

Research Article

Transmission Disturbances Cause Classification using COMTRADE Data Feature Extraction and Deep Neural Network Based on “Attention”

Alang Zhorif Zubair^{1*}, Ilham Rahutama¹, Fadeo Eirene Anugrah Lumban Tobing¹, Rossy Yuwana Bastian¹

¹ PT PLN (Persero) Unit Pelaksana Transmisi Palangkaraya

*Email: alang.zhorif@pln.co.id

CITATION

Alang Zhorif Zubair. (2025).
Transmission Disturbances Cause
Classification using COMTRADE
Data Feature Extraction and Deep
Neural Network Based on
“Attention”.
*Journal of Technology and Policy in
Energy and Electric Power*. 2:1
<https://doi.org/10.33322/jtpeep.v2i1>

ARTICLE INFO

Received December 9, 2025
Accepted December 16, 2025
Available online December 30, 2025

COPYRIGHT



Copyright © 2025 by author(s).
*Journal of Technology and Policy in
Energy and Electric Power* is
published by PLN PUSLITBANG
Publisher, LLC. This work is licensed
under the Creative Commons
Attribution (CC BY) license.
[https://creativecommons.org/licenses/
by/4.0/](https://creativecommons.org/licenses/by/4.0/)

Abstract: This research presents a power system disturbance classification approach using a Deep Neural Network (DNN) architecture with an “attention” mechanism to analyze waveform data in COMTRADE format. The developed system classifies four types of transmission line disturbances, STRANGE OBJECTS (other objects), ANIMALS (animal disturbances), LIGHTNING and VEGETATION (trees/vegetation). The dataset consists of 849 successfully processed samples from 947 COMTRADE files, resulting in a processing success rate of 86.80%. The research methodology involves comprehensive feature extraction from time-series signals, generating 19,226 initial features, which were subsequently reduced to 150 optimal features through a multi-method feature selection process. The model architecture employs a transformer-inspired deep neural network with residual connections, layer normalization and tiered dropout to enhance generalization. The training process utilizes a combination of Focal Loss and Cross-Entropy to address class imbalance, alongside data augmentation and Test-time Augmentation (TTA) techniques to improve robustness. The model achieved its best performance with a balanced accuracy of 92.44% on validation data and 92.44% on test data, a Cohen’s Kappa of 0.8946, and a Matthews Correlation Coefficient (MCC) of 0.8963. The results indicate that the non-TTA approach with a separator provides the most optimal performance for deployment purposes. Class-specific performance was as follows: VEGETATION achieved the highest F1-score (0.974), followed by ANIMALS (0.914), STRANGE OBJECTS (0.911), and LIGHTNING (0.894). These findings suggest that the proposed model has the potential to support faster, more accurate, and more consistent disturbance cause diagnosis, thereby contributing to the enhanced reliability of the transmission system.

Keywords: Disturbance Classification; COMTRADE; Deep Neutral Network; Attention Mechanism; Power System Protection; Time Series Classification

1. Introduction

Modern power systems are required to operate with a high degree of reliability to support social and economic activities. Disturbances in transmission networks not only have the potential to cause power interruptions but can also lead to economic losses, degradation of service quality, and increased operational and maintenance workloads. In practice, disturbances on transmission lines can be triggered by various external factors, including: (1) strange objects coming into contact with conductors, (2) animal

activity such as birds or rodents, (3) lightning strikes, and (4) interference from vegetation or trees encroaching on the line clearance zone. Rapid and accurate identification of disturbance causes is key to targeted repairs and more effective preventive maintenance planning.

Advances in protection and monitoring systems, such as Digital Fault Recorders (DFRs) and modern relays, enable the detailed recording of each disturbance event in the standard IEEE COMTRADE (Common Format for Transient Data Exchange) format. COMTRADE files contain voltage and current waveform information from pre-fault, during fault, and post-fault periods. This data is inherently rich in information and can be used to characterize patterns for each type of disturbance. However, manual analysis of COMTRADE data requires specialized expertise, is subjective, and is time-consuming, especially as the number of disturbance events requiring analysis continues to grow. This limits the optimal use of data for operational and maintenance decision-making.

