International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

MACHINE LEARNING BASED FAULT LEVEL ANALYSIS FOR DISTRIBUTION SUBSTATIONS

M. Vishal[1], S. K. Vetrivel[2], M. S. Sivagama Sundari[3*]

[1]UG Student, Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

[2]UG Student, Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

[3*]Assistant Professor, Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India, ms_sivagamasundari@cb.amrita.edu

Corresponding Author: M. S. Sivagama Sundari

Abstract

Fault level analysis in the distribution substations is an important and key part of the power system's protection and reliability. Disturbances like three-phase, line-to-line, and line-to-earth faults can drastically affect system performance, damage the equipment, and cause extended interruptions. An accurate fault identification is needed to select appropriate protection devices and ensure prompt fault fixing to minimize downtime. Traditional analysis methods heavily rely on electrical parameters, which may not be available in many functioning environments. This paper presents a machine-learning-based approach for predicting and categorising the fault types in distribution substation studies using the available parameters in hand. The proposed method helps power companies find faults more easily, make their systems stronger, and take steps to fix problems before they happen.

Keywords: Fault level analysis, Distribution substations, Machine Learning, Fault classification, Power system reliability, Predictive maintenance.

INTRODUCTION

The maintenance of the power distribution network's reliability is essential for ensuring a steady and uninterrupted electricity supply. At present, disturbances in power distribution such as three-phase, line to line, and line-to-earth faults can cause significant amount of damage to the framework, interrupt supply, and reduces the lifespan of the electrical equipment[1,7]. Detecting

Received: August 01, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

these faults early and classifying them accurately allows resources to restore service faster, fine-tune protection schemes and schedule preventive maintenance more efficiently[5,11].

Conventional fault analysis in substations often depends on detailed electrical measurements like current, voltage and impedance[1,8]. Though these parameters are valuable but they are not always accessible. In the olden days, during post fault investigations or in some areas with limited monitoring features, such data of real-time electrical maybe incomplete and unreliable. Such limitations made the first step to the exploration of an alternative, data-centric techniques that interpret existing indicators to assess and predict fault conditions.

This paper presents a machine-learning-based model that predicts fault types in distribution substations without the actual need of direct electrical parameters. Rather, it uses, operational and practical parameters like substation attributes, causes of fault, time taken for restoration and other system level details to train supervised learning models that can understand and determine the nature of a fault.Moreover, by integrating Machine Learning into this fault analysis, utilities can get a clearer view of fault occurrence, predicts the potential risks and also improves for daily operations and emergency responses. The proposed method helps move towards a smarter and more flexible distribution system that can learn from past faults and handle future problems better and more efficient[2].

LITERATURE REVIEW/RELATED WORK

Conventional Fault Analysis

Old methods for fault level analysis use symmetrical component theory, short-circuit current calculations, and relay coordination studies[6]. They need correct readings of voltage, current, and impedance from PMUs or protection relays. While these methods give good results, they also need high-quality real-time data and costly equipment. Their accuracy can drop if there are measurement mistakes, transformer saturation, or delays in wide-area monitoring systems. They also need trained staff to set up and understand the results, which makes the work more complex. On top of that, these methods do not work well when only limited or combined data is available, like in small substations or rural power networks.

Fault Classification in Power Systems

Many research works have used signal processing and statistical methods to identify different fault types. Commonly used methods include wavelet transforms, Fourier analysis, and impedance-based algorithms[8]. These methods take information from current and voltage waveforms and pull out certain features[9]. The features are then compared with pre-set limits to decide which type of fault has happened. In some situations, time–frequency domain analysis is used to detect the short-lived signals that appear right after the fault starts. Other techniques use

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

modal transformation or traveling wave analysis to make fault location more accurate in complicated networks[13]. These approaches can give very good results when the conditions are ideal. However, they often need high sampling rates and synchronized measurements, which may not be possible for all distribution networks[15,16]. Their accuracy can also be affected by noise in the system, non-linear loads, and changes in fault resistance[3]. Because of these factors, such methods may not always work as well in real operating conditions.

Machine Learning Applications in Fault Analysis

In the last few years, people started using machine learning for fault finding and fault type classification [2,10]. Some of the common ones are decision trees, random forests, and SVM. They work well with large data and can notice patterns we might miss. A few papers also tried artificial neural networks [12]. Some even went for deep learning like CNN so the system finds features on its own. But still, most of these studies are on transmission lines. They also need proper voltage and current readings [3, 14]. In many substations, that data is not even there or is not complete, so the same methods cannot be used directly.

