Article

Distribution Grids Fault Location employing ST based Optimized Machine Learning Approach

Md Shafiullah [1] , M. A. Abido [1,*] and Taher Abdel-Fattah [2]

[1] Electrical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; shafiullah@kfupm.edu.sa
[2] Electrical Engineering Department, Menoufia University, Shebin El-Kom 32511, Egypt; shafi.buet03@gmail.com
* Correspondence: mabido@kfupm.edu.sa; Tel.: +96-650-875-7838

Abstract: Precise information of fault location plays a vital role in expediting the restoration process, after being subjected to any kind of fault in power distribution grids. This paper proposed the Stockwell transform (ST) based optimized machine learning approach, to locate the faults and to identify the faulty sections in the distribution grids. This research employed the ST to extract useful features from the recorded three-phase current signals and fetches them as inputs to different machine learning tools (MLT), including the multilayer perceptron neural networks (MLP-NN), support vector machines (SVM), and extreme learning machines (ELM). The proposed approach employed the constriction-factor particle swarm optimization (CF-PSO) technique, to optimize the parameters of the SVM and ELM for their better generalization performance. Hence, it compared the obtained results of the test datasets in terms of the selected statistical performance indices, including the root mean squared error (RMSE), mean absolute percentage error (MAPE), percent bias (PBIAS), RMSE-observations to standard deviation ratio (RSR), coefficient of determination (R^2), Willmott's index of agreement (WIA), and Nash–Sutcliffe model efficiency coefficient (NSEC) to confirm the effectiveness of the developed fault location scheme. The satisfactory values of the statistical performance indices, indicated the superiority of the optimized machine learning tools over the non-optimized tools in locating faults. In addition, this research confirmed the efficacy of the faulty section identification scheme based on overall accuracy. Furthermore, the presented results validated the robustness of the developed approach against the measurement noise and uncertainties associated with pre-fault loading condition, fault resistance, and inception angle.

Keywords: distribution grid; extreme learning machine; fault location; feature extraction; multilayer perceptron neural network; s-transform; support vector machine

1. Introduction

The detailed information regarding the faulty area and fault location, plays a vital role in expediting the restoration process in electric utilities after being subjected to any kind of fault. Consequently, there is a growing research interest to locate faults and identify faulty sections in distribution grids efficiently, to reduce customer minute losses and revenue losses for the utilities. The available fault location techniques for distribution grids can be categorized into three major groups namely the impedance, the traveling wave, and the knowledge-based techniques [1–3]. The impedance-based technique, evaluates fault location using voltage and current measurements available at the substation end, as well as technical information of the distribution grids, including grid topology, load, and line data. This approach analyzes the fault location either in phase domain [4–6] or using symmetrical components [7–9]. However, the impedance-based fault location approach,

frequently leads towards multiple estimations and they are generally based on a hectic iterative process. Additionally, the dynamic behavior of loads, the presence of laterals, inherent properties of distribution grids, measurement noise, and fault resistance and inception angle uncertainties, heavily degrade the performance of these techniques [1–3].

However, the development of advanced measurement infrastructure, communication systems, and the willingness of the electric system decision makers paved the way for a faster and more accurate fault location technique, called the traveling wave technique. The technique is based on the characteristic frequencies of the traveling waves associated with faults in the electrical grids [10], and can be ramified into A, B, and C-type [11]. The A-type and the B-type fault location techniques, are based on online measurements and use single/double ended recorded traveling waves originated by faults, whereas the C-type technique injects traveling waves manually to locate faults in electricity grids [12]. These techniques are very useful for long transmission/distribution lines, but distribution lines are short by nature. Consequently, they cannot effectively locate faults in short length distribution lines. The third category fault location technique namely, the knowledge-based approach, offers auspicious prospects in dealing with distribution grid faults considering their intrinsic complexities. In Reference [13], the artificial neural network (ANN) was employed to locate distribution grid faults based on voltage, current, and active power measurements collected from the substation end. The authors of Reference [14–17], combined the wavelet transform and different machine learning tools to locate faults in distribution grids. However, the wavelet transform (WT) based decompositions, do not uphold the phase information of the original signal and are sometimes sensitive to the presence of measurement noise [18–20].

However, most of the referenced fault location techniques did not consider the presence of measurement noise and uncertainties associated with pre-fault loading condition, fault resistance, and inception angle in locating faults and identifying faulty sections in power distribution grids. Additionally, an advanced signal processing technique (SPT) namely, the Stockwell transform that combines the advantages of wavelet transform and short time Fourier transform was employed recently to detect and classify faults in distribution grids in Reference [21], but not employed in locating faults and identifying faulty sections in distribution grids. Therefore, this research proposes the hybrid fault location and faulty section identification approach, for power distribution grids, combining the Stockwell transform and different machine learning tools. The proposed approach starts with the extraction of useful features from the ST decomposed faulty current signals, collected from different branches of the grids. Then it fetches the extracted characteristic features, as inputs of the different machine learning tools, including the MLP-NN, SVM, and ELM to get decisions on fault location and faulty section. Additionally, the proposed technique employs the CF-PSO to optimize the SVM and ELM key parameters, to achieve better accuracy in locating distribution grid faults. The presented results, demonstrated the independence of the proposed approach in the presence of measurement noise, pre-fault loading condition, fault resistance, and inception angle.

2. Background Theories

The combination of SPT and MLT, received significant attention in diagnosing power systems transients in recent years [22–24], where the researchers used SPT to collect characteristic features related to the electrical transients, and fetched them as inputs to MLT. The following parts of this section, briefly describe the employed SPT and MLT in this research.

2.1. Stockwell Transform

Among many SPT, the Fourier transform (FT), is one of the most widely used techniques to transform time domain signals into the frequency domain to analyze harmonics and design filters. However, the FT loses the temporal information and provides erroneous results for non-stationary signals. In response to this deficiency, Dennis Gabor (1946) employed a small sampling window of the regular interval to map a signal into a two-dimensional function of time and frequency,

and the adaptation is known as short time Fourier transform (STFT) [25]. Unfortunately, the STFT cannot provide better time-frequency resolution simultaneously, where a better frequency resolution leads towards a worse time resolution, and vice versa. The wavelet transform (WT), resolves the resolution problem by employing larger windows at lower frequencies and smaller windows at higher frequencies [26]. However, the WT is sensitive to the presence of measurement noise and does not uphold the phase information of the non-stationary signals. Consequently, Stockwell et al. [18–20] developed the ST combining the benefits of the STFT and WT, that effectively upholds the referenced frequency and phase information. The ST of a signal $x(t)$ can be defined as:

$$S(\tau, f) = S\{x(t)\} = \int_{-\infty}^{\infty} x(t)\, w(\tau - t, f)\, e^{-j2\pi ft} dt \tag{1}$$

where, $w(t, f) = \frac{|f|}{\sqrt{2\pi}} e^{-\frac{t^2 f^2}{2}}$ is the Gaussian window function, f represents the frequency, and t and τ are the time variables.

The ST transforms the time domain signals, into both the frequency and time domains. The coefficients of this transform produce complete resolutions for each designated frequency and cover the whole temporal axis by picking up the possible values of τ where the τ indicates the center of the window function. Different values of f are employed to adjust the sizes of the Gaussian windows, with a view to realizing multiresolution over different frequencies over the temporal axis. However, the ST on a signal provides an $M \times N$ matrix, commonly known as the *S-matrix*, where all the elements are complex numbers. The rows of the matrix pertain to frequency characteristics and the columns pertain to time characteristics. However, the following equation gives the energy matrix of the same signal:

$$E = |S_{M \times N}|^2 \tag{2}$$

The proposed fault location approach, generates three *S-matrices* and three *E-matrices*, from a three-phase current signal through ST. Hence, it creates new vectors namely S_{cmax}, S_{rmax}, and E_{max} from the *S-matrices* and the *E-matrices*. The S_{cmax}-vectors contain the maximum absolute values of the columns of the *S-matrices*, whereas the S_{rmax}-vectors contain the maximum absolute values of the rows of the *S-matrices*, and the E_{cmax}-vectors contain the absolute maximum values of the columns of the energy matrices. The approach also produces another set of vectors called $S_{c\text{-}phase\text{-}max}$, that contain the phase angles of the *S-matrices* associated with the elements of S_{cmax}. Eventually, the proposed approach applies standard statistical techniques on the produced vectors and extracts characteristic features that contain fault signature. Among many features, the amplitude, the gradient of amplitude, mean value, standard deviation, entropy, skewness, kurtosis, time of occurrence, and energy of different harmonics are widely used features for the analysis of power quality transients [21,27–29].

2.2. Multilayer Perceptron Neural Network

Artificial neural networks have become very popular for many engineering applications, due to their parallel computing abilities and adaptiveness to external disturbances. The MLP-NN is one of the widely used ANN and consists of input, hidden, and output layers. The hidden layer uses squashing functions to process the inputs before sending them to the output layer. In order to minimize training error, the supervised learning algorithm optimizes randomly initiated connecting weights and biases [30–32]. In this paper, the total number of hidden neurons for the MLP-NN are chosen through a systematic trial and error approach.

2.3. Support Vector Machine

The SVM was first introduced by Boser et al. [33] with a view to efficient analysis of data, and was primarily restricted to classification problems only. With the passage of time, the SVM was extended to solve regression problems that construct an optimal geometric hyperplane to distinguish the available data, and to map them into the higher dimensional feature space [34]. It forms a separation

surface employing various functions including sigmoidal, polynomial, and radial basis functions. This paper employed a MATLAB/SIMULINK based SVM-KM toolbox [35], to locate distribution grid faults. It considered the mean absolute percentage error (MAPE) of the test dataset as the objective function, while optimizing SVM parameters including regularization coefficients (λ and *C*), tolerance of termination criterion (ε), and kernel option (K_o) employing CF-PSO.