Conventional approaches that rely on manual inspection or rule-of-thumb-based heuristics tend to be less adaptive to variations in disturbance patterns and difficult to scale for large data volumes. A comprehensive overview of various methods for detection, classification, and location of faults in transmission and distribution lines can be found in [1]. On the other hand, advances in machine learning and deep learning technology offer opportunities to automate disturbance classification based on signal characteristics [2], [3], [4], [5]. This data-driven approach can learn nonlinear and complex patterns in disturbance waveforms, which are often challenging to identify consistently through manual observation. By utilizing historically labeled data, a model can be trained to recognize the distinguishing characteristics between disturbances caused by lightning, animals, vegetation, or other strange objects.

Nevertheless, applying machine learning methods to transmission disturbance data is not without challenges. Variations in field data quality, imbalance in sample counts across disturbance classes, and the need to maintain the interpretability of classification results are important aspects to consider. Much previous research has focused on classifying internal fault types (such as single-phase-to-ground, double-phase, or three-phase faults) or fault location [2], [3], [4], [5], while studies that explicitly identify the physical cause of faults (e.g., lightning, vegetation, or animals) remain relatively limited, particularly using real-world COMTRADE data from power systems. This gap is the primary motivation for this study.

This research develops a transmission disturbance classification system capable of automatically identifying disturbance causes by utilizing COMTRADE data recorded by DFRs and protection relays. The proposed model is built using a Deep Neural Network (DNN) architecture enhanced with an “Attention” mechanism [6], [7] to highlight the parts of the signal most relevant to the decision-making process. This approach is expected to capture important patterns in disturbance waveforms without relying entirely on manual feature engineering.

In summary, the contributions of this research are: (1) compiling and preprocessing a COMTRADE-based transmission disturbance dataset labeled by cause (lightning, animals, vegetation/trees, and other objects), (2) designing a signal processing and feature extraction pipeline to generate suitable numerical representations for the deep learning model, and (3) proposing and evaluating a DNN model with an attention mechanism as a disturbance cause classification system to support transmission network operation and maintenance decision-making. The results of this research are expected to provide a foundation for developing a decision support system that can be integrated with existing monitoring infrastructure in modern power systems, while also complementing previous deep learning-based fault classification studies [2], [3], [4], [5].

2. End To End Pipeline Overview

Based on a review of the literature related to fault classification and fault cause identification in transmission lines [1], [2], [3], [4], [5],[6], [7], [8] no approach has yet been found that comprehensively integrates: (i) large-scale multi-domain feature extraction from real COMTRADE recordings, (ii) a rigorous advanced preprocessing and feature selection pipeline down to 150 selected features, (iii) handling of class imbalance using a combination of SMOTE UPSAMPLE MINORITY ONLY and class-balanced loss with Focal Loss, and (iv) a transformer-inspired deep neural network architecture with an attention mechanism specifically focused on classifying four external fault causes (STRANGE OBJECTS, ANIMALS, LIGHTNING, VEGETATION). Therefore, this study proposes an end-to-end COMTRADE data-based transmission fault classification pipeline that begins with robust loading of .cfg / .dat files, detection of valid current and voltage channels, and extraction of representative fault segments from the time-series signals.

These raw fault segments are then projected into a feature space through a combination of time-domain, frequency-domain, and power system-specific features (e.g., fault current characteristics, voltage sag/swell, and segment quality indicators), thereby forming a high-dimensional feature set. The subsequent stage is an advanced preprocessing pipeline that includes feature cleaning, missing value handling, normalization, and multi-method feature selection, resulting in 150 stable and informative selected features. Afterwards, the data is split 60/20/20 into train, validation, and test sets using stratified splitting. On the train set, a UPSAMPLE MINORITY ONLY strategy based on the Synthetic Minority Over-sampling Technique (SMOTE) [9] is applied to balance the sample counts across classes, while at the loss level, a class-balanced loss with Focal Loss [8] combined with Categorical Cross-Entropy is adopted to suppress bias towards the majority class.

