

# Data-Driven for Neutral-to-Ground Voltage: An Evaluation of Dataset Using Deep Learning

Mohd Ezwan Mahadan\*, Ahmad Farid Abidin, Rijalul Fahmi Mustapa, Mohd Abdul Talib Mat Yusoh and Muhammad Asraf Hairuddin

**Abstract**— Excessive neutral-to-ground voltage (NTGV) poses safety and operational risks in TT-grounded secondary distribution systems, particularly when caused by ground faults. This study investigates the effectiveness of data structure and density on a deep learning-based approach for classifying NTGV events as originating from either upstream or downstream fault locations. A deep learning model using LSTM is trained and tested using real-world datasets collected from multiple locations in a TT system. Results show that input data structure S3 which incorporating ground and phase current signals alongside NTGV measurements significantly improves model performance, achieving an accuracy of 98.15% and an F1-score of 97.85%. Notably, the proposed model maintains high classification accuracy even when trained using only 10% to 30% of the available dataset, demonstrating strong robustness under limited data conditions. These results indicate that feature-rich input structures enable reliable and data-efficient NTGV source classification with minimal additional computational cost, supporting practical deployment in real-world distribution networks.

**Index Terms**— Deep learning, LSTM Classification, Neutral-to-Ground Voltage (NTGV), TT Grounding System

## I. INTRODUCTION

Excessive neutral-to-ground voltage (NTGV) is a critical power quality (PQ) disturbance that compromises the safety, reliability, and performance of low-voltage secondary distribution systems (SDS), particularly in networks with TT grounding configurations. In these systems, ground faults can introduce elevated NTGV levels throughout the network, even when the fault occurs at a specific location. This behavior arises from the interconnected nature of local earthing points, which allows fault-induced voltages to propagate beyond the fault origin [1]. Such conditions may lead to equipment malfunction and insulation degradation, and hazardous touch voltages at end-user premises [2].

This manuscript is submitted on August 1, 2025, revised on January 15, 2026, accepted on February 10, 2026, and published on April 30, 2026.

Mohd Ezwan Mahadan, Rijalul Fahmi Mustapa and Muhammad Asraf Hairuddin are from the Faculty of Electrical Engineering, Universiti Teknologi MARA, Kampus Johor, Cawangan Pasir Gudang, Johor.

Ahmad Farid Abidin and Mohd Abdul Talib Mat Yusoh are from the Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam.

\*Corresponding author  
Email address: ezwan7101@uitm.edu.my

1985-5389/© 2026 The Authors. Published by UiTM Press. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Timely and accurate localization of NTGV events is essential for effective fault response and risk mitigation. However, conventional methods often face limitations when applied to TT systems, where distributed earthing introduces ambiguity in fault traceability [5]. These challenges have led to increased interest in data-driven approaches that leverage time-series analysis of electrical signals for automated classification and localization.

Recent advances in deep learning (DL) have demonstrated substantial potential for power quality monitoring, with models such as convolutional neural networks (CNNs), long short-term memory (LSTM), and gated recurrent units (GRUs) showing success in recognizing patterns within voltage and current waveforms [6][7]. These approaches have been successfully used to classify a wide range of PQ disturbances, including voltage sags, swells, harmonics, and transient events[8]. Nevertheless, the majority of existing studies focus on generic PQ disturbances under controlled or simulated conditions, often assuming balanced systems and abundant labelled datasets[9].

In the specific context of NTGV analysis, existing works remain limited and predominantly address detection rather than source localization or rely on rule-based or threshold-driven techniques that are difficult to generalize to complex grounding environments [10]. Furthermore, few studies explicitly consider TT-grounded secondary distribution systems, where the behavior of NTGV is strongly influenced by local earthing characteristics and real-world operating conditions. In our previous study [11], we introduced a LSTM-based model specifically designed for classifying NTGV source locations, demonstrating its effectiveness using real-world measurements from TT-grounded systems. Despite these advances, two critical aspects remain insufficiently explored in the literature. First, existing PQ on NTGV classification studies have never provide a systematic analysis of how different input feature configurations affect model performance. Second, most DL-based approaches implicitly assume the availability of large, well-balanced training datasets, an assumption that is often unrealistic for NTGV events caused by ground faults, which are rare and difficult to capture in practice [12], [13]. The combined impact of input data structure and limited training data availability on NTGV source classification performance has not been adequately investigated, especially using real-world measurements from TT-grounded systems.