2.4. Extreme Learning Machine

Huang et al. [36] introduced the extreme learning machines for single hidden layer feedforward neural networks, that are a thousand-fold faster over the traditional feedforward neural networks, in attaining generalization performance efficiently. The ELM evaluates the output weights analytically, by picking the input weights randomly [37]. This paper employs an ELM toolbox [38], developed in the MATLAB/SIMULINK platform to locate distribution grid faults. Like the SVM, this research considers the MAPE of the test dataset as the objective function, while optimizing ELM parameters including the regularization coefficient (C_R) and the kernel option (K_p) with the aid of CF-PSO.

2.5. Constriction-Factor Partcile Swarm Optimization

The population-based stochastic optimization approach widely known as the particle swarm optimization (PSO) mimics the social behavior of swarms [39]. Initially, it was considered that the inertia weight of the PSO technique gets decreased monotonously in each generation to update the velocity, hence to update the position of the swarms. This monotonicity sometimes leads towards fall of the quality of the solutions. Consequently, Clerc [40] introduced the idea of constriction factor to update the velocity by ensuring system convergence and effective diversified search. The following steps briefly discuss the CF-PSO technique:

Step 1: Initialization

The CF-PSO employs the following equations to initialize positions and velocities of the swarms randomly from the search space, for a pre-specified number of particles:

$$x_{i,j} = U\left(x_{min,j},\ x_{maxj}\right), \qquad \forall\ i,\ j \tag{3}$$

$$v_{i,j} = U\left(v_{min,j},\ v_{maxj}\right), \qquad \forall\ i,\ j \tag{4}$$

where, *U* represents the uniform distribution; *i* and *j* represent particle size and dimension of the problem; x_{min} and x_{max} refer to the lower and upper boundaries of the solution space; v_{min} and v_{max} represent the lower and upper boundaries of the velocities of the particles.

Step 2: Fitness assessment and updating the best solution

This step assesses the fitness of swarms, based on their randomly initiated positions. Hence, it stores the fitness values of all swarms as the individual best solutions for the first generation. From the second generation, this step compares the new fitness of the swarms with the fitness of the previous generation and updates the individual best solutions accordingly. Besides this, the step picks the global best solution from the individual best solutions in each generation. Finally, the CF-PSO chooses the optimal solution from the global best solutions.

Step 3: Checking the stopping criteria

The CF-PSO checks the stopping criteria after a certain number of generations, to avoid premature convergence. It stops the optimization algorithm if the objective function does not change for a pre-specified number of generations, or it reaches the targeted number of generations.

Step 4: Updating inertia weights, velocities, and positions

This step upgrades the inertia weights and the velocity of each swarm, with the aid of the following equations in each generation. Hence, this step upgrades the positions of the swarms based on the revised velocities.

$$w^t = w_{max} - \frac{w_{max} - w_{min}}{t^{max}} \times t \tag{5}$$

$$v_{j,k}^t = c_f \left(w^t \times v_{j,k}^{t-1} + c_1 r_1^t \times \left[x_{j,k}^{*(t-1)} - x_{j,k}^{(t-1)} \right] + c_2 r_2^t \times \left[x_{j,k}^{**(t-1)} - x_{j,k}^{(t-1)} \right] \right) \tag{6}$$

$$c_f = \frac{2}{\left| 2 - \varphi - \sqrt{(\varphi^2 - 4\varphi)} \right|}; \; c_1 + c_2 = \varphi; \; |c_1| \leq |c_2|; \; \varphi > 4 \tag{7}$$

$$x_{j,k}^t = v_{j,k}^t + x_{j,k}^{(t-1)} \tag{8}$$

where, w^t is the inertia weight of tth generation; $v_{j,k}^t$ is the velocity of jth particle in t^{th} generation; $v_{j,k}^{(t-1)}$ and $x_{j,k}^{(t-1)}$ are the velocity and the position of jth particle in $(t - 1)^{th}$ generation $x_{j,k}^{**(t-1)}$; and $x_{j,k}^{*(t-1)}$ are the global and individual best positions of $(t - 1)^{th}$ generation. w_{max} and w_{min} are the maximum and minimum values of inertia weight; c_f is the constriction factor; r_1 and r_2 are uniformly distributed random numbers in [0 1]; and c_1 and c_2 are known as the cognitive and social parameters, respectively.

Figure 1 shows the complete flowchart of the presented CF-PSO algorithm. However, this paper sets v_{max} and v_{min} to 4 and −4, respectively, and starts checking of the stopping criteria after 30 generations, through a systematic trial and error approach, and based on the experience as described in References [41–43]. Likewise, this paper sets other parameters as $w_{max} = 1.2$, $w_{min} = 0.1$, $4.05 \leq \varphi \leq 4.15$, $2.00 \leq c_1 \leq 2.05$.
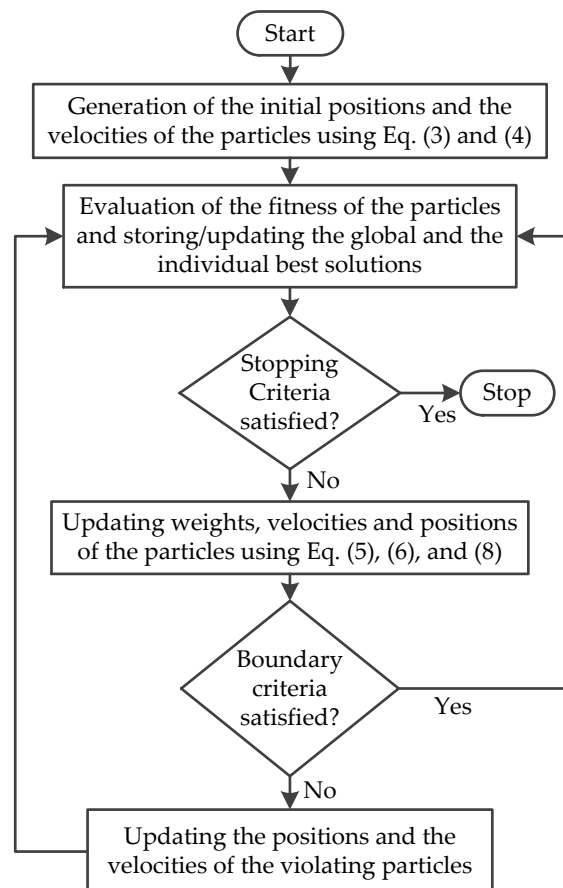

Figure 1. The flowchart of the constriction factor particle swarm optimization.

3. Proposed Fault Location Technique

The proposed fault location technique comprises of three steps namely the data generation, ST based feature extraction and selection, and training and testing of the machine learning tools to locate faults/identify faulty sections in the distribution grids. The following parts of this section describe the mentioned steps:

3.1. Data Generation

The proposed fault location/faulty section identification scheme, applies hundreds of faults on the selected test distribution feeders by varying the pre-fault loading condition, fault resistance, class, and inception angle. For each faulty case, the proposed technique records three-phase current signals from different locations/branches of the distribution feeders, with a sampling frequency of 20 kHz. The technique also keeps track of the fault locations and faulty sections, for future use to train different machine learning tools as the regression and classification problems, respectively. However, the ST extracted features are employed as inputs, and the fault locations and faulty sections are considered as the outputs.

3.2. ST Based Feature Extraction and Selection

To select the useful features, the proposed fault location technique decomposes one pre and one post cycle current signals recorded from an applied phase-A-to-ground (AG) fault, on the four-node test distribution feeder employing the ST. Hence, the developed approach calculates the following statistical features from the produced S_{cmax}, S_{rmax}, E_{max}, and $S_{c\text{-}phase\text{-}max}$ vectors, as discussed in Section II (A):

- F_1 to F_4: Standard deviations of the vectors.
- F_5 to F_8: Maximum values of the vectors.
- F_9 to F_{12}: Minimum values of the vectors.
- F_{13} to F_{16}: Mean values of the vectors.
- F_{17} to F_{20}: Entropies of the vectors.
- F_{21} to F_{24}: Skewness of the vectors.
- F_{25} to F_{28}: Kurtosis of the vectors.

Therefore, the proposed approach collects 28 features for the phase-A-current and repeats the same step for 500 similar faulty cases. Eventually, the technique ends up with a matrix of 500×28 for the phase-A-current. Then it selects twelve useful features, and removes twelve more redundant and insignificant features, based on their correlation factor. In addition, the technique also removes four more features, as they possess constant/zero values for all 500 faulty cases. Consequently, the developed technique selects a total of thirty-six (= 12×3) features, for each three-phase branch currents. Table 1 summarizes the feature selection and removal process.

Table 1. ST Extracted Feature Selection and Removal Process Summary.

Selected Features	Correlation Factor/Comment	Action
F_5 & F_6	1.00	Remove feature F_6
F_9 & F_{13}	1.00	Remove feature F_{13}
F_{21} & F_{24}	1.00	Remove feature F_{24}
F_{25} & F_{28}	1.00	Remove feature F_{28}
F_{21} & F_{25}	0.9999	Remove feature F_{25}
F_5 & F_9	0.9997	Remove feature F_9
F_{11} & F_{15}	0.9996	Remove feature F_{15}
F_5 & F_8	0.9965	Remove feature F_8
F_{22} & F_{26}	0.9958	Remove feature F_{26}
F_7 & F_{11}	0.9957	Remove feature F_7
F_8, F_{12} & F_{16}	1.00	Remove features F_{12} & F_{16}
F_{17}, F_{19} & F_{20}	Most of the cases they are zero	Remove all of them
F_{27}	It is constant for all cases	Remove F_{27}
The selected features: F_1, F_2, F_3, F_4, F_5, F_{10}, F_{11}, F_{14}, F_{18}, F_{21}, F_{22}, and F_{23}		

3.3. Training and Testing of the Machine Learning Tools

As mentioned earlier, the proposed technique records three-phase faulty current signals from different locations/branches of the selected test distribution feeders and extracts thirty-six features employing the ST for each three-phase branch current signal, where the faulty cases are generated by varying the pre-fault loading condition, fault resistance, inception angle, class, and location. Then the technique employs the selected ST extracted features to the machine learning tools as inputs, and the fault locations/faulty sections as outputs. The proposed ST based MLT approach, considers the fault location problem as a regression problem, and the faulty section identification problem as a classification problem. It is worth mentioning, that the proposed approach employs 70% of the generated faulty cases for the training purpose and uses rest of the cases for testing purposes.

