Hybrid regrouping PSO based wavelet neural networks for characterization of acoustic signals due to surface discharges on H.V. glass insulators

Nasir A. Al-geelani a,*, M. Afendi M. Piah a,*, Zuraimy Adzis a, Munir A. Algeelani b

a Institute of High Voltage and High Current, Universiti Teknologi Malaysia, 81310 Johor, Malaysia
b Department of Electrical Engineering, University of Aden, Aden, Yemen

A R T I C L E   I N F O
Article history:
Received 4 May 2012
Received in revised form 30 May 2013
Accepted 8 July 2013
Available online 2 August 2013

Keywords:
Acoustic signal
Dry bands
Glass insulator
RBF-NN
Surface discharge and wavelet transform
PSO and RegPSO

A B S T R A C T
A hybrid algorithm combining Regrouping Particle Swarm Optimization (RegPSO) with wavelet radial basis function neural network referred to as (RegPSO-WRBF-NN) algorithm is presented which is used to detect, identify and characterize the acoustic signals due to surface discharge activity and hence differentiate abnormal operating conditions from the normal ones. The tests were carried out on clean and polluted high-voltage glass insulators by using surface tracking and erosion test procedure of international electro-technical commission 60,587. A laboratory experiment was conducted by preparing the prototypes of the discharges. A very important step for the WRBF network training is to decide a proper number of hidden nodes, centers, spreads and the network weights can be viewed as a system identification problem. So PSO is used to optimize the WRBF neural network parameters in this work.

Therefore, the combination method based on the WRBF neural network is adapted. A regrouping technique called as a Regrouping Particle Swarm Optimization (RegPSO) is also used to help the swarm escape from the state of premature convergence, RegPSO was able to solve the stagnation problem for the surface discharge dataset tested and approximate the true global minimizer. Testing results indicate that the proposed approach can make a quick response and yield accurate solutions as soon as the inputs are given. Comparisons of learning performance are made to the existing conventional networks. This learning method has proven to be effective by applying the wavelet radial basis function based on the RegPSO neural network in the classification of surface discharge fault data set. The test results show that the proposed approach is efficient and revealed a very high classification rate.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Atmospheric elements when accumulated on the insulator’s surface, form a layer of pollutant over time. The dielectric properties of the insulator do not diminish significantly, due to the pollutant layer especially when the layer is dry, but due to high humidity, light rain and even fog, it gets wet, and generates a leakage current resulting in a flash-over which eventually leads to a disaster in service reliability [1,2].

The acoustic technology for target detection has developed very rapidly in the past few years. So strong tools are required, such as signal processing and feature extraction for the detection of such a condition [3,4]. Several researchers successfully used a method of acoustic detection for studying the characteristics of electrical discharges on insulators [5,6]. Many techniques on signal analysis have been used such as the Fourier transform, wavelet transform (WT) as well as neural network in order to characterize and classify the electrical discharge signals [7,8]. But no work has been done up to now on the combined effect of Regrouping Particle Swarm Optimization wavelet transform with radial basis function neural network for characterization of surface discharge (SD).

Wavelet transform (WT) has been successfully employed in various fields of chemistry for signal processing and shape optimization for improving the quality characteristics of the products. A lot of research have been done regarding WT, employed mainly for signal processing in various fields of analytical chemistry, including flow injection analysis (FIA), high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), infrared spectrometer (IR), ultraviolet–visible spectrometer (UV–vis), mass spectrometry (MS), nuclear magnetic resonance spectrometer (NMR), electroanalytical chemistry, and X-ray diffraction [9,10].

Some researchers developed the immune algorithm part of the neural network to optimize machining parameters for milling operations. A new hybrid optimization approach was developed by hybridizing the immune algorithm with hill climbing local search
algorithm to maximize the total profit rate in milling operations [11,12].

Among different structures of artificial neural networks (ANNs), the multilayer perceptron with the error-back-propagation training algorithm called back-propagation network (BPN) is the most popular one. However, due to its multilayered structure and the greedy nature of the back-propagation algorithm, the training process often settles in the undesirable local minima of error surface or converges slowly. Recently radial basis function neural networks (RBF-NNs) have been found to be very attractive for many problems. An important property of the RBF-NNs is that they form a unifying link among many different research fields such as function approximation, regularization, noisy interpolation, pattern recognition, and medicine. The increasing popularity of the RBF-NNs is partly due to their simple topological structure, their locally tuned neurons, and their ability to have a fast learning algorithm in comparison with other multilayer feed forward neural networks [13,14].

