[1]Wael Abdulhasan Atiyah

[2*]Shahram Karimi

[3]Mohamad Moradi

A Novel Approach for Diagnosing Transformer Internal Defects and Inrush Current Based on 1DCNN and LSTM Deep Learning

Journal of Electrical Systems

Abstract: **-** In power systems, power transformer (Pt) protection plays a vital role in ensuring that customers have a reliable power supply. Correctly recognizing inrush currents from internal defects and preventing differential relay malfunctions are two of biggest challenges facing the differential protection of power transformers. Although previous approaches suggested to overcome these issues have promising outcomes, increasing the accuracy and reducing the execution time and complexity of transformer differential relays are still interesting topics for researchers. Accordingly, a new fault diagnostic method based on wavelet transform (WT) and deep learning is introduced in paper. In the proposed approach, Discrete WT is used to extract the features of differential currents, and combined one-dimensional convolutional neural networks and long short-term memory (1DCNN-LSTM)) is applied for classify internal faults from other abnormal events. High accuracy, no need for any thresholds or transformer parameters and fast fault detection are the main advantages of the proposed approach. The simulation results for a 132/11 kV, 63 MVA power transformer approved the proposed method for its ability to accurately differentiate between inrush currents and internal defects after 5 ms, as well as its accuracy for abnormal event classification of about 99.4%.

Keywords: differential relay, inrush current, wavelet, transform, convolutional neural network.

1. INTRODUCTION

Power transformers (Pts) are among the most costly and essential components of high-voltage substations. Therefore, in case of internal faults that may cause serious damage to the transformer, the transformer must be disconnected quickly in order to check and fix the failure. However, unintended disconnection of power transformers can compromise the dependability of the power network. As a result, transformer protection has a significant impact on network reliability as well as preventing damage to the transformer. Differential relay is the first protection strategy that uses the results of subtraction of currents from both sides of the transformer to ensure the correct operation of the power transformer. Any faults inside the transformer will affect the amounts of this subtraction, which refer to faults.

Differential relay is used for protect power transformers with capacities over 10MVA [1]. Failure to recognition between inrush current and internal defects eventually damages the transformer, which then impairs system performance and threatens worker safety in addition to significant financial losses. The main issue of differential protection is improper operation due to inrush currents. The amount of inrush current is dependent on the power transformer's rating capacity, core specification, residual fluxes in the transformer core, and voltage level during energization. However, the inrush current level is typically very large to cause the differential relay to operate incorrectly.

Different differential protection methods have been suggested to recognition between inrush current and internal defects. The total harmonic distortion (THD) is used in[2] due to the larger THD value in inrush currents. However, applying harmonic criteria to differential protection could result in false blocking or performance slowdown. Harmonic components are determined by various methods such as the passive filter, Fourier transform, sine-rectangular transform, Haar function, Walsh function, extended Kalman filter, least squares algorithm and Debauches function [3]. Gap detection refers to the technique of identifying the duration of time intervals when the differential current is close to zero. Typically, this time interval or gap is greater than 1/4 cycle for inrush currents and less than 1/4 cycle for internal faults. Therefore, by measuring this gap, inrush currents can be distinguished from internal defects[4][5]. The inrush current may be identified because the leakage flux measured when the no-load transformer is energized differs significantly from the leakage flux observed during normal operation of the transformer[6]. Multiple types of criteria for differential protection approaches based on FL is proposed in[7]. Authors in[8] proposed the boundary stationary wavelet transform (WT). The high computational burden and reliance on the threshold value are the shortcomings of this method. In[9], a method based on the WT

[1]*Department of Electrical Engineering, Razi University, Kermanshah, Iran, waaeleng79@yahoo.com*

[2]*. Department of Electrical Engineering, Razi University, Kermanshah, Iran, shahramkarimi@razi.ac.ir*

[3]*. Department of Electrical Engineering, University of Kurdistan, Sanandaj, Iran., mmoradi@uok.ac.ir*

**(corresponding author)*

and support vector machine (SVM) is presented to identify the type of defects in power transformers. In[10], the primary currents are estimated using extended Kalman filter (EKF) and SVM to categorize the defects. However, because of the lengthy data training process, large datasets will result in inadequate SVM performance for data classification. Additionally, this method has high computational complexity and is susceptible to noisy data. This paper proposes an adaptive neuro-fuzzy inference system-based wavelet-based classification technique[11]. In[12], differential protection has been improved using the convolutional neural network (CNN). In [13], a novel approach to differential protection is suggested that uses multi-resolution analysis to decompose the differential current signal into various energy levels before determining which feature is best to feed the bi recurrent unit with gates for event classification. Fuzzy membership function is used in multi-scale multivariate fuzzy entropy, which separates fault currents from normal cases[14]. In[15], a method based on back-propagation neural network (BPNN) and DWT is presented for detecting the internal defects in Pts. In [16], a new transformer fault diagnostic approach based on an ensemble machine learning and Internet of Things (IoT) monitoring system is given. A data-measuring subsystem and a data-receiving subsystem constitute the monitoring system based on IoT technology. An approach based on WT and probabilistic neural network is also proposed for distinguishing external and internal defects in power transformers [17]. An approach built on a Clark-based transform to extract signal characteristics and a modified hyperbolic S-transform for fault classification are proposed in[18].This study, does not take into account the impacts of inrush current or saturation of current transformer (CT). An approach based on the combination of WT and SVM is presented in [19], where PSO has been utilized to adjust the SVM parameters. Hyperbolic S-transform [20] and Hilbert transform [21] are also used for diagnosing inrush currents from internal defects. In[22],differential protection is based on Clark and WT in the time domain, with only one differential unit per power transformer and automatic settings available.

Although reviewed approaches have promising outcomes, increasing the accuracy and reducing the execution time and complexity of transformer differential relays are still challenging topics for researchers. According to the mentioned issues, this paper present, a novel fault diagnostic technique based on DWT and deep learning is developed. In the proposed method, DWT is used to extract the features of differential currents, and (1DCNN-LSTM) is applied to classify internal faults from other abnormal events. The combination of CNN and LSTM overcomes the drawbacks of using CNN to process series data and increases the robustness and accuracy of the model. Since the proposed method does not require any threshold, it is not sensitive to noisy data. Results of simulation approved that the proposed method for its ability to accurately recognition internal defects, inrush current and external faults under saturation of CT after 5 ms (1/4 cycle).

The novelty and contribution of this study can be described as follows:

• Introducing an innovative differential relay for power transformers: this paper introduces a novel differential protection method based on DWT and 1DCNN-LSTM to recognizing inrush currents from internal defects of 3-ph power transformers.

• Employing spectral energy of detail coefficients: to minimize the computational burden of the proposed model, by using spectral energy of detail coefficients instead of detail coefficients is as the input data of the proposed deep learning model during each moving window.

• Not needing any thresholds or transformer parameters: the proposed differential protection technique does not require any threshold or transformer parameters. Therefore, it is not sensitive to noisy data and can be used to protect power transformers with any capacity and voltage level.

• Robust fault detection with high accuracy and short detection time: by training the developed deep learning model using data collected from the simulation of various normal operations and abnormal events, after 5 milliseconds, the proposed approach can reliably identify internal defects and restrain for cases of inrush currents and external faults with saturation of CT.

The other sections in this article are arranged: Structure of DWT and 1DCNN-LSTM are explained 2nd section. Deep learning-based proposed approach is presented in 3rd section. The specification of the simulated power system and results of simulation are elucidated in 4th section and 5th section, respectively. Finally, in 6th Section, conclusions are presented.

2. DWT and 1DCNN-LSTM Structure

2.1. DWT Structure

A frequency analysis technique with complex properties that outperforms more conventional methods is wavelet signal analysis. There are three common types of WT: continuous WT (CWT), DWT, and packet WT (PWT). WT is a powerful mathematical tool for frequency analysis. Short-term transients include high-frequency components. The intrinsic benefit of WT is that it requires a short time interval to extract the features of short-term transient signals. Therefore, WT is a suitable tool for fast detection of transient events. The accuracy of extracting signal features depends on the type of mother wavelet. Due to the accuracy of Daubechies with degree four (db4) for choosing mother wavelet to extract the differential current characteristics of power transformers during abnormal events, as reported in[8], this mother wavelet is selected in this paper.