Research Gaps

Very few papers talk about using other types of data, like basic substation info or operation details, to find faults in distribution substations. In many places, there is no full set of electrical readings. Sometimes the data is missing, sometimes it is not measured at all. So, there is a need for models that can work with the data that is actually there.

2. Fault Analysis Assumptions

a. The points below are the basic things assumed while doing the fault level analysis

- All generator EMFs are taken as in phase before the fault, which means the system is in synchronism at that time [1].
- For a three-phase(L–L–L) fault, short-circuit current is there in all three lines at the same time.
- For a line-to-line (L–L) fault, the current flows only between the two faulted phases [7].
- In a line-to-earth (L–E) fault, one phase is connected to the ground and this brings in zero-sequence currents [7].
- The number of phases involved in the fault does not change while the fault is there [1].
- Voltages and fault impedances are taken as constant during the fault period [1].

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- Depending on the fault type, positive, negative, and zero sequence impedances are considered [1].
- The inductive reactance of the system is also taken into account in the fault current calculations [4].

b. Factors affecting Line-Line Fault(L-L)

- Insulation between two phases gets damaged.
- Lightning strikes or switching causes sudden high voltage.
- Conductors get physically damaged because of wind, trees falling, or accidents [8].
- Conductors are too close to each other, especially in wet or dirty places.
- Old cables and connectors that have worn out over time.

c. Factors affecting Line-Earth Fault(L-E)

- Insulation breaks down and one phase touches the earth.
- Water gets inside underground cables.
- A conductor falls on the ground or on something connected to the earth.
- The earthing system is bad or the earth connection is rusty.
- Animals touch both the live wire and the earth at the same time[8].

d. Factors affecting Three-Phase Fault(L-L-L)

- Serious mechanical damage happens in equipment like busbars or switches.
- Insulation fails badly in all three phases at once.
- Metal objects accidentally get inside switchgear and cause short circuits.
- Transformers or other equipment break down completely.
- Very strong lightning causes all three phases to flashover together.

e. Block Diagram

Received: August 01, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

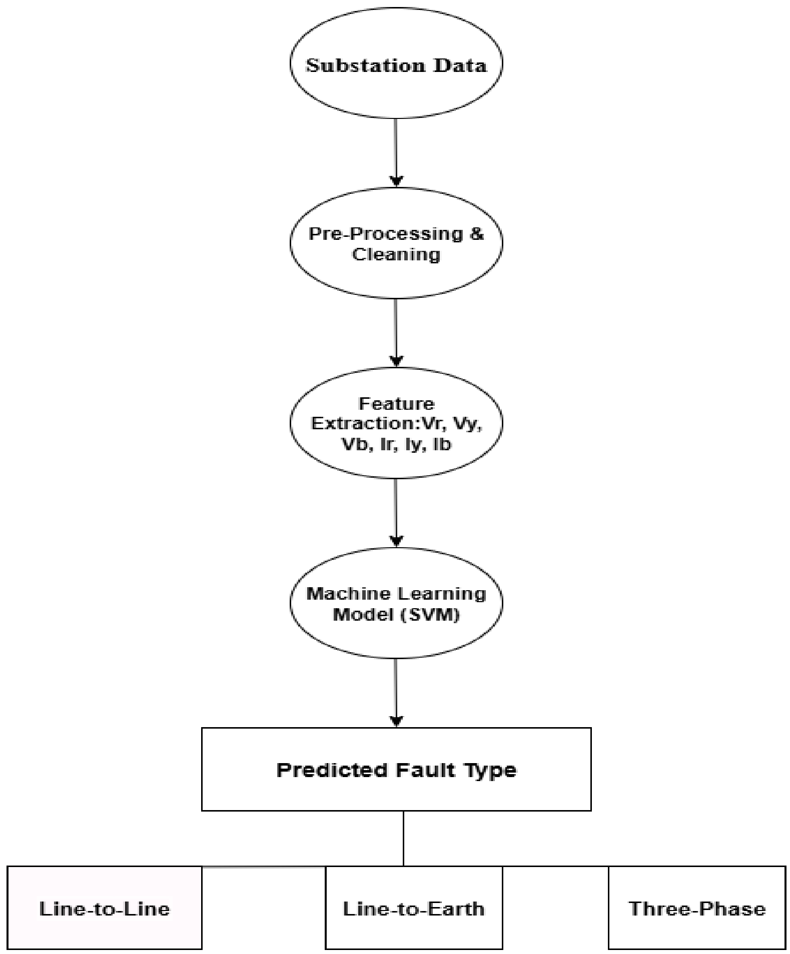

Fig 1. Schematic representation of the fault analysis using machine learning

DATASET DESCRIPTION

The dataset has values for three-phase voltages **Vr, Vy, Vb** and three-phase currents **Ir, Iy, Ib.** These were taken from a substation when faults happened. Along with these readings, types of faults are given - line-to-line fault, line-to-earth fault, and three-phase fault.The voltages and currents are the input features. The fault type is the output that we want to predict. The data is split into two parts. One part is for training the model and the other is for testing it.

