-2}	8.48×10^{-2}	1.95×10^{-2}	0.5981	**70.62**
	50	$\mathbf{1.78}\times10^{-2}$	$\mathbf{4.82}\times10^{-2}$	$\mathbf{1.15}\times10^{-2}$	**0.8339**	74.71
ADAM	10	5.61×10^{-2}	1.79×10^{-1}	4.08×10^{-2}	0.6419	**71.29**
	20	5.64×10^{-2}	1.76×10^{-1}	4.05×10^{-2}	0.6601	72.94
	30	3.33×10^{-2}	1.02×10^{-1}	2.34×10^{-2}	0.4206	74.13
	40	2.97×10^{-2}	8.85×10^{-2}	2.06×10^{-2}	0.5396	74.16
	50	$\mathbf{1.82}\times10^{-2}$	$\mathbf{4.84}\times10^{-2}$	$\mathbf{1.16}\times10^{-2}$	**0.8274**	74.43
RMSprop	10	5.13×10^{-2}	1.68×10^{-1}	3.79×10^{-2}	0.3732	**72.47**
	20	5.13×10^{-2}	1.75×10^{-1}	3.90×10^{-2}	0.3762	73.32
	30	4.50×10^{-2}	1.51×10^{-1}	3.38×10^{-2}	0.0566	73.74
	40	$\mathbf{3.10}\times10^{-2}$	$\mathbf{9.16}\times10^{-2}$	$\mathbf{2.14}\times10^{-2}$	**0.4995**	75.56
	50	3.73×10^{-2}	1.33×10^{-1}	2.92×10^{-2}	0.2722	73.08

When varying the hyperparameters of the LSTM model, there was no significant improvement, considering that the difference between the best and worst results was closer than the other compared models. Table 3 presents the results of the variation of the maximum number of neurons in the GMDH model, these parameters were evaluated because the model is adaptive and it is only needed to define the maximum number of layers and neurons.

There was not much variation in the results when the maximum number of neurons used by the GMDH was changed, the best result was obtained using the maximum of 80 neurons. This result shows that GMDH is promising for this application, given that even

Table 3. Evaluation of GMDH hyperparameters.

Max Neurons	RMSE	MAPE	MAE	R^2	Time (s)
10	4.15×10^{-3}	7.79×10^{-4}	2.04×10^{-4}	0.9910	0.26
20	4.58×10^{-3}	5.61×10^{-3}	1.31×10^{-3}	0.9891	0.34
30	4.49×10^{-3}	1.41×10^{-3}	2.92×10^{-4}	0.9903	0.25
40	4.59×10^{-3}	5.89×10^{-3}	1.30×10^{-3}	0.9890	0.24
50	4.30×10^{-3}	2.57×10^{-3}	6.23×10^{-4}	0.9903	0.24
60	4.22×10^{-3}	4.44×10^{-4}	1.28×10^{-4}	0.9907	0.24
70	4.42×10^{-3}	3.43×10^{-3}	8.15×10^{-4}	0.9905	0.26
80	$\mathbf{4.09}\times10^{-3}$	$\mathbf{3.75}\times10^{-4}$	$\mathbf{1.22}\times10^{-4}$	**0.9912**	0.26
90	4.45×10^{-3}	3.70×10^{-3}	8.18×10^{-4}	0.9897	0.24
100	4.34×10^{-3}	1.08×10^{-3}	2.80×10^{-4}	0.9902	**0.23**

varying the hyperparameters of the model, the results remain with lower error compared to LSTM, high value of the coefficient of determination, being one of the fastest models to converge. These superior results of this model are given by the fact that it is an optimized model, in which neurons that do not help in the learning phase are disregarded.

The next model in which the configuration of the hyperparameters was evaluated is the ANFIS model considering the subtractive clustering structure, the results of this evaluation regarding training form and influence radius are presented in Table 4.

There was a minor variation in the ANFIS subtractive clustering model by changing the influence radius using a hybrid optimization method. The lowest MAPE and MAE values occurred using the hybrid method with a radius of influence of 0.4, and using this method the lowest RMSE value was achieved with a radius of influence of 0.2, ranking among the best coefficient of determination values. Considering that the MAPE and MAE using the classical backpropagation method were higher and the difference in RMSE and coefficient of determination were not high, the hybrid method proved to be more promising.

Table 4. Evaluation of ANFIS subtractive clustering hyperparameters.

Method	Radius	RMSE	MAPE	MAE	R^2	Time (s)
Hybrid	0.2	$\mathbf{4.15}\times10^{-3}$	5.92×10^{-4}	9.44×10^{-5}	**0.9910**	28.34
	0.4	4.48×10^{-3}	$\mathbf{3.29}\times10^{-5}$	$\mathbf{3.43}\times10^{-5}$	0.9895	28.81
	0.6	4.16×10^{-3}	4.86×10^{-4}	7.06×10^{-5}	**0.9910**	29.39
	0.8	4.16×10^{-3}	1.25×10^{-3}	2.39×10^{-4}	**0.9910**	**25.68**
	1.0	4.17×10^{-3}	6.83×10^{-4}	1.14×10^{-4}	0.9909	25.74
Backpropag.	0.2	7.60×10^{-3}	3.57×10^{-2}	6.41×10^{-3}	0.9698	25.18
	0.4	4.09×10^{-3}	2.00×10^{-3}	3.98×10^{-4}	0.9913	**24.96**
	0.6	$\mathbf{4.06}\times10^{-3}$	$\mathbf{1.74}\times10^{-3}$	$\mathbf{3.02}\times10^{-4}$	**0.9914**	25.12
	0.8	7.66×10^{-3}	3.58×10^{-2}	6.52×10^{-3}	0.9693	30.52
	1.0	8.17×10^{-3}	3.89×10^{-2}	7.12×10^{-3}	0.9651	25.51

Table 5 shows the results of using the L1QP, ISDA, and SMO optimizers for the ensemble bagging model. For these optimizers the linear, RBF, and polynomial Kernel functions are evaluated.