Figure 1, shows the schematic block diagram of a DWT with three-level decomposition. In this figure, blocks A1 to A3 represented low-pass filter (LPF) and D1 to D3 represented high-pass filter (HPF), that output sampling frequency is half of their input sampling frequency. The output of A and D blocks are approximation and detail coefficients, respectively. Since each filter's cut-off frequency is 1⁄2 that of its predecessor, each filter's output has 1⁄2 the input frequency band, which doubles the frequency resolution.

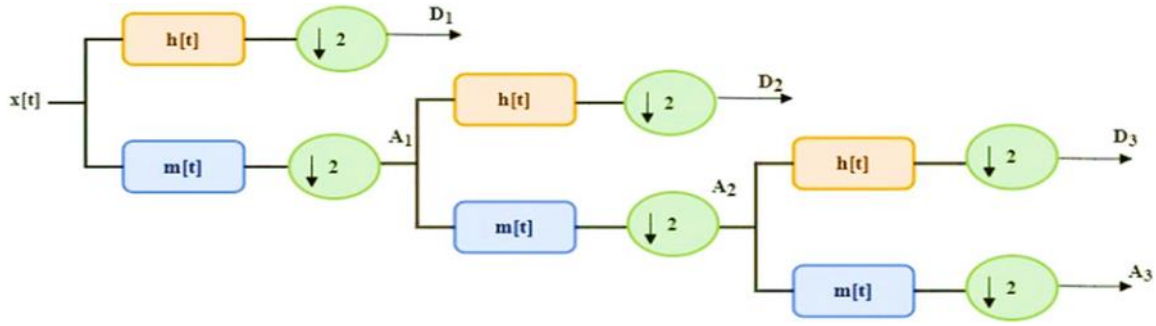


Figure1. Multi-Resolution for extraction of approximation and detail coefficients

2.2. CNN Structure

CNN is a general feedforward deep neural network with powerful classification capabilities inspired by primate perception techniques [23]. It is primarily intended for processing data with a known network layout. CNN uses multiple filters to extract representative features from the input data while integrating sparse connectivity and parameter weight-sharing techniques to downsample and refine information dimensions in time and space. By doing so, the number of training parameters is decreased and algorithm overfitting is effectively avoided. Figure 2, displays the general architecture diagram of CNN [23]. Convolution, pooling, activation and fully connected layers are typically present in the CNN structure. The foundation of CNN is made up of convolutional layers. Data in one dimension is the only input for one-dimensional convolution. As a result, both the dimensionality of its kernel and its output are one-dimensional. Figure 3 illustrates how the one-dimensional convolution works [24]. In each convolutional steps, a pooling operation follows Convolutional filters that can be learned. These convolutional filters extract the high-level characteristics from the input signal through the combination of an input signal with a set of weights and applying a non-linear activation function. After that, the outputs have been sent to a pooling operation, which makes the features found by the convolutional filters less spatially detailed while highlighting the most important ones. As the input moves through the convolutional steps (from left to right in figure 3), the network picks up more signal features.

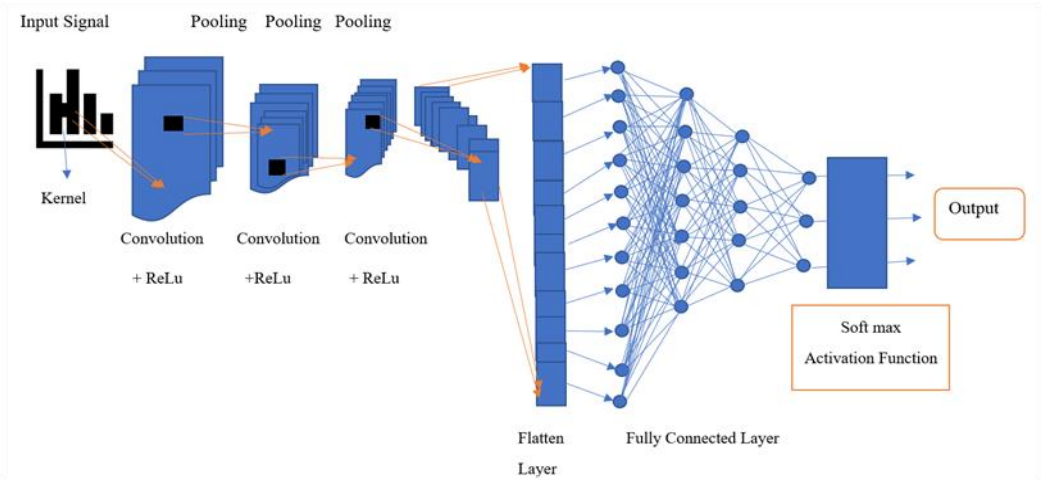


Figure 2. General structure of CNN

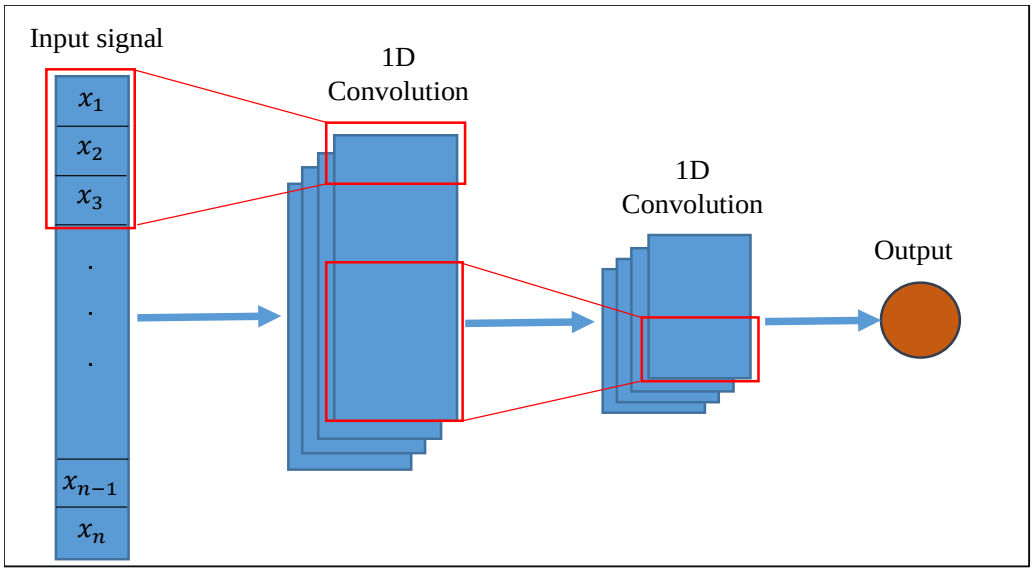


Figure 3. One-dimensional convolution

CNN has a typical architecture for image classification (as presented in figure 4), where the image is processed through a series of convolutional, activation, and pooling layers, and then finally classified using a series of dense layers. It appears to be suitable for relatively small images and a dataset with 4 distinct classes. The trainable layers of a deep CNN for image processing are organized hierarchically as arranged below:

• Layer of Image Input: The location of the picture size is on an image input layer.

• Layer of Convolution: Filter size, the first argument in the convolutional layer, determines the width and height of the filters which the training function applies to images while it scans them. Common filter sizes are 3x3 and 5x5. Convolutional filters combine input signals with a set of weights.

• Layer of Batch Normalization: This layer uses a transformation which keeps activation average deviation near to 1 and the mean of activation near to 0, normalization of activations of the last layer at each batch.

• ReLu Layer: After completing the batch normalizing layer, activation is carried out nonlinearly. When creating a Relu layer, rectified linear units (RLUs) are the most commonly used activation function.

• Max Pooling Layer: The spatial dimension of the input map of features are reduced by a factor of two using a 2D maximum pooling layers using a pool dimension of two-by-two pixels and a step of 2.

• Layer of Fully Connected: a dense layer that is entirely connected, connecting all of the nodes of the previous one to its own nodes. It probably represents the classes or categories the CNN is trying to classify the inputs.

• SoftMax Layer: the output of the preceding layer is subjected to the SoftMax function in this manner, producing a probability distribution across the classes.

• Layer of Classification: the outputs category label layer with the highest probability.