The proposed fault location technique (regression problem), evaluates different statistical performance indices, including the mean absolute percentage error (MAPE), root mean squared error (RMSE), percent bias (PBIAS), RMSE-observations to standard deviation ratio (RSR), the coefficient of determination (R^2), Willmott's index of agreement (WIA), and the Nash–Sutcliffe model efficiency coefficient (NSEC) on the test datasets, to investigate the efficacy of the employed machine learning tools. According to Lewis [44], any model is considered highly accurate for forecasting, if the developed model gives a MAPE less than 10%. Additionally, lower values of the RMSE and RSR indicate the strength of the model [45], whereas the ideal value of the PBIAS is zero, which says the developed model can predict desired outputs accurately. Negative and positive values of PBIAS, demonstrate the overestimation and underestimation of the predicted outputs, respectively. Moreover, the values of R^2, WIA, and NSEC vary from 0 to 1. The value 1 refers to a perfect match between the predicted and actual outputs, whereas the value 0 demonstrates that the output cannot be predicted at all from available inputs [45,46]. Like the statistical performance indices, the scatter plot provides a comprehensive summary of a set of bivariate data (two variables) and is often employed to determine the potential associations between the variables. The resulting pattern, demonstrates the type and strength of the relationship between two variables. The more data points located in the neighborhood of the identity line (i.e., $y = x$), the more the data sets agree with each other. If the model output and actual data are the same, then all data points fall on the identity line. Based on the evaluated statistical performance indices of the test datasets, the proposed technique selects the fault location (regression problem) models. Conversely, the faulty section identification technique (classification problem) selects the efficient model based on the overall accuracy. However, after selection of the efficient models, the proposed ST based MLT approach diagnoses the faults based on the recorded three-phase current signals, as illustrated graphically in Figure 2.
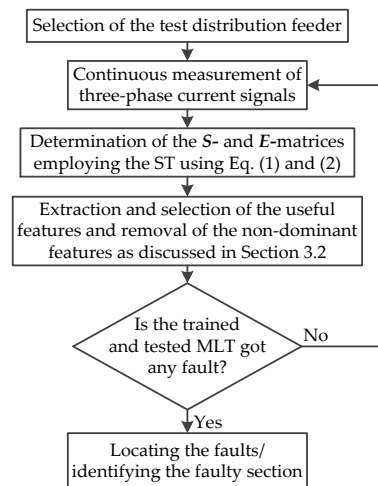
Figure 2. Fault location/faulty section identification flowchart employing the ST based MLT approach.

3.4. Cross Validation

Cross-validation is a popular strategy for assessing how the results of statistical analysis techniques generalize to an independent data set [47]. To select an accurate model, the researchers developed several cross validation techniques, including the *k*-fold cross-validation, jackknife cross-validation, and independent data test [48–53]. This research work, employed the *k*-fold cross validation technique to select the appropriate machine learning predictor. The employed validation technique partitions the available data into *k* segments or folds (roughly equal) and employs *k*-1 segments for training the predictor model, and the remaining segment for validation of the model.

4. Results and Discussions

The proposed fault location/faulty section identification techniques based on the ST and MLT, were tested on two different test distribution feeders. The following parts of this section, briefly introduce the test distribution feeders and provide essential discussions on the simulation results.

4.1. Example 1: Four-Node Test Distribution Feeder

The four-node test distribution feeder consists of two distribution transformers, a distribution line, a lumped load, and a source, as presented in Figure 3, whereas Table 2 presents the technical specifications of the test feeder. This research modeled the test distribution feeder for Example 1 in the MATLAB/SIMULINK environment, considering its unbalanced loading condition.
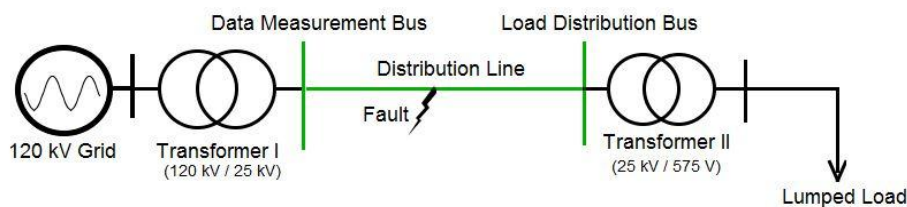

Figure 3. Four-node test distribution feeder used for fault location purpose.

Table 2. Four-node test distribution feeder specifications.

Item	Details
Transformer I	12 MVA, 120/25 kV, Yg-D
Transformer II	12 MVA, 25 kV/575 V, Yg-D
Length of distribution line	30 km
System frequency	60 Hz
Load (Unbalanced)	Phase A (1 MW, 0.48 MVAR) Phase B (2 MW, 0.65 MVAR) Phase C (2.5 MW, 1.2 MVAR)
Load variations	$\pm 10\%$

4.1.1. Fault Location with Un-Optimized MLT

The proposed fault location technique, generated 1200 cases of AG faults by varying the pre-fault loading condition, fault resistance, inception angle, and location. It recorded a three-phase current signal from the data measurement bus for each faulty case and employed the ST to extract the features from the recorded current signals. After selection of the previously mentioned thirty-six features for each faulty case, the proposed fault location technique fetched them as inputs to different machine learning tools, where their outputs were the corresponding fault locations. The proposed fault location scheme chose the SVM and ELM control parameters randomly, as summarized in Table 3 whereas the number of hidden neurons of the MLP-NN were chosen through a systematic trial and error approach. However, the proposed technique employed 70% of the available data for training and the rest of the data for testing purposes.

Table 3. Randomly selected machine learning tools (MLT) control parameters.

MLT	Values of the Control Parameters
MLP-NN	Number of hidden neuron = 6
SVR	$C = 1000$, $\lambda = 10^{-8}$, $\varepsilon = 0.1$, $K_O = 100$ and *kernel* = Gaussian RBF
ELM	$C_R = 10^{10}$, $K_p = 1.5 \times 10^8$ and *kernel* = Gaussian RBF

Figure 4 presents the targeted and the ST based MLT predicted fault locations, for forty randomly selected observations from the test dataset of AG faults, for comparison purpose. It visually appears that the MLP-NN technique adequately predicted the desired outputs in most of the cases. Conversely, the fault location accuracies of the other two techniques (the SVM and ELM) were poor, compared to the MLP-NN approach. Figure 5 presents the scatter plots of the test dataset, which depicts that the data points are closer to the identity line for the MLP-NN approach as expected for an efficient model. On the other hand, the scatter plots of the SVM and ELM models illustrated their inefficiency in locating faults in the four-node test distribution feeder. The dissatisfactory performance of these two models requires further investigation through the optimization of their key parameters, employing the well-known stochastic approaches to achieve better performance. However, like AG faulty cases, the proposed approach collected features for other types of faults including phase-B-to-ground (BG), phase-C-to-ground (CG), phase-A-phase-B-to-ground (ABG), phase-B-to-phase-C-ground (BCG), phase-C-to-phase-A-ground (CAG), and phase-A-to-phase-B-to-phase-C-to-ground (ABCG). Instead of presenting similar figures and discussions, this research summarized the operational times along with the selected statistical performance indices, for the test datasets of the ST based MLT techniques, for all seven types of faults in Table 4.
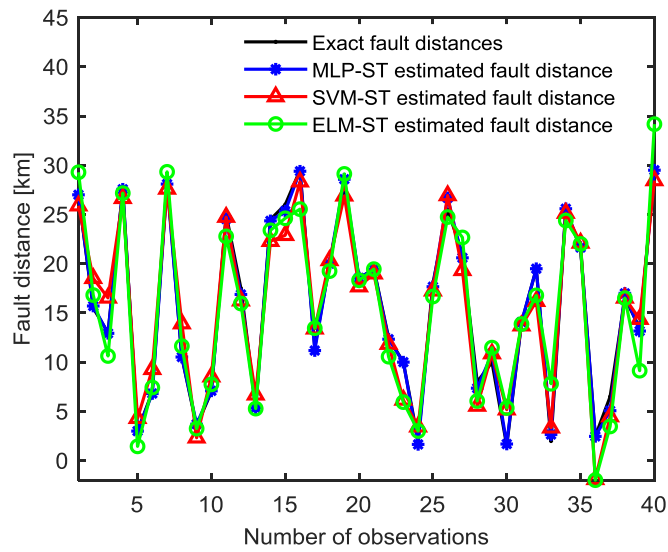
Figure 4. Comparison of the Stockwell transform (ST) based non-optimized MLT predicted and actual fault distances, for the test dataset of applied phase-A-to-ground (AG) fault.
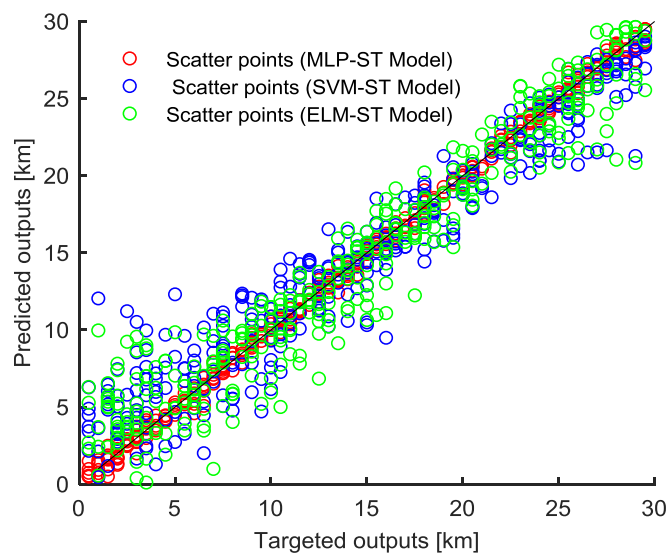

Figure 5. Scatter plot of the targeted and the ST based non-optimized MLT predicted outputs, for the test dataset of AG fault.

