Università degli studi di Udine

Analysis of the ultrasonic signal in polymeric contaminated insulators through ensemble learning methods

(Article begins on next page)

04 May 2024

Analysis of the Ultrasonic Signal in Polymeric Contaminated Insulators Through Ensemble Learning Methods

STÉFANO FRIZZO STEFENON[1,2]**, RAFAEL BRUNS**[3]**, ANDREZA SARTORI**[3,4]**,**
LUIZ HENRIQUE MEYER[3]**, (Senior Member, IEEE), RAÚL GARCÍA OVEJERO**[5]**,**
AND VALDERI REIS QUIETINHO LEITHARDT[6,7]**, (Member, IEEE)**

[1]Fondazione Bruno Kessler, Povo, 38123 Trento, Italy
[2]Department of Computer Science and Artificial Intelligence, University of Udine, 33100 Udine, Italy
[3]Department of Electrical Engineering, Regional University of Blumenau, Blumenau 89030-000, Brazil
[4]Department of Information Systems and Computing, Regional University of Blumenau, Blumenau 89030-903, Brazil
[5]Expert Systems and Applications Laboratory, E.T.S.I.I. of Béjar, Universidad de Salamanca, 37700 Salamanca, Spain
[6]COPELABS, Lusófona University of Humanities and Technologies, 1749-024 Lisboa, Portugal
[7]VALORIZA, Research Center for Endogenous Resources Valorization, Instituto Politécnico de Portalegre, 7300-555 Portalegre, Portugal

Corresponding author: Stéfano Frizzo Stefenon (sfrizzostefenon@fbk.eu)

ABSTRACT Outdoor insulators may experience stress due to severe environmental conditions, such as pollution and contamination. Through the identification of partial discharges by ultrasonic noise, it is possible to assess the possibility of a power grid failure occurring. In this paper, ensemble models are used to analyze an ultrasonic signal from an ultrasonic microphone Pettersson M500. As the insulators are susceptible to developing irreversible failures, it will be evaluated whether the ultrasonic signal will remain over time, so that it is possible to assess whether the discharges being captured can result in a failure in contaminated polymeric insulators, evaluated in a high voltage laboratory under controlled conditions. The ensemble models were used in this paper because they typically require less computational effort than techniques based on deep learning and have acceptable performance for the problem at hand. The bagging, boosting, random subspace, bagging plus random subspace, and stacked generalization ensemble models are evaluated, and the best result of each model is used to compare the differences between the models. The bagging ensemble learning model proved to be faster and have lower error than other ensemble models, long short-term memory (LSTM), and nonlinear autoregressive (NAR).

INDEX TERMS Deep learning, machine learning, insulator testing.

I. INTRODUCTION

To guarantee the supply of electricity to consumers, it is necessary to have a reliable power grid. Electrical power system inspections are the key to early fault identification and predictive maintenance [1]. Usually, during the insulation degradation process, there are various types of manifestations that can be caught with appropriate equipment. By monitoring such manifestations in regular field inspections, the reliability of the system can be increased [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Fabio Mottola.

There are several types of insulators degradation processes, one of them that is common to be observed in inspections is the build up of contamination on the surface of the insulators in rural or polluted areas [3]. When contamination is strongly attached to the surface of the insulator, its conductivity may increase, making this component more vulnerable to flashovers [4]. Some techniques to identify the contamination can be applied for the diagnosis of the insulator's condition [5].

According to Lv, Zhao, and Song [6] the non-soluble deposit density in a fog environment can be analyzed using the finite element method (FEM) to evaluate the electric field

in transmission line insulators. The FEM stands out for the flexibility of the evaluation of parameter changes in electrical equipment, thus, profile variations can be evaluated without the need to develop several prototypes [7].

The composite materials shows advantages in this application and are being widely used, then, it becomes increasingly important to assess the influence of contamination on these new components [8]. A major difference of composite materials in relation to porcelain insulators is their hydrophobicity capacity, reducing the levels of the leakage current, which may also reduce the degradation process, thus improving anti-pollution performance [9].

The prediction of time series can be performed using several models [10]–[12]. There is no model that is suitable for all signal variations. Thus, to reach a better accuracy is necessary to compare the models that are used for this purpose [13]. Based on this statement the contributions of this paper are:

- The ensemble learning methods are adequate for time series prediction because they require less computational effort compared to deep learning structures. To evaluate that, a comparison between the ensemble models and the long short-term memory (LSTM) is presented. The LSTM is an algorithm applied in deep learning that is widely used for time series prediction.
- The second contribution is related to the possibility of variation in the structure, based on different arrangements it is possible to apply a more efficient structure. In this paper the bagging, boosting, random subspace, bagging plus random subspace, and stacked generalization models are evaluated.
- Finally, the third contribution is related to the analysis of a chaotic time series. Signals with high noise are difficult to be analyzed as the high frequencies can have a nonlinear pattern. For this reason, the algorithm needs to be robust to have a reliable result. The necessary feature to deal with this type of data is found in the ensemble models that are presented in this paper.

The continuation of this paper is organized as follows: In Section II related works are presented. Section III discusses the characteristics of polymeric insulators and presents the laboratory test, carried out to detect the ultrasonic noise of contaminated insulators. In Section IV the Ensemble Learning Models are presented. In Section V the results of the analysis are discussed and finally, in Section VI the conclusion is described.

II. RELATED WORK

To improve the diagnosis of possible failures in electrical power components, an approach that has proven to be successful is the prediction of failures [14], which is the specific subject addressed in this paper. Among the techniques used to improve the predictive capacity of the model, the wavelet transform stands out for having the ability to reduce the noise in the signal without losing its characteristic [15]. Other techniques widely used nowadays are the LSTM and the

support vector machine (SVM) that can be applied to different type of problems [16].

One of the practical difficulties in identifying an insulator with damaged properties is that defects can be hidden under the mooring cables or under contamination [17]. Identifying a failed component during an inspection of the electrical power system requires great experience from the operator [18], even then the human factor may bring uncertainties to the process [19].

Based on the growing use of polymeric insulators and the need to diagnose their conditions [20], this paper presents an evaluation of the time series prediction, to determine the development of a fault by verifying the signal variation. The evaluation presented in this paper is related in terms of the ability to predict a chaotic signal emitted by an ultrasound detector, which is a specific equipment for inspection of the electrical power system.

Increased contamination on the surface of an insulator results in a cumulative loss of its insulating properties, resulting in a greater likelihood of discharges occurring in its surroundings. Partial discharges, which usually occur during the degradation process, emit ultrasonic noise that can be identified with specific equipment [21]. From the prediction of increased discharges or a variation in the ultrasonic signal, it is possible to predict a flashover before it happens.

Pre-processing the ultrasonic signal is a strategy to improve the ability to predict a failure, however, many algorithms that filter the signal can reduce relevant information. This becomes a major challenge for prediction models, considering that with more nonlinearities it is more difficult to train the model properly. Some studies have been carried out to classify the condition of insulators using acoustic signals [22]. In addition to classification, it is promising to evaluate the capabilities of predicting the signal of a contaminated insulator, considering that it is more susceptible to failure.

Nowadays, one of the approaches that are being increasingly used is the ensemble learning method, mainly because of their lower computational effort compared to other techniques. When the problem is divided into smaller problems to be solved by simpler combined models, an efficient framework is obtained for dealing with complex problems [23]. As it is possible to combine the weak models in different ways to obtain a robust structure, it is possible to carry out variations in the architecture until a suitable model is obtained to be used in the problem in question. Because of this ability to adapt the model structure to the problem and require less computational effort, the ensemble learning models stand out to predict chaotic time series [24].

III. OUTDOOR POLYMERIC INSULATORS

High voltage polymeric insulators that are used outdoors are susceptible to tracking and erosion due to contamination, since hidrophobicity eventually is lost [25]. From the beginning of these effects there is a tendency to continue until there is a flashover and the insulator is damaged [26].
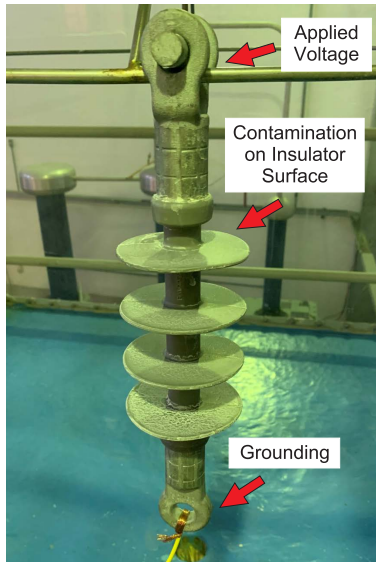
FIGURE 1. Contaminated long rod polymer insulator.

FIGURE 2. Laboratory analysis setup.