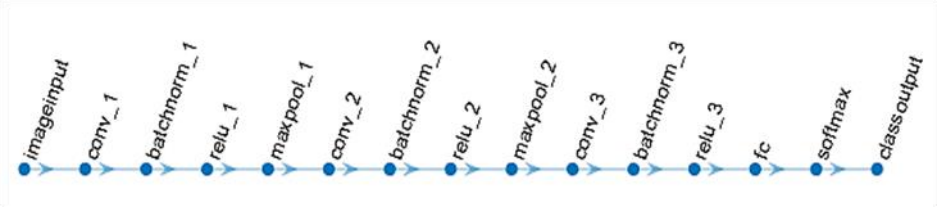


Figure 4. CNN layers

2.3. LSTM Network

A class of deep neural network is a recurrent neural network (RNN), which enables information sharing in the time dimension by adding connections (i.e., weights) between neurons in the same layer. Therefore, RNN is appropriate for dealing with time series challenges. LSTM is a unique variation of RNNs. By incorporating a "gate" structure in the right position, it overcomes the issue of information redundancy by enabling information to be selectively remembered or lost as it passes through each neuron. LSTM has the ability to increase the weight of the original data while decreasing the weight of unrelated data, which submits solutions to the problems of conventional RNNs such as gradient vanishing, gradient expansion, and the inability to manage long-range dependencies. The extraction of characteristics from multidimensional temporal correlation data using LSTM is also very well done. A LSTM includes a gate for input, a forget gate, and a gate for output to address an RNN problem. Figure 5 depicts the internal architecture of an LSTM [25].

The equations for calculating the vectors of activation of a gate for input, a forget gate, and a gate for output are:

$$I_t = \sigma(w_i * X_t + z_i * Y_{t-1} + b_i) \quad (1)$$

$$F_t = \sigma(w_f * X_t + z_f * Y_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(w_o * X_t + z_o * Y_{t-1} + b_o) \quad (3)$$

A Sigmoid activation function forms the activation function. Therefore, the range of I_t, F_t and o_t is between 0 to 1. w_i, w_f, w_o, z_i, z_f, and z_o are the weight matrices of the gates. The bias vectors are b_i, b_f, and b_o. The subscripts *i*, *f*, *o*, and *t* denote a gate for input, a forget gate, and a gate for output, and time step, respectively.

A hyperbolic tangent activation function (*th*) is also applied to the input data as follows:

$$N_t = th(w_n * X_t + z_n * Y_{t-1} + b_n) \quad (4)$$

whereas w_n and z_n are the matrices of weights, and b_n is the vector of bias. The intermediate variable M_t can be calculated as (5).

$$M_t = I_t * N_t \quad (5)$$

The unit information, L_t, and the output state, Y_t, are calculated as (6) and (7), respectively.

$$L_t = M_t + F_t * L_{(t-1)} \quad (6)$$

$$Y_t = th(L_t) * O_t \quad (7)$$

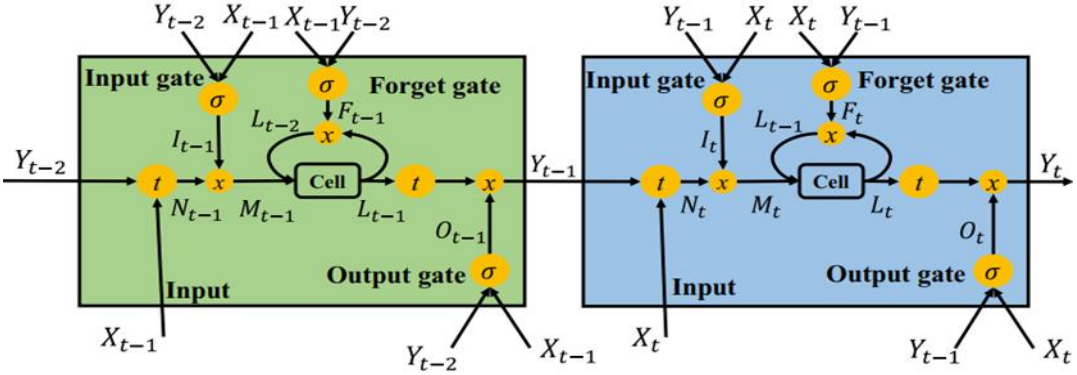


Figure 5. LSTM internal architecture

As shown in figure 6, the layers of an LSTM network are:

• Sequence Input: this layer is designed to handle sequence data. This means that the data fed to the neural network has an inherent order or temporal aspect, which is important for tasks like time series forecasting.

• Lstm Layer: the LSTM layer is a crucial component that enables the network to process and learn from time series data or any sequential data. LSTMs are specifically designed to remember and recognize patterns over long sequences and are less susceptible to the vanishing gradient problem that affects RNNs.

• Fully Connected (fc) Layer: All of the neurons (or units) from the preceding layer are connected to all of the neurons in its layer by the fully connected layer. This layer acts as a bridge between the LSTM layer and the regression output, ensuring that the high-dimensional output from the LSTM is transformed into a regression shape suitable.

• layer of Regression output: layer of the regression is used as the final layer in a neural network intended for regression tasks. In contrast, the objective of task classification is to allocate the input data to one of multiple predetermined categories, regression tasks involve predicting a continuous value.

Figure 6. LSTM layers

3. Proposed Algorithm

The proposed fault discrimination algorithm in this study is based on 1DCNN-LSTM. quickly and correctly distinguishes between internal defects and inrush currents of Pts under different conditions. In the proposed algorithm, distinctive features are extracted from one-dimensional data using a 1DCNN that can learn fault features on its own. An LSTM is then used to learn the correlation between distinctive features (as shown in figure 7). The advantage of combining 1DCNN and LSTM to create the 1DCNN-LSTM algorithm is that it can achieve greater discrimination and overcome the drawbacks of conventional methods. Table 1 presents the functions and specifications of 1DCNN and LSTM.The proposed algorithm flowchart is depicted in figure 8. Processing of creating differential protection algorithm includes the following steps:

Step 1: measuring the differential current for three phases and applying WT for feature extraction.

Step 2: calculating the spectral energy for three details during each moving window.

Step 3: converting the spectral energies to image and collecting all data in dataset groups for different conditions and events.

Step 4: training the proposed 1DCNN-LSTM using the saved dataset groups.

Step 5: evaluating the proposed algorithm to recognition between inrush currents and internal defects using the test datasets.

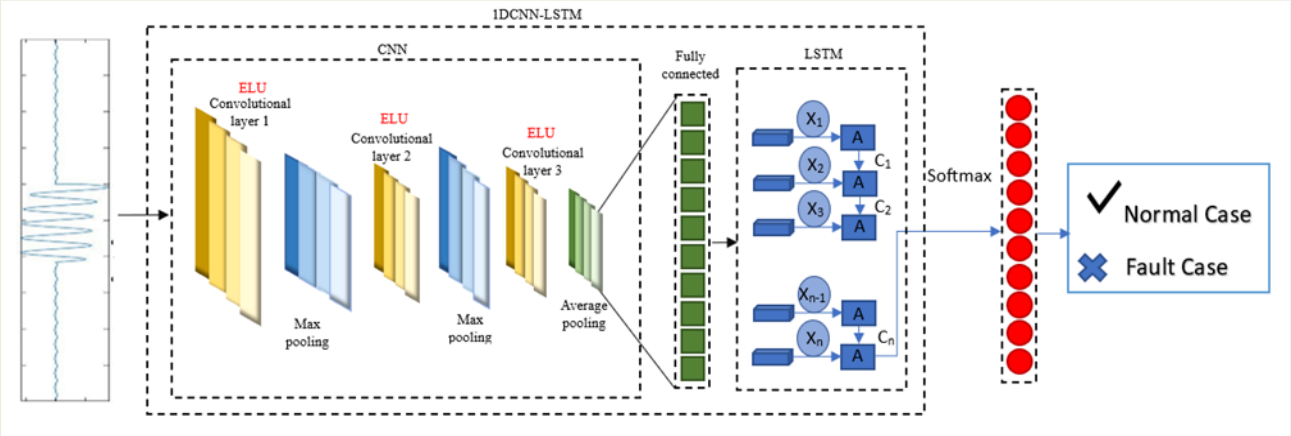


Figure 7. Proposed 1DCNN-LSTM construction

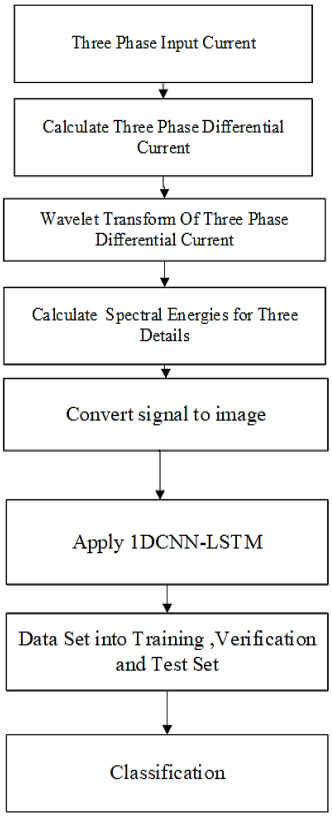


Figure 8. Flowchart of the proposed algorithm

Table1.Specifications of 1DCNN and LSTM algorithm

Specification	1DCNN	LSTM
Function of training	sgdm	adam
Total number of layers	fifteen	three
Number of convolution layers	three	---
Number of LSTM layers	---	one
Initial learn rate	0.005	0.005
Maximum Epochs	20	200
Learn rate drop factor	0.1	0.1

