Measurement 177 (2021) 109330

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

# Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems

Soufiane Belagoune [a,*], Noureddine Bali [a], Azzeddine Bakdi [b], Bousaadia Baadji [c], Karim Atif [a]

[a] Electrical and Industrial Systems Laboratory (LSEI), Department of Electrical Engineering, Faculty of Electronics and Computer Science, University of Sciences and Technology of Houari Boumediene (USTHB), BP.32 El-Alia, Bab Ezzouar, 16111, Algiers, Algeria
[b] Department of Mathematics, University of Oslo, 0851 Oslo, Norway
[c] Signals and Systems Laboratory (SSL), Department of Power and Control, Institute of Electrical and Electronics Engineering, IGEE, (ex: INELEC), University of M'hamed Bougara, Boumerdes, 35000, Boumerdes, Algeria

ARTICLE INFO

*Keywords:*
Multi-machine power system
Power transmission lines
Short-circuit fault
Long short-term memory
Fault detection and isolation
Sequential deep learning

ABSTRACT

Fault detection, diagnosis, identification and location are crucial to improve the sensitivity and reliability of system protection. This maintains power systems continuous proper operation; however, it is challenging in large-scale multi-machine power systems. This paper introduces three novel Deep Learning (DL) classification and regression models based on Deep Recurrent Neural Networks (DRNN) for Fault Region Identification (FRI), Fault Type Classification (FTC), and Fault Location Prediction (FLP). These novel models explore full transient data from pre- and post-fault cycles to make reliable decisions; whereas current and voltage signals are measured through Phasor Measurement Units (PMUs) at different terminals and used as input features to the DRNN models. Sequential Deep Learning (SDL) is employed herein through Long Short-Term Memory (LSTM) to model spatiotemporal sequences of high-dimensional multivariate features to achieve accurate classification and prediction results. The proposed algorithms were tested in a Two-Area Four-Machine Power System. Training and testing data are collected during transmission lines faults of different types introduced at various locations in different regions. The presented algorithms achieved superior detection, classification and location performance with high accuracy and robustness compared to contemporary techniques.

## 1. Introduction

Modern power systems exhibit very fast increase in both size with the integration of renewable energy resources and complexity of generation, transmission and distribution parts to meet the increasing demand of energy [1] and guarantee continuous delivering of energy. However, faults occur frequently and they are inevitable due to various random causes; this severely impacts the performance of power systems, interrupts the delivery of energy [2], and degrades the system health as well as efficiency and reliability. Therefore, the main goal today is the successful earliest detection of these faults to eliminate them rapidly which enables maintenance of faulted element; it also reduces the downtime via recovering the system main function in a minimum time and keeps the components reliabilities closer to initial ones. Efficient protection and maintenance schemes are required to meet this goal of mitigating these faults and restoring proper system operation [3]. One kind of these faults is electrical short-circuit faults that may occur in generation, transmission, or distribution systems; particularly, faults occur at generators, transformers, insulations, HVDC converter, feeder bus, underground cables and transmission lines. Electrical short-circuit faults have negative effects on power system operation and they cause the loss of its main function [2]. It was mentioned in [4] that most of the faults in power system transmission part occur on transmission lines. Short-circuit faults are common in transmission lines and considered as the worst type of faults which cause high risks on the lines such as reducing components remaining useful lifetime, increasing the power loss and the heat at cables, and damage the insulators [5]. Various types of short-circuit faults are encountered during daily operation and mainly classified into symmetrical and unsymmetrical faults. The symmetrical or balanced faults, which keep the system balanced, consist of triple line

* Corresponding author. Tel.: +213 699 606 826.
*E-mail addresses:* belaggounesofiane@gmail.com, sbelagoune@usthb.dz (S. Belagoune), bali_nordine@yahoo.fr (N. Bali), bkdaznsun@gmail.com (A. Bakdi), b_baadji@yahoo.com, lss@univ-boumerdes.dz (B. Baadji), atif_karim2001@yahoo.fr (K. Atif).

https://doi.org/10.1016/j.measurement.2021.109330
Received 18 December 2020; Received in revised form 18 March 2021; Accepted 21 March 2021
Available online 28 March 2021

to ground (LLLG), and triple line (LLL) faults, and have a small occurrence probability of but they are the most severe type of short-circuit faults due to their big effect and damage to system equipment. The asymmetrical or unbalanced faults that make the power system unbalanced during the fault consist of double line to ground (LLG), line to ground (LG), and line-to-line (LL) faults. Even though they are less severe compared to the balanced ones, they have a greater occurrence probability of 0.80 due to single line to ground faults [6].

The fault impact on the system depends on its type, location, and duration thereby faster and reliable detection, identification, and localization of faults greatly improve the protection and maintenance strategies of the system [3,4] and aid in maintaining power systems quality and quantity performance [7]. During short-circuit faults, the amplitude, phase, and frequency of signals in the system such as voltages, currents, and rotor angles will undergo transient variations based on the fault type and location. Their temporal patterns provide a set of covariate features to be extracted from the signals to distinguish among them and other disturbances.

Owing to the development of wide area measurement systems (WAMs) due to the emergence of PMUs, massive data becomes available which allow the application of data-driven approaches to become more widespread in power systems such as in [8,9]. Modern power systems are becoming digital and data-rich; many variables can be measured at various terminals using PMUs such as current, voltage, and frequency and are available as inputs to intelligent techniques for detection and classification. Furthermore, with PMUs emergence, recent works aim to address the problem of fault detection, diagnosis, and location in transmission lines through measurements of variables at different points in the power system. [10] presented a simple method for fault detection and location based on the concept of post-fault change in the injected current at various buses. Thus, in case of no fault at the power system, the change in injected current at all buses is equal to zero where in case of a fault at transmission line, the change of injected currents at the two buses in transmission lines ends is far from zero. In [11], the authors proposed a method based on the pilot impedance for transmission lines fault detection to improve the sensitivity and reliability of the protection by defining the fault components of positive, negative and zero sequence currents and voltages. Authors in [12] introduced a method for transmission lines fault identification and location based on three-phase estimation framework considering measurement chain error model, the method is suitable for both symmetrical and unsymmetrical networks. However, the limitation of these methods is due to ignoring the transient performance where fault detection and location are mainly explained from a single timestamp measurement rather than explaining faults by their full transient multivariate patterns.

Recently machine learning techniques gain more interest with the advent of WAMs and big data availability [13]. Authors in [14] suggested Fuzzy-Neuro method to detect and classify faults at transmission lines. In the proposed Fuzzy-Neuro method, current and voltage samples have been used as inputs and harmonics were removed via Fast Fourier Transform (FFT) through back propagation with fuzzy controllers; the method has been tested on 220 kV, 177.4 km, 50 Hz transmission line and particularly single line to ground faults and double line to ground fault. Similarly, In [15], Discrete Wavelet Transforms (DWT) and Artificial Neural Networks (ANNs) were used for fault detection and identification in a single transmission line with limitations of unsuccessful voltage sags detection and classification of single line to ground faults. Authors in [16] showed that the results for fault detection and classification can be improved using a DWT and self-organized ANN to decompose the signal into three levels and 3960 fault cases were taken for training and validation. A Bayesian classifier and adaptive wavelet method for fault detection and identification has been proposed in [17]. The main shortcoming for the previous approaches is their limitation to single transmission line fault detection and classification.

Conventional machine learning techniques are limited to modelling single sample of features to classification and regression. Advanced deep learning techniques have the advantage of modelling both spatial and temporal relationships in multidimensional features observed sequentially in time. These techniques include Recurrent Neural Network (RNN) and particularly Long Short-Term Memory (LSTM), Stacked Auto Encoder (SAE) and Convolutional Neural Networks (CNNs). Through image processing, authors in [18] showed a great performance of applying the deep CNN method based on faster Region-CNNs for fault detection and classification of high-voltage line by locating the broken insulators and bird nests based on image features. In [19], authors showed another deep learning method for fault classification based on deep belief neural networks (DBNs) in underground cable distribution system by extracting features from the fault signal and classifying them into various categories. [20] showed that results can be improved for fault line selection using adaptive CNNs applied for distribution network with satisfactory classification performance. Another deep learning algorithm used for fault detection and classification of transmission lines based on convolutional SAEs was introduced in [21] by considering many conditions including noise and measurement errors. Authors in [22] showed the performance of a deep learning method for three-phase fault detection and classification in transmission lines based on multilayer Perceptron DNN with wavelet transform.

RNN and particularly LSTM show a great efficiency in feature sequence extraction and data classification in many applications. The advantage of recurrent networks is that they can handle sequences of data where a long pattern of multidimensional features is modelled to classify a particular example in its correct class or mapping the full sequence into a predicted scalar in regression problems. LSTM was proved useful in [23] for automatic feature extraction for photovoltaic array fault diagnosis with high accuracy. [24] proved the superiority of LSTM for fault diagnosis in wind turbines from multivariate time series. In [25], LSTM have been used for sequential fault diagnosis in Tennessee Eastman chemical process. [26] Used LSTM for railway track circuit fault diagnosis. In [27], the authors presented a mixed building of CNN with LSTM trained to predict the fault location distance in a 2-bus single line test system of 220 km length with single-ended voltage and current measurements, this method proved to be superior in classifying transmission line faults compared with other schemes. Authors in [28] presented a combination of LSTM with calibration training filter for transmission lines fault classification, the method was tested on 2-bus single line test system of 300 km length. Authors in [29] presented a wavelet technique with artificial neural networks to detect, classify and locate different types of faults in transmission line of 300 km length, the technique showed a great performance in terms of average error and percentage maximum error of the fault prediction. [30] presented a support vector machine technique with wavelet transform for features extraction to classify and locate short-circuit faults in transmission lines. This technique has been tested on single line, 69 kV, 29.4 km length. Authors in [31] presented a transmission line fault classification based on support vector machine (SVM) with different trained types models. The method has been tested on a single 400-kV, 50-Hz, 300 km transmission line and it achieved a superior fault classification. However, these methods presented in [27], [28] and [29] have been accomplished in a two-bus, single-line power systems. Large-scale multi-machine power systems are more general and more complex and require the completion of three LSTM novel models, to identify the defective region, classifying the type of the fault and to find the distance of the fault location in that faulted region. Beside the fact that LSTM is assumption-free algorithm that easily handles complex nonlinear dynamics in higher-dimensional noisy space. LSTM is also a powerful machine learning tool that has a remarkable advantage compared to classical neural network structures since it captures all important information in both features and time directions coordinately. It captures long-term dependencies in addition to the standard parallel dependencies between input features. It is used to extract the maximum information from the full transient period using high-dimensional input features. To the authors' knowledge, LSTM sequential deep learning was not

investigated before for transmission line fault detection, diagnosis, and location in large-scale multi-machine power systems. Also, both FRI and FTC models were not reported in previous works in large-scale multi-machine power systems and this paper is the first paper which addresses these models and provides information on the accuracy of both faulted region identification and fault type classification. This work presents novel algorithms to determine the fault region, diagnose the fault, and determine its precise location after the fault is detected. Observing and modelling the full sequence, which represents the entire transient phase, which reflects the unique system response for a fault, grants the presented algorithms an increased accuracy and robustness.