Building on this foundation, the present study focuses on evaluating two critical factors that influence model performance: the structure of input data and the density of

training data. The input data structure refers to the combination of features provided to the model. For instance, NTGV signals may be used alone or in combination with ground current and three-phase current measurements, depending on the instrumentation available. The training data density pertains to the proportion of available labeled data used for model training, which is a particularly important consideration given that NTGV events caused by ground faults are rare and difficult to capture under field conditions.

To address this limitation, this study presents a detailed evaluation of the LSTM-based model introduced in [11], trained and tested under various configurations of input structure and training data density. The classification task is formulated as a binary problem, in which each NTGV event is labeled according to whether its source is located upstream or downstream relative to the monitoring point. Real-world measurements from TT-grounded systems experiencing ground fault conditions are used to develop and validate the model.

The main contributions of this paper are threefold. First, the study systematically quantifies the impact of different input feature configurations on NTGV source classification performance. Second, it evaluates the robustness of the proposed model under reduced training data densities, reflecting realistic field deployment constraints. Third, it provides practical insights for the design of data-efficient and reliable NTGV monitoring systems tailored to TT-grounded secondary distribution networks. The rest of this paper is organized as follows: Section 2 details the methodology and experimental design; Section 3 presents the results; Section 4 discusses the implications and relevance of the findings; and Section 5 concludes with a summary and directions for future research.

## II. METHODOLOGY

This section outlines the methodological framework used to evaluate the effect of input data structure and training data density on the performance of a LSTM-based deep learning model for localizing the source of excessive NTGV in TT-grounded SDS. The model architecture used in this study is based on our previously validated design presented in [11]. The methodology includes data acquisition and preprocessing, construction of input structures and data density scenarios, dataset partitioning, model training, and performance evaluation. Fig. 1 depicts the overall process in this work to evaluate the NTGV data.

### A. Data Acquisition and Preprocessing

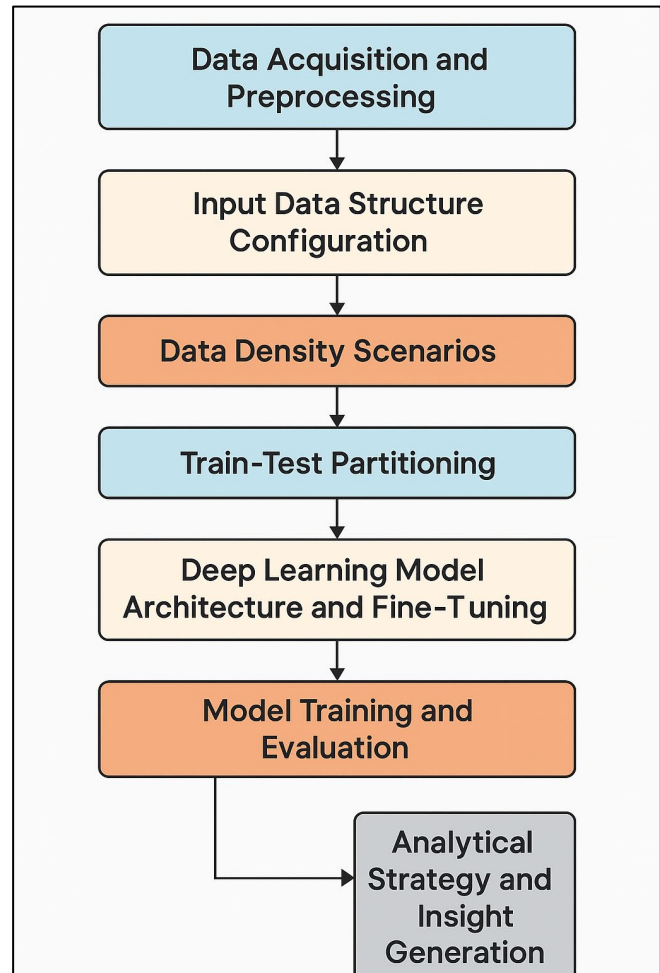
The dataset used in this study was obtained from a TT-grounded SDS under uncontrolled ground fault conditions. The system was subjected to a series of single-line-to-ground faults at various locations that generates NTGV disturbances. Voltage and current waveforms were recorded using high-resolution data logger positioned at key nodes in the distribution network.