3.1. Spectral Energy Calculation of Wavelet Details

By using wavelet analysis to examine the frequency content of the differential currents, it is possible to determine the detail coefficients spectral energy. We use the detail coefficients spectral energy because it preserves the wavelet properties of differential currents. The spectral energy for three details of three wavelet signals are calculated as follows:

$$P_{a,b,c-det\,i} = \sum_{k=1}^{n} I^2_{a,b,c-det\,i}(k)\Delta t \quad i=1, 2, 3 \qquad (8)$$

Where $P_{a,b,c-det\,i}$ is the spectral energy of each wavelet details (D1-D3) for three phases a, b, and c. The parameter *n* is samples numbering in the moving window, and Δt is time step length.

In the proposed algorithm, nine spectral energies are calculated during the moving window moving on the wavelet signals. These nine spectral energies will be represented in a figure, which will be fed to the 1DCNN after converting it into an image.

4. Simulated Power System

In order to assess the proposed differential relay's effectiveness, complete power system depicted in figure 9 is simulated using the MATLAB/Simscape Electrical toolbox. The power system includes a three-phase power supply, a three-phase Ynd1 power transformer, a three-phase load, three CTs on transformer both sides, and two three-phase circuit breakers. Table 2 illustrates specifications of current transformers and Pts.

Table 2. Power transformer and CT specifications

Parameter	Value
Rated power	63 MVA
Voltage ratio	132/11 KV
Rated frequency	50 Hz
Percentage impedance	18.7%
CT ratio for primary side	300/5
CT ratio for secondary side	2000/5

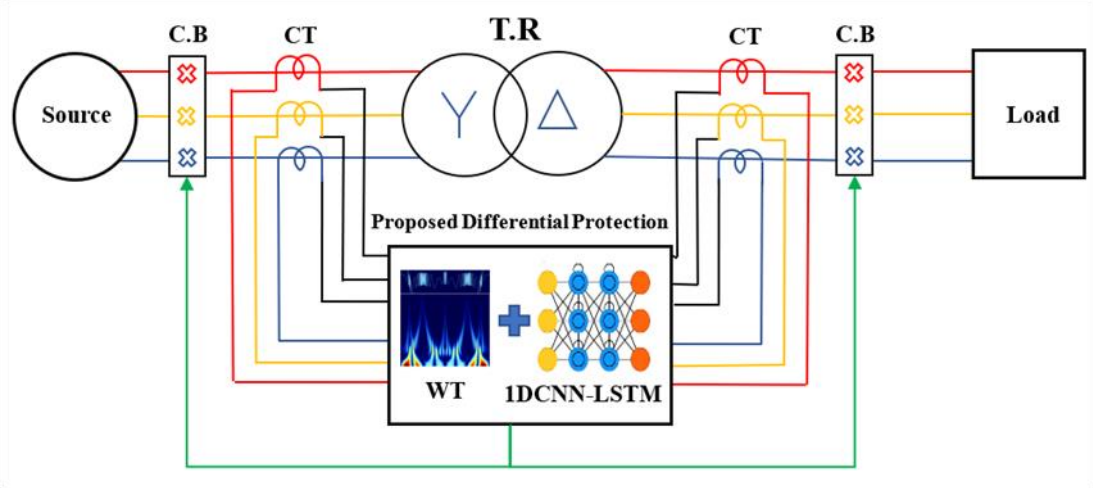


Figure 9. Power system modelling

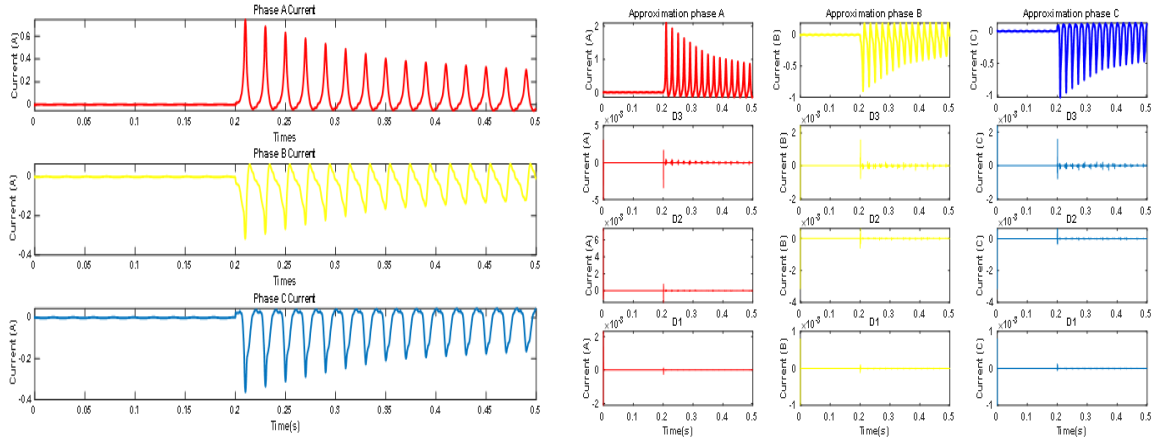
5. Simulation Results

To train and evaluate the proposed deep learning-based algorithm, 533 cases divided into four groups (magnetizing inrush currents, transformer internal defects, transformer external faults under saturation of CT, and transformer normal operation) are simulated. For each case, the detail coefficients are extracted using the WT of differential currents and then related spectral energies are calculated. After converting detail coefficients spectral energy to the corresponding image, the stored images are used to train and evaluate the proposed 1DCNN-LSTM.

5.1 Feature Extraction Using WT

5.1.1. Inrush Current Case

The magnetizing inrush current flows on the high voltage side of Pts and can even reach 7–8 times the rated current. In figure 10a, the power transformer under study exhibits differential currents associated with magnetizing inrush currents. The wavelet approximation and detail coefficients of inrush currents are presented in figure 10b. As seen in this figure, the detail signals (D1-D3) experience several narrow pulses with large amplitude when inrush currents pass. This feature can be utilized for recognize inrush currents from internal defects.

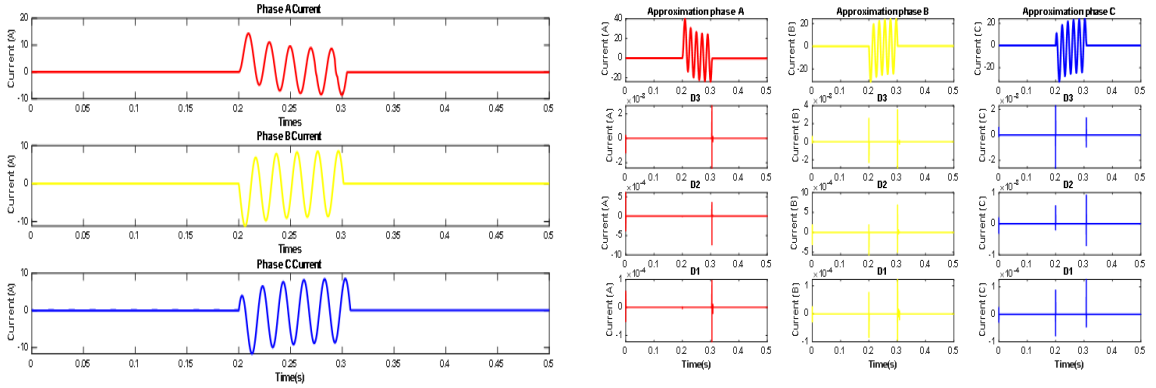


(a) (b)

Figure10. Inrush current case: (a) differential currents (b) approximation and details coefficients

5.1.2. Internal Fault Case

Different internal faults are simulated in this research. The differential currents associated with a symmetrical three-phase-to-ground fault (ABC-G) are illustrated in figure 11a. The wavelet approximation and detail coefficient of the differential currents are shown in figure 11b. In contrast to inrush currents, where many narrow pulses linger at little attenuation and don't decay to zero, in this case the narrow pulses instantly decay to zero and the detail remains zero during the internal fault.



