Fast GRNN-Based Method for Distinguishing Inrush Currents in Power Transformers

Afrasiabi, Shahabodin; Afrasiabi, Mousa; Parang, Benyamin; Mohammadi, Mohammad; Samet, Haidar; Dragicevic, Tomislav

Published in:
IEEE Transactions on Industrial Electronics

Link to article, DOI:
10.1109/TIE.2021.3109535

Publication date:
2022

Document Version
Peer reviewed version

Fast GRNN-Based Method for Distinguishing Inrush Currents in Power Transformers

Shahabodin Afrasiabi, Mousa Afrasiabi, Benyamin Parang, Mohammad Mohammadi, *Member, IEEE*, Haidar Samet, *Member, IEEE*, Tomislav Dragicevic, *Senior Member, IEEE*

***Abstract*—Differential protection, as the key protection element in the power transformers, has always been threatened with sending false trips subjected to external transient disturbances. As a result, differential protection needs an additional block to distinguish between internal faults and external transient disturbances. The protection system should i) be able to perform based on raw data, ii) be able to learn fully temporal features and sudden changes in the transient signals, and iii) impose no assumption on noise. To address these challenges, a fast recurrent neural network, namely fast gated recurrent neural network (FGRNN). By removing reset gate in the gated recurrent unit (GRU), the proposed network is capable of learning abrupt changes in addition to significantly reducing the computational time. Furthermore, a loss function based on an information theory concept is formulated in this paper to enhance the learning ability as well as robustness against non-Gaussian/Gaussian noises. A generalized form of mutual information is also adopted to form a noise model-free loss function, then incorporated with the designed deep network. Simulated and experimental examinations engaging various external factors, in addition to comparison between the proposed fast GRNN, GRU and seven firmly-established methods indicates the faster and more reliable performance of the proposed algorithm.**

***Index Terms*—Differential protection, Power transformers, Inrush current, Internal fault, Deep learning, Fast gated recurrent neural network (FGRNN)**

I. Introduction

A. Motivation

Power transformers are essential devices in power system. Therefore, protection of power transformers is crucial to prevent large energy outages as a catastrophic effect of faults [1]. Differential protection has been the main protection scheme in power transformers because of its simple principle and efficiency . When a power transformer is energized initially or experiences a sudden change in its terminals, a typically non-sinusoidal transient current flows in transformer windings which abruptly reaches a large value in the first half cycle [2].Although differential protection has been widely studied in the previous investigations, there are several main challenges that motivated us to present the current work, including: i) performance in an extremely short-time period (less than 10 ms); ii) capturing the fully temporal information of the transient differential current signal without neglecting the transient changes; iii) being able to perform in the real conditions without imposing any assumption on measurement noises; iv) can perform accurately considering eternal factors such as current transformer (CT) saturation, series capacitor compensation, and fault current limiter (FCL) [3].

M.Afrasiabi, S.Afrasiabi, B.Parang, M.Mohammadi and H.Samet are with the School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran (e-mail: shpower77@yahoo, musa.afra@shirazu.ac.ir, benyamin.parang@shirazu.ac.ir, m.mohammadi@shirazu.ac.ir, samet@shirazu.ac.ir).

M.Afrasiabi is also with Department of Electrical Engineering, School of Energy Systems, Lappeenranta University of Technology, Lappeenranta, Finland.

H.Samet is also with Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands (Corresponding Author).

T.Dragicevic is with the Center of Electric Power and Engineering, Technical University of Denmark, 2800 Kgs.Lyngby, Denmark (e-mail: tomdr@elektro.dtu.dk).

B. Brief literature Review

Distinguishing internal faults from inrush currents have been widely investigated in researches using various methods. In term of the basic principles, we can classify them into three categories: signal processing, model-based, and artificial intelligence methods. A popular group is based on analyzing measurement signals based on spectral analysis techniques [1]. Model-based methods include establishing and accurate estimation using least square [4] and extended Kalman filter [5]. In spite of fast performance and easy implementation, these two families have deficiencies such as sensitivity to noises and threshold values and dependency on different models in varying conditions.

The original idea of using intelligent classifiers to identify transformer inrush and faults was published in 1994 [6]. Since 1994, several shallow-based structures have been presented to distinguish between internal faults and inrush currents. For instance, artificial intelligence neural network (ANN) [7], relevance vector machine (RVM) [8], support vector machine (SVM) [9], learning vector quantization (LVQ) [10], and k-nearest neighbor (KNN) [11], are some of the previously presented investigations. Their shallow structures cannot distinguish internal faults when data have a complex structure. For instance, they will experience a decrease in accuracy when the fault data is affected by external factors such as FCL which limits the magnitude of fault current to a value close to the magnitude of inrush current and may cause problems for differential protection in detecting the fault. As the effect of external factors is inevitable in practice, we need more reliable methods. Machine learning methods with shallow architecture are not able to extract features from raw data. Therefore, they need to be attached to feature extractors like WT. Since there are various types of feature extractors, these methods cannot work with different kinds of data-sets generally.

In conclusion, there are serious challenges for a practical supplementary material in the in power transformers protection that are neglected:

- Establishing a general tool to distinguish the internal faults and other disturbances in a half-cycle time period (less than 10ms).
- Capturing the fully temporal information of the transient differential current signal without neglecting the transient changes.
- Being able to perform in the real conditions without imposing any assumption on measurement noises.
- General solution to perform accurately as well as fast considering external factors CT saturation, FCL, and series capacitor compensation.

To overcome these challenges and above-mentioned problem of the previous methods, using deep learning structures can be a potential solution. Deep Learning is a fairly new established area of machine learning that can serve in many fields especially fault diagnosis and fault classification [12]–[15]. Machine learning methods with deep architecture are capable of capturing information from raw data by establishing several information processing layers. Deep learning models might potentially overcome the mentioned deficiencies by establishing lots of information processing layers which are able to capture information from data without any preprocessing. It is more powerful and efficient than shallow architecture neural networks because it approximates complex problems by learning the deep nonlinear structure. Generally, there are four type of deep neural networks including deep auto-encoder (DAE), deep Boltzmann machine (DBM), convolutional neural network (CNN), and recurrent neural network (RNN) [16]. Although DAE and DBM improve the capability of learning from raw data through a dimensional reduction procedure, the main disadvantage of DAE and DBM is the disability in understanding long sequences. CNN is a widely used deep structure method in time series analysis. Despite the great performance in capturing spatial features, CNN is unable to fully realize the temporal feature, especially in long-tailed time series associated with abrupt changes. Besides, CNN, DAE, and DBM cannot detect the transient disturbances in an extremely short-term period and perform in real-time manner. To this end, in [17] an accelerated convolutional neural network (ACNN) has been presented for the differential protection of the power transformers. In [17], the quantization process has been utilized to speed up CNN, however, the quantization process is highly sensitive to noise. In presence of Gaussian/non-Gaussian noises, ACNN has adverse effects and poor performs. As in the real applications where noise is an unavoidable phenomenon, the application of ACNN for realistic power transformers is questionable. Recurrent neural networks are a variation of deep neural networks in which the recurrent connections allow the hidden units to see their own previous outputs and shape their next output based on them. That is what gives the RNNs memory [18]. Long short-term memory (LSTM) neural networks are alternative forms of RNNs which use gates. Gates enable the network to add, remove, or send information through the hidden state layer. This gives LSTM a long-term memory. Removing one of the three gates in LSTM structure, forms another structure called gated recurrent unit (GRU) which is faster and more accurate [19]. GRU generally consists of two main gates, i.e. update and reset gate.

