Electrical Power and Energy Systems 133 (2021) 107102

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

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

A deep learning based intelligent approach in detection and classification of transmission line faults

Shahriar Rahman Fahim [a,*], Subrata K. Sarker [c], S.M. Muyeen [b], Sajal K. Das [c], Innocent Kamwa [d]

[a] Department of Electrical and Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[b] School of Electrical Engineering Computing and Mathematical Sciences, Curtin University, Perth, Australia
[c] Department of Mechatronics Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[d] Hydro-Québec Research Institute (IREQ), Varennes, Quebec, Canada

ARTICLE INFO

Keywords:
Capsule network
Discrete wavelet transform
Dynamic routing
Energy matrix
Sparse filtering

ABSTRACT

Detection and classification of transmission line (TL) faults are key factors for the fault root cause analysis and rapid restoration of the power network. Deep learning can extract representative features automatically from big data, thus becoming an important breakthrough in fault analysis. Existing deep learning-based fault diagnosis models rely on a large number of data obtained from a variety of fault conditions to be generalized. Indeed, in the TL domain, it is difficult to define and, therefore to collect the fault incepted features of all possible fault scenarios. This paper proposes an unsupervised framework for fault detection and classification (FDC) of TL based on a capsule network (CN). Instead of using the baseline CN, an extension to this with a sparse filtering technique is adopted in this work. The capsule network with sparse filtering (CNSF) voluntarily learns the expensive fault features and significantly improves the model performance without involving a large number of data. The proposed scheme receives 1/2 cycle post-fault three-phase signals and encodes them into a single image that is defined as the input for the proposed CNSF model. The effectiveness of the proposed CNSF model is corroborated by four different TL topology confirming the model's adaptability to a topology change in response to intended control action or the switching actions due to cascading faults. Further assessment of the model's performance against noise, high impedance faults (HIF), and line parameter variations are also carried out in order to confirm the high reliability of the proposed model. In addition, a rigorous comparative study is conducted to guarantee the state-of-the-art performance of the proposed model.

1. Introduction

An electric power system network (EPSN) is consists of three fundamental zones: the generation, the transmission, and the distribution [1]. Because of the growing electric power demand, the EPSN is approaching to become more complex and susceptible to electrical disturbances or faults [2]. Amid the entire EPSN, the transmission lines (TL) are subjected to 80% of power system faults as they exposed to various environmental, as well as animal or human contacts [3,4]. The majority of the faults that the TL experiences are the short circuit faults [5]. During a fault event, the voltage and the current signals of a three-phase TL deviate from their reference values and result in catastrophic consequences if not cleared in time [6]. Thus, fault analysis has become an influential research platform for power engineers.

To assist in the maintenance and reparation of the power network, a protection system capable of accurate detecting, classifying, positioning, and eliminating the faults is evident [7]. Despite a number of approaches, consolidating all the features of the protection system into a single framework is still challenging [8]. Therefore, this work is concerned with specialized intelligent FDC tasks for achieving higher noise immune performance and reliability. It should be noted that identifying the fault location is out of the scope of this paper.

1.1. Literature review

In the prior literature, the experts analyze the pre and post-fault signals to activate a protection system. The post fault voltage and current (V&C) waveforms of a TL contain valuable signature information for short circuit faults. Some early attempts in the direction of FDC set some preset values of these waveforms to declare a fault type [9,10]. The

* Corresponding author.
E-mail address: sr.fahim@ieee.org (S.R. Fahim).

https://doi.org/10.1016/j.ijepes.2021.107102
Received 4 October 2020; Received in revised form 21 January 2021; Accepted 10 April 2021
Available online 18 June 2021

literature has also presented some proposals using signal processing based methods that use a set of mathematical operation for analyzing the signals in different natures. These methods can be used directly for fault analysis. As an example of a fault analysis scheme that follows this theme, the work proposed in [11] uses wavelets for the FDC purpose. In addition, some approaches combine the Stockwell transform either with Hilbert transform [12,13] or Wigner distribution function [14] for fault recognition. Furthermore, a transmission line protection scheme is reported in [15], which uses the Wigner distribution function and alienation coefficient. A similar approach is presented for the protection of power system with solar energy penetration [16]. The blend of alienation coefficient and Wigner distribution function can also be used for the protection of the hybrid power system network [17]. These approaches require to analyze the transform components and/or coefficients with a set of heuristically defined rules for fault analysis. Besides, the decision thresholds are needed to be defined. Defining the conditions in a power system is tricky because various parameters may produce similar results and the quantity of possible parameter values is innumerable. Thus, if the working condition is changed, it is difficult to make use of these methods.

Other than the direct analysis of faults with the signal processing based method, these methods can also be used for feature extraction purposes. In later, the FDC algorithms become complex and require a feature extraction process to synthesize the essential information from the raw signals [18]. The feature extraction process purposefully digs out the underlying message bought by the fault transients. With the feature extraction technique, the FDC algorithms become more accurate and robust than they were before [19].

With the evaluation of FDC algorithms in the power TL, the feature extraction tools also witness a similar development. When a fault occurs, the frequency characteristics of V&C waveforms within various frequency bands experience a sharp change [20]. A basic mathematical tool to analyze these changes in the frequency domain is the discrete Fourier transform (DFT). The methods based on DFT use half-cycle DFT [21] or full-cycle DFT [22] to estimate the phasors of faulty signals for classifying the faults. Among the signal processing based fault feature extraction techniques, the discrete wavelet transform (DWT) is one of the most popular ones [19,23,24]. The DWT decomposes a waveform into the approximation coefficient at one level and detail coefficient at multiple levels. In [23] authors use the wavelet coefficient at the detail level to generate the fault features. Besides, a combination of WT and chebysev neural network (ChNN) can also be used as a feature [25].

An alternative way of using the coefficients is to estimate the energy content at the detail level coefficient [26]. Furthermore, Shanon's entropy with the wavelet transform is also considered as an effective measure of representing the features [27]. As a further improvement, a combination of the different features mined from the wavelet coefficient is introduced in [28], which contributes greatly to the classification accuracy. These aforementioned methods return the fault features as a numerical value which may gain desirable performance to some extent but lack the generalization of performance. With the idea of deploying recent computer vision and pattern recognition algorithms into fault analysis tasks, representing the signals into visual clues has raised much attention [29,30]. The image representation of the data can reveal more expensive fault features which in turn confirms the generalization of classification performance.

After representing the raw faulty signals into some set of rules by the feature extraction technique, the FDC task is triggered. The detection stage can be implemented prior to the classification stage [31] or concurrently [29]. Considering the feature learning stage in the classification model, there presents no essential difference whether the detection stage triggers earlier or not.

Due to the various uncertainties in the TL domain, computational intelligence-based techniques have become a powerful candidate for the deployment in FDC because of their strong generalization ability [8]. An approach in this direction is mentioned in [32], that uses the feed-forward neural network (FNN) to detect the TL disturbances. Thereafter, replacing the FNN's sigmoid activation function with the radial basis function (RBF), another network is introduced to build a good classification effect [19]. Like the RBF network, another version of the FNN is the probabilistic neural network (PNN) [33]. The PNN uses an exponential activation function and can be trained faster as well as achieve 10% higher accuracy than that of the FNN [34]. In addition, an ultrafast DWT aided FDC method is illustrated in [35]. The neural network models mentioned in these approaches have shallow architecture and thus they can not maximize the benefit of multiple non-linear features with perfection due to the shortage of hidden layers. In the TL domain, the fuzzy-logic based implementations make it adequate to deal with the non-linearity and uncertainty [36]. Though the fuzzy-based implementations deliver a desirable outcome, their performance depends on expert's interaction which is not available all the time.

With the transmission and distribution system networks stepping into the new era of development, the FDC method having more robustness, and the discriminative property is required. In need of these properties, the autoencoder (AE), an unsupervised feature extraction model is introduced in [37,29] for FDC. The AE represents the input to a latent space and regenerates it from this representation with its single layer [38] or stacked-layer [37] structure. Unlike with the AE, the deep belief network can also deal with the non-linearity and can be efficiently applied in the FDC task [39,40]. Another data-driven FDC method is introduced in [41], which adopts long and short term memory to achieve robust classification performance. Although these models are well developed and can perform effectively in the power TL domain, the challenge is still remaining as they hardly capture the position information.