Based on this, there will be a failure in the electrical power system that may cause a disconnection from the grid and the components will need to be replaced through corrective maintenance, leaving the system off [27]–[29]. Therefore, it is necessary to detect these effects before a disruptive fault occurs. The ultrasound can be used to assess the level of contamination, making possible to perform predictive maintenance, cleaning the electrical power grid, or changing the component that may develop the failure [30].

The development of the fault can be linked to the increase in a measure of the insulator's condition. The leakage current is a measure that can be used to evaluate the time series measured according to the increase of contamination, thus being possible to forecast the development of a failure by the variation of this measure [2]. Regarding the time series forecasting, several approaches are being studied to obtain a better performance with lower computational cost, such as the ensemble learning models which will be presented in the next section.

According to Meyer and Pintarelli [31] the polymeric insulators are being used in electricity distribution networks in Brazil, considering that they are lighter and therefore easily installed. A long rod 24.2 kV class polymer insulator for anchoring the electrical distribution network [32], with surface contamination was evaluated. This component is presented in Figure 1, where the electrical potential is applied to the top of the insulator and ground is connected to the bottom.

A. LABORATORY SETUP

To simulate the contamination found in the field, a solid contamination method was used with Kaolin inside the salt spray chamber (as shown in Figure 2). Kaolin is an ore composed of hydrated aluminium silicates that are used to simulate contamination in outdoor insulators. Initially, the
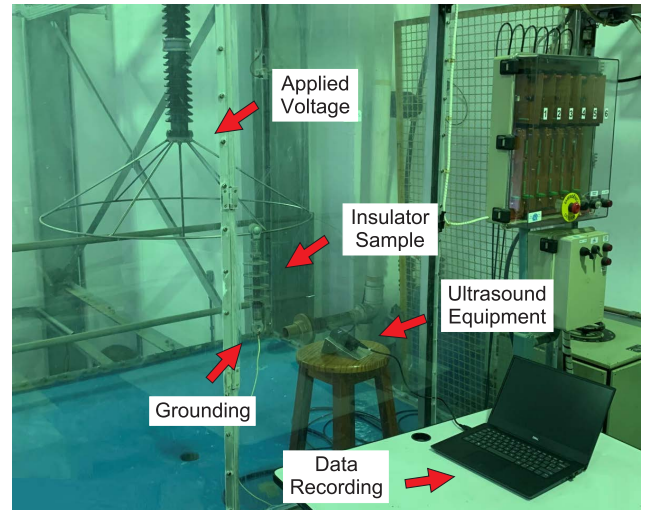
insulators are cleaned with isopropyl alcohol and dried in an oven, the kaolin is weighed with a precision scale to have the exact concentration to simulate contamination. The insulator is immersed in a glass beaker containing a slurry of Kaolin. The definition of the quantity of salt used for contamination is determined by measuring the conductivity during sample preparation [33].

After the contamination is evenly distributed over the insulation surface, the insulator is dried in the oven and installed in the chamber for measurements. The chamber has a volume of 8 m^3, being 2 m high, 2 m wide and 2 m deep. The salt spray chamber is designed for laboratory analysis to be carried out under controlled conditions. To perform the artificial contamination contamination the international electrotechnical commission (IEC-507) [34] was used, which is specific for determining the characteristics of the tolerably of artificial pollution. This standard is developed by the insulator studies commission for overhead lines and substations, which is specific for tests on artificial pollution in high voltage insulators [35].

The chamber has a conductive metal arc that is energized with high voltage to simulate the electrical cables, this arc is connected to a porcelain bushing, which in turn is connected to an external muffle which is connected to an isolated medium voltage cable, which in its other termination also has an external muffle connected. The muffle is connected to the high voltage of a single-phase transformer. The low voltage of the transformer is connected to a voltage regulator that is connected to the mains. The voltage is measured by a multimeter connected to a high voltage probe at the high voltage terminal of the transformer.

The chamber was used because it is completely sealed and helps to reduce possible external interference in the ultrasound measurements. In this way, the insulator was connected to the ring with electrical voltage applied and the other side of the insulator was grounded. The microphone was positioned inside the chamber directed to the subject
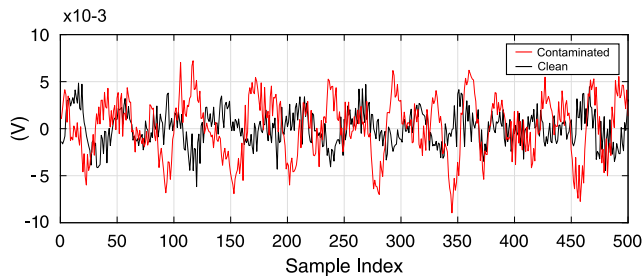
FIGURE 3. Ultrasonic signal recorded in the laboratory of the clean and contaminated insulators.

insulator, at a distance of approximately 50 cm from the insulator. The microphone was connected to a notebook and then measurements were made.

In this paper, 10,000 samples recorded in the measurement were considered. In this way, all ultrasonic frequencies are within the analyzed spectrum and the signal becomes processable [36]. Due to the high amount of the data, the hold-out approach [37] is adopted in this paper. For training and testing 70% and 30% were respectively used in the machine learning models.

For comparative purposes, Figure 3 presents the image of the signal recorded by the ultrasound detector of an insulator in good condition (which is clean and new), and an insulator that is contaminated. A window of the recorded signals was presented with only 500 samples, then it is possible to better visualize the differences. In this work, in particular, the signal evaluated was that of the contaminated insulator.

It can be observed that the signal of the contaminated insulator has a greater amplitude as well as variations containing lower frequencies. The visual difference between the signals is small, which makes it difficult to identify a fault in the field. When an insulator develops its leakage current the signal amplitude becomes even higher with frequencies that have its origin in partial discharges and/or dry band arcing [38].

Due to the large amount of non-linearities in the signal caused by discharges, forecasting the time series of a signal based on ultrasonic noise is a difficult task, making it necessary to use advanced prediction models such as deep learning strategies or techniques that combine several weaker learners to obtain a stronger model, such as ensemble-based techniques that will be the focus of this paper.

The identification that there is an increase in the amplitude of the signal is possible through advanced artificial intelligence techniques [39]. The choice of which technique is most suitable for the problem is difficult, since some techniques have a high computational effort [40]. Based on the problem presented in this section, ensemble models are used to predict the signal produced by ultrasound, these models are presented in the next section.

IV. ENSEMBLE LEARNING MODEL

The ensemble learning models are approaches that combine weak learners to obtain an algorithm with greater regression capability. These methods have currently stood out in applications related to the electrical system, as their main advantage is their superior convergence speed compared to deep learning strategies [41].

In this paper, the weak learners used for the ensemble models are the support vector regression (SVR) type [42]. The SVR to perform the relationship between input and output of the data is given by:

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}) + b \qquad (1)$$

where $f(\mathbf{x})$ is the forecasting values and ϕ is the mapping of the input vector $\mathbf{x}$ [43]; $\mathbf{w}$ and b are adjustable the coefficients calculated by minimizing the risk function (R):

$$R(f) = \frac{1}{N}\sum_{i=1}^{N}L_{\varepsilon}\left(\mathbf{y}_i, \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_i) + b\right) \qquad (2)$$

where L_{ε} is the loss function to penalize the training errors, calculated by:

$$L_{\varepsilon}(\mathbf{y}, f(\mathbf{x})) = \begin{cases} 0 & if\ |\mathbf{y} - f(\mathbf{x})| \leq \varepsilon \\ |\mathbf{y} - f(\mathbf{x})| - \varepsilon & otherwise \end{cases} \qquad (3)$$

The use of the loss function in the regularized function leads to a quadratic programming problem [44]. The minimization of the regularized function can be rewritten as the equivalent optimization problem, which is often called the *primal* problem:

$$\min_{(w,b,\xi,\xi^*)}\ R(w, b, \xi, \xi^*)$$

$$= C\sum_{i=1}^{n}(\psi_s(\xi_i) + \psi_s(\xi_i^*)) + \frac{1}{2}\|w\|^2 \qquad (4)$$

$$\text{subject to}\ \begin{cases} y_i - \langle w \cdot \phi(x_i)\rangle - b \leq (1-\beta)\epsilon + \xi_i; \\ \langle w \cdot \phi(x_i)\rangle + b - y_i \leq (1-\beta)\epsilon + \xi_i^*; \\ \xi_i^* \geq 0;\quad \xi_i^* \geq 0\ \forall i \end{cases} \qquad (5)$$

where

$$\psi_s(\xi) = \begin{cases} \dfrac{\xi^2}{4\beta\epsilon} & if\ \xi \in [0, 2\beta\epsilon]; \\ \xi - \beta\epsilon & if\ \xi \in [2\beta\epsilon, +\infty]. \end{cases} \qquad (6)$$

Simplifying the *dual* problem:

$$\min_{(\alpha,\alpha^*)}\ R(\alpha, \alpha^*)$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_i^*)K(x_i, x_j)$$

$$- \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i + \sum_{i=1}^{n}(\alpha_i + \alpha_i^*)(1 - \beta)\epsilon$$

$$+ \frac{\beta\epsilon}{C}\sum_{i=1}^{n}(\alpha_i^2 + \alpha_i^{*2}) \qquad (7)$$

subject to $0 \leq \alpha_i \leq C,\quad \forall_i, 0 \leq \alpha_i^* \leq C,\ \forall_i$

$$and\ \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0. \qquad (8)$$

Given the Karush-Kuhn-Tucker (KKT) conditions for the *primal* problem [45], the *dual* form of the regression function is:

$$f(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x_j}) + b \qquad (9)$$

wherein α_i and α_i^* are the Lagrangian multipliers [46].

