Fault Location in Transmission Lines based on LSTM Model

L. A. Ensina[1,2], P. E. M. Karvat[1], E. C. de Almeida[1], L. E. S. de Oliveira[1]

[1] Federal University of Paraná, Brazil
patrickkarvat.pk@ufpr.br, eduardo.almeida@ufpr.br, luiz.oliveira@ufpr.br
[2] Federal University of Technology - Paraná, Brazil
leandroa@utfpr.edu.br

Abstract. Transmission lines are fundamental components of the electric power system, demanding special attention from the protection system due to the vulnerability of these lines. This article presents a method for fault location in transmission lines using data for a single terminal without requiring explicit feature engineering by a domain expert. The fault location task provides an approximate position of the point of the line where the failure occurred, serving as information to the operators to dispatch a maintenance staff to this location to reclose the transmission line with better reliability and safety. In our method, we extract two post-fault cycles of the three-phase current and voltage signals to serve as input to a model based on the LSTM algorithm. We defined the model's architecture with empirical experiments searching for the best structure to estimate the fault distance. For this purpose, we used a dataset with diversified failure events, also available to the scientific community. The results demonstrate the effectiveness of the proposed method with a mean error of 0.1309 km ± 0.4897 km, representing 0.0316% ± 0.1183% of the transmission line extension.

CCS Concepts: • **Information systems** → **Data mining**; • **Computing methodologies** → **Machine learning**; • **Applied computing** → **Engineering**.

Keywords: artificial intelligence, deep learning, fault diagnosis, recurrent neural network

1. INTRODUCTION

The electrical power system is a complex structure that encloses electricity generation, transmission, and distribution. In particular, transmission lines are fundamental elements of this structure, linking the electricity power production to the customers, such as industries, businesses, and residences. Consequently, these elements demand special attention since they are exposed to several adverse situations that can negatively affect their normal operating state, such as storms and vegetation contact with the lines [Singh and Vishwakarma 2015].

These disturbances in a transmission line result in faults that can interrupt the electricity supply temporarily or permanently. A fault can be defined as an abnormal condition in the components of a power system, such as an increase in the current flow to one or more phases [Yadav and Dash 2014]. Fig. 1 shows an example of the effects of a fault on the current waveform in a transmission line. In this scenario, a digital fault recorder of the protection system, located in the electrical substations, samples data composed of voltage and current signals over time, containing relevant information for failure diagnosis, such as fault location. In particular, fault location is a crucial task that enables the operators to dispatch a maintenance staff to the location of the fault occurrence.

Several approaches are proposed in the specialized literature for fault analysis in transmission lines, especially based on three methods: traveling waves, impedance-based, and Artificial Intelligence (AI). The traveling wave technique is complex, requiring a high sampling rate and high computational cost

This study was financed in part by the Energy Company of Paraná (Copel) within the Research and Development Program of the Brazilian Electricity Regulatory Agency (ANEEL) – project number PD-06491-0407/2015.

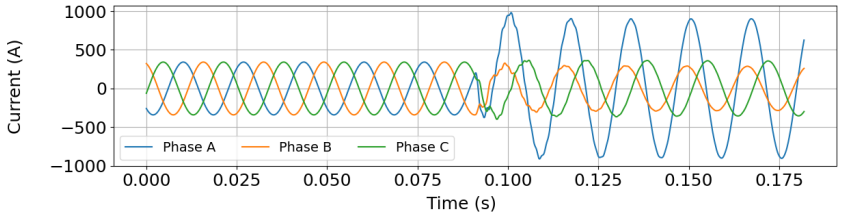

Fig. 1. Waveform of a faulty transmission line: amplitude of the phases changes after the fault occurred in the transmission line close to 0.1 seconds.

[Dong et al. 2009], while the impedance-based approach can be affected by the variation of the fault parameters, mainly for faults with high resistances [Tang et al. 2000; Furse et al. 2021].

Nowadays, methods based on AI have gained more attention, even though they require a significant amount of data for model training. The increase in its choice is mainly due to its high performance and adaptability to different fault conditions and parameters [Chen et al. 2016; Raza et al. 2020].