Addressing the problems of uncertainty in the power TL domain, the convolutional neural network (CNN) is introduced in some state-of-the-art literature [42,30]. The CNN based FDC model exhibits impressive performance in fault classification by efficiently extracting the local features from the time series image. However, straightforward CNN architecture still struggles while dealing with the limited amount of data that needs to be considered in the case of the power TL domain. As the nature of the faults are unpredictable and the TL parameters are prone to change, the CNN requires a big data set to deal with this situation. Deepening the CNN model for FDC may mitigate this limitation, this still imposes challenges as this is hard to converge which in turn limits the classification performance.

Thus, training a model with a large dataset for the generalization is not a practical solution. This time, the capsule network (CN) based implementation comes in handy [43]. The "capsule" refers to an activity vector whose length offers the probability of a class's existence [44]. The CN also utilizes the dynamic routing between the capsules for passing the information from one layer to another rather than the max-pooling layer on CNN. The max-pooling layer strips the spatial relation between the parts while the CN significantly improves the discriminative property. To improve the model, some modifications on CN is done earlier [45,46]. In turn, instead of using the straightforward CN, the proposed FDC model utilizes a capsule network with sparse filtering (CNSF). Besides taking the advantages of baseline CN, the sparse filtering offers more robust performance under a low signal to noise ratio (SNR).

1.2. Contribution

The main contributions of this research are as follows,

- Proposed a new robust intelligent approach for power TL FDC based on the CN combined with the sparse filtering for insight diagnosis of TL faults.
- An unsupervised model is developed, which requires no manual labeling of the data for training and testing.

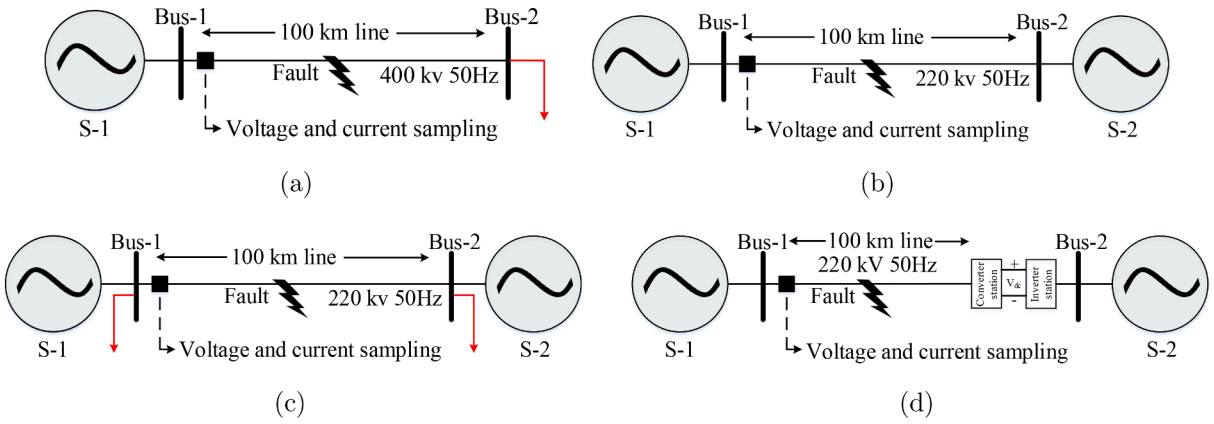

Fig. 1. Studied transmission line systems (a) TL-1, (b) TL-2, (c) TL-3, (d) TL-4.

Table 1
System parameters for considered transmission lines.

	System parameters	Types or Values			
		TL-1	TL-2	TL-3	TL-4
Fault	Types of fault	a-g, b-g, c-g, ab-g, bc-g, ac-g, ab, bc, ac, abc-g, non-faulty			
	Fault distance (km)	1 to 99			
	Fault resistance (Ω)	0.1, 1, 10, 50, 100			
	Fault inception angle (degree)	0 to 180 in increment of 30			
Transmission line	Positive and zero sequence resistances (Ω/km)	0.01273 and 0.3864			
	Positive and zero sequence inductances (H/km)	0.9337e-3 and 4.1264e-3			
	Positive and zero sequence capacitances (F/km)	12.74e-9 and 7.751e-9			
Source	Phase to phase voltage	400 kV	220 kV	220 kV	215 kV
	Base voltage	400 kV	220 kV	220 kV	215 kV
	Base power	60 MVA			
Load	Active power (W)	200 e^6/100e^6	nil	100 e^6	nil

- The performance of the proposed model against line parameter variation and faults with high resistance is considered to count the effectiveness.
- Noises are considered to justify the robustness of the proposed CNSF model.
- The performance of the proposed CNSF model is compared with the baseline CN and existing state-of-the-art FDC methods to clarify the superior performance in the case of fault diagnosis.

2. Test systems

For benchmarking the CNSF based FDC scheme, four types of power TL models are simulated and investigated on the MATLAB/Simulink environment. Switching the network topology is a common industry practice to improve operational efficiency. As a result, the FDC model performance may be affected due to the change in network short-circuit values following a topology switch [47]. Thus, in order to make the proposed FDC model performable over the topology change, four

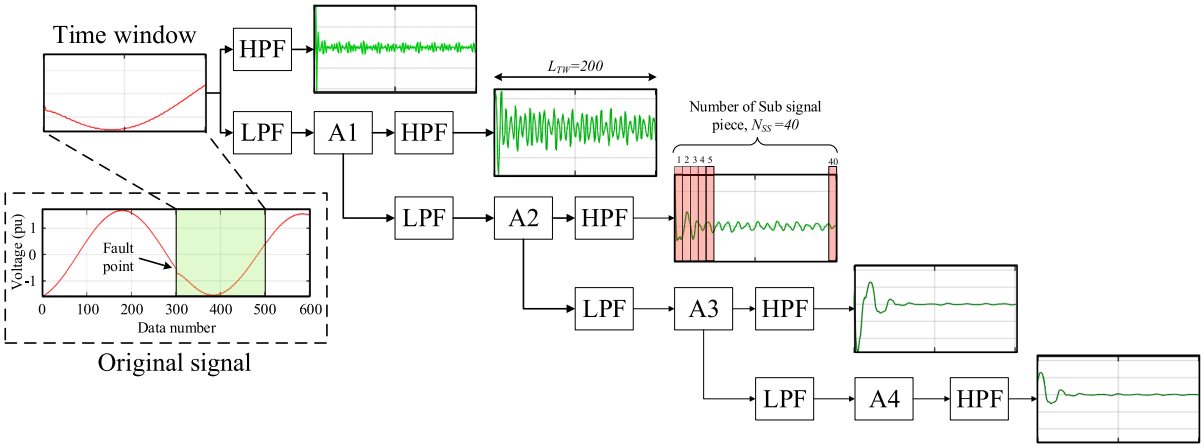

Fig. 2. Analyzing procedure of original time series signal with wavelet transform.

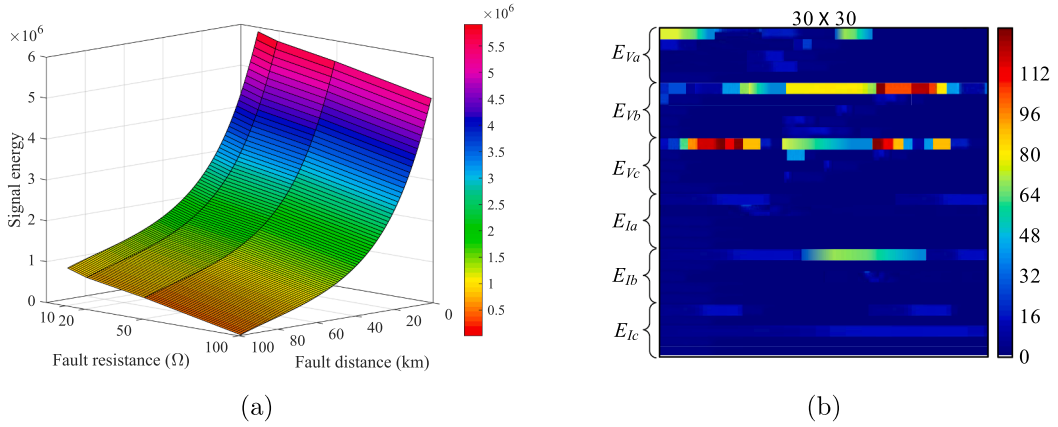

Fig. 3. (a) 3D-surface showing the changes in energy with fault distance and fault resistance, (b) Finalized wavelet detail level coefficient energy matrix.

common topologies are considered as they contain basic power system components, i.e., loads, buses, sources, and transmission lines [31]. The single line diagram of the investigated power TL models is illustrated in Fig. 1. The considered TL models are denoted as "TL-1", "TL-2", "TL-3", and "TL-4" as shown in Fig. 1(a), (b), (c), and (d), respectively. The TL models TL-1, TL-2, and TL-3 are set for the variable, heavy, and no loading conditions. The TL-4 is a 500 kV DC (HVDC) system with a converter utilized in [48]. The short circuit faults involving the AC TLs are only studied in this paper to assess the CNSF performance. All the considered systems are simulated against the variations of system parameters as listed in Table 1 for ten different types of short circuit faults and non-faulty conditions. In Table 1, the a, b, and c are three phases of the power line and g denotes the ground. The V&C signals from the phases are sampled at the vicinity of bus-1 at a 20 kHz sampling resolution.