All signals were sampled at 12.8 kHz and segmented into fixed-length windows representing two power frequency cycles (40 ms). Each instance in the dataset was labelled as either upstream or downstream, depending on the relative location of

the fault source with respect to the measurement point.

Subsequently, the cumulative number of labelled samples totalled 510 upstream instances and 411 downstream instances. These quantities reflect the natural imbalance in ground fault propagation as observed in TT networks and serve as the basis for developing a binary classification model.

This preprocessing framework ensured that the dataset was structured and appropriately labelled, thereby enabling effective training and evaluation of the deep learning model under varying data and feature scenarios.



**Fig. 1.** Flowchart illustrates the overall evaluation framework, including data acquisition, input feature configuration, dataset partitioning, DL model training, and performance evaluation.

### B. Input Data Structure Configuration

Three progressively complex data structures were defined to investigate the effect of feature richness on classification performance. Each structure was composed of synchronized time-series waveforms with varying dimensionality, as summarized in Table 1.

This progressive design allows for evaluating how richer signal content enhances the model's ability to distinguish between upstream and downstream faults.

TABLE 1. INPUT DATA STRUCTURES FOR LSTM MODEL

Structure	Input Features	Description
S1	NTGV	Only neutral-to-ground voltage waveform
S2	NTGV, Ground Current	Adds ground current to represent fault current behavior
S3	NTGV, Ground Current, Three-Phase Line Currents	Full feature set representing voltage and current asymmetries across phases

### C. Data Density Scenarios

To evaluate the impact of data availability on model performance, the dataset was down sampled into five distinct training densities: 10%, 30%, 50%, 70%, and 90% of the total labeled data. For each input structure, the training subset was selected randomly to match the desired density level, while the remaining portion of the dataset was retained for testing, as detailed in the subsequent section.

Table 2 summarizes the data density levels used during training along with the number of training samples at each level.

TABLE 2. TRAINING DATA DENSITY LEVELS AND CORRESPONDING SAMPLE SIZES

Density Level	Percentage of Dataset	Number of Samples
D1	10%	93 samples
D2	30%	278 samples
D3	50%	461 samples
D4	70%	645 samples
D5	90%	829 samples

This configuration allows a systematic investigation into how the quantity of labeled data interacts with input feature richness to influence deep learning model performance for NTGV classification.

### D. Train-Test Partitioning

All data configurations were partitioned into training and testing sets using an 80:20 split. To ensure generalizability and prevent data leakage, no overlapping segments were allowed

between the training and testing sets. For each configuration, five-fold cross-validation was conducted on the training data, and the average results were reported.

This partitioning strategy ensured that the evaluation metrics accurately reflect the model's ability to generalize to unseen NTGV events.

### E. Deep Learning Model Architecture and Training

The deep learning model employed in this study is based on an LSTM architecture previously developed and validated in [11]. LSTM networks are particularly well-suited for capturing temporal dependencies and long-range patterns in time-series data, making them effective for analysing voltage and current waveforms associated with PQ disturbances such as excessive NTGV.