The input feature space spans thousands of training and validation examples each includes a long sequence of high-dimensional vector using amplitude and phase of 3-phase voltage and current signals measure at different locations as shown in Fig. 1. This work presents three novel models, the first model is a deep LSTM classifier to detect a fault and identify the faulty region. For each region, a deep LSTM classifier is constructed to classify the faults based on their type and it is used for fault diagnosis in that region. The third deep LSTM regression model is used to determine the exact distance at which the fault occurred within the region.

The rest of this article is organized as follows, section 2 provides a review of RNN and particularly LSTM methods for time series sequence classification and regression; Section 3 then describes the benchmark system used for validation in this work, data collection and faults experiments; The proposed transmission line fault detection, diagnosis, and location methods are presented and described in Section 4; the obtained results are presented and discussed in Section 5; and important conclusions are finally drawn in Section 6.

## 2. Overview on LSTM Cell-Based sequence processing

The recurrent neural network (RNN) is a type of deep neural networks (DNNs). It is a developed and an advanced version of artificial neural networks (ANNs) [23]. RNN is based sequence model that can learn and extract features directly from inputs in time series domain and it can be used in a vast range of applications where inputs are in time series sequence. It is used herein for fault diagnosis in transmission lines where the measured current and voltage data at the end of transmission lines are time series sequences. However, RNN has some disadvantages due to gradient evanescence and it has a limited capability to learn long-term temporal correlations because of the exponentially fast decreasing gradient norm toward zero and due to blast problems. To overcome the vanishing gradient problem [32], LSTM neural network has been proposed. A single LSTM cell consists of an input gate ($i$), output gate ($o$), forget gate ($f$) and the cell candidate ($c$) as shown in Fig. 7. Each gate has an activation function with two weighted inputs; the previous hidden state $h_{t-1}$ of the LSTM cell is weighted by a recurrent weight $RW$ and the current input $X_t$ is weighted by an input weight $IW$ where forget, input and output gates have a sigmoid activation function and the cell candidate gate has a tangent hyperbolic function (TanH) with biases$b$. LSTM cell has two outputs; the memory cell state $C_t$ and the hidden state$h_t$.

The output gates values [23] at time $t$ are computed as follows:

$$i_t = Sigmoid(IW_i \times X_t + RW_i \times h_{t-1} + b_i) \quad (1)$$

$$f_t = Sigmoid(IW_f \times X_t + RW_f \times h_{t-1} + b_f) \quad (2)$$

$$o_t = Sigmoid(IW_o \times X_t + RW_o \times h_{t-1} + b_o) \quad (3)$$

$$c_t^* = TanH(IW_c \times X_t + RW_c \times h_{t-1} + b_c)) \quad (4)$$

Where: $X_t$ is the input feature at the timestamp $t$, $h_{t-1}$ is the previous hidden state,$RW$ and $IW$ and $b$ are the recurrent weights, input weights and biases respectively.

The activation function is a smooth nonlinear function that defines

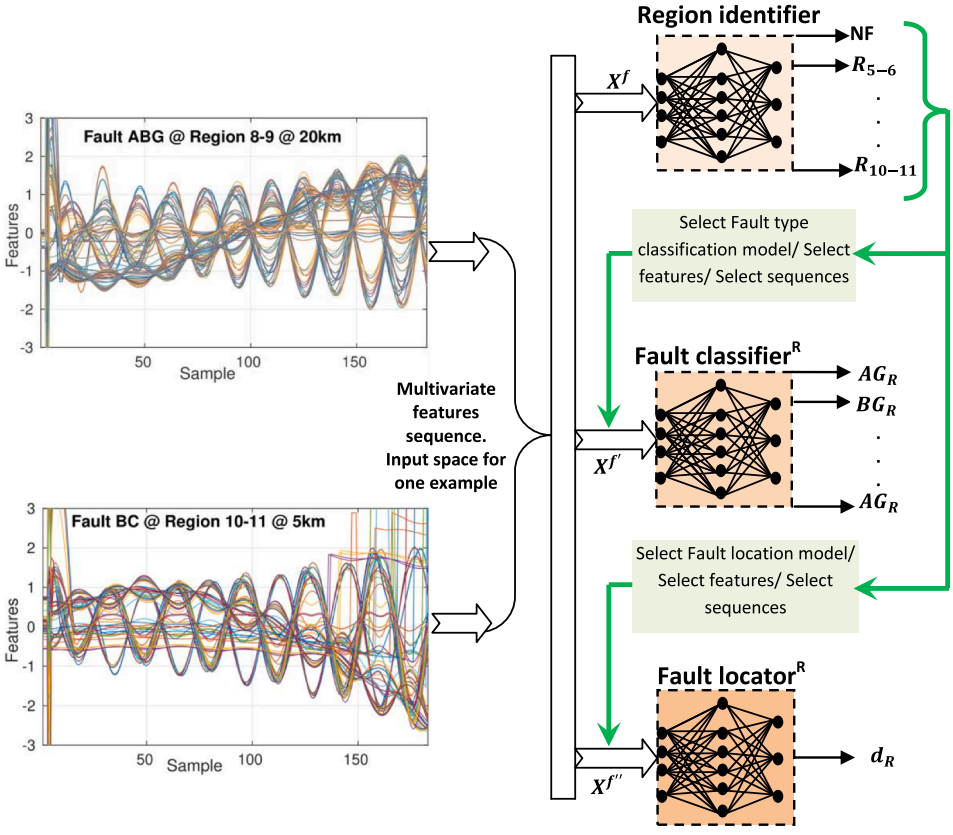


**Fig. 1.** An overview of the presented classification and regression methods.

the output of the neuron given an input or a set of inputs. Hyperbolic Tangent (TanH) and sigmoid functions are activation functions used for LSTM structure and they are defined in table 1 as follows:

The memory cell state $c_t$ and the LSTM hidden state $h_t$ are updated as follows:

$$c_t = c_{t-1} \otimes f_t + c_t^* \otimes i_t \tag{5}$$

$$h_t = o_t \otimes \mathrm{TanH}(c_t) \tag{6}$$

where: $c_{t-1}$ represents the previous memory cell state value. The operation $\times$ is matrix vector multiplication and the operation $\otimes$ is the Hadamard (Element-Wise) product.

## 3. System and faulty scenarios description for data collection

The Two-Area Four-Machine system is used as a case study in this work. The system is publicly available in MATLAB and it is a well-known benchmark that has been used widely in conducting power system studies and it has been detailed in [33]. It consists of two fully symmetrical areas linked together by two transmission lines of 220 km length as shown in Fig. 2. The synchronous generators are represented by their detailed models. The system parameters can be found in [33]. For the analysis of detection, classification, and localization of short-circuit faults, this system is composed of 6 different regions (region from bus 5 to bus 6, region from bus 6 to bus 7, region from bus 7 to bus 8, region from bus 8 to bus 9, region from bus 9 to bus 10 and region from bus 10 to bus 11) in which any fault type can occur at different locations.

The leading important task in machine learning is creating a data set of sufficient and good-quality training data. This should ensure maximum information coverage with minimum redundancy [34,35]. Moreover, it is important that the collected training data are both informative and representative [34]. The current and voltage patterns are selected as inputs directly measured via PMUs whereas other variables such as rotor speed angles and rotor speed deviations are estimated. Moreover, short-circuit faults directly affect the currents and voltages at system buses. Hence, the magnitudes and phases of voltages and currents at the system buses are used as inputs for the following LSTM models, except for generator buses 1, 2, 3, and 4 since these buses are connected to adjacent buses through step up transforms. Those inputs can be obtained from PMUs measurements in real time as well as simulations. Faults occur on the transmission lines even though faults can occur on other components such as transformers, generators, incipient short-circuit faults at underground cables in the distribution network or other equipment faults. However, this work focuses on short-circuit faults on transmission lines. For the offline training of LTSM, a comprehensive data set is constructed to test the LSTM-based models by simulating 10 types of short-circuit faults {AB, AC, BC, ABG, ACG, BCG, AG, BG, CG and ABC}. These faults are classified as line to ground (LG), line to line (LL), line to line to ground faults (LLG), and line to line-to-line faults (LLL), only LLL fault is symmetrical and the remaining are asymmetrical faults as shown in Fig. 3. Where A, B, C, and G stand for Phase A, Phase B, phase C, and ground respectively. These short-circuit faults considered are simulated using Simcape toolbox in Simulink environment in a publicly available test system presented in [33].

The simulation duration is set to 10*s* and that value is considered since it is long enough to ensure most symptoms of the faults are manifested without system loss of synchronism. The various faults were introduced after $t = 1s$ of normal operation. The criteria adopted in this work for loss of synchronism are whenever the difference between the rotor angles of any two generators exceeds 180°. All generated voltages and currents signals are sampled every $1/60s$ over 601 timestamps. Each signal has two features; phase and magnitude time series with time intervals of 0.016 s. Hence for each region $l$ ($l = 1,2, \ldots,6$) at timestamp $t$, four input features are sent to the LTSM that are; the current angles $\theta_{I_n}(\varphi_i)_t$, the current magnitude $I_n(\varphi_i)_t$, voltages angle $\theta_{V_n}(\varphi_i)_t$ and the voltage magnitude $V_n(\varphi_i)_t$, in different phases $\varphi_i$, $i = 1,2,3$, and at buses $n = \{5,6,7,8,9,10,11\}$. In total, the system is monitored through a set of $m = 84$ covariate features in $m$-dimensional data matrix. Each region is subject to a set of faults of different fault types at different locations, and the generated training data are stored as multi-dimensional array of $X \in \mathbb{R}^{m \times n}$ of $m = 84$ features of $n = 601$ samples for each experimental scenario; the full array represent only a single training example in this framework. A total of $N = N_f \times DR_l$ scenarios are conducted at each region to generate $N$ different examples from $N_f = 11$ fault-free and 10 various faults at $DR_l$ various distances for the $l^{th}$ region$l \in \{1,2,3,4,5,6\}$. A key point is the incremental distances of faults. In this work, we consider a 0.8 km to 5 km incremental distance depending on the line length. As these distances were practically chosen to minimize the time required for maintenance. Naturally, large selected distance intervals in training process will result in obtaining large intervals of predicted distances that will negatively affect the maintenance strategy, as the workforce will search for the location of the fault over large distance after determining the faulty region and the type of the fault. This will increase the time required for maintenance that should be minimum. Conversely, selecting very short distances in the training process will greatly complicate the proposed models in which the number of examples will increase. This will affect the proposed models whose practical implementation would be very expensive due to the increase in the number of neurons. Moreover, the large number of examples will lead to increased learning time and stressful computations due to the complexity of these proposed models. More details about the fault distance increment and the number of locations are presented in table 2 and Fig. 4. The training data is hence of dimension$\mathbb{R}^{m \times n \times N_f \times DR_l}$.