However, a standard GRU-based method has three major deficiencies to apply in the real-time differential protection system, i) performance time of GRU is still high for implementation in the real-time differential protection system, ii) reset gates can reset the network during the training process and might neglect a set of sudden changes in transient phenomena such as inrush current and internal faults in the power transformers, iii) they are not able to perform efficiently in highly noise conditions.

C. Contributions and Organization

In order to propose a diagnosis scheme to implement a real-time differential protection, this paper develops a GRU-based structure. To decrease the computational burden as well as improving the accuracy and reliability of the differential protection, in the proposed deep network, namely fast gated recurrent unit (FGRNN), the reset gate has been removed. By removing the reset gate, the computational complexity has significantly reduced (almost 42% in the diagnosis process). Furthermore, transient phenomena such as inrush current and internal faults in the power transformers usually follow a set of sudden changes that might be neglected in the training process of the GRU networks by completely resetting the network. In the designed FGRNN, the sudden changes due to transient behavior of internal faults and other transient disturbances are not neglected. Learning sudden changes in the power transformers, that can be beneficial in the discrimination process, is completely fulfilled in the designed FGRNN network, as well as learning fully temporal features.

Noise is an unavoidable and undesirable phenomena in power systems. Based on [20], the majority of noises during voltage and currents in the power systems follow non-Gaussian noises. As best of our knowledge, previously presented works on the discrimination between internal faults and other disturbances neglected the noise impact [21], or only considered Gaussian noises [5]. Thus, it is the first paper that investigates the non-Gaussian noise impact on the differential protection in the power transformers. This paper develops a new probabilistic loss function in which the proposed network can perform without any assumption on the noise model. To this end, an information theory concept based on mutual information is utilized to enhance the robustness against noises as well as improving the learning ability of the designed FGRNN.

Furthermore, the proposed differential protection scheme preserves accuracy in case of CT saturation, series capacitor, and the presence of superconducting fault current limiter (SFCL). In the mentioned conditions, discrimination between internal faults and other transient disturbances is a difficult task due to decreasing the internal fault magnitude in presence of SFCL, the error of CT saturation corrector, and higher current magnitude of external disturbances in presence of series compensators.

Thus, we can summarize the contributions and novelties of this work, as:

- A GRU-based deep network is designed for the real-time differential protection. The standard (GRU) is modified by removing the reset gate, that enhances the accuracy and reliability of the proposed network by learning the transient abrupt changes. The computational complexity has been significantly reduced as well.
- Performance of the proposed deep network is very accurate and fast using only half-cycle of raw data considering external factors, e.g. SFCL, CT saturation, and series compensators.
- A probabilistic loss function is formulated based on information theory principles to improve the learning ability as well as robustness without any assumption on noise model (Gaussian or non-Gaussian).

The remainder of the paper is organized as follows. Section II represents the principles of inrush current and differential protection scheme. The FGRNN based method is described in Section III. Section IV and V represents the simulation results and comparison of methods. Finally, Section VI concludes the paper.

II. FGRNN Architecture

Different parts of the structure of the proposed FGRNN technique in differential protection is presented in this section.

A. Standard GRU

The standard GRU consists of update and reset gate. The reset gate removes the unnecessary detected features by neglecting abrupt and sudden changes in a time series. The update gate makes each of the recurrent units capture dependencies of different time scales and sequences adaptively. In particular, the standard GRU architecture is defined by the following equations:

$$u(m,L,t) = f\{W_u y(m,L,t) + R_u h(L,t-1) + B_u\} \quad (1)$$

$$r(m,L,t) = f\{W_r y(m,L,t) + R_r h(L,t-1) + B_r\} \quad (2)$$

$$h(L,t) = [1-u(m,L,t)] \odot h^{'}(L,t) + ... \\ u(m,L,t).h(L,t-1) \quad (3)$$

$$h^{'}(L,t) = f\{W_h y(m,L,t) + ... \\ R_h(h(L,t-1)\dot{r}(m,L,t)) + B_h\} \quad (4)$$

$u(m,L,t)$ represents the update gate for m^{th} output map at the L^{th} layer in the t time interval, while $h(L,t)$ and $h'(L,t)$ are the hidden state and the candidate state, respectively. W and R show the weight matrices corresponding to the input vectors and recurrent parameters. Subscript u, and h relate to update gate and hidden state.

The utilized activation function in this paper is the rectified linear unit (ReLU). ReLU is an activation function to resolve vanishing gradient problems and prevent significant saturations in pre-training [22] which is presented in [23].

B. Removing Reset Gate

As mentioned before, reset gate is a useful tool in a time series with discontinuities. In a differential current signal, the abrupt changes can determine the difference between transient disturbances and internal faults. The rest gate might ignore sudden changes and therefore reduce the accuracy and reliability. By removing the reset gate, the candidate hidden state is reformed as:

$$h^{'}(L,t) = f\{W_h y(m,L,t) + R_h h(L,t-1) + B_h\} \quad (5)$$

Removing reset gate in the typical GRU leads to a sufficient and compact model in the real-time classifier.

C. Batch Normalization

Distribution of activation function might change during the training process due to the change in FGRNN parameters known as the internal covariate shift. Batch normalization is proposed to resolve this problem [18], [24]. Batch normalization can be extended to FGRNN parameters and weight matrices. In this paper, batch normalization is used in weight matrixes [18].The batch normalization $Bn(L)$ at the L^{th} layer presented in [24] can be described as

$$Bn(L) = s \odot \frac{L-\mu_b}{\sqrt{{\sigma_b}^2+\xi}} + \beta(L) \quad (6)$$

where s , μ_b, σ_b, ξ and $\beta(L)$ represent the scale parameter of the training, minibatch mean and variance, stability enhancement parameter, and shifting parameters, respectively. Scale parameter s, and shifting parameters $\beta(L)$ are used to restore the network capacity and remove the bias matrices to reduce computational complexity.

D. FGRNN

Consequently, by removing the reset gate and applying the batch normalization, the proposed FGRNN is described as follows:

$$u(m,L,t) = f\{Bn[W_u y(m,L,t)] + R_u h(L,t-1)\} \quad (7)$$

$$h(L,t) = [1-u(m,L,t)] \odot h^{'}(L,t) + ... \\ u(m,L,t).h(L,t-1) \quad (8)$$

$$h^{'}(L,t) = f\{Bn[W_h y(m,L,t) + R_h h(L,t-1)]\} \quad (9)$$

E. Dense Layer

In the last layer of the proposed FGRNN technique, the dense network is added to connect all hidden states in the FGRNN and control the dimension of the FGRNN output. The m^{th} output map at the L^{th} layer passes through a dense layer to construct the final output map as

$$y(m,L) = W_f.f[W.y(m,L)] \quad (10)$$

where W_f represents the weight matrix of the dense layer.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2021.3109535, IEEE Transactions on Industrial Electronics

F. Loss Function

The basis of the proposed loss function is information theory, where for each random variables R, the Shannon entropy has defined as:

$$f^{Shan}(R)=\sum_{r=1}^{R}P[R=r]\log P[R=r] \tag{11}$$

where $f^{(Shan)}$ measures the uncertainty associated with random variable R. Based on the Shannon entropy, the mutual information (MI) has been defined, and for two random variables R_1 and R_2 is:

$$f^{MI}(R_1,R_2)= \sum_{r_1,r_2}P[R_1=r_1,R_2=r_2]\log\frac{P[R_1=r_1,R_2=r_2]}{P[R_1=r_1]P[R_2=r_2]} \tag{12}$$

Property 1: For three different random variables R_1, R_2, and R_3, where R_3 is less informative than R_1 and R_2, thus:

$$f^{MI}(R_3,R_2)\le f^{MI}(R_1,R_2) \tag{13}$$

Since R_2 is the most informative variables for itself:

$$f^{MI}(R_1,R_2)\le f^{MI}(R_2,R_2)\triangleq\Lambda(R_2) \tag{14}$$

The data-driven network for internal fault detection in the power transformers, $\Theta(X)$ can be constructed based on $f^{MI}[\Theta(X),Y]$.