Table 4. Operation time and statistical performance indices for test datasets with randomly selected MLT parameters.

Fault Type	Technique	Time		Statistical Performance Measures						
		Training	Testing	RMSE	MAPE	RSR	PBIAS	R^2	WIA	NSCE
AG	ANN	11.5469	0.0156	0.3333	0.0442	0.0378	0.0642	0.9993	0.9996	0.9986
	SVR	28.6094	0.0469	2.6432	0.4612	0.2999	0.0569	0.9576	0.9775	0.9100
	ELM	0.0625	0.0156	2.2721	0.4152	0.2578	−1.238	0.9671	0.9834	0.9335
BG	ANN	11.5000	0.0156	0.4419	0.0816	0.0517	−0.193	0.9987	0.9993	0.9973
	SVR	27.4688	0.0313	2.9087	0.5246	0.3400	1.5522	0.9443	0.9711	0.8844
	ELM	0.0781	0.0625	2.2018	0.4499	0.2574	0.5975	0.9664	0.9834	0.9338
CG	ANN	12.2344	0.0313	0.3327	0.0589	0.0385	−0.151	0.9993	0.9996	0.9985
	SVR	28.7656	0.0469	2.2030	0.4376	0.2550	1.1159	0.9686	0.9837	0.9350
	ELM	0.1250	0.0469	1.6798	0.3493	0.1944	0.4669	0.9810	0.9905	0.9622
ABG	ANN	11.0000	0.0156	0.2627	0.0337	0.0298	0.0849	0.9996	0.9998	0.9991
	SVR	33.4219	0.0156	3.1566	0.6123	0.3579	0.6687	0.9741	0.9680	0.8719
	ELM	0.0938	0.0156	1.8594	0.3139	0.2108	−0.384	0.9776	0.9889	0.9555
BCG	ANN	12.3750	0.0156	1.9296	0.1843	0.2240	1.2046	0.9748	0.9875	0.9498
	SVR	29.6094	0.0156	3.0060	0.5706	0.3489	1.1242	0.9759	0.9696	0.8783
	ELM	0.0469	0.0156	1.8044	0.3047	0.2094	0.0152	0.9779	0.9890	0.9561
CAG	ANN	11.1406	0.0469	0.2019	0.0261	0.0228	−0.030	0.9997	0.9999	0.9995
	SVR	31.6563	0.0156	3.1285	0.6507	0.3533	0.5950	0.9732	0.9688	0.8752
	ELM	0.1250	0.0156	1.9358	0.3451	0.2186	0.9108	0.9759	0.9881	0.9522
ABCG	ANN	12.6406	0.0313	0.3496	0.0327	0.0425	−0.002	0.9991	0.9995	0.9982
	SVR	29.2500	0.0469	3.6239	0.7279	0.4409	1.7376	0.9867	0.9514	0.8056
	ELM	0.0625	0.0313	2.5005	0.4327	0.3042	−0.535	0.9529	0.9769	0.9075

It showed the SVM technique required an almost three-fold time to get trained, compared to the MLP-NN technique, whereas the training process of the ELM technique was a hundred times faster than the MLP-NN approach. The test times for all the approaches were almost equal, and are less than few cycles of a 60 Hz power system network. Additionally, the RMSE, MAPE, and RSR values were quite low, whereas the R^2, WIA, and NSCE were almost unity for the MLP-NN approach; indicating the strength and effectiveness of the MLP-NN approach in locating all types of faults. Furthermore, the PBIAS values were positive for a few cases and negative for the others, which demonstrates underestimations and overestimations of the fault distances, respectively. However, these values were very small and closer to zero; hence, the smaller underestimations and overestimations could be neglected for the ST based MLP-NN approach. On the contrary, the statistical performance indices of the other two approaches (the SVM and ELM), were not satisfactory for all seven types of faults. Consequently, they required further investigation through the optimization of their key parameters, employing the well-known stochastic approaches to achieve better accuracy.

4.1.2. Fault Location with Optimized MLT

The ST based MLP-NN technique, showed satisfactory performance in locating faults in four-node test feeders, but the performance of the other two MLT techniques were dissatisfactory. Consequently, this section optimizes their key parameters employing the constriction factor-based particle swarm optimization. The paper presented the detailed procedures of the CF-PSO, in Section 2.5. The objective function of the CF-PSO was to minimize the MAPE of the actual and the predicted fault locations, for the test datasets. After going through 100 generations with a population size of 20 individuals, the CF-PSO ended up with the optimized SVM control parameters. Similarly, the CF-PSO optimized the control parameters of the ELM in 100 generations, with a population size of 80 individuals. Table 5 summarizes the obtained optimal control parameters of both machine learning approaches. It is worth mentioning, that this paper neither optimized the number of hidden neurons nor the connecting weights and biases of the MLP-NN, due to its better performance with the systematically selected number of neurons.

Table 5. CF-PSO optimized MLT control parameters.

MLT	Values of the Control Parameters
MLP-NN	Number of hidden neuron = 6
SVR	$C = 1000$, $\lambda = 8.5 \times 10^{-3}$, $\varepsilon = 4.75 \times 10^{-1}$, $K_O = 100$ and *kernel* = Gaussian RBF
ELM	$C_R = 5 \times 10^{11}$, $K_p = 3.5 \times 10^9$ and *kernel* = Gaussian RBF

Figure 6 presents the targeted and the ST based MLT predicted fault locations, for forty randomly selected observations from the test dataset of AG faults, where the SVM and ELM techniques used their CF-PSO optimized key parameters, as presented in Table 5. It visually appeared that all the MLT techniques adequately predict the desired outputs in almost every case. Furthermore, Figure 7 presents the scatter plots of the developed MLT for the test dataset, which depicts that the data points are closer to the identity line for the employed techniques. Hence, the presented results confirmed the superiority of the optimized MLT, over the MLT with randomly picked parameters. Table 6 summarizes the operational times, along with the selected statistical performance indices of the test datasets of the ST based optimized MLT approaches, for all seven types of faults. The training and testing times for the selected machine learning tools were like that of Table 4. As can be observed from the table, the RMSE, MAPE, and RSR values were quite low, whereas the R^2, WIA, and NSCE values were almost unity, indicating the strength and effectiveness of the proposed approach in locating faults. However, the PBIAS values were positive for a few cases and negative for the others, which demonstrated underestimations and overestimations of the fault distances, respectively. However, these values were closer to zero; hence, the smaller underestimations and overestimations could be neglected.

Table 6. Operation time and statistical performance indices for test datasets with CF-PSO optimized MLT parameters.

| Fault Type	Technique	Time		Statistical Performance Measures						
		Training	Testing	RMSE	MAPE	RSR	PBIAS	R^2	WIA	NSCE
	ANN	11.5469	0.0156	0.3333	0.0442	0.0378	0.0642	0.9993	0.9996	0.9986
AG	SVR	26.2813	0.0156	0.3646	0.0094	0.0414	0.0304	0.9991	0.9996	0.9983
	ELM	0.1250	0.0469	0.2517	0.0266	0.0286	0.0388	0.9996	0.9998	0.9992
	ANN	12.5000	0.0156	0.4419	0.0816	0.0517	−0.193	0.9987	0.9993	0.9973
BG	SVR	26.3281	0.0156	0.2271	0.0058	0.0265	0.1045	0.9996	0.9998	0.9993
	ELM	0.0781	0.0625	0.3172	0.0335	0.0371	−0.035	0.9993	0.9997	0.9986
	ANN	12.2344	0.0313	0.3327	0.0589	0.0385	−0.151	0.9993	0.9996	0.9985
CG	SVR	29.8438	0.0156	0.2263	0.0030	0.0262	0.2070	0.9997	0.9998	0.9993
	ELM	0.0938	0.0313	0.1668	0.0221	0.0193	−0.001	0.9998	0.9999	0.9996
	ANN	11.0000	0.0156	0.2627	0.0337	0.0298	0.0849	0.9996	0.9998	0.9991
ABG	SVR	28.1250	0.0156	0.4898	0.0814	0.0555	0.0985	0.9992	0.9992	0.9969
	ELM	0.0781	0.0156	0.1899	0.0209	0.0215	−0.012	0.9998	0.9999	0.9995
	ANN	12.3750	0.0156	1.9296	0.1843	0.2240	1.2046	0.9748	0.9875	0.9498
BCG	SVR	30.1875	0.0313	0.4868	0.0799	0.0565	0.1896	0.9991	0.9992	0.9968
	ELM	0.0625	0.0469	0.1638	0.0201	0.0190	0.0309	0.9998	0.9999	0.9996
	ANN	12.1406	0.0469	0.2019	0.0261	0.0228	−0.030	0.9997	0.9999	0.9995
CAG	SVR	26.5156	0.0313	0.4878	0.0874	0.0551	0.0111	0.9991	0.9992	0.9970
	ELM	0.1094	0.0313	0.1714	0.0188	0.0194	−0.012	0.9998	0.9999	0.9996
	ANN	11.6406	0.0313	0.3496	0.0327	0.0425	−0.001	0.9991	0.9995	0.9982
ABCG	SVR	30.7813	0.0469	0.5138	0.0852	0.0625	0.3396	0.9994	0.9990	0.9961
	ELM	0.0625	0.0313	0.2942	0.0255	0.0358	0.2254	0.9994	0.9997	0.9987
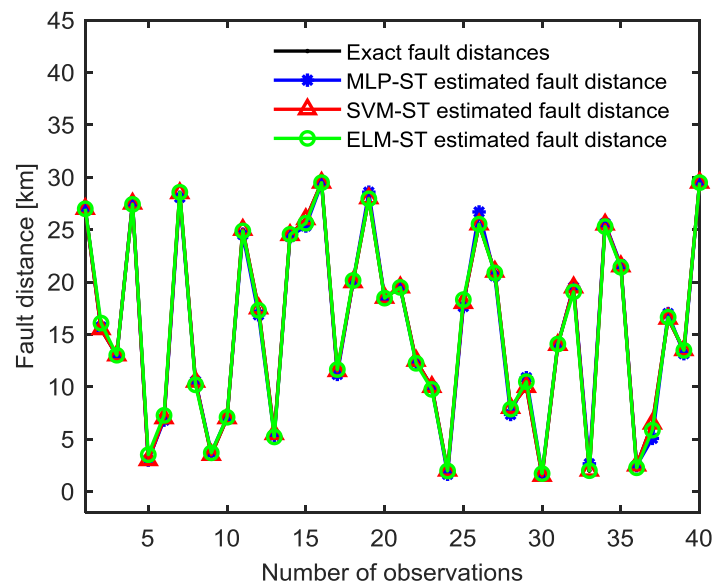
Figure 6. Comparison of the ST based constriction factor-based particle swarm optimization (CF-PSO) optimized MLT predicted and actual fault distances for the test dataset of AG fault.
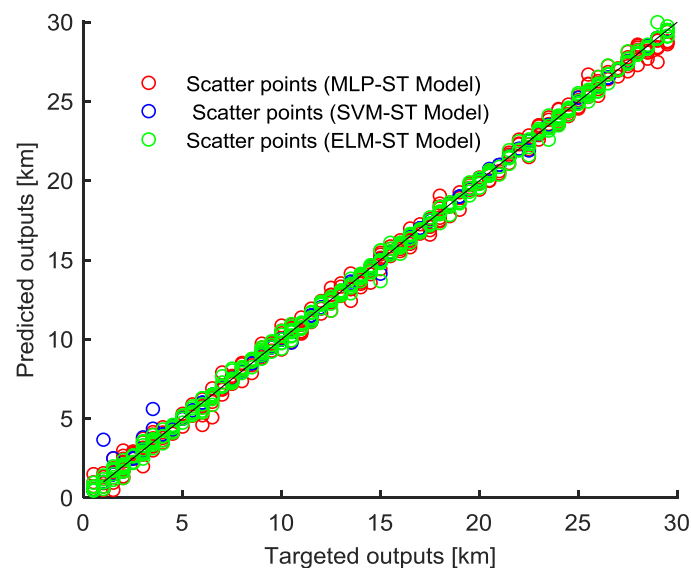