To ensure the effectiveness of the proposed models and avoid overfitting, separate testing data are generated with the parameters indicated in table 3 to obtain a new set of examples – different and independent from the training data – to fairly test the models. Notice also that the testing data includes a number of 1481 examples larger than the training data set of 1081 examples; moreover, the testing data includes scenarios at new distances which are unseen for the training stage.

The training data will be used to train three models: two LSTM classification models for faulted region identification and fault type diagnosis besides an LSTM regression model for fault location (distance) prediction. The data are then first pre-processed before being fed into the inputs of each of the three LSTM models, since each LSTM model needs a specific interval and sampling of the data to produce its output as described in detail in Fig. 5. For faulted region detection, the training data have been down sampled to a rate of 1/3 and the sequence starts from the 55th timestamp and increments by 3 timestamps until the 601st timestamp. Each fault type in each region at each distance has its data of voltage and current angles and magnitudes at all the different 7 buses and all the different 3 phases and that gives $(7 \times 3 \times 4) = 84$ features for region classification. Each feature has 183 different timestamps. The dimension of data for one fault at one distance in each region is $\mathbb{R}^{m_1 \times n_1}$. Then the dimension of all the data at each region is $\mathbb{R}^{m_1 \times n_1 \times N_f \times DR_l}$ where $m_1 = 84, n_1 = 183$ and$N_f \times DR_l = 1081$. The motivation behind the down sampling is that the high frequency data is not necessary for faulted region identification. For fault type classification, the data have been extracted from the full resolution of 601 timestamps starting from 60th to the 100th timestamp and the length of the data is $n_2 = 41$

**Table 1**
Activation functions description.

| The activation function | The formula |
|---|---|
| Sigmoid | $\mathrm{Sigmoid}(x) = \frac{1}{1+e^{-x}}$ |
| TanH | $Tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$ |

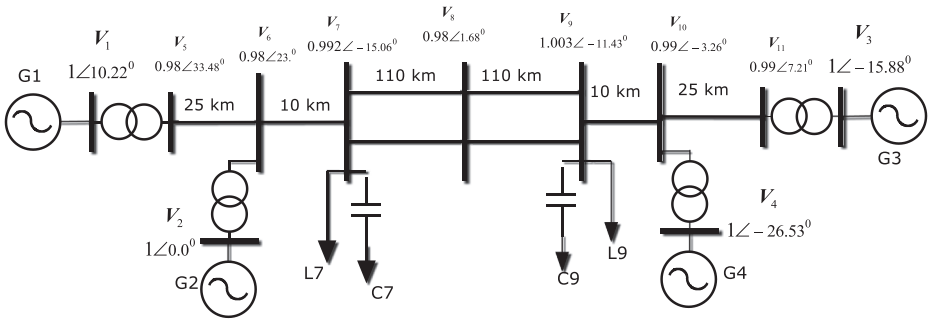


**Fig. 2.** Single Line Diagram of Four-Machine Two-Area Test Power System.

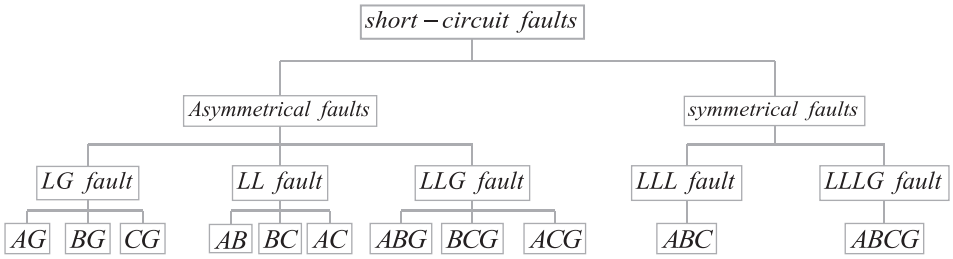


**Fig. 3.** Short circuit fault types classification.

**Table 2**
Training data description.

| Region | $R_{5\text{-}6}$ ($l$ = 1) | $R_{6\text{-}7}$ ($l$ = 2) | $R_{7\text{-}8}$ ($l$ = 3) | $R_{8\text{-}9}$ ($l$ = 4) | $R_{9\text{-}10}$ ($l$ = 5) | $R_{10\text{-}11}$ ($l$ = 6) | No fault |
|---|---|---|---|---|---|---|---|
| Fault distance increment (km) | 1 | 1 | 5 | 5 | 1 | 1 | – |
| Number of locations$DR_l$ | 24 | 9 | 21 | 21 | 9 | 24 | – |
| Number of examples | 240 | 90 | 210 | 210 | 90 | 240 | 1 |

timestamps. The fault classification is challenging and hence the full resolution is required for the LSTM models to extract information; however, the total sequence length is unnecessary for this task and only the first 41 samples are used. After identifying the faulted region, the classification input feature space now spans the difference in voltage angles $\Delta\left[\theta_{V_{n,n+1}}(\varphi_i)_t\right]$, current angles $\Delta\left[\theta_{I_{n,n+1}}(\varphi_i)_t\right]$, voltage magnitudes $\Delta\left[V_{n,n+1}(\varphi_i)_t\right]$, and current magnitudes $\Delta\left[I_{n,n+1}(\varphi_i)_t\right]$between two successive buses $n$ and $n+1$ at the identified phase $\varphi_i$ for line $i$ at timestamp $t$. That gives $m_2 = 3 \times 4 = 12$ features for fault type classification. Then the size of all the data at each region is $\mathbb{R}^{m_2 \times n_2 \times N_f \times DR_l}$.

For fault location, the training data have been down sampled to a rate of 1/3 and the sequence starts from the 80th timestamp and increments by 3 timestamps until the 601st timestamp and the length of the data is $n_3 = 174$. The fault location feature space include full sequence length with down sampling, the input feature space is similar to the fault classification space and it spans the difference of voltage angles $\Delta\left[\theta_{V_{n,n+1}}(\varphi_i)_t\right]$, current angles $\Delta\left[\theta_{I_{n,n+1}}(\varphi_i)_t\right]$, voltage magnitudes $\Delta\left[V_{n,n+1}(\varphi_i)_t\right]$ and current magnitudes $\Delta\left[I_{n,n+1}(\varphi_i)_t\right]$ between two successive buses $n$ and $n+1$ at the identified phase $\varphi_i$ for line $i$ at timestamp $t$ and that gives $m_3 = m_2 = 12$ features for fault each fault location. Then the size of all the data at each region is $\mathbb{R}^{m_3 \times n_3 \times 11 \times DR_l}$. From a practical aspect, the proposed models in section.4 enable the identification of faulted region then the fault type classification and after that the fault location, so it will greatly reduce the time required for maintenance. In addition, with the development of the PMUs the implementation of control strategies, protection schemes, monitoring in real time becomes possible. The PMUs are capable of extracting measurements (voltages, currents, frequencies … etc.) at high rates which allows to provide signals in real time for applications in power systems. PMUs provide signals with 80 to 250 samples per nominal cycles for both 50 and 60 Hz [36], which allow the implementation of our models in real time since the considered interval, in our paper, is 0.016 s or 1.6 ms (equivalently to 60 samples per second), that is practical to obtain through down sampling as clarified in Fig. 5. Therefore, the practical application of the proposed models depends strongly on the advancement of synchro-phasors. The proposed models depend on the PMUs and it is not practical to apply the proposed models if the data extraction is done through SCADA. It is worth noting that with the traditional SCADA, the samples are proceeds each minutes. Thus, with slower data extraction, the accuracy of the fault detection will decrease. SCADA is capable of extracting data at low rates with low sampling frequency and it has been considered too weak for performing a precise fault diagnosis [37] and eventually it will not be practical to apply the proposed models, thereby the protection and maintenance strategies cannot be implemented.

## 4. LSTM models for faulted region detection, fault classification and fault location

Three separate but dependent models have been developed in this work; a classification model for faulted region identification (M1), fault type classification model (M2), and a regression model for fault location (M3). The models are connected hierarchically as presented in Fig. 1. Each of the deep models contains five layers; Input layer, Hidden LSTM layer, LSTM layer, fully connected layer, Softmax layer (for fault detection and fault classification models M1 and M2 respectively) or a Regression layer (for fault location model M3), followed by the output layer. The overall flowchart of the proposed fault detection, classification and location algorithms is shown in Fig. 6. This algorithm consists of three stages: faulted region identification, fault type classification and fault distance prediction.

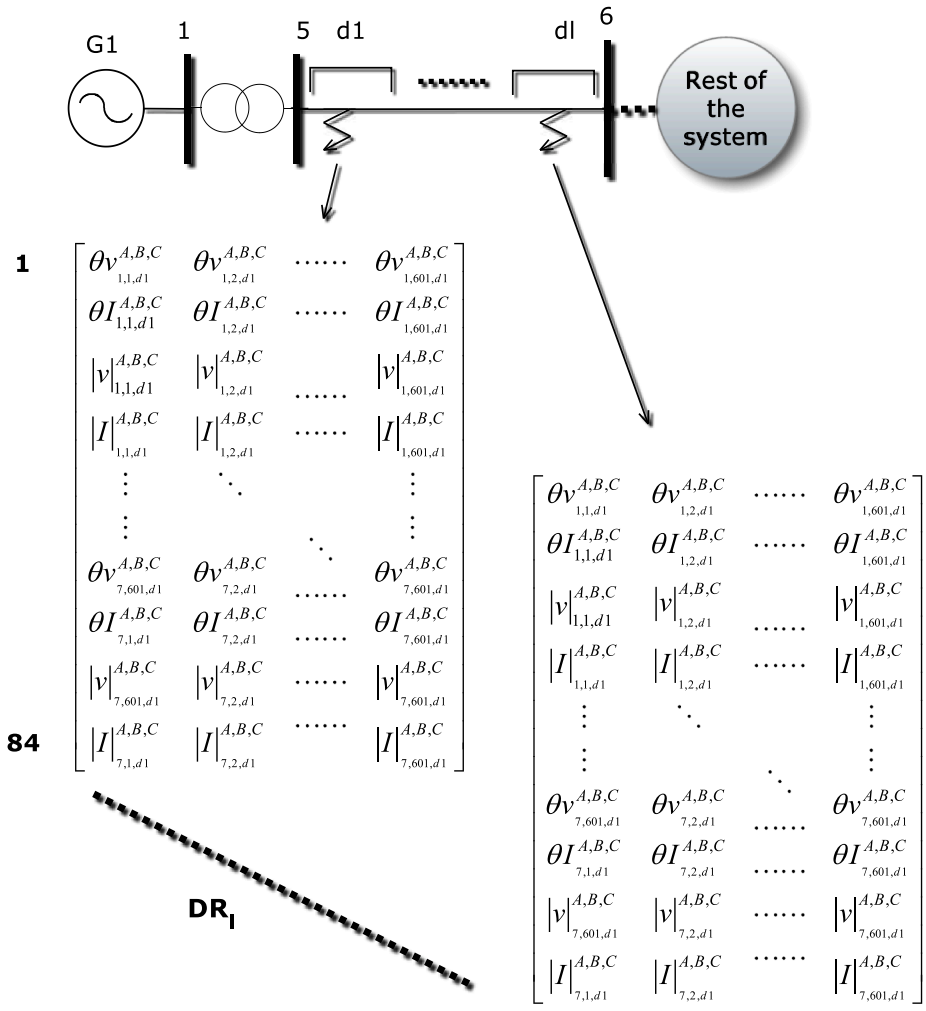


**Fig. 4.** Data structure for one fault type at one region along all locations.

**Table 3**
Testing data description.

| Region | $R_{5\text{-}6}$ ($l=1$) | $R_{6\text{-}7}$ ($l=2$) | $R_{7\text{-}8}$ ($l=3$) | $R_{8\text{-}9}$ ($l=4$) | $R_{9\text{-}10}$ ($l=5$) | $R_{10\text{-}11}$ ($l=6$) | No fault |
|---|---|---|---|---|---|---|---|
| Fault distance increment (km) | 0.8 | 0.8 | 3.5 | 3.5 | 0.8 | 0.8 | – |
| Number of locations$DR_l$ | 31 | 12 | 31 | 31 | 12 | 31 | – |
| Number of examples | 310 | 120 | 310 | 310 | 120 | 310 | 1 |

### 4.1. LSTM and hidden LSTM layers

LSTM networks are robust to the known learning problem of vanishing gradient problem, they are also advantageous in applications of unknown-duration delays in time series data. The main parameters that are selected are the number of cells per LSTM layer. The LSTM layers are used to model both parallel dependencies (correlation between input features) and the temporal dependencies (serial correlations) together from the available data. Since the numbers of (sequences, features per sequence, time length of the sequence) varies between the three problems, their structures should vary respectively.