Property 2: Considering measurement and process noise in power systems, $\Theta(X)$ should construct a set of $\hat{Y}$ (fault/non-fault) based on the noisy outputs $\tilde{Y}$ and $\tilde{X}$ inputs (measured data by CTs and process noises in the power transformers). There is a possibility that $\forall\Theta,\Theta'$, if $f^{MI}[\Theta(X),Y]>f^{MI}[\Theta'(X),Y]$, might not lead to $f^{MI}[\Theta(X),\hat{Y}]>f^{MI}[\Theta'(X),\hat{Y}]$, hence, the general form of MI cannot directly implement in the differential protection of the power transformers.

To tackle this problem, a generalized form of MI has been adopted from [25]. The general MI (GMI) is defined for two different random variables, R_1 and R_2:

$$f^{GMI}(R_1,R_2)=|\det(J_{r_1,r_2})| \tag{15}$$

where J_{r_1,r_2} shows the joint distribution of R_1 and R_2 in the matrix form.

Property 3: GMI represents non-negative and symmetric outputs, where for three different random variables, R_1, R_2, and R_3 (R_3 is less informative than R_2), then:

$$f^{GMI}(R_2,R_3)= f^{GMI}(R_2,R_1)\left|\det\left[P\begin{pmatrix}R_3=r_3|\\ R_1=R_1\end{pmatrix}\right]\right| \tag{16}$$

Thus, GMI can be generalized for classification issues, because:

$$f^{GMI}[\Theta(X),Y]>f^{GMI}[\Theta'(X),Y]\Leftrightarrow f^{GMI}\left[\Theta(X),\hat{Y}\right]>f^{GMI}\left[\Theta'(X),\hat{Y}\right] \tag{17}$$

To this end, an informatics-theory based loss function is defined as:

$$f^{loss}\left\{P^J\left[h(X),\hat{Y}\right]\right\}=-\log\left\{f^{GMI}\left[h(X),\hat{Y}\right]\right\}= -\log\left[\left|\det\left(P^J\left[h(X),\hat{Y}\right]\right)\right|\right] \tag{18}$$

The joint distribution function is shown by $P^J(\bullet)$, whose size depends on the class, if each class (fault/non-fault) is shown by the c, the size of $P^J(\bullet)$ is $c\times c$. To resolve the scaling problem due to the matrix form of $P^J(\bullet)$, the log function has been used. In each iteration of the training process, raw data obtained by the CTs have been sampled associated with noisy labels, $(X_k,\hat{Y}_k)\ \forall k=1,...,N$, where k is the number of the samples and N shows the total number of samples in one record that is dependent on the sampling frequency of the CTs. Let be denote to the output of the FGRNN classifier, $f^{out}(\bullet)$, the output of the FGRNN is the distribution probability based on the proposed probabilistic loss function. By taking noisy labels a value between zero and one, showing by M^n with the size of $1\times N$ and defined as:

$$M_i^n=\mathbb{C}[y_i=\hat{c}] \tag{19}$$

where $\mathbb{C}[y_i=\hat{c}]$ is one, when y_i replaces in class $\hat{c}$; otherwise is a zero.

The output of the proposed FGRNN based on the probabilistic loss function is defined as:

$$f^{out}(c,\hat{c})=P\frac{1}{N}\sum_{j=1}^{N}f_{FGRNN}^{out}M_j^n= \frac{1}{N}\sum_{j=1}^{N}\Theta(X_i)\,\mathbb{C}[y_i=\hat{c}] \tag{20}$$

The proposed loss function is robust against non-Gaussian/Gaussian noises due to:

Proof: For classifier Θ with proposed loss function:

$$\arg\min_{c\in\Theta}f^{GMI}[\Theta(X''),Y'']=\arg\min_{c\in\Theta}f^{GMI}[\Theta(X),Y] \tag{21}$$

where X'' and Y'' is the noisy input and output, respectively. Therefore, in the training process:

$$f^{GMI}[\Theta(X''),Y'']=f^{GMI}[\Theta(X),Y]+\kappa \tag{22}$$

where κ is the constant. Accordingly, the proposed loss function cannot be influenced by the noises, regardless of whether the noise is Gaussian or not.

G. Training Process

To find the optimal learning weight of the designed FGRNN network based on the proposed loss function, an iterative based process has been conducted:

1) Determining the epoch number, learning rate, and batch size.
2) Pre train the FGRNN based on the generated dataset and cross-entropy loss function.

3) Randomly select a set of the batch normalized sample from the training dataset
4) Evaluate the samples based on the proposed probabilistic loss function based GMI
5) Update the learning weights
6) Check the batch numbers, if all of them has been evaluated go to step 7; otherwise back to step 3
7) Sort the learning weights based on the proposed loss function and select the minimum values

H. Overall Structure

As shown in Fig. 1, the proposed structure of FGRNN is composed of different layers. The differential current sample dataset are revealed in different layers of the proposed FGRNN. Different outputs are constructed for each layer. Differential current sample is converted to $s \times 56$ (number of samples in a half cycle). Then, in terms of arranging samples vectors are shaped to 4D tensors. Three FGRNN layers are used to extract the temporal dependency in $s \times 64$ vector. The $s \times 64$ vector passes through a dense layer with ReLU activation function. Consequently, internal fault is identified in the dense layer based on the zero or one outputs of *sigmoid* activation function. A one means that the event is diagnosed as fault by the network.

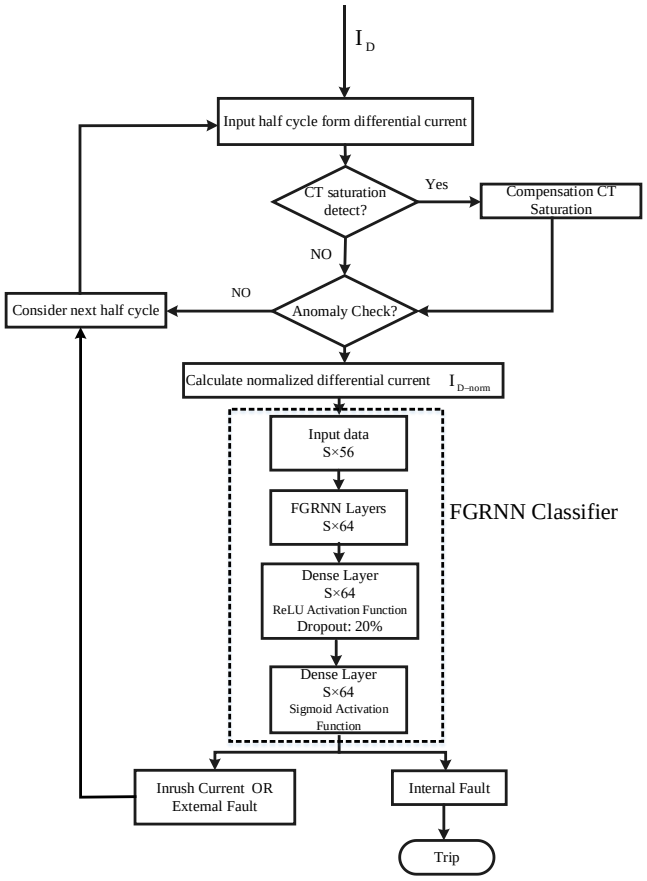


Fig. 1. Structure of the proposed FGRNN for differential protection scheme

I. Proposed Differential Protection Scheme

The proposed differential protection scheme of the power transformers have shown in Fig 1. To implement the proposed differential protection scheme, the following steps should be conducted:

Step1: A half-cycle of current signal is measured by CTs (112 samples per cycle [4]).

