MDPI

Article

Classification of Contaminated Insulators using *k*-Nearest Neighbors based on Computer Vision

Marcelo Picolotto Corso [1], ∗, Fabio Luis Perez [1], Stéfano Frizzo Stefenon [2], Kin-Choong Yow [2], Raúl García Ovejero [3], and Valderi Reis Quietinho Leithardt [4,5]

1 Electrical Engineering Graduate Program, Regional University of Blumenau. Blumenau, Brazil; fabiotek@furb.br (F.L.P.).
2 Faculty of Engineering and Applied Science, University of Regina. Regina, Canada; stefanostefenon@gmail.com (S.F.S.), kin-choong.yow@uregina.ca (K.C.Y.).
3 Expert Systems and Applications Lab., E.T.S.I.I. of Béjar, University of Salamanca, Salamanca, Spain; raulovej@usal.es (R.G.O.).
4 VALORIZA, Research Center for Endogenous Resources Valorization, Instituto Politécnico de Portalegre. Portalegre, Portugal; valderi@ipportalegre.pt (V.R.Q.L.).
5 COPELABS, Universidade Lusófona de Humanidades e Tecnologias. Lisboa, Portugal.
* marcelopcorso@gmail.com (M.P.C.)

Abstract: The contamination on the insulators may increase its surface conductivity and, as a consequence, electrical discharges occur more frequently, which can lead to interruptions in the power supply. To maintain reliability in the electrical distribution power system, components that have lost their insulating properties must be replaced. Identifying the components that need maintenance, is a difficult task as there are several levels of contamination that are hardly noticed during inspections. To improve the quality of inspections, this paper proposes to use the *k*-nearest neighbors (*k*-NN) to classify the levels of insulator contamination, based on the image of insulators at various levels of contamination simulated in the laboratory. Using computer vision features such as mean, variance, asymmetry, kurtosis, energy, and entropy are used for training the *k*-NN. To assess the robustness of the proposed approach, statistical analysis and a comparative assessment with well-consolidated algorithms such as decision tree, ensemble subspace, and support vector machine models are presented. The *k*-NN showed results of up to 85.17 % accuracy using the *k*-fold cross-validation method, with an average accuracy higher than 82 % for multi-classification of the contamination of the insulators, being superior to the compared models.

Keywords: Classification of insulators; electrical power system; *k*-nearest neighbors; computer vision.

Citation: Corso, M. P.; Perez, F. L.; Stefenon, S. F.; Yow, K.-C.; Ovejero R. G. Leithardt, V. R. Q. Classification of Contaminated Insulators using *k*-Nearest Neighbors based on Computer Vision. *Computers* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Electrical power distribution systems are responsible for providing electricity to consumers. Due to population growth and improved access to energy it is necessary to have the reliability of the electricity distribution grid [1]. Saline contamination in coastal regions is a problem that must be monitored, if the insulator has greater surface conductivity it is more likely to develop a failure. The identification of faults in the electrical power system is a difficult task that requires experience from the operator [2].

Researchers have conducted promising studies regarding the evaluation of contamination in insulators installed outdoors using artificial intelligence strategies [3–5]. According to Abouzeid *et al.* [6], the flashover of the insulator occurs when contaminants cover the surface of the insulator resulting in a reduction in the surface resistance. The leakage current can be used as an evaluation parameter to predict the level of contamination, given by the equivalent salt deposit density (ESDD), on the external surface of the insulators and thereby leave the system operators alert to possible failures.

As presented by Soltani and El-Hag [7], an artificial neural network (ANN) in curve fitting can be used to denoising of different types of measured signals emitted from partial discharges sources. The ANN can be applied to distinguish electrical discharges in outdoor insulators caused by surface contamination, corona near the insulator surface, and other common failure types of these components [8]. Defects in insulators give rise to the onset of partial discharges, which has a detrimental effect on the life of the insulator. It is important identify the defective components as early as possible, so appropriate strategies can be implemented [9]. Currently, many fault detection techniques are based on the identification of partial discharges [10], as these may be associated with the presence of faults that are difficult to identify without the use of specific equipment [**?**].

The application of ANN to assess the contamination in insulators can assist in the identification of faults in the electrical network and prevent the shutdown of the electrical power system. The use of adaptive neuro-fuzzy inference system (ANFIS) is a promising strategy that can stand out from classic models. The ANFIS combined with the wavelet transform can improve the accuracy in the evaluation of contaminated insulators installed outdoors next to the rural roads [11].

The use of the wavelet transform combined with other algorithms is a promising alternative for the evaluation of noise signals [12]. The group method of data handling (GMDH) model can be used from data filtered by the wavelet transform and outperform well-established algorithms such as long short term memory (LSTM) and ANFIS [13]. The results of that analysis were supported by comparisons to wavelet LSTM and wavelet ANFIS, being the fastest GMDH and with equivalent accuracy in benchmarks. The LSTM is an architecture used in the field of deep learning, which has gained space in forecasting and classification research [14]. The accuracy of the LSTM is high compared to classic approaches, however, the training process could requires a great computational effort, which increases the training time [15].

Currently, many authors have used computer vision to identify failures in the electrical power system. As example, Nguyen *et al*. [16] developed a research about power line inspections and Manninen *et al*. [17] made an assessment of the network infrastructure based on computer vision. Some authors have done specific assessments on failure conditions in insulators. Shi and Huang [18] present a strategy for detecting faults in insulators using geometric constraints, that can be applied effectively to recognize a damaged insulator. For an accurate classification of the condition it is necessary to have a large number of images recorded during field inspections. The strategy based on a deep learning model could achieves an accuracy of up to 92.86 % to identify insulators that do not have their insulating surface.

The Sampedro *et al*. [19] present a system for diagnosing a set of insulators based on deep learning. The analysis is performed using a convolutional neural network (CNN) that aims to identify the absence of disk insulators that are installed in series. In this paper, the strategy is based on modeling the similarity between adjacent disks, so it is possible to identify various types of defects from the same model. For robustness in the model, it is necessary to extensively evaluate several video sequences of inspections of the high voltage electrical network.

The evaluation of the lack of insulators in transmission lines is a problem that has been intensively studied using computer vision, most studies are based on controlled situations in which lighting conditions and the background of the image are determined. Specific field assessments during inspections of the electrical system are rare due to the scarcity of images of failures. Data augmentation techniques, such as transformation, segmentation, blur, and brightness changes can be a strategy for training the model for field evaluation. According to Tao *et al*. [20] from a CNN model, using a technique to increase the database, the accuracy to fault detection can reach up to 91 % based on a set of standard insulators under various conditions.