Figure 7. Scatter plot of the targeted and the ST based CF-PSO optimized MLT predicted outputs for the test dataset of AG fault.

4.1.3. Fault Location with Optimized MLT in the Presence of Measurement Noise

The presented results of the previous section, confirmed the superiority of the optimized MLT approaches over their non-optimized counter parts, where this section provides the results of the test datasets in the presence of measurement noise. Figure 8 shows the targeted and the ST based MLT models predicted fault locations, for forty randomly selected observations from the test dataset of AG faults, in the presences of measurement noise (30 dB SNR). It visually appeared that all ST based MLT approaches, estimated the fault distances with similar accuracy. Moreover, the scatter plots of Figure 9 of the developed MLT models for the test dataset, also depicts that the data points were closer to the identity line as expected for any efficient model. Additionally, Figure 10 presents the comparison of the predicted fault distances with the actual fault distances; and Figure 11 presents the scatter plot of the proposed MLT models, for the test dataset of AG faults, in the presence of 20 dB SNR. Figure 8 to

Figure 11 illustrate the efficacy of proposed ST based MLT models in predicting the fault distances efficiently, even with the presence of measurement noise. Furthermore, Tables 7 and 8 present the selected statistical performance indices, along with the required operational time of the proposed ST based MLT approaches, in the presence of 30 dB and 20 dB SNR, respectively. It can be observed from both tables, that the training and testing times, for the selected machine learning tools, are like that of Table 4.
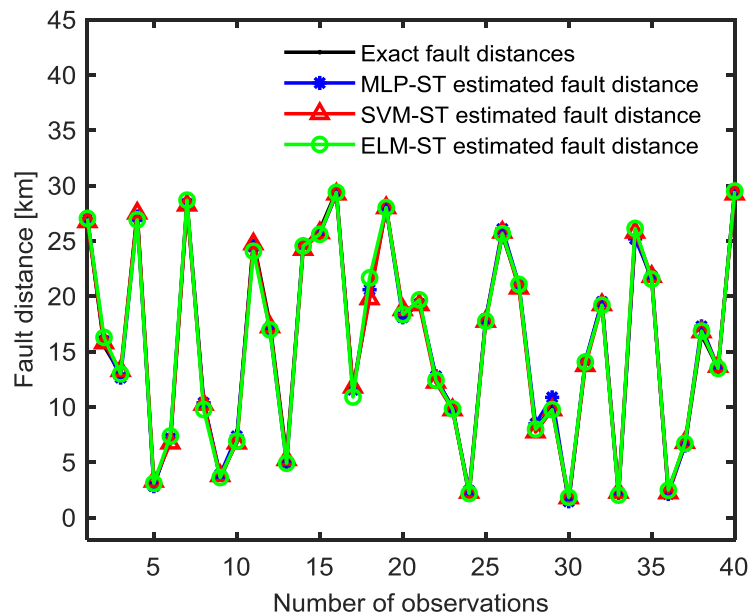

Figure 8. Comparison of the ST based optimized MLT predicted and actual fault distances, for the test dataset, in the presence of 30 dB SNR.
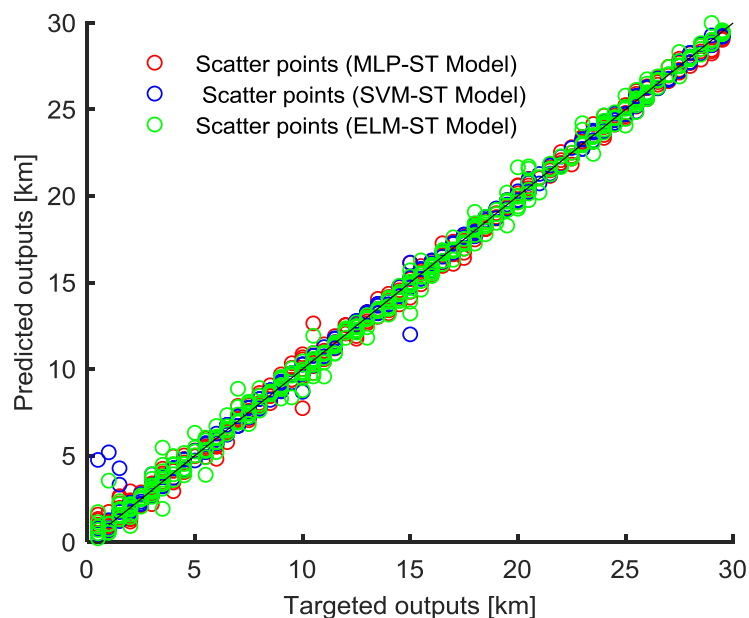

Figure 9. Scatter plot of the targeted and the ST based optimized MLT predicted outputs, for the test dataset, in the presence of 30 dB SNR.
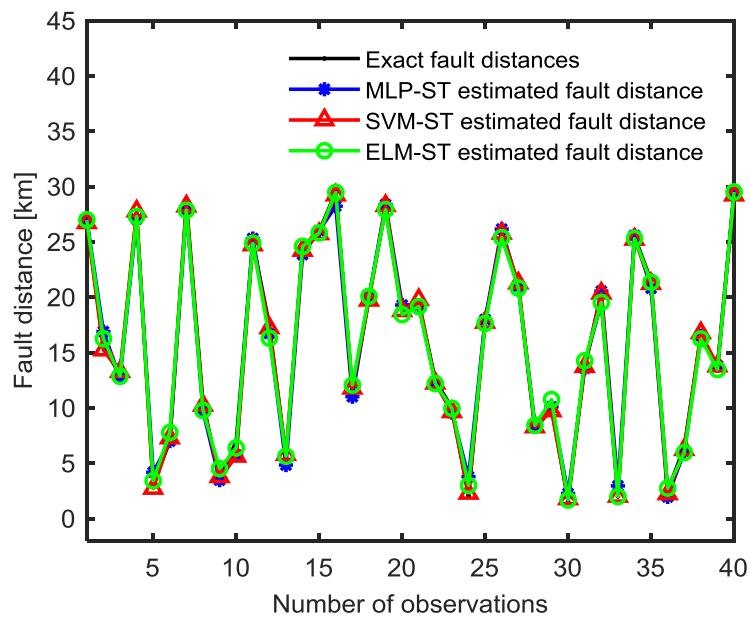
Figure 10. Comparison of the ST based optimized MLT predicted and actual fault distances, for the test dataset, in the presence of 20 dB SNR.