Step 2: The measured current by the CTs are checked for saturation. If the CT saturation is detected, the discorded current are corrected by the presented method [17].

Step 3: Differential current is computed based on (3) to detect the abnormality. If the abnormal condition recognize, then the input data should send to the designed robust FGRNN block; otherwise the next half-cycle back to Step 1.

Step 4: The robust FGRNN block detects internal fault rom other transient disturbance; if the internal fault is diagnosed, then a trip signal send; otherwise the next half-cycle have considered and back to Step1.

III. Numerical Results

In this section, five case studies are discussed to investigate the superiority of the proposed differential protection technique. The case studies are briefly as follows:

1) A simple case without external factors.
2) Differential protection in the presence of CT saturation.
3) Series capacitor compensation.
4) Transformer with SFCL on the neutral point.
5) Experimental prototype.

The following reliability metrics are defined to evaluate the performance of the proposed protection technique [26].

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}\% \tag{23}$$

$$DEP = \frac{TN}{TN + FP}\% \tag{24}$$

$$\text{SEC} = \frac{TP}{FN + TP}\% \tag{25}$$

$$SAF = \frac{TP}{TP + FP}\% \tag{26}$$

where *TP*, *TN*, *FP*, and *FN* show true positive, true negative, false positive, and false negative, respectively. ACC, DEP, SEC, and SAF stand for accuracy, dependability, security, and safety which illustrate the correct detection percentage, diversity of the proposed classifier, false trip ratio, and reliability of the correct decisions in the proposed protection technique, respectively.

Different fault types, fault locations, switching angles, transformer winding types, and source impedances are used to generate datasets for the mentioned case studies. The datasets are generated in PSCAD/EMTDC software package by simulating a 230 kV network, showed in Fig. 2. Sampling frequency is 112 samples per second. The generated datasets are processed in Keras package [27] on a computer with a 3.4 GHz processor, 32 GB memory, and a GeForce GTX 1080 TI graphical processing unit (GPU). Based on the procedure

presented in [28] the samples are divided into random training and testing sets with 70% to 30% ratio to train the network sufficiently, avoid fitting problems, and yet have enough data samples to evaluate the performance of the trained network.

To evaluate the performance of FGRNN-based protection scheme, a half cycle from each sample are extracted and normalized as follows

$$I_{D-norm} = \frac{I_P - I_S}{\max(|I_P - I_S|)} \tag{27}$$

where I_{D-nom} is the normalized differential current, which is considered as the input of the FGRNN method. The time duration for one cycle is 20 ms (nominal frequency is 50 Hz) and sampling frequency is 112 samples per cycle.

The number of samples in cases 1 ,3, 4, and 5 is 4488 samples. In case 2, overall, 2064 samples are generated. The samples are divided into random training and testing sets with 70% to 30% ratio. The time duration for one cycle is 20 ms (nominal frequency is 50 Hz) and sampling frequency is 112 samples per cycle. To train and test the proposed method and other method, a half cycle from each sample are extracted with size , therefore, each sample length is 56. Similar to each multi-class classification problem, the output labels are binary labels. In cases 1, 3, 4, and 5, LG fault, LL fault, LLL fault, P-S fault, inrush phenomena, and external faults are assigned by [0 0 1], [0 1 0], [0 1 1], [1 0 0], [1 0 1], [0 0 0], and [1 1 0], respectively. In case 2, the output labels are zero and one. Table I shows the parameters and labeling of the designed robust FGRNN.

The comprehensive evaluation of each case is presented in this section. To this end, the overall results obtained from the simulation and experiments are presented in several metrics, which are discussed individually in the following sub-sections.

TABLE I
PARAMETER OF THE DESIGNED ROBUST FGRNN

Case	Case Description	Output Labels
Case 1	4488 Samples 70%:30% Train:Test Ratio 1x56 Sample size 56 sample length	Inrush Current: [0 0 0]
		LG Fault: [0 0 1]
		LL Fault: [0 1 0]
		LLL Fault: [0 1 1]
		P-S Fault: [1 0 0]
		External Fault: [1 0 1]
Case 2	2064 Samples 70%:30% Train:Test Ratio 1x56 Sample size 56 sample length	Internal Fault Saturated 0
		Inrush Current and External Fault Saturated 1
case 3	4488 Samples 70%:30% Train:Test Ratio 1x56 Sample size 56 sample length	Inrush Current: [0 0 0]
		LG Fault: [0 0 1]
		LL Fault: [0 1 0]
		LLL Fault: [0 1 1]
		P-S Fault: [1 0 0]
		External Fault: [1 0 1]
Case 4	4488 Samples 70%:30% Train:Test Ratio 1x56 Sample size 56 sample length8	Inrush Current: [0 0 0]
		LG Fault: [0 0 1]
		LL Fault: [0 1 0]
		LLL Fault: [0 1 1]
		P-S Fault: [1 0 0]
		External Fault: [1 0 1]
Case 5	4488 Samples 70%:30% Train:Test Ratio 1x56 Sample size 56 sample length8	Inrush Current: [0 0 0]
		LG Fault: [0 0 1]
		LL Fault: [0 1 0]
		LLL Fault: [0 1 1]
		P-S Fault: [1 0 0]
		External Fault: [1 0 1]

A. Case 1

In the first case, the dataset contains 4488 samples for fault current, inrush current and external fault excluding any external factors. Fault currents are generated considering different fault types (line to ground (LG), line to line (LL), three phase (LLL), and primary to secondary (P-S)), winding connections, source impedances, fault inception angles, and fault locations. different residual fluxes, winding connections, source impedances, switching instant fault inception angles, and loads are involved in fault simulation. External fault data set contains 8 fault types, 12 fault inception angles, 4 impedance values, and 6 fault locations. The information on fault locations are represented as the distance from the corresponding transformer in form of percentage of corresponding line. The distances are 5%, 25%, 45%, 65%, 75%, and 90%.

The numerical results of detecting four internal faults, inrush current and external faults using the proposed method is represented in Table II. The results indicate the precision in detecting LG faults, inrush current and, and external faults, while for other internal faults the performance metrics are at least 98.47%. As this case is a simple benchmark, the other cases are designated for further evaluation of the proposed protection scheme. The proposed protection algorithm represent 100% performance in case of external faults which shows it is immune against cross-country faults.

B. Case 2

When short circuit fault or inrush phenomena occur, the measured currents experience a considerable distortion caused by CT saturation. The protection systems operate based on the secondary side current of CTs. Therefore it makes it difficult for differential protection system to detect the internal fault current or distinguish it from inrush current [29]. The effect of CT saturation is simulated in this case to take the large DC component of inrush current and decay DC component of short-circuit fault [30] into account. A reliable differential protection should be able to compensate the saturation effect to minimize measurement errors. Therefore, A CT saturation compensation stage is required prior to the abnormal condition detection. The CT saturation compensation method used in our paper is a least square (LS) based wave shape-independent method comprised of two filters based on [30]. The first filter reproduces the deformed saturated CT output while the second one compensate DC offset.