As illustrated in Fig. 2, the model receives multivariate time-series input through a sequence input layer, followed by an LSTM layer comprising 150 hidden units configured with an output mode set to "sequence." This is followed by a dropout layer with a dropout rate of 0.5 to reduce overfitting by randomly deactivating neurons during training. A second LSTM layer, with 100 hidden units, is added with an output mode set to "last" to extract the final temporal representation. A second dropout layer with a dropout rate of 0.1 is applied before passing the extracted features to a fully connected layer with two output neurons representing the binary classes (upstream and downstream). The final classification is performed through a softmax layer and a cross-entropy classification layer. The model was trained and evaluated using MATLAB R2025a on a Windows-based workstation equipped with an Intel Core i7-8750H CPU, 8 GB of RAM, and an NVIDIA GeForce GTX 1050 Ti GPU. The training process was configured using the Adam optimizer with a maximum of 200 epochs. The initial learning rate was set to 0.0001, and a gradient threshold of 1 was applied to avoid exploding gradients. The model was trained using sequences of the longest available length, with data shuffled at every epoch to improve generalization. Training verbosity was disabled to streamline execution, and no training progress plots were displayed.

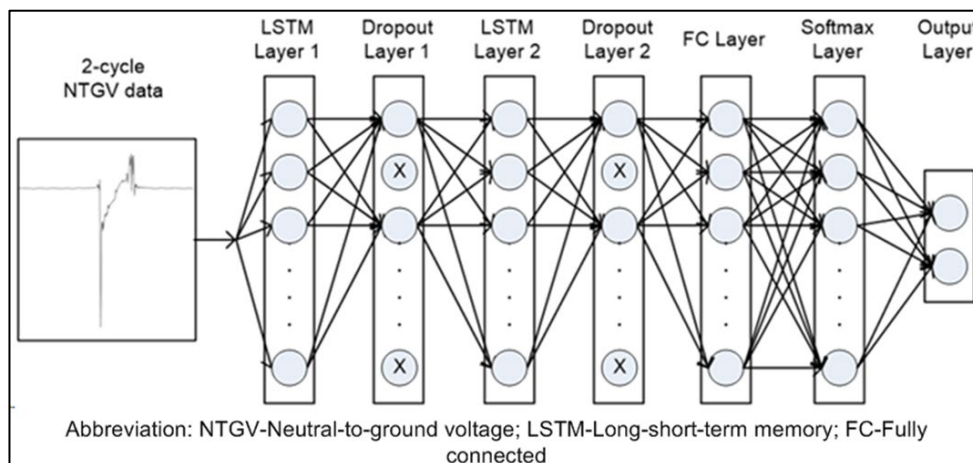


Fig. 2. Architecture of the LSTM-based deep learning model used for classifying the source location of NTGV events as upstream or downstream. [11].

This configuration ensured a stable and consistent training process across all experimental scenarios. Each model was trained using stratified subsets of the dataset defined by different data density levels and input structures, as described in earlier sections. The best-performing model under each condition was evaluated on the corresponding test set to assess classification accuracy and F1-score.

### III. EXPERIMENT RESULTS AND ANALYSIS

This section presents the results of the LSTM-based model applied to the binary classification of excessive NTGV events originating from either upstream or downstream locations in a TT-grounded SDS. The analysis focuses on the influence of input data structure and training data density on classification performance, using metrics such as accuracy and F1-score.

#### A. Impact of Input Data Structure

To assess the influence of feature richness on classification performance, the LSTM model was evaluated using three different input structures: S1 (NTGV only), S2 (NTGV + ground current), and S3 (NTGV + ground current + three-phase line currents). All models were trained using 100% of the dataset, and performance was measured using accuracy, F1-score, and training time.

As illustrated in Fig. 3, both accuracy and F1-score improved progressively with increasing input complexity. For S1, the model achieved an accuracy of 97.39% and an F1-score of 97.19%. While acceptable, this structure provided limited spatial and directional information, which constrained the model’s ability to fully distinguish between upstream and downstream fault scenarios.

By incorporating ground current in S2, the model’s performance increased to 97.72% accuracy and 97.46% F1-score. The additional ground current features captured important fault-related dynamics, leading to more balanced and confident predictions.