Additionally, the RMSE, MAPE, RSR, and PBIAS were quite low, whereas the R^2, WIA, and NSCE were almost unity for the optimized MLT approaches, indicating the strength and the effectiveness of the approaches, for all types of faults, even in the presence of measurement noise. However, the performance of the developed scheme degrades a bit in the presence of measurement noise, but it can still estimate fault locations with acceptable accuracy.
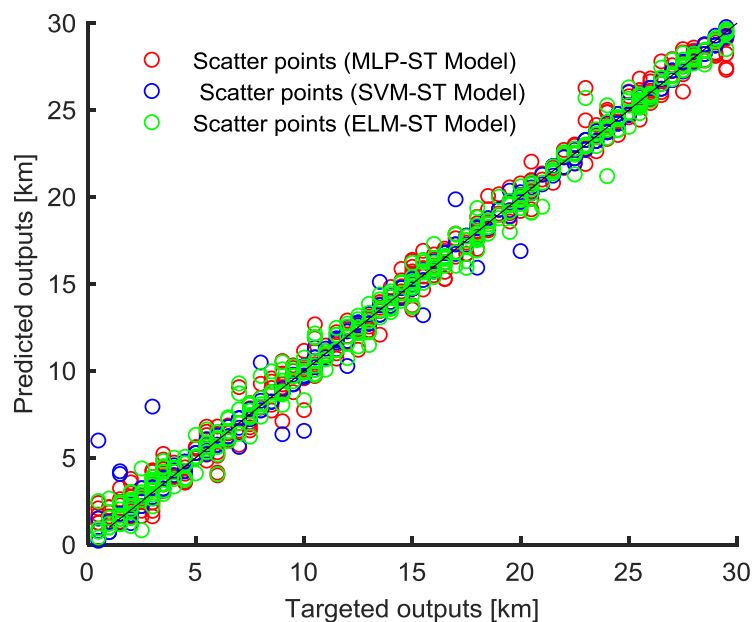

Figure 11. Scatter plot of the targeted and the ST based optimized MLT predicted outputs, for the test dataset, in the presence of 20 dB SNR.

Table 7. Operation time and the statistical performance indices, for the test datasets with CF-PSO optimized MLT, in the presence of 30 dB SNR.

Fault Type	Technique	Time		Statistical Performance Measures						
		Training	Testing	RMSE	MAPE	RSR	PBIAS	R^2	WIA	NSCE
	ANN	12.3906	0.0156	0.4775	0.0761	0.0542	−0.075	0.9986	0.9993	0.9971
AG	SVR	30.4688	0.0313	0.4179	0.0564	0.0474	−0.006	0.9989	0.9994	0.9978
	ELM	0.1250	0.0156	0.4959	0.0606	0.0563	0.1173	0.9984	0.9992	0.9968
	ANN	12.0000	0.0156	0.5551	0.0993	0.0649	−0.443	0.9979	0.9989	0.9958
BG	SVR	26.6406	0.0313	0.2184	0.0042	0.0255	−0.056	0.9997	0.9998	0.9993
	ELM	0.0781	0.0156	0.3641	0.0529	0.0426	−0.243	0.9991	0.9995	0.9982
	ANN	12.2813	0.0313	0.2729	0.0495	0.0316	−0.145	0.9995	0.9998	0.9990
CG	SVR	27.3438	0.0313	0.7008	0.1291	0.0811	0.5864	0.9969	0.9984	0.9934
	ELM	0.0781	0.0625	0.2815	0.0357	0.0326	0.1662	0.9995	0.9997	0.9989
	ANN	12.9531	0.0156	0.2220	0.0264	0.0252	0.0921	0.9997	0.9998	0.9994
ABG	SVR	31.4063	0.0156	0.4898	0.0814	0.0555	0.0758	0.9991	0.9992	0.9969
	ELM	0.0625	0.0156	0.2551	0.0307	0.0289	0.0261	0.9996	0.9998	0.9992
	ANN	11.9375	0.0313	0.1421	0.0229	0.0165	−0.094	0.9999	0.9999	0.9997
BCG	SVR	33.1094	0.0156	0.4876	0.0802	0.0566	0.1567	0.9991	0.9992	0.9968
	ELM	0.0781	0.0156	0.2271	0.0244	0.0264	0.0301	0.9997	0.9998	0.9993
	ANN	12.2813	0.0469	0.1599	0.0265	0.0181	0.0164	0.9998	0.9999	0.9997
CAG	SVR	28.9844	0.0156	0.4880	0.0876	0.0551	0.0313	0.9991	0.9992	0.9970
	ELM	0.0938	0.0313	0.2265	0.0241	0.0256	0.0912	0.9997	0.9998	0.9993
	ANN	10.7344	0.0156	0.1960	0.0252	0.0238	0.0435	0.9997	0.9999	0.9994
ABCG	SVR	29.3906	0.0156	0.5138	0.0852	0.0625	0.3203	0.9993	0.9990	0.9961
	ELM	0.0469	0.0469	0.4437	0.0408	0.0540	0.1201	0.9985	0.9993	0.9971

Table 8. Operation time and the statistical performance indices, for the test datasets with CF-PSO optimized MLT, in the presence of 20 dB SNR.

Fault Type	Technique	Time		Statistical Performance Measures						
		Training	Testing	RMSE	MAPE	RSR	PBIAS	R^2	WIA	NSCE
	ANN	12.7344	0.0313	0.6308	0.1032	0.0716	0.1823	0.9975	0.9987	0.9949
AG	SVR	26.7344	0.0313	0.4518	0.0635	0.0513	0.0368	0.9987	0.9993	0.9974
	ELM	0.1250	0.0313	0.5413	0.0684	0.0614	−0.190	0.9981	0.9991	0.9962
	ANN	11.3438	0.0156	0.7414	0.1215	0.0867	0.0410	0.9962	0.9981	0.9925
BG	SVR	31.5781	0.0313	0.5687	0.0144	0.0665	0.4296	0.9978	0.9989	0.9956
	ELM	0.1094	0.0156	0.6717	0.0775	0.0785	1.1419	0.9971	0.9985	0.9938
	ANN	10.6563	0.0313	0.4570	0.0625	0.0529	−0.129	0.9986	0.9993	0.9972
CG	SVR	33.8906	0.0313	0.7456	0.1352	0.0863	0.1454	0.9964	0.9981	0.9926
	ELM	0.0781	0.0313	0.3366	0.0425	0.0390	−0.118	0.9992	0.9996	0.9985
	ANN	11.6250	0.0313	0.3126	0.0357	0.0354	0.0672	0.9994	0.9997	0.9987
ABG	SVR	32.5938	0.0156	0.4989	0.0823	0.0566	0.1467	0.9991	0.9992	0.9968
	ELM	0.0625	0.0156	0.3678	0.0360	0.0417	−0.032	0.9991	0.9996	0.9983
	ANN	11.4375	0.0313	0.4005	0.0778	0.0465	0.0445	0.9989	0.9995	0.9978
BCG	SVR	28.4531	0.1920	0.4934	0.0801	0.0573	0.1787	0.9990	0.9992	0.9967
	ELM	0.0781	0.0313	0.3747	0.0346	0.0435	0.0331	0.9991	0.9995	0.9981
	ANN	11.8594	0.0156	0.2753	0.0407	0.0311	0.0033	0.9995	0.9998	0.9990
CAG	SVR	27.5938	0.4898	0.4972	0.0890	0.0562	−0.146	0.9990	0.9992	0.9968
	ELM	0.1094	0.1920	0.3388	0.0392	0.0383	0.1323	0.9993	0.9996	0.9985
	ANN	10.3281	0.0313	0.4188	0.0453	0.0509	0.1053	0.9987	0.9994	0.9974
ABCG	SVR	29.1250	0.0156	0.5161	0.0854	0.0628	0.3574	0.9994	0.9990	0.9961
	ELM	0.0938	0.4898	0.6710	0.0504	0.0816	0.2233	0.9967	0.9983	0.9933

4.1.4. Validating the Developed Fault Location Technique

In order to validate the developed ST based fault location scheme for the four-node test distribution feeder, the following faults were applied on the distribution line and predicted the fault location employing the trained and tested machine learning tools. It is worth mentioning, the proposed approach generated the faulty cases by varying the pre-fault loading condition, presence of measurement noise, and fault resistance and inception angle. Then it employed the ST to extract the useful features and fetched them to the trained machine learning tools as inputs, to

get a decision on the applied faults. As can be seen from Table 9, the developed signal processing based MLT approach located all applied faults with satisfactory accuracy (<1%), which validated the effectiveness of the employed approach.

Table 9. Fault location results obtained employing the ST based machine learning tools, for the four-node test distribution feeder.

Applied Fault Information		Employed Technique	Predicted Fault Information	
Type	Location (km)		Location (km)	Accuracy (%)
AG	9.0	ANN	8.9999	0.0003
		SVR	9.0071	0.0237
		ELM	8.9961	0.0130
BG	27.5	ANN	27.3093	0.6357
		SVR	27.4908	0.0307
		ELM	27.5084	0.0280
CG	17.5	ANN	17.5247	0.0823
		SVR	17.2249	0.09170
		ELM	17.5371	0.1237
ABG	3.5	ANN	3.5169	0.0563
		SVR	3.5071	0.0237
		ELM	3.5000	0.0000
BCG	21.5	ANN	21.5281	0.0937
		SVR	21.5071	0.0237
		ELM	21.5001	0.0003
CAG	10.0	ANN	10.0691	0.2303
		SVR	10.0071	0.0237
		ELM	10.0002	0.0007
ABCG	24.0	ANN	23.7609	0.7970
		SVR	23.7249	0.9170
		ELM	23.9999	0.0003