TABLE II
PERFORMANCE OF PROPOSED FGRNN METHOD IN CASE 1

Cases	ACC%	DEP%	SEC%	SAF%
LG Fault	100	100	100	100
LL Fault	99.88	99.23	100	99.86
LLL Fault	99.88	100	99.86	100
P-S Fault	98.91	98.47	99.00	99.71
Inrush	100	100	100	100
External Fault	100	100	100	100

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2021.3109535, IEEE Transactions on Industrial Electronics

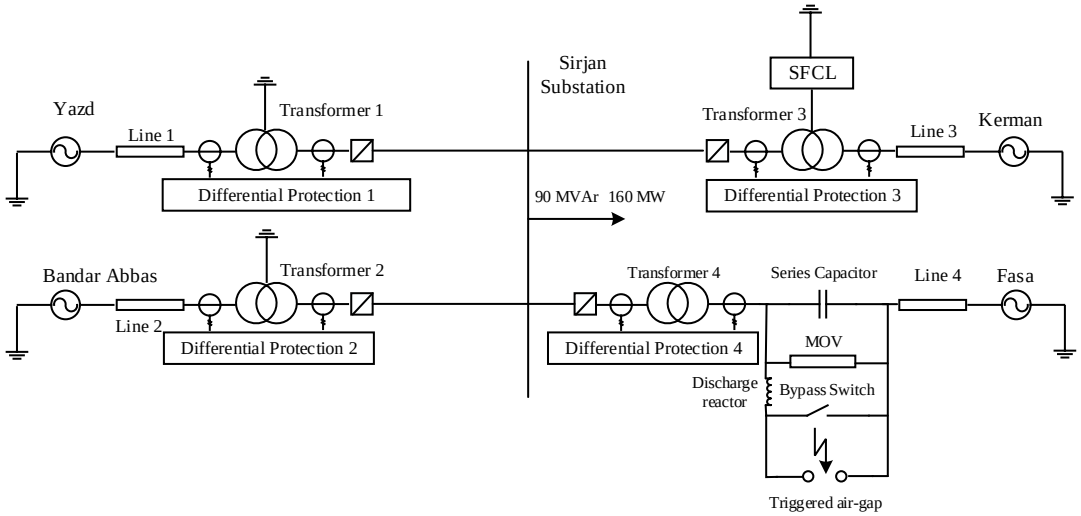


Fig. 2. Single line diagram of a modified part of the 230 kV Iranian power network

Jiles-Atherton model for CT [31] is used to simulate short circuit fault and inrush current occurrence in Transformer 2. The effect of inrush current on CT saturation is survayed in previous studies very rarely. 1200 fault current CT saturation samples are generated in different conditions containing 4 different CT burdens, 5 fault types, 12 fault inception angles, and 5 different fault resistances including 0, 2, 4, 8, and 10 ohm fault resistances. 864 samples are generated for inrush current CT saturation considering 12 switching inception angles, 4 CT burdens, 3 different load levels, 2 winding types, and 3 different source impedance values.

Table III illustrates the impacts of CT saturation compensation method on FGRNN-based protection scheme. Despite the inevitable error in saturation compensation, the method operates with 99.19% accuracy in these two conditions. Compared to Case 1, CT saturation does not have a significant effect on the FGRNN-based protection scheme performance. DEP values are higher than 99.17% which proves that the FGRNN-based method is highly capable of isolating the internal fault in conditions such as CT saturation by inrush current or internal fault. The results show that the proposed protection algorithm is able to cover the errors of saturation compensation errors to prevent any malfunction in differential protection system. Again, in this case it has been observed that the proposed algorithm performs fairly well in presence of cross-country faults.

TABLE III
PERFORMANCE OF PROPOSED FGRNN METHOD IN CASE 2

Cases	Internal Fault saturated	Inrush Current and External Fault Saturated
ACC%	99.19	99.19
DEP%	99.44	99.84
SEC%	99.84	99.44
SAF%	99.22	99.17

C. Case 3

Series capacitor compensation is a common way to improve power transferring capacity, voltage regulation, stability, and loss. Therefore, this case examined the differential protection scheme in presence of series capacitor. Series capacitor compensation can cause sudden changes in angle and sequence of three phase currents and lead to an increase in the fault current. The simulation results for Transformer 4 and Protection system 4 are presented for this case. The datasets for inrush current and external faults are similar to Case 1 and for internal faults four different distances to Transformer 4, three different compensation rates, 12 fault inception angels, three different impedances, and five fault types are taken into account.

Table IV shows the performance of the proposed FGRNN method in presence of series capacitor compensation. The results prove the reliable performance of the proposed method. As can be seen, the lowest performance metric is 98.66% which shows the great performance of the FGRNN method. The minimum and maximum values of SAF are 100% and 99.71%, respectively. These values show the ability of the FGRNN-based method in isolating hazardous short circuit faults in presence of series capacitor compensation.

D. Case 4

A high impedance in neutral point of the power transformer added by SFCL can limit fault current for less than a half cycle. Accordingly, the differential relay cannot detect the fault to send the trip signal. Therefore, in this case, a variable

TABLE IV
PERFORMANCE OF PROPOSED FGRNN METHOD IN CASE 3

Cases	ACC%	DEP%	SEC%	SAF%
LG Fault	99.28	98.51	99.43	99.71
LL Fault	99.04	98.62	99.13	99.76
LLL Fault	100	100	100	100
P-S Fault	99.71	98.92	99.46	99.00
Inrush	99.40	99.23	99.43	99.86
External Fault	99.52	100	99.14	100

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2021.3109535, IEEE Transactions on Industrial Electronics

TABLE V
PERFORMANCE OF PROPOSED FGRNN METHOD IN CASE 4

Cases	ACC%	DEP%	SEC%	SAF%
LG Fault	99.52	99.23	99.57	99.86
LL Fault	98.68	96.97	99.00	99.43
LLL Fault	99.04	99.23	99.01	99.86
P-S Fault	98.93	97.63	99.29	99.57
Inrush	98.98	98.99	98.97	99.22
External Fault	99.96	98.39	99.98	99.98

TABLE VI
PERFORMANCE OF PROPOSED FGRNN METHOD IN CASE 5

Cases	ACC%	DEP%	SEC%	SAF%
LG Fault	98.99	98.57	99.07	99.73
LL Fault	98.59	97.16	98.87	99.43
LLL Fault	99.15	99.23	99.71	99.86
P-S Fault	99.40	98.46	99.57	99.71
Inrush	99.18	98.31	99.08	99.48
External Fault	99.48	99.12	99.51	99.92

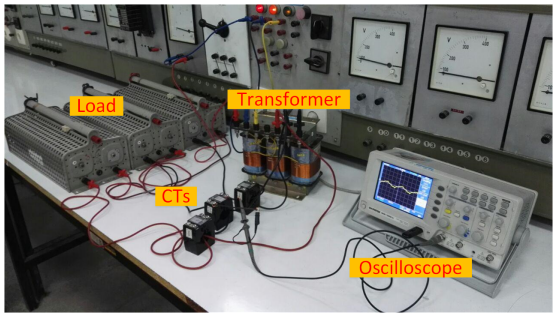


Fig. 3. Experimental prototype

resistance is used to fix the problem. This problem can happen during an inrush current, too. The SFCL model is added to Transformer 3 in Fig. 2. Parameters of SFCL are set based on the procedure presented in [21].

Similar dataset to Case 1 is used to evaluate the performance of the proposed FGRNN-based classifier. As can be seen in Table V, the lowest value of the accuracy is 99.04, which shows the effectiveness and robustness of FGRNN in presence of SFCL. The main concern about the implementation of SFCL in differentioal protection is the false trip for non-fault conditions. In this case, the fault current hs the least difference with inrush current. This makes the SEC an important factor in evaluating the protection scheme. SEC is between 98.97 and 99.98 which indicates the great performance of the proposed protection method in presence of SFCL.