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

METHODOLOGY

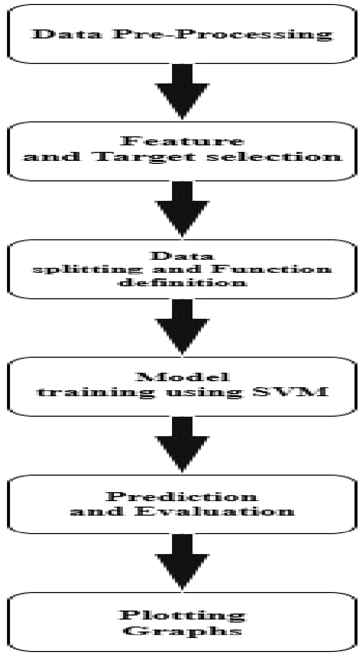

Fig 2. Representation of code analysis

f. Data Pre-Processing

In data pre-processing, the dataset is imported into the model and prepared for analysis. The missing values, noise, and data normalization is performed here.

g. Feature and Target selection

The features are extracted from the dataset, including voltage, current, and other derived quantities. The target is the fault type (Line-to-Line fault, Line-to-Earth fault, Three-phase fault).

h. Data splitting and Function definition

The SVM model splits the dataset into an 80:20 ratio to train and test. The functions are defined for training, prediction, and evaluation (accuracy, precision).

i. Model training using SVM

The SVM model is trained using all four kernels (Linear, Polynomial, RBF, and Sigmoid) separately to check accuracy. Hyperparameters are tuned to balance the bias and variance.

Received: August 01, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

j. Prediction and Evaluation

After the training is over, the model tests the values for unseen data. The performance of the model is then measured using accuracy and precision, and will be differentiated.

k. Plotting Graphs

The Graphical visualization is generated to understand the behavior of the model and fault classification.

RESULTS AND DISCUSSIONS

The various types of faults were analysed for 11KV distribution substations are as follows.

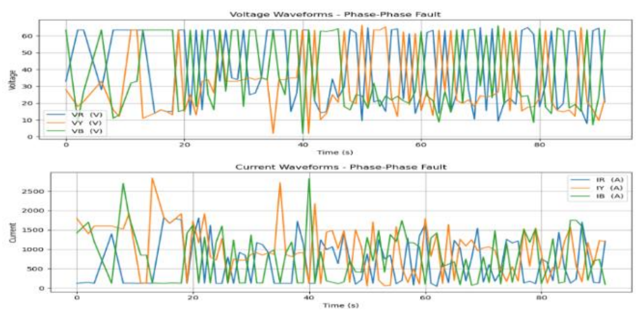

Fig 3(a) Voltage and Current waveform – Phase to Phase

Fig 3(a)Two phases show the voltage dips and the sharp current rise, which confirms that a phase-to-phase short happened. The third phase does remain quite stable. Identification of faults is helped by the unbalanced nature for fault current that is lower.

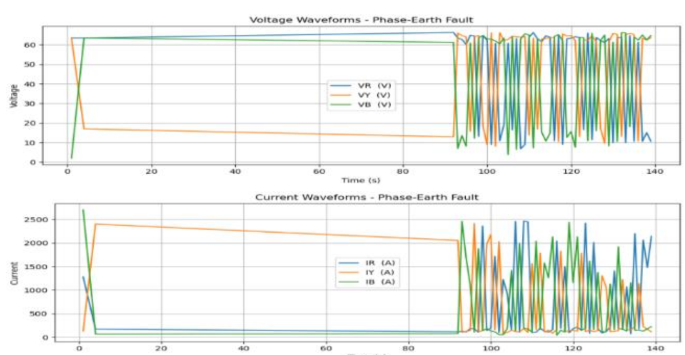

Fig 3(b) Voltage and Current waveform – Phase to Earth

Received: August 01, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Fig 3(b) shows how one phase alone experiences a current spike plus voltage drop while the rest are unaffected. This frequent single-phase-to-ground fault happens due to insulation breakdown and is simply found using the clear one-phase change.