An optimization of the architecture of the Multilayer Perceptron Neural Network (MLP) is made using an implementation of Genetic Algorithm (GA) in the SAS system (SAS Institute Inc., 2010), in order to improve the predictive power of the credit risk scorecards. The study showed an area under the ROC curve of the GA (71–25%) is significantly larger than that of the MLP neural network using the default parameters of SAS Enterprise Miner (68–09%) and the logistic regression (65–92%). This difference indicates that the GA in the MLP neural network has a greater predictive power at all risk levels. The only alternative that slightly exceeds the GA performance is the global optimum (71–26%). The GA used to optimize the MLP neural network architecture in 9.3 h (559 min) in a 30 iteration run and made 274 function calls while the model used to find the global optimum took 139.2 h (8,356 min) and made 4,096 function calls [15].

Genetic algorithms (GA) offer an efficient search method for a complex problem space and can be used as powerful optimization tools. By selecting suitable parameters to control the GA, high efficiency and good performance can be achieved. Benchmark problems have been applied to evaluate the performance and properties of the evolutionary trained net. The well-known exclusive-or (XOR) problem has been selected. Based on a multitude of experiments evaluating different parameters to control the genetic algorithm. In general, genetic algorithms are inherently slower than back-propagation (BP). This could be expected due to their global search technique compared to the highly directed gradient descent learning of back-propagation. But even this makes it valuable and helpful in those cases BP fails [16].

A hybrid method which combines neural network classifiers by genetic algorithm is presented, which also considers the difference in performance of each network by combining the networks is also presented. This method utilizes the weight parameters, which are determined by genetic algorithm, to obtain the combined output. The superiority in performance of the proposed method and compared with conventional methods by thorough experiments in a real-world pattern recognition problem was demonstrated. Handwritten digit database of Concordia University of Canada, which consists of 6000 unconstrained digits originally collected from dead letter envelopes by the US Postal Services at different locations in the US were used. Although the network learned the training set almost perfectly, the performances on the test sets were quite different. Furthermore, the performance did not improve by training a large network with considering all the features used by each network. This is a strong evidence that the multiple neural network might produce a better result than conventional single network approach. Actually, the method has a small, but statistically significant, advantage in recognition rates obtained by the conventional methods. The error rate of the method was 2.10%, which was the highest, which was a big improvement compared with those of the previous methods [17].

In some applications, electrical surface discharge detection methods are not very effective, typically as a result of excessive interfering signal [18,19]. Acoustic method has been used, which has advantages over the electrical surface discharge detection methods in that they are immune and non-invasive to electromagnetic noise. For the signal analysis, wavelet signal processing was used to de-noise the surface discharge acoustic signal by discarding the noise [20–22].

Furthermore, gating techniques was used to raise the efficiency of surface discharge acoustic signal extraction. Time and frequency measurements were considered in both domains. Acoustic signals are familiar with their dominating features that are frequency, phase and amplitude in which; key to many signal analysis solutions is the frequency feature.

Orthonormal basis function local in time can be provided by wavelet transforms. The beauty of WT is that it can nearly give a signal without distortion. The WT can also provide a multi-resolution concept, which is used in signal processing, identification of acoustic signals, numerical analysis and weak signal detection.

How to determine the structure and parameters of the neural networks promptly and efficiently has been a difficult point all the time in the field of RBF neural networks research, these parameters are the position of RBF centers, the width of RBFs, and the weights. The main disadvantage of this method is that it is very difficult to quantify how many numbers of the centers should be adequate to cover the input vector space. Furthermore, the training algorithm is prone to getting stuck in a local minimum. The selection of these optimal parameters plays an important role in various training algorithms. A single parameter choice has a tremendous effect on the rate of convergence. In this paper, the optimal parameters are determined by the PSO algorithm to calculate the number of the centers and all other parameters for the WRBF-NN and also to overcome the premature convergence to help the swarm escape from the state of stagnation, the idea of regrouping technique was implemented [23].

This work describes and portrays the great capabilities of WT to extract unique features from the signal of the SD which when combined with radial basis function neural network (RBF-NN) and optimized using a regrouping particle swarm algorithm gives very high classification accuracy.

The organization of this paper is as follows. In Section 2, the concept of WT applied for SD detection is described first. Although there are many types of ANNs, we focus our study on the commonly seen feed-forward network, namely, RBF-NN. A brief introduction to RBF-NN is also given in Section 3. Section 4 throws a light upon the PSO algorithm and also explains the regrouping technique with the optimal selection of parameters. Section 5 reports the experimental setup obtained by developing a model to detect the SD acoustic signals. The feature extraction and the feature vector from the normalized inputs is being generated in Section 6. The analysis of results and discussions are given in Section 7. Section 8 presents some comparison with other related works and the conclusion is made in Section 9.

2. The Wavelet Transform

The acoustic signals have some non-linear characteristics due to the surface discharge, and that makes some difficulties to deal with because of nonlinear and the random like behavior of the system. The problem of non-linearity of the acoustic signal is overcome by using wavelet transform, which is a strong tool for feature picking-up. It is equivalent to filters. Details \( d_n \) are produced by high-pass filters and approximations \( a_n \) are produced by low-pass
filters. Due to the multi-dimensional characters which the wavelets possess, there are able to adjust their scale to the nature of the signal features [24]. It can zoom in or zoom out the required details just like a microscope.