Fault detection in insulators can be automated by identifying components with erosion, such as those found in silicone rubber insulators. The use of pre-processing

methods for extracting characteristics can improve the classification of failures. In a laboratory analysis presented by Ibrahim *et al*. [21], was possible to identify insulators with the level of erosion following the IEC-60587 standard from a CNN. The application of deep CNN based on the degree of erosion in silicone rubber showed better results than classical shallow feedforward ANN approaches. The results of the paper show that the proof of concept is valid and can be applied in future tests in an external environment.

An assessment of the impact of contamination caused by rail vehicles on insulators is presented by Kang *et al*. [22], in which the computer vision through CNN was applied for automatic inspection of the insulators, with the objective of improving the safety of the railway operation. The identification of failures in these locations is difficult, due to the low failure rate that exists, which results in few images for the algorithm to be trained. The anomaly classification is determined from a deep multitasking ANN, which has a material classifier and a denoising autoencoder. The results presented using CNN indicate that the proposed can achieve high precision for fault identification along the railway line.

One method that has currently stood out for classification and regression is the *k*-nearest neighbors (*k*-NN) [23]. Despite being a method that has been used for several years, many variations of this algorithm are currently being evaluated to improve its capacity. The papers from Mailagaha Kumbure *et al*. [24] and González *et al*. [25] combine the *k*-NN algorithm with systems based on fuzzy logic to improve its accuracy. As presented by Sharma and Seal [26], the way of calculating the distances of this algorithm is an important parameter to be evaluated, for this reason the distance calculation method will be a parameter evaluated in this paper.

Following in this paper is presented: In Section 2, the problem related to contamination will be presented, followed by the explanation of the analysis in the laboratory for the production of the samples and feature extraction. In Section 3, the *k*-nearest neighbors model is presented. In Section 4, the results obtained are discussed and compared with well-established models. Finally, in Section 5, a conclusion is presented with a general discussion about the application of the used algorithm, and possible future works.

2. Insulator Contamination

Contaminated insulators are a common problem in electrical power distribution and transmission networks. There are several types of contamination, that depend on where the network is installed [27]. Insulators installed near unpaved streets may have accumulation of dust and organic residues, insulators that are close to regions with a high population density may have industrial contamination or automobile combustion residues [28].

Another very common problem is the saline contamination, which is present in coastal regions. These contaminations must be evaluated and monitored by the energy utilities, since the salinity combined with an high relative humidity of the air, increases the surface conductivity of the insulators and can cause failures in the electrical power system [29].

2.1. Contaminated Insulator Samples

The insulators used in this paper are 15 kV pin type made of porcelain, from the manufacturer Germer. These components meet the criteria of IEC 60383 [30] and are widely used in medium voltage distribution lines in southern Brazil. The contamination procedure was performed in the high voltage laboratory at the Regional University of Blumenau, Brazil.

The experimental procedure consists of 3 stages, which are: artificial contamination; photographic capture, and measurement of the level of contamination. To perform the artificial contamination process, the insulators were immersed in a contamination solution based on IEC 60507 [31], which deals specifically with artificial contamination tests on high-voltage ceramic insulators. To create the variations defined by the stan-

dard, different contamination conditions were used, and these were obtained from the variation of the mass of Kaolin added to distilled water.

The photographic capture of the insulators under different levels of contamination was carried out with an LG H-818 sensor that has a focal aperture of 1.8. For standardization of photo registration, the exposure time was set at 3.33 ms, light sensitivity at ISO 150, and focal length of 4 mm. A photograph of a contaminated insulator is presented in Fig. 1.

Figure 1. Contaminated insulator sample.

The place where the photographs were recorded was standardized, to guarantee that the luminous intensity applied to the insulator and the background were the same in all photographic captures. The insulators were fixed on a metal pedestal at a distance of 1 m at the same height that the camera was positioned. For each contaminated sample, 8 captures are made, 2 every 90° around the insulator, thus comprising 4 different viewing directions, the photos on each side were taken using a zoom of 2.5 times and 3.5 times to ensure a greater level of detail.

For the determination of the intensity of the non-soluble deposit density (NSDD) IEC 60815 (Annex C) [32] was used, which is specific for selection and dimensioning of high-voltage insulators under polluted conditions. In this analysis, the residues are extracted from the insulating surface into a container with distilled water and the contents of the container go through a filtering process. The variation in the mass of the filter results in the amount of NSDD on the insulator, given by the equation:

$$NSDD = (W_f - W_i)/A, \tag{1}$$

where W_f is the final mass of the filter after the process, W_i is the mass of the filter before the process and A is the surface area of the insulator, so the value of NSDD is given in g/cm^2.

For the data set, 5 different insulators of the same profile were used in 8 contamination classes. Thus, a set of data was obtained from 40 conditions of analysis. For comparison, 3 insulators without contamination were used, resulting in 4 levels of NSDD for multi-classification. The concentration of Kaolin used was 6, 8, 10, 16, 20, and 25 (g/l). For 2 classes the concentrations of 40 and 60 (S/cm) of salt were used, from these variations 8 classes were obtained. The 4 levels of NSDD were: NSDD equals zero, NSDD greater than 0 and less than 1.0, NSDD from 1.0 to 2.0, NSDD greater than 2.0.

2.2. Image pre-processing

To perform pre-processing of the image, arithmetic transformations are performed, the first conversion is applied to transform the color image to a grayscale image [33]. In this step, each transformation combines different color channels of the image to result in a grayscale image. After converting the image to grayscale, the image segmentation is applied. Segmentation is performed to partition the image in regions that are based on their characteristics of the image pixels [34].

The segmentation active contours region technique was used to segment the image into foreground and background. Starting from the active contour technique, the initial curves of an image are specified, and then the active contour function evolves the curves towards the object's limits. Using this segmentation technique, the mask argument is a binary image that specifies the initial state of the active contour [35]. The limits of the object's regions in the mask define the starting position of the contour used for the evolution of the contour to segment the image.

The resulting image is a binary image in which the foreground is white and the background is black. From this conversion resulting in a black background, it is possible to focus on the classification specifically of the insulator, disregarding image noise. Fig. 2 presents the result of the conversion and segmentation of the insulator presented in Fig. 1.

Figure 2. Conversion of the photograph of the insulator sample after segmentation.

2.3. Feature Extraction

The purpose of the feature extraction is that the image information is segmented to perform the classification. For this purpose, the information is encoded in a feature vector, given by:

$$\vec{x} = [x_1, x_2, x_3, \cdots, x_n]^T, \tag{2}$$

representing the image signature, where n is the total number of attributes considered.