In this context, deep learning (DL) algorithms back promising methods for fault distance estimation [Raza et al. 2020; Mishra et al. 2020]. The main advantage is reducing the need for feature engineering, which requires specialized knowledge from domain experts for feature extraction. Although DL algorithms require a large amount of data, the dataset described in this article with diversified failure events solves this limitation for fault diagnosis. Also, the Recurrent Neural Network (RNN), a particular type of DL, can efficiently represent data with complex sequential dependencies among the attributes, such as time series of current and voltage signals. One of the most successful variants of RNN is the Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber 1997], achieving good performance in several domains.

This work presents a method for fault location in transmission lines, using only two post-fault cycles for the three-phase current and voltage signals for a single end (terminal) of the line. These segmented waveforms, standardized by the Z-Score technique, are used as input for an LSTM-based model to locate the fault in the transmission line.

2. RELATED WORK

Different methods are found in the specialized literature for fault location based on AI, especially using DL algorithms [Kanagasabapathy 2021]. In the remainder of this section, we describe some works based on DL algorithms for fault location. In turn, the main advantages of our method compared to these works are described in Section 4.

[Chen et al. 2018] proposed a framework with integrated feature extraction based on the Summation-Wavelet Extreme Learning Machine (SW-ELM), as well as an extension of the SW-ELM algorithm called Summation-Gaussian ELM (SG-ELM). The algorithms use as input a single cycle of the three-phase current signal differences measured at both terminals connected to the power line. The average errors reported were around 2.90 km for the SW-ELM and 2.77 km for the SG-ELM, equivalent to 2.90% and 2.77%, respectively, for the analyzed transmission line with 100 km of extension.

[Fan et al. 2019] presented a mixed Convolutional Neural Network (CNN) with LSTM structure to predict the fault distance. The input data contain six channels (three-phase voltages and currents), where each channel consists of around six cycles. The absolute error was 0.184 km $\pm$ 0.146 km, representing about 0.092% $\pm$ 0.073% for a 200 km transmission line length.

[Zhang et al. 2020] proposed a method for fault location using the current waveform data for 0.1 seconds before and 0.1 seconds after the fault inception for both terminals of the transmission line, representing about 12 cycles in total (i.e., six pre-fault cycles and six post-fault cycles). These data

were used as input to the Bidirectional Gated Recurrent Unit (Bi-GRU) algorithm, indicating the percentage of the line where the fault occurred. The average error was 0.087%. However, the authors do not specify the length of the transmission lines evaluated, and neither the details about the fault data used to assess their approach.

[Belagoune et al. 2021] presented an approach using current and voltage signals for the three phases from both terminals of the transmission line. The authors employed all data from each simulation without segmenting into pre or post-fault cycles. The distance to a fault is estimated by a model based on the LSTM algorithm, determining an average error of around 0.071 km ± 0.065 km, which is equivalent to 0.213% ± 0.196% for the transmission line with 300 km used by the authors.

3. THE METHOD FOR FAULT LOCATION

3.1 Fault Analysis Database

The Fault Analysis Database (FADb) is a public dataset with several fault simulations developed by our research group. These events were based on the IEEE 9-bus power system provided by the ATPDraw tool [Høidalen et al. 2019], which reproduces a production power system. The modeled system and all simulation data are available for the scientific community[1].

The examined transmission line consists of the following properties: 500 kV, 414 km, and 60 Hz. This specification represents the longest transmission line in the Copel network, a Brazilian electric utility company. The data for each simulation comprises voltage and current signals for each of the three phases at both ends of the line for a sampling rate of 10 kHz, representing about 167 samples per cycle (10 kHz divided by 60 Hz). The oscillography of each simulation starts without the presence of a failure, which occurs in distinct instants inserted into the same cycle representing the uncertainty related to the time when the fault begins. The fault parameters used in the simulations are as follows:

—Type: AG, BG, CG, AB, AC, BC, ABG, ACG, BCG, ABC;
—Location: 1 to 100% of line extension, with intervals of 1%;
—Resistance: 0.01 to 200 Ω, with intervals of 10 Ω;
—Inception time, in seconds (s): 0.091 s, 0.093 s, 0.095 s, 0.097 s, 0.099 s, 0.101 s, 0.103 s, 0.105 s.

In particular, the letters A, B, and C represent each of the three phases of the transmission line, while the letter G corresponds to the ground action in a fault. Thus, the initials AG indicates a failure involving phase A and the ground, as well as AB represents a fault between phases A and B without the action of the ground.