In faulted region identification model, both LSTM layers include $N_1 = n_1 = 183$ LSTM cells since a rough number around 150 to 200 is needed with the aim to model the full sequence in a multi-class classification problem. This model is the largest since it is for the entire multi-machine power system. Each LSTM cell has 4 inputs assigned to its 4 gates as shown in Fig. 7. In the first layer, the 84 input features are weighted through the input weight vector of $IW_k^1 \in \mathbb{R}^{(4\times m_1)\times 1}$ for each LSTM cell $k$. So, the overall feature space is inputted to the first LSTM layer through an input weight$IW^1 \in \mathbb{R}^{(4\times n_1)\times N_1}$. The outputs of the $k^{th}$ LSTM cell in the first hidden layer at timestamp $t$ are the cell hidden state $h_k^1(t)$ and the memory cell state $h_k^1(t)$. The output vectors $H^1(t)$ and $C^1(t)$ of the first layer are both of dimension $\mathbb{R}^{N_1\times 1}$ and they are both fed back as inputs to the same layer while weighted through the four gates recurrent weights $RW^1 \in \mathbb{R}^{(4\times N_1)\times N_1}$. The second LSTM layer has a similar structure. However, the first hidden LSTM layer passes the full sequence of $\mathbb{R}^{N_1\times n_1}$ of hidden states vector over all timestamps $t = 1, 2, \cdots, n_1$ to the second LSTM layer, which passes only the last output $H^2(T) \in \mathbb{R}^{N_1\times 1}$ to the next layer.

In fault type classification model, both LSTM layers include $N_2 = 82$ LSTM cells. Here 82 cells have been selected instead of 41 because it was verified to be insufficient and this classification problem is more challenging since it is difficult to separate the fault classes which have similarities (AB, BC, AC, … etc.) and a more complex model was required. In fault location prediction, the LSTM layers include $N_3 = 100$ LSTM cells. Deepest and widest possible models have been selected and provided that the training data are enough to fit the model properly. Of course, the deeper (more layers) and the wider (more cells per layer) the network is, the better the performance will be. Unfortunately, this comes

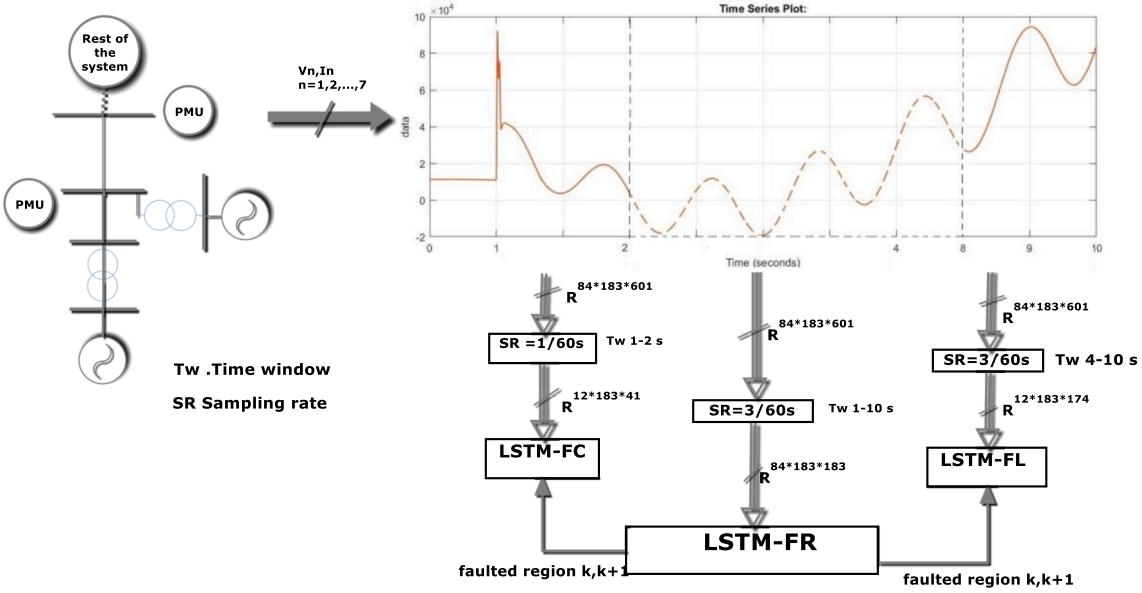


**Fig. 5.** The real-time response of the proposed LSTM models.

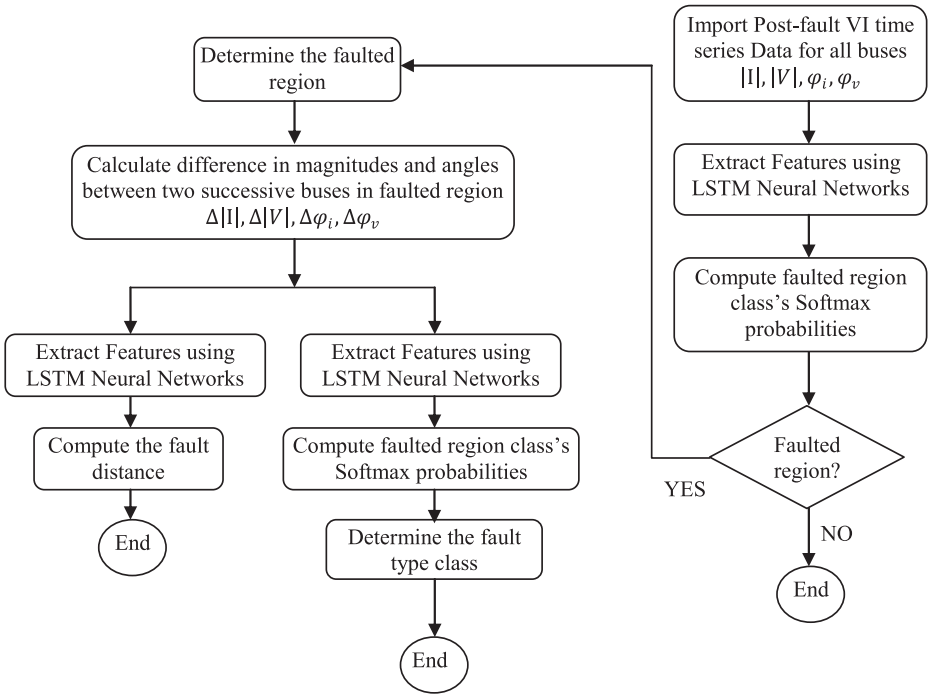


**Fig. 6.** flowchart of the proposed approach.

at the cost of increased complexity of the model in addition to falling into a high-dimensional problem where the number of unknown parameters (degrees of freedom in the model) exceeds greatly the number of available training observations (sequences, features, time points). In either of the cases, it is difficult to fit the model, the first causes an over fitted model with uncontrolled variance, the second causes an insufficiently trained model where error bias cannot be controlled. A sort of tradeoff should be reached to ensure acceptable performance without complexity and over/under fitting, this is what has been done, as shown in subsection 5.3 and subsection 5.4, by the prediction results where the mean prediction error is around zero and the error deviation is not too far. The input $X_t$ represents the time series input with $m$ channels of voltage and current magnitudes and angles as $t \in \{1, 2, 3, \cdots, T\}$, T is the total number of timestamps where $T = n_1 = 183$ in region classification, $T = n_2 = 41$ for fault type classification and $T = n_3 = 174$ for fault location. For fault region identification, $X_t$ is $[\theta_{I_n}(\varphi_i)_t, \theta_{V_n}(\varphi_i)_t, I_n(\varphi_i)_t, V_n(\varphi_i)_t]$ and for fault type classification or for fault location; $X_t$ is $\left[\Delta\left[\theta_{I_{n,n+1}}(\varphi_i)_t\right], \Delta\left[\theta_{V_{n,n+1}}(\varphi_i)_t\right], \Delta\left[I_{n,n+1}(\varphi_i)_t\right], \Delta\left[V_{n,n+1}(\varphi_i)_t\right]\right]$.