Furthermore, wavelets can decompose a signal to give dilations and translation parameters, so the information in the signal is presented with these parameters in the form of frequencies. The Matlab wavelet toolbox is used to verify the algorithm where discrete wavelet transforms (DWT) is used to analyze the signals. The coefficients are generated, and the features from the signal are extracted. Wavelet is a good tool to analyze the non-linear signals as it represents the features both in time and frequency domains [25,26]. The WT analyses the non-periodic surface discharge signal and adopt the principle of linking of frequency scales. Generally, the DWT is used for this mission.

The equation for non-static signal for a DWT is shown below [27,28].

\[
f(t) = \sum_{k} c_{0,k} \phi_{0,k} (t) + \sum_{j \geq 0} \sum_{k} w_{jk} 2^{j/2} \psi(2^{j} t - k)
\]  

where: \( \psi \) – Mother wavelet function; \( j \) – Dilation or level index; \( k \) – Translation or scaling index, \( \phi_{0,k} \) – scaling function of the coarse scale coefficients, \( c_{0,k} \), \( W_{jk} \) scaling function of detail (fine) coefficients.

One of the capabilities of DWT is that, it produces details to show high-frequency information and approximations to show low frequency information. The most suitable mother wavelet for detecting SD acoustic signals is the Daubichies (Db) wavelets transform, which is capable of detecting short duration, fast decaying, high-frequency and low amplitude signals. The decomposition process in the WT consists of many numbers of filters from (Db2 to Db44), so the most promising number depends upon how they minimize the aliasing. Basically, in the first stage the captured signal is divided into two by the frequency bandwidth, which is then passed to high-pass and low-pass filters [29]. After that the output signal from the low-pass filter is further subdivided into two of the frequency bandwidth and sent to the following stage [30]. This procedure continues until the predetermined number of levels is reached. The output of the final stage represents the same captured signal but at different frequency bands [31,32]. The suitable selection of mother wavelet depends on the application. Among the various de-noising techniques, from the point of view of the de-noising effect and the computing time the DWT method is the most suitable.

Finally, the Daubichies wavelet is the most appropriate for treating SD [33,34]. In this study, the adaptability of the Daubichies wavelets of orders 2 has been evaluated, and results have shown the superiority. It is befitting to select a suitable number of breakup levels based on the nature of the signal. Based on acoustic signal features, it is seen that six levels of decomposition is the best choice, because it has described the SD acoustic signal in a more mindful and symptomatic way. This decision is mainly due to the low-frequency band (approximation), which is the most valuable part of the acoustic signal [35].

3. Radial basis function networks

Radial basis function (RBF) networks have certain advantages over other types of ANNs and have been widely applied in many science and engineering fields. It is a three layered feed-forward and fully connected network. The output layer is nonlinear, and the connections of the output layer are only weighted. The connections from the input to the hidden layer are not weighted. It is a feed-forward network with a single layer of hidden units, called radial basis functions (RBFs). RBF outputs show the maximum value at its center point and decrease its output value as the input leaves the center. Typically, the Gaussian function is used for the activation function. The RBF network is constructed with three layers: input layer, hidden layer and output layer. In input layer, the number of neurons is the same with the number of input dimensions. In the case of Gaussian function, this value represents a measure of the quality of the match between the input vector and the location of the center in the input space. Each hidden node, therefore, can be considered as a local detector in the input data space [36,37].

One of the unique features of Radial basis networks is that they can be represented by simple functions. They could cope with any type of model, linear or nonlinear and to any network single layer or multi-layer. A RBF-NN is said to be nonlinear, if there exists more than one hidden layer or if the basis functions can move or change the size. The activation function of a hidden unit is predicted by the distance between the input vector and a prototype vector [38].

\[
z_{k}(\mathbf{x}) = \sum_{j=0}^{M} w_{kj} \phi_{j}(\mathbf{x}) = \sum_{j=0}^{M} w_{kj} \phi_{j}(\mathbf{x})
\]

where: \( \mathbf{x} \) – is the input vector, \( \phi_{j} \) is the activation of one of the RBFs, \( w_{kj} \) is the weight of each RBF, \( M \) – is the number of RBFs. \( z_{k} \) – is the output (linear sum of radial basis function).