The ensemble bagging model took a much higher time to converge using the L1QP optimizer, this shows that an inadequate configuration can result in low performance and a high computational effort. The Kernel function that had the best coefficient of determination and lowest error results was the linear function, so the combination between the ISDA optimizer and the linear function had the best result in this analysis considering the metrics evaluated.

The best setup results in the hyperparameters evaluated in this section were: LSTM with 50 hidden units with one deeper layer using SGDM optimizer, GMDH with maximum

Table 5. Evaluation of ensemble bagging hyperparameters.

Optimizer	Kernel	RMSE	MAPE	MAE	R^2	Time (s)
L1QP	Linear	$\mathbf{4.19}\times10^{-3}$	$\mathbf{1.88}\times10^{-3}$	$\mathbf{3.16}\times10^{-4}$	**0.9909**	**4,443.30**
	RBF	1.05×10^{-1}	3.42×10^{-1}	7.76×10^{-2}	0.7926	5,256.65
	Polynomial	2.98×10^{-2}	1.43×10^{-2}	3.18×10^{-3}	0.5363	5,763.47
ISDA	Linear	$\mathbf{4.17}\times10^{-3}$	$\mathbf{1.27}\times10^{-3}$	$\mathbf{1.98}\times10^{-4}$	**0.9909**	25.33
	RBF	9.70×10^{-2}	3.14×10^{-1}	7.14×10^{-2}	0.9143	**11.52**
	Polynomial	9.68×10^{-2}	3.14×10^{-1}	7.13×10^{-2}	0.8935	11.83
SMO	Linear	$\mathbf{4.23}\times10^{-3}$	$\mathbf{3.19}\times10^{-3}$	$\mathbf{5.74}\times10^{-4}$	**0.9906**	8.09
	RBF	1.06×10^{-1}	3.44×10^{-1}	7.82×10^{-2}	0.8680	**6.06**
	Polynomial	1.06×10^{-1}	3.44×10^{-1}	7.80×10^{-2}	0.8132	6.67

2 layers and 80 neurons, ANFIS subtractive clustering with hybrid method and influence radius of 0.2, and ensemble bagging with the ISDA optimizer with linear Kernel function. These settings were used for the following analysis in this paper, the result of the statistical evaluation using these settings for the RMSE is presented in Table 6. In this evaluation, the ANFIS and ensemble models had the better results with lower error, smaller variance, and standard deviation than GMDH and LSTM.

Table 6. Statistical assessment.

Model	Mean	Median	Std. Dev.	Variance
LSTM	2.17×10^{-2}	2.11×10^{-2}	3.30×10^{-3}	1.09×10^{-5}
GMDH	4.45×10^{-3}	4.38×10^{-3}	3.02×10^{-3}	9.12×10^{-6}
ANFIS	$\mathbf{4.15}\times10^{-3}$	$\mathbf{4.15}\times10^{-3}$	$\mathbf{8.72}\times10^{-18}$	$\mathbf{7.60}\times10^{-35}$
Ensemble	4.20×10^{-3}	4.19×10^{-3}	5.77×10^{-5}	3.33×10^{-9}

4.3. Application of Wavelet Transform

The results of the use of the Wavelet transform on insulator 1 are presented in Figure 6. These results are presented in relation to a generic index, considering that after the use of the downsample algorithm the samples do not follow the time sequence of the experiment, since the focus of the evaluation is the variation in the amplitude of the leakage current.

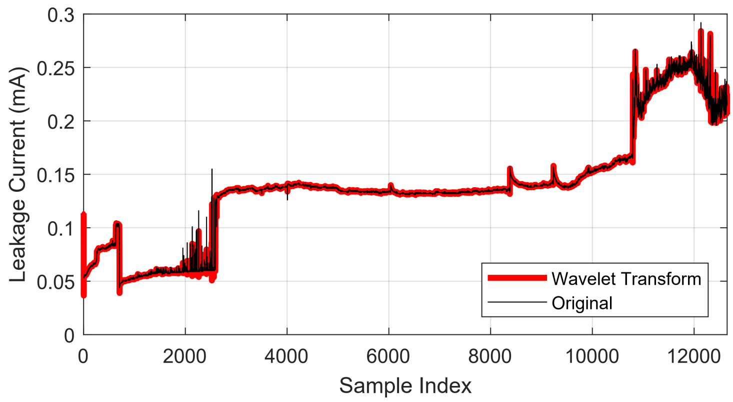

Figure 6. Wavelet Transform evaluation.

To evaluate the use of the Wavelet transform, all models were tested using different combinations of depth for the Wavelet packet tree. The results of this evaluation are presented in Table 7. The use of more than 1 node results in a loss of time series features, so this configuration was standardized.

The results of using the Wavelet transform to reduce signal noise were promising in all models except for the LSTM, in which the original signal prediction achieved better results

 doi:10.20944/preprints202208.0031.v1

of determination coefficients, time to convergence, and lower error. In GMDH, the use of the Wavelet transform with 2 depth levels had better results than predicting the original signal. In the ANFIS model, using 2 levels of depth, the RMS was close to the best result and had the best performance considering the other error metrics, for this reason, a depth of 2 levels was used in the ANFIS model, as in GMDH.

Table 7. Analysis using the Wavelet transform.