(a) (b)

Figure 11. Internal fault (ABC-G) case: (a) differential currents (b) approximation and details coefficients

5.1.3. External Fault Under Saturation of CT

During the normal operation of Pts and in the event of an external fault, differential protection does not operate. However, if CTs saturate during external fault, conventional differential protection can cause a false trip. Therefore, it's important to recognize internal defects from all types of external faults under CT saturation. Figure 12a shows the differential currents with 3ph-ground fault (ABC-G) with a ground fault resistance of 95 Ω, outside the protection zone of the power transformer. As obvious in the figure 12b, the narrow pulses with the similar shape and magnitude are seen in detail signals during CT saturation. These features can distinguish external faults from other events.

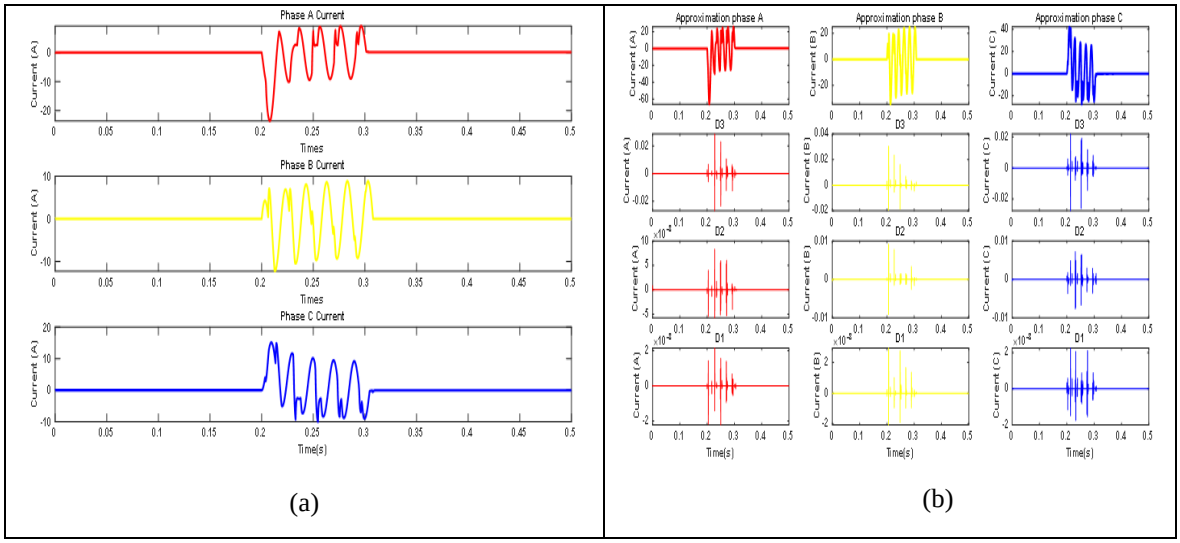


Figure 12. External fault under CT saturation (ABC-G): (a) differential currents (b) approximation and details coefficients

5.2. Spectral Energies Calculation

In the proposed model, the window of spectral energies is divided into three segments. Each segment corresponds to a phase and contains three spectral energies (related to D1-D3). Figure 13 illustrates the window of spectral energies for different events. Figure 13 (a, b, c) show the spectral energies for three different inrush currents. Figure 13 (d, e, f) depicts the window of spectral energies for BC, AB, and B-G internal faults, respectively. As can be seen in these figures, the spectral energies have a small range for the healthy phases and increase for the faulty phases. the window of spectral energies for three different external faults with CT saturation are shown in Figure 13 (g, h, i).

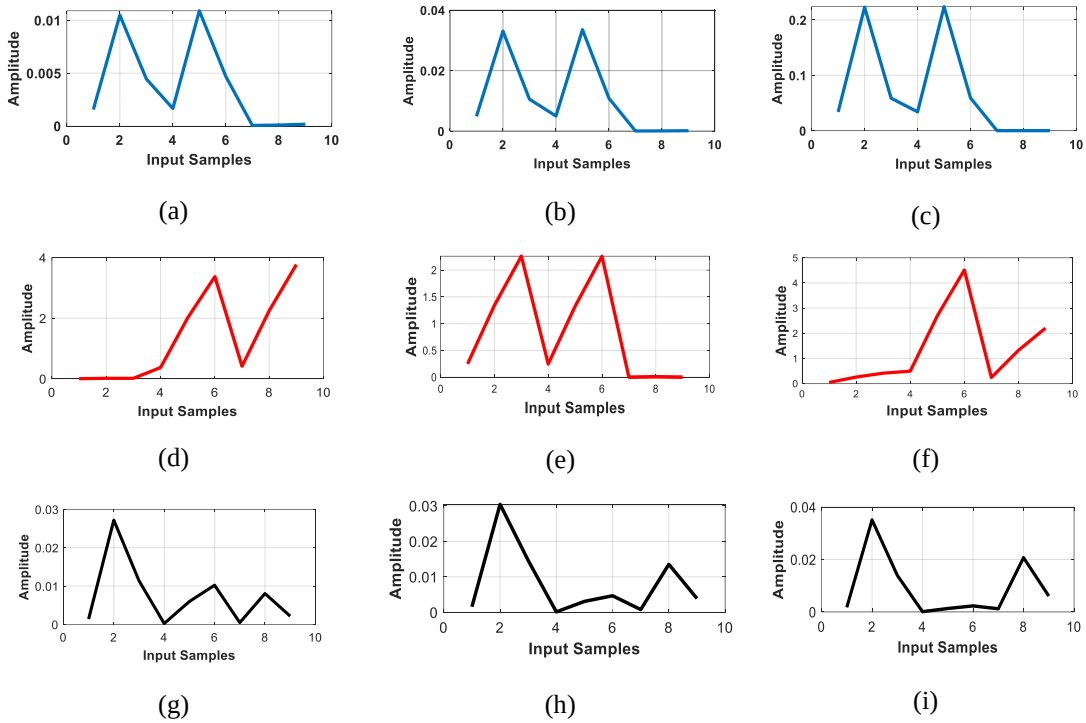


Figure 13. (a, b, c) spectral energies for different inrush currents, (d, e, f) spectral energies for different internal faults, (g, h, i) spectral energies for different external faults under CT saturation

5.3. 1DCNN and LSTM Training

5.3.1. 1DCNN Training

Out of 533 simulated cases, about 70% (373 cases) are used as training datasets. As seen in figure 14 on the right side of the plot, which depicts the training curve, the network converges with accuracy 93.51% after 110 epochs. The training process shows the amount of time that has passed, the total number of iterations and the total number of iterations for each epoch, maximum number of iterations, the learning rate, and the learning method. A crucial step in neural networks training is iterative minimization of a loss function. To reduce the loss function, a gradient descent strategy should be used. In each iteration of the descent process, the weights are updated, and the gradient of the loss function is evaluated. The training process for the CNN model takes only 59 seconds to complete because it only has 15 layers. The training results of the proposed CNN are listed in figure 15.