The repository contains 168,000 fault events, combining the previously mentioned parameters. In addition, each of these parameter combinations results in different fault signatures. The archives available in the repository, i.e., simulation data and the modeled system, ensure reproducibility of the results and the generation of new fault events considering other values of the parameters, e.g., failure resistances with values higher than 200 Ω. Also, this dataset can be used for different purposes in machine learning-based applications, such as anomaly detection in time series (fault detection), classification (fault type classification), and regression tasks (fault location).

3.2 Input Transformation

Input transformation corresponds to the conversion of the oscillography data (voltage and current waveforms for the three phases) into the input space for fault location. Therefore, there are two steps in our method for this purpose, which are described in the remainder of this section.

[1] https://1drv.ms/u/s!ArMEeMx4MYDNimHVxiDx3b4CI3iL?e=8GfXg7

The former step in our method is to extract two post-fault cycles of the current and voltage signals individually, as represented in Fig. 2. We used only the data of a single terminal, avoiding the need for data synchronization between the transmission line terminals.

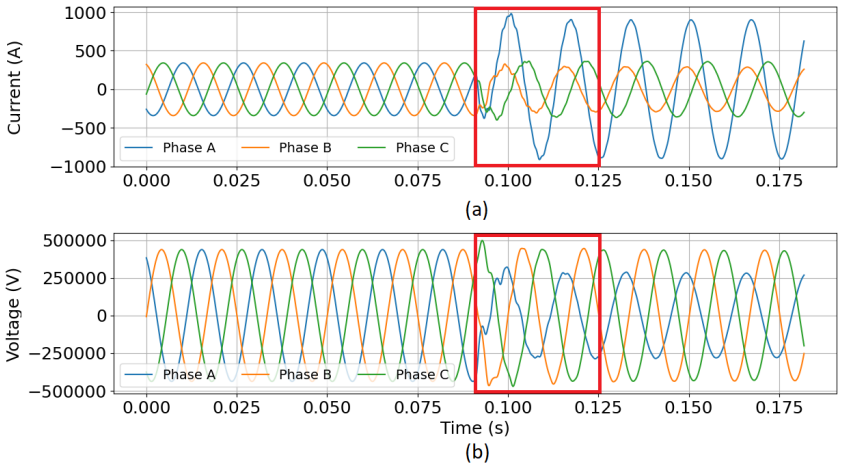

Fig. 2. Example of two post-fault cycles demarcated in a fault simulation of AG type for the (a) current and (b) voltage waveforms.

In the latter step, the data were standardized by the Z-Score technique, as defined in Equation 1.

$$Z_i = \frac{X_i - \bar{X}}{S} \tag{1}$$

where Z_i represents the standardized value of X_i in the instant i, while $\bar{X}$ and S represent the mean and the standard deviation, respectively, for a set of values X.

Thus, the input space consists of a matrix of 6×334, where each row represents a single signal type (current followed by voltage) for a specific phase (A, B, and C, in this respective order), and each column is a data sample concerning the two post-fault cycles (334 samples) after the Z-Score standardization. To make it clear, the order of the rows is as follows: current A, current B, current C, voltage A, voltage B, and voltage C.

3.3 Experimental Setup

The LSTM algorithm can handle complex temporal relationships with sequential dependencies between attributes [Hochreiter and Schmidhuber 1997], such as between current and voltage waveform data samples. In our method, an LSTM-based model performed the fault location task. Table I presents the architecture of our model.

Table I. Proposed model for fault location in transmission lines.

Layer	**Units**	**Return sequences**	**Input shape**
LSTM	1000	True	$(6, 334)$
LSTM	1000	True	-
LSTM	1000	False	-
Dense	1	-	-

To train the proposed model, we used the Adam optimizer, the Mean Squarer Error (MSE) metric, a learning rate of 0.001, and a batch size of 1680. The output of these models consists of a real value between [0, 1], corresponding to the percentage of the transmission line length where the fault was located. This value is converted to distance in kilometers (km) after the prediction by Equation 2.

$$distance_{km} = TL_{length} * distance_{pred} \tag{2}$$

where $distance_{km}$ is the fault location in km, TL_{length} is the transmission line length in km (414 km in FADb dataset - Section 3.1), and $distance_{pred}$ is the output of the model.