Model	Depth	RMSE	MAPE	MAE	R^2	Time (s)
Wavelet LSTM	1	3.83×10^{-2}	1.18×10^{-1}	2.72×10^{-2}	0.2314	86.24
	2	3.72×10^{-2}	1.13×10^{-1}	2.61×10^{-2}	0.2776	**70.33**
	3	$\mathbf{3.28}\times10^{-2}$	$\mathbf{1.12}\times10^{-1}$	$\mathbf{2.48}\times10^{-2}$	**0.4376**	73.87
	4	4.01×10^{-2}	1.46×10^{-1}	3.17×10^{-2}	0.1595	77.55
Wavelet GMDH	1	3.98×10^{-3}	5.79×10^{-3}	1.35×10^{-3}	0.9917	**0.22**
	2	$\mathbf{3.01}\times10^{-3}$	$\mathbf{4.09}\times10^{-3}$	$\mathbf{9.50}\times10^{-4}$	**0.9953**	0.29
	3	5.81×10^{-3}	1.01×10^{-2}	2.30×10^{-3}	0.9824	0.23
	4	3.94×10^{-3}	6.11×10^{-3}	1.42×10^{-3}	0.9919	0.24
Wavelet ANFIS	1	$\mathbf{1.56}\times10^{-3}$	5.31×10^{-4}	1.17×10^{-4}	**0.9987**	35.93
	2	1.58×10^{-3}	$\mathbf{5.48}\times10^{-5}$	$\mathbf{1.28}\times10^{-5}$	**0.9987**	42.54
	3	1.57×10^{-3}	1.76×10^{-4}	3.90×10^{-5}	**0.9987**	33.86
	4	1.57×10^{-3}	2.03×10^{-4}	4.52×10^{-5}	**0.9987**	**33.05**
Wavelet Ensemble	1	2.64×10^{-3}	8.46×10^{-3}	1.71×10^{-3}	0.9964	21.03
	2	3.62×10^{-3}	1.44×10^{-2}	2.88×10^{-3}	0.9931	**17.65**
	3	$\mathbf{2.46}\times10^{-3}$	$\mathbf{7.01}\times10^{-3}$	$\mathbf{1.43}\times10^{-3}$	**0.9968**	20.08
	4	3.12×10^{-3}	1.15×10^{-2}	2.31×10^{-3}	0.9949	19.90

The ensemble model had a promising result using 3 levels in the Wavelet transform. Only MAE and MAPE had superior results on the original signal after using the downsample algorithm. The use of the Wavelet transform had promising results when it was applied after the donwsample algorithm, when the analysis was performed using the downsample before the Wavelet transform all models had inferior results, not being a suitable strategy for this analysis.

Considering the use of the best configuration of the Wavelet transform, 100 runs were performed and the statistical results regarding RMSE are presented in Table 8. As can be observed the models that had the best performance are the ANFIS model and the ensemble, which were the same models that had the best performance without the use of the Wavelet transform.

Table 8. Statistical evaluation of models with Wavelet transform.

Model	Mean	Median	Std. Dev.	Variance
Wavelet LSTM	2.08×10^{-2}	2.06×10^{-2}	3.48×10^{-3}	1.21×10^{-5}
Wavelet GMDH	4.39×10^{-3}	4.29×10^{-3}	1.45×10^{-3}	2.10×10^{-6}
Wavelet ANFIS	$\mathbf{1.58}\times10^{-3}$	$\mathbf{1.58}\times10^{-3}$	$\mathbf{2.18}\times10^{-19}$	$\mathbf{4.75}\times10^{-38}$
Wavelet Ensemble	2.94×10^{-3}	2.91×10^{-3}	3.12×10^{-4}	9.76×10^{-8}

An interesting result that can be observed is that the LSTM had the worst performance in both analyses, being a model that is not suitable for this evaluation. Many authors have used the LSTM, however, this model may not be the most appropriate for forecasting depending on the signal used, as presented in this paper.

5. Conclusion

The use of time series prediction models to evaluate the development of faults in insulators based on leakage current shows promise, several models can be successfully applied to accomplish this task. The increase in contamination results in a consequent increase in leakage current until an electrical discharge occurs, which is an adequate way of assessing the development of the adverse condition to result in a failure. For this reason, leakage current must be monitored to keep the electrical power system operational.

The ANFIS subtractive clustering and the ensemble bagging stood out in predicting the leakage current, considering that they had lower error and better coefficient of determination. The results show that some hyperparameters have a major influence on the model performance, then it is necessary to perform a comparative analysis of all variations of the model to have an optimized algorithm. The statistical results showed that ANFIS is a stable model, resulting in low variance when several simulations are performed.

The application of the Wavelet transform resulted in an improvement in the predictive ability of most of the evaluated models, being a promising technique for noise reduction without the loss of signal characteristics. Most of the models presented stability when several simulations were performed, showing that these methods are reliable for the application presented in this paper. Based on the results, future work can be done to develop an embedded system for monitoring leakage current and indicating vulnerability to a disruptive discharge in distribution insulators. The leakage current proves to be a suitable indicator for monitoring the conditions of the power electrical system; this measurement can be applied to other insulating components of the electrical power grid.

Author Contributions: Investigation, writing-original draft, formal analysis, methodology, Nemesio Fava Sopelsa Neto; Software, validation, Stefano Frizzo Stefenon; Supervision, writing-review and editing, Luiz Henrique Meyer; Supervision, project administration, Valderi Reis Quietinho Leithardt.

Funding: This work was supported in part by the Spanish Agencia Estatal de Investigación, and in part by the Project Monitoring and Tracking Systems for the Improvement of Intelligent Mobility and Behavior Analysis (SiMoMIAC) under Grant PID2019-108883RB-C21/AEI/10.13039/501100011033.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data can be used upon request to the authors.