E. Case 5

In this case, the proposed method is evaluated by an experimental prototype. The experimental data is gathered using a 1 kVA, 50 Hz, and 380/380 V three-phase transformer shown in Fig. 3. Several points inside the windings are accessible through terminals which are used to create internal faults. 0.6-3kV, 2.5VA CTs are used for current measurement 4488 case studies were implemented and measurements were recorded by a digital storage oscilloscope with 112 samples per second rate. The proposed approach uses normalized data, so we can compare the simulated and experimental results without concerning about the difference between rated values of simulated and experimental prototype.

Table VI shows the performance of the proposed method using experimental results. As we can see, the proposed method performs well using experimental data. For all disturbances the reliability indices are at least 97.16% which show the applicability of the FGRNN-based protection scheme. The experimental tests represent a great performance in skipping the external faults which shows the immunity of proposed algorithm against cross-country faults.

IV. COMPARATIVE STUDY

To evaluate the proposed FGRNN method, the results obtained in each case are compared with eight other widely used methods which are briefly described as follows.

- GRU: ReLU activation function, 256 epochs, and 128 hidden states.
- ACNN: ACNN is utilized with 20 feature maps, 128 filter, 256 epochs, and maximum pooling size of three, with sigmoid activation function [17].
- CNN: 20 feature maps, 128 hidden states, 256 epochs, maximum pooling size of 3, and sigmoid activation function.
- LSTM: ReLU activation function, 256 epochs, and 128 hidden states.
- SVM: radial basis function (RBF) kernel and cross validation.
- KNN: 1 nearest neighbor based on Euclidean distance and Baye's discussion rule.
- LVQ: based on Euclidean distance, with 6 input and 6 hidden layers with corresponding rate of 1 and 2 neurons in output layer.
- ANN: 86 neurons in input layer, two parallel hidden layers with 25 and 13 neurons, and 1 neuron in the output layer.
- Restricted second harmonic: based on [32].
- Gradient-vector method: based on [33]

For the sake of the comparison, and also to show the robustness of the proposed method in noisy conditions, three different types of noise are added to the raw data. Then, the proposed method and other methods were tested on the noisy dataset. As we mentioned before, a large share of noises in the power system do not obey Gaussian probability distribution. Therefore, three types of noises are considered in this section: i)Gaussian noise with SNR =20, ii) Gaussian mixture noise consists of three Gaussian noises, each of them with SNR = 30, and iii) Laplace probability distribution noise follows chi-square distribution with degree of freedom 4. As an example, Fig. 4 shows the raw signal, three types of noises, and the noisy signals.

The difference between the performances of shallow architecture (SVM, KNN, LVQ, and ANN) methods and deep ones (GRU, ACNN, CNN, and LSTM) can easily be noticed in Table VI where the performances of the five deep architecture methods are better. Among the shallow architecture methods,

TABLE VII
COMPARISON BETWEEN RELIABILITY INDICES OF FGRNN AND SHALLOW AND DEEP STRUCTURE CLASSIFIERS

	Case 2				Case 3				Case 4				Case 5			
	ACC %	DEP %	SAF %	SEC %	ACC %	DEP %	SAF %	SEC %	ACC %	DEP %	SAF %	SEC %	ACC %	DEP %	SAF %	SEC %
The Proposed	**99.35**	**99.72**	**98.84**	**99.61**	**98.52**	**98.46**	**98.57**	**98.57**	**98.59**	**98.31**	**98.43**	**98.85**	**99.26**	**99.23**	**99.28**	**99.28**
GRU	92.56	93.33	91.47	90.77	95.33	94.92	95.29	95.70	95.40	95.08	95.43	95.70	96.15	96.00	96.28	96.28
ACNN	89.71	98.84	89.53	86.19	94.81	94.47	94.99	95.13	94.32	93.91	94.43	94.70	94.81	94.62	94.99	94.99
CNN	91.59	93.06	89.53	90.23	95.55	95.38	95.70	95.70	95.54	95.08	95.56	95.96	95.85	95.88	96.25	95.70
LSTM	90.61	91.67	89.15	88.46	94.51	95.08	95.53	93.98	94.29	94.62	94.93	93.98	95.26	95.38	95.68	95.14
SVM	82.04	85.56	77.13	79.28	88.98	85.51	89.05	89.43	87.39	85.39	86.57	89.29	89.34	89.40	89.80	89.29
KNN	80.68	84.08	75.97	77.47	87.61	86.02	87.11	89.11	87.82	86.96	87.31	88.68	88.17	88.62	87.56	88.68
LVQ	79.77	83.06	75.19	76.08	87.90	86.46	88.20	89.16	87.09	85.85	87.01	88.25	88.58	87.38	88.42	89.68
ANN	75.86	74.44	65.12	64.62	76.39	75.38	77.62	77.30	79.37	78.28	79.91	80.38	78.93	78.62	79.91	79.23
Restricted second harmonic	80.10	85.00	73.51	78.49	88.91	88.46	89.32	89.32	53.00	47.21	54.39	58.36	89.31	88.77	89.56	89.81
Gradient vector	81.33	84.17	77.34	77.65	90.24	90.15	90.83	90.31	77.12	75.85	78.16	78.27	87.19	81.61	82.61	93.13

TABLE VIII
COMPARISON OF AVERAGE PERFORMANCE TIME OF THE DEEP AND SHALLOW BASED METHODS

Methods	**FGRNN**	**GRU**	**ACNN**	**CNN**	**LSTM**	**SVM**	**KNN**	**LVQ**	**ANN**
Test time (ms)	6.42	11.2	4.12	13.56	14.48	2.46	2.42	2.56	3.86

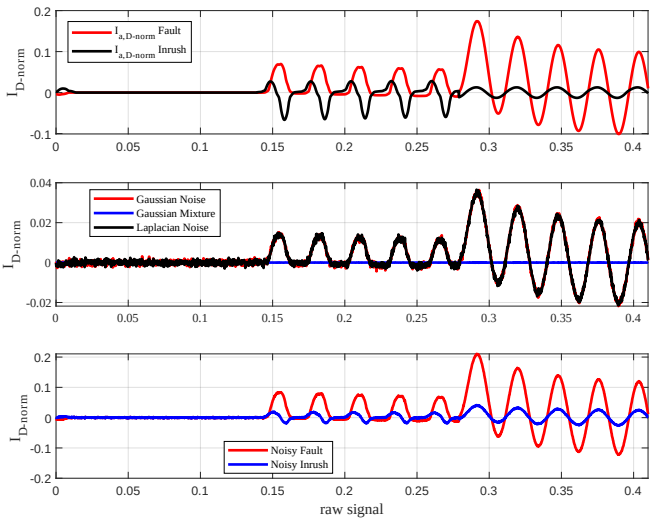


Fig. 4. Raw signal, three type of noise, and noisy signals

ANN has the least reliable performance, while KNN, LVQ, and SVM provides better performances. According to the numbers presented in Table VII, the proposed FGRNN is significantly more reliable than a restricted second harmonic and gradient vector. In particular, in case 4 where the signals are distorted by SFCL as an external factor, the restricted second harmonic method experiences a severe decrease in reliability. In this case, the reliability indexes are between 47.21% and 58.31%. in presence of non-Gaussian/Gaussian noises. It is important to notice that the restricted second harmonic method is used in most of the real differential relays in power transformers. The proposed robust FGRNN has a significantly better performance than the GRU in all cases. Furthermore, the noise impact on other methods is significant in terms of accuracy and reliability, the obtained metrics proposed method is at least 98.31%, while the obtained metrics by the GRU, ACNN, CNN, and LSTM are at least 92.56%, 89.71%, 89.53%, and 88.46%, respectively. It is also can be noted that the performance of the ACNN is highly influenced by the noises, in particular, non-Gaussian noises, which is inferior to other deep networks in terms of accuracy and reliability.