In a 3-layer RBF-NN, conversion from the input room to the hidden room uses a nonlinear function and linear transformation takes place between the hidden and the output layer. The hidden units used in radial basis functions usually take the form of [38]:

\[
\phi_{j}(\|\mathbf{x} - \mathbf{\mu}_{j}\|)
\]

The distance between the input vector \( x \) and a vector \( \mu_{j} \) of the function usually depends on Euclidean distance.

\[
\|\mathbf{x} - \mathbf{\mu}\|^{2} = \sum_{i} (x_{i} - \mu_{ji})^{2}
\]

The most ordinary form of basis function used is the Gaussian function.

\[
\phi_{j}(\mathbf{x}) = \exp \left( - \frac{\|\mathbf{x} - \mathbf{\mu}_{j}\|^{2}}{2\sigma^{2}} \right)
\]

\( \mathbf{\mu} \) and \( \sigma^{2} \) are the \( j \)th center vector and the width parameter, respectively. A hidden neuron is more susceptible to data points near its center. This sensitivity may be adjusted by tuning the width \( \sigma \). Larger width leads to less sensitivity. For a given input vector, typically only a few numbers of hidden units will have notable activations [39]. RBF neural networks can design complicated mappings compared to multilayer perceptron. In addition, RBF-NN has two-layer weights only and a simple learning algorithm, that makes it very fast in training speed compared to multilayer perceptron.

MATLAB provides two commands that can be used to design RBF neural network. Newrb and newrbe. Newrb adds neurons step by step until the goal is hit, with long training time and a little error, while newrbe very quickly designs a network with zero error [40,41]. In the training process, the following steps should be fulfilled:

(a) The hidden layer’s number of neurons.
(b) The coordinates of the center of RBF function.
(c) The radius (spread) of each RBF function in each dimension.

4. The PSO algorithm

The particle swarm optimization (PSO) was first introduced by Kennedy and Eberhart in 1995. Through the simulation of a simplified social system, the behavior of PSO can be treated as an
optimization process [23]. A swarm in PSO consists of a number of particles. Each particle represents a potential solution of the optimization task. All the particles iteratively discover the probable solution, and generate a position according to the new velocity and the previous positions of the cell, and is compared with the best position which was generated by previous particles in the cost function. The best solution is then kept, that is each particle accelerating in the directions of not only the local best (pbest) solution but also the global best (gbest) position. If a particle discovers a new probable solution, other particles will move closer to it so as to explore the region more completely in the process [42].

PSO is an evolutionary computation technique motivated by the simulation of social behavior. In the PSO system, each agent makes his decision according to his own experiences and other agent experiences. The system initially has a population of random solutions. Each potential solution, called a particle (agent), is given a random velocity and is flown through the problem space. The basic concept of the PSO technique lies in accelerating each agent toward its pbest and gbest locations, with a random weighted acceleration at each time step. Searching procedures by PSO can be described as follows: a flock of agents optimizes a certain objective function; each agent knows its best value so far (pbest) and its position. Moreover, each agent knows the best value in the group (gbest) among pbest, namely the best value so far of the group. The modified velocity and the distance from pbest and gbest as shown below [43]:

\[ v_{ik}^{k+1} = w_{ik} v_{ik}^k + c_1 r_{and1} (pbest_{ik} - x_{ik}^k) + c_2 r_{and2} (gbest - x_{ik}^k) \]

(6)

where: \( v_{ik}^k \) denotes current velocity of agent \( i \) at iteration \( k \); \( v_{ik}^{k+1} \) denotes modified velocity of agent \( i \) at iteration \( k+1 \); \( r_{and1}, r_{and2} \) denote a random number between 0 and 1; \( x_{ik}^k \) denotes current position of agent \( i \) at iteration \( k \); pbest denotes pbest of agent \( i \); gbest denotes gbest of the group; \( w_{ik} \) denotes weight function for velocity of agent \( i \); \( c_1, c_2 \), weight coefficients for each term; \( k \) denotes the current iteration number.

Using the above equation, a certain velocity that gradually gets close to pbest and gbest can be calculated. The current position can be modified by the following equation:

\[ x_{ik}^{k+1} = x_{ik}^k + v_{ik}^{k+1} \]

(7)

The inertia weight \( w \) is introduced to improve PSO performance. Suitable selection of inertia weight \( w \) provides a balance between global and local exploration and exploitation. The inertia weight \( w \) is set according to the following equation.

\[ w_{ik} = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times k \]

(8)

where \( w_{max} \) initial weight, \( w_{min} \) final weight and \( iter_{max} \) maximum iteration number [43].

The basic PSO is influenced by a number of control parameters, namely the dimension of the problem, number of particles, acceleration coefficients, inertia weight, neighborhood size, number of iterations, and the random values that scale the contribution of the cognitive and social components. Additionally, if velocity clamping or constriction is used, the maximum velocity and constriction coefficient also influence the performance of the PSO. The velocity clamping effect was introduced to avoid the phenomenon of “swarm explosion.” With no restriction on the maximum velocity of the particles, a simple one-dimensional analysis of the swarm dynamic concludes that the particle velocity can grow unbounded while the particle oscillates around an optimum, increasing its distance to the optimum on each iteration. In several studies, the velocity clamping parameter was set to the maximum values of the positions in the search space. Further studies demonstrated that this mechanism was not enough to properly control the particle velocities. When compared to evolutionary computation (EC) techniques the PSO quickly identified the region where the optimum was located, but had trouble to adjust the velocity to lower values to perform a fine-grained search of the area. The constriction coefficient, \( \chi \), evaluates to a value in the range \([0,1]\), which implies that the velocity is reduced at each time step. Eberhart and Shi showed empirically that if velocity clamping, and constriction are used together; faster convergence rates can be obtained. Best results were achieved when both the constriction and the velocity clamping mechanisms were combined, using \( V_{max} = \chi \times V_{max} \).