In our experiments, we used the FADb dataset, as described in Section 3.1, randomly divided into training, validation, and testing sets. The training set comprises 60% of all examples (100,800 simulations), while validation and testing sets contain 20% of all cases each (33,600 simulations). Each set was standardized separately, avoiding data leakage from the validation and testing sets to the training step [Kaufman et al. 2012]. The benefit of this protocol is that we guarantee that the data used to test the proposed method have never been seen directly or indirectly by the model during the training process.

In this scenario, the usage of the validation set focused on establishing a proper configuration for the algorithm (e.g., number of layers and batch size), while the testing set was used to report the method's performance since the model had no contact with it before. So, the results reported in Section 4 represent MSE (Equation 3), Mean Absolute Error (MAE) (Equation 4), and maximum (max) error considering the testing set with three repetitions for all experiments.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

where n is the number of examples, y_i is the real fault localization, and $\hat{y}_i$ is the predicted fault localization for the instant i.

4. RESULTS AND DISCUSSION

We used voltage and current data in our experiments because both waveforms performed better together than individually. Although the current signal presents the most critical information for short circuit faults [Yadav and Dash 2014], complementing the voltage signal improves the accuracy of the results. Likewise, we applied two cycles because the model with fewer data made the results worse. On the other hand, using more data did not reveal a performance improvement.

Using data from a single terminal aims to avoid data synchronization between the transmission line terminals. The availability of data from both terminals simultaneously is not guaranteed, which can compromise the usability of a method that requires data from both, as in some related works presented in Section 2 [Chen et al. 2018; Zhang et al. 2020; Belagoune et al. 2021]. Although [Fan et al. 2019] also only requires data for a single terminal, our method requires less data than they do as we only use two post-fault cycles. In contrast, this related work needs about six cycles, as reported in Section 2. In a real scenario, only a few post-fault cycles may be available since the protection system acts as quickly as possible to identify a failure onset and isolate it from the rest of the system. On the Copel network, for example, there are fault files that contain only about three post-fault cycles,

which would make methods that require more than three cycles unfeasible (e.g., [Fan et al. 2019] and [Zhang et al. 2020]).

We also evaluated our method using the Gated Recurrent Unit (GRU) algorithm, but the LSTM-based model performed best. We do not present the detailed performances of GRU and only show the overall outcome for the GRU algorithm to demonstrate the superior performance of the LSTM. The results reported in Table II represent the performances of the LSTM model using only the testing set. We divided the results into ten partitions that represent 10% of the line extension each.

Table II. Performance of the proposed model for different segments of the transmission line for the testing set.

Line extension	MAE	Max error
4.14 to 41.4 km	0.2803 km ± 1.4645 km	74.3131 km
41.4 to 82.8 km	0.0942 km ± 0.1212 km	2.8580 km
82.8 to 124.2 km	0.0822 km ± 0.0750 km	1.9251 km
124.2 to 165.6 km	0.0857 km ± 0.0828 km	2.2534 km
165.6 to 207.0 km	0.0831 km ± 0.0746 km	0.8582 km
207.0 to 248.4 km	0.0878 km ± 0.0834 km	1.3314 km
248.4 to 289.8 km	0.1053 km ± 0.1477 km	4.2323 km
289.8 to 331.2 km	0.1273 km ± 0.1318 km	1.5404 km
331.2 to 372.6 km	0.1932 km ± 0.2855 km	5.3079 km
372.6 to 414.0 km	0.1732 km ± 0.3044 km	6.1312 km
TOTAL	0.1309 km ± 0.4897 km	74.3131 km

The experiments revealed that the algorithm demonstrated an overall MSE of 0.2569 and MAE of 0.1309 km ± 0.4897 km, representing 0.0316% ± 0.1183% of the length of the transmission (414 km). Despite this average value showing a low error rate, there is a high maximum error of 74.3131 km, corresponding to 17.95% of the line length. However, the highest errors are concentrated near the terminal where the data (voltage and current signals) were used as input to the algorithm, as shown in Fig. 3. Most methods proposed in the literature, including those mentioned in Section 2, evaluate their approaches generally with fault distances from 5% to 95% of the transmission line length, whereas we use faults of 1% to 99% of the line length. In other words, these works avoid evaluating faults that occur very close to the transmission line terminals, where the highest errors can occur.