The initial values are set to vector of zeros as $C^1(0) = \theta$ and $H^1(0) = \theta$ and the operator $\otimes$ denotes the Hadamard product. The time series data, of set of features sequences$X_t$ in terms of timestamps$t$, pass sequentially through an LSTM layer where the next hidden state vector $H^1(t)$ is computed at each timestamp $t$ in order to be as a hidden state vector for the next LSTM unit $k$. The process continues until scanning a one sequence for one fault type at a certain distance and then the next

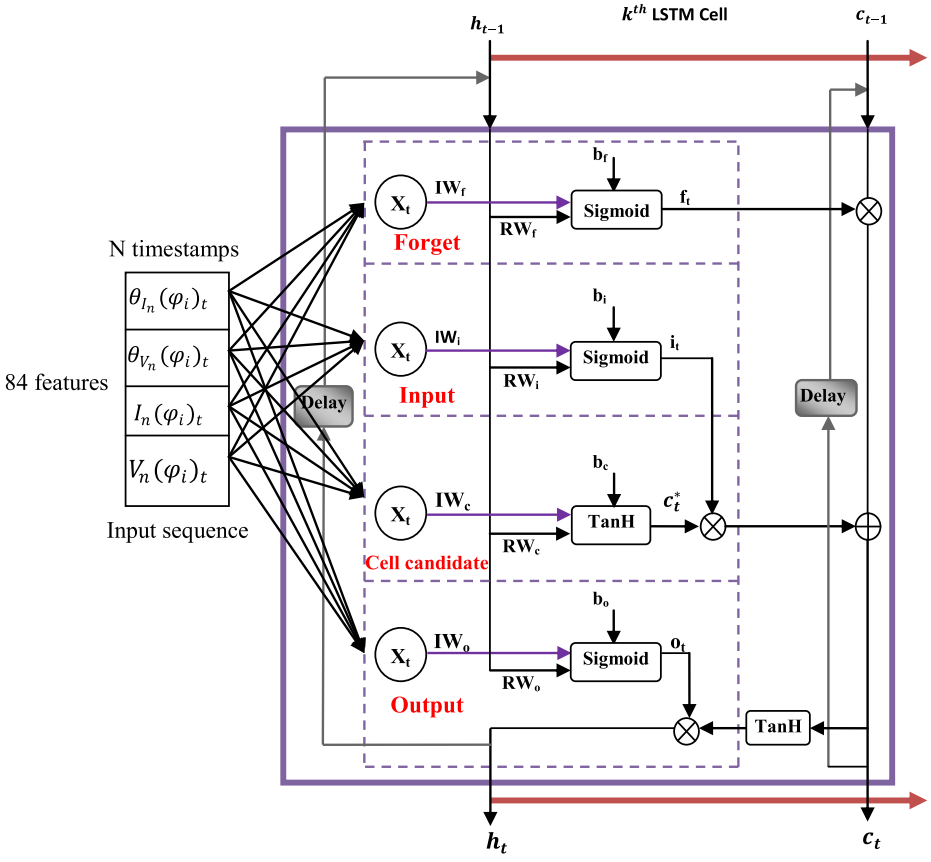


**Fig. 7.** Input sequences connections with $k^{th}$ LSTM cell in an LSTM layer.

sequence starts to flow. The cell memory state vector $C_1$ and the cell next hidden state vector $H^1(t)$ of the hidden LSTM layer 1 at timestamp$t$ are determined in Eq. (7) and Eq. (8) respectively:

$$C^1(t) = \left[c_1^1(t), c_2^1(t), \cdots, c_k^1(t), \cdots, c_N^1(t)\right]^T \tag{7}$$

$$H^1(t) = \left[h_1^1(t), h_2^1(t), \cdots, h_k^1(t), \cdots, h_N^1(t)\right]^T \tag{8}$$

where: $C^1(t)$ and$H^1(t) \in R^{N\times 1}$.

### 4.2. Fully connected, softmax, classification and regression layers

In the three models, the output of hidden LSTM layer $H^2(t)$ is put into the input of the fully connected layer, which contains a number of simple neurons. Each neuron has its input of the previous hidden LSTM layer output vector $H^2(T)$ with dimension of $R^{N\times 1}$ and multiplied by weights $IW_j^3$ with dimension of $R^{1\times N}$ and added to the layer bias $b_j^3$.

The fully connected layer (third layer) outputs vector $o_j^3$at each unit $j$ is computed as follows [38]:

$$o_j^3 = f\left(\sum_{k=1}^{N} IW_{j,k}^3 \times H_k^2(t) + b_j^3\right) \tag{9}$$

where $f$ is the activation function, it is a smooth nonlinear function such as the sigmoid function ($\sigma$) or a hyperbolic tangent function.$IW_{3\,j,k}$ are the fully connected layer weights of neuron $j$multiplied by the $k^{th}$ value of the LSTM layer output vector $H_k^2(T)$, $b_j^3$ is the bias of neuron $j$,$IW_j^3 \in \mathrm{R}^{1\times N}$, The vector$IW^3$of all the input weights in fully connected layer is with dimension of $R^{J\times N}$ and $\mathrm{b}^{(3)} \in \mathrm{R}^{J\times 1}$ is the bias vector of all neurons biases.

In faulted region identification, the fully connected layer has 7 neurons since the total number of classes is $J = 7$ and the class number$j \in \{1, 2, 3, 4, 5, 6, 7\}$.

In fault type classification, a total number of 11 neurons have been used according to total number of faults classes $J = 11$ in which each class number $j \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$.

In faulted location, the FCL has only one neuron as$J = 1$ and $j = 1$.

The vector $o_j^3$ is called the logits vector or the scores vector, $\mathrm{o}^3$ is the predicted location of one fault example.

For both faulted region identification and the fault type classification models, the output of the fully connected layer is fed into the input of a Softmax layer.

$$o_j^4 = \mathrm{Softmax}\left(o_j^3\right) = \frac{e^{o_j^3(t)}}{\sum_{q=1}^{J} e^{o_q^3(t)}} = P\left(C_j|x,\theta\right) \tag{10}$$

where: $o_j^4 \in \mathrm{R}^{J\times 1}$ and J is the total number of classes which represents the number of neurons in the fully connected layer (which is equal to the number of units in the Softmax layer); J = 7 for faulted region identification, $J = 11$ for fault type classification and$j = \{1, 2, \cdots, J\}$. This quantity represents the probability of class $C_j$, given the sample $x$ and the neural network parameters $\theta$.

The cross entropy loss function has been used by the classification layer based on the results of the Softmax cross entropy [39]. The cross entropy loss function is defined as follows [40]:

$$Loss = -\sum_{i=1}^{M}\sum_{j=1}^{J} I_{ij}\ln\left(P\left(C_j|x_i,\theta\right)\right) \tag{11}$$

where: $M$ is the number of all examples of sequences, and $J$ is the number of classes. $I_{ij}$ is an indicator of the desired class, $I_{ij} =$

$$\begin{cases} 1 & \textit{for correct class} \\ 0 & \textit{for other class} \end{cases}.$$

The loss is then reversely proportional to the probability of making correct classifications for all input sequences, and hence the challenge of the deep sequential is to minimise this loss. The training phase determines the optimal input and recurrent weights besides biases for all the units in all the layers to ensure minimum loss [41 42] by the neural network model.

The fault location is a regression problem where a regression layer is used after the fully connected layer instead to calculate the root mean-squared-error loss for the regression fault location problem. The root-mean-squared-error loss is computed as follows [43]:

$$Loss = \sum_{p=1}^{M} \left( y_p - o^3(p) \right)^2 \tag{12}$$

where: $y_p$ is the real location, $o^3(p)$ is the predicted location of fault example$p$ and $M$ is the total number of examples.

On a set of training sequences, LSTM units can be trained in a supervised manner using an optimisation algorithm, like Stochastic Gradient Descent Method (SGDM) using back propagation through time to compute. Using the gradient, the large parameters vector $\theta$ is pushed closer to optimal values in each iteration [44 45]:

$$\theta_{it+1} = \theta_{it} - \eta \frac{\partial Loss}{\partial \theta} \tag{13}$$

where $\theta$ is a large vector of all layers' parameters in the neural network model, $\eta$ is the learning rate, $it+1$ is the next iteration. In batch training, the stochastic gradient is used to approximate the real gradient by calculating the gradient of stochastic loss over a randomly selected subset of $m_{it+1} < M$ training examples instead of the global set of overall examples.

## 5. Results and discussion

The proposed method has been tested on the test system described in section 3. The model output is either a class label of identified faulted region and fault type classification or a real number representing the distance for fault location.

### 5.1. Faulted region identification (FRI) model

The FRI model was not reported before for large-scale multi-machine power systems. This paper is the first to report on the FRI model and provide the accuracy of region classification in large-scale multi-machine power systems. A multi-class classification problem of 7-class has been presented for FRI model. The number of epochs used for training is set to 500 and the maximum number of iterations is 12,000 where 24 iterations have been used for one epoch. Over a single CPU, the models' parameter tuning is achieved using SGDM with back propagation through time; A piecewise learning rate schedule has been used, the choice of the learning rate is selected according to the stopping criteria, here it is the number of epochs and the number of iterations per epoch (for SGDM). The learning rate is generally selected in the interval of 1% to 10%. Large learning rate implies faster convergence to optimal tuning but with larger oscillations around it, and the opposite for smaller learning rates. The learning rate is set to decay exponentially from 5% by a factor of 0.97 each 10 epochs; the high training rate at the beginning speeds up the convergence toward optimal parameters, this rate is gradually reduced to reduce the chattering effects and limit the oscillations around the optimal parameters when approaching the end of training. The training process reached 100 percent during a number of 7905 iterations and at a learning time of 1397 min and 16 s and the learning rate reached 1.8865% at this number of iterations. The training procedure has given a great performance. The accuracy achieved nearly 100% and the loss has achieved nearly zero after 8000 iterations. For testing performance, the confusion matrix is shown in Fig. 8; the matrix has been translated into table.4. The deep LSTM classification achieved a very high accuracy and the misidentified examples are those very close to regions buses. The number of faults, the correct identified faults, the misidentified faults and the accuracy of faulted region identification are listed in table 4, which shows the great performance of faulted region detection and identification where the No-fault accuracy is 100%. In which a total number of 24 faults have been correctly identified. The accuracy for regions 10–11, 5–6 and 6–7 is very high where just nine faults have been misidentified in region 10–11 and have been attributed to region 9–10 from a total number of 341 faults with accuracy of 97.36%. Only ten faults have been wrongly identified in region 5–6 and have been assigned to region 6–7 from a total number of 341 faults with accuracy of 97.06% and six faults have been misidentified in region 6–7 from 132 faults and have been attributed to region 5–6 with accuracy of 95.45%. While the accuracy is reasonable and good in regions 7–8, 8–9 and 9–10, a number of 249 faults have been correctly identified and have been attributed to region 7–8 from a number of 341 faults with accuracy of 73.02% and the wrongly identified faults have been assigned to regions 5–6, 6–7 and 8–9. A number of 255 faults have been correctly assigned to region 8–9 from a total number of 341 faults with accuracy of 74.78% and the other misidentified faults have assigned to all other regions. A total number of 98 faults have been correctly identified in region 9–10 from a total number of 132 faults with accuracy of 74.24% and the misidentified faults have been assigned to region 10–11.