\( V_{max} \) – velocity clamping parameter. \( X_{max} \) – the positions in the search space [44].

### 4.1. Regrouping technique

The idea of regrouping (Reg) is to help the swarm escape from the state of premature convergence, which is primarily troublesome on nonlinear problems. A regrouping mechanism is developed by which to liberate particles from the state of premature convergence so that exploration can continue. This regrouping mechanism will make use of the state of the swarm when premature convergence is detected in order to re-organize the swarm according to information inferred from the swarm state. The regrouping mechanism should work better than simply restarting on the same search space repeatedly and should still be applicable to a variety of problem types. RegPSO was able to solve the stagnation problem for the SD tested and approximate the true global minimizer with each trial conducted [45,46].

### 4.2. PSO parameters

The selection of optimal parameters plays an important role in various training algorithms. A single parameter choice has a tremendous effect on the rate of convergence. For this work, the optimal parameters are determined by trial and error experiments. The search space range available for \( X_1 \) is defined as \((-1, 10, 0, 0)\), \( X_2 \) is defined as \((0, 10, 0)\), and \( X_3 \) is defined as \((1, 10, 0)\), that is, the particle cannot move out of this range in each dimension. The other settings of the PSO algorithm are as follows: swarm size was set to 5, \( c_1 \) and \( c_2 \) were both set to 1.49618, and the inertia weight linearly varied from 0.9 to 0.4 per grouping. The parameter settings of BP-NN, RBF-NN and RegPSO are presented in Table 1.

#### Table 1: Parameters settings for the algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Number of hidden layers</th>
<th>Activation functions of all neurons</th>
<th>Training method</th>
<th>Mean squared error goal</th>
<th>Max epochs</th>
<th>Number of layers</th>
<th>Number of neurons</th>
<th>Mean squared error goal</th>
<th>Spread of radial basis functions</th>
<th>Maximum number of neurons</th>
<th>Number of neurons to add between displays</th>
<th>Number of layers</th>
<th>Number of hidden layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP-NN</td>
<td>1</td>
<td>Tan Sigmoid (tansig)</td>
<td>Levenberg–Marquardt (trainlm)</td>
<td>0.0002</td>
<td>10,000</td>
<td>3</td>
<td>4</td>
<td>0.0002</td>
<td>0.11</td>
<td>30</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>RBF-NN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reg-PSO</td>
<td>Population size</td>
<td>20</td>
<td>Iterations</td>
<td>10,000</td>
<td>W</td>
<td>C1</td>
<td>C2</td>
<td>Velocity clamping percentage λ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.9, 0.4]</td>
<td>1.49618</td>
<td>1.49618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4625
5. Experimental setup

The test setup consists of two main parts: the circuit loop (AC source, transformer, connections and insulator), the measurement and acquisition system (earthing resistor, wideband antennas and high-resolution digital oscilloscope, model Tektronix TDS55000B series). Fig. 1 shows the experimental set-up for generating surface discharge as well as detecting the consequent acoustic signal of surface discharge. The test was designed based on the IEC 60587 standard test procedures. A point to plane electrode configuration was used and mounted at the top and bottom side of the glass insulator surface. The surface discharges were generated across the electrodes by applying high-voltage stress across them.

The glass insulator was fixed on a wooden base, and the electrodes were fixed by some arrangement on the insulator. A series of experiments were performed on H.V glass insulator, which are extensively used in transmission lines. Before the tests, the insulator surface was cleaned by washing with isopropyl alcohol and rinsing with distilled water, in order to remove any trace of dirt and grease. To reproduce saline pollution typical of coastal areas, the insulators were sprayed with a solution, consisting of Nacl and distilled water, with different degrees of salinity from (10 Nacl g/l to 30 Nacl g/l). A peristaltic pump was used to continuously deliver the electrolyte at a fixed flow-rate of 0.60 ml/min. Eight layers of filter paper were used between the top electrode and the sample, which act as an electrolyte reservoir to ensure proper flow of electrolyte along the insulating material surface. The contaminant must flow from the quill hole at the bottom of the top electrode and should not squirt out of the side or top of the filter paper. The specimen was adjusted so that the electrolyte ran down as near as possible to the centerline of the specimen.

An ultrasound detector (USD) with a parabolic antenna was used to detect the acoustic signals resulted from the surface discharge activity and was placed in a suitable position from the specimen. The USD was then connected to a digital oscilloscope to capture and record the acoustic signals. The recorded signals were then processed using MATLAB platform. Many trials were done using the experimental set-up and the most logical data was finalized.