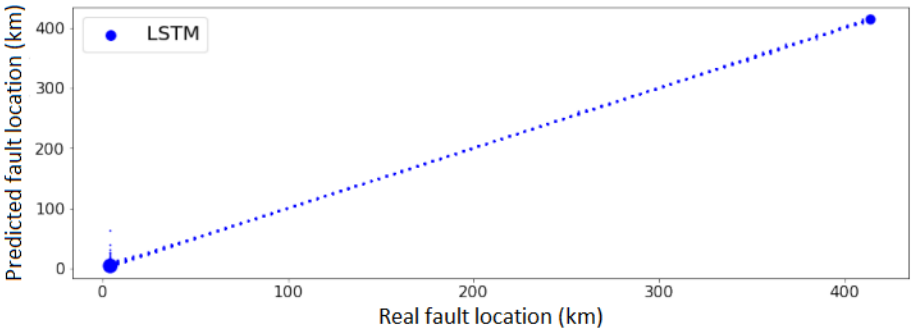

Fig. 3. Representation of the proposed method prediction errors.

Errors bigger than 10 km represent only 10 out of 33,600 failure events in the testing set, which only corresponds to about 0.03% of all test examples, despite the high maximum error value for the LSTM-based model. Fig. 4 shows the first experimental repetition (1 of 3 repetitions). The other two repetitions revealed similar distributions of forecast errors. On the other hand, most errors are less than or equal to 0.5 km and correspond to 97.71% of all test examples. Furthermore, this analysis can also justify the high variability of errors with a coefficient of variation of about 374%.

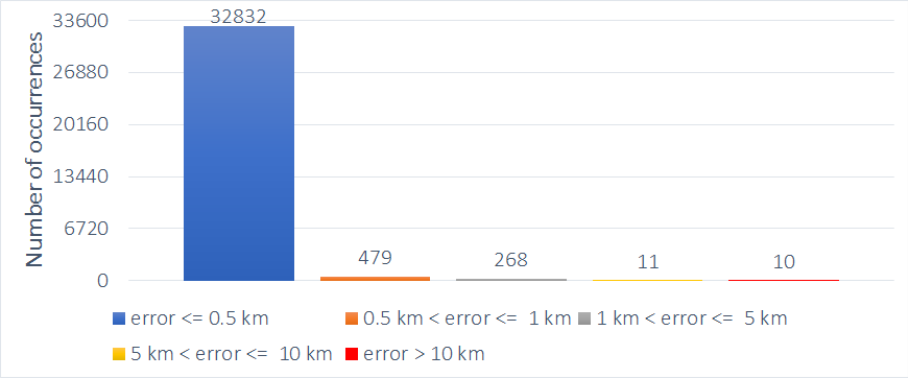

Fig. 4. Histogram of errors for the first experimental repetition. The rest of the repetitions presented similar error distributions.

In turn, the general MAE for GRU corresponds to 0.5 km $\pm$ 1.187 km, with a maximum error of 172.2453 km, demonstrating inferior performance compared to LSTM, as mentioned above, and concentrating even more significant errors close to the terminal versus LSTM.

We intend to continue investigating the improvement of our method concerning faults near the transmission line terminals. A promising prospect is related to the combination of several models (e.g., distinct DL-based models), which are experts in specific feature spaces [Cruz et al. 2018]. In other words, we may have models that can locate faults precisely to the middle of the line, while other models can locate faults accurately close to the terminals with the lowest error rate.

Thus, we can dynamically select the most competent model(s) for each new test pattern on-the-fly, combining their predictions by mean or median measures, for example. This method is known as Dynamic Regressor Selection when a single model is selected, or Dynamic Ensemble Selection, when a set of models is chosen [Mendes-Moreira et al. 2009]. In this scenario, we can achieve a method with fewer errors and avoid discrepant forecasts, mainly for faults closer to the transmission line terminals. In addition, it should result in less variability of errors in failure predictions.

5. CONCLUSION

This article presented a method for fault location in transmission lines, using samples of voltage and current signals for a single end of the line. We evaluated the proposed method with failures for all points of the transmission line, demonstrating its effectiveness with an MAE of 0.1309 km $\pm$ 0.4897 km (0.0316% $\pm$ 0.1183%). So, the results show that the proposed method complies with our objective. It is also important to mention that we used only one dataset because there are no other datasets available in the literature for fault location task.