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

References

1. Kim, T.; Sanyal, S.; Koo, J.B.; Son, J.A.; Choi, I.H.; Yi, J. Analysis of Thermal Sensitivity by High Voltage Insulator Materials. *IEEE Access* **2020**, *8*, 75586–75591. doi:10.1109/ACCESS.2020.2987705.
2. 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 Generation, Transmission & Distribution* **2022**, *16*, 1865–1876. doi:10.1049/gtd2.12425.
3. Meyer, L.; Pintarelli, R. Inclined Plane Test for Erosion of Polymeric Insulators under AC and DC Voltages. *IEEE Latin America Transactions* **2020**, *18*, 1455–1461. doi:10.1109/TLA.2020.9111682.
4. Ilomuanya, C.; Nekahi, A.; Farokhi, S. A Study of the Cleansing Effect of Precipitation and Wind on Polluted Outdoor High Voltage Glass Cap and Pin Insulator. *IEEE Access* **2022**, *10*, 20669–20676. doi:10.1109/ACCESS.2022.3148709.
5. Wang, S.; Wu, Y. Effect of salty fog on flashover characteristics of OCS composite insulators. *Chinese Journal of Electrical Engineering* **2019**, *5*, 59–66. doi:10.23919/CJEE.2019.000021.
6. 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. *International Journal of Electrical Power & Energy Systems* **2022**, *134*, 107336. doi:10.1016/j.ijepes.2021.107336.
7. Nan, J.; Li, H.; Wan, X.; Xu, T.; Huo, F.; Lin, F. Study on Fast Contamination Characteristics of Cap and Pin Insulators in Straight Flow Wind Tunnel Simulation. *IEEE Access* **2021**, *9*, 125912–125919. doi:10.1109/ACCESS.2021.3110802.
8. Sezavar, H.R.; Fahimi, N.; Shayegani-Akmal, A.A. An Improved Dynamic Multi-Arcs Modeling Approach for Pollution Flashover of Silicone Rubber Insulator. *IEEE Transactions on Dielectrics and Electrical Insulation* **2022**, *29*, 77–85. doi:10.1109/TDEI.2022.3146531.
9. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Coelho, L.D.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. doi:10.1109/ACCESS.2021.3076410.
10. Alqudsi, A.Y.; Ghunem, R.A.; David, E. Analyzing the Role of Filler Interface on the Erosion Performance of Filled RTV Silicone Rubber under DC Dry-band Arcing. *IEEE Transactions on Dielectrics and Electrical Insulation* **2021**, *28*, 788–796. doi:10.1109/TDEI.2021.009337.
11. Wang, S.; Liu, Y.; Qing, Y.; Wang, C.; Lan, T.; Yao, R. Detection of Insulator Defects With Improved ResNeSt and Region Proposal Network. *IEEE Access* **2020**, *8*, 184841–184850. doi:10.1109/ACCESS.2020.3029857.