Four test conditions were conducted in the laboratory, including a first condition in which the insulator was kept clean, the second condition in which the insulator was lightly contaminated by a layer of Nacl solution (10 g/l), a third condition in which the insulator was medium contaminated by a layer of Nacl solution (20 g/l) and the fourth condition in which the insulator was heavily contaminated by a layer of Nacl solution (30 g/l). In the first condition the insulator was kept clean and the result is shown in Fig. 2(a). This signal is being processed by using wavelet analysis to remove the noise. It can clearly be seen from the black in color signal that the clean insulator has no or very small surface discharge activity. So this pattern could be considered to be the reference, default pattern or the target to RBF-NN. Fig. 2(b)–(d) shows the second condition, the third condition and the fourth condition and its de-noise signals respectively.

6. Feature extraction

The wavelet transform is well suited in identifying sharp edge transitions. The decomposition of a signal using the wavelet basis has an inherent adaptation to the signal spatial characteristics. In this work, the SD signals that were collected in the experimental process were processed using the DWT to obtain a feature vector that will be used in the next stage of classification. The approach of using the DWT to extract a feature vector, apart from utilizing the properties of the transform itself in representing the signal, has the advantage that it can be combined with sensitivity improvement and noise rejection in a single step. The DWT due to its time frequency localization, unlike the Fourier transform where all time information about the signal is lost, can able to successfully handle such randomly occurring of SDs [47,48].
6.1. Processing the data

(1) The goal of preprocessing is to reduce the number of parameters to face the challenge of “curse of dimensionality.”

(2) The preprocessing has a huge impact on performances of neural networks.

(3) The unwanted field noise and all the interference are filtered off, which massages the data to unique features.

6.2. Decomposition of surface discharge signals

In this work the feature vector is obtained by applying wavelet transform. For a single decomposition, the number of coefficients produced by Db2 is 7 that is 1 approximation coefficient and 6 detailed coefficients. This approximation coefficient that is a low-frequency component is further decomposed into approximation and detailed coefficients. In using the DWT certain features of the signal that is not immediately obvious in the time domain become more apparent through this multi scale differential operator. The features selected to represent the most important part of the data are:

Mean \( \eta \), standard deviation \( \sigma \), normalized skewness \( \gamma \) and normalized kurtosis \( k \), at each decomposition node, were used as a fingerprint for SD and as an input to the classifier. The mean \( \eta \), the standard deviation \( \sigma \), the normalized skewness \( \gamma \) and the normalized kurtosis \( k \), were estimated from the following equations [48]:

\[
\eta = \frac{1}{N} \sum_{n=1}^{N} (x[n])
\]

\[
\sigma = \left( \frac{1}{N} \sum_{n=1}^{N} (x[n] - \eta)^2 \right)^{1/2}
\]

\[
\gamma = \frac{1}{N \sigma^2} \sum_{n=1}^{N} (x[n] - \eta)^4
\]

\[
k = \frac{1}{N \sigma^4} \sum_{n=1}^{N} (x(n) - \eta)^4
\]

where \( x[n] \) is the wavelet coefficients at positive \( n \) and \( N \) is the total number of wavelet coefficients used in each scale. By taking 6-detailed coefficients and 1-approximation coefficients from the wavelet analysis for the 4-conditions mentioned in Section 5, we will have the feature vector which uses four descriptors for each scale; therefore, a feature vector of dimensions was needed to represent the data. So the feature vector is of dimension \( 42 \times 4 \), i.e. 21 samples for the 3-contaminated conditions and other 21 samples for the clean condition that totally gives us 42 dimensional vectors each having 4-features. These coefficients are used to train the RBF neural network, so that it should discriminate SD activity from normal operating conditions.

7. Results and discussion

The four inputs are the main features of the surface discharge, which are kurtosis, mean, standard deviation and skewness extracted by using wavelet transform technique. The target vector consists of two classes that are:

The first class is denoted by (0) which shows no or very small SD activity. The second class is denoted by (1) which shows high
The normalized inputs and target vectors are randomly divided into three sets. 60% of the vectors are used to train the network. 20% of the vectors is used to validate how well the network generalized. Finally, the last 20% of the vector provide an independent test of network generalization to data that the network has never seen.

The goal of regression analysis is to determine the values of parameters for a function that cause the function to best fit a set of data observations that is provided. In linear regression, the function is a linear (straight-line) equation. Regression analysis is widely used for prediction and forecasting, where its use has been substantial overlap with the field of machine learning. The regression in Fig. 3 is perfectly 1 that means the fitting is perfect and completely overlapped the target. Fig. 3 depicts the optimization processes of the best fitness value obtained are due to applying the regrouping technique for particle swarm optimizer algorithms and we can see how effectively it avoids stagnation. The output tracks the targets very well for training, testing, and validation, and the regression value is over 1 of the total response. For added analysis, the NN training goal is set as a black dotted line which is denoted by $Y$. The stopping conditions of RegPSO-WRBFF-NN were set to a minimum error of 0.0002 or a maximum iteration of 10,000.