4.2. Example 2: Thirteen-Node Test Distribution Feeder

The proposed ST based machine learning approach was also employed to identify the faulty sections of the IEEE 13-node test distribution feeder. The selected test distribution feeder operates at 4.16 kV and displays most of the characteristic features of the power distribution grids, as shown in Figure 12 [54]. Consequently, it is usually used to test common features of the distribution system analysis software. The highly loaded test feeder consists of a single voltage regulator at the substation, an in-line transformer, shunt capacitor banks, overhead distribution lines, and underground cables of various configurations, and several unbalanced spot and distributed loads. Moreover, the test feeder contained single-phase, double-phase, and three-phase laterals. This research modeled the test feeder in an RSCAD environment and implemented the developed model in a Real Time Digital Simulator (RTDS) machine. The proposed ST based faulty section identification approach generated 900 faulty cases by varying the pre-fault loading condition, fault resistance, inception angle, type, and location, through employing the '*batch-mode*' operation option of the RSCAD software. After generation of different fault scenarios, the proposed approach recorded three-phase current signals from eight different branches/locations of the test feeder, as indicated in Figure 12, with a sampling frequency of 20 kHz. Then it processed the recorded signals, employing ST in the MATLAB/SIMULINK environment and extracted useful features, as discussed in Section 3.2. It is worth mentioning, that the proposed approach collected 288 features (~8 branches × 36 features per branches), for each fault scenario. Furthermore, this research divided the test distribution feeder into nine sections, as indicated by the rectangular boxes in Figure 12. Consequently, this research prepared the faulty section identification problem as a classification problem.
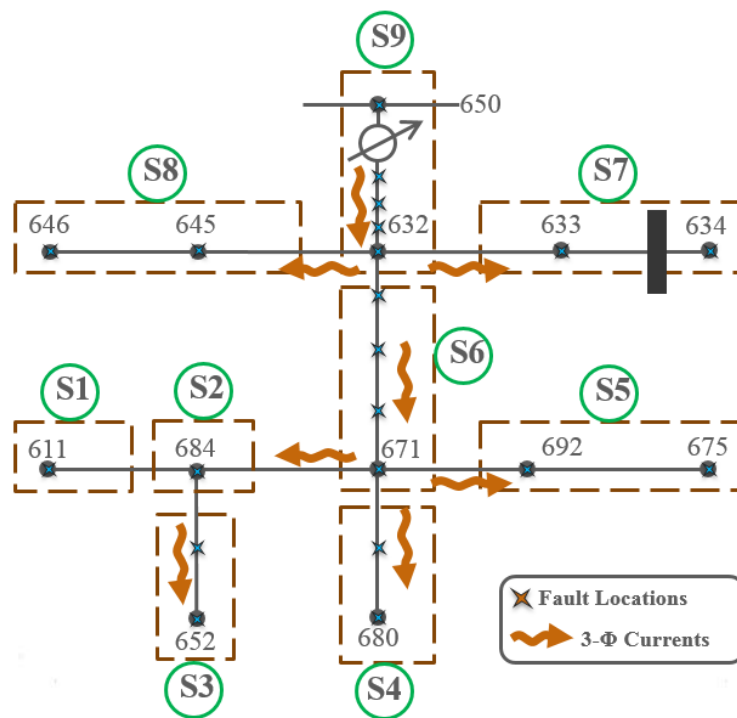
Figure 12. IEEE thirteen-node test distribution feeder.

4.2.1. Faulty Section Identification in Noise-Free Environment

After collection of useful features, the proposed approach trained and tested the MLP-NN with a different number of neurons, where the best neural network consisted of eleven hidden neurons, in terms of minimum mean squared error and overall accuracy. In addition, the proposed approach systematically chose the tan-sigmoid as a squashing function and the resilient backpropagation technique as the training algorithm. Like the fault location technique of the four-node test distribution feeder, the faulty section identification approach also employed 70% of the available data for training purposes and the rest of the data for testing purposes. Table 10 summarizes the technical features related to the MLP neural networks employed for faulty section identification purposes. Table 11 presents a 9×9 confusion matrix obtained from the MLP neural networks, as they identified nine faulty sections employing the proposed approach in a noise-free environment. The diagonal and off-diagonal elements of the confusion matrix, represent the successful and unsuccessful identification of faulty sections, respectively. As can be observed from the table, the proposed technique identified the faulty sections with an accuracy of almost 100%, which demonstrated the successful implementation of the adopted ST based machine learning approach.

Table 10. Multilayer perceptron neural networks (MLP-NN) parameters for the faulty section identification problem.

Item	Specifications
Number of hidden neurons	11
Squashing function	Tan-sigmoid
Training algorithm	Resilient backpropagation
Objective function	Mean squared error (MSE)
Training data	70%
Testing data	30%

Table 11. Faulty section identification results for noise-free measurement.

Symbol	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
S_1	900	0	0	0	0	0	0	0	0
S_2	0	900	0	0	0	0	0	0	0
S_3	0	1	899	0	0	0	0	0	0
S_4	0	0	0	900	0	0	0	0	0
S_5	0	0	0	0	900	0	0	0	0
S_6	0	0	0	0	0	900	0	0	0
S_7	0	0	0	0	0	0	899	1	0
S_8	0	0	0	0	0	0	0	900	0
S_9	0	0	0	0	0	0	0	0	900
Overall accuracy = 99.975%									

4.2.2. Faulty Section Identification in the Presence of Measurement Noise

The proposed technique added a different level of additive white gaussian noise to the recorded three-phase branch current signals, with a view to testing the efficacy of the proposed faulty section identification approach, in the presence of measurement noise. Then it extracted the same features from the noisy signals, and employed them as inputs to the MLP-NN. Table 12 summarizes the faulty section identification results of the ST based MLP-NN approach, in the presence of measurement noise. As can be seen, the ST based approach ended up with the overall accuracies of 99.876%, 99.741%, and 99.235% for the 40 dB, 30 dB, and 20 dB SNR, respectively. Consequently, the obtained results validated the efficacy of the proposed ST based faulty section identification technique, for both noise-free and noisy conditions. In addition, the developed approach confirmed its independence on the pre-fault loading condition, fault type, resistance, and inception angle.

Table 12. Faulty section identification results under noisy measurement.

Faulty Section	Samples Classified Successfully		
	40 dB SNR	**30 dB SNR**	**20 dB SNR**
S_1	900	897	895
S_2	899	896	893
S_3	898	899	894
S_4	897	896	891
S_5	900	900	896
S_6	900	895	892
S_7	898	899	895
S_8	900	900	893
S_9	898	895	889
Overall accuracy (%)	**99.876**	**99.741**	**99.235**

4.2.3. Validating the Developed Faulty Section Identification Technique

To validate the developed ST based faulty section identification technique for the thirteen-node test distribution feeder, the following faults were applied on different locations of the feeder and predicted the faulty section employing the trained and tested MLP-NN. It is worth mentioning, the proposed approach generated the faulty cases by varying the pre-fault loading condition, presence of measurement noise, fault resistance, and inception angle. Then it employed the ST to extract the useful features and fetched them to the trained MLP-NN as inputs, to get a decision on the applied faults. As can be seen from Table 13, the developed signal processing based MLP-NN approach identified all applied faults accurately, which validated the effectiveness of the employed approach.

Table 13. Faulty section identification results obtained employing the ST based machine learning tools, for the IEEE thirteen-node test distribution feeder.

Applied Fault Information			ST Based MLP-NN Predicted Faulty Section
Type	Node	Faulty Section	
AG	652	3	3
BG	671	6	6
CG	611	1	1
ABG	692	5	5
BCG	646	8	8
CAG	680	4	4
ABCG	650	9	9

5. Conclusions and Future Scope

This paper presented advanced signal processing-based machine learning tools to locate faults and identify faulty sections in distribution grids. The proposed approach decomposed three-phase current signals recorded from different branches/locations of the distribution grids, employing the ST to extract the useful characteristic features of faulty cases. Then it employed the extracted features as inputs to different machine learning tools and trained them to locate the faults, and to identify the faulty section in the distribution grids. Additionally, this research employed the CF-PSO technique to optimize the parameters of the support vector and extreme learning machines; whilst it optimized the MLP-NN through a systematic approach to achieve better generalization performance. The presented results validated the efficacy of the developed fault diagnosis approach. Additionally, the optimized machine learning tools outperformed the non-optimized tools in locating distribution grid faults, in terms of the selected statistical performance indices, i.e., the RMSE, MAPE, PBIAS, RSR, R^2, WIA, and NSEC. Moreover, the ST based ELM approach required lesser training time compared to the SVM and MLP-NN techniques. However, the trained MLT diagnosed the applied faults in a fraction of a second, signaling the real-time implementation of the developed fault diagnosis scheme. Furthermore, the developed faulty section identification approach, identified faulty sections with almost 100% accuracy. Finally, the presented results confirmed the independency of the proposed ST based MLT approach, in the fault inception angle and resistance, pre-fault loading conditions, and presence of measurement noise. As a future extension, the signal processing based MLT approach can be applied to diagnose faults in distribution grids, with the incorporation of renewable energy resources, considering their associated uncertainties. The proposed approach can also be extended to diagnose simultaneous faults at distribution level.

Author Contributions: All authors contributed to this work through collaboration. M.S. is the first author in this manuscript. All authors revised and approved for the publication.

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

References

1. Bahmanyar, A.; Jamali, S.; Estebsari, A.; Bompard, E. A comparison framework for distribution system outage and fault location methods. *Electr. Power Syst. Res.* **2017**, *145*, 19–34. [CrossRef]
2. Shafiullah, M.; Abido, M.A. A Review on Distribution Grid Fault Location Techniques. *Electr. Power Components Syst.* **2017**, *45*, 807–824. [CrossRef]
3. Awalin, L.; Mokhlis, H.; Bakar, A. Recent developments in fault location methods for distribution networks. *Prz. Elektrotech.* **2012**, *88*, 206–212.

4. Salim, R.; Resener, M.; Filomena, A.D.; De Oliveira, K.R.C.; Bretas, A.S. Extended fault-location formulation for power distribution systems. *IEEE Trans. Power Deliv.* **2009**, *24*, 508–516. [CrossRef]

5. Filomena, A.D.; Resener, M.; Salim, R.H.; Bretas, A.S. Fault location for underground distribution feeders: An extended impedance-based formulation with capacitive current compensation. *Int. J. Electr. Power Energy Syst.* **2009**, *31*, 489–496. [CrossRef]

6. Salim, R.H.; Salim, K.C.O.; Bretas, A.S. Further improvements on impedance-based fault location for power distribution systems. *IET Gener. Transm. Distrib.* **2011**, *5*, 467. [CrossRef]

7. Girgis, A.A.; Fallon, C.M.; Lubkeman, D.L. A fault location technique for rural distribution feeders. *IEEE Trans. Ind. Appl.* **1993**, *29*, 1170–1175. [CrossRef]

8. Dashti, R.; Sadeh, J. Accuracy improvement of impedance-based fault location method for power distribution network using distributed-parameter line model. *Int. Trans. Electr. Energy Syst.* **2014**, *24*, 318–334. [CrossRef]

9. Das, S.; Karnik, N.; Santoso, S. Distribution Fault-Locating Algorithms Using Current Only. *IEEE Trans. Power Deliv.* **2012**, *27*, 1144–1153. [CrossRef]

10. Tang, J.; Yin, X.; Zhang, Z. Traveling-Wave-Based Fault Location in Electrical Distribution Systems With Digital Simulations. *Telecommun. Comput. Electron. Control* **2014**, *12*, 297.