12. 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. doi:10.1109/ACCESS.2022.3161506.
13. Polisetty, S.; El-Hag, A.; Jayram, S. Classification of common discharges in outdoor insulation using acoustic signals and artificial neural network. *High Voltage* **2019**, *4*, 333–338. doi:10.1049/hve.2019.0113.
14. Jin, L.; Tian, Z.; Ai, J.; Zhang, Y.; Gao, K. Condition Evaluation of the Contaminated Insulators by Visible Light Images Assisted With Infrared Information. *IEEE Transactions on Instrumentation and Measurement* **2018**, *67*, 1349–1358. doi:10.1109/TIM.2018.2794938.
15. Pereira, F.; Crocker, P.; Leithardt, V.R. PADRES: Tool for PrivAcy, Data REgulation and Security. *SoftwareX* **2022**, *17*, 100895. doi:10.1016/j.softx.2021.100895.
16. 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 Generation, Transmission & Distribution* **2020**, *14*, 5667–5674. doi:10.1049/iet-gtd.2020.0814.
17. Kadiyala, A.; Kumar, A. Applications of python to evaluate the performance of bagging methods. *Environmental Progress & Sustainable Energy* **2018**, *37*, 1555–1559. doi:10.1002/ep.13018.
18. Sauer, J.; Mariani, V.C.; dos Santos Coelho, L.; Ribeiro, M.H.D.M.; Rampazzo, M. Extreme gradient boosting model based on improved Jaya optimizer applied to forecasting energy consumption in residential buildings. *Evolving Systems* **2021**, *Preprint*, 1–12. doi:10.1007/s12530-021-09404-2.
19. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Yow, K.C.; Mariani, V.C.; dos Santos Coelho, L.; Seman, L.O. Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam. *Electric Power Systems Research* **2022**, *202*, 107584. doi:10.1016/j.epsr.2021.107584.
20. Saha, S.; Saha, M.; Mukherjee, K.; Arabameri, A.; Ngo, P.T.T.; Paul, G.C. Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India. *Science of The Total Environment* **2020**, *730*, 139197. doi:10.1016/j.scitotenv.2020.139197.
21. da Silva, R.G.; Ribeiro, M.H.D.M.; Moreno, S.R.; Mariani, V.C.; dos Santos Coelho, L. A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting. *Energy* **2021**, *216*, 119174. doi:10.1016/j.energy.2020.119174.
22. Palangar, M.F.; Amin, U.; Bakhshayesh, H.; Ahmad, G.; Abu-Siada, A.; Mirzaie, M. Identification of Composite Insulator Criticality Based on a New Leakage Current Diagnostic Index. *IEEE Transactions on Instrumentation and Measurement* **2021**, *70*, 1–10. doi:10.1109/TIM.2021.3096869.
23. Ghunem, R.A.; El-Hag, A.H.; Marzinotto, M.; Nazir, M.T.; Wong, K.L.; Jayaram, S.H. Overhead Lines and Wildfires: Role of Outdoor Insulators: Prepared by a Task Force of the IEEE DEIS Outdoor Insulation Technical Committee. *IEEE Electrical Insulation Magazine* **2022**, *38*, 14–25. doi:10.1109/MEI.2022.9797260.
24. Yeh, C.T.; Thanh, P.N.; Cho, M.Y. Real-Time Leakage Current Classification of 15kV and 25kV Distribution Insulators Based on Bidirectional Long Short-Term Memory Networks With Deep Learning Machine. *IEEE Access* **2022**, *10*, 7128–7140. doi:10.1109/ACCESS.2022.3140479.
25. Gouda, O.E.; Darwish, M.M.F.; Mahmoud, K.; Lehtonen, M.; Elkhodragy, T.M. Pollution Severity Monitoring of High Voltage Transmission Line Insulators Using Wireless Device Based on Leakage Current Bursts. *IEEE Access* **2022**, *10*, 53713–53723. doi:10.1109/ACCESS.2022.3175515.
26. Liu, Y.; Du, B.X.; Farzaneh, M. Self-Normalizing Multivariate Analysis of Polymer Insulator Leakage Current Under Severe Fog Conditions. *IEEE Transactions on Power Delivery* **2017**, *32*, 1279–1286. doi:10.1109/TPWRD.2017.2650214.
27. Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2019**, *323*, 203–213. doi:10.1016/j.neucom.2018.09.082.
28. Elbaz, K.; Shen, S.L.; Zhou, A.; Yin, Z.Y.; Lyu, H.M. Prediction of Disc Cutter Life During Shield Tunneling with AI via the Incorporation of a Genetic Algorithm into a GMDH-Type Neural Network. *Engineering* **2021**, *7*, 238–251. doi:10.1016/j.eng.2020.02.016.
29. Kardani, N.; Bardhan, A.; Kim, D.; Samui, P.; Zhou, A. Modelling the energy performance of residential buildings using advanced computational frameworks based on RVM, GMDH, ANFIS-BBO and ANFIS-IPSO. *Journal of Building Engineering* **2021**, *35*, 102105. doi:10.1016/j.jobe.2020.102105.
30. Meira, E.; Cyrino Oliveira, F.L.; de Menezes, L.M. Forecasting natural gas consumption using Bagging and modified regularization techniques. *Energy Economics* **2022**, *106*, 105760. doi:10.1016/j.eneco.2021.105760.
31. Lu, H.; Cheng, F.; Ma, X.; Hu, G. Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: A case study of an intake tower. *Energy* **2020**, *203*, 117756. doi:10.1016/j.energy.2020.117756.
32. Talukdar, S.; Eibek, K.U.; Akhter, S.; Ziaul, S.; Towfiqul Islam, A.R.M.; Mallick, J. Modeling fragmentation probability of land-use and land-cover using the bagging, random forest and random subspace in the Teesta River Basin, Bangladesh. *Ecological Indicators* **2021**, *126*, 107612. doi:10.1016/j.ecolind.2021.107612.
33. 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. *International Journal of Electrical Power & Energy Systems* **2022**, *136*, 107712. doi:10.1016/j.ijepes.2021.107712.
34. Zhang, K.; Gençay, R.; Ege Yazgan, M. Application of wavelet decomposition in time-series forecasting. *Economics Letters* **2017**, *158*, 41–46. doi:10.1016/j.econlet.2017.06.010.