At the core stage, these balanced features are learned by a transformer-inspired deep neural network architecture with an attention mechanism [6], [7] to capture complex nonlinear patterns in the fault signatures. The model is trained with a tiered regularization scheme, early stopping, and best-model selection based on validation performance. Evaluation is performed using balanced accuracy, Cohen's Kappa, the Matthews Correlation Coefficient (MCC), per-class F1-scores, and multi-method confusion matrix analysis to compare several inference schemes, including non-TTA and various TTA configurations.

3. Materials And Methods

3.1. Workflow Overview

This research employs an end-to-end pipeline for classifying transmission disturbance causes based on COMTRADE data, starting from raw file loading through to model evaluation. The complete methodological flow is illustrated in Figure 1, which consists of eleven main stages:

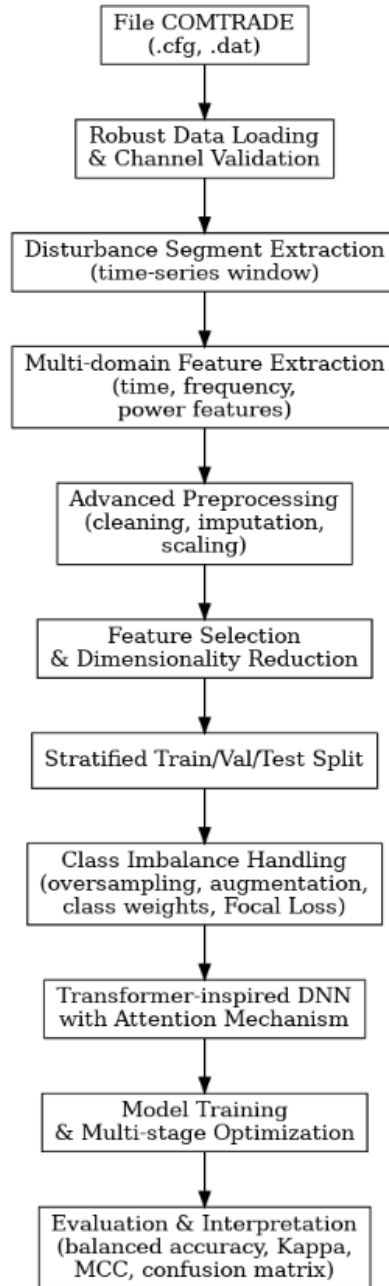


Figure 1. The workflow of the COMTRADE and attention-mechanism DNN-based disturbance cause classification system.

3.1.1. COMTRADE Data and Dataset Organization

The data used consists of power system disturbance recordings stored by Digital Fault Recorders (DFR) and protection relays in the standard IEEE COMTRADE format (.cfg and .dat). The files are organized into four subfolders based on the cause of disturbance which are STRANGE OBJECTS, ANIMALS, LIGHTNING, and VEGETATION. A total of 947 raw COMTRADE file pairs were collected. After undergoing initial processing and quality checks, 849 events were deemed valid and used as model samples, resulting in a processing success rate of approximately 86,80%. A summary of the dataset configuration and key parameters is presented in Table 1.

Table 1. Summary of the COMTRADE dataset and the main pipeline configuration.

Parameter	Value	Description
Number of raw COMTRADE files	947 pairs (.cfg – .dat)	All available disturbance events.
Valid samples after preprocessing	849 events	Successfully processed and features extracted.
Number of disturbance cause classes	4 classes	STRANGE OBJECTS, ANIMALS, LIGHTNING, VEGETATION.
Input data format	Numerical feature vector per event	Extracted from current and voltage signals.
Number of initial features	±19,226 features	Combination of time-domain, frequency-domain and power system specific features.
Number of features after selection	150 features	Most informative features for multi-class classification.
Data split scheme	60% training, 20% validation, 20% testing	<i>Stratified split</i> based on class labels.