The highest performance was achieved with S3, where the inclusion of three-phase line currents further enriched the input representation. The model attained an accuracy of 98.15% and an F1-score of 97.85%, reflecting improved feature discrimination and classification precision. These results affirm the benefit of using comprehensive current and voltage features for NTGV source localization in TT-grounded systems.

In addition to absolute training times, the relative computational overhead between input configurations was evaluated by expressing the increase in training duration as a percentage as depicted in Fig. 4. Compared to the baseline configuration S1, the training time increased from 204.1 s to 214.3 s for S2, corresponding to an increase of approximately 5.0%. For the most feature-rich configuration S3, the training time increased to 221.2 s, representing an overall increase of approximately 8.4% relative to S1. These results indicate that, although additional input features lead to a measurable increase in training time, the relative overhead remains modest. When considered alongside the substantial improvement in classification performance achieved by S3, the added computational cost is justified and remains practical for offline

training and periodic model updates in real-world NTGV monitoring applications. Overall, the results demonstrate that enriching the input structure significantly enhances model performance. Even though additional features require slightly more computation time, the gain in classification accuracy and F1-score justifies the inclusion of current-based features in practical NTGV monitoring systems.

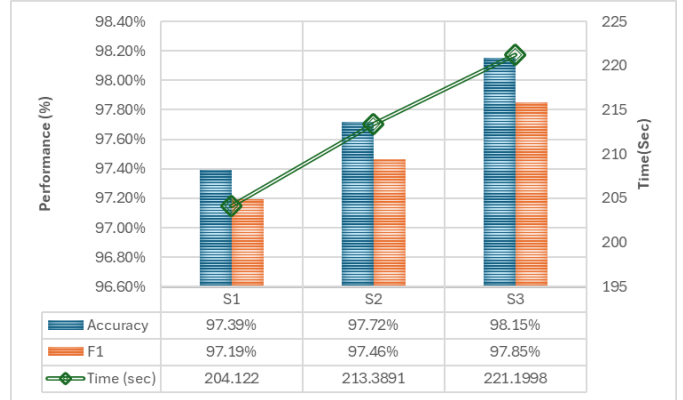


Fig. 3. Impact of data structure on DL performance

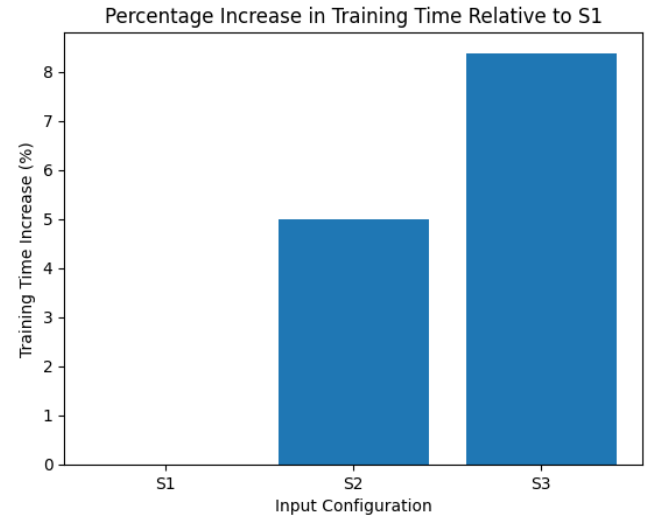


Fig. 4. Percentage increase in model training time for different input configurations relative to the baseline configuration S1.

#### B. Influence of Training Data Density

To evaluate the robustness of the DL-based model under varying data availability, the classification accuracy was assessed across six training data densities: 10%, 30%, 50%, 70%, and 90%. This evaluation was repeated for each input structure to understand how data quantity interacts with input complexity. As shown in Fig. 5, accuracy improved consistently with increasing data density for all structures. However, the rate of improvement and overall performance varied significantly depending on the richness of the input features.