35. 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 Generation, Transmission & Distribution* **2021**, *16*, 1096–1107. doi:10.1049/gtd2.12353.
36. 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. doi:10.1109/ACCESS.2020.2993630.
37. Salem, A.A.; Abd-Rahman, R.; Rahiman, W.; Al-Gailani, S.A.; Al-Ameri, S.M.; Ishak, M.T.; Sheikh, U.U. Pollution Flashover Under Different Contamination Profiles on High Voltage Insulator: Numerical and Experiment Investigation. *IEEE Access* **2021**, *9*, 37800–37812. doi:10.1109/ACCESS.2021.3063201.
38. 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. *Electrical Engineering* **2022**, *104*, 615–622. doi:10.1007/s00202-021-01332-3.
39. Sopelsa Neto, N.F.; Stefenon, S.F.; Meyer, L.H.; Bruns, R.; Nied, A.; Seman, L.O.; Gonzalez, G.V.; Leithardt, V.R.Q.; Yow, K.C. A Study of Multilayer Perceptron Networks Applied to Classification of Ceramic Insulators Using Ultrasound. *Applied Sciences* **2021**, *11*, 1592. doi:10.3390/app11041592.
40. 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 Latin America Transactions* **2021**, *19*, 1278–1285. doi:10.1109/TLA.2021.9475858.
41. 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. *Journal of Intelligent & Fuzzy Systems* **2022**, *42*, 3285–3298. doi:10.3233/JIFS-211126.
42. Maraaba, L.S.; Soufi, K.Y.A.; Alhems, L.M.; Hassan, M.A. Performance Evaluation of 230 kV Polymer Insulators in the Coastal Area of Saudi Arabia. *IEEE Access* **2020**, *8*, 164292–164303. doi:10.1109/ACCESS.2020.3022521.
43. Azizi, S.; Momen, G.; Ouellet-Plamondon, C.; David, E. Performance improvement of EPDM and EPDM/Silicone rubber composites using modified fumed silica, titanium dioxide and graphene additives. *Polymer Testing* **2020**, *84*, 106281. doi:10.1016/j.polymertesting.2019.106281.
44. Yamashita, T.; Ishimoto, R.; Furusato, T. Influence of series resistance on dry-band discharge characteristics on wet polluted insulators. *IEEE Transactions on Dielectrics and Electrical Insulation* **2018**, *25*, 154–161. doi:10.1109/TDEI.2018.007005.
45. Madi, A.M.; He, Y.; Jiang, L. Design and testing of an improved profile for silicone rubber composite insulators. *IEEE Transactions on Dielectrics and Electrical Insulation* **2017**, *24*, 2930–2936. doi:10.1109/TDEI.2017.006170.
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 Sensors Journal* **2019**, *19*, 5266–5273. doi:10.1109/JSEN.2019.2902192.
47. Yin, C.; Guo, Y.; Zhang, X.; Huang, G.; Wu, G. A Novel Method for Visualizing the Pollution Distribution of Insulators. *IEEE Transactions on Instrumentation and Measurement* **2021**, *70*, 1–8. doi:10.1109/TIM.2021.3098793.
48. 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. doi:10.3390/computers10090112.
49. Salem, A.A.; Abd-Rahman, R.; Al-Gailani, S.A.; Salam, Z.; Kamarudin, M.S.; Zainuddin, H.; Yousof, M.F.M. Risk Assessment of Polluted Glass Insulator Using Leakage Current Index Under Different Operating Conditions. *IEEE Access* **2020**, *8*, 175827–175839. doi:10.1109/ACCESS.2020.3026136.
50. 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. doi:10.3390/s22134859.
51. Lan, L.; Zhang, G.; Wang, Y.; Wen, X.; Wang, W.; Pei, H. The Influence of Natural Contamination on Pollution Flashover Voltage Waveform of Porcelain Insulators in Heavily Polluted Area. *IEEE Access* **2019**, *7*, 121395–121406. doi:10.1109/ACCESS.2019.2936868.
52. 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 Latin America Transactions* **2017**, *15*, 884–889. doi:10.1109/TLA.2017.7910202.
53. Ilhan, S.; Cherney, E.A. Comparative tests on RTV silicone rubber coated porcelain suspension insulators in a salt-fog chamber. *IEEE Transactions on Dielectrics and Electrical Insulation* **2018**, *25*, 947–953. doi:10.1109/TDEI.2018.006968.
54. 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. *Journal of Intelligent & Fuzzy Systems* **2021**, *40*, 1083–1096. doi:10.3233/JIFS-201279.
55. Tealab, A. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal* **2018**, *3*, 334–340. doi:10.1016/j.fcij.2018.10.003.
56. Abbasimehr, H.; Shabani, M.; Yousefi, M. An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering* **2020**, *143*, 106435. doi:10.1016/j.cie.2020.106435.
57. Salem, A.A.; Lau, K.Y.; Abdul-Malek, Z.; Al-Gailani, S.A.; Tan, C.W. Flashover voltage of porcelain insulator under various pollution distributions: Experiment and modeling. *Electric Power Systems Research* **2022**, *208*, 107867. doi:10.1016/j.epsr.2022.107867.
58. Cao, J.; Li, Z.; Li, J. Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications* **2019**, *519*, 127–139. doi:10.1016/j.physa.2018.11.061.