One of the possible ways to features extraction of an image is through the histogram h analysis. To obtain the histogram, normalization is performed to obtain the probability density p of the image. This way,

$$p(i) = h(i)/NM, \tag{3}$$

where i is each index of the histogram and NM is the dimensions of the image. From p, considering $i \in 0, \cdots, L-1$, wherein, L is the total amount of intensity in the image; other measures are obtained, such as mean (4), variance (5), asymmetry (6), kurtosis (7), energy (8), and entropy (10).

All the features presented here were used to perform ANN training and testing:

$$\mu_1 = \sum_{i=0}^{L-1} ip(i). \tag{4}$$

$$\mu_2 = \sigma^2 = \sum_{i=0}^{L-1} (i-\mu_1)^2 p(i). \tag{5}$$

$$\mu_3 = \sigma^{-3} \sum_{i=0}^{L-1} (i-\mu_1)^3 p(i). \tag{6}$$

$$\mu_4 = \sigma^{-4} \sum_{i=0}^{L-1} (i - \mu_1)^4 p(i) - 3. \tag{7}$$

$$En = \sum_{i=0}^{L-1} [p(i)]^2. \tag{8}$$

$$Et = -\sum_{i=0}^{L-1} p(i) \log_2 [p(i)]. \tag{9}$$

To make it possible to calculate the entropy logarithm [36] when the value of $p(i)$ is equal to zero, the value of a constant c is added, in this case it's equal to 0.1, and we have:

$$Et = -\sum_{i=0}^{L-1} p(i) \log_2 [p(i) + c]. \tag{10}$$

3. Nearest Neighbors Method

This section presents the k-nearest neighbors (k-NN) method that was used to classify the insulators evaluated in this paper. The presented method is based on the evaluation of similar data considering the hypothesis of being concentrated in the same region of input space and non-similar data will be distant from each other.

The method of the nearest neighbor has variations defined by the number of neighbors considered, for this reason, the number of neighbors is an evaluation parameter discussed in this paper. In the k-NN model, the k objects of the training set closest to x_t are evaluated [37]. When k is greater than one, the neighboring k is obtained for each test point. Fig. 3 shows the impact of using the value of k on the k-NN model. As can be seen, considering $k = 3$ the test would be classified with a failed (defective) insulator, while for k= 5 the insulator would be classified as a component in good condition.

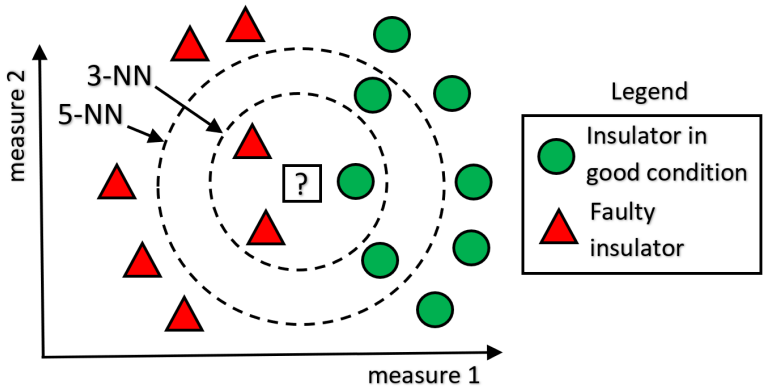

Figure 3. Impact of the k value on the k-NN algorithm.

For this reason, choosing the value of k is not a trivial decision and its variation must be evaluated. As it is a classification problem, it is convenient that the value of k is odd, thus avoiding draws, which makes classification difficult [38]. For the classification problem, the weighted mode is given by:

$$y_t = \arg\max_{c \in Y} \sum_{i=1}^{k} w_i I(c, y_i) \tag{11}$$

wherein,

$$w_i = \frac{1}{d(x_t, x_i)} \tag{12}$$

and $I(a,b)$ is a function that returns 1 if $a = b$ [39]. Since y_i is the class of example x_i, w_i is the weight associated with the example of x_i and c is the class with the most weighted mode.

3.1. Model Architecture

The k-NN is a memory-based algorithm, so all computation is postponed until the classification phase, since the learning process consists of memorizing objects [40]. One of the advantages of k-NN is that the model is simple, thus it requires less computational effort than methods based on deep learning because during training the algorithm only stores objects.

The k-NN constructs the approximation of the objective function, which is different for each new data to be stored. This feature can be advantageous when the objective function is complex and can be described as a collection of less complex local approaches [41]. Based on its characteristics, k-NN can be applied to complex problems, being an incremental algorithm, that is, when new data are available, it is only necessary to store it in memory.

An important aspect to be considered about the k-NN is related to its behavior at the limit if considered:

- e = optimal Bayes classifier error;
- $e_{1nn}(D)$: error of 1-NN;
- $e_{knn}(D)$: error of k-NN.

Thus, the following theorems are proved:

- $lim_{n\to\infty}\, e_{1nn}(D) <= 2 \times e$;
- $lim_{n\to\infty,\, k\to n}\, e_{knn}(D) = e$.

So for an infinite number of objects, the error of 1-NN is increased by twice of the optimal Bayes error, and the error of k-NN tends to the error of the optimal Bayes.

The disadvantages in the k-NN are that the model does not obtain a compact representation of the objects, so there is no explicit model from the training data. To classify an object, k-NN requires calculating the distance from that object to all training objects. Like other algorithms that perform the calculation based on distance, k-NN is affected by the presence of redundant attributes and/or irrelevant [42].

One way to improve the capacity of the model is related to investigating the reduction of the problem space or evaluate how the variation of the distance is calculated. For a complete assessment about the k-NN, in this paper many methods to calculate the distance of the neighbors are evaluated, as described in the next section.