11. Yan, F.; Liu, W.; Tian, L. Fault location for 10kV distribution line based on traveling wave-ANN theory. In Proceedings of the 2011 IEEE Power Engineering and Automation Conference, Wuhan, China, 8–9 September 2011.

12. Sheng-nan, Y.; Yi-han, Y.; Hai, B. Study on fault location in distribution network based on C-type traveling-wave scheme. *Relay* **2007**, *35*, 1–5.

13. Dehghani, F.; Nezami, H. A new fault location technique on radial distribution systems using artificial neural network. In Proceedings of the 22th International Conference and Exhibition on Electricity Distribution (CIRED), Stockholm, Sweden, 10–13 June 2013.

14. Pourahmadi-Nakhli, M.; Safavi, A.A. Path Characteristic Frequency-Based Fault Locating in Radial Distribution Systems Using Wavelets and Neural Networks. *IEEE Trans. Power Deliv.* **2011**, *26*, 772–781. [CrossRef]

15. Rafinia, A.; Moshtagh, J. A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ANN and FLS. *Int. J. Electr. Power Energy Syst.* **2014**, *55*, 261–274. [CrossRef]

16. Ray, P.; Mishra, D. Artificial Intelligence Based Fault Location in a Distribution System. In Proceedings of the 2014 International Conference on Information Technology, Bhubaneswar, India, 22–24 December 2014.

17. Shafiullah, M.; Ijaz, M.; Abido, M.A.; Al-Hamouz, Z. Optimized support vector machine & wavelet transform for distribution grid fault location. In Proceedings of the 2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), Cadiz, Spain, 4–6 April 2017.

18. Stockwell, R.G.; Mansinha, L.; Lowe, R.P. Localization of the complex spectrum: The S transform. *IEEE Trans. Signal Process.* **1996**, *44*, 998–1001. [CrossRef]

19. Mansinha, L.; Stockwell, R.G.; Lowe, R.P. Pattern analysis with two-dimensional spectral localisation: Applications of two-dimensional S transforms. *Phys. A Stat. Mech. Its Appl.* **1997**, *239*, 286–295. [CrossRef]

20. Wang, Y. Efficient stockwell transform with applications to image processing. Ph.D. Thesis, University of Waterloo, Waterloo, ON, Canada, 20 May 2011.

21. Shafiullah, M.; Abido, M.A. S-Transform Based FFNN Approach for Distribution Grids Fault Detection and Classification. *IEEE Access* **2018**, *6*, 8080–8088. [CrossRef]

22. Khokhar, S.; Mohd Zin, A.A.B.; Mokhtar, A.S.B.; Pesaran, M. A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances. *Renew. Sustain. Energy Rev.* **2015**, *51*, 1650–1663. [CrossRef]

23. Ijaz, M.; Shafiullah, M.; Abido, M.A. Classification of power quality disturbances using Wavelet Transform and Optimized ANN. In Proceedings of the 2015 18th International Conference on Intelligent System Application to Power Systems (ISAP), Porto, Portugal, 11–16 September 2015.

24. Adewole, A.C.; Tzoneva, R.; Behardien, S. Distribution network fault section identification and fault location using wavelet entropy and neural networks. *Appl. Soft Comput.* **2016**, *46*, 296–306. [CrossRef]

25. Sejdić, E.; Djurović, I.; Jiang, J. Time–frequency feature representation using energy concentration: An overview of recent advances. *Digit. Signal Process.* **2009**, *19*, 153–183. [CrossRef]

26. Shafiullah, M.; Abido, M.A.; Al-Hamouz, Z. Wavelet-based extreme learning machine for distribution grid fault location. *IET Gener. Transm. Distrib.* **2017**, *11*, 4256–4263. [CrossRef]

27. Mishra, S.; Bhende, C.N.; Panigrahi, B.K. Detection and Classification of Power Quality Disturbances Using S-Transform and Probabilistic Neural Network. *IEEE Trans. Power Deliv.* **2008**, *23*, 280–287. [CrossRef]

28. Roy, N.; Bhattacharya, K. Detection, Classification, and Estimation of Fault Location on an Overhead Transmission Line Using S-transform and Neural Network. *Electr. Power Components Syst.* **2015**, *43*, 461–472. [CrossRef]

29. Krishnanand, K.R.; Dash, P.K.; Naeem, M.H. Detection, classification, and location of faults in power transmission lines. *Int. J. Electr. Power Energy Syst.* **2015**, *67*, 76–86. [CrossRef]

30. Masiur Rahman, S.; Khondaker, A.N.; Imtiaz Hossain, M.; Shafiullah, M.; Hasan, M.A. Neurogenetic modeling of energy demand in the United Arab Emirates, Saudi Arabia, and Qatar. *Environ. Prog. Sustain. Energy* **2017**, *36*, 1208–1216. [CrossRef]

31. Ansari, H.R.; Zarei, M.J.; Sabbaghi, S.; Keshavarz, P. A new comprehensive model for relative viscosity of various nanofluids using feed-forward back-propagation MLP neural networks. *Int. Commun. Heat Mass Transf.* **2018**, *91*, 158–164. [CrossRef]

32. Rana, M.J.; Shahriar, M.S.; Shafiullah, M. Levenberg–Marquardt neural network to estimate UPFC-coordinated PSS parameters to enhance power system stability. *Neural Comput. Appl.* **2017**, 1–12. [CrossRef]

33. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (COLT), Pittsburgh, PA, USA, 27–29 July 1992.

34. Vapnik, V.; Golowich, S.E.; Smola, A.J. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. *Adv. Neural Inf. Process. Syst.* **1997**, *9*, 281–287.

35. Canu, S.; Grandvalet, Y.; Guigue, V.; Rakotomamonjy, A. SVM and kernel methods matlab toolbox. *Percept. Systmes Inf.* **2003**, *69*, 70.

36. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: a new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2004.

37. Cambria, E.; Huang, G.B.; Kasun, L.L.C.; Zhou, H.; Vong, C.M.; Lin, J.; Yin, J.; Cai, Z.; Liu, Q.; Li, K.; et al. Extreme Learning Machines [Trends & Controversies]. *IEEE Intell. Syst.* **2013**, *28*, 30–59.

38. MATLAB Codes of ELM Algorithm. Available online: http://www.ntu.edu.sg/home/egbhuang/elm_kernel.html (accessed on 29 Feburary 2016).

39. Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. Proceedings of Sixth International Symposium on Micro Machine and Human Science (MHS), Nagoya, Japan, 4–6 October 1995.

40. Clerc, M. The swarm and the queen: towards a deterministic and adaptive particle swarm optimization. In Proceedings of the 1999 Congress on Evolutionary Computation-CEC99, Washington, DC, USA, 6–9 July 1999.

41. Malekpour, A.; Seifi, A. Application of constriction factor particle swarm optimization to optimum load shedding in power system. *Mod. Appl. Sci.* **2010**, *4*, 188–196. [CrossRef]

42. Patwardhan, R.P.; Mhetre, S.L. Effect of constriction factor on minimization of transmission power loss using Particle Swarm Optimization. In Proceedings of the 2015 International Conference on Energy Systems and Applications, Pune, India, 30 October–1 November 2015.

43. Alam, M.N.; Kumar, K.; Mathur, A. Economic load dispatch considering valve-point effects using time varying constriction factor based particle swarm optimization. In Proceedings of the 2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON), Allahabad, India, 4–6 December 2015.

44. Lewis, C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*; Butterworth–Heinemann: Oxford, London, 1982; ISBN 0408005599.

45. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]

46. Willmott, C.J.; Robeson, S.M.; Matsuura, K. A refined index of model performance. *Int. J. Climatol.* **2012**, *32*, 2088–2094. [CrossRef]

47. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. [CrossRef]

48.   Manavalan, B.; Shin, T.H.; Lee, G. PVP-SVM: Sequence-Based Prediction of Phage Virion Proteins Using a Support Vector Machine. *Front. Microbiol.* **2018**, *9*, 476. [CrossRef] [PubMed]

49.   Dao, F.-Y.; Yang, H.; Su, Z.-D.; Yang, W.; Wu, Y.; Hui, D.; Chen, W.; Tang, H.; Lin, H. Recent Advances in Conotoxin Classification by Using Machine Learning Methods. *Molecules* **2017**, *22*, 1057. [CrossRef] [PubMed]

50.   Manavalan, B.; Shin, T.H.; Kim, M.O.; Lee, G. AIPpred: Sequence-Based Prediction of Anti-inflammatory Peptides Using Random Forest. *Front. Pharmacol.* **2018**, *9*. [CrossRef] [PubMed]

51.   Lin, H.; Liang, Z.-Y.; Tang, H.; Chen, W. Identifying sigma70 promoters with novel pseudo nucleotide composition. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *1*, 1–6. [CrossRef] [PubMed]

52.   Manavalan, B.; Subramaniyam, S.; Shin, T.H.; Kim, M.O.; Lee, G. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. *J. Proteome Res.* **2018**, *17*, 2715–2726. [CrossRef] [PubMed]

53.   Manavalan, B.; Shin, T.H.; Lee, G. DHSpred: support-vector-machine-based human DNase I hypersensitive sites prediction using the optimal features selected by random forest. *Oncotarget* **2018**, *9*, 1944–1956. [CrossRef] [PubMed]

54.   Distribution Test Feeders. Available online: https://ewh.ieee.org/soc/pes/dsacom/testfeeders/ (accessed on 6 January 2018).