59. Wang, W.; Du, Y.; Chau, K.; Chen, H.; Liu, C.; Ma, Q. A Comparison of BPNN, GMDH, and ARIMA for Monthly Rainfall Forecasting Based on Wavelet Packet Decomposition. *Water* **2021**, *13*, 2871. doi:10.3390/w13202871.
60. Zardkoohi, M.; Fatemeh Molaeezadeh, S. Long-term prediction of blood pressure time series using ANFIS system based on DKFCM clustering. *Biomedical Signal Processing and Control* **2022**, *74*, 103480. doi:10.1016/j.bspc.2022.103480.
61. Zhang, S.; Chen, Y.; Zhang, W.; Feng, R. A novel ensemble deep learning model with dynamic error correction and multi-objective ensemble pruning for time series forecasting. *Information Sciences* **2021**, *544*, 427–445. doi:10.1016/j.ins.2020.08.053.
62. Liu, F.; Cai, M.; Wang, L.; Lu, Y. An Ensemble Model Based on Adaptive Noise Reducer and Over-Fitting Prevention LSTM for Multivariate Time Series Forecasting. *IEEE Access* **2019**, *7*, 26102–26115. doi:10.1109/ACCESS.2019.2900371.
63. Abbasimehr, H.; Paki, R. Improving time series forecasting using LSTM and attention models. *Journal of Ambient Intelligence and Humanized Computing* **2022**, *13*, 673–691. doi:10.1007/s12652-020-02761-x.
64. Smyl, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting* **2020**, *36*, 75–85. M4 Competition, doi:10.1016/j.ijforecast.2019.03.017.
65. Sagheer, A.; Kotb, M. Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems. *Scientific reports* **2019**, *9*, 1–16. doi:10.1038/s41598-019-55320-6.
66. 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 Science, Measurement & Technology* **2020**, *14*, 953–961. doi:10.1049/iet-smt.2020.0083.
67. Smagulova, K.; James, A.P. A survey on LSTM memristive neural network architectures and applications. *The European Physical Journal Special Topics* **2019**, *228*, 2313–2324. doi:10.1140/epjst/e2019-900046-x.
68. Han, Y.; Wang, C.; Ren, Y.; Wang, S.; Zheng, H.; Chen, G. Short-Term Prediction of Bus Passenger Flow Based on a Hybrid Optimized LSTM Network. *ISPRS International Journal of Geo-Information* **2019**, *8*, 366. doi:10.3390/ijgi8090366.
69. Chang, Z.; Zhang, Y.; Chen, W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy* **2019**, *187*, 115804. doi:10.1016/j.energy.2019.07.134.
70. Lin, J.; Li, H.; Liu, N.; Gao, J.; Li, Z. Automatic Lithology Identification by Applying LSTM to Logging Data: A Case Study in X Tight Rock Reservoirs. *IEEE Geoscience and Remote Sensing Letters* **2021**, *18*, 1361–1365. doi:10.1109/LGRS.2020.3001282.
71. Dodangeh, E.; Panahi, M.; Rezaie, F.; Lee, S.; Tien Bui, D.; Lee, C.W.; Pradhan, B. Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. *Journal of Hydrology* **2020**, *590*, 125423. doi:10.1016/j.jhydrol.2020.125423.
72. Roshani, M.; Sattari, M.A.; Muhammad Ali, P.J.; Roshani, G.H.; Nazemi, B.; Corniani, E.; Nazemi, E. Application of GMDH neural network technique to improve measuring precision of a simplified photon attenuation based two-phase flowmeter. *Flow Measurement and Instrumentation* **2020**, *75*, 101804. doi:10.1016/j.flowmeasinst.2020.101804.
73. Roshani, M.; Phan, G.; Hossein Roshani, G.; Hanus, R.; Nazemi, B.; Corniani, E.; Nazemi, E. Combination of X-ray tube and GMDH neural network as a nondestructive and potential technique for measuring characteristics of gas-oil–water three phase flows. *Measurement* **2021**, *168*, 108427. doi:10.1016/j.measurement.2020.108427.
74. Al-qaness, M.A.; Ewees, A.A.; Fan, H.; Abualigah, L.; Elaziz, M.A. Boosted ANFIS model using augmented marine predator algorithm with mutation operators for wind power forecasting. *Applied Energy* **2022**, *314*, 118851. doi:10.1016/j.apenergy.2022.118851.
75. Suparta, W.; Samah, A.A. Rainfall prediction by using ANFIS times series technique in South Tangerang, Indonesia. *Geodesy and Geodynamics* **2020**, *11*, 411–417. doi:10.1016/j.geog.2020.08.001.
76. Ribeiro, G.T.; Mariani, V.C.; dos Santos Coelho, L. Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Engineering Applications of Artificial Intelligence* **2019**, *82*, 272–281. doi:10.1016/j.engappai.2019.03.012.
77. Ribeiro, M.H.D.M.; Mariani, V.C.; dos Santos Coelho, L. Multi-step ahead meningitis case forecasting based on decomposition and multi-objective optimization methods. *Journal of Biomedical Informatics* **2020**, *111*, 103575. doi:10.1016/j.jbi.2020.103575.
78. Divina, F.; Gilson, A.; Goméz-Vela, F.; García Torres, M.; Torres, J.F. Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. *Energies* **2018**, *11*, 949. doi:10.3390/en11040949.
79. Gao, X.; Shan, C.; Hu, C.; Niu, Z.; Liu, Z. An Adaptive Ensemble Machine Learning Model for Intrusion Detection. *IEEE Access* **2019**, *7*, 82512–82521. doi:10.1109/ACCESS.2019.2923640.
80. da Silva, R.G.; Ribeiro, M.H.D.M.; Mariani, V.C.; dos Santos Coelho, L. Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables. *Chaos, Solitons & Fractals* **2020**, *139*, 110027. doi:10.1016/j.chaos.2020.110027.
81. Tan, M.; Yuan, S.; Li, S.; Su, Y.; Li, H.; He, F. Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning. *IEEE Transactions on Power Systems* **2020**, *35*, 2937–2948. doi:10.1109/TPWRS.2019.2963109.
82. Sun, S.; Wang, S.; Zhang, G.; Zheng, J. A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Solar Energy* **2018**, *163*, 189–199. doi:10.1016/j.solener.2018.02.006.
83. Ribeiro, M.H.D.M.; Stefenon, S.F.; de Lima, J.D.; Nied, A.; Mariani, V.C.; Coelho, L.S. Electricity Price Forecasting Based on Self-Adaptive Decomposition and Heterogeneous Ensemble Learning. *Energies* **2020**, *13*, 5190. doi:10.3390/en13195190.
84. Boiroux, D.; Jørgensen, J.B. Sequential L1 Quadratic Programming for Nonlinear Model Predictive Control. *IFAC Symposium on Dynamics and Control of Process Systems* **2019**, *52*, 474–479. doi:10.1016/j.ifacol.2019.06.107.