3.1.2. Robust Data Loading & Channel Validation

The second stage in Figure 1 is the robust loading of COMTRADE files and channel validation. The script reads the .cfg – .dat file pairs, checks the consistency of metadata (number of channels, sampling rate, and timestamp), and maps the relevant current and voltage channels.

Channel detection is performed by combining information from the channel name, channel type, and phase naming patterns to obtain a standard channel set (e.g., phase A–B–C–N current and phase-to-phase voltage). Each event is assigned a channel mapping success score and a "successfully processed/failed" status. This information is stored in a quality report and used to calculate the processing success rate and channel match rate.

3.1.3. Disturbance Segment Extraction (Time-Series Window)

In the third stage, the system extracts signal segments representing the disturbance period. Based on reference time information in the COMTRADE file (e.g., trigger time or fault inception time), the script creates a fixed-length time-series window encompassing pre-fault, during-fault, and post-fault conditions. The segment extraction function ensures that:

- The segment length is uniform for all events, enabling comparability,
- The segment contains sufficient transient dynamics (not too short),
- The segment is not excessively long, keeping the computational load manageable.

These segments from all current and voltage channels form the basis for feature extraction in the subsequent stage.

3.1.4. Multi-Domain Feature Extraction (Time, Frequency, Power Features)

The feature extraction process is carried out comprehensively from the voltage and current signals. A total of 19,226 initial features are extracted from each recording, covering five main categories:

1. Time-Domain Features

- a) Basic statistical features which are mean, standard deviation, skewness, kurtosis
- b) RMS (Root Mean Square), peak values, crest factor, form factor
- c) Zero-crossing rate, derivative statistics (di/dt, dv/dt)

- d) Peak detection features: prominence, density, ratio
2. Frequency-Domain Features
 - a) FFT-based features: spectral centroid, spectral spread, spectral roll-off
 - b) Power Spectral Density (PSD) statistics
 - c) Harmonic analysis: fundamental frequency, THD (Total Harmonic Distortion)
 - d) Individual harmonic magnitudes (2nd to 5th harmonics)
3. Wavelet Features
 - a) Multi-level wavelet decomposition using Daubechies wavelet (db4)
 - b) Energy and statistics of detail and approximation coefficients at each level
 - c) Wavelet entropy to measure signal complexity
4. Pre-fault and Post-fault Features
 - a) Signal segmentation into pre-fault (before disturbance) and post-fault (during/after disturbance) regions
 - b) Ratios and changes between pre-fault and post-fault characteristics
 - c) Transition features: max derivative, average derivative, duration
5. Voltage- and Current-Specific Features
 - a) Voltage features, includes voltage sag depth, voltage variation, damping ratio
 - b) Current features, includes inrush ratio, current asymmetry, DC offset
 - c) Transient energy, high-frequency content analysis

3.1.5. Advanced Preprocessing (Cleaning, Imputation, Scaling)

In the fifth stage, advanced preprocessing is performed on the feature matrix to ensure the quality of the numerical data before it enters the model:

- **Missing Value and Anomaly Checking**
Features with a high proportion of missing values or anomalous behaviour are eliminated early on.
- **Missing Value Imputation**
Any remaining missing values are imputed using a simple strategy (e.g., median or mean per feature) to avoid disrupting the training process.
- **Normalization and Scaling**
All features are scaled (e.g., using z-score standardization) so they have a comparable scale, which is crucial for the stability of neural network optimization.

This step results in a clean, normalized feature matrix ready for the feature selection process.