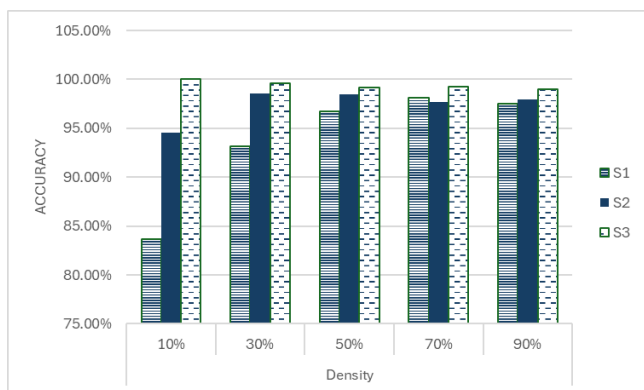
For Structure 1 (S1), the model showed limited generalization when trained on sparse data, achieving only

83.4% accuracy at 10% density. As data density increased, accuracy gradually improved, reaching approximately 96.5% at 90%. This sensitivity to training volume indicates that NTGV-only input lacks sufficient redundancy or spatial cues to perform well under data-constrained conditions.

Structure 2 (S2) demonstrated better resilience to limited data. With just 10% of the data, the model achieved 94.6% accuracy, an 11.2% improvement over S1 highlighting the value of ground current features in supporting early-stage learning. As data density increased, accuracy continued to improve steadily, reaching approximately 97.7% at 90% density.

Structure 3 (S3) consistently outperformed the other configurations across all density levels. Remarkably, with only 10% of the data, the model achieved near-perfect classification with 100% accuracy and remained above 98% accuracy across all other density levels. This demonstrates the strong representational power of the combined NTGV and current signals, which enabled the model to learn class distinctions effectively, even under severe data limitations.

The diminishing accuracy gap between structures at higher densities suggests that while additional features are especially critical under sparse data conditions, they also support more stable learning across different dataset sizes. These results underscore the practical importance of incorporating richer input structures in real-world NTGV classification systems, particularly when labelled data are limited due to operational, safety, or economic constraints.



**Fig. 5.** Impact of data density to the accuracy of the DL model.

In addition to classification accuracy, the time for each input structure was also recorded across all data densities to assess the computational cost associated with increased data volume and feature complexity. As illustrated in Fig. 6, training time rose with both input richness and data density, but the rate of increase varied across structures.

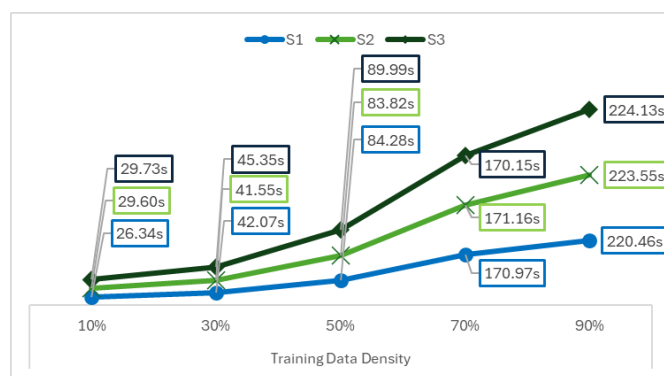
For Structure 1 (S1), which includes NTGV data only, training time grew significantly as the data density increased. The model required only 26.34 seconds at 10% density but expanded to 220.46 seconds at 90%. This steep rise is attributed to the increased number of samples processed, despite the relatively simple input structure. The results suggest that while S1 is computationally light at small data volumes, its training

cost becomes comparable to more complex structures as the dataset grows.

Structure 2 (S2), which includes NTGV and ground current, showed a similar trend, with training time increasing from 29.60 seconds at 10% to 223.55 seconds at 90%. The additional input channel introduced modest computational overhead, but the efficiency remained comparable to S1, especially at lower densities, making S2 an attractive compromise between performance and training efficiency.

Structure 3 (S3), which integrates NTGV, ground current, and three-phase line currents, incurred the highest training times overall. At 10% density, training took 29.73 seconds, rising to 224.13 seconds at 90%. The increase reflects the additional input dimensions, which contribute to longer sequence processing and weight updates in the LSTM layers. Although S3 incurred slightly higher computational cost, the time difference between S1, S2, and S3 at 90% density was marginal, while the performance benefits were substantial. The results suggest that enriching input features yields meaningful gains in classification accuracy and robustness, with relatively minor increases in training time, particularly when operating in offline or non-time-critical environments.