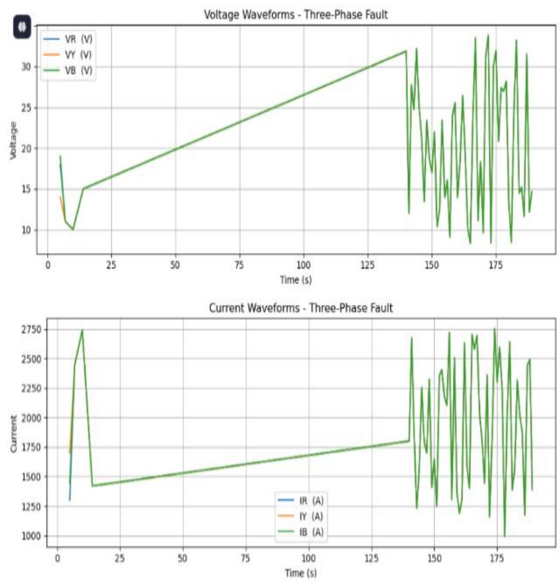

Fig 3(c) Voltage and Current waveform – Three Phase

Fig 3(c) shows equal voltage drop with current surge across all three phases because it indicates a balanced plus severe short circuit. The greatest amount of current comes from this type. Immediate isolation is required to prevent major equipment damage.

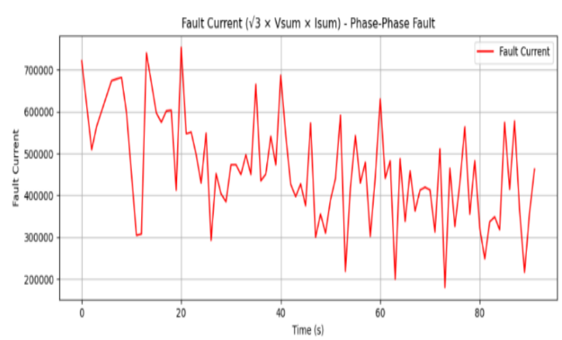

Fig 3(d) Fault current – Phase to Phase fault

Received: August 01, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Fig 3(d) shows Two phases reveal matching fault patterns when there are significant voltage dips as well as current spikes, though the third one stays steady. This corresponds with double-line contact scenarios. These cases are milder than faults with three phases yet still harmful.

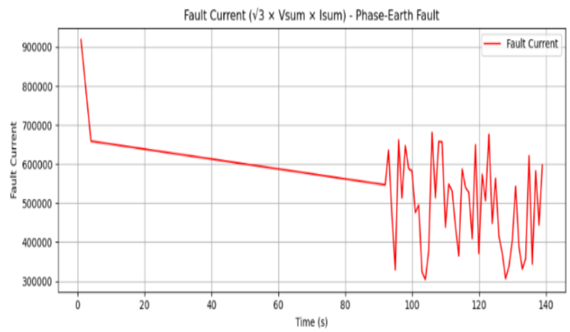

Fig 3(e) Fault current – Phase to Earth fault

Fig 3(e) shows how one phase displays a sudden current surge as well as voltage drop, and also other phases are unaffected, and that confirms a ground fault. It is able to cause localized issues and it must be cleared quickly though it is less severe than three-phase faultsprovided by the drop down menu to differentiate the head from the text.

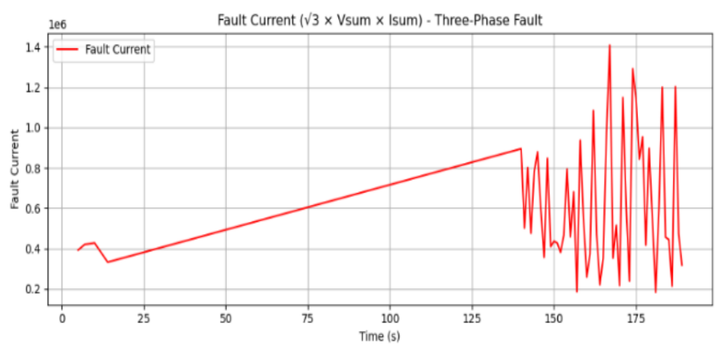

Fig 3(f) Fault current – Three phase fault

Fig 3(f) shows that All phases have equal current surges and voltage drops again because they match the balanced total short signature. Quick disconnection avoids system instability that is

Received: August 01, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

common.The graphs collectively depict the signatures,electrically distinct, from line-to-earth, line-to-line, and three-phase faults in a power system. Three-phase faults can display symmetrical voltage collapse along with current surge across all phases, marking them as being the most severe. Rapid isolation is needed for protection of equipment because they are of the most severe kind. Line-to-line faults can affect only two phases here, with sharp voltage dips also with current spikes in those phases while the third one remains stable, and this also aids in that quick fault recognition. Line-to-earth faults usually disturb just one phase while the other phases are not affected because insulation breaks down, so isolation and detection are simpler.It becomes easier to identify the type as well as severity of a fault in real time via comparing these patterns. Then, protection actions targeted protection actions can then be enabled. As voltage and current waveforms are reliable indicators for distinguishing between balanced and unbalanced faults, they ensure more rapid decision-making for system protection. By preventing fault escalation, this analysis minimizes downtime also supports predictive maintenance but confirms fault type.