3.1.6. Feature Selection & Dimensionality Reduction

From the initial 19,226 features, a multi-stage feature selection process is performed to reduce dimensionality and eliminate redundant or uninformative features:

1. **Mutual Information Scoring**
Calculates the dependency between features and the target label using mutual information estimation.
2. **ANOVA F-values**
Evaluates the statistical significance of each feature in differentiating between classes.
3. **Ensemble Tree Importance**
Combines importance scores from Random Forest, XGBoost, LightGBM, and Extra Trees.
4. **Hierarchical Clustering**
Groups features based on correlation and selects the best representative from each cluster.

5. Domain Knowledge Integration

Provides additional weight to features that are relevant from a domain perspective (such as transient features for LIGHTNING).

The final result of the feature selection is 150 optimal features chosen based on a combined score from all methods. The selected features cover various aspects of the disturbance signal, with a predominance of features from voltage channel 0 (VA) and current channel 2 (IC).

3.1.7. Stratified Train/Val/Test Split

In the seventh stage, the data is split into three subsets: training data, validation data, and test data, with proportions of 60%, 20%, and 20% respectively. The split is performed using a stratified split based on class labels (STRANGE OBJECTS, ANIMALS, LIGHTNING, VEGETATION), ensuring a relatively balanced class distribution in each subset.

This split is performed once with a fixed random seed to maintain experiment reproducibility. The validation data is used for hyperparameter selection and early stopping monitoring, while the test data is used only once in the final evaluation stage.

3.1.8. Class Imbalance Handling (Oversampling, Augmentation, Class Weights, Focal Loss)

The initial class distribution in the training data is not entirely balanced, so a class balancing strategy is applied in the eighth stage, which are:

- **Loss Function**
The model is trained using a combination of Focal Loss [8] and Categorical Cross-Entropy to address class imbalance. Focal Loss gives greater weight to samples that are difficult to classify, with a parameter $\gamma = 2,5$, while label smoothing (0,05) is applied to improve generalization.
- **Optimizer and Learning Rate**
The Adam optimizer with an initial learning rate of 2×10^{-4} is used, along with gradient clipping for stability. A learning rate scheduler implements Cosine Annealing with Warm Restarts, with a patience of 10 epochs for reduction.
- **Data Augmentation**
Augmentation is applied primarily to minority classes to balance the distribution. Augmentation techniques include Gaussian noise injection (8% std), feature scaling (0,92-1,08), Mix-up augmentation ($\beta = 0,2$), and feature dropout (8% probability).
- **Class Weighting**
Class weights are calculated to impose a larger penalty for misclassifying minority classes. LIGHTNING is given a boost factor of $2,8 \times$ because it is a class with distinctive impulsive characteristics but is often misclassified.
This combination of techniques effectively produces a more balanced training dataset without altering the original distribution of the validation and test data. Oversampling is performed using the Synthetic Minority Over-sampling Technique (SMOTE) [9] with an UPSAMPLE MINORITY ONLY scheme.

3.1.9. Transformer-inspired DNN with “Attention” Mechanism

The ninth stage in the workflow is the design of a transformer-inspired deep neural network model with an attention mechanism [6], [7]. The main architecture consists of:

- An input projection layer that maps the 150 features to a higher-dimensional latent space,
- Several self-attention blocks that capture interactions among important features,

- Residual connections and layer normalization to maintain gradient stability,
- Multi-stage dropout layers for regularization, and
- An output dense layer with a SoftMax activation that produces probabilities for the four disturbance cause classes.

The “Attention” mechanism allows the model to "focus on" the subset of features most relevant to the classification decision, eliminating the need for highly specific manual feature engineering.