3. Data processing

The three-phase waveforms sampled from a sub-station during a fault event is the signature of that event. This indicates that the exact phase causing the fault event can be determined using the signature information. In essence, the signatures are the function of the type of fault events, line parameters, and network topology. Considering the raw three-phase waveforms are difficult as the raw signals contain a large volume of sampled data points. Thus, to generate the most effective input features, this study converts the raw time-domain transients into a wavelet detail level coefficient energy (WDLCE) image via the wavelet transform (WT). The detailed process of the WT on the three-phase waveform is shown in Fig. 2. The signal energy [28] of three-phase V&C transients are unique to the line parameters. For example, the variation of the phase-a signal energy against the fault distances and fault resistance during "a-g" fault for TL-2 is conferred in Fig. 3(a) using a 3-D surface. A similar result can also be achieved for the rest of the studied system due to the variation of parameters reported in Table 1. Intuitively, for the detection and classification problem to be solved in this study, the post fault half power-frequency cycle (i.e., 200 data points at 20 kHz sampling frequency and 50 Hz system frequency) waveforms are considered. Longer waveforms are not considered here as it increases the training period for the model.

3.1. Wavelet transform analysis

This study analyzes the faulty V&C transients as they are the main components to declare a fault event. The considered transients are non-stationary which obviates the use of a signal processing tool that can handle the time-frequency localization of a fault transient. This, in turn, the WT comes in handy which decomposes a fault transient into a set of components. Each wavelet components portrays a time-range transient that covers a distinct frequency band representing information in a more detailed manner. Practically the WT is implemented using the DWT [28]. In DWT, the digital filters of different threshold frequencies are utilized to analyze a non-stationary transient at different scales. A time-domain signal $S(t)$ is passed through a chain of high-pass filter (HPF) to decompose into approximations (A) using father wavelet $\delta_m(t)$. Likewise, the mother wavelet (MW) $\sigma_m(t)$ decomposes the signal into detailed coefficients through a chain of a low-pass filter (LPF).

$$\delta_{mk}(t) = 2^{-\frac{m}{2}}\delta\left(2^{-j}t - n\right), \sigma_{mk}(t) = 2^{-\frac{m}{2}}\sigma\left(2^{-j}t - n\right). \quad (1)$$

Here, $n \in Z$. In the prior literature, the Daubechies (db) MW is found to be one of the most suitable ones for the TL signal analysis. Thus, the db MW at $4th$ order ($db4$) has been chosen for the implementation. Because of not having the discrete implementation [25], the *Marr* (Mexican hat), *Meyer*, and *Morlet* wavelets are not studied in this work. The m and k in (1) are integers, n denotes the units of time by which the functions are translated to and 2^m is the factor by which the functions are scaled. The filters which are given rise to two-scale functions are:

$$\delta(t) = \sum_n h(n)\sqrt{2}\delta(2t - n), \sigma(t) = \sum_n g(n)\sqrt{2}\sigma(2t - n). \quad (2)$$

where $h()$ and $g()$ are the filter coefficients. If M decomposition level is considered, the signal using wavelet transform can be expressed as:

$$S(t) = \sum_{k=0}^{2^{M-m-1}} a_{mk}\delta_{mk}(t) + \sum_m \sum_{k=0}^{2^{M-m-1}} d_{mk}\sigma_{mk}(t). \quad (3)$$

The output from HPFs and LPFs are down-sampled by a factor of 2 and give rise to approximation a_{mk} and detail coefficients d_{mk} at j decomposition level. The aforementioned process is iterated till the A and D coefficient at level five is retrieved. The wavelet decomposition tree up to level five ($M = 5$) for the input signal $S(t)$ is shown in Fig. 2.

3.2. Generating wavelet detail level coefficient energy (WDLCE) image

To generate the necessary training and testing datasets, the signal segments correspond to the same time scale are analyzed. In this light, the post-fault half-cycle signal is cut off from the original simulated signal as the most important information lies at the early stage of faults. For 50 Hz system frequency and 20 kHz sampling resolution, the number of data within a full power frequency cycle is 400. Thus, the length of the considered time window, $N = 200$ for the half-power frequency cycle. From the wavelet decomposition tree, only the data points at the detail level coefficients are considered in this study. The data points in each detail coefficients are split into N_{ss} equal pieces referred to as sub-signal pieces along the time axis. Therefore, each sub-signal piece contains P

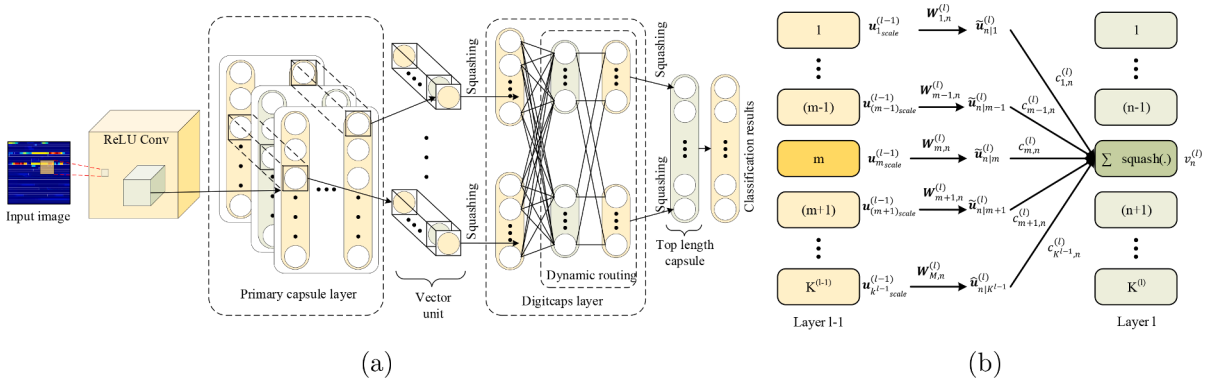

Fig. 4. (a) Capsule network architecture, (b) Dynamic routing between capsule.

data points where $P = \frac{N}{N_{ss}}$. Then, the energy content in each sub-signal piece is, $e_{dl} = \sum_{c=1}^{P} |a_c|^2$. Where, $l = 1, 2, 3, \ldots, N_{ss}$ at d^{th} $(d = 1, 2, 3, \ldots, M)$ decomposition level. Using the energy of each sub-signal piece, the WDLCE matrix, E for a single signal can be constructed as defined in (4).

$$E = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1l} \\ e_{21} & e_{22} & \cdots & e_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ e_{d1} & e_{d2} & \cdots & e_{dl} \end{pmatrix}, \quad E_f = \begin{pmatrix} E_{V_a} \\ E_{V_b} \\ E_{V_c} \\ E_{I_a} \\ E_{I_b} \\ E_{I_c} \end{pmatrix}. \tag{4}$$

In this study, the variables used to construct the matrix are assumed to $M = 5, N = 200, N_{ss} = 40, P = 5$. The dimension of the WDLCE matrix for each time window of the three-phase signal is 5×40. The complete process of converting a time-series signal into the WDLCE matrix is depicted in Fig. 2. Six WDLCE matrices of V&C signals are merged to construct the full energy matrix E_f with the dimension of 30×40 as defined in (4).