The kernel functions $K(x_i, x_j)$ used in this paper were linear (LIN) (10), radial basis function (RBF) (11), and polynomial (POLY) (12).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \prime \, \mathbf{x}_j. \qquad (10)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2). \qquad (11)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left(1 + \mathbf{x}_i \prime \, \mathbf{x}_j\right)^q. \qquad (12)$$

The optimizers used in the SVR were soft-margin minimization via quadratic programming (L1QP), iterative single data algorithm (ISDA), and sequential minimal optimization (SMO) [47].

Using the L1QP a linear approximation is considered in the space resource. The approximation function is calculated by minimizing the approximation error for training and response data [48]. This is accomplished by minimizing:

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{M} \zeta_i \qquad (13)$$

subject to

$$y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i \qquad (14)$$

for

$$i = 1, \ldots, M \qquad (15)$$

where M is the number of training data, w is the weight vector, x_i is the training data, y_i is the response variable, C is the margin parameters, and ζ_i is the positive slack variable.

ISDA is designed to avoid the use of typical solvers [49]. The important feature of the algorithm is that it deals with one data point at a time to develop the objective function. SMO systematically solves many small optimization problems that are divided into subsets, including only 2 Lagrange multipliers at a time [50].

One of the great difficulties in using ensemble models is that there is a wide range of approaches and defining the best strategy is a difficult task [51]. Based on this premise, this paper aims to compare variations of these algorithms for a chaotic time series prediction problem, specifically, bagging, boosting, random subspace, and stacking ensemble learning models will be evaluated. The differences between these models are mainly given by the way that the weak learners are organized, being usually these structures bagging, boosting, random subspace, and stacked generalization.

The bagging ensemble learning will focus on getting an ensemble model with less variance than its components, with stacking mainly trying to produce strong models less skewed than its components [52]. To fit several independent models
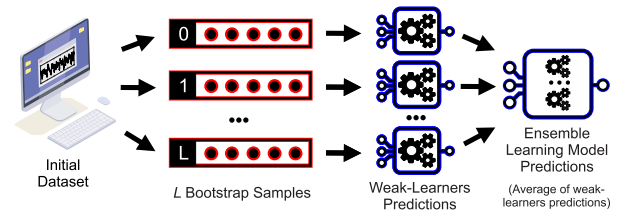

FIGURE 4. Bagging ensemble learning model.

and calculate the average of their predictions to obtain less variance would be necessary to have a very large dataset, then the bagging model is considered to approximate properties of bootstrap samples to fit the model, which are almost independent [53].

Initially, examples are created for each new bootstrap that acts as other approximately independent datasets taken from the true distribution. Thus, each weak learners is adjusted for the samples and these are aggregated obtaining an average of their results [54]. In Figure 4 the bagging ensemble learning model is presented, it is possible to see that the strategy consists of adjusting several base models in different bootstrap samples to build a model that is the average of these results.

The meta-learner is the combination of the weak learners, which can be realized in several ways besides bagging ensemble learning. From an ensemble learning framework, the use of SVR for regression the meta-learner is given by:

$$S(\cdot) = \frac{1}{L} \sum_{l=1}^{L} w_l(\cdot) \qquad (16)$$

here the final meta-learner function is $S(\cdot)$ and each weak-learner regression is given by $w_l(\cdot)$, where L is the number of weak learners [49].

Some strategies like boosting and stacking work similarly to aggregate the weak learners and obtain a model with better performance [55]. The boosting model consists of sequentially adjusting several weak learners in an adaptive way, so more importance is given to observations that were poorly handled by previous models in the sequence [56].

The boosting ensemble learning presented in Figure 5 focuses its efforts on observations that are more difficult to fit, therefore, the resultant has less bias. Weak models tend to have low variance and high bias [49]. Based on this characteristic the models have little degrees of freedom when parameterized. As the fit of models cannot be done in parallel, it can be computationally expensive to sequentially fit complex models [57].

To perform a comparison between the ensemble learning models, SVRs weak learners were used. Thus, it is possible to perform a comparison between the ensemble structures to assess which model is more suitable for the problem in question. The SVR determines support vectors close to a hyperplane that maximize the margin between the two-point classes obtained in relation to the difference between a target value and a threshold value [51].
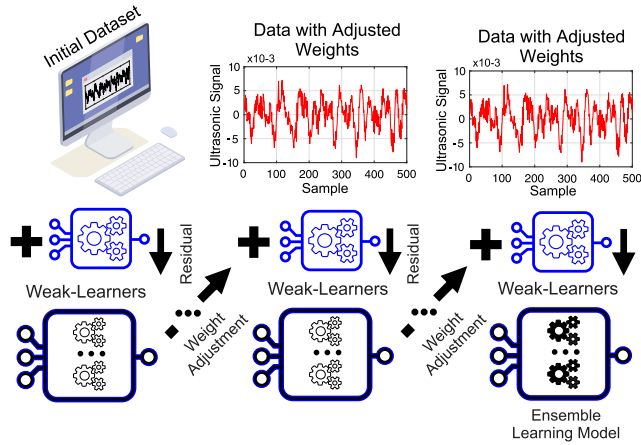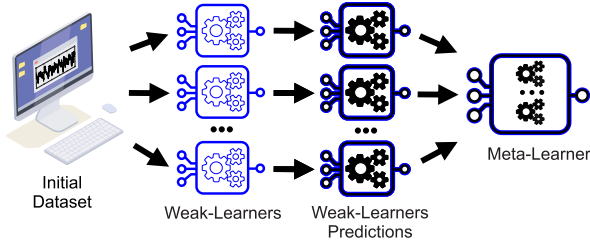
FIGURE 5. Boosting ensemble learning model.

FIGURE 6. Stacking ensemble learning model.

FIGURE 7. Random subspace ensemble learning model.

optimization process, according to:

$$s_l(\cdot) = s_{l-1}(\cdot) + c_l w_l(\cdot) \qquad (17)$$

where the weights c_l and w_l are chosen so that s_l is the model that fits better the training data (improving s_{l-1}). The process is done until convergence, when $S(\cdot) = s_l(\cdot)$.

The random subspace method is similar to bagging, except those resources are randomly sampled for each learner. This causes individual learners to not focus on features that appear to be predictive or descriptive in the training set [69]. Thus, in this model, random subspaces are a promising choice for large problems where there are more resources than training dataset [70].

The random subspace ensemble learning model, shown in Figure 7, was developed to deal with high-dimensional problems. This approach combines weak learners trained in random subspaces in an iterative process, which results in a suitable approach to problems with a large number of resources [71].

For comparison purposes, the LSTM [72] and the nonlinear autoregressive (NAR) [73] models will be used to compare to the ensemble approach. The LSTM is used as a deep learning strategy and the NAR is used as a classical approach to perform a complete comparison.

A. LONG SHORT-TERM MEMORY

The LSTM model stands out for applications in time series, considering its ability to deal with non-linear variations of the system [74]. The LSTM can be calculated through the equations:

$$
\begin{aligned}
i_t &= \sigma_g(W_i x_t + R_i h_{t-1} + b_i), \\
f_t &= \sigma_g(W_f x_t + R_f h_{t-1} + b_f), \\
o_t &= \sigma_g(W_o x_t + R_o h_{t-1} + b_o).
\end{aligned}
\qquad (18)
$$

where R and W are earning matrices and b is the polarization matrix, whose values are assigned by the net training. The σ_g is the activation function of gate. To achieve the predicted values of future time steps, the responses of the training sequences are shifted by one time step. Thus, for each input time step, the network learns to forecast the value of the next time step [75].

For global analysis, two optimizers will be used, these being the stochastic gradient descent with momentum (SGDM) [76] and the adaptive moment estimation (ADAM) [75]. The SGDM is a classic optimizer that has been used by several researchers due to its simplicity and

As with other strategies, the stacking ensemble performs a combination of weak models to result in a model with greater processing capacity (meta-model). For a classification task, beyond the SVR the weak learners can be, for instance, support vector machine (SVM) [58]–[60], k-nearest neighbors (k-NN) [59], [61], [62], or decision trees [63]–[65]. The artificial neural network will take as inputs the results of the weak learners and it will return the final predictions based on these [66]. The structure of the stacking ensemble model is shown in Figure 6.