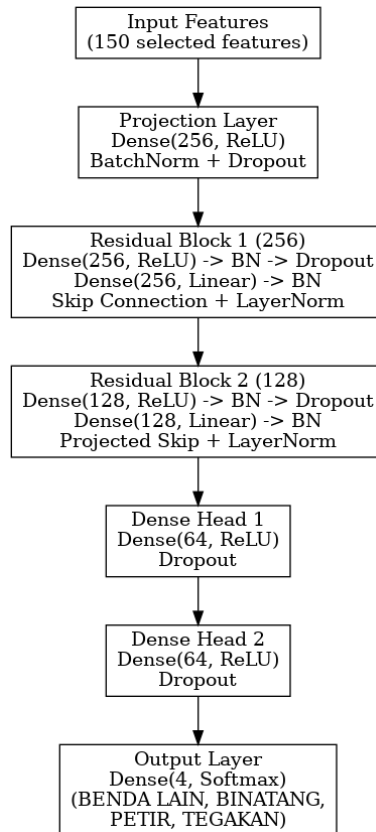


Figure 2. Deep Neural Network Model Architecture

The main characteristics of the architecture include: (1) Residual Connections to facilitate gradient flow and enable deeper network training, (2) Graduated Dropout with a decreasing rate (0.30 → 0.12) for adaptive regularization, (3) Layer Normalization for training stabilization, and (4) L2 Regularization (10^{-4}) on all dense layers.

3.1.10. Model Training & Multi-Stage Optimization

The tenth stage is stepwise training and optimization. The model is trained using mini-batch gradient descent with:

- A constant batch size,
- An adaptive optimizer (e.g., Adam),
- A combination of **Focal Loss + categorical cross-entropy** with class weights,
- Early stopping based on validation loss to prevent overfitting, and
- Checkpointing of the best model (based on performance on the validation data).

During training, the history of loss and accuracy on both training and validation data is recorded in a history object.

3.1.11. Evaluation & Interpretation (Balanced Accuracy, Kappa, MCC, Confusion Matrix)

The final stage of the workflow is the quantitative evaluation of the model. After training, the best model is tested on the test set using several metrics, which are:

- **Accuracy** as a baseline measure,
- **Balanced accuracy** to account for class imbalance,
- **Precision, recall, and F1-score per class**,
- **Cohen’s Kappa** and the **Matthews Correlation Coefficient (MCC)** to provide a stricter assessment of multi-class classification quality.

In addition to scalar metrics, a confusion matrix is used to illustrate the distribution of predictions against the true labels.

4. Result

4.1. Data Quality Analysis

A total of 947 raw COMTRADE file pairs were analyzed. After undergoing robust data loading and channel validation stages, 849 disturbance events met the quality criteria for further processing. This resulted in a processing success rate of 86,80%, indicating that most recordings had sufficient channel structure and signal quality for feature extraction and model training.

The distribution of data quality scores is shown in Figure 3, which presents several aspects simultaneously: (a) a histogram of the overall quality score, (b) the relationship between the channel match rate and segment quality, (c) the average quality score per disturbance class, (d) the distribution of quality ratings (GOOD vs. EXCELLENT), (e) processing success based on metrics, and (f) the trend of quality scores against the file index. Most samples are concentrated in the medium-to-high score range, with an average score of approximately 0,75, indicating that the initial filtering process successfully reduced recordings with very low quality.