A. *Computation Time*

Table VIII compares the average performance time of all data-driven methods with deep and shallow structures. As can be seen in Table VIII, the methods with shallow architecture provide faster performance than the proposed method. However, the proposed FGRNN method is more reliable and performs significantly better in terms of accuracy, which verifies that its computational complexity is perfectly fit for real-time application. The ACNN method is slightly faster than the proposed FGRNN; however, the FGRNN performs much more robustly in noisy conditions, while ACNN is sensitive to noise.

Considering some conditions can help to verify the generality of the proposed. For instance, power electronic devices by frequently changing of switching angles can strength the amplitude of harmonics. Thus, considering different switching angles in data generation shows that the proposed differential protection scheme can also perform properly in presence of power electronic infrastructure.

B. *Role of Hyper-Parameters*

To investigate the hyperparameter setting on the performance of the designed network, four different conditions have been considered: i) the proposed network with 3 dense layers, where a dense layer is added between the first and second dense layer in the proposed network and has the same parameters as the first dense layer (C1), ii) the proposed

TABLE IX
PERFORMANCE OF PROPOSED ROBUST FGRNN METHOD WITH DIFFERENT STRUCTURE BASED ON EXPERIMENTAL CASE (CASE5)

Cases	ACC%	DEP%	SEC%	SAF%
The Proposed	99.26	99.23	99.28	99.28
C1	98.52	98.46	98.58	98.58
C2	97.92	97.85	97.99	97.99
C3	97.16	97.08	97.24	97.24
C4	96.84	96.77	96.97	96.97

TABLE X
PERFORMANCE OF PROPOSED ROBUST FGRNN METHOD WITH DIFFERENT LOSS FUNCTIONS IN THE EXPERIMENTAL CASE (CASE5)

Cases	ACC%	DEP%	SEC%	SAF%
The Proposed	99.26	99.23	99.28	99.28
Robust loss function from [34]	97.40	96.62	98.14	96.89
Cross-entropy loss function	95.77	97.69	97.85	97.98

TABLE XI
PERFORMANCE OF PROPOSED METHOD WITH RESET GATE AND WITHOUT RESET GATE

Cases	ACC%	DEP%	SEC%	SAF%
Without Reset Gate	99.26	99.23	99.28	99.28
With Reset Gate	97.77	96.84	97.45	97.45

network with 2 FGRNN layers, where the second FGRNN has the same parameters as the first FGRNN layer (C2), iii) the proposed network without dropout technique (C3), and iv) the proposed network by using sigmoid activation function instead of ReLU activation function (C4). The results are discussed considering data generated in case 5. The results are given in Table IX. As can be seen, the proposed method is slightly better in comparison with other conditions (C1-4).

C. Role of the Proposed Loss Function

In order to show the role of the proposed loss function, the proposed method is compared with two different conditions, i) cross-entropy loss function, and ii) the robust loss function presented in [34], where they are formulated for the noises with Gaussian distribution function. The comparison in Table X verify the role of the proposed loss function in robustness against noise and learning ability improvement of the deigned robust deep network. The major reason for the superiority of the proposed method is its robustness in noisy conditions without any assumption on noise model and improving the training process based on the formulated informative loss function.

D. Role of removing reset gate

To address the role of the removing the reset gate, two comparison results should be shown: 1) with reset gate; 2) without reset gate. In both conditions, the proposed loss function is integrated with the deep networks. The comparison in Table XI demonstrates the superiority of the proposed network FGRNN with GRU network with same loss function, batch normalization, and activation function.

V. CONCLUSION AND FUTURE WORK

Conventional machine learning methods have deficiencies such as computational burden, sensitivity to noises, and dependency on a model. These deficiencies decrease the reliability of these methods in the differential protection scheme of power transformers to discriminate between inrush current and internal faults. Using the GRU method which is a deep architecture method is a proper solution to overcome these deficiencies. Furthermore, we proposed a robust and fast GRU to speed up the performance as well as improve the accuracy and robustness against non-Gaussian/Gaussian noises of the differential protection scheme. To this end, the update gate in the conventional GRU has been removed to enhance the computational efficiency and capture sudden changes in transient disturbances such as inrush current and internal fault in the power transformers. Besides, an informatics theory-based loss function is developed in this paper to handle non-Gaussian/Gaussian noises. The proposed method was applied to a real system in both simulated and experimental cases. External factors i.e. series capacitor compensation, SFCL, and CT saturation were taken into account to evaluate the proposed method as a practical solution. The results of the robust FGRNN method shows that the proposed method has an average computational time less than 6.5 ms. The results of the robust FGRNN method were compared to 10 other methods. The comparisons proved that the FGRNN method improved the diagnosis accuracy by at least 3.34% compared with the deep networks and 10.72% compared with shallow networks. Thus, the comparisons proved that the FGRNN method makes a differential protection scheme more reliable and faster.

The investigations on the proposed FGRNN-based differential protection reveal that it would be worthwhile to evaluate the FGRNN-based differential protection method using real time digital simulator (RTDS) or embed the proposed scheme in a hardware for a real power transformers. We can also involve different factors and events such as sympathetic and inrush condition to generalize it to real systems.

REFERENCES

[1] D. Guillen, J. Olivares-Galvan, R. Escarela-Perez, D. Granados-Lieberman, and E. Barocio, "Diagnosis of interturn faults of single-distribution transformers under controlled conditions during energization," *Measurement*, vol. 141, pp. 24 – 36, 2019.

[2] H. Esponda, E. Vázquez, M. A. Andrade, and B. K. Johnson, "A setting-free differential protection for power transformers based on second central moment," *IEEE Transactions on Power Delivery*, vol. 34, no. 2, pp. 750–759, 2019.

[3] G. Ziegler, *Numerical differential protection: principles and applications*. John Wiley & Sons, 2012.

[4] F. Naseri, H. Samet, T. Ghanbari, and E. Farjah, "Power transformer differential protection based on least squares algorithm with extended kernel," *IET Science, Measurement & Technology*, vol. 13, pp. 1102–1110(8), 2019.

[5] F. Naseri, Z. Kazemi, M. M. Arefi, and E. Farjah, "Fast Discrimination of Transformer Magnetizing Current From Internal Faults: An Extended Kalman Filter-Based Approach," *IEEE Transactions on Power Delivery*, vol. 33, no. 1, pp. 110–118, 2018.

[6] L. G. Perez, A. J. Flechsig, J. L. Meador, and Z. Obradovic, "Training an artificial neural network to discriminate between magnetizing inrush and internal faults," *IEEE Transactions on Power Delivery*, vol. 9, no. 1, pp. 434–441, 1994.

[7] L. Tightiz, M. A. Nasab, H. Yang, and A. Addeh, "An intelligent system based on optimized anfis and association rules for power transformer fault diagnosis," *ISA Transactions*, vol. 103, pp. 63 – 74, 2020.

[8] D. Patel, N. G. Chothani, K. D. Mistry, and M. Raichura, "Design and development of fault classification algorithm based on relevance vector machine for power transformer," *IET Electric Power Applications*, vol. 12, no. 4, pp. 557–565, 2018.

[9] A. M. Shah and B. R. Bhalja, "Discrimination Between Internal Faults and Other Disturbances in Transformer Using the Support Vector Machine-Based Protection Scheme," *IEEE Transactions on Power Delivery*, vol. 28, no. 3, pp. 1508–1515, 2013.

[10] G. Mokryani, M. R. Haghifam, H. Latafat, P. Aliparast, and A. Abdollahy, "Detection of inrush current based on wavelet transform and LVQ neural network," in *IEEE PES T&D 2010*, 2010, pp. 1–5.