3.2. Neighbor Distance Method

For a complete evaluation of the variation of the result in relation to the function for the calculation of neighbor distance, in this paper several functions were used. From an my-by-n data matrix Y, which is treated as my (1-by-n) row vectors $y1, y2, ..., y_{my}$, and an mx-by-n data matrix X, which is treated as mx (1-by-n) row vectors $x1, x2, ..., x_{mx}$, the distances between the vector x_s and y_t could be defined according to: Euclidean distance (13), cosine distance (14), correlation distance (15a), chebychev distance (16), city block distance (17), spearman distance (18), standardized euclidean distance (19), minkowski distance (20), mahalanobis distance (21).

$$d_{st}^2 = (x_s - y_t)(x_s - y_t)'. \tag{13}$$

$$d_{st} = (1 - \frac{x_s y_t'}{\sqrt{(x_s x_s')(y_t y_t')}}). \tag{14}$$

$$d_{st} = 1 - \frac{(x_s - \bar{x_s})(y_t - \bar{y_t})'}{\sqrt{(x_s - \bar{x_s})(x_s - \bar{x_s})'}\sqrt{(y_t - \bar{y_t})(y_t - \bar{y_t})'}} \tag{15a}$$

wherein,

$$\bar{x_s} = \frac{1}{n}\sum_j x_{sj} \tag{15b}$$

and

$$\bar{y_t} = \frac{1}{n}\sum_j y_{tj}. \tag{15c}$$

$$d_{st} = \max_j\{|x_{sj} - y_{tj}|\}. \tag{16}$$

$$d_{st} = \sum_{j=1}^{n}|x_{sj} - y_{tj}|. \tag{17}$$

$$d_{st} = 1 - \frac{(r_s - \bar{r_s})(r_t - \bar{r_t})'}{\sqrt{(r_s - \bar{r_s})(r_s - \bar{r_s})'}\sqrt{(r_t - \bar{r_t})(r_t - \bar{r_t})'}}, \tag{18}$$

where r_s and r_t are the coordinate-wise rank vectors of x_s and y_t [43].

$$d_{st}^2 = (x_s - y_t)V^{-1}(x_s - y_t)', \tag{19}$$

wherein V is the n-by-n diagonal matrix whose the diagonal element is $S(j)^2$, being S is a vector of scaling factors for each dimension.

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n}|x_{sj} - y_{tj}|^p}. \tag{20}$$

$$d_{st}^2 = (x_s - y_t)C^{-1}(x_s - y_t)', \tag{21}$$

where C is the covariance matrix. Considering that the calculation of the distance from neighbors is of paramount importance for the algorithm, a wide range of functions results in a more complete assessment of the k-NN [44].

3.3. Holdout

In the holdout approach, the division of data is carried out considering a proportion of data ρ for training and a proportion $(1 - \rho)$ for testing. To make the results less dependent on the partition made, random partitions can be used to obtain an average performance using holdout [45].

3.4. Cross-validation

In the process of cross-validation, the data set is divided into subsets k of approximately equal size. The k-1 objects are used in the training of the predictor, which is then tested in the remaining partition. The process is repeated k times, using a different partition in each cycle until all have been used. The predictor's performance are given by the average of the performance observed in each test set [46].

In the classification problem in question, k-fold maintains in each partition the proportion of examples from each class similar to the proportion contained in the total data set [47]. This means that, for example, if the original dataset contains 20 % of objects in class 1 and 80 % in class 2, each partition will also maintain this proportion.

3.5. Benchmarking

After evaluating the best configuration of the model, a benchmarking with the decision tree [48], ensemble [49,50] and support vector machine (SVM) [51–53] models is presented. The pictures of the insulators were taken before the measurement of the NSDD, so that there was no influence from the operator on the contamination, if the contamination did not meet the requirement of IEC 60815 (Annex C) [32] the process would be repeated since the start.

The results presented in this paper were evaluated from Intel Core I5-7400, 20 GB of random-access memory, with the MATLAB software. The application of the proposed method in an embedded system could be a solution for the electrical inspections, improving the reliability of the electrical power network [54–57].

4. Analysis of Results

In this section, the results of the multi-classification of contaminated insulators are presented and discussed. As the used model is supervised, the first analysis to be performed is related to the NSDD result, which is considered the desired output from the network in the classification process.

Table 1 shows the variation of NSDD depending on the Kaolin concentration used in the analysis. It is noticeable that the NSDD value is directly influenced by the Kaolin concentration used in the experiment, the range of variation of the NSDD can be high which makes the classification of the data more difficult, since this distribution is not linear. As discussed in subSection 2.1, the NSDD value is obtained by the difference of W_f and W_i in relation to the area, according to Eq. (1).

Table 1: NSDD variation according to Kaolin concentration.

Kaolin (g/l)	6	8	10	16	20	25
Max. W_f	2.778	3.279	3.483	4.297	4.201	4.485
Min. W_f	2.408	2.542	2.615	2.946	3.580	3.956
Max. W_i	2.226	2.415	2.204	2.522	2.385	2.461
Min. W_i	2.121	2.134	2.088	2.123	2.088	2.145
Max. NSDD	0.868	1.503	1.688	2.618	2.637	2.656
Min. NSDD	0.377	0.393	0.633	1.029	1.568	1.661

In addition to the 6 variations in Kaolin concentration, 2 classes were analyzed with the inclusion of salt (to evaluate ESDD), resulting in 6 concentration variations and 8 evaluated classes. In Fig. 4, NSDD results are presented in mg/cm^2 for the 40 analyses performed, being divided into 8 classes (presented in different colors) with 5 insulators in each class.

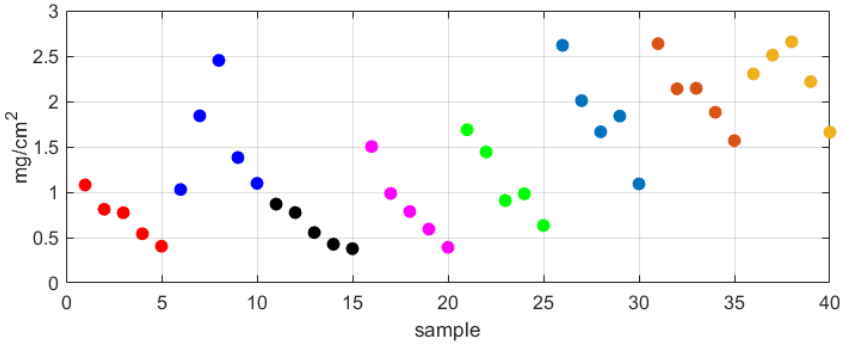

Figure 4. Non-soluble deposit density (NSDD) for the samples analyzed.

As can be seen, there is a grouping of data that are related to the levels of NSDD. This result is expected since for each insulator of the same class the same concentration

 doi:10.20944/preprints202108.0282.v1

of contaminants was used. It was noticed during the initial laboratory evaluation that ESDD does not generate visual variations, that is, the conductivity that occurs due to the concentration of salt in the analysis does not represent a large visual variation and for this reason, the ESDD was not evaluated. The NSDD variation caused by the Kaolin concentration generates great visual differences and for this reason, the NSDD was the focus of the classification analysis.