Where E_{V_a}, E_{V_b}, and E_{V_c} denote the WDLCE matrices of three-phase voltages and E_{I_a} to E_{I_c} represents the same for three-phase line currents, respectively. For convenience, the matrix E_f is resized to 30×30. The first three and last seven columns of the matrix E_f are dropped for this operation. Under various fault conditions, the maximum value of the matrices E_{V_a} to E_{I_c} is found out. Thereafter, the elements of these matrices are normalized following the maximum obtained value ensuring that they fall within the range 0 to 1. The final image representation of the time-series signal is then obtained by transferring the normalized matrix elements to a proportional color value. As an illustration, the final WDLCE image for the"a-g" fault with the color scale is shown in Fig. 3(b).

For every specific fault type, the combination of the system parameters mentioned in Table 1 creates 3465 distinct fault cases for each power transmission system referred to in Fig. 1. Therefore, for all ten types of fault and non-fault condition, a sum total of 38,115 fault cases which implies 38,115 WDLCE images have been generated. A learning stage is a prerequisite to the classification task for the proposed FDC algorithm. Besides, for justifying its robust performance, the proposed model needs to be tested with the fault cases other than the training cases. With this intention, the total generated data are split into training and testing set with a percentage ratio of 70:30. It is worth noticing that the training pattern is completely different from the pattern used for testing. The performance of the proposed FDC algorithm with the testing dataset is discussed in Section 5.

4. Proposed FDC framework

This section overviews the details of the proposed CNSF based FDC method. The proposed paradigm of the CN framework is illustrated in Fig. 4(a). A CN has a three-layer architecture: the convolutional layer, the primary-capsule layer, and the digit-capsule layer. This three-layer architecture utilizes the overall length of the capsule (or a group of neurons) to delegate the probability of classification and the direction of the capsule (vector) to designate a list of class's properties. The CN attempts to overcome the notions and limits associated with the CNN such as losing valuable fault information in pooling operation.

The capsule network begins with a classical convolutional layer whose aim is to arrange the WDLCE image into features and fed it to the succeeding primary capsule layer. The convolutional layer generates the patches $p_{i,j} \in \mathbb{R}^{d \times d \times C}$ plucked from the original time-series image by applying a kernel of size $k^{(1)} \times k^{(1)} \times q^{(1)}$ (where, $q^{(1)} = C$). This process is followed by a non-linear batch normalization stage to speed up the convergence using a leaky rectified linear unit (leaky RelU) in the convolutional channel as,

$$f(a) = \begin{cases} a & \text{if } a \geqslant 0 \\ \gamma a & \text{if } a < 0 \end{cases} \tag{5}$$

Here $\gamma = 0.01$. Thus, the total output volume $\mathbf{O}^{(1)} \in \mathbb{R}^{H^{(1)} \times W^{(1)} \times K^{(1)}}$ from CN layer-1 is composed of $K^{(1)}$ feature map of size $H^{(1)} \times W^{(1)}$ (*height* × *width*).

The working mode of the second layer of the CN or the primary capsule layer is quite similar to the CNN's kernel operation yet it shows special behavior that is different from the traditional convolution process. The primary capsule-layer can be considered as a babushka doll analogy, where the primary capsule layer is made up of $K^{(2)}$ convolutional capsules which, in addition, are a set of $Z^{(2)}$ convolutional unit of size $k^{(2)} \times k^{(1)} \times q^{(1)}$. Where $k^{(1)} = q^{(2)}$. The primary capsule layer replaces the scalar convolutional units to vector $\mathbf{u}_m^{(2)} \in \mathbb{R}^{Z^{(2)}} = [u_{m,1}^{(2)}, u_{m,2}^{(2)}, \ldots, u_{m,Z^{(2)}}^{(2)}]$ by applying $Z^{(2)}$ units of m^{th} capsule over $O^{(1)}$ volume. The intension behind this vector representation is to store additional detailed information regarding the features that improve the reliability of a neural network. Hence, each individual element of $\mathbf{u}_m^{(2)}$ signifies different attributes of the same entity [44].

An interesting aspect of these set of the neuron is that it detects features as well as detects its variants, in that enrich the capsule network with equivariance properties. Thus, even a slight variation in the input will result in a slight variation in the respective capsule's output. The orientation and length of the m^{th} layer activity vector $\mathbf{u}_m^{(l)}$ at whatever layer l are equally important as they represent the properties and probability of a class's existence. In order to adequately portray this property, the activity vector is scaled down to $\mathbf{u}_{m_{scale}}^{(l)}$ in between 0 to 1. A non-linear squashing function as in (6) performs this scaling operation without affecting the orientation of the vectors.

$$\mathbf{u}_{m_{scale}}^{(l)} = \frac{\left\|\mathbf{u}_m^{(l)}\right\|^2}{\varepsilon + \left\|\mathbf{u}_m^{(l)}\right\|^2} \frac{\mathbf{u}_m^{(l)}}{\left\|\mathbf{u}_m^{(l)}\right\|} \tag{6}$$

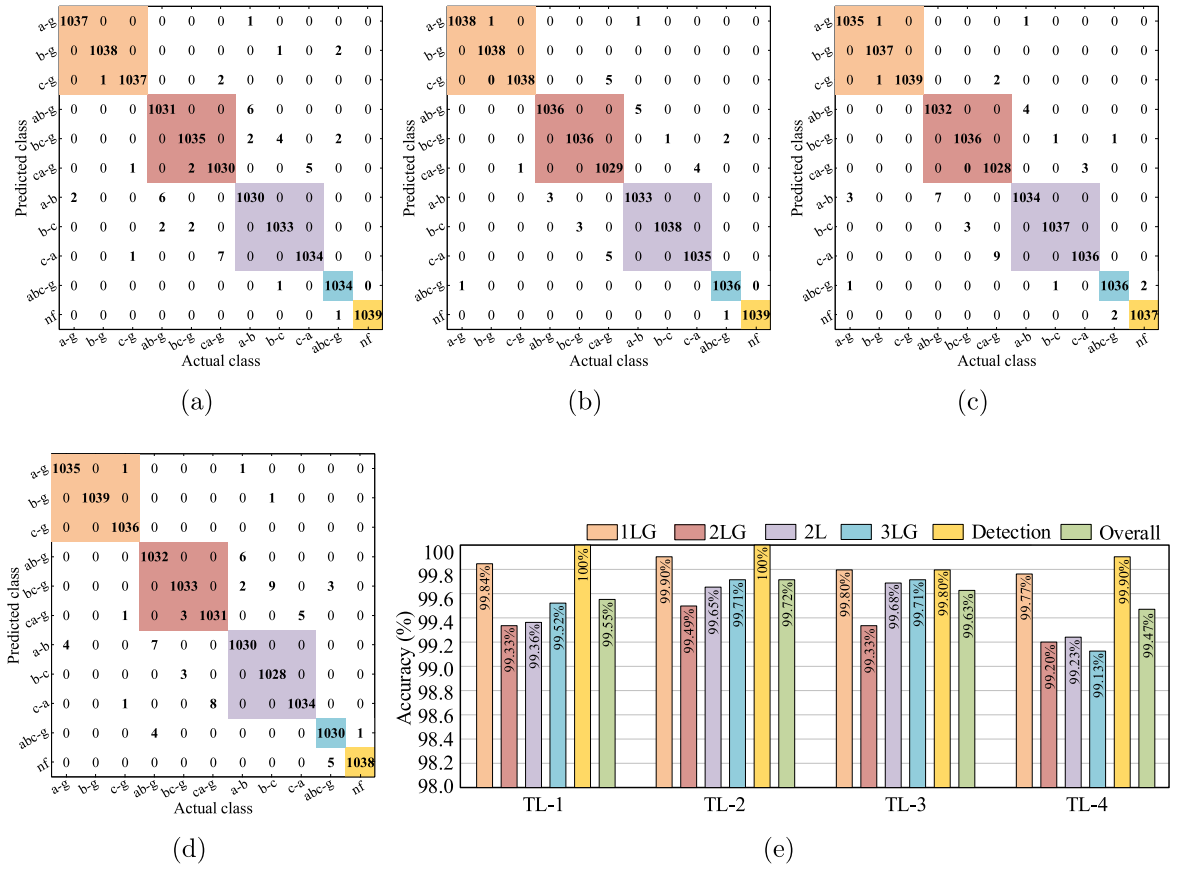

Fig. 5. Confusion matrix for (a) TL-1, (b) TL-2, (c) TL-3, (d) TL-4, and (e) overall accuracies reported from the confusion matrix.

Here ε is the factor with respect to which short vectors get scaled to almost 0 and the long vectors get scaled to almost 1 and come in focus.