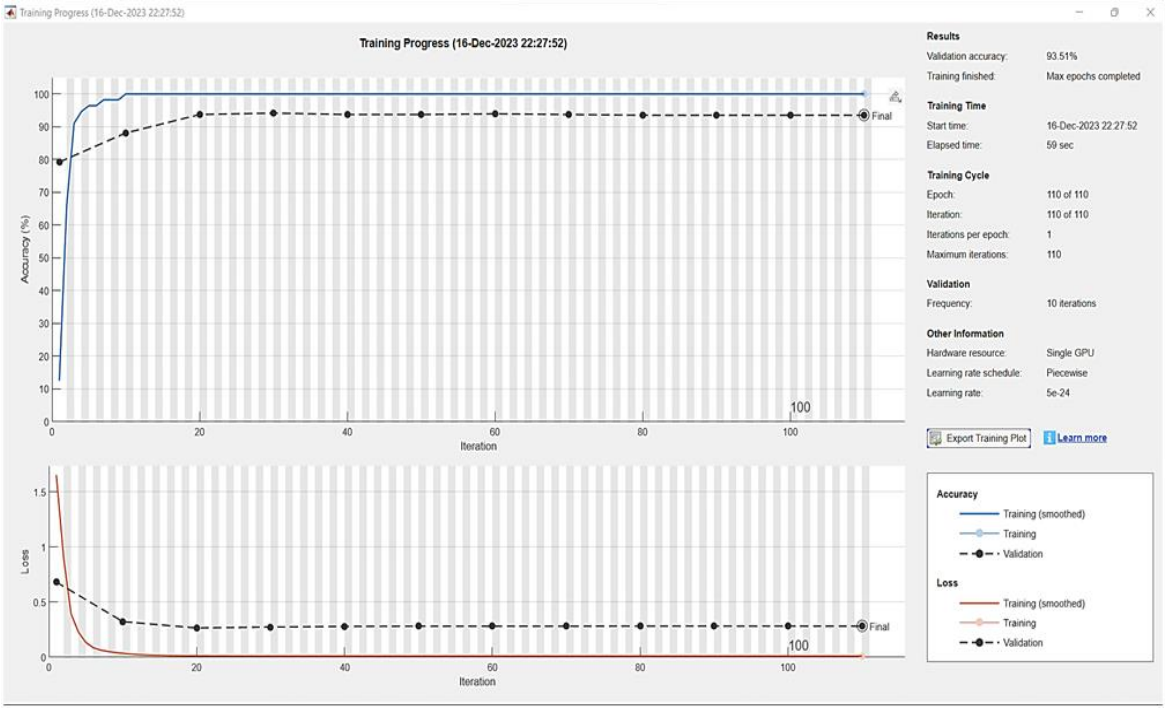


Figure 14. CNN network training and validation analysis

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation Loss	Base Learning Rate
1	1	00:00:46	12.50%	79.22%	1.6524	0.6817	0.0050
10	10	00:00:48	100.00%	88.10%	0.0354	0.3228	0.0005
20	20	00:00:49	100.00%	93.72%	0.0121	0.2655	5.0000e-06
30	30	00:00:51	100.00%	94.16%	0.0103	0.2742	5.0000e-08
40	40	00:00:52	100.00%	93.72%	0.0101	0.2792	5.0000e-10
50	50	00:00:53	100.00%	93.72%	0.0100	0.2816	5.0000e-12
60	60	00:00:54	100.00%	93.94%	0.0100	0.2816	5.0000e-14
70	70	00:00:55	100.00%	93.72%	0.0100	0.2820	5.0000e-16
80	80	00:00:56	100.00%	93.51%	0.0100	0.2827	5.0000e-18
90	90	00:00:57	100.00%	93.51%	0.0100	0.2828	5.0000e-20
100	100	00:00:58	100.00%	93.51%	0.0100	0.2825	5.0000e-22
110	110	00:00:59	100.00%	93.51%	0.0100	0.2822	5.0000e-24

Figure 15. Evaluation of the CNN network training and validation accuracy at each epoch and iteration level

5.3.2 LSTM Training

By adjusting the weight matrices and the bias vectors during training, in order to minimize the error, the proposed LSTM gains the ability to identify patterns and temporal dependencies in the sequences of inputs. During the training process, the correctness of the proposed LSTM is verified by defining validation data. Activity and performance of model is evaluated using the root-mean-square error (RMSE). To ensure the robustness of the model, three types of optimizers are tested to select the appropriate optimizer.

The training analysis of the proposed LSTM network using the Sgdm optimizer is displayed in figure 16. As can be seen, using this optimizer, the model achieves a loss of 0.5 and an RMSE of 0.98. Figure 17 shows the LSTM training analysis when the RMSprop optimizer is applied. In this case, the loss and RMSE value are 2.5e-02 and 0.22, respectively. The training progress using the Adam optimizer is illustrated in figure 18. Using this optimizer, the loss and RMSE value reaches 0.012 and 0.15, respectively. It can be observed from figures 16 to 18 that compared to the Sgdm and RMSprop optimizers, the Adam optimizer provides higher accuracy for the proposed LSTM.