To perform the multi-classification, the results were evaluated in relation to the NSDD concentration level. Table 2 shows the percentage of insulators that fit each condition evaluated in the multi-classification. In addition, to the 40 samples evaluated, 3 samples without contamination were included for the classification, which were considered to be scale background, so the classification occurs for 4 different conditions.

Table 2: Percentage of insulators by contamination level for the classification.

Kaolin (g/l)	NSDD < 1.0 (mg/cm^2)	$1.0 <$ NSDD < 2.0 (mg/cm^2)	NSDD > 2.0 (mg/cm^2)
6	100 %	0 %	0 %
8	80 %	20 %	0 %
10	60 %	40 %	0 %
16	0 %	40 %	60 %
20	0 %	40 %	60 %
25	0 %	20 %	80 %

After defining the desired output from the network, the classification process begins using the *k*-NN algorithm. The parameters of the algorithm are evaluated dynamically, from the variation of the number of neighbors and variation of the function for the calculation of the distance in *k*-NN. To assess the influence of the data separation method, the first analysis is performed with the data directly separated, through the holdout approach, presented in Fig. 5.

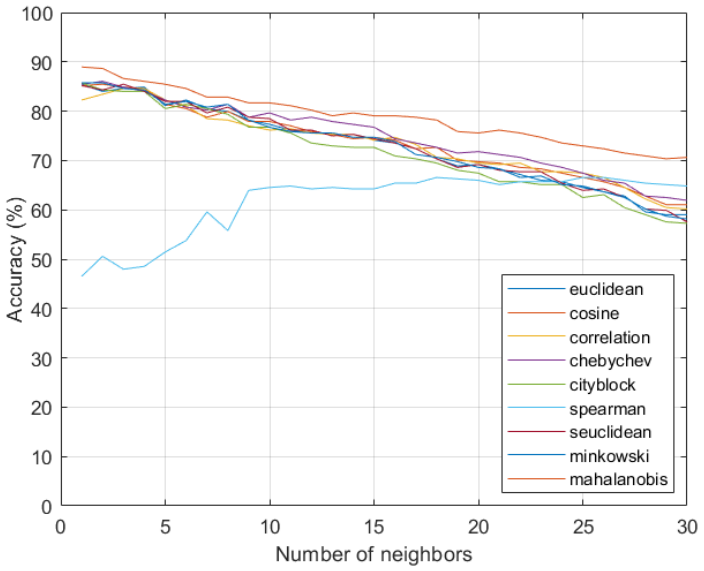

Figure 5. Analysis of the accuracy of the algorithm using holdout.

In this initial assessment, in most ways to distance calculation, the greater the number of neighbors in the *k*-NN, the lower the accuracy of the algorithm. The exception occurs in calculating the distance using the Euclidean equation, however from 10 neighbors, there is no significant improvement in accuracy through this function. The method of calculation that results in better accuracy for the division of data by the holdout approach is the mahalanobis function. In the following analysis, the evaluation

 doi:10.20944/preprints202108.0282.v1

is performed with the validation approach from randomly separated data, these results are shown in the Fig. 6.

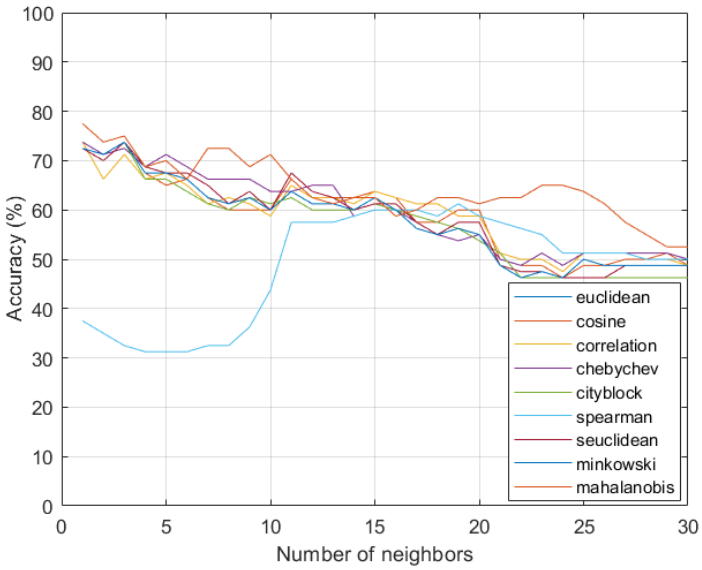

Figure 6. Analysis of the accuracy of the model using holdout with randomly separated data.

Using the holdout approach with the data randomly separated, there is a greater variation in accuracy depending on the variation in the number of neighbors. There is also a greater variation using different functions for calculating the distance in the *k*-NN. Despite the greater variation, the characteristics of better accuracy in the mahalanobis function and greater variation of the spearman function remain, as well as those found without the use of randomness. Fig. 7 presents an analysis of the use of *k*-fold cross-validation method.

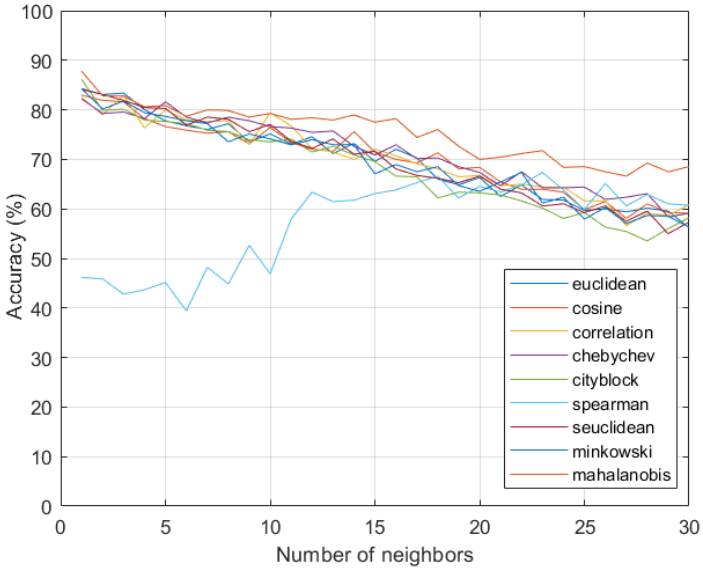

Figure 7. Analysis of the accuracy of the algorithm using cross-validation.

As well as the results applying the holdout approach from a random data set, in the *k*-fold method, the best accuracy result is maintained using the mahalanobis function. From the mahalanobis function there is an improvement in accuracy depending on the way the data set is used in the training process, the best result was obtained using the *k*-fold from 2 neighbors.