The first step in this algorithm is to adjust the data referring to the input of the network and the second step adjust the meta-model using predictions made by previous weak learners [67]. The data division is performed in such a way that the training of weak learners is not relevant for the training of the meta-model, only the combination of its results has an influence on the training process [68]. To obtain better reliability in the model, cross-validation can be applied to separate the dataset. For the stacking approach the combination of weak learners is presented in Figure 6.

To prevent overfitting a regularization parameter is used. In the boosting structure, new models are iteratively trained, focusing on observations that previous models had greater difficulty in predicting, making this structure predictive. As the goal is to reduce the bias of the simplest predictors, it is suitable to use a simpler model with high bias and low variance [49].

To generate the meta learner in the boosting approach, weak learners are added one by one in an iterative
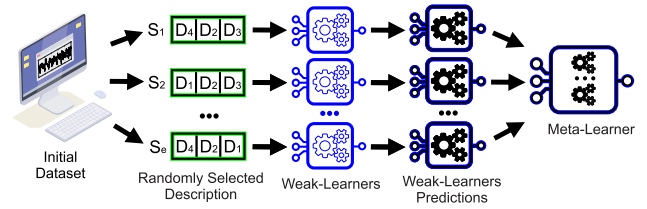
satisfactory result, ADAM is a modern optimizer that is standing out for artificial intelligence applications [77].

In addition to the analysis of the optimizers, the use of deeper layers is evaluated, being a strategy that is currently being widely researched given the popularization of the deep learning approach. For comparison purposes the hyperparameters were set: 200 hidden units, initial learn rate of 0.005, gradient threshold of 1, learn rate drop period of 125, and learn rate drop factor of 0.2. The model used in this paper is a standard structure sequence-to-sequence regression LSTM network, available in MathWorks of Matlab.[1]

B. NONLINEAR AUTOREGRESSIVE

To perform a comparative analysis with shallow learning structures [78], the NAR was used to predict the time series, given by:

$$f(x) = f(y(t-1), \dots, y(t-d),) + \epsilon_t, \quad (19)$$

where $y(t)$ is the predicted output of the model given d past values and another series of $x(t)$ [79].

Like the compared models, NAR is applied for multi-step prediction of a sequence of values in a time series.[2] In this approach, when external feedback is missing, the closed-loop can continue to predict using internal feedback.

To evaluate the configuration parameters, the Levenberg-Marquardt (LM) [80] and bayesian regularization (BR) [81] optimizers were applied. All evaluated models were compared using the same settings, following the performance measures presented in the next subsection.

C. PERFORMANCE MEASURES

The signal error is calculated by the difference in the observed value y_i to the predicted output $\hat{y}_i$ [82]. Thus, for an overall evaluation in relation to the forecast error, mean-square error (MSE), root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were used, which are calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (21)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (22)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (23)$$

where n is the length of the original signal [83]. In this paper, simulations were performed using an Intel Core I5-7400,

20 GB of Random-Access Memory (RAM), with Matlab software.

V. ANALYSIS OF RESULTS

In this section, the evaluation of the parameters change in the considered models will be presented and discussed. The learner kernel base change in the SVR and solver optimizer will be evaluated. The configuration that obtains the best results will be used to compare the differences between each model. The best results for each metric for the used model will be highlighted in bold. In Table 1 is presented the evaluation of ensemble learning models.

The best configuration for the bagging model was obtained using the SMO solver from the LIN kernel as learner base, considering that using this configuration there was a significant improvement in MSE and RMSE, and the time to convergence was among the best results. Using SMO, the best MAPE was obtained with the RBF kernel, however, there was a wide variation of results using this metric, which does not make its analysis reliable. The best convergence time was also obtained with the RBF kernel, however, using the ISDA solver.

In this initial analysis, it was noticed that the metrics do not necessarily indicate the same solver and configuration as the best choice. This requires a deeper analysis of the results to determine which metric is more interesting to be evaluated for more consistent analysis.

The boosting model had similar results to the bagging model. The best MSE and RMSE were obtained using LIN kernel from the SMO solver, being among the best speed results for convergence. In this analysis, the MAPE values were more approximate, with the exception of the RBF kernel using the L1QP, which resulted in a MAPE considerably higher than the other configurations. This result is not promising, as there is no stability regarding this metric.

In the evaluation of the random subspace ensemble model, there was also a great variation of MAPE, and the best result was obtained using the SMO solver with the POLY kernel function. The best result of the speed for convergence of the algorithm were also obtained using the SMO solver. Considering that the best MSE and RMSE were obtained using L1QP with kernel function LIN with a median convergence time, this setting was considered for the final evaluation.

In the bagging plus random subspace model, the best result considering the MSE and RMSE was obtained using the SMO solver with the were also obtained using the SMO solver, however with the RBF kernel function. In this evaluation, there was less MAPE variance in relation to the configuration change of the analyzed model, and the best result obtained in this metric was using the ISDA solver with the LIN kernel function.

The best values for the MSE and RMSE of the stacking ensemble learning model were obtained using the L1QP solver with the RBF kernel function. The best MAPE value was obtained using the ISDA solver and in this case, there

---

[1] https://www.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-learning.html

[2] https://www.mathworks.com/help/deeplearning/ug/multistep-neural-network-prediction.html

TABLE 1. Results of ensemble learning model.

Model	Solver	Kernel Funct.	MSE	RMSE	MAE	MAPE	Time (s)
Bagging	L1QP	LIN	1.06×10^{-6}	1.03×10^{-3}	5.88×10^{-5}	2.65	13.56
		RBF	1.29×10^{-6}	1.14×10^{-3}	6.68×10^{-5}	11.62	7.49
		POLY	1.39×10^{-6}	1.18×10^{-3}	$\mathbf{9.35 \times 10^{-6}}$	4.13	6.93
	ISDA	LIN	1.06×10^{-6}	1.03×10^{-3}	5.77×10^{-5}	1.27	5.00
		RBF	1.29×10^{-6}	1.14×10^{-3}	6.35×10^{-5}	6.03	$\mathbf{1.92}$
		POLY	1.38×10^{-6}	1.18×10^{-3}	1.03×10^{-5}	3.73	120.12
	SMO	LIN	$\mathbf{8.97 \times 10^{-7}}$	$\mathbf{9.47 \times 10^{-4}}$	5.67×10^{-5}	36.71	2.36
		RBF	1.31×10^{-6}	1.14×10^{-3}	6.01×10^{-5}	$\mathbf{1.19}$	2.02
		POLY	1.23×10^{-6}	1.11×10^{-3}	9.73×10^{-6}	5.48	62.57
Boosting	L1QP	LIN	1.13×10^{-6}	1.06×10^{-3}	$\mathbf{3.11 \times 10^{-6}}$	0.46	17.44
		RBF	1.57×10^{-6}	1.25×10^{-3}	4.19×10^{-5}	37.00	15.70
		POLY	2.45×10^{-6}	1.57×10^{-3}	2.15×10^{-4}	4.39	17.51
	ISDA	LIN	1.23×10^{-6}	1.11×10^{-3}	2.86×10^{-5}	1.48	$\mathbf{2.82}$
		RBF	1.42×10^{-6}	1.19×10^{-3}	4.16×10^{-5}	0.77	2.98
		POLY	1.76×10^{-6}	1.33×10^{-3}	1.23×10^{-4}	$\mathbf{0.39}$	249.62
	SMO	LIN	$\mathbf{1.11 \times 10^{-6}}$	$\mathbf{1.06 \times 10^{-3}}$	1.55×10^{-5}	1.33	3.03
		RBF	1.39×10^{-6}	1.18×10^{-3}	1.24×10^{-4}	1.83	2.84
		POLY	1.85×10^{-6}	1.36×10^{-3}	2.37×10^{-4}	0.65	131.05
Random subspace	L1QP	LIN	$\mathbf{1.24 \times 10^{-6}}$	$\mathbf{1.11 \times 10^{-3}}$	7.70×10^{-5}	23.43	15.51
		RBF	1.32×10^{-6}	1.15×10^{-3}	7.97×10^{-5}	2.61	16.84
		POLY	1.35×10^{-6}	1.16×10^{-3}	3.80×10^{-5}	3.61	16.98
	ISDA	LIN	1.37×10^{-6}	1.17×10^{-3}	1.64×10^{-4}	5.43	3.14
		RBF	1.58×10^{-6}	1.26×10^{-3}	1.67×10^{-4}	64.41	2.79
		POLY	1.32×10^{-6}	1.15×10^{-3}	1.93×10^{-4}	21.82	98.18
	SMO	LIN	1.30×10^{-6}	1.14×10^{-3}	$\mathbf{2.17 \times 10^{-5}}$	34.45	3.51
		RBF	1.48×10^{-6}	1.22×10^{-3}	1.05×10^{-4}	0.96	$\mathbf{2.39}$
		POLY	1.26×10^{-6}	1.12×10^{-3}	2.27×10^{-5}	$\mathbf{0.27}$	52.63
Bagging plus random subspace	L1QP	LIN	1.21×10^{-6}	1.10×10^{-3}	8.27×10^{-5}	3.79	7.89
		RBF	1.35×10^{-6}	1.16×10^{-3}	1.69×10^{-4}	7.04	8.05
		POLY	1.45×10^{-6}	1.21×10^{-3}	1.87×10^{-4}	2.37	7.56
	ISDA	LIN	1.27×10^{-6}	1.13×10^{-3}	5.49×10^{-6}	$\mathbf{0.42}$	2.52
		RBF	1.40×10^{-6}	1.19×10^{-3}	9.84×10^{-5}	0.88	2.89
		POLY	1.29×10^{-6}	1.14×10^{-3}	1.85×10^{-5}	18.21	26.90
	SMO	LIN	1.32×10^{-6}	1.15×10^{-3}	1.27×10^{-4}	3.43	2.44
		RBF	1.40×10^{-6}	1.18×10^{-3}	$\mathbf{3.29 \times 10^{-6}}$	8.94	$\mathbf{1.87}$
		POLY	$\mathbf{1.13 \times 10^{-6}}$	$\mathbf{1.06 \times 10^{-3}}$	4.71×10^{-5}	4.83	11.35
Stacking	L1QP	LIN	1.88×10^{-6}	1.37×10^{-3}	3.94×10^{-5}	3.16	4.43
		RBF	$\mathbf{1.40 \times 10^{-6}}$	$\mathbf{1.18 \times 10^{-3}}$	1.14×10^{-4}	4.80	3.46
		POLY	1.71×10^{-6}	1.31×10^{-3}	7.61×10^{-5}	4.35	3.52
	ISDA	LIN	1.83×10^{-6}	1.35×10^{-3}	9.34×10^{-5}	$\mathbf{0.18}$	1.51
		RBF	1.61×10^{-6}	1.27×10^{-3}	6.42×10^{-5}	0.93	1.58
		POLY	1.97×10^{-6}	1.40×10^{-3}	5.50×10^{-5}	0.45	15.83
	SMO	LIN	1.65×10^{-6}	1.29×10^{-3}	9.92×10^{-5}	2.02	$\mathbf{1.49}$
		RBF	1.79×10^{-6}	1.34×10^{-3}	$\mathbf{2.50 \times 10^{-5}}$	0.53	1.50
		POLY	1.97×10^{-6}	1.40×10^{-3}	4.27×10^{-5}	1.08	8.48