This analysis highlights a practical trade-off between computational load and model performance. When data availability is limited or training efficiency is critical, S2 may offer a balanced solution. However, for applications prioritizing accuracy and resilience especially under sparse fault data conditions, S3 remains the most effective configuration with acceptable computational cost. The results affirm that selecting an appropriate input structure should consider both the availability of sensor data and computational constraints, especially when deploying NTGV monitoring systems in resource-constrained environments or edge-based platforms.



**Fig. 6.** Consumption time in seconds for each dataset

#### IV. DISCUSSION AND PRACTICAL IMPLICATIONS

The results presented in this study demonstrate the effectiveness of a deep learning-based framework for classifying excessive NTGV events in TT-grounded SDS. The classification was based on a binary model that distinguishes whether the ground fault is located upstream or downstream relative to the monitoring point. Two key factors were evaluated: input data structure and training data density, both of which significantly impacted model performance and usability.

### A. Importance of Input Data Structure

The input feature structure emerged as a primary determinant of classification accuracy and robustness. As shown in Section 3.1, enriching the input beyond NTGV alone with ground and phase currents (S2 and S3) yielded considerable gains in accuracy and F1-score. The inclusion of current signals, particularly in S3, enabled the model to capture more discriminative fault-related patterns associated with spatial and load asymmetries within the SDS.

This improvement is especially critical in TT systems, where the interconnection of local earth electrodes can lead to voltage rise at multiple locations, complicating fault source identification. The results affirm that multi-modal input signals, especially those related to fault current paths, play a key role in disambiguating the fault origin. These findings are consistent with earlier work in PQ disturbance classification but represent a novel application in the context of NTGV localization due to ground faults.

### B. Data Density and Model Robustness

Data availability is a known constraint in NTGV-related studies due to the infrequency of ground faults and the complexity of conducting controlled experiments. This study systematically evaluated model performance across a range of data densities and found that richer input structures help mitigate the adverse effects of limited data. In particular, the S3 configuration delivered high accuracy even when trained on only 10–30% of the dataset, making it suitable for real-world scenarios where labelled data is scarce.

This resilience highlights the value of designing input structures that offer maximal informational content per sample, allowing the model to generalize effectively from fewer training examples. Such a strategy is essential for practical deployment, particularly in environments with limited data storage, transmission capacity, or ground fault event history.

### C. Practical Deployment Considerations

From a deployment perspective, the slight increase in computational time associated with feature-rich input structures is minimal and manageable. As shown in Section 3.2, the training time difference between S1 and S3 at full data density was approximately 4 seconds, indicating that the additional training cost does not present a significant barrier, especially in offline or scheduled retraining scenarios.

Furthermore, the trained model's lightweight inference nature, once deployed, makes it suitable for implementation on embedded devices or edge computing platforms that continuously monitor NTGV behaviour and alert maintenance personnel when abnormal patterns are detected.

## V. CONCLUSION

This study evaluated an LSTM-based deep learning model for classifying the source location of excessive neutral-to-ground voltage (NTGV) events in TT-grounded secondary distribution systems. By analyzing different input structures and data densities, the results showed that adding ground and phase current signals significantly improved classification

accuracy and robustness, even with limited training data. Enriching the input features, particularly through the S3 configuration, yielded considerable gains in both accuracy and F1-score. The S3 configuration achieved an accuracy of 98.15% and an F1-score of 97.85%, and maintained high classification performance even when trained using only 10 to 30 percent of the available dataset. These findings highlight the value of using rich input features and support the development of practical, data-driven NTGV monitoring systems. This data-efficient behavior makes the proposed approach well suited for real-world NTGV monitoring scenarios, where labeled fault data are inherently scarce and difficult to obtain. Future work will explore the model's adaptability to other grounding configurations, such as TN and IT systems, and investigate transfer learning to enhance generalization across different network environments

## ACKNOWLEDGEMENT

The authors would like to express their profound gratitude to the Faculty of Electrical Engineering at Universiti Teknologi MARA (UiTM), Cawangan Johor, Kampus Pasir Gudang, for their persistent support and important assistance throughout this research project. Their assistance and resources have been critical to the effective completion of this project.