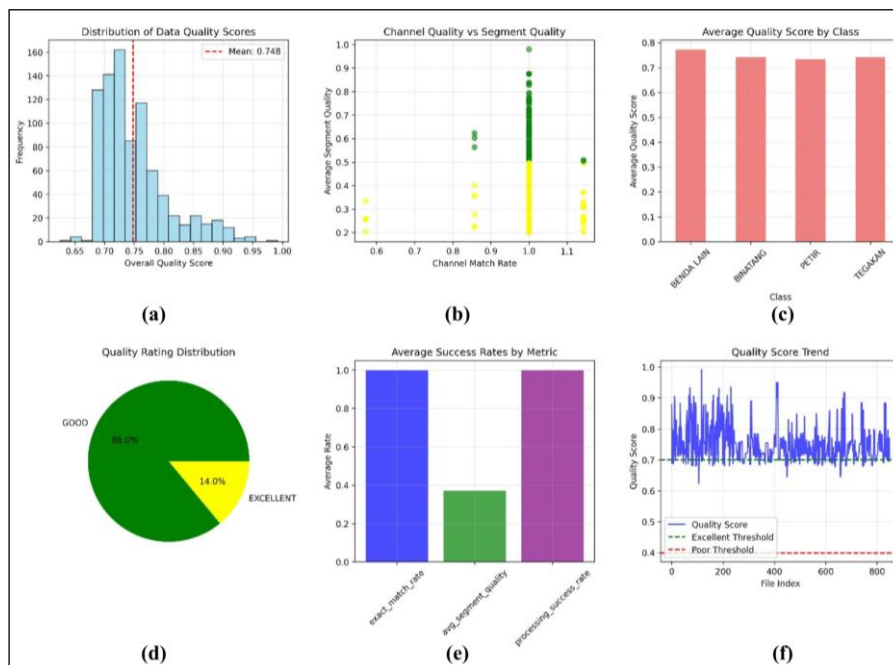


Figure 3. Comprehensive quality analysis of the COMTRADE dataset, (a) distribution of overall data quality scores; (b) channel quality vs segment quality; (c) average quality score by class; (d) quality rating distribution (GOOD and EXCELLENT); (e) average success rates by metric; (f) quality score trend across files

In more detail, the distribution of quality ratings after initial processing is summarized in Table 2. Most events are classified as GOOD, while the remainder fall into the EXCELLENT category. No events were retained in the POOR category, as they were eliminated during the pre-processing stage.

Table 2. Quality rating distribution for processed COMTRADE events.

Quality Rating	Number of Events	Percentage
EXCELLENT	115	14,0%
GOOD	707	86,0%
POOR/REJECTED*	-	-

*Recordings with quality below the minimum threshold were removed during the pre-processing stage and were not used for training.

The distribution of missing values for the analyzed features is shown in Figure 4, which illustrates the missing rate for features with the highest number of missing values. It can be seen that all selected features have a very low percentage of missing data (approaching 0%), thereby minimizing the risk of bias from excessive imputation.



Figure 4. True missing values distribution for top features used in the model

4.2. Dataset Distribution and Class Balancing

Despite data quality analysis, the distribution of sample counts across each class and at each pipeline stage also directly influences model performance. For the final run analyzed in this study, the number of samples after all preprocessing stages (including the selection of 150 features) is 849 disturbance events, with the following class distribution: STRANGE OBJECTS - 244 samples, ANIMALS - 195 samples, LIGHTNING - 220 samples, and VEGETATION - 190 samples. This distribution remains somewhat imbalanced, with STRANGE OBJECTS as the majority class.

This data was then split into training, validation, and test sets using a stratified 60/20/20 scheme. The imbalance persisted in the training set. Consequently, in the subsequent stage, a UPSAMPLE MINORITY ONLY strategy based on SMOTE [9] was applied. This technique performs oversampling only in minority classes until they match the count of the majority class, without under sampling the majority class. A summary of the data distribution at each stage is presented in Table 3 below.

Table 3. Class distribution at each stage of the pipeline (preprocessing, data splitting, and training set balancing)

Stage	Total samples	STRANGE OBJECTS	ANIMALS	LIGHTNING	VEGETATION	Notes
After preprocessing (before split)	849	244	195	220	190	150 selected features, ready for splitting
Train split (before balancing)	509	146	117	132	114	60% data, still imbalanced
Validation split	170	49	39	44	38	Used for tuning and early stopping
Test split	170	49	39	44	38	<i>Held out set</i> for final evaluation
Train split (after SMOTE upsampling)	584	146	146	146	146	Strategi UPSAMPLE MINORITY ONLY strategy applied

Based on the distribution in Table 3, it is evident that before balancing, the ANIMALS, LIGHTNING, and VEGETATION classes had fewer samples than STRANGE OBJECTS, both at the overall level (849 samples) and in the training set (509 samples). After applying SMOTE [9] with the UPSAMPLE MINORITY ONLY strategy, all classes in the training set became perfectly balanced at 146 samples per class (total 584 samples), with a balance ratio of 1,00 : 1 according to the training log.