85. Vijayan S, V.; Mohanta, H.K.; Pani, A.K. Support vector regression modeling in recursive just-in-time learning framework for adaptive soft sensing of naphtha boiling point in crude distillation unit. *Petroleum Science* **2021**, *18*, 1230–1239. doi:10.1016/j.petsci.2021.07.001.
86. Gu, B.; Shan, Y.; Quan, X.; Zheng, G. Accelerating Sequential Minimal Optimization via Stochastic Subgradient Descent. *IEEE Transactions on Cybernetics* **2021**, *51*, 2215–2223. doi:10.1109/TCYB.2019.2893289.
87. Lv, Y.; Peng, S.; Yuan, Y.; Wang, C.; Yin, P.; Liu, J.; Wang, C. A classifier using online bagging ensemble method for big data stream learning. *Tsinghua Science and Technology* **2019**, *24*, 379–388. doi:10.26599/TST.2018.9010119.
88. Tanha, J.; Abdi, Y.; Samadi, N.; Razzaghi, N.; Asadpour, M. Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data* **2020**, *7*, 1–47. doi:10.1186/s40537-020-00349-y.
89. Yaman, M.A.; Subasi, A.; Rattay, F. Comparison of Random Subspace and Voting Ensemble Machine Learning Methods for Face Recognition. *Symmetry* **2018**, *10*, 651. doi:10.3390/sym10110651.
90. Wen, L.; Hughes, M. Coastal Wetland Mapping Using Ensemble Learning Algorithms: A Comparative Study of Bagging, Boosting and Stacking Techniques. *Remote Sensing* **2020**, *12*, 1683. doi:10.3390/rs12101683.
91. Kim, D.; Baek, J.G. Bagging ensemble-based novel data generation method for univariate time series forecasting. *Expert Systems with Applications* **2022**, *203*, 117366. doi:10.1016/j.eswa.2022.117366.
92. Truong, X.L.; Mitamura, M.; Kono, Y.; Raghavan, V.; Yonezawa, G.; Truong, X.Q.; Do, T.H.; Tien Bui, D.; Lee, S. Enhancing Prediction Performance of Landslide Susceptibility Model Using Hybrid Machine Learning Approach of Bagging Ensemble and Logistic Model Tree. *Applied Sciences* **2018**, *8*, 1046. doi:10.3390/app8071046.
93. Jiang, Y.; Ge, H.; Zhang, Y. Quantitative analysis of wheat maltose by combined terahertz spectroscopy and imaging based on Boosting ensemble learning. *Food Chemistry* **2020**, *307*, 125533. doi:10.1016/j.foodchem.2019.125533.
94. Kumari, P.; Toshniwal, D. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *Journal of Cleaner Production* **2021**, *279*, 123285. doi:10.1016/j.jclepro.2020.123285.
95. Basaran, K.; Özçift, A.; Kılınç, D. A new approach for prediction of solar radiation with using ensemble learning algorithm. *Arabian Journal for Science and Engineering* **2019**, *44*, 7159–7171. doi:10.1007/s13369-019-03841-7.
96. Liang, W.; Sari, A.; Zhao, G.; McKinnon, S.D.; Wu, H. Short-term rockburst risk prediction using ensemble learning methods. *Natural Hazards* **2020**, *104*, 1923–1946. doi:10.1007/s11069-020-04255-7.
97. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1998**, *20*, 832–844. doi:10.1109/34.709601.
98. Pham, B.T.; Tien Bui, D.; Prakash, I.; Dholakia, M. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149*, 52–63. doi:10.1016/j.catena.2016.09.007.
99. Pham, B.T.; Prakash, I.; Tien Bui, D. Spatial prediction of landslides using a hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. *Geomorphology* **2018**, *303*, 256–270. doi:10.1016/j.geomorph.2017.12.008.
100. Li, G.; Zheng, Y.; Liu, J.; Zhou, Z.; Xu, C.; Fang, X.; Yao, Q. An improved stacking ensemble learning-based sensor fault detection method for building energy systems using fault-discrimination information. *Journal of Building Engineering* **2021**, *43*, 102812. doi:10.1016/j.jobe.2021.102812.
101. Wu, T.; Zhang, W.; Jiao, X.; Guo, W.; Alhaj Hamoud, Y. Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Computers and Electronics in Agriculture* **2021**, *184*, 106039. doi:10.1016/j.compag.2021.106039.
102. 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. doi:10.3390/en13020484.
103. El-Hendawi, M.; Wang, Z. An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting. *Electric Power Systems Research* **2020**, *182*, 106265. doi:10.1016/j.epsr.2020.106265.
104. Sui, X.; Wan, K.; Zhang, Y. Pattern recognition of SEMG based on wavelet packet transform and improved SVM. *Optik* **2019**, *176*, 228–235. doi:10.1016/j.ijleo.2018.09.040.
105. Tayab, U.B.; Zia, A.; Yang, F.; Lu, J.; Kashif, M. Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform. *Energy* **2020**, *203*, 117857. doi:10.1016/j.energy.2020.117857.
106. Stefenon, S.F.; Ribeiro, M.H.D.M.; Nied, A.; Mariani, V.C.; Coelho, L.S.; da Rocha, D.F.M.; Grebogi, R.B.; Ruano, A.E.B. Wavelet group method of data handling for fault prediction in electrical power insulators. *International Journal of Electrical Power & Energy Systems* **2020**, *123*, 106269. doi:10.1016/j.ijepes.2020.106269.
107. Yesilli, M.C.; Khasawneh, F.A.; Otto, A. On transfer learning for chatter detection in turning using wavelet packet transform and ensemble empirical mode decomposition. *CIRP Journal of Manufacturing Science and Technology* **2020**, *28*, 118–135. doi:10.1016/j.cirpj.2019.11.003.
108. Wang, L.; Liu, Z.; Cao, H.; Zhang, X. Subband averaging kurtogram with dual-tree complex wavelet packet transform for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing* **2020**, *142*, 106755. doi:10.1016/j.ymssp.2020.106755.
109. Arshad.; Ahmad, J.; Tahir, A.; Stewart, B.G.; Nekahi, A. Forecasting Flashover Parameters of Polymeric Insulators under Contaminated Conditions Using the Machine Learning Technique. *Energies* **2020**, *13*, 3889. doi:10.3390/en13153889.
110. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing* **2020**, *86*, 105837. doi:10.1016/j.asoc.2019.105837.

111. Kasburg, C.; Stefenon, S.F. Deep Learning for Photovoltaic Generation Forecast in Active Solar Trackers. *IEEE Latin America Transactions* **2019**, *17*, 2013–2019. doi:10.1109/TLA.2019.9011546.
112. de Lima, R.R.; Fernandes, A.M.R.; Bombasar, J.R.; da Silva, B.A.; Crocker, P.; Leithardt, V.R.Q. An Empirical Comparison of Portuguese and Multilingual BERT Models for Auto-Classification of NCM Codes in International Trade. *Big Data and Cognitive Computing* **2022**, *6*, 8. doi:10.3390/bdcc6010008.
113. 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. *Journal of Intelligent & Fuzzy Systems* **2022**, *6*, 6221–6234. doi:10.3233/JIFS-212788.
114. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* **2021**, *7*, e623. doi:10.7717/peerj-cs.623.