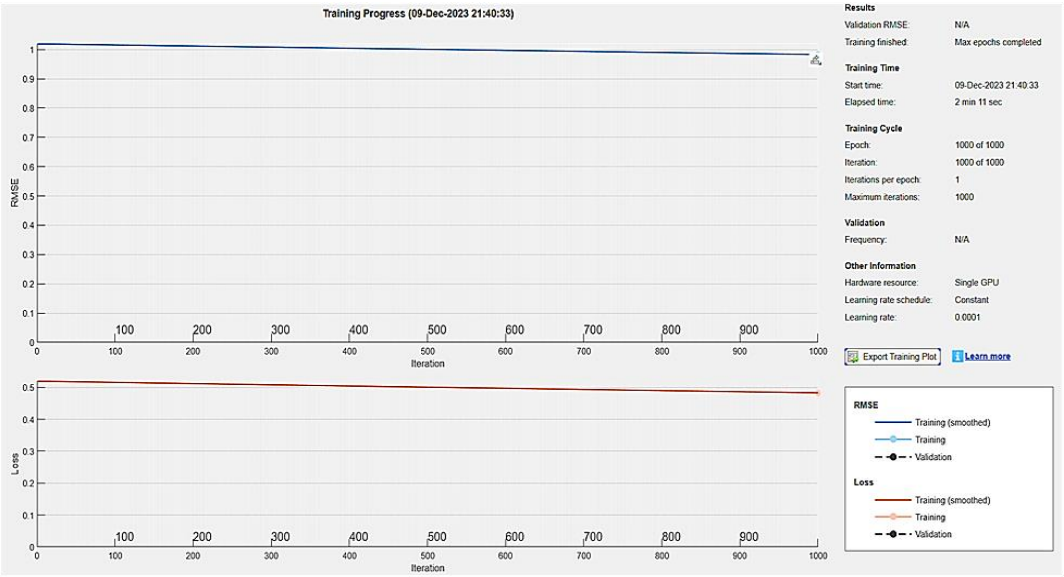


Figure 16. Proposed LSTM training using Sgdm optimizer

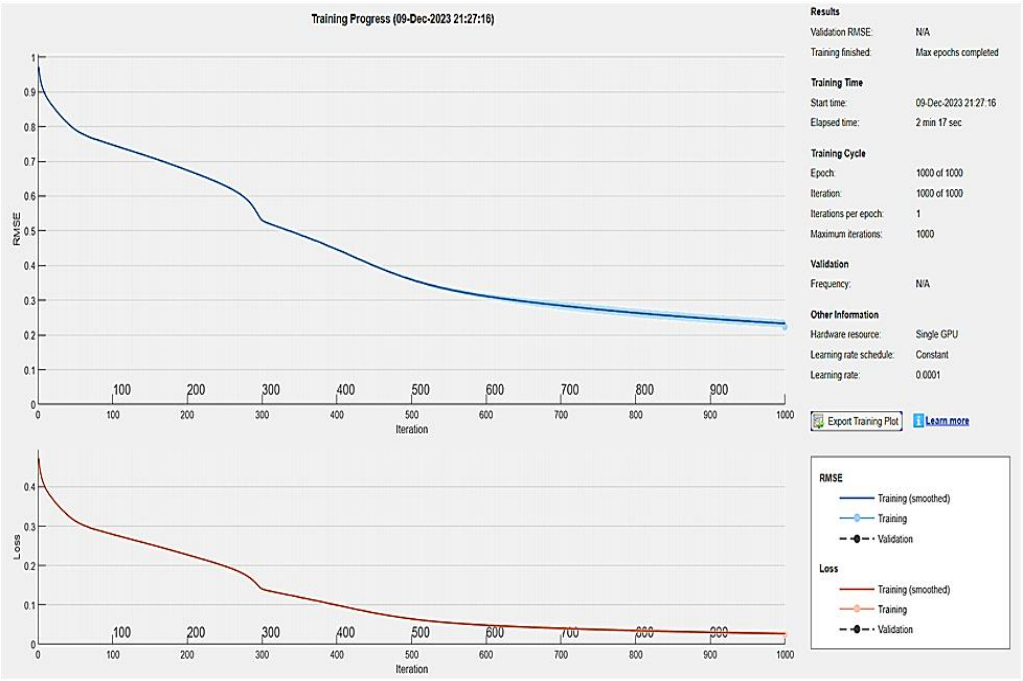


Figure 17. Proposed LSTM training using RMSprop optimizer

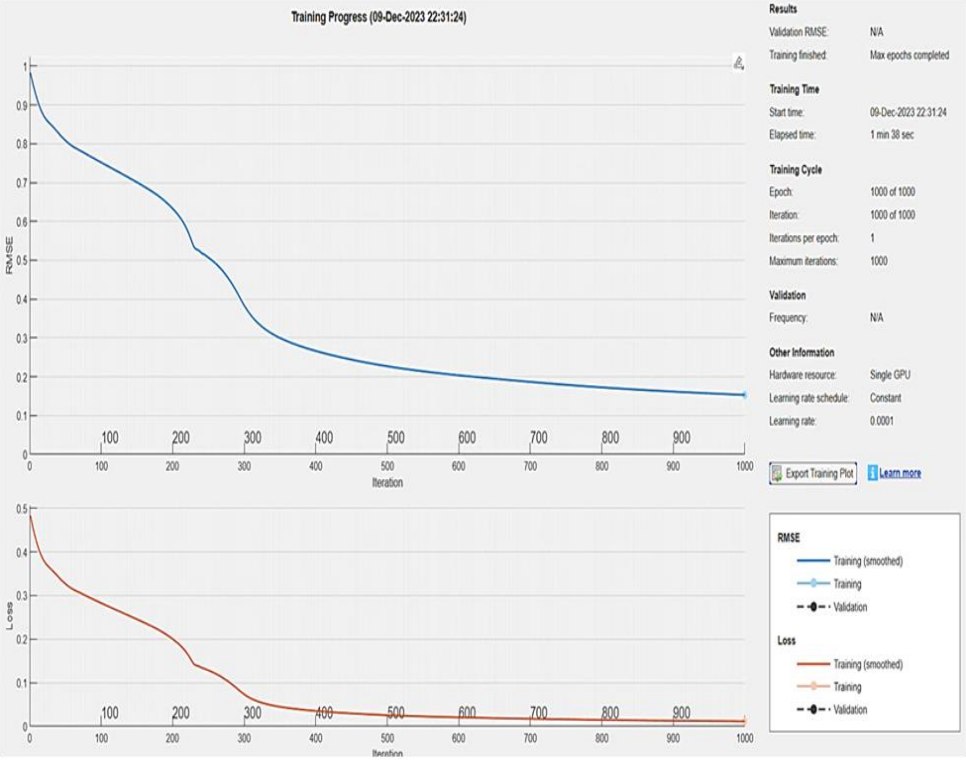


Figure 18. Proposed LSTM training using Adam optimizer

5.3.3 Classification Results

The test data is fed into the model after the training is finished. Out of 533 simulated cases, about 30% (160 cases) are used as test data. The idea is to use the proposed 1DCNN-LSTM to classify an image, which seemingly comes from some sort of processed time series. Here, a unique identifier (which represents one of the inrush current data) is converted to a PNG image. The model is required to reduce the image's size to 64 by 86 pixels. The model is used to classify the image, and the classify function returns the class label, as depicted in figure19. Convolutional layer activations on a picture are used to investigate features by comparing them to the equivalent areas in the source images. From a few layers, visual information can be gleaned. The earliest layers record basic visual characteristics.

Each activation is scaled such that it has an appropriate minimum value of 0 and maximum value of 1. Each channel in the layer has its own image, which makes up 64 activation images. Figure 20 illustrates an 8 by 8 grid where the activation is shown. Observing whether areas of image convolutional layers are active and comparing them to the corresponding areas of the original images allows one to assess features.

The results of validation test show that among 160 cases including different normal operations, inrush currents, internal faults and external faults under CT saturation, only one event related to an external fault under CT saturation is wrongly identified as an inrush current and all other cases are correctly identified and classified. It can be demonstrated that the accuracy of the proposed approach is about 99.4%. It is important to note that using the suggested algorithm, 100% internal faults have been identified. Compared to the recent approaches to recognition internal defects from others normal conditions presented in [10] and [15], the accuracy of the proposed approach is higher.

The accuracy of the suggested method is almost same as the accuracy (99.7%) of the method presented in [13]. However, the approach presented in [13] is dependent on the threshold value, and it is difficult to determine a suitable threshold that can be applied to different situations. If an inappropriate threshold level is chosen, it may cause false blocking or delayed differential relay operation. While our proposed approach is not dependent on a threshold value.

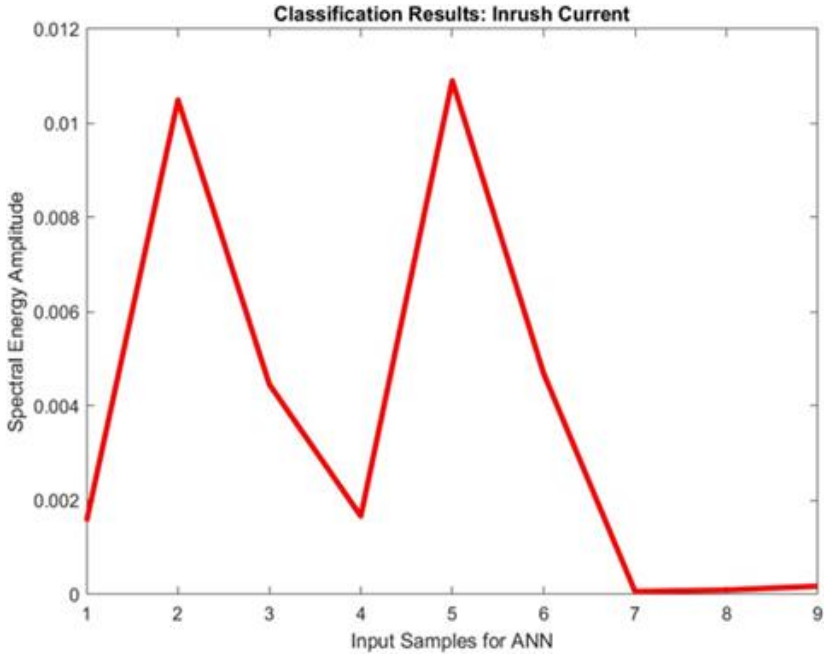


Figure 19. Classification result (Inrush Current)