was not such an expressive variation as in the other models. The best conference time was obtained using the SMO solver with the LIN kernel function. The subsection V-A provides a comparison between all models in order to obtain an overall performance comparison.

A. OVERALL COMPARISON OF THE ENSEMBLE METHODS
Using the best optimizer configuration and kernel function, all simulations were performed again to obtain a comparison between the best results of each model, this evaluation is presented in Table 2.

As can be seen, bagging and boosting models have similar results in all metrics evaluated in this work. They can be considered the best models for the application in question, since they have the lowest values of RMSE, MSE, MAPE, and time to convergence. Comparing all the results, the best

TABLE 2. Results of all compared ensemble learning models.

Ensemble Model	MSE	RMSE	MAPE	Time (s)
Bagging	$\mathbf{1.12 \times 10^{-6}}$	$\mathbf{1.06 \times 10^{-3}}$	3.74	$\mathbf{0.92}$
Boosting	1.13×10^{-6}	$\mathbf{1.06 \times 10^{-3}}$	$\mathbf{2.76}$	1.48
Random Sub.	1.26×10^{-6}	1.12×10^{-3}	49.66	12.61
B+RS	1.28×10^{-6}	1.13×10^{-3}	11.62	11.53
Stacking	1.63×10^{-6}	1.28×10^{-3}	2.85	1.06

model for this application was the bagging invest which had the best MSE, being a faster model for convergence in this evaluation. The forecast result using the bagging model is shown in Figure 8.

B. LONG SHORT-TERM MEMORY RESULTS
For a comparison using another approach in Table 3 the results of the LSTM are presented, which is an algorithm that
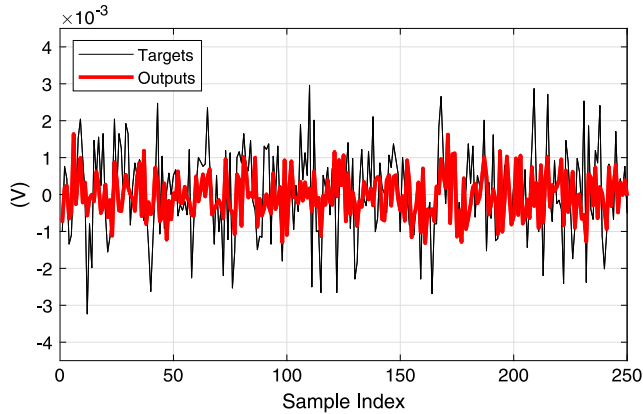
FIGURE 8. Results of the predicted signal based on the observed data.

TABLE 3. Results of long short-term memory model.

Optimizer	DL	MSE	RMSE	MAPE	Time (s)
SGDM	1	**1.30**×10^{-6}	**1.14**×10^{-3}	2.44	**30.93**
	2	1.31×10^{-6}	1.14×10^{-3}	1.53	51.64
	3	1.31×10^{-6}	1.14×10^{-3}	1.64	76.01
	4	1.33×10^{-6}	1.15×10^{-3}	**0.87**	98.44
ADAM	1	1.38×10^{-6}	1.17×10^{-3}	**0.33**	**30.99**
	2	1.34×10^{-6}	1.16×10^{-3}	1.48	55.24
	3	**1.32**×10^{-6}	**1.15**×10^{-3}	6.65	75.73
	4	1.33×10^{-6}	1.15×10^{-3}	3.81	105.51

has stood out for time series prediction due to its ability to deal with non-linear data, being an approach widely applied in deep learning. In this evaluation, the SGDM and ADAM optimizers were used and the use of deeper layers (DL) in the LSTM structure were evaluated.

The first observation regarding the results is that the LSTM needs a longer time for convergence considering that it uses more computational effort. This observation becomes even clearer when deeper layers are used. The best result obtained in this evaluation was using the SGDM optimizer, being the best MSE, RMSE, and time to convergence. From this optimizer, the inclusion of deeper layers did not result in a reduction of the error and resulted in a considerable increase in the time needed for convergence.

Using the ADAM optimizer, the best result was obtained with 3 deeper layers. However, this result was lower than the values obtained using the SGDM optimizer and the time needed for convergence was considerably higher with this number of deeper layers. Comparing these results with the bagging ensemble model, the biggest noticeable difference is in relation to the time to convergence, as this model was much faster. Another difference is the error, given by lower MSE and RMSE, making it clear that this model is more appropriate to be used in the problem in question.

C. NONLINEAR AUTOREGRESSIVE RESULTS

For a comparative analysis, using a nonlinear autoregressive model for time series forecasting, Table 4 presents the results of the use of different optimizers and variations in the number of hidden neurons (HN).

TABLE 4. Results of nonlinear autoregressive model.

Optimizer	HN	MSE	RMSE	MAPE	Time (s)
LM	1	1.97×10^{-3}	4.43×10^{-2}	2.64	48.96
	2	1.98×10^{-3}	4.45×10^{-2}	0.65	**27.40**
	3	1.97×10^{-3}	4.44×10^{-2}	1.44	32.51
	4	**1.95**×10^{-3}	**4.42**×10^{-2}	**0.62**	27.51
BR	1	**1.95**×10^{-3}	**4.42**×10^{-2}	0.65	57.24
	2	1.97×10^{-3}	4.44×10^{-2}	**0.61**	**28.18**
	3	1.97×10^{-3}	4.43×10^{-2}	2.65	31.41
	4	1.96×10^{-3}	4.43×10^{-2}	1.65	61.46

Although the NAR model is usually fast in the training process, for this dataset the improvement in the training time was not so significant compared to LSTM. The training time in all compared ensemble models was faster. Furthermore, all error results calculated by the MSE and RMSE were superior, the NAR model did not show promise for this evaluation. For this approach, the variation of parameters did not result in expressive variations in the results.

VI. CONCLUSION

Improved fault detection in electrical system inspections can help electrical utility companies increase reliability in the electrical power system. Ultrasound is promising equipment for fault identification, considering that indirect detection is directional equipment. From a failure prediction model based on an analysis of equipment that are on the threshold of breakdown, it will be possible to identify failures before they occur and perform maintenance. From predictive maintenance, it is possible to reduce the maintenance costs of the network and mainly to reduce the need for corrective maintenance, which in some situations can leave the system off and reduce the reliability in the power supply by the electric power utility.