Based on this configuration, the variation in the number of folds (k) is evaluated in Table 3 for 3 distance weight calculation methods. The evaluation was performed using 2 to 12 folds, however, the best results were obtained between 5 to 10, for this reason only these values are presented. The best accuracy result was found using 9 folds in the 3 methods of calculating the distance weight. By using this configuration, 100 analyses were performed and the statistical result is shown in Table 4.

Table 3: Evaluation of the variation in the number of folds.

Distance Weight	Accuracy (%)					
	5-fold	6-fold	7-fold	8-fold	9-fold	10-fold
Equal	82.85	81.69	82.85	81.69	82.85	79.94
Inverse	79.36	81.98	82.85	80.52	**84.59**	81.69
Sq. Inver.	81.69	84.30	82.85	83.14	84.58	83.14

Table 4: Statistical evaluation of the k-NN algorithm.

Distance Weight	Accuracy (%)			Std. Dev.
	Max.	Min.	Mean	
Equal	**85.17**	79.65	82.39	8.70×10^{-3}
Inverse	**85.17**	77.91	82.26	1.11×10^{-2}
Sq. Inverse	84.88	79.36	82.23	9.6×10^{-3}

The best accuracy found was 85.17 %, for calculating the distance weight from 100 analyzes. It is possible to realize that this variation is statistically low based on the values of standard deviation; this shows that the algorithm is robust with low variation even with a large number of analyzes.

4.1. Benchmarking

For a comparative evaluation, the same configuration used for the cross-validation of the k-NN was applied in all compared models, the statistical results are presented in Table 5. Models variations are presented for a complete assessment. The decision tree algorithm uses split criterion Gini's diversity index (gdi) and deviance. In the SVM the coding "one vs one" and "all pairs" are applied.

Table 5: Benchmarking with statistical evaluation.

Method	Accuracy (%)			Std. Dev.
	Max.	Min.	Mean	
Decision Tree (gdi)	56.98	47.09	52.92	1.26×10^{-2}
Decision Tree (deviance)	61.63	56.40	60.02	7.43×10^{-3}
Ensemble (subspace)	**67.44**	63.95	65.53	6.42×10^{-3}
SVM (onevsone)	47.67	43.31	45.51	6.34×10^{-3}
SVM (allpairs)	47.38	42.73	45.40	7.84×10^{-3}

As can be seen in Table 5, the decision tree, ensemble (subspace) and SVM algorithms had lower accuracy results than the k-NN for the evaluation presented in this paper. The differences between the maximum and minimum values remained low in all evaluated algorithms, considering 100 simulations with the same configuration using the cross-validation of the data. The best result found in benchmarking was using the Ensemble (subspace), which resulted in an accuracy of 67.44 % in the best case.

5. Conclusion

The use of the k-NN presented in this paper demonstrated that it is possible to classify contamination in insulators as a promising accuracy. Based on this classification, it would be possible to define strategies for maintaining the distribution network

according to the levels of contamination, which are found on the surface of the insulating components. Through predictive maintenance, it will be possible to improve the reliability of the network by reducing power outages.

The analysis showed that cross-validation, a method covered in several articles, is superior to the holdout method for this application. The best accuracy results were found from a lower number of neighbors, this shows that it is necessary to make a complete evaluation of the algorithm to find its best configuration, which in consequence results in higher accuracy. The functions for calculating the distance from neighbors had results that followed the same trend, with the exception of the euclidean function, which had an inverse result in relation to the other functions. The most accurate function for calculating neighbors was the mahalanobis for all validation methods evaluated.

From the best configuration found for the model, the statistical analysis showed a low variation, considering that 100 analyzes were performed using random weights. For all distance calculation methods, the variation between the worst result and the best result was less than 10 % and in the best case, the accuracy reached 85.17 %. This result was superior to that of well-consolidated algorithms such as decision tree, ensemble subspace and SVM. Based on the promising results found in this paper, it will be possible in the future to carry out analyzes to classify the conditions of the insulators in the field. Other ways of extracting characteristics can be applied to compare the influence of the characteristics on the classification results. In the future, the registration of the photos may be performed using drones, facilitating the visualization and classification of the conditions of the insulators through algorithms based on artificial intelligence, as well as presented in this paper.

Author Contributions: Conceptualization, M.P.C. and S.F.S.; methodology, S.F.S.; software, M.P.C.; validation, S.F.S.; formal analysis, M.P.C. and S.F.S.; investigation, M.P.C.; resources, F.L.P.; writing—original draft preparation, M.P.C.; writing—review and editing, S.F.S.; supervision, F.L.P., K.C.Y., R.G.O. and V.R.Q.L.; funding acquisition, R.G.O. and V.R.Q.L.

Funding: This work was supported by national funds through the Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) by the project UIDB/05064/2020 (VALORIZA – Research Centre for Endogenous Resource Valorization). Seed Funding ILIND–Instituto Lusófono de Investigação e Desenvolvimento, COPELABS (COFAC/ILIND/COPELABS 2020) and was partially supported by Fundação para a Ciência e a Tecnologia under Project UIDB/04111/2020 and FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020.

Al Proyeto: Uso de algoritmos y protocolos de comunicación en dispositivos con énfasis en la privacidad de los datos.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: For future comparisons, the data are available at: https://github.com/SFStefenon/InsulatorsDataSet.

Acknowledgments: The authors would like to thank the Emerging Leaders in the Americas Program (ELAP), Canadian Bureau for International Education (CBIE), Government of Canada, who provided a scholarship for Visiting Graduate Research Student in the University of Regina, Canada. The authors are thankful to the Coordination for the Improvement of Higher Education Personnel (CAPES) for Master's scholarship for Marcelo P. Corso.