CONCLUSION

Fault analysis remains vital to power distribution system's stability and safety maintenance. It can accurately detect balanced faults such as three-phase faults, and it can differentiate unbalanced faults such as line-to-line and line-to-earth faults. From the voltage and current waveforms, the severity and nature of faults can be determined. Such a decision allows correct selection and coordination for protective devices such as relays and circuit breakers. In this paper, MATLAB simulations were analysed for an 11 kV power distribution feeder. Waveform-based analysis, as a result, does support faster fault identification and isolation, as well as improved system reliability. Prediction accuracy that is greater can be achieved by improving the developed model given more operational data collected over time. It would allow for real-time suggestions of solutions for or corrective measures with more precise fault detection. Outage durations could be reduced greatly as a result of such improvements that would optimize maintenance scheduling as well as strengthening power distribution network resilience.

REFERENCES

[1]G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] S.Saravanan, R.Senthil Kumar and P.Balakumar,"Binary firefly algorithm based reconfiguration for maximum power extraction under partial shading and machine learning

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

approach for fault detection in solar PV arrays," Applied Soft Computing Journal 154 (2024) 111318.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] Shyni P Nair, M.S. Sivagama Sundari,"Optimizing day ahead energy management with demand response in a PV – diesel battery system using a hybrid GOA-SSN strategy" Journal of Energy Storage,Vol 76, 2024.

[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopystudies on magneto-optical media and plastic substrate interface," IEEE Trans. J. Magn., vol. 2, pp. 740–741, August 1987.

[6] P. Edwards and J. Berry, "Characterisation of 11 kV fault level contributions based on substation load profile," in 24th International Conference & Exhibition on Electricity Distribution (CIRED), Glasgow,UK, 2017, pp. 2153-2156.

[7] T. Thakur, "Three Phase Faults Analysis of Power System," GlobalJournal of Researches in Engineering: F Electrical and ElectronicsEngineering, vol. 16, no. 5, 2016.

[8] N. Gana, N. F. Ab Aziz, Z. Ali, H. Hashim, and B. Yunus, "AComprehensive Review of Fault Location Methods for Distribution PowerSystem," Indonesian Journal of Electrical Engineering and ComputerScience, vol. 2, no. 3, pp. 185-192, Jun. 2016.

[9] A. Farughian et al., "Power line signalling based earth fault location,"The Journal of Engineering, vol. 2018, no. 15, pp. 1155-1159, 2018.

[10] Y. Liu, M. Razeghi-Jahromi, and J. Stoupis, "Unsupervised High Impedance Fault Detection Using Autoencoder and Principal Component Analysis," DOI:10.48550/arXiv.2301.01867, 2023.

[11] K. Moloi, N. W. Ndlela, and I. E. Davidson, "Fault Classification andLocalization Scheme for Power Distribution Network," Applied Sciences,vol. 12, no. 23, p. 11903, Nov. 2022.

[12] M. Caporuscio, A. Dupuis, and W. Löwe, "Data-Driven Ground-Fault Location Method in Distribution Power System With Distributed Generation," ar https://doi.org/10.48550/arXiv.2402.14894, 2024.

[13] T. F. Moraes, L. Lovisolo, and L. F. C. Monteiro, "Fault location indistribution systems from analysis of the energy of sequence componentwaveforms," IET Generation, Transmission & Distribution, vol. 12, no. 9,pp. 1951–1960, May 2018.

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[14] Praveen Rai, Narendra D. Londhe and Rithesh Raj, “Fault classification in power distribution network integrated with distributed generators usingCNN, Electric Power Systems Research, vol 192, 2021, 106914, ISSN 0378-7796.

[15] Balan A, Srujan T.L, Manitha P.V, Deepa K,”Detection and Analysis of Faults in Transformer using Machine Learning”, IDCIOT 2023 –International Conference on Intelligent Data Communication Technologies and Internet of Things, Proceedings 2023, pp. 477-482.

[16] Prashanth S,R Sucheta, Vishva R, Ganesh kumar T R, NavyaMohan,”BIST Based Aging Fault Prediction Using Machine Learning”, 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC) 10.1109/ICESC51422.2021.9532739,2021.