Based on the results obtained from the ensemble learning models analysis, it was found that this type of structure was able to handle highly non-linear data to perform the prediction. Among the compared models the bagging and boosting resulted in lower error values considering MSE and shorter time to convergence than the other models, with a MSE of 1.12×10^{-6}, and 1.13×10^{-6}, and a convergence time of 0.92 and 1.48 seconds respectively.

It was also observed that the random subspace and bagging plus random subspace models obtained similar results, regarding the MSE of approximately 1.3×10^{-6}, which makes it clear that depending on the fact that there may be great variation in the results, it is necessary to evaluate different structures to obtain promising results for forecasting time series.

Comparatively, ensemble models are faster to converge than LSTM in all variations of the approaches, considering the use of the best configuration for each one of them, since the LSTM with a deep layer using the SGDM optimizer had a time to convergence of 30.93 seconds, being the fastest configuration in this category, while the bagging, boosting and stacking ensemble models need less than 2 seconds

to converge. Another point to be noted is that despite the greater computational effort of the LSTM, the error values considering the MSE and RMSE were higher than the ensemble models, except the stacking model, which resulted in a greater error than the LSTM. The best result using the LSTM was a RMSE of 1.14×10^{-3} while the bagging and boosting models had a RMSE of 1.06×10^{-3}.

Considering the promising results presented in this work, it becomes feasible to use filters to further improve its processing capacity as high-frequency noise can increase the forecast error. The wavelet transform, among other filters, can reduce signal noise, and then improve the predictive power of the model. With this, there is room to use hybrid models in future works, thus improving signal prediction.

REFERENCES

[1] M. P. Corso, S. F. Stefenon, V. F. Couto, S. H. L. Cabral, and A. Nied, "Evaluation of methods for electric field calculation in transmission lines," *IEEE Latin Amer. Trans.*, vol. 16, no. 12, pp. 2970–2976, Dec. 2018.

[2] A. Medeiros, A. Sartori, S. F. Stefenon, L. H. Meyer, and A. Nied, "Comparison of artificial intelligence techniques to failure prediction in contaminated insulators based on leakage current," *J. Intell. Fuzzy Syst.*, vol. 42, no. 4, pp. 3285–3298, 2022.

[3] L. S. Maraaba, K. Y. A. Soufi, L. M. Alhems, and M. A. Hassan, "Performance evaluation of 230 kV polymer insulators in the coastal area of Saudi Arabia," *IEEE Access*, vol. 8, pp. 164292–164303, 2020.

[4] A. A. Salem, R. Abd-Rahman, W. Rahiman, S. A. Al-Gailani, S. M. Al-Ameri, M. T. Ishak, and U. U. Sheikh, "Pollution flashover under different contamination profiles on high voltage insulator: Numerical and experiment investigation," *IEEE Access*, vol. 9, pp. 37800–37812, 2021.

[5] S. F. Stefenon, J. P. Americo, L. H. Meyer, R. B. Grebogi, and A. Nied, "Analysis of the electric field in porcelain pin-type insulators via finite elements software," *IEEE Latin Amer. Trans.*, vol. 16, no. 10, pp. 2505–2512, Oct. 2018.

[6] Y. Lv, W. Zhao, and Q. Song, "Simulation on deposition characteristics of contamination particles in fog-haze environment," *IEEE Access*, vol. 8, pp. 26119–26128, 2020.

[7] S. F. Stefenon, L. O. Seman, C. S. F. Neto, A. Nied, D. M. Seganfredo, F. G. D. Luz, P. H. Sabino, J. T. González, and V. R. Q. Leithardt, "Electric field evaluation using the finite element method and proxy models for the design of stator slots in a permanent magnet synchronous motor," *Electronics*, vol. 9, no. 11, p. 1975, 2020.

[8] F. Yin, P. Liu, H. Mei, L. Wang, L. Li, and M. Farzaneh, "The shed hole influence on the electrical performance of composite insulator," *IEEE Access*, vol. 8, pp. 217447–217455, 2020.

[9] Y. Hao, Y. Fu, J. Wei, L. Yang, G. Mao, Z. Yang, and L. Li, "Internal temperature detections of contaminated silicone rubber under discharge conditions based on fiber Bragg gratings," *IEEE Access*, vol. 7, pp. 161167–161174, 2019.

[10] B. B. Ustundag and A. Kulaglic, "High-performance time series prediction with predictive error compensated wavelet neural networks," *IEEE Access*, vol. 8, pp. 210532–210541, 2020.

[11] F. Liu, Y. Lu, and M. Cai, "A hybrid method with adaptive sub-series clustering and attention-based stacked residual LSTMs for multivariate time series forecasting," *IEEE Access*, vol. 8, pp. 62423–62438, 2020.

[12] S. Lu and S. Huang, "Segmentation of multivariate industrial time series data based on dynamic latent variable predictability," *IEEE Access*, vol. 8, pp. 112092–112103, 2020.

[13] J. Jia and W. Qiu, "Research on an ensemble classification algorithm based on differential privacy," *IEEE Access*, vol. 8, pp. 93499–93513, 2020.

[14] Y. Wang, X. Wang, Y. Wu, and Y. Guo, "Power system fault classification and prediction based on a three-layer data mining structure," *IEEE Access*, vol. 8, pp. 200897–200914, 2020.

[15] S. F. Stefenon, M. H. D. M. Ribeiro, A. Nied, V. C. Mariani, L. D. S. Coelho, D. F. M. da Rocha, R. B. Grebogi, and A. E. D. B. Ruano, "Wavelet group method of data handling for fault prediction in electrical power insulators," *Int. J. Electr. Power Energy Syst.*, vol. 123, Dec. 2020, Art. no. 106269.

[16] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2018.

[17] S. F. Stefenon, M. P. Corso, A. Nied, F. L. Perez, K.-C. Yow, G. V. Gonzalez, and V. R. Q. Leithardt, "Classification of insulators using neural network based on computer vision," *IET Gener., Transmiss. Distrib.*, vol. 2021, pp. 1–12, Dec. 2021.

[18] J. Zhu, Y. Guo, F. Yue, H. Yuan, A. Yang, X. Wang, and M. Rong, "A deep learning method to detect foreign objects for inspecting power transmission lines," *IEEE Access*, vol. 8, pp. 94065–94075, 2020.

[19] S. F. Stefenon, M. C. Silva, D. W. Bertol, L. H. Meyer, and A. Nied, "Fault diagnosis of insulators from ultrasonic detection using neural networks," *J. Intell. Fuzzy Syst.*, vol. 37, no. 5, pp. 6655–6664, Nov. 2019.

[20] M. Ramesh, L. Cui, Y. Shtessel, and R. Gorur, "Failure studies of polymeric insulating materials using sliding mode observer," *Int. J. Electr. Power Energy Syst.*, vol. 126, Mar. 2021, Art. no. 106539.

[21] A. Ibrahim, A. Dalbah, A. Abualsaud, U. Tariq, and A. El-Hag, "Application of machine learning to evaluate insulator surface erosion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 314–316, Feb. 2020.

[22] S. Polisetty, A. El-Hag, and S. Jayram, "Classification of common discharges in outdoor insulation using acoustic signals and artificial neural network," *High Voltage*, vol. 4, no. 4, pp. 333–338, Dec. 2019.

[23] A. Galicia, R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martínez-Álvarez, "Multi-step forecasting for big data time series based on ensemble learning," *Knowl.-Based Syst.*, vol. 163, pp. 830–841, Jan. 2019.

[24] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers Comput. Sci.*, vol. 14, no. 2, pp. 241–258, Apr. 2020.

[25] S. F. Stefenon, L. O. Seman, B. A. Pavan, R. G. Ovejero, and V. R. Q. Leithardt, "Optimal design of electrical power distribution grid spacers using finite element method," *IET Gener., Transmiss. Distrib.*, pp. 1–12, Feb. 2022.

[26] K. Sit, A. K. Das, D. Mukherjee, N. Haque, S. Deb, A. K. Pradhan, S. Dalai, and B. Chatterjee, "Condition monitoring of overhead polymeric insulators employing hyperbolic window stockwell transform of surface leakage current signals," *IEEE Sensors J.*, vol. 21, no. 9, pp. 10957–10964, May 2021.

[27] S. F. Stefenon, C. S. F. Neto, T. S. Coelho, A. Nied, C. K. Yamaguchi, and K.-C. Yow, "Particle swarm optimization for design of insulators of distribution power system based on finite element method," *Electr. Eng.*, vol. 104, pp. 615–622, 2022.

[28] M. P. Corso, F. L. Perez, S. F. Stefenon, K.-C. Yow, R. G. Ovejero, and V. R. Q. Leithardt, "Classification of contaminated insulators using K-nearest neighbors based on computer vision," *Computers*, vol. 10, no. 9, p. 112, Sep. 2021.