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

Abbreviations

The following abbreviations are used in this manuscript:

ANFIS	adaptive neuro-fuzzy inference system
ANN	artificial neural network
CAPES	Coordination for the Improvement of Higher Education Personnel
CBIE	Canadian Bureau for International Education
CNN	convolutional neural network
ELAP	Emerging Leaders in the Americas Program
ESDD	equivalent salt deposit density
GMDH	group method of data handling
k-NN	*k*-nearest neighbors
LSTM	long short term memory
NSDD	non-soluble deposit density
SVM	support vector machine

References

1. Ribeiro, M.H.D.M.; Stefenon, S.F.; Lima, J.D.; Nied, A.; Marini, 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.
2. Corso, M.P.; Stefenon, S.F.; Couto, V.F.; Cabral, S.H.L.; Nied, A. Evaluation of Methods for Electric Field Calculation in Transmission Lines. *IEEE Latin America Transactions* **2018**, *16*, 2970–2976. doi:10.1109/TLA.2018.8804264.
3. Stefenon, S.F.; Branco, N.W.; Nied, A.; Bertol, D.W.; Finardi, E.C.; Sartori, A.; Meyer, L.H.; Grebogi, R.B. Analysis of training techniques of ANN for classification of insulators in electrical power systems. *IET Generation, Transmission & Distribution* **2020**, *14*, 1591–1597. doi:10.1049/iet-gtd.2019.1579.
4. Khafaf, N.; El-Hag, A. Bayesian regularization of neural network to predict leakage current in a salt fog environment. *IEEE Transactions on Dielectrics and Electrical Insulation* **2018**, *25*, 686–693. doi:10.1109/TDEI.2017.006936.
5. 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.
6. Abouzeid, A.; El-Hag, A.; Assaleh, K. Equivalent Salt Deposit Density Prediction of Silicone Rubber Insulators Under Simulated Pollution Conditions. *Electric Power Components and Systems* **2018**, *46*, 1123–1133. doi:10.1080/15325008.2018.1488303.
7. Soltani, A.; El-Hag, A. Denoising of Radio Frequency Partial Discharge Signals Using Artificial Neural Network. *Energies* **2019**, *12*, 3485. doi:10.3390/en12183485.
8. 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.
9. Anjum, S.; Jayaram, S.; El-Hag, A.; Jahromi, A. Detection and classification of defects in ceramic insulators using RF antenna. *IEEE Transactions on Dielectrics and Electrical Insulation* **2017**, *24*, 183–190. doi:10.1109/TDEI.2016.005867.
10. Cabral, S.H.L.; Bertoli, S.L.; Medeiros, A.; Hillesheim, C.R.; De Souza, C.K.; Stefenon, S.F.; Nied, A.; Leithardt, V.R.Q.; González, G.V. Practical Aspects of the Skin Effect in Low Frequencies in Rectangular Conductors. *IEEE Access* **2021**, *9*, 49424–49433. doi:10.1109/ACCESS.2021.3069821.
11. 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.
12. Li, W.; Chen, J.; Li, J.; Xia, K. Derivative and enhanced discrete analytic wavelet algorithm for rolling bearing fault diagnosis. *Microprocessors and Microsystems* **2021**, *82*, 103872. doi:10.1016/j.micpro.2021.103872.
13. 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.
14. 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.
15. 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.
16. Nguyen, V.N.; Jenssen, R.; Roverso, D. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power & Energy Systems* **2018**, *99*, 107–120. doi:10.1016/j.ijepes.2017.12.016.
17. Manninen, H.; Ramlal, C.J.; Singh, A.; Rocke, S.; Kilter, J.; Landsberg, M. Toward automatic condition assessment of high-voltage transmission infrastructure using deep learning techniques. *International Journal of Electrical Power & Energy Systems* **2021**, *128*, 106726. doi:10.1016/j.ijepes.2020.106726.
18. Shi, C.; Huang, Y. Cap-Count Guided Weakly Supervised Insulator Cap Missing Detection in Aerial Images. *IEEE Sensors Journal* **2021**, *21*, 685–691. doi:10.1109/JSEN.2020.3012780.
19. Sampedro, C.; Rodriguez-Vazquez, J.; Rodriguez-Ramos, A.; Carrio, A.; Campoy, P. Deep Learning-Based System for Automatic Recognition and Diagnosis of Electrical Insulator Strings. *IEEE Access* **2019**, *7*, 101283–101308. doi:10.1109/ACCESS.2019.2931144.

20. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **2020**, *50*, 1486–1498. doi:10.1109/TSMC.2018.2871750.
21. Ibrahim, A.; Dalbah, A.; Abualsaud, A.; Tariq, U.; El-Hag, A. Application of Machine Learning to Evaluate Insulator Surface Erosion. *IEEE Transactions on Instrumentation and Measurement* **2020**, *69*, 314–316. doi:10.1109/TIM.2019.2956300.
22. Kang, G.; Gao, S.; Yu, L.; Zhang, D. Deep Architecture for High-Speed Railway Insulator Surface Defect Detection: Denoising Autoencoder With Multitask Learning. *IEEE Transactions on Instrumentation and Measurement* **2019**, *68*, 2679–2690. doi:10.1109/TIM.2018.2868490.
23. Dong, Y.; Ma, X.; Fu, T. Electrical load forecasting: A deep learning approach based on K-nearest neighbors. *Applied Soft Computing* **2021**, *99*, 106900. doi:10.1016/j.asoc.2020.106900.
24. Mailagaha Kumbure, M.; Luukka, P.; Collan, M. A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean. *Pattern Recognition Letters* **2020**, *140*, 172–178. doi:10.1016/j.patrec.2020.10.005.
25. González, S.; García, S.; Li, S.T.; John, R.; Herrera, F. Fuzzy k-nearest neighbors with monotonicity constraints: Moving towards the robustness of monotonic noise. *Neurocomputing* **2021**, *439*, 106–121. doi:10.1016/j.neucom.2019.12.152.
26. Sharma, K.K.; Seal, A. Spectral embedded generalized mean based k-nearest neighbors clustering with S-distance. *Expert Systems with Applications* **2021**, *169*, 114326. doi:10.1016/j.eswa.2020.114326.
27. 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.
28. 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.
29. 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.
30. IEC 60383, T. Insulators for overhead lines with a nominal voltage above 1000 V - Part 1: Ceramic or glass insulator units for a.c. systems - Definitions, test methods and acceptance criteria. *International Standard* **1993**, *1*, 1–111.
31. IEC 60507, T. Artificial pollution tests on high-voltage ceramic and glass insulators to be used on a.c. systems. *International Standard* **2013**, *1*, 1–77.
32. IEC 60815, T. Selection and dimensioning of high-voltage insulators intended for use in polluted conditions: Definitions, information and general principles. *International Standard* **2008**, *1*, 1–53.
33. Siriapisith, T.; Kusakunniran, W.; Haddawy, P. Pyramid graph cut: Integrating intensity and gradient information for grayscale medical image segmentation. *Computers in Biology and Medicine* **2020**, *126*, 103997. doi:10.1016/j.compbiomed.2020.103997.
34. Al-Musawi, A.K.; Anayi, F.; Packianather, M. Three-phase induction motor fault detection based on thermal image segmentation. *Infrared Physics & Technology* **2020**, *104*, 103140. doi:10.1016/j.infrared.2019.103140.
35. Zhang, Q.; Lin, C.; Li, F. Application of binocular disparity and receptive field dynamics: A biologically-inspired model for contour detection. *Pattern Recognition* **2021**, *110*, 107657. doi:10.1016/j.patcog.2020.107657.
36. Wang, Z.; Li, Y.; Li, D.; Zhu, Z.; Du, W. Entropy and gravitation based dynamic radius nearest neighbor classification for imbalanced problem. *Knowledge-Based Systems* **2020**, *193*, 105474. doi:10.1016/j.knosys.2020.105474.
37. Kahraman, H.T. A novel and powerful hybrid classifier method: Development and testing of heuristic k-NN algorithm with fuzzy distance metric. *Data & Knowledge Engineering* **2016**, *103*, 44–59. doi:10.1016/j.datak.2016.02.002.
38. Tripoppoom, S.; Ma, X.; Yong, R.; Wu, J.; Yu, W.; Sepehrnoori, K.; Miao, J.; Li, N. Assisted history matching in shale gas well using multiple-proxy-based Markov chain Monte Carlo algorithm: The comparison of K-nearest neighbors and neural networks as proxy model. *Fuel* **2020**, *262*, 116563. doi:10.1016/j.fuel.2019.116563.
39. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Wang, R. Efficient k-NN Classification With Different Numbers of Nearest Neighbors. *IEEE Transactions on Neural Networks and Learning Systems* **2018**, *29*, 1774–1785. doi:10.1109/TNNLS.2017.2673241.
40. Rahmatian, M.; Vahidi, B.; Ghanizadeh, A.; Gharehpetian, G.; Alehosseini, H. Insulation failure detection in transformer winding using cross-correlation technique with ANN and k-NN regression method during impulse test. *International Journal of Electrical Power & Energy Systems* **2013**, *53*, 209–218. doi:10.1016/j.ijepes.2013.04.020.
41. Liu, Z.G.; Zhang, Z.; Liu, Y.; Dezert, J.; Pan, Q. A new pattern classification improvement method with local quality matrix based on k-NN. *Knowledge-Based Systems* **2019**, *164*, 336–347. doi:10.1016/j.knosys.2018.11.001.
42. Čech, P.; Lokoč, J.; Silva, Y.N. Pivot-based approximate k-NN similarity joins for big high-dimensional data. *Information Systems* **2020**, *87*, 101410. doi:10.1016/j.is.2019.06.006.
43. Abu Alfeilat, H.A.; Hassanat, A.B.; Lasassmeh, O.; Tarawneh, A.S.; Alhasanat, M.B.; Eyal Salman, H.S.; Prasath, V.S. Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big Data* **2019**, *7*, 221–248. doi:10.1089/big.2018.0175.
44. Song, B.; Tan, S.; Shi, H.; Zhao, B. Fault detection and diagnosis via standardized k nearest neighbor for multimode process. *Journal of the Taiwan Institute of Chemical Engineers* **2020**, *106*, 1–8. doi:10.1016/j.jtice.2019.09.017.
45. Lee, L.C.; Liong, C.Y.; Jemain, A.A. Validity of the best practice in splitting data for hold-out validation strategy as performed on the ink strokes in the context of forensic science. *Microchemical Journal* **2018**, *139*, 125–133. doi:10.1016/j.microc.2018.02.009.

46. Wong, T.; Yeh, P. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering* **2020**, *32*, 1586–1594. doi:10.1109/TKDE.2019.2912815.
47. Saud, S.; Jamil, B.; Upadhyay, Y.; Irshad, K. Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach. *Sustainable Energy Technologies and Assessments* **2020**, *40*, 100768. doi:10.1016/j.seta.2020.100768.
48. Biedrzycki, J.; Burduk, R. Weighted Scoring in Geometric Space for Decision Tree Ensemble. *IEEE Access* **2020**, *8*, 82100–82107. doi:10.1109/ACCESS.2020.2990721.
49. Zhang, Y.; Chen, H.C.; Du, Y.; Chen, M.; Liang, J.; Li, J.; Fan, X.; Yao, X. Power transformer fault diagnosis considering data imbalance and data set fusion. *High Voltage* **2021**, *6*, 543–554. doi:10.1049/hve2.12059.
50. Yin, C.; Xiao, Z.; Guo, Y.; Shi, C.; Zhang, X.; Wu, G. Method for detecting the pollution degree of naturally contaminated insulator based on hyperspectral characteristics. *High Voltage* **2021**. doi:10.1049/hve2.12076.
51. Mao, X.; Wang, Z.; Crossley, P.; Jarman, P.; Fieldsend-Roxborough, A.; Wilson, G. Transformer winding type recognition based on FRA data and a support vector machine model. *High Voltage* **2020**, *5*, 704–715. doi:10.1049/hve.2019.0294.
52. Zang, Y.; Qian, Y.; Wang, H.; Xu, A.; Sheng, G.; Jiang, X. Method of GIL partial discharge localization based on natural neighbour interpolation and ECOC-MLP-SVM using optical simulation technology. *High Voltage* **2021**, *6*, 514–524. doi:10.1049/hve2.12071.
53. Sun, J.; Zhang, H.; Li, Q.; Liu, H.; Lu, X.; Hou, K. Contamination degree prediction of insulator surface based on exploratory factor analysis-least square support vector machine combined model. *High Voltage* **2021**, *6*, 264–277. doi:10.1049/hve2.12019.
54. Montagna, F.; Buiatti, M.; Benatti, S.; Rossi, D.; Farella, E.; Benini, L. A machine learning approach for automated wide-range frequency tagging analysis in embedded neuromonitoring systems. *Methods* **2017**, *129*, 96–107. doi:10.1016/j.ymeth.2017.06.019.
55. Pinto, H.S.; Américo, J.P.; Leal, O.E.S.; Stefenon, S.F. Development of Measurement Device and Data Acquisition for Electric Vehicle. *Revista GEINTEC* **2021**, *11*, 5809–5822. doi:10.7198/geintec.v11i1.12030,.
56. Leithardt, V.; Santos, D.; Silva, L.; Viel, F.; Zeferino, C.; Silva, J. A Solution for Dynamic Management of User Profiles in IoT Environments. *IEEE Latin America Transactions* **2020**, *18*, 1193–1199. doi:10.1109/TLA.2020.9099759.
57. Cerutti, G.; Prasad, R.; Brutti, A.; Farella, E. Compact Recurrent Neural Networks for Acoustic Event Detection on Low-Energy Low-Complexity Platforms. *IEEE Journal of Selected Topics in Signal Processing* **2020**, *14*, 654–664. doi:10.1109/JSTSP.2020.2969775.