The main advantages of our method correspond to the use of data from a single terminal of the transmission line without requiring explicit feature engineering by a domain expert. So, our approach does not demand data synchronization between the terminals, requiring only two post-fault cycles of the voltage and current signals from a single end of the line. Despite the accurate results, our method revealed high prediction errors for failures close to the terminals.

Future works include (1) the refinement of the method to accurately locate faults that are close to the terminals, such as combining multiple models; (2) the evaluation of the method using lower sampling rates than 10 kHz; (3) the evaluation with real fault events; and (4) conduct additional experiments to compare our method with other state-of-the-art methods.

REFERENCES

Belagoune, S., Bali, N., Bakdi, A., Baadji, B., and Atif, K. Deep learning through lstm classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. *Measurement* vol. 177, pp. 109330, 2021.

Chen, K., Huang, C., and He, J. Fault detection, classification and location for transmission lines and distribution systems: a review on the methods. *High Voltage* 1 (1): 25–33, 2016.

Chen, Y. Q., Fink, O., and Sansavini, G. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction. *IEEE Transactions on Industrial Electronics* 65 (1): 561–569, 2018.

Cruz, R. M., Sabourin, R., and Cavalcanti, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion* vol. 41, pp. 195–216, 2018.

Dong, X., Kong, W., and Cui, T. Fault classification and faulted-phase selection based on the initial current traveling wave. *IEEE Transactions on Power Delivery* 24 (2): 552–559, 2009.

Fan, R., Yin, T., Huang, R., Lian, J., and Wang, S. Transmission line fault location using deep learning techniques. In *2019 North American Power Symposium (NAPS)*. IEEE, Wichita, KS, USA, pp. 1–5, 2019.

Furse, C. M., Kafal, M., Razzaghi, R., and Shin, Y.-J. Fault diagnosis for electrical systems and power networks: A review. *IEEE Sensors Journal* 21 (2): 888–906, 2021.

Høidalen, H. K., Prikler, L., and Peñaloza, F. *ATPDraw version 7.0 for Windows - Users' Manual*. ATPDraw, 2019.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural Computation* 9 (8): 1735–1780, 1997.

Kanagasabapathy, O. Fault location in transmission line through deep learning—a systematic review. In *Inventive Systems and Control*, V. Suma, J. I.-Z. Chen, Z. Baig, and H. Wang (Eds.). Springer Singapore, Singapore, pp. 223–238, 2021.

Kaufman, S., Rosset, S., Perlich, C., and Stitelman, O. Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data* 6 (4): 1–21, 2012.

Mendes-Moreira, J., Jorge, A. M., Soares, C., and de Sousa, J. F. Ensemble learning: A study on different variants of the dynamic selection approach. In *Machine Learning and Data Mining in Pattern Recognition*, P. Perner (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 191–205, 2009.

Mishra, M., Nayak, J., Naik, B., and Abraham, A. Deep learning in electrical utility industry: A comprehensive review of a decade of research. *Engineering Applications of Artificial Intelligence* vol. 96, pp. 104000, 2020.

Raza, A., Benrabah, A., Alquthami, T., and Akmal, M. A review of fault diagnosing methods in power transmission systems. *Applied Sciences* 10 (4): 1–27, 2020.

Singh, S. and Vishwakarma, D. N. Intelligent techniques for fault diagnosis in transmission lines — an overview. In *International Conference on Recent Developments in Control, Automation and Power Engineering (RDCAPE)*. IEEE, Noida, India, pp. 280–285, 2015.

Tang, Y., Wang, H.-f., Aggarwal, R. K., and Johns, A. T. Fault indicators in transmission and distribution systems. In *International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*. IEEE, London, UK, pp. 238–243, 2000.

Yadav, A. and Dash, Y. An overview of transmission line protection by artificial neural network: Fault detection, fault classification, fault location, and fault direction discrimination. *Advances in Artificial Neural Systems* vol. 2014, pp. 1–20, 2014.

Zhang, F., Liu, Q., Liu, Y., Tong, N., Chen, S., and Zhang, C. Novel fault location method for power systems based on attention mechanism and double structure gru neural network. *IEEE Access* vol. 8, pp. 75237–75248, 2020.