[29] L. Yang, J. Fan, Y. Liu, E. Li, J. Peng, and Z. Liang, "A review on state-of-the-art power line inspection techniques," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9350–9365, Dec. 2020.

[30] S. F. Stefenon, R. Z. Freire, L. H. Meyer, M. P. Corso, A. Sartori, A. Nied, A. C. R. Klaar, and K.-C. Yow, "Fault detection in insulators based on ultrasonic signal processing using a hybrid deep learning technique," *IET Sci., Meas. Technol.*, vol. 14, no. 10, pp. 953–961, Dec. 2020.

[31] L. Meyer and R. Pintarelli, "Inclined plane test for erosion of polymeric insulators under AC and DC voltages," *IEEE Latin Amer. Trans.*, vol. 18, no. 8, pp. 1455–1461, Aug. 2020.

[32] S. F. Stefenon, M. H. D. M. Ribeiro, A. Nied, V. C. Mariani, L. D. S. Coelho, V. R. Q. Leithardt, L. A. Silva, and L. O. Seman, "Hybrid wavelet stacking ensemble model for insulators contamination forecasting," *IEEE Access*, vol. 9, pp. 66387–66397, 2021.

[33] M. Albano, R. T. Waters, P. Charalampidis, H. Griffiths, and A. Haddad, "Infrared analysis of dry-band flashover of silicone rubber insulators," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 1, pp. 304–310, Feb. 2016.

[34] *Artificial Pollution Tests on High-Voltage Insulators to be Used on A.C. Systems*, Standard C. IEC 507, 1991, pp. 1–7.

[35] S. F. Stefenon, R. B. Grebogi, R. Z. Freire, A. Nied, and L. H. Meyer, "Optimized ensemble extreme learning machine for classification of electrical insulators conditions," *IEEE Trans. Ind. Electron.*, vol. 67, no. 6, pp. 5170–5178, Jun. 2020.

[36] S. F. Stefenon, L. O. Seman, N. F. S. Neto, L. H. Meyer, A. Nied, and K.-C. Yow, "Echo state network applied for classification of medium voltage insulators," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107336.

[37] G. Maillard, S. Arlot, and M. Lerasle, "Aggregated hold-out," *J. Mach. Learn. Res.*, vol. 22, pp. 1–20, 2021.

[38] Z. Wu, B. Lyu, Q. Zhang, L. Liu, and J. Zhao, "Phase-space joint resolved PD characteristics of defects on insulator surface in GIS," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 27, no. 1, pp. 156–163, Feb. 2020.

[39] N. F. S. Neto, S. F. Stefenon, L. H. Meyer, R. Bruns, A. Nied, L. O. Seman, G. V. Gonzalez, V. R. Q. Leithardt, and K.-C. Yow, "A study of multilayer perceptron networks applied to classification of ceramic insulators using ultrasound," *Appl. Sci.*, vol. 11, no. 4, p. 1592, Feb. 2021.

[40] S. F. Stefenon, N. W. Branco, A. Nied, D. W. Bertol, E. C. Finardi, A. Sartori, L. H. Meyer, and R. B. Grebogi, "Analysis of training techniques of ann for classification of insulators in electrical power systems," *IET Gener. Transmiss. Distrib.*, vol. 14, pp. 1591–1597, Apr. 2020.

[41] J. Yang and F. Wang, "Auto-ensemble: An adaptive learning rate scheduling based deep learning model ensembling," *IEEE Access*, vol. 8, pp. 217499–217509, 2020.

[42] M. W. Ahmad, J. Reynolds, and Y. Rezgui, "Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees," *J. Cleaner Prod.*, vol. 203, pp. 810–821, Dec. 2018.

[43] H. Zhong, J. Wang, H. Jia, Y. Mu, and S. Lv, "Vector field-based support vector regression for building energy consumption prediction," *Appl. Energy*, vol. 242, pp. 403–414, May 2019.

[44] W. Chu, S. S. Keerthi, and C. J. Ong, "A general formulation for support vector machines," in *Proc. 9th Int. Conf. Neural Inf. Process. (ICONIP)*, vol. 5, Nov. 2002, pp. 2522–2526.

[45] M. Li, "Generalized Lagrange multiplier method and KKT conditions with an application to distributed optimization," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 2, pp. 252–256, Feb. 2019.

[46] Z. Zhang, W.-C. Hong, and J. Li, "Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm," *IEEE Access*, vol. 8, pp. 14642–14658, 2020.

[47] M. L. D. Dias and A. R. R. Neto, "Training soft margin support vector machines by simulated annealing: A dual approach," *Expert Syst. Appl.*, vol. 87, pp. 157–169, Nov. 2017.

[48] A. Ayodeji and Y.-K. Liu, "SVR optimization with soft computing algorithms for incipient SGTR diagnosis," *Ann. Nucl. Energy*, vol. 121, pp. 89–100, Nov. 2018.

[49] S. F. Stefenon, M. H. D. M. Ribeiro, A. Nied, K.-C. Yow, V. C. Mariani, L. D. S. Coelho, and L. O. Seman, "Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam," *Electr. Power Syst. Res.*, vol. 202, Jan. 2022, Art. no. 107584.

[50] A. T. Mohamed, "On the prediction accuracy of the mid-term electricity prices using optimized support vector machine," Dalhousie Univ., Tech. Rep., Halifax, NS, Canada, 2016.

[51] M. Ribeiro, S. Stefenon, J. de Lima, A. Nied, V. Mariani, and L. Coelho, "Electricity price forecasting based on self-adaptive decomposition and heterogeneous ensemble learning," *Energies*, vol. 13, no. 19, p. 5190, Oct. 2020.

[52] K. Mehmood, H. T. Ul Hassan, A. Raza, A. Altalbe, and H. Farooq, "Optimal power generation in energy-deficient scenarios using bagging ensembles," *IEEE Access*, vol. 7, pp. 155917–155929, 2019.

[53] M. H. D. M. Ribeiro and L. dos Santos Coelho, "Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105837.

[54] Y. Wang, Y. Guo, X. Zeng, J. Chen, Y. Kong, and S. Sun, "Stator single-line-to-ground fault protection for bus-connected powerformers based on S-transform and bagging ensemble learning," *IEEE Access*, vol. 8, pp. 88322–88332, 2020.

[55] M. H. D. M. Ribeiro, R. G. da Silva, S. R. Moreno, V. C. Mariani, and L. D. S. Coelho, "Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 136, Mar. 2022, Art. no. 107712.

[56] M. H. D. M. Ribeiro, V. C. Mariani, and L. D. S. Coelho, "Multi-step ahead meningitis case forecasting based on decomposition and multi-objective optimization methods," *J. Biomed. Informat.*, vol. 111, Nov. 2020, Art. no. 103575.

[57] A. H. Osman and H. M. A. Aljahdali, "An effective of ensemble boosting learning method for breast cancer virtual screening using neural network model," *IEEE Access*, vol. 8, pp. 39165–39174, 2020.

[58] M. Liu, Z. Cao, J. Zhang, L. Wang, C. Huang, and X. Luo, "Short-term wind speed forecasting based on the jaya-SVM model," *Int. J. Electr. Power Energy Syst.*, vol. 121, Oct. 2020, Art. no. 106056.

[59] W. Ding, X. Li, H. Yang, J. An, and Z. Zhang, "A novel method for damage prediction of stuffed protective structure under hypervelocity impact by stacking multiple models," *IEEE Access*, vol. 8, pp. 130136–130158, 2020.

[60] G. Wang, G. Zhang, K.-S. Choi, and J. Lu, "Deep additive least squares support vector machines for classification with model transfer," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 7, pp. 1527–1540, Jul. 2019.

[61] M. Rahmatian, B. Vahidi, A. J. Ghanizadeh, G. B. Gharehpetian, and H. A. Alehosseini, "Insulation failure detection in transformer winding using cross-correlation technique with ANN and k-NN regression method during impulse test," *Int. J. Electr. Power Energy Syst.*, vol. 53, pp. 209–218, Dec. 2013.

[62] W. Jia, J. Cheng, and H. Hu, "A cluster-stacking-based approach to forecasting seasonal chlorophyll—A concentration in coastal waters," *IEEE Access*, vol. 8, pp. 99934–99947, 2020.

[63] S. R. Samantaray, "Ensemble decision trees for high impedance fault detection in power distribution network," *Int. J. Electr. Power Energy Syst.*, vol. 43, no. 1, pp. 1048–1055, Dec. 2012.

[64] S. Zian, S. A. Kareem, and K. D. Varathan, "An empirical evaluation of stacked ensembles with different meta-learners in imbalanced classification," *IEEE Access*, vol. 9, pp. 87434–87452, 2021.