The third layer of the CN is formed by n_{class} capsules, where $n_{class} = 11$ (number of total fault class and non-fault condition to be classified). For a distinct fault class, there presents an activity vector. The probability of each input patch associated with a fault class will be encoded by the module of this activity vector. The prediction vector $\tilde{\mathbf{u}}_{n|m}^{(l)}$ can be expressed as the weighted multiplication amid the output of the m^{th} capsule at layer $l-1$ and its corresponding transformation matrix $\mathbf{W}_{m|n}^{(l)}$ and adding the biases $B_n^{(l)}$ of n^{th} capsule at the current layer l as illustrated in (7) with $m = 1,2,3,\ldots,K^{(l-1)}$.

$$\tilde{\mathbf{u}}_{n|m}^{(l)} = \mathbf{W}_{m|n}^{(l)}\mathbf{u}_{m_{scale}}^{(l-1)} + B_n^{(l)} \tag{7}$$

In that, the former primary capsule layer connects with the current digit-capsule layer. Moreover, as observed from (7) the prediction vector $\tilde{\mathbf{u}}_{n|m}^{(l)}$ is a prediction of m^{th} capsule at $l-1$ layer about the activity vector of n^{th} capsule at l layer. This process allows that the higher-levels capsules can be predicted with the capsules at lower levels. For the later, the concept of "*agreement*" has been emerged were the numerous predictions agree at different levels, their mutual connections become stronger and a higher level capsule got activation for a more complex feature. This concept is reinforced by coupling-coefficient $\tilde{\mathbf{c}}_{m,n}^{(l)}$ which is a dynamic routing element that calculates final input $s_n^{(l)}$ of n^{th} capsule at layer l and relates it with the m^{th} capsule at layer $l-1$ as:

$$s_n^{(l)} = \sum_{m}^{K^{(l-1)}} \tilde{\mathbf{c}}_{m,n}^{(l)}\tilde{\mathbf{u}}_{n|m}^{(l)} \tag{8}$$

Again, $s_n^{(l)}$ is squashed by utilizing (6) so as to obtain the final activity vector $v_n^{(l)}$, where the vector's length expresses the probability of the target class. Directing the attention again on the dynamic routing element $\tilde{\mathbf{c}}_{m,n}^{(l)}$, this parameter initializes with the same probability for all connections amid the n^{th} capsule in l and m^{th} capsule in $l-1$. Thus, all the dynamic routing elements of m^{th} capsule must be summed as 1 and can be calculated by the routing softmax as:

$$\tilde{\mathbf{c}}_{m,n}^{(l)} = \frac{exp\left(\tilde{b}_{m,n}\right)}{\sum_i^{K^{(l)}} exp\left(\tilde{b}_{m,i}\right)} \tag{9}$$

where, the expression $\sum_i^{K^{(l)}}\tilde{\mathbf{c}}_{m,i}^{(l)} = 1$ must satisfies. Also, $\tilde{b}_{m,n}$= log-prior probability of m^{th} capsule will activate n^{th} capsule. $\tilde{b}_{m,n}$ initializes from 0 and updated in each iteration as follows:

$$\tilde{b}_{m,n(i)} \leftarrow \tilde{b}_{m,n(i-1)} + \tilde{a}_{m,n(i-1)} \tag{10}$$

Here, $\tilde{a}_{m,n_{(i-1)}}$ denotes the degree of agreement or $(\tilde{\mathbf{v}}_n^{(l)}.\ \tilde{\mathbf{u}}_{n|m}^{(l)})$ in the previous iteration $i-1$. If $\tilde{\mathbf{v}}_n^{(l)}$ and $\tilde{\mathbf{u}}_{n|m}^{(l)}$ are in agreement their scalar product will be 1. Conceptually it implies that the capsules at a layer can predict the capsules at a higher layer. The capsules with similar results are grouped together through the dynamic routing process as illustrated in Fig. 4(b) to obtain a more clear output.

For each input set, the capsule network receives a collection of n number of classes n_{class} activity vectors, where $\tilde{\mathbf{v}}_C^{(l)}$ denotes the capsule for class C and its modulus $||\tilde{\mathbf{v}}_C^{(l)}||$indicates the probability of a class's existence. In order to tune the output and increase the goodness of the capsule network the margin loss L_m is chosen as the loss function as

defined in (11).

$$L_m = \sum_{C}^{n_{class}} T_C max\left(0, m^+ - \left\|\tilde{\mathbf{v}}_C^{(l)}\right\|\right)^2 + \lambda(1 - T_C) max\left(0, \left\|\tilde{\mathbf{v}}_C^{(l)}\right\| - m^-\right)^2 \tag{11}$$

where, λ is a regularization parameter valued as 0.5. In addition m^+ and m^- is upper and lower boundary of probability valued as 0.9 and 0.1.

The sparse filtering as an unsupervised feature-learning algorithm combined with the capsule network shows superior performance over the straightforward capsule network [45]. This filtering operation optimizes a filter's output applied to the prior layer which, in turn, expresses a good intra and inter-sample sparsity as well as the regularity of features through the blend of L_1 and L_2 norm. The fundamental operation of the sparse capsule is similar to the baseline capsule network except in the process of updating the weight matrix. The baseline capsule network employs the backpropagation algorithm for this purpose, in there the sparse capsule replaces it with a sparse filtering technique. In this technique, each feature is normalized by dividing it with its L_2 norm and normalizing the row R of the feature matrix and column O of the same matrix can be defined as:

$$R = \frac{\widetilde{\mathbf{u}}_{n|m}}{||\widetilde{\mathbf{u}}_{n|m}||_2}, O = \frac{R}{||R||_2} \tag{12}$$

With the matrix normalization, each feature normalizes into the activation value and falls within the sphere surface of two norms. For $M \times N$ dimensional image the objective function could be defined by:

$$\min_{arg\mathbf{W}_{m|n}} L(\mathbf{W}_{m|n}) = \sum_{p=1}^{N} \sum_{q=1}^{M} O \tag{13}$$

Solving (13) for $\mathbf{W}_{m|n}$, the weight matrix in (7) can be updated with the sparse filtering technique.

5. Results of detection and classification

This section assesses the performance of the proposed fault detection and classification framework. Besides reporting the overall accuracy, a precision and sensitivity analysis of results are also carried out in this section to better evaluate the CNSF model performance.

5.1. Overall accuracy

The task of FDC can be done after generating a sufficient amount of training and testing samples by the time series imaging technique. For bringing the simplicity in operation, instead of performing the FDC separately, they are performed concurrently. In this way, the non-faulty (nf) condition is considered as a fault class. If the system output switches to a distinct fault class other than the "nf", a fault is supposed to be triggered.

The performance of the proposed CNSF classification model on the test dataset is shown in Fig. 5 with the confusion matrix for all TL topologies. The confusion matrix tells how the model confuses within the fault classes. The diagonal values of this matrix indicate the number of correctly classified classes and the non-diagonal values designate the number of miss classified classes. As the test data are taken by shuffling the main dataset and not previously been used for training the model, the confusion matrix reflects the classification results for different types of unseen data. The confusion matrices in Fig. 5 show that for the considered dataset the large majority of the fault data are classified correctly. The lowest inter-class miss-classification is observed for the no-load condition (TL-2). For the heavily loaded condition (TL-3), the number of miss-classification increased slightly. During the variable loading condition, the transmission line experiences voltage variation and thus a decrement in classification performance is observed for TL-1. Compared to the other three models, the TL-4 performs poorly as the TL with converters shows non-linear characteristics. The overall accuracies (OAs) for all four test systems are calculated as:

$$OA = \frac{N_{correct}}{N_{total}} \times 100\% \tag{14}$$

Here, $N_{correct}$ and N_{total} denote the number of truly classified class and number of total class respectively. The OAs reported in Fig. 5(e) are taken by repeatedly performing the training and testing for five times and averaging the values. The reported OAs for all the studied system is satisfactory, as truly classified about more than 99% of the fault cases in the test dataset. In order to better analyze how the proposed classifier confuses the fault type, the OAs for different categories are also calculated from the confusion matrix using (14) and reported in Fig. 5(e). As can seen from the confusion matrix, the proposed model confuses more between 2-line-ground (2LG) and line-line (2L) faults. One reason for this is that two phases are involved in both cases and, to some extent, erroneously plots in another fault category. Also, as the sound condition and 3-line-ground (3LG) fault conditions are balanced one, a negligible miss-classification between them is observed. In the rest of the cases, the miss-classification is witnessed where similar phases are involved. Furthermore, the classification accuracy for the non-fault condition is considered as the detection accuracy as reported in Fig. 5(e) for all four TL model. Nevertheless, the OAs show that the proposed FDC framework is proficient in detecting and classifying the faulty waveforms effectively.