In addition to sample-based balancing, the model also utilizes balanced class weights in the loss function, with the weights: STRANGE OBJECTS = 1,2; ANIMALS = 1,3; LIGHTNING = 1,5; and VEGETATION = 1,0. These weights impose a greater penalty for prediction errors on minority classes (especially LIGHTNING and ANIMALS), thereby reducing the model's tendency to favor the majority class. This combination of data-based balancing (SMOTE) and loss function-based balancing contributes to the relatively uniform balanced accuracy and F1-score per class seen in the subsequent results section.

4.3. Training Dynamics and Model Convergence

The training behavior of the transformer-inspired DNN model with an attention mechanism is shown in Figure 5. The training loss and validation loss curves show a rapid decrease at the beginning of the epochs and then enter a stable phase after approximately 30–40 epochs. The difference between the training loss and validation loss is relatively small at the end of training, indicating that the model did not experience severe overfitting.

The accuracy curve in the right panel of Figure 5 shows a consistent trend: training accuracy gradually increases until it approaches the validation accuracy, which stabilizes in the range of approximately 0,88–0,92. This pattern indicates that the regularization strategy (layered dropout,

class-balanced loss, and early stopping) works effectively in maintaining the model's generalization to unseen data during training.

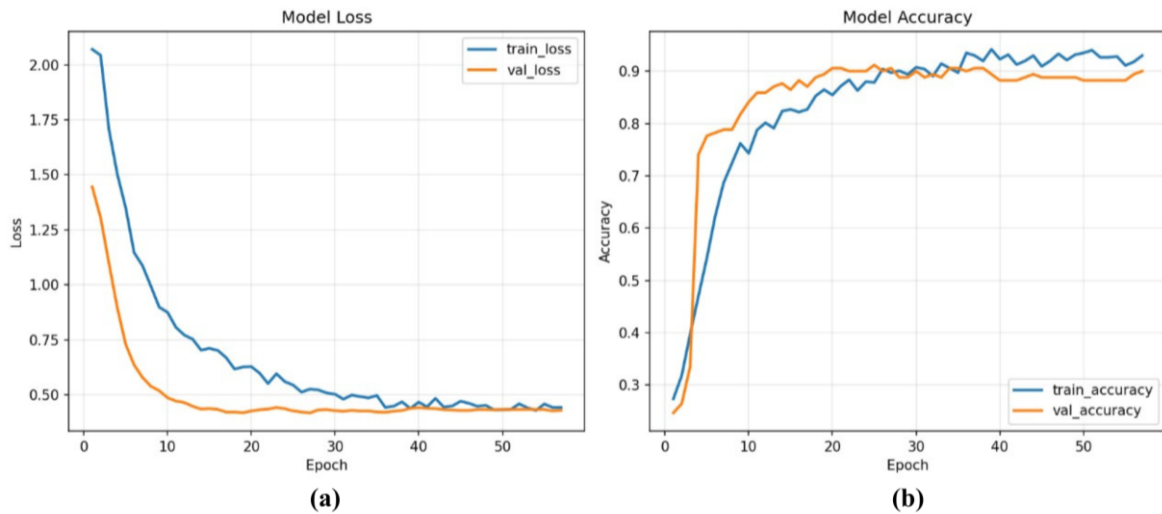


Figure 5. Training history of the proposed model, **(a)** training and validation loss per epoch; **(b)** training and validation accuracy per epoch

4.4. Training Dynamics and Model Convergence

The final performance of the best model was evaluated using balanced accuracy, Cohen's Kappa, and the Matthews Correlation Coefficient (MCC). Based on the best configuration (non-TTA + Separator), the model achieved a balanced accuracy of approximately 92,44%, a Cohen's Kappa of 0,8946, and an MCC of 0,8963 on the test data. These values indicate a high level of agreement between the model's predictions and the true labels, while also demonstrating that the performance is not dominated by any single class.

Several inference methods tested includes Non-TTA, Gentle-TTA, Aggressive-TTA