[65] J. Song, Y. Wang, Z. Fang, L. Peng, and H. Hong, "Potential of ensemble learning to improve tree-based classifiers for landslide susceptibility mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4642–4662, 2020.

[66] R. G. da Silva, M. H. D. M. Ribeiro, S. R. Moreno, V. C. Mariani, and L. D. S. Coelho, "A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting," *Energy*, vol. 216, Feb. 2021, Art. no. 119174.

[67] H. Hou, X. Chen, M. Li, L. Zhu, Y. Huang, and J. Yu, "Prediction of user outage under typhoon disaster based on multi-algorithm stacking integration," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107123.

[68] M. H. D. M. Ribeiro, R. G. Silva, V. C. Mariani, and L. S. Coelho, "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," *Chaos, Solitons Fractals*, vol. 135, Jun. 2020, Art. no. 109853.

[69] I. Rida, S. Al-Maadeed, A. Mahmood, A. Bouridane, and S. Bakshi, "Palmprint identification using an ensemble of sparse representations," *IEEE Access*, vol. 6, pp. 3241–3248, 2018.

[70] A. Mosavi, A. Shirzadi, B. Choubin, F. Taromideh, F. S. Hosseini, M. Borji, H. Shahabi, A. Salvati, and A. A. Dineva, "Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping," *IEEE Access*, vol. 8, pp. 145968–145983, 2020.

[71] Z. Wang, R. Chu, M. Zhang, X. Wang, and S. Luan, "An improved selective ensemble learning method for highway traffic flow state identification," *IEEE Access*, vol. 8, pp. 212623–212634, 2020.

[72] F. Fernandes, S. F. Stefenon, L. O. Seman, A. Nied, F. C. S. Ferreira, M. C. M. Subtil, A. C. R. Klaar, and V. R. Q. Leithardt, "Long short-term memory stacking model to predict the number of cases and deaths caused by COVID-19," *J. Intell. Fuzzy Syst.*, pp. 1–14, Dec. 2021.

[73] A. A. Alsumaiei, "A nonlinear autoregressive modeling approach for forecasting groundwater level fluctuation in urban aquifers," *Water*, vol. 12, no. 3, p. 820, Mar. 2020.

[74] Y. Bai, J. Xie, C. Liu, Y. Tao, B. Zeng, and C. Li, "Regression modeling for enterprise electricity consumption: A comparison of recurrent neural network and its variants," *Int. J. Electr. Power Energy Syst.*, vol. 126, Mar. 2021, Art. no. 106612.

[75] S. F. Stefenon, C. Kasburg, A. Nied, A. C. R. Klaar, F. C. S. Ferreira, and N. W. Branco, "Hybrid deep learning for power generation forecasting in active solar trackers," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 23, pp. 5667–5674, Dec. 2020.

[76] Y. Guan, Y. Fu, L. Chen, G. Liu, and L. Sun, "Belief-rule-base inference method based on gradient descent with momentum," *IEEE Access*, vol. 9, pp. 34487–34499, 2021.

[77] C. Kasburg and S. F. Stefenon, "Deep learning for photovoltaic generation forecast in active solar trackers," *IEEE Latin Amer. Trans.*, vol. 17, no. 12, pp. 2013–2019, Dec. 2019.

[78] T. Ahmad and H. Chen, "Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems," *Sustain. Cities Soc.*, vol. 45, pp. 460–473, Feb. 2019.

[79] P. Varahram, J. Dooley, K. Finnerty, and R. Farrell, "A digital pre-distortion based on nonlinear autoregressive with exogenous inputs," *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 5, pp. 364–366, May 2016.

[80] H. X. Nguyen, H. Q. Cao, T. T. Nguyen, T. N.-C. Tran, H. N. Tran, and J. W. Jeon, "Improving robot precision positioning using a neural network based on Levenberg Marquardt–APSO algorithm," *IEEE Access*, vol. 9, pp. 75415–75425, 2021.

[81] I. Khan, M. A. Z. Raja, M. Shoaib, P. Kumam, H. Alrabaiah, Z. Shah, and S. Islam, "Design of neural network with Levenberg-Marquardt and Bayesian regularization backpropagation for solving pantograph delay differential equations," *IEEE Access*, vol. 8, pp. 137918–137933, 2020.

[82] S. F. Stefenon, C. Kasburg, R. Z. Freire, F. C. S. Ferreira, D. W. Bertol, and A. Nied, "Photovoltaic power forecasting using wavelet neuro-fuzzy for active solar trackers," *J. Intell. Fuzzy Syst.*, vol. 40, no. 1, pp. 1083–1096, Jan. 2021.

[83] W. Xuan, W. Shouxiang, Z. Qianyu, W. Shaomin, and F. Liwei, "A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems," *Int. J. Electr. Power Energy Syst.*, vol. 126, Mar. 2021, Art. no. 106583.
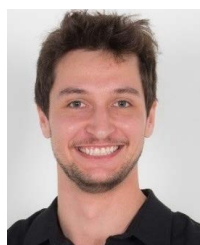
STÉFANO FRIZZO STEFENON received the B.E. and M.E. degrees in electrical engineering (power systems) from the Regional University of Blumenau, Brazil, in 2012 and 2015, respectively, and the Ph.D. degree in electrical engineering (artificial intelligence) from the State University of Santa Catarina, Brazil, in 2021.

During his Ph.D. period, he developed a Research Project in the field of deep learning applied to computer vision with the Faculty of Engineering and Applied Science, University of Regina, Canada. He is currently working with Fondazione Bruno Kessler, Italy, in co-foundation with the University of Udine in the field of computer science and artificial intelligence. His research interests include artificial intelligence for fault identification in electrical power systems, time series forecasting, deep learning, computer vision, and wavelet transform.

RAFAEL BRUNS received the B.E. and M.E. degrees in electrical engineering (power systems) from the Regional University of Blumenau, Brazil, in 2017 and 2020, respectively.

His research interests include artificial neural networks, machine learning, decision tree, electrical insulators, and ultrasound.

ANDREZA SARTORI received the degree in information systems from the University Center of Brusque, in 2007, the master's degree from Åbo Akademi University, Finland, under the Erasmus Program, the master's degree in NetEconomy: technology and information and knowledge management from the University of Trento, Italy, in 2011, and the Ph.D. degree in informatics from the International Doctoral School on Information and Communication Technology, University of Trento, in 2015.

During her doctoral period, she worked at Telecom Italia's research center, SKIL Laboratory (Semantics and Knowledge Innovation Laboratory), associated with EIT ICT Labs. She did part of her doctorate at Boğaziçi University, Turkey. Her doctoral thesis discusses the use of computer vision and machine learning techniques to analyze emotions in abstract paintings. She has been a full-time Professor with the Regional University of Blumenau (FURB), since 2016, where she teaches for the undergradate in computer science and graduate programs in electrical engineering. Her research interests include automatic emotion recognition, computer vision, machine learning, and pattern recognition of industrial systems.

LUIZ HENRIQUE MEYER (Senior Member, IEEE) received the B.E. degree in electrical engineering and the M.Sc. degree in power systems from the Federal University of Santa Catarina, Brazil, in 1990 and 1994, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada, in 2003.

He has been a full-time Professor with the Regional University of Blumenau (FURB), since 1994, where he teaches for the undergraduate and graduate programs in electrical engineering. In 1999, he was the Director of the Technological Research Institute of Regional University of Blumenau, Brazil. He is focused on dielectric materials for outdoor high voltage applications, ceramic and polymeric insulating apparatus, and used in transmission and distribution networks.

RAÚL GARCÍA OVEJERO received the Industrial Technical Engineering degree in electricity, specializing in industrial electronics, in 1997, the master's degree in occupational risks, in 2006, and the Ph.D. degree in industrial engineering from the Department of Mechanics, University of Salamanca (USAL), Spain, in 2014.

He became an Industrial Engineer, in 2001. Currently, he is a contracted Professor Doctor with USAL and a Researcher with the Expert Systems and Applications Laboratory (ESALab). Throughout his training, he has followed a line of research linked to the engineering field, mainly focused on smart fabrics. He has coauthored articles published in several JCR indexed journals.

VALDERI REIS QUIETINHO LEITHARDT (Member, IEEE) received the Ph.D. degree in computer science from INF-UFRGS, Brazil, in 2015.

He is currently an Adjunct Professor with the Polytechnic Institute of Portalegre and a Researcher integrated with the VALORIZA Research Group, School of Technology and Management (ESTG). He is also a Collaborating Researcher with the following research groups, such as COPELABS, Universidade Lusófona de Lisboa, Portugal; and the Expert Systems and Applications Laboratory, University of Salamanca, Spain. His research interests include distributed systems with a focus on data privacy, communication, and programming protocols, involving scenarios and applications for the Internet of Things, smart cities, big data, cloud computing, and blockchain.

• • •