5.2. Precision and sensitivity analysis

In practice, some faults are very common in power lines and some fault events are rare. Thus, the data obtained for different fault classes during the training stage is not equal. Therefore, the FDC task is an imbalanced classification problem as the number of data for each fault

Table 2
Precision, sensitivity, and F1-scores for CNSF based FDC scheme.

Fault	TL-1			TL-2			TL-3			TL-4		
class	Pr	Se	F1	Pr	Se	F1	Pr	Se	F1	Pr	Se	F1
a-g	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
b-g	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
c-g	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ab-g	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
bc-g	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
ac-g	0.99	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99
a-b	0.99	0.99	0.99	1.00	0.99	1.00	0.99	1.00	0.99	0.99	0.99	0.99
b-c	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	0.99
a-c	0.99	1.00	0.99	1.00	1.00	1.00	0.99	1.00	0.99	0.99	1.00	0.99
abc-g	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
nf	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3
Average fault detection time for different transmission lines.

TL Model	Detection time (cycle)									
	a-g	b-g	c-g	ab-g	bc-g	ca-g	a-b	b-c	c-a	abc-g
TL-1	0.53	0.52	0.53	0.54	0.54	0.55	0.52	0.52	0.53	0.57
TL-2	0.54	0.54	0.55	0.57	0.55	0.56	0.52	0.53	0.54	0.60
TL-3	0.53	0.53	0.53	0.57	0.55	0.55	0.54	0.52	0.52	0.58
TL-4	0.55	0.55	0.53	0.57	0.58	0.57	0.56	0.56	0.58	0.64

class is not equal. In this case, the overwhelming majority of a distinct fault class drives the accuracy of the classifier. In order for a meaningful evaluation of performance, precision (Pr), and sensitivity (Se) analysis are also conducted. Both the precision and sensitivity are based on the measure of relevance. The precision can be defined as (15), which expresses the portion of the relevant data point among the retrieved data points. The sensitivity (as defined in (15)) indicates the ratio of the amount of the model's correct prediction to the total number of true samples.

$$Precision = \frac{TP}{TP + FP}, Sensitivity = \frac{TP}{TP + FN}. \tag{15}$$

Here, the false positives (FPs) are the actual negative cases the model declares as positive. The true positive (TP) indicates the amount of data declared as positive by the model which are actually positive and the false positives (FPs) are the actual negative cases the model declares as positive. Individually maximizing the precision causes the expense of the sensitivity or vice versa. Thus, with the aim of finding the optimal blend of these two measures the f1-measure is calculated. Ranging between 0 to 1, the f1-measure indicates the harmonic mean of both precision and sensitivity. The higher f1-measure designates the better classification ability of the model.

$$f1 - measure = \frac{2 \times precision \times sensitivity}{precision + sensitivity} \tag{16}$$

The scores for precision, recall, and f1-measure (F1) at the first trial for different types of faults incepted on the different power system models are listed in Table 2. The table indicates that the model shows the maximum f1-score for different fault data. In addition to that, the average fault detection time for online implementation of the proposed method is listed in Table 3. Randomly 50 samples for each fault class from the test dataset are selected and the detection time is recorded. The detection time only for the correctly classified classes is considered and taking their average listed in Table 3 in terms of power frequency cycle. The proposed method considers half-cycle post fault signal information for data processing and about another half cycle for declaring a fault type. In Table 3, the system response time for all fault classes is within 1 cycle after fault inception point which is well accepted for the transmission line protection. Despite the fact that the proposed scheme requires a time-consuming training stage which involves 70% of the data from the original dataset and necessitates only once at the beginning of the operation. On the other hand, the detection time is the model's response time to declare a fault class for one fault case. Once the model is trained with a sufficient amount of data, the model is ready to classify the faults within minimal time.

6. Practical ability

6.1. Effect of noise

This section investigates the effect of different levels of noise on the classification performance of the proposed protection scheme. The noise is the high-frequency component which randomly mixes with the sampled signals as the sensors are exposed to harsh environmental conditions. Thus to evaluate the model performance in more realistic settings the test samples are contaminated with different levels of white Gaussian noise (WGN) while the training samples remain un-corrupted. Six different signal-to-noise ratios (SNRs) are considered (as per [29]) to measure the OAs shown in Fig. 6 for each four TL topology.

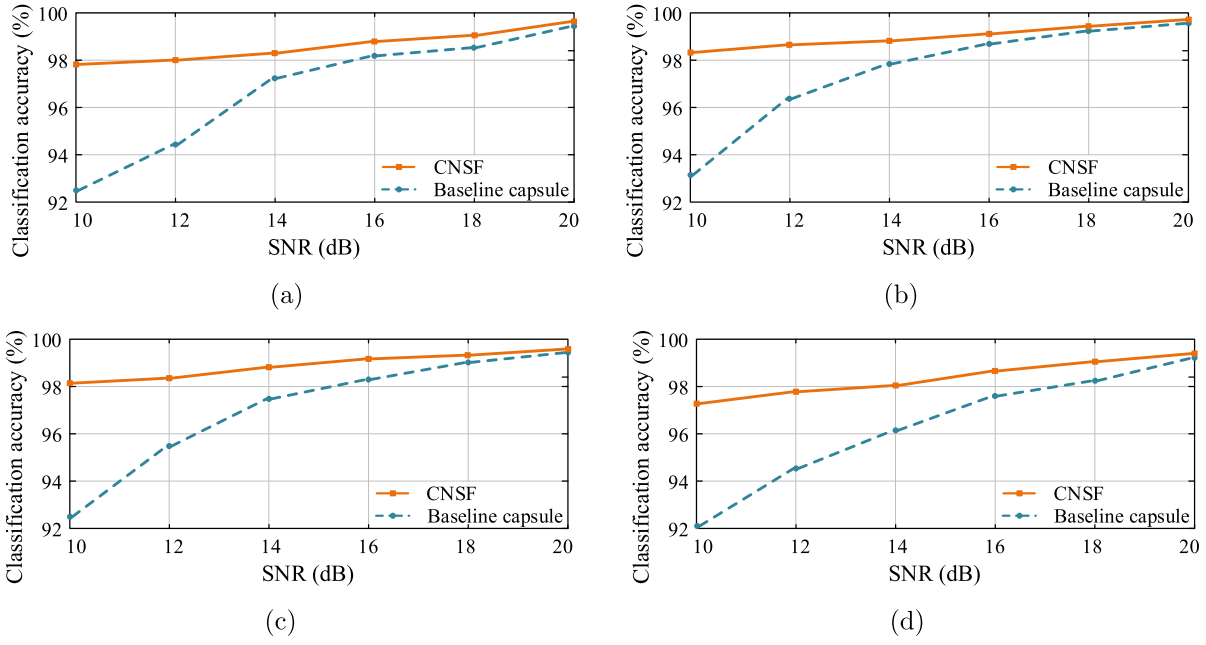

Fig. 6. Performance of the proposed FDC scheme and its comparison with baseline CN against noise for (a) TL-1, (b) TL-2, (c) TL-3, and (d) TL-4.

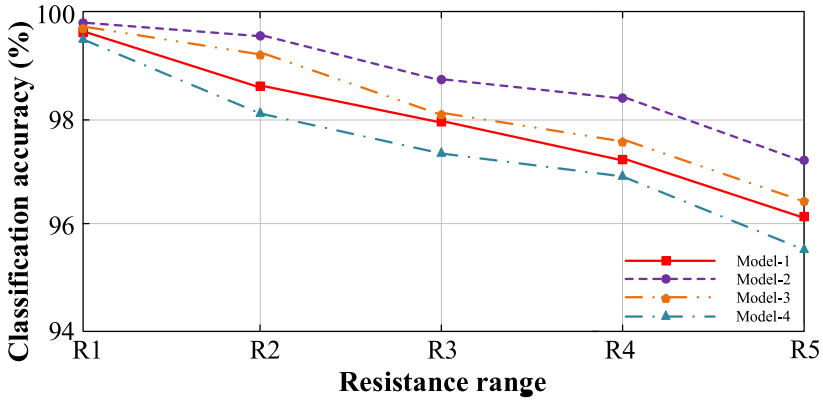

Fig. 7. Performance of the proposed model with HIF.