[11] P. B. Thote, M. B. Daigavane, P. M. Daigavane, and S. P. Gawande, "An Intelligent Hybrid Approach Using KNN-GA to Enhance the Performance of Digital Protection Transformer Scheme," *Canadian Journal of Electrical and Computer Engineering*, vol. 40, no. 3, pp. 151–161, 2017.

[12] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Real-time bearing fault diagnosis of induction motors with accelerated deep learning approach," in *2019 10th International Power Electronics, Drive Systems and Technologies Conference (PEDSTC)*, 2019, pp. 155–159.

[13] S. Afrasiabi, M. Afrasiabi, B. Parang, M. Mohammadi, M. M. Arefi, and M. Rastegar, "Wind turbine fault diagnosis with generative-temporal convolutional neural network," in *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I CPS Europe)*, 2019, pp. 1–5.

[14] J. Long, S. Zhang, and C. Li, "Evolving deep echo state networks for intelligent fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4928–4937, 2020.

[15] J. Long, J. Mou, L. Zhang, S. Zhang, and C. Li, "Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots," *Journal of Manufacturing Systems*, vol. In Press, 2020.

[16] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, and M. Khazaei, "Deep-based conditional probability density function forecasting of residential loads," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3646–3657, 2020.

[17] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Integration of accelerated deep neural network into power transformer differential protection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 865–876, 2020.

[18] M. Ravanelli, P. Brakel, M. Omologo, and Y. Bengio, "Light Gated Recurrent Units for Speech Recognition," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 92–102, apr 2018.

[19] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," dec 2014.

[20] S. Afrasiabi, A. Saffarian, and E. Mashhour, "Dynamic state estimation of power systems using intelligent particle filtering based on ant colony optimisation for continuous domains," *IET Generation, Transmission Distribution*, vol. 13, no. 13, pp. 2627–2636, 2019.

[21] A. Sahebi and H. Samet, "Discrimination between internal fault and magnetising inrush currents of power transformers in the presence of a superconducting fault current limiter applied to the neutral point," *IET Science, Measurement & Technology*, vol. 10, no. 5, pp. 537–544, 2016.

[22] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1, 2013, p. 3.

[23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, feb 2015.

[25] Y. Kong, "Dominantly truthful multi-task peer prediction with a constant number of tasks," in *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '20. USA: Society for Industrial and Applied Mathematics, 2020, p. 2398–2411.

[26] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y. J. Shin, "High-Impedance Fault Detection in the Distribution Network Using the Time-Frequency-Based Algorithm," *IEEE Transactions on Power Delivery*, vol. 30, no. 3, pp. 1260–1268, 2015.

[27] "Keras: Deep learning library for theano and tensorflow." [Online]. Available: https://keras.io/

[28] I. Guyon, "A scaling law for the validation-set training-set size ratio," in *AT & T Bell Laboratories*, 1997.

[29] U. Rudez and R. Mihalic, "A reconstruction of the wams-detected transformer sympathetic inrush phenomenon," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 724–732, 2018.

[30] E. Hajipour, M. Vakilian, and M. Sanaye-Pasand, "Current-Transformer Saturation Compensation for Transformer Differential Relays," *IEEE Transactions on Power Delivery*, vol. 30, no. 5, pp. 2293–2302, 2015.

[31] U. D. Annakkage, P. G. McLaren, E. Dirks, R. P. Jayasinghe, and A. D. Parker, "A current transformer model based on the Jiles-Atherton theory of ferromagnetic hysteresis," *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 57–61, 2000.

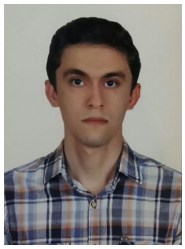
[32] A. Guzman, S. Zocholl, G. Benmouyal, and H. J. Altuve, "A current-based solution for transformer differential protection. ii. relay description and evaluation," *IEEE Transactions on Power Delivery*, vol. 17, no. 4, pp. 886–893, 2002.

[33] R. J. N. Alencar, U. H. Bezerra, and A. M. D. Ferreira, "A method to identify inrush currents in power transformers protection based on the differential current gradient," *Electric Power Systems Research*, vol. 111, pp. 78–84, 2014.

[34] M. Afrasiabi, S. Afrasiabi, M. Mohammadi, and B. Parang, "Fault localisation and diagnosis in transmission networks based on robust deep gabor convolutional neural network and pmu measurements," *IET Generation, Transmission & Distribution*, vol. In Press, 2021.

Shahabodin Afrasiabi received his B.Sc. degree from Semenan University, Semnan, Iran, in 2014, and his M.Sc. degree from the Shahid Chamran University, Ahvaz, Iran, in 2017. He is currently working at smart grid laboratory, Shiraz University. His research interests include power system dynamic analysis, machine learning, fault diagnosis,energy managment, electric machine design, state estimation, time series forecasting and power system probabilistic analysis.

Mousa Afrasiabi received his B.Sc. degree from University of Guilan, Rasht, Iran, in 2008, and his M.Sc. degree from the K.N. Toosi University of Technology, Tehran, Iran, in 2011, and his Ph.D. from the Shiraz University, Shiraz, Iran. all in electrical power engineering. His research interests include energy managment, time series forecasting, machine learning, power system dynamic analysis, electric machine design, and power system probabilistic analysis.



Benyamin Parang received the B.Sc. degree from Yazd University, Yazd, Iran, in 2015, and his M.Sc. degree from the Shiraz University , Shiraz, Iran, in 2018. He is currently working in smart grid laboratory, Shiraz University. His research interests include power system dynamic analysis, machine learning, fault diagnosis,energy managment, and power system probabilistic analysis.

Mohammad Mohammadi received his B.Sc. degree from Shiraz University, Shiraz, Iran, in 2000, and the M.Sc. and Ph.D. degrees from the Amirkabir University of Technology, Tehran, Iran, in 2002 and 2008, respectively. He is currently a Professor at the Department of Power and Control Engineering, Shiraz University. His research interests include power system probabilistic analysis, power system security assessment, machine learning, and power system dynamic analysis.

Haidar Samet (M'15) received the Ph.D. degree in electrical engineering from Isfahan University of Technology. Currently, he is a professor in Shiraz University and a postdoctoral researcher in Eindhoven University of Technology. His main research interest is application of DSP techniques in power systems.

Tomislav Dragicevic (S'09-M'13-SM'17) received the M.Sc. and the industrial Ph.D. degrees in Electrical Engineering from the Faculty of Electrical Engineering, University of Zagreb, Croatia, in 2009 and 2013, respectively. From 2013 until 2016 he has been a Postdoctoral researcher at Aalborg University, Denmark. From 2016 until 2020 he was an Associate Professor at Aalborg University, Denmark. Currently, he is a Professor at the Technical University of Denmark.

He made a guest professor stay at Nottingham University, UK during spring/summer of 2018. His research interest is application of advanced control, optimization and artificial intelligence inspired techniques to provide innovative and effective solutions to emerging challenges in design, control and diagnostics of power electronics intensive electrical distributions systems and microgrids. He has authored and co-authored more than 330 technical publications (more than 150 of them are published in international journals, mostly in IEEE), 10 book chapters and a book in this field, as well as filed for several patents.

He serves as an Associate Editor in the *IEEE Transactions on Industrial Electronics*, *IEEE Transactions on Power Electronics*, *IEEE Emerging and Selected Topics in Power Electronics* and *IEEE Industrial Electronics Magazine*. Prof. Dragicevic is a recipient of the Koncar prize for the best industrial PhD thesis in Croatia, a Robert Mayer Energy Conservation award, and he is a winner of an Alexander von Humboldt fellowship for experienced researchers.