### 5.2. Fault type classification (FTC) model :

The FTC model is first proposed in multi-machine power systems in this work. A multi-class classification problem of 11-class has been presented for FTC model. The fault type classification model is trained through 300 epochs and the maximum number of iterations is 2100 where 7 iterations have been used for one epoch. Over a single CPU, a piecewise learning rate schedule has been used for our case of study and the learning rate is initialized at 0.07 with a decaying factor of 0.97 each 10 epochs. The training process reached 100 percent during a number of 1469 iterations and at a learning time of 5 min and 23 s where the learning rate achieved 3.8066% at this number of iterations, the training procedure has given a great performance. The accuracy achieved almost 100% and the loss has achieved nearly zero at a number after 1400 iterations. For testing performance, the confusion matrix is shown in Fig. 9 and it has been translated into table.5 where only few faults are

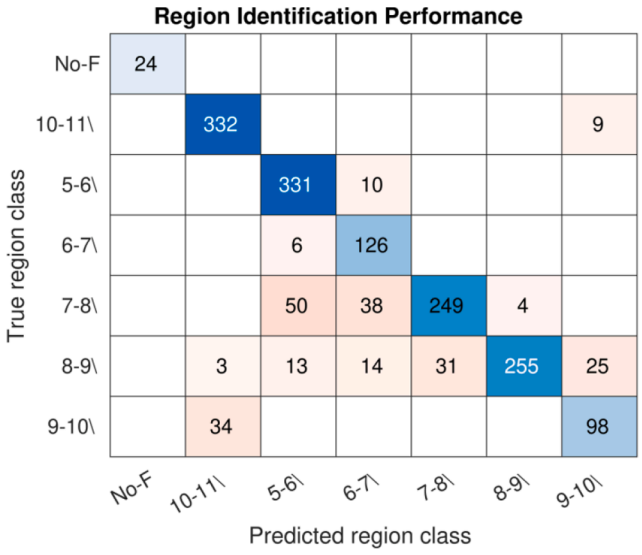


**Fig. 8.** The testing confusion matrix of the faulted region identification model.

**Table 4**
The testing accuracy of the faulted region identification model.

| Region | Number of faults | Correctly identified faults | Misidentified faults | Accuracy (%) |
|---|---|---|---|---|
| No-Fault | 24 | 24 | 0 | 100 |
| 10–11 | 341 | 332 | 9 | 97.36 |
| 5–6 | 341 | 331 | 10 | 97.06 |
| 6–7 | 132 | 126 | 6 | 95.45 |
| 7–8 | 341 | 249 | 92 | 73.02 |
| 8–9 | 341 | 255 | 86 | 74.78 |
| 9–10 | 132 | 98 | 34 | 74.24 |

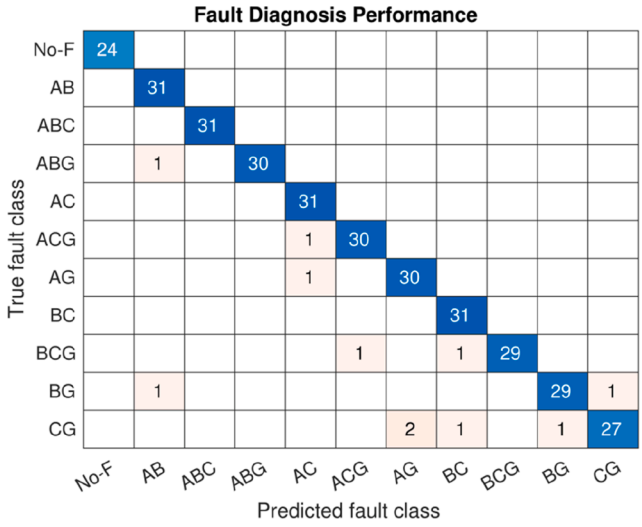


**Fig. 9.** The testing confusion matrix of the fault-type classification model.

**Table 5**
The testing accuracy of the fault-type-classification model in region 10–11.

| Fault type | Number of faults | Correctly classified faults | Misclassified faults | Accuracy (%) |
|---|---|---|---|---|
| No-F | 24 | 24 | 0 | 100 |
| AB | 31 | 31 | 0 | 100 |
| ABC | 31 | 30 | 0 | 100 |
| ABG | 31 | 30 | 1 | 96.77 |
| AC | 31 | 31 | 0 | 100 |
| ACG | 31 | 30 | 1 | 96.77 |
| AG | 31 | 30 | 1 | 96.77 |
| BC | 31 | 31 | 0 | 100 |
| BCG | 31 | 29 | 2 | 93.54 |
| BG | 31 | 29 | 2 | 93.54 |
| CG | 31 | 27 | 4 | 87.09 |

wrongly classified. In region 10–11, No-Faults, AB faults, ABC faults and AC faults have been totally and correctly classified with 100% of accuracy from a total number of 24 faults for No-Fault condition and 31 faults for fault condition. ABG faults, ACG faults and AG faults have been correctly classified with 96.77% of accuracy from a total number of 31 faults where one fault has been classified as AB fault for ABG faults case, one fault has been classified as AC fault for ACG faults case and one fault has been classified as AC fault for AG faults case. BCG and BG faults have been correctly classified with 93.54% of accuracy of a full number of 31 faults in which only two faults have been classified as BC and ACG faults for BCG faults case and two faults have been classified as CG and AB faults for BG faults case. CG faults have been classified with 87.09% accuracy and only 4 faults have been misclassified from a full number of 31 faults where two faults have been classified as AG faults and the others have classified as BC and BG faults. The number of faults, the correct classified faults, the misclassified faults and the accuracy of fault type classification in region 10–11 are listed in table 5, which shows the great performance of fault type classification.

### 5.2.1. *Comparison results of fault type classification*

In this subsection, the fault type classification model is compared with the SVM fault classifier model presented in [31] including its two training methods; Poly SVM ($\nu = 2$) and Gaussian SVM ($\sigma = 0.5$). [31] used binary classification only $y \in \{-1, 1\}$ to classify faults $y = 1$ from no-fault $y = -1$. The binary fault classification is simpler since it groups all faults in one class and it ignores the type pf the fault. Three SVM binary classifiers have been used; SVM-1 for fault type classification, SVM-2 for ground fault detection and SVM-3 for section identification. A total number of 89 test cases, for SVM classification, have been used for L-G, LL-G, LL and LLL faults types whereas a total number of 334 test cases have been used in our case for LSTM-based classification models. Table 6 shows the different classifiers used for FTC, parameter values of different SVM models reported in [31] and our LSTM models, number of test cases and fault type classification accuracy.

In terms of binary fault classification, i.e. classifying all faults in one class from no-faults in an alternative class, the proposed LSTM-based FTC greatly outperforms the results of the SVM model; The no-fault cases, labelled NF in the first row in Fig. 9, are purely classified from the faults in the other rows. In addition, Fig. 8 shows that all types of faults in all regions are purely classified from the no-fault cases. The 100% binary classification accuracy of the LSTM classifiers is because fault symptoms are easily detected using sequential deep learning; classifying faults from no-fault cases reduces to a trivial problem since the symptoms of faults are apparent; The LSTM – FTC only struggles in distinguishing the symptoms of one fault from the symptoms of another equally-severe fault in order to classify the region then the types of these faults, this was not reported in previous works. In addition to the multi-class FTC problem, the proposed LSTM – FTC model has been tested on large-scale multi-machine power systems while taking into account spatiotemporal dependencies between all signals in all transmission lines. SVM fault classifiers have been tested on simple power system of

**Table 6**
Comparison of results of different fault type classification models.

| Fault type | Method Name | Number of test cases | Classification rates (%) |
|---|---|---|---|
| No-Fault | Poly-SVM[(a)], $\nu = 2$ | – | – |
| | Gaussian-SVM[(a)], $\sigma = 0.5$ | – | – |
| | LSTM[(b)] | 24 | 100 |
| L-G | Poly-SVM[(a)], $\nu = 2$ | 15 | 96.52 |
| | Gaussian-SVM[(a)], $\sigma = 0.5$ | 13 | 95.23 |
| | LSTM[(b)] | 93 | 92.47 |
| LL-G | Poly-SVM[(a)], $\nu = 2$ | 9 | 96.27 |
| | Gaussian-SVM[(a)], $\sigma = 0.5$ | 7 | 97.51 |
| | LSTM[(b)] | 93 | 95.70 |
| LL | Poly-SVM[(a)], $\nu = 2$ | 11 | 96.84 |
| | Gaussian-SVM[(a)], $\sigma = 0.5$ | 9 | 95.99 |
| | LSTM[(b)] | 93 | 100 |
| LLL | Poly-SVM[(a)], $\nu = 2$ | 14 | 96.28 |
| | Gaussian-SVM[(a)], $\sigma = 0.5$ | 11 | 95.68 |
| | LSTM[(b)] | 31 | 100 |

[(a)] : used a total number of 89 Data sets, tested on a simple 2-bus test power system; 400-kV, 50-Hz, 300 km transmission line; The binary classifier classifies faults (class I) from no-fault (class II).
[(b)] : used a total number of 334 Data sets, tested on large-scale multi-machine test power system; (Four-Machine Two-Area 11-Bus Test Power System), the model classifies all fault types in addition to fault and no-fault operation.

single 400-kV, 50-Hz, 300 km transmission line and there are no other transmission lines that may affect this test transmission line. Yet, our approach still maintained its high accuracy and excellence despite the presence of many transmission lines that influence each other and this may affect the accuracy of the classification process. This strongly demonstrates the robustness of our method for classifying faults and that the proposed LSTM – FTC model accuracy is unaffected by the complexity of power systems.

Table 7 shows the total performance of our proposed LSTM-FTC model. The proposed model totally performs better than both Poly-SVM ($\nu = 2$) and Gaussian-SVM ($\sigma = 0.5$) fault classifiers in which the total LSTM-FTC model accuracy is 96.71% while it is 96.48% and 95.92% for both Poly-SVM ($\nu = 2$) and Gaussian-SVM ($\sigma = 0.5$) fault classifiers despite the fact that our proposed LSTM-FTC model is tested on large-scale multi-machine power system and SVM fault classifiers have been tested on simple power system and this work considered a multi-class fault type classification whereas [31] considered binary fault classification only.

Ground faults in [31], are treated separately and another SVM block has been used for detecting the ground faults. In our case of study, only one FTC model classifies all faults types from one another beyond classifying just fault class from no-fault class or the presence of ground fault from its absence. However, to be totally fair, it is also important to mention that the superior LSTM based models are more complex than the SVM models; the advantage of SVM is that it can be implemented on advanced protection devices whereas the deep LSTM-based models require serious computation resources and they can only be used in surveillance and control centres to monitor the entire system.

### 5.3. Fault location (FL) model

This model is trained to predict the fault location using 1000 epochs with a maximum number of 8000 iterations where 8 iterations have been used for one epoch. Over a single CPU, a piecewise learning rate schedule has been used for our case of study and the learning rate is set to 0.07 and its decaying factor is set to 0.997 at each epoch. The training process reached final number of 8000 iterations at a training time of 30 min and 33 s where the learning rate achieved 0.34799%, the training procedure has given a great performance. The root mean square error (RMSE) and the loss function are very small just after a number of 3000 iterations. For training performance, the distance prediction errors, which are a difference between the true fault location distances and the predicted fault location distances, are shown in Fig. 10. Fig. 10 explains the errors in predicting distances for each fault type. From the figure, standard deviations of distance errors for ABG, ACG and BCG fault types are close to each other and smaller than standard deviations of distance errors for CG, BG and AG faults that are also close to each other. Standard deviations of distance errors for CG, BG and AG faults are smaller than standard deviations of distance errors for AB, AC and BC faults that are close to each other. The standard deviation of distance errors for ABC fault type is the greatest one. Table 9 shows statistical results of the prediction error of the fault location distance during training for each

**Table 7**
Comparison results of FTC models.

| Model | Number of test cases | Classification rates (%) |
|---|---|---|
| Poly-SVM[(a)], $\nu = 2$ | 49 | 96.48 |
| , $\sigma = 0.5$ | 40 | 95.92 |
| LSTM[(b)] | 334 | 96.71 |

[(a)] : used a total number of 89 Data sets, tested on a simple 2-bus test power system; 400-kV, 50-Hz, 300 km transmission line; The binary classifier classifies faults (class I) from no-fault (class II).
[(b)] : used a total number of 334 Data sets, tested on large-scale multi-machine test power system; (Four-Machine Two-Area 11-Bus Test Power System), the model classifies all fault types in addition to fault and no-fault operation.