As can be seen from the result, the sparse-capsule based implementation yields higher than 99% accuracy within the range of various noise levels. With an extremely lower value of SNR (10 dB), the performance is observed to degrade slightly, which is also in the region of acceptance. In addition, a comparative performance with the baseline capsule network in terms of noise immunity is also shown in Fig. 6. From the figure the TL-4 is observed to be more sensitive to noise. Also, the baseline capsule-based implementation ended up with significantly lower accuracy over the considered SNR range. From the above analysis, it can be manifest that the effective immunization of the fault classification can be achieved with the CNSF based implementation.

6.2. Performance on HIF

The high impedance faults (HIF) imposes challenges in the detection task as the current magnitude during HIF is generally low. These forms of short circuit faults are not effectively detectable with the traditional over current protection. The current magnitude of HIF is close to the load current level which is due to the arc between the lines or the high ground resistance. To evaluate the model performance for HIFs, five ranges of fault resistances denoted as, R1[100 Ω to 1000Ω], R2[1001 Ω to 2000Ω], R3[2001 Ω to 3000Ω], R4[3001 Ω to 4000Ω], and R5[4000 Ω to 5000Ω] are considered as per [4]. Each type of fault is simulated for 50 random fault resistance within each range. Therefore 550 new samples within each resistance range are generated. The newly generated samples are separately injected into the previous dataset to form the new datasets. Thus a total of five datasets are formed in which the model is trained and tested separately while maintaining 70:30 training and testing ratio. The aforementioned procedure is performed for four considered power system and the accuracies are reported in Fig. 7. Referring to the result, it is observed that the TL-2 performs against HIF with higher than 97% accuracy while the other models struggle to perform the fault classification at higher fault resistance but their accuracies still remain higher than 95%. Consequently, the achieved accuracies validated the efficacy of the proposed FDC scheme for HIF.

6.3. Effect of errors in line parameters

Another uncertainty present in the TL domain and affect the fault classification accuracy is the errors in the line parameters. Such errors may arise due to the aging effect of the transmission line or any inaccurate estimation of the line parameters. At the error state, the line parameters are subjected to +- 10% deviation about the real values. Therefore, to evaluate the classification performance of the proposed FDC algorithm +-1%, +-5%, and +-10% (as per [6]) variation in impedance and admittance have been introduced in this study. Considering the error level, fault location (at 20%, 40%, 60%, and 80% line length), and fault resistance (10, 50, 100 Ω), a total of 72 fault events for each individual fault class and transmission line model are simulated. After necessary data processing, the model is trained with a newly generated dataset while keeping the training stage as before and the model performance is listed in Table 4. From the table, the maximum drop in classification accuracy is observed at −10% error for TL-1 which is 2.36% less than the accuracy without error. This slight deviation in classification performance is in line with the expectation and confirms the robust classification performance within the considerable error range.

7. Comparison with alternative approaches

This section portrays a comparative analysis of the proposed CNSF based FDC framework in terms of the baseline capsule network and other alternative approaches. Comparing the obtained results with the alternative approaches are not thoroughly consistent due to the diversity of the TL topologies, system parameters, and operating conditions adopted in different approaches. Moreover, the proposed approach is tested on four different TL topologies and the overall prediction accuracies are reported in Fig. 5. Also, the result analysis and the fault conditions (such as sensitivity analysis, HIF detection, influence of noise, impact of line parameters variation) carried out in this research are not reported concurrently in the prior literature. To put this in context, only one transmission line model (TL-2) is used and the comparison of the proposed model with the existing approach is performed

Table 4
Performance of the proposed method with errors in line parameters.

	Overall accuracy (%)					
Error (%)	**-10**	**-5**	**-1**	**1**	**5**	**10**
TL-1	97.19	98.24	98.69	98.80	98.51	97.49
TL-2	98.11	98.45	98.97	99.04	98.56	98.18
TL-3	98.00	98.30	98.70	98.89	98.41	97.98
TL-4	97.18	98.17	98.67	98.77	98.37	97.40

Table 5
Quantitative comparison with existing FDC approaches.

Method	Reference	Data type	Overall accuracy
DBN	[39]	Numerical	96.70%
WT+ChNN	[25]	Numerical	99.43%
Self-attention CNN	[42]	Image	99.58%
Baseline capsule	--	Image	99.64%
CNSF	**–**	**Image**	**99.72%**

Table 6
Qualitative comparison with existing FDC approaches.

Method	DBN	CNN	SAT-CNN	Proposed
Reference	[39]	[30]	[42]	**---**
Signal considered	Voltage	Voltage + Current	Voltage + Current	**Voltage + Current**
Data representation	Numerical	Image	Image	**Image**
Learning type	Unsupervised + Supervised	Supervised	Supervised	**Unsupervised**
No. of TL Topology considered	1	1	1	**4**
Noise resiliency	No	Yes	91.26% accuracy at 10 dB SNR	**98% accuracy at 10 dB SNR**
Characteristics	Greedy layer-by-layer training	Automatic feature extraction.	Attentive to important fault oriented features.	**1) Reduce training set % 2) Adapt to topology change 3) Immune to noise and error**

from both quantitative, as well as qualitative perspectives. The quantitative performance of the proposed model with respect to the overall accuracy is reported in Table 5. It is important to note that the accuracies in that table are presented here only to give an overall idea of the results, they are not necessarily assessed for the same transmission line topology.

Moving to the algorithm, a qualitative comparison of the proposed method and the existing method is shown in Table 6. In need of specialized feature extraction from the sequential time series fault signals, the proposed approach encodes the signal into the image rather than using the raw ones. This image representation of the data helps to capture the more detailed fault information and condense the local patterns that in other cases spread over time. The previous approaches reported in Table 6 utilizes numerical data to represent the fault features where the signatures remain unexplored and the training objective will not fulfill. This, in turn, affects their classification accuracy which is clearly inferior to the proposed one. The proposed unsupervised approach fuse the benefit of capsule network and sparse filtering. One important benefit of the capsule network over the other deep learning architecture is that the superior level features can be directed by agreement within the inferior-level features, which sets up a complex connection to characterize the faults with higher accuracy. In addition, the sparse filtering technique restrains the effect of signal distortion and reduce the size of the data sample set required for pre-training. The aforementioned analysis regarding the data representation, algorithm used and accuracies reported confirm that the proposed approach is comparable with state-of-the-art methods.

8. Conclusion

This paper proposes a novel deep learning-based intelligent model to detect and classify the power system transmission line faults. The proposed model is a capsule network combined with sparse filtering where the capsule network enables the model to work with the limited number of available system data and the sparse filtering empowers the model to become immune against the system noises. The model is designed to account for the faulted feature into a single attribute by encoding the time-series signal into a GAF image that passes through the discrete wavelet transform to generate the required faulted feature. The proposed model is independent of the manual labeling of the data required during the training and testing of the model. The efficacy of the model is tested first in detecting and classifying faults of four different transmission lines and the results indicate that the proposed model has a performance accuracy of not less than 99.47% and the highest of 99.72% within the single cycle time. The testing of the model is further carried out against the noises in raw data, high impedance faults, and line parameters variations. It is observed that the proposed CNSF model achieves the accuracy of 99% against the noises and more than 97% against the high impedance faults and line parameters variation. Additionally, a comparative performance analysis between the proposed model and some of the existing methods available in the literature shows that the proposed model has higher accuracy than that of those methods. The presence of real field data in the dataset and deployment of the model in the complex transmission line topologies will be interesting to work with in the future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Morais J, Pires Y, Cardoso C, Klautau A. A framework for evaluating automatic classification of underlying causes of disturbances and its application to short-circuit faults. IEEE Trans Power Deliv 2010;25(4):2083–94.
[2] Freire JCA, Castro ARG, Homci MS, Meiguins BS, De Morais JM. Transmission line fault classification using hidden markov models. IEEE Access 2019;7:113499–510.
[3] Aleem SA, Shahid N, Naqvi IH. Methodologies in power systems fault detection and diagnosis. Energy Syst 2015;6(1):85–108.