Fig. 4 shows the receiver operating characteristic (ROC) curve, which is another measure used to test the classifier sensitivity. For each class of a classifier, ROC applies threshold values across the interval $[0, 1]$ to outputs. For each threshold, two values are calculated, the True Positive Ratio (the number of outputs greater or equal to the threshold, divided by the number of one targets), and the False-Positive Ratio (the number of outputs less than the threshold, divided by the number of zero targets). The red-colored lines on each axis represent the ROC curves for each of the two categories of this test problem. ROC curve is a plot of the true-positive rate (sensitivity) versus the false-positive rate ($1 - \text{specificity}$) as the threshold is varied. This is a perfect test which shows points in the upper-left corner, with 100% sensitivity and 100% specificity. For this problem, the network performs almost perfectly.

A dynamic 3-D scatter diagram with the $x$-axis being PSO dimension one, the $y$-axis being PSO dimension two, and the $z$-axis being the NN performance dimension 3. Fig. 5(a–d) shows the complete operation of particles searching for the best position until they reached the desired or the NN training goal. The number and range of PSO dimensions are determined by the NN weights and bias values. Therefore, each PSO dimension represents one NN weight or bias.

Each PSO particle is shown in this plot as a red marker. For each PSO iteration, the position of each PSO particle is updated. The PSO particle positions over all iterations are displayed. This is useful to visually present the convergence or divergence of PSO particles in the search space.

The NN goal is shown as a straight line in the search space to indicate an approximate area were the PSO particles are desired to converge to. So by looking at the Fig. 5(a–d) we can see that the particles update their positions and a clear progress in the performance can be realized where in Fig. 5(d) the five particles overlap on with the target giving a very high performance.

Fig. 6 illustrates the efficient regrouping mechanism by which the swarm can escape from local minima and continue its search without needing to restart on the entire search space. It can be noted that shortly after the first regrouping and immediately after the second regrouping, solution quality was improved as particles escaped premature convergence in order to continue onwards toward the true global minimizer rather than simply stagnating in place and so on to third and fourth regrouping trying to overcome the stagnation.

Mean squared error (mse) is a network performance function. It measures the network’s performance according to the mean of squared errors. The unique feature of RBF-NN is depicted in Fig. 7 which shows that the mean squared error (mse) required only 25 neurons to attain zero value. This is the beauty of combining the WT with RBF-NN optimized with RegPSO.

8. Comparison with other related research works

In order to check the reliability of the proposed hybrid model RegPSO-WRBFF-NN is being compared with the previous works conducted and mentioned in the literature review. The same SD dataset is being used to compare with the Wavelet Back-Propagation neural network (WBP-NN), Regrouping Particle Swarm Optimization Wavelet Back-Propagation Neural Network (RegPSO-WBP-NN) and finally, Wavelet radial basis function neural network (WRBF-NN).

The summary of the results is shown in Table 2, although the table section provides some insight into the behavior of the different algorithms, more information can be gleaned from the plot in
Fig. 5. (a) PSO particles in the search space, (b) PSO particles updating positions, (c) PSO particles updating positions and approaching the target, (d) PSO particles hit the target.

Fig. 6. The Regrouping PSO mechanism.

Fig. 7. The mean square error.

RegPSO-WBP-NN can achieve a higher classification rate as can be seen from Fig. 10 than WBP-NN in the training process since

![Image](https://via.placeholder.com/150)

**Fig. 8.** The WBP-NN network converges to the best value after 30 iterations. The plot in Fig. 8 shows three lines, because the normalized inputs and target vectors are randomly divided into three sets. 60% of the vectors are used to train the network, 20% of the vectors is used to validate how well the network generalized. Finally, the last 20% of the vector provide an independent test of network generalization to data that the network has never seen. Training on the training vectors continues as long as the training reduces the network error on the validation vectors. After the network memorizes the training set (at the expense of generalizing more poorly), training is stopped. By training the WBP-NN with the normalized inputs, we can see the result which shows us training performance did not reach the goal perfectly. The result here is reasonable, because the test set error and the validation set error have similar characteristics, and it do not appear that any significant overfitting has occurred. An updating algorithm usually causes the training process to converge slowly, and the classification rate is very poor as observed in Fig. 9.