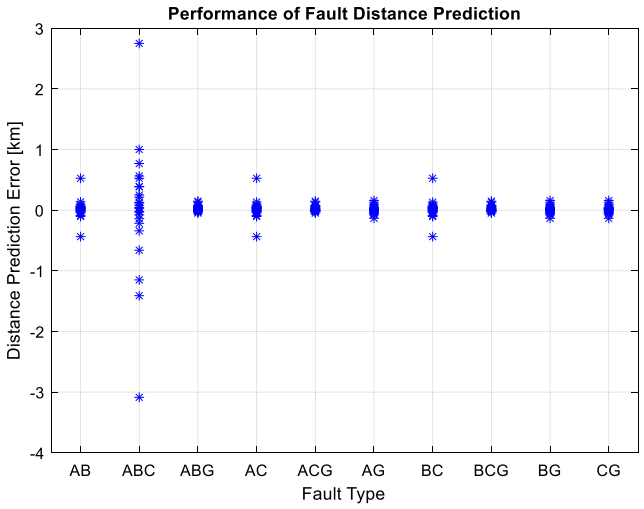


**Fig. 10.** The distance prediction errors of all faults location distance predictions during training in region 10–11.

fault type in region 10–11 and it explains the Fig. 10. The true locations, the predicted locations and the predicted errors of ABG fault in region 10–11 are listed in table 8, which shows the great performance of fault location prediction; a negative sign means that the true value is greater than the predicted value.

#### 5.3.1. Comparison results of fault location

The results of the proposed fault location model have been compared statistically with other techniques presented in the introduction. These techniques have been tested on single line 2-bus small-scale simple power systems. In that way it is assured that the problem of fault detection for larger system with greater lines lengths is more complicated than in ones of smaller sizes. Thereby it further demonstrates the superiority of the proposed models as they are able to achieve better accuracy in larger systems. This allows for a very fair comparison with other techniques as the transmission lines, in very complex power systems, are affected by each other and this adversely affects the accuracy of prediction error, unlike very simple power systems that contain one

**Table 8**
The distance predictions of ABG faults location during training in region 10–11.

| The true distance (km) | The predicted distance (km) | The distance prediction error (km) |
|---|---|---|
| 1 | 1.0595 | 0.0595 |
| 2 | 1.9518 | −0.0482 |
| 3 | 3.0275 | 0.0275 |
| 4 | 4.0773 | 0.0773 |
| 5 | 4.9936 | −0.0064 |
| 6 | 5.9946 | −0.0054 |
| 7 | 7.0094 | 0.0094 |
| 8 | 7.9761 | −0.0239 |
| 9 | 8.9894 | −0.0106 |
| 10 | 10.1480 | 0.1480 |
| 11 | 11.0161 | 0.0161 |
| 12 | 12.1349 | 0.1349 |
| 13 | 13.0505 | 0.0505 |
| 14 | 13.9887 | −0.0113 |
| 15 | 15.0217 | 0.0217 |
| 16 | 16.0807 | 0.0807 |
| 17 | 16.9744 | −0.0256 |
| 18 | 18.0436 | 0.0436 |
| 19 | 18.9974 | −0.0026 |
| 20 | 20.0311 | 0.0311 |
| 21 | 20.9930 | −0.0070 |
| 22 | 21.9892 | −0.0108 |
| 23 | 23.0017 | 0.0017 |
| 24 | 23.9761 | −0.0239 |

**Table 9**
Statistical results of prediction errors of fault location distances for each fault type in region 10–11.

| Fault type | Standard deviation of Errors | Mean of Errors | Fault type | Standard deviation of Errors | Mean of Errors |
|---|---|---|---|---|---|
| CG | 0.0627 | 0.0107 | ACG | 0.0494 | 0.0220 |
| BG | 0.0627 | 0.0107 | AC | 0.1520 | 0.0191 |
| BCG | 0.0496 | 0.0221 | ABG | 0.0494 | 0.0219 |
| BC | 0.1519 | 0.0190 | ABC | 1.0145 | 0.0047 |
| AG | 0.0629 | 0.0106 | AB | 0.1520 | 0.0189 |

transmission line that is not affected by any other lines. Proving the proposed fault location model to be effective in complex power systems will prove its strength compared to other techniques that have been tested on very simple power systems. The fault location absolute error (%), which is expressed in Eq.(14), has been used in table 11 for comparison of our location model against the approach presented in [30] and it specifies the percentage of absolute prediction error over the total length of the transmission line. This fault location absolute error (%) is a measure that allows to make a comparison using different frames of references because it is, in fact, a ratio of a distance prediction absolute error to the total length of the line. Table 10 shows the comparison results of the proposed fault location model with wavelet-ANN technique presented in [29] for locating the faults in transmission line.

$$Absolute\ error\ (\%) = \frac{|Predicted\ location\ distance - True\ location\ distance|}{Total\ line\ length} \times 100\% \tag{14}$$

The fault location error (%) and the fault location absolute average error (%), used for generating table 8, are expressed in Eq.(15) and Eq. (16):

$$e_k(\%) = \frac{Predicted\ location\ distance_k - True\ location\ distance_k}{Total\ line\ length} \times 100\% \tag{15}$$

$$Absolute\ Average\ error\ (\%) = \left|\frac{\sum_{k=1}^{M} e_k(\%)}{M}\right| \tag{16}$$

where, $e_k(\%)$ is the $k^{th}$ error (%) and $M$ is the number of data.

The proposed method works well to locate the faults compared to other wavelet-ANN method, as it surpasses it in many cases except in the case of line-to-line faults. The absolute average error rate does not exceed 0.1 in all cases, and this proves the superiority of our proposed method despite the fact that the power system, used for training and testing, is very complex and all lines are affected by each other. Table 11

**Table 10**
Comparison of statistical results of fault location distances prediction error for different types.

| Method | LSTM [a] | Wavelet-ANN [b] [29] |
|---|---|---|
| Fault type | **Average error (%)** | **Average error (%)** |
| LG faults | 0.0427 | 0.10 |
| LLG faults | 0.0880 | 0.16 |
| LLG faults | 0.0189 | – |
| LL faults | 0.0761 | 0.01 |

[a] : Results from complex power system: Kundur Four-Machine Two-Area 11-Bus Test Power System.
[b] : Results from simple power system: Single line 2-Bus Test Power System

**Table 11**
Comparison of statistical results of distance prediction error for LG fault type location.

| Method | LSTM [a] | | Wavelet-SVM [b] [30] | |
|---|---|---|---|---|
| True distance (km) | Predicted distance (km) | Absolute error (%) | Predicted distance (km) | Absolute error (%) |
| 2 | 2.034 | 0.136 | 2.178 | 0.605 |
| 4 | 4.031 | 0.124 | 3.899 | 0.343 |
| 6 | 6.001 | 0.004 | 6.019 | 0.064 |
| 8 | 7.940 | 0.240 | 8.117 | 0.398 |
| 10 | 10.089 | 0.356 | 10.149 | 0.506 |
| 12 | 12.032 | 0.128 | 12.146 | 0.496 |
| 14 | 14.041 | 0.164 | 14.119 | 0.404 |
| 16 | 15.866 | 0.536 | 16.060 | 0.204 |
| 18 | 18.158 | 0.632 | 17.964 | 0.122 |
| 20 | 20.016 | 0.064 | 19.835 | 0.561 |
| 22 | 21.984 | 0.064 | 21.710 | 0.986 |
| 24 | 23.973 | 0.108 | 23.703 | 1.010 |

[a] : Results from Complex power system: Region 10–11, Kundur Four-Machine Two-Area 11-Bus Test Power System.
[b] : Results from very simple power system: Single line −69 kV − 29.4 km Test Power System.

shows another comparison with another technique called Wavelet-SVM, the comparison results are statistical results of distance prediction error for LG fault type location.

Table 11 shows that the proposed method performs very well as it outperforms in most cases a method of wavelet-SVM. This is evidenced by table 12, which shows a statistical comparison of mean, standard deviation (STD), minimum value and maximum value of the absolute error (%). It can be found that the proposed method excels the method wavelet-SVM in all statistical results.

## 6. Conclusion

This paper considered the problem of fault detection, identification, and diagnosis in transmission lines in a data-driven approach for large-scale multi-machine power systems in which an extensive data validation across numerous scenarios in a Two-Area Four-Machine Power System has been performed. The tacked challenge is data driven faulted zone identification, accurate fault type classification, and exact fault location prediction. Novel three deep learning models – based on LSTM – are designed for intelligent fault identification, classification and location in transmission lines of the Two-Area Four-Machine Power System. The neural network models are developed in both classification and regression frameworks; namely faulted region identification model, fault type classification model, and fault location model. The faulted region identification (FRI) and fault type classification (FTC) models have not been verified before in Large-Scale Multi-Machines Power Systems and this paper is the major one for addressing this work and

**Table 12**
Comparison of absolute error statistical results of distance prediction error for LG fault type location.

| Error type | Absolute Error (%) | | | |
|---|---|---|---|---|
| Method | Mean | Std | Min (%) | Max (%) |
| LSTM | 0.213 | 0.196 | 0.004 | 0.632 |
| Wavelet-SVM [30] | 0.475 | 0.297 | 0.064 | 1.010 |

providing accuracy results of both FRI and FTC models. The novel models explore whole transient data of angles and magnitudes of both current and voltage signals from pre- and post-fault cycles while current and voltage signals are measured through Phasor Measurement Units (PMUs), that are of high rates (high sampling frequencies/very low intervals), at different buses and used as input features to recurrent deep neural network models. These models rely on features self-extraction directly from the voltage and current input patterns over a period of time without the need for additional techniques to extract the features. The presented sequential learning algorithms extract maximum spatio-temporal information from the sequential features to model the system behavior. This ensures the highest classification and prediction accuracy and robustness. Both region identification and fault type classification models show high accuracy in fault detection and fault type classification, especially the greatest accuracy in identifying of the type of the fault. The accuracy and robustness of the results obtained from predictions of fault location distance have been discussed through a statistical study by both the mean and the standard deviation respectively of the prediction error of the fault location distance. The results show that the LSTM-based models are accurate, reliable and very effective for identifying, classifying, and locating the faults in power systems transmission lines. This contribution, of proposing three novel deep learning models working in tandem with each other, greatly improves the performance of maintenance strategies in large-scale power systems. In addition, the proposed models depend strongly on the advancement of synchro phasors and they will contribute a lot with the progress of synchro phasors. In future works, the proposed method will be tested on larger power systems.