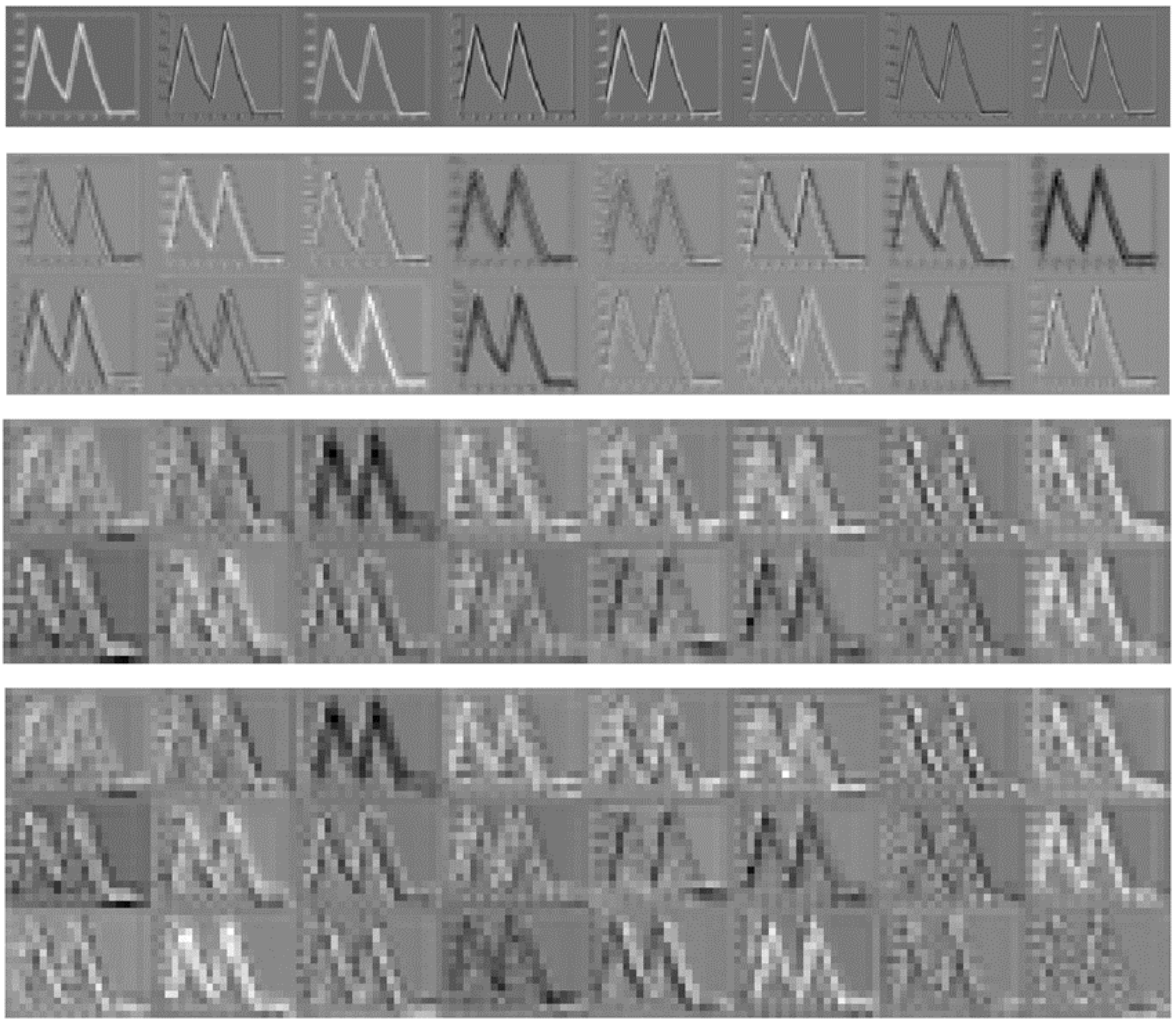


Figure 20. Activation conventional layers

6. Conclusion

To differentiate between inrush currents and internal defects in Pts, a novel approach based on DWT and 1DCNN-LSTM is introduced in this study. Based on the proposed approach, detail coefficients are extracted using db4 mother wavelet. Then, nine spectral energies of detail coefficients are calculated during 5 ms moving window. Each spectral energy moving window after converting into an image feeds the proposed 1DCNN-LSTM. Each moving window of spectral energy is considered as the input of the proposed 1DCNN-LSTM after being converted into an image. The suggested deep learning-based model is trained using 533 simulated cases including magnetizing inrush currents, transformer internal defects, transformer external faults under saturation of CT, and transformer normal operation. The effectivity of the proposed approach is verified using 160 simulated cases as test data. The results demonstrate that the proposed approach can precisely differentiate between inrush currents and internal defects, and its accuracy for the classification of simulated events is about 99.4%. Not needing any thresholds or transformer parameters and fast internal fault detection (after 5 ms) are other advantages of the proposed approach.

REFERENCES

[1] Stanley H. Horowitz and Arun G. Phadke, *Power System Relaying*, Fourth edi. Chennai, India, 2014.

[2] M. Raichura, N. Chothani, D. Patel, and K. Mistry, "Total Harmonic Distortion (THD) based discrimination of normal, inrush and fault conditions in power transformer," *Renew. Energy Focus*, vol. 36, no. March, pp. 43–55, 2021.

[3] EZECHUKWU O A. PhD, MNSE, MNIEEE, "DIFFERENTIAL PROTECTION-APPLICATION OF ZERO SEQUENCE CURRENT TRAP."

[4] A. Sahebi, H. Askarian-Abyaneh, S. H. H. Sadeghi, H. Samet, and O. P. Malik, "Efficient practical method for differential protection of power transformer in the presence of the fault current limiters," *IET Gener. Transm. Distrib.*, vol. 17, no. 17, pp. 3861–3871, 2023.

[5] H. Dashti, M. Davarpanah, M. Sanaye-Pasand, and H. Lesani, "Discriminating transformer large inrush currents from fault currents," *Int. J. Electr. Power Energy Syst.*, vol. 75, pp. 74–82, 2016, doi: 10.1016/j.ijepes.2015.08.025.

[6] X. Gong, Y. Zheng, S. Pan, C. Wu, and J. Deng, "Identification Methods for Transformer Turn-to-Turn Faults and Inrush Current Based on Electrical Information," *J. Phys. Conf. Ser.*, vol. 2215, no. 1, 2022.

[7] A. Wiszniewski and B. Kasztenny, "A Multi-Criteria Differential Transformer Relay Based on Fuzzy Logic," *IEEE Trans. Power Deliv.*, vol. 10, no. 4, pp. 1786–1792, 1995.

[8] R. P. Medeiros and F. B. Costa, "A Wavelet-Based Transformer Differential Protection: Internal Fault Detection during Inrush Conditions," *IEEE Trans. Power Deliv.*, vol. 33, no. 6, pp. 2965–2977, 2018.

[9] Maya P., S. V. Shree, Roopasree K., and K. P. Soman, "Discrimination of Internal Fault Current and Inrush Current in a Power Transformer Using Empirical Wavelet Transform," *Procedia Technol.*, vol. 21, pp. 514–519, 2015.

[10] Z. Kazemi, F. Naseri, M. Yazdi, and E. Farjah, "An EKF-SVM machine learning-based approach for fault detection and classification in three-phase power transformers," *IET Sci. Meas. Technol.*, vol. 15, no. 2, pp. 130–142, 2021.

[11] M. Y. Suliman and M. T. Al-Khayyat, "Discrimination Between Inrush and Internal Fault Currents in Protection Based Power Transformer Using Dwt," *Int. J. Electr. Eng. Informatics*, vol. 13, no. 1, pp. 1–21, 2021.

[12] M. Afrasiabi, S. Afrasiabi, B. Parang, and M. Mohammadi, "Power transformers internal fault diagnosis based on deep convolutional neural networks," *J. Intell. Fuzzy Syst.*, vol. 37, no. 1, pp. 1165–1179, 2019.

[13] R. Afsharisefat, M. Jannati, and M. Shams, "Discrimination of inrush current and internal faults incorporating the MRA and BIGRU techniques in power transformers," *Electr. Power Syst. Res.*, vol. 219, no. March, p. 109255, 2023.

[14] C. Li, N. Zhou, J. Liao, and Q. Wang, "Multiscale multivariate fuzzy entropy-based technique to distinguish transformer magnetising from fault currents," *IET Gener. Transm. Distrib.*, vol. 13, no. 12, pp. 2319–2327, 2019.

[15] P. Chiradeja and A. Ngaopitakkul, "Winding-to-ground fault location in power transformer windings using combination of discrete wavelet transform and back-propagation neural network," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, 2022.

[16] C. Zhang, Y. He, B. Du, L. Yuan, B. Li, and S. Jiang, "Transformer fault diagnosis method using IoT based monitoring system and ensemble machine learning," *Futur. Gener. Comput. Syst.*, vol. 108, pp. 533–545, 2020.

[17] P. Chiradeja *et al.*, "Application of probabilistic neural networks using high-frequency components' differential current for transformer protection schemes to discriminate between external faults and internal winding faults in power transformers," *Appl. Sci.*, vol. 11, no. 22, 2021.

[18] A. Behvandi, S. G. Seifossadat, and A. Saffarian, "A new Method for Discrimination of Internal Fault from Other Transient States in Power Transformer using Clarke's Transform and Modified Hyperbolic S-Transform," *Electr. Power Syst. Res.*, vol. 178, 2020.

[19] S. Jazebi, B. Vahidi, and M. Jannati, "A novel application of wavelet based SVM to transient phenomena identification of power transformers," *Energy Convers. Manag.*, vol. 52, no. 2, pp. 1354–1363, 2011.

[20] L. A. Yaseen, A. Ebadi, and A. A. Abdoos, "Discrimination between Inrush and Internal Fault Currents in Power Transformers Using Hyperbolic S-Transform," *Int. J. Eng. Trans. C Asp.*, vol. 36, no. 12, pp. 2184–2189, 2023.

[21] A. A. Nazari, P. D. Student, F. Razavi, and A. F. Associate, "A novel method to differentiate internal faults and inrush current in power transformers using adaptive sampling and Hilbert transform," *Iran. Electr. Ind. J.*, vol. Vol.11 / N, 2022.

[22] R. P. Medeiros, F. B. Costa, K. Melo Silva, J. D. J. C. Muro, J. R. L. Junior, and M. Popov, "A Clarke-Wavelet-Based Time-Domain Power Transformer Differential Protection," *IEEE Trans. Power Deliv.*, vol. 37, no. 1, pp. 317–328, 2022.

[23] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, p. 107398, 2021.

[24] A. Shenfield and M. Howarth, "A novel deep learning model for the detection and identification of rolling element-bearing faults," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–24, 2020.

[25] Z. Li, N. Qu, X. Li, J. Zuo, and Y. Yin, "Partial discharge detection of insulated conductors based on CNN-LSTM of attention mechanisms," *J. Power Electron.*, vol. 21, no. 7, pp. 1030–1040, 2021.