RegPSO-WBP-NN can achieve a higher classification rate as can be seen from Fig. 10 than WBP-NN in the training process since

<table>
<thead>
<tr>
<th>Model</th>
<th>Computational time (s)</th>
<th>Number of iterations</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBP-NN</td>
<td>1.069211</td>
<td>30</td>
<td>75.109</td>
</tr>
<tr>
<td>RegPSO-WBP-NN</td>
<td>13.907</td>
<td>254</td>
<td>92.71</td>
</tr>
<tr>
<td>WRBF-NN</td>
<td>0.945416</td>
<td>25</td>
<td>96.98</td>
</tr>
<tr>
<td>RegPSO-WRBF-NN</td>
<td>15.898</td>
<td>253</td>
<td>100</td>
</tr>
</tbody>
</table>

Notation:
- WBP-NN: wavelet back propagation neural network
- RegPSO-WBP-NN: Regrouping Particle Swarm Optimization wavelet back propagation neural network.
- WRBF-NN: Wavelet radial basis function neural network.
- RegPSO-WRBF-NN: Regrouping Particle Swarm Optimization wavelet radial basis function neural network.
Fig. 8. The training performance of WBP-NN.

Fig. 9. The linear regression between the network outputs and the corresponding targets of WBP-NN.

Fig. 10. The linear regression between the network outputs and the corresponding targets of RegPSO-WBP-NN.

Fig. 11. The training performance of WRBF-NN.

Fig. 12. The linear regression between the network outputs and the corresponding targets of WRBF-NN.

its parameters are optimized and applies the regrouping technique for particle swarm optimizer algorithms and we can see how effectively it avoids stagnation and reaches somehow a better classification rate.

Compared with the WRBF-NN, the training performance can be seen in Fig. 11 where the training curve hit the goal within 25 epochs only giving a good performance. The regression plot in Fig. 12 reports that the output overlaps the target in a good manner and provides an acceptable classification rate. Since it can update the parameters, at the same time it does not need the previous updating values. It is found that global searching may easily be trapped in the local minimum; therefore, this manner makes the estimation errors in the BP-NN model difficult to be minimized under the same system parameters as that in RBF-NN. The major problem of traditional WRBF-NN is how to choose the optimal number of weights, bias and even the centres which if not suitably and carefully selected can lead to a long and time-consuming process, so to overcome this deficiency the basic PSO method is used. PSO is a form of swarm intelligence a type of multi-agent system whereby individual agents behave according to simple rules but interact to produce a surprisingly capable collective behavior, but the only
fact that basic PSO suffers is that for nonlinear signals as SD signal it sticks in the local minimum, so no progress is being made toward a better solution since the swarm has stagnated, so to make a progress and take out the particles from the stagnation condition regrouping technique was implemented, regrouping helps liberate the swarm from the state of premature convergence in order to enable continued progress toward true global minimum, that can be seen from the results that it has achieved a very high quality of accuracy in terms of performance and classification rate as could be seen from Table 2. Although the computational time is somehow high compared to the other models which is not a major concern in the field of the power system as long as the classification rate is very high, this aspect is very important to avoid false detection of surface discharge activity, which can lead to unnecessary shut-downs of electrical power, which may cause a loss of thousands of dollars to companies. One algorithm’s weakness may be another’s strength, so that no method can be said to be irrelevant to the study of optimization algorithms.

9. Conclusion

Having done all these analyses, we can conclude that acoustic detection of surface discharge is possible by observing the results which are quite good, explainable and logical. The proposed hybrid model has proven to characterize the SD activity with a high degree of integrity, which is attributed to the combined effect of the WT and RBF-NN optimized by RegPSO. The high rate of classification accuracy acquired by the RBF-NN is due to the preprocessing of data by the WT and the great capabilities of the RegPSO. Based on the same network structure, the proposed method provides a more efficient search scheme to determine the related parameters of the RBF network than the other mentioned methods and RegPSO was able to solve the stagnation problem for the surface discharge tested and approximate the true global minimizer with each trial conducted. However, the proposed method classifies the surface discharge with a very high rate which reaches to almost 100% compared to the other methods.

we can also see that the RegPSO-RBF-NN algorithm has more smooth mean recognition rate. This shows that the proposed algorithm is more stable than the other algorithms. Although the computational time is somehow high compared to the other models which is not a major concern in the field of a power system as long as the classification rate is very high, this aspect is very important to avoid false detection of surface discharge activity which can lead to unnecessary shut-downs of electrical power, which may cause a loss of thousands of dollars to companies.

The model is quite versatile for a wide range of applications in the field of power system analyses. As future work four types of partial discharges will be generated in a high-voltage laboratory, namely corona discharge in the air, floating discharge in oil, internal discharge in oil and surface discharge in the air, at different applied voltages will be recorded and a feature vector will be extracted by using wavelet transform, which will be used to train the RegPSO-RBF-NN.

Acknowledgments

This work is supported by Universiti Teknologi Malaysia (UTM) under the Research University Grant Scheme of Vote number (1) Q,J130000.2523.03H86 and (2) Q,J130000.7123.01H69. The authors would like to thank Research Management Center (RMC) Universiti Teknologi Malaysia, for the research activities. The authors would also like to thank the anonymous reviewers who have contributed sincerely in this work.

References