## CRediT authorship contribution statement

**Soufiane Belagoune:** Conceptualization, Formal analysis, Investigation, Resources, Validation, Writing - original draft, Writing - review & editing. **Noureddine Bali:** Supervision, Project administration. **Azzeddine Bakdi:** Conceptualization, Formal analysis, Methodology, Resources, Software, Visualization, Writing - review & editing. **Bousaadia Baadji:** Data curation, Investigation, Validation, Visualization, Writing - review & editing. **Karim Atif:** Supervision, Project administration.

## Declaration of Competing Interest

The authors declared that there is no conflict of interest.

## References


[1] H. Fathabadi, Ultra high benefits system for electric energy saving and management of lighting energy in buildings, Energy Convers. Manag. 80 (2014) 543–549.

[2] A.J. Mazon, I. Zamora, J.F. Miñambres, M.A. Zorrozua, J.J. Barandiaran, K. Sagastabeitia, A new approach to fault location in two-terminal transmission lines using artificial neural networks, Electr. Power Syst. Res. 56 (3) (Dec. 2000) 261–266, https://doi.org/10.1016/S0378-7796(00)00122-X.

[3] S. El Safty, A. El-Zonkoly, Applying wavelet entropy principle in fault classification, Int. J. Electr. Power Energy Syst. 31 (10) (2009) 604–607, https://doi.org/10.1016/J.IJEPES.2009.06.003.

[4] T.A. Kawady, N.I. Elkalashy, A.E. Ibrahim, A.-M.I. Taalab, Arcing fault identification using combined Gabor Transform-neural network for transmission lines, Int. J. Electr. Power Energy Syst. 61 (2014) 248–258, https://doi.org/10.1016/J.IJEPES.2014.03.010.

[5] H. Fathabadi, Novel filter based ANN approach for short-circuit faults detection, classification and location in power transmission lines, Int. J. Electr. Power Energy Syst. 74 (2016) 374–383, https://doi.org/10.1016/J.IJEPES.2015.08.005.

[6] M. Farshad, J. Sadeh, Fault locating in high-voltage transmission lines based on harmonic components of one-end voltage using random forests, Iran. J. Electr. Electron. Eng. 9 (3) (2013) 158–166.

[7] Z. He, S. Lin, Y. Deng, X. Li, Q. Qian, A rough membership neural network approach for fault classification in transmission lines, Int. J. Electr. Power Energy Syst. 61 (2014) 429–439, https://doi.org/10.1016/J.IJEPES.2014.03.027.

[8] M.R. Aghamohammadi, M. Abedi, DT based intelligent predictor for out of step condition of generator by using PMU data, Int. J. Electr. Power Energy Syst. 99 (2018) 95–106.

[9] Y. Zhou, Q. Guo, H. Sun, Z. Yu, J. Wu, L. Hao, A novel data-driven approach for transient stability prediction of power systems considering the operational variability, Int. J. Electr. Power Energy Syst. 107 (2019) 379–394.

[10] S. Das, S.P. Singh, B.K. Panigrahi, Transmission line fault detection and location using Wide Area Measurements, Electr. Power Syst. Res. 151 (2017) 96–105, https://doi.org/10.1016/j.epsr.2017.05.025.

[11] X. Tong, H. Wen, A novel transmission line fault detection algorithm based on pilot impedance, Electr. Power Syst. Res. 179 (June) (2020), https://doi.org/10.1016/j.epsr.2019.106062.

[12] A. Ghaedi, M.E. Hamedani Golshan, M. Sanaye-Pasand, Transmission line fault location based on three-phase state estimation framework considering measurement chain error model, Electr. Power Syst. Res. 178 (October) (2020), https://doi.org/10.1016/j.epsr.2019.106048.

[13] J. Hu, A.V. Vasilakos, Energy big data analytics and security: challenges and opportunities, IEEE Trans. Smart Grid 7 (5) (2016) 2423–2436.

[14] Huisheng Wang and W. W. L. Keerthipala, 'Fuzzy-neuro approach to fault classification for transmission line protection,' IEEE Trans. Power Deliv., vol. 13, no. 4, pp. 1093 – 1104, Oct. 1998, doi : 10.1109/61.714467.

[15] K.M. Silva, B.A. Souza, N.S.D. Brito, Fault detection and classification in transmission lines based on wavelet transform and ANN, IEEE Trans. Power Deliv. 21 (4) (2006) 2058–2063, https://doi.org/10.1109/TPWRD.2006.876659.

[16] S.E. Cheng Hong, A B-Spline Wavelet Based Fault Classification Scheme for High Speed Protection Relaying, Electr. Mach. Power Syst. 28 (4) (2000) 313–324, https://doi.org/10.1080/073135600268289.

[17] F.E. Pérez, E. Orduña, G. Guidi, Adaptive wavelets applied to fault classification on transmission lines, IET Gener. Transm. Distrib. 5 (7) (2011) 694–702, https://doi.org/10.1049/iet-gtd.2010.0615.

[18] X. Lei, Z. Sui, Intelligent fault detection of high-voltage line based on the Faster R-CNN, Measurement 138 (2019) 379–385, https://doi.org/10.1016/j.measurement.2019.01.072.

[19] X. Qin, et al., A cable fault recognition method based on a deep belief network, Comput. Electr. Eng. 71 (2018) 452–464, https://doi.org/10.1016/j.compeleceng.2018.07.043.

[20] J. Liang, T. Jing, H. Niu, J. Wang, Two-Terminal Fault Location Method of Distribution Network Based on Adaptive Convolution Neural Network, IEEE Access 8 (2020) 54035–54043, https://doi.org/10.1109/ACCESS.2020.2980573.

[21] K. Chen, J. Hu, J. He, Detection and Classification of Transmission Line Faults Based on Unsupervised Feature Learning and Convolutional Sparse Autoencoder, IEEE Trans. Smart Grid 9 (3) (2018) 1748–1758, https://doi.org/10.1109/TSG.2016.2598881.

[22] M.N. Mahmud, M.N. Ibrahim, M.K. Osman, Z. Hussain, A robust transmission line fault classification scheme using class-dependent feature and 2-Tier multilayer perceptron network, Electr. Eng. 100 (2) (2018) 607–623, https://doi.org/10.1007/s00202-017-0531-5.

[23] A.Y. Appiah, X. Zhang, B.B.K. Ayawli, F. Kyeremeh, Long Short-Term Memory Networks Based Automatic Feature Extraction for Photovoltaic Array Fault Diagnosis, IEEE Access 7 (2019) 30089–30101, https://doi.org/10.1109/ACCESS.2019.2902949.

[24] J. Lei, C. Liu, D. Jiang, Fault diagnosis of wind turbine based on Long Short-term memory networks, Renew. Energy 133 (2019) 422–432, https://doi.org/10.1016/j.renene.2018.10.031.

[25] H. Zhao, S. Sun, B. Jin, Sequential Fault Diagnosis Based on LSTM Neural Network, IEEE Access 6 (2018), https://doi.org/10.1109/ACCESS.2018.2794765.

[26] T. de Bruin, K. Verbert, R. Babuška, Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks, IEEE Trans. Neural Netw. Learn. Syst. 28 (3) (2017) 523–533, https://doi.org/10.1109/TNNLS.2016.2551940.

[27] R. Fan, T. Yin, R. Huang, J. Lian, and S. Wang, 'Transmission Line Fault Location Using Deep Learning Techniques,' in 2019 North American Power Symposium (NAPS), Oct. 2019, pp. 1–5, doi: 10.1109/NAPS46351.2019.9000224.

[28] M. Li, Y. Yu, T. Ji, and Q. Wu, 'On-line Transmission Line Fault Classification using Long Short-Term Memory,' in 2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), Aug. 2019, pp. 513–518, doi: 10.1109/DEMPED.2019.8864831.

[29] A.G. Shaik, R.R.V. Pulipaka, A new wavelet based fault detection, classification and location in transmission lines, Int. J. Electr. Power Energy Syst. 64 (2015) 35–40, https://doi.org/10.1016/j.ijepes.2014.06.065.

[30] S. Ekici, Support Vector Machines for classification and locating faults on transmission lines, Appl. Soft Comput. 12 (6) (2012) 1650–1658, https://doi.org/10.1016/j.asoc.2012.02.011.

[31] P.K. Dash, S.R. Samantaray, G. Panda, Fault Classification and Section Identification of an Advanced Series-Compensated Transmission Line Using Support Vector Machine, IEEE Trans. Power Deliv. 22 (1) (2007) 67–73, https://doi.org/10.1109/TPWRD.2006.876695.

[32] W. Kong, Z.Y. Dong, Y. Jia, D. Hill, Y. Xu, Y. Zhang, Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network, IEEE Trans. Smart Grid PP (2017) 1, https://doi.org/10.1109/TSG.2017.2753802.

[33] P. (Prabha) Kundur, Power system stability and control. New York: McGraw-Hill.

[34] A. Jacobsson and C. Gustavsson, 'Prediction of the Number of Residue Contacts in Proteins Using LSTM Neural Networks,' no. January, pp. 1–61, 2003.

[35] U. Hobohm, M. Scharf, R. Schneider, C. Sander, Selection of representative protein data sets, Protein Sci. Publ. Protein Soc. 1 (3) (1992) 409–417, https://doi.org/10.1002/pro.5560010313.

[36] C. Brunner, G. Lang, F. Leconte, F. Steinhauser, Implementation guideline for digital interface to instrument transformers using IEC 61,850–9–2, Tech Rep (2004).

[37] W. Yang, '18 - Condition monitoring of offshore wind turbines,' in Offshore Wind Farms, C. Ng and L. Ran, Eds. Woodhead Publishing, 2016, pp. 543–572.
[38] K. Liu, G. Kang, N. Zhang, B. Hou, Breast cancer classification based on fully-connected layer first convolutional neural networks, IEEE Access 6 (2018) 23722–23732.
[39] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Process. 100 (2018) 439–453.
[40] R. Yu, Y. Wang, Z. Zou, and L. Wang, 'Convolutional neural networks with refined loss functions for the real-time crash risk analysis,' Transp. Res. Part C Emerg. Technol., vol. 119, p. 102,740, 2020.
[41] S. Fernández, A. Graves, J. Schmidhuber, 'Sequence labelling in structured domains with hierarchical recurrent neural networks', in *IN PROC. 20TH INT*, JOINT CONF. ON ARTIFICIAL INTELLIGENCE, IJCAI (2007) 774–779.
[42] A. Graves, S. Fernández, F. Gomez, Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks, Proceedings of the International Conference on Machine Learning (2006) 369–376.
[43] J. Wang, W. Wan, Optimization of fermentative hydrogen production process using genetic algorithm based on neural network and response surface methodology, Int. J. Hydrog. Energy 34 (1) (2009) 255–261.
[44] J. F. Kolen and S. C. Kremer, 'Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies,' Field Guide Dyn. Recurr. Netw., no. March 2003, 2010, doi: 10.1109/9780470544037.ch14.
[45] Y. H. Eom, J. W. Yoo, S. B. Hong, and M. S. Kim, 'Refrigerant charge fault detection method of air source heat pump system using convolutional neural network for energy saving,' Energy, vol. 187, p. 115,877, 2019.