[4] Chen K, Hu J, Zhang Y, Yu Z, He J. Fault location in power distribution systems via deep graph convolutional networks. IEEE J Sel Areas Commun 2019;38(1):119–31.
[5] Gururajapathy S, Mokhlis H, Illias H. Fault location and detection techniques in power distribution systems with distributed generation: A review. Renew Sustain Energy Rev 2017;74:949–58.
[6] Saber A, Emam A, Elghazaly H. A backup protection technique for three-terminal multisection compound transmission lines. IEEE Trans Smart Grid 2017;9(6):5653–63.
[7] Zaki MI, El Sehiemy RA, Amer GM, El Enin FMA. Sensitive/stable complementary fault identification scheme for overhead transmission lines. IET Gener Transmiss Distrib 2019;13(15):3252–63.
[8] Ferreira V, Zanghi R, Fortes M, Sotelo G, Silva R, Souza J, et al. A survey on intelligent system application to fault diagnosis in electric power system transmission lines. Electr Power Syst Res 2016;136:135–53.
[9] Sachdev M, Baribeau M. A new algorithm for digital impedance relays. IEEE Trans Power Apparatus Syst 1979;6:2232–40.
[10] Girgis AA. A new kalman filtering based digital distance relay. IEEE Trans Power Apparatus Syst 1982;(9):3471–80.
[11] Rathore B, Mahela OP, Khan B, Alhelou HH, Siano P. Wavelet-alienation-neural based protection scheme for statcom compensated transmission line. IEEE Trans Ind Informat 2021;17(4):2557–65.
[12] Mahela OP, Sharma J, Kumar B, Khan B, Alhelou HH. An algorithm for the protection of distribution feeder using stockwell and hilbert transforms supported features. CSEE J Power Energy Syst. https://doi.org/10.17775/CSEEJPES.2020.00170.
[13] Yogee GS, Mahela OP, Kansal KD, Khan B, Mahla R, Haes Alhelou H, et al. An algorithm for recognition of fault conditions in the utility grid with renewable energy penetration. Energies 2020;13(9):2383.
[14] Kulshrestha A, Mahela OP, Gupta MK, Gupta N, Patel N, Senjyu T, et al. A hybrid fault recognition algorithm using stockwell transform and wigner distribution function for power system network with solar energy penetration. Energies 2020;13(14):3519.
[15] Ola SR, Saraswat A, Goyal SK, Jhajharia S, Rathore B, Mahela OP. Wigner distribution function and alienation coefficient-based transmission line protection scheme. IET Gener Transmiss Distrib 2020;14(10):1842–53.
[16] Ram Ola S, Saraswat A, Goyal SK, Jhajharia S, Khan B, Mahela OP, et al. A protection scheme for a power system with solar energy penetration. Appl Sci 2020;10(4):1516.
[17] Ram Ola S, Saraswat A, Goyal SK, Sharma V, Khan B, Mahela OP, et al. Alienation coefficient and wigner distribution function based protection scheme for hybrid power system network with renewable energy penetration. Energies 2020;13(5):1120.
[18] Prasad A, Edward JB, Ravi K. A review on fault classification methodologies in power transmission systems: Part-i. J Electr Syst Inform Technol 2018;5(1):48–60.
[19] Chen K, Huang C, He J. Fault detection, classification and location for transmission lines and distribution systems: a review on the methods. High voltage 2016;1(1):25–33.
[20] Bo Z, Jiang F, Chen Z, Dong X, Weller G, Redfern M. Transient based protection for power transmission systems. In: 2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 00CH37077), vol. 3. IEEE; 2000. p. 1832–7.
[21] Jamehbozorg A, Shahrtash SM. A decision-tree-based method for fault classification in single-circuit transmission lines. IEEE Trans Power Deliv 2010;25(4):2190–6.
[22] Hashemi SM, Sanaye-Pasand M, Shahidehpour M. Fault detection during power swings using the properties of fundamental frequency phasors. IEEE Trans Smart Grid 2017;10(2):1385–94.
[23] Malathi V, Marimuthu N, Baskar S. Intelligent approaches using support vector machine and extreme learning machine for transmission line protection. Neurocomputing 2010;73(10–12):2160–7.
[24] Reddy MJ, Mohanta DK. Adaptive-neuro-fuzzy inference system approach for transmission line fault classification and location incorporating effects of power swings. IET Gener Transmiss Distrib 2008;2(2):235–44.
[25] Vyas B, Das B, Maheshwari RP. An improved scheme for identifying fault zone in a series compensated transmission line using undecimated wavelet transform and chebyshev neural network. Int J Electr Power Energy Syst 2014;63:760–8.
[26] Gao W, Ning J. Wavelet-based disturbance analysis for power system wide-area monitoring. IEEE Trans Smart Grid 2011;2(1):121–30.
[27] Liu Z, Han Z, Zhang Y, Zhang Q. Multiwavelet packet entropy and its application in transmission line fault recognition and classification. IEEE Trans Neural Netw Learn Syst 2014;25(11):2043–52.
[28] Mishra DP, Samantaray SR, Joos G. A combined wavelet and data-mining based intelligent protection scheme for microgrid. IEEE Trans Smart Grid 2015;7(5):2295–304.
[29] Chen K, Hu J, He J. Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. IEEE Trans Smart Grid 2016;9(3):1748–58.
[30] Guo M-F, Yang N-C, Chen W-F. Deep-learning-based fault classification using hilbert–huang transform and convolutional neural network in power distribution systems. IEEE Sens J 2019;19(16):6905–13.
[31] Chen YQ, Fink O, Sansavini G. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction. IEEE Trans Industr Electron 2017;65(1):561–9.
[32] Zin AAM, Saini M, Mustafa MW, Sultan AR, et al. New algorithm for detection and fault classification on parallel transmission line using dwt and bpnn based on clarke's transformation. Neurocomputing 2015;168:983–93.
[33] Reyes-Archundia E, Guardado JL, Gutiérrez-Gnecchi J, Moreno-Goytia E, Guerrero-Rodriguez N. Fault analysis in tcsc-compensated lines using wavelets and a pnn. Neural Comput Appl 2018;30(3):891–904.
[34] Taheri MM, Seyedi H, Mohammadi-ivatloo B. Dt-based relaying scheme for fault classification in transmission lines using modp. IET Gener Transmiss Distrib 2017;11(11):2796–804.
[35] Abdullah A. Ultrafast transmission line fault detection using a dwt-based ann. IEEE Trans Ind Appl 2017;54(2):1182–93.
[36] Malik H, Sharma R. Transmission line fault classification using modified fuzzy q learning. IET Gener Transmiss Distrib 2017;11(16):4041–50.
[37] Mahdi M, Genc VI. Post-fault prediction of transient instabilities using stacked sparse autoencoder. Electr Power Syst Res 2018;164:243–52.
[38] Chen K, Hu J, He J. A framework for automatically extracting overvoltage features based on sparse autoencoder. IEEE Trans Smart Grid 2016;9(2):594–604.
[39] Zhang Y, Mei W, Dong G, Gao J, Wang P, Deng J, et al. A cable fault recognition method based on a deep belief network. Comput Electr Eng 2018;71:452–64.
[40] Gashteroodkhani O, Majidi M, Etezadi-Amoli M. A combined deep belief network and time-time transform based intelligent protection scheme for microgrids. Electr Power Syst Res 2020;182:106239.
[41] Zhang S, Wang Y, Liu M, Bao Z. Data-based line trip fault prediction in power systems using lstm networks and svm. IEEE Access 2017;6:7675–86.
[42] Fahim SR, Sarker Y, Sarker SK, Sheikh MRI, Das SK. Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification. Electr Power Syst Res 2020;187:106437.
[43] Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. In: Advances in neural information processing systems; 2017. p. 3856–66.
[44] Paoletti ME, Haut JM, Fernandez-Beltran R, Plaza J, Plaza A, Li J, et al. Capsule networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 2018;57(4):2145–60.
[45] Liu M, Liao G, Yang Z, Song H, Gong F. Electromagnetic signal classification based on deep sparse capsule networks. IEEE Access 2019;7:83974–83.
[46] Chen T, Wang Z, Yang X, Jiang K. A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals. Measurement 2019;148:106857.
[47] Tasdighi M, Kezunovic M. Impact analysis of network topology change on transmission distance relay settings. In: 2015 IEEE Power & Energy Society General Meeting. IEEE; 2015. p. 1–5.
[48] Muyeen S, Hasanien HM. Operation and control of hvdc stations using continuous mixed p-norm-based adaptive fuzzy technique. IET Gener Transmiss Distrib 2017;11(9):2275–82.