## REFERENCES

- [1] D. Paul, "Low-voltage power system surge overvoltage protection," *IEEE Transactions on Industry Applications*, vol. 37, no. 1, pp. 223–229, 2001, doi: 10.1109/28.903152.
- [2] J. D. Bouford, S. Member, D. J. Hurley, and S. Member, "Across Publicly Accessible Surfaces," vol. 29, no. 1, pp. 48–51, 2014.
- [3] R. O. Peruzzi, "Forensic engineering analysis of a swimming pool electric shock injury," *Journal of the National Academy of Forensic Engineers*, vol. 38, no. 1, pp. 131–139, 2021, doi: 10.51501/jotnafe.v38i1.64.
- [4] G. van Schaik *et al.*, "Possible causes of aberrations in adverse grouping behavior of dairy cows: A field study," *Journal of Dairy Science*, vol. 104, no. 6, pp. 7000–7007, 2021, doi: 10.3168/jds.2020-19269.
- [5] *IEEE Std. 1695-Guide to Understanding, Diagnosing, and Mitigating Stray and Contact Voltage*. 2016.
- [6] M. E. Mahadan, A. F. Abidin, M. A. T. M. Yusoh, M. A. Hairuddin, and N. D. K. Ashar, "End-to-end deep learning on classification of neutral-to-ground voltage in secondary distribution systems," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 10, Dec. 2024, doi: 10.1016/j.prime.2024.100795.
- [7] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 209–220, 2021, doi: 10.17775/CSEEJPES.2020.02700.
- [8] H. Liu, F. Hussain, S. Yue, O. Yildirim, and S. J. Yawar, "Classification of multiple power quality events via compressed deep learning," *International Transactions on Electrical Energy Systems*, vol. 29, no. 6, Jun. 2019, doi: 10.1002/2050-7038.12010.
- [9] R. Machlev, A. Chachkes, J. Belikov, Y. Beck, and Y. Levron, "Open source dataset generator for power quality disturbances with deep-learning reference classifiers," *Electric Power Systems Research*, vol. 195, no. March, p. 107152, 2021, doi: 10.1016/j.epsr.2021.107152.
- [10] D. Stetzer, A. M. Leavitt, C. L. Goeke, and M. Havas, "Monitoring and remediation of on-farm and off-farm ground current measured as step potential on a Wisconsin dairy farm: A case study," *Electromagnetic Biology and Medicine*, vol. 35, no. 4, pp. 321–336, 2016, doi: 10.3109/15368378.2015.1089888.
- [11] M. E. Mahadan, A. F. Abidin, M. A. T. M. Yusoh, N. D. K. Ashar, R. F. Mustapa, and M. A. Hairuddin, "LSTM-based Deep Learning Method for Excessive Neutral-to-ground Voltage (NTGV) Localization in Secondary Distribution Systems," in *2024 IEEE 4th International*

*Conference in Power Engineering Applications: Powering the Future: Innovations for Sustainable Development, ICPEA 2024*, IEEE, 2024, pp. 361–366. doi: 10.1109/ICPEA60617.2024.10498326.

- [12] J. Hong, Y. H. Kim, H. Nhung-Nguyen, J. Kwon, and H. Lee, “Deep-Learning Based Fault Events Analysis in Power Systems,” *Energies (Basel)*, vol. 15, no. 15, pp. 1–16, 2022, doi: 10.3390/en15155539.
- [13] R. Machlev, A. Chachkes, J. Belikov, Y. Beck, and Y. Levron, “Open source dataset generator for power quality disturbances with deep-learning reference classifiers,” *Electric Power Systems Research*, vol. 195, no. March, p. 107152, 2021, doi: 10.1016/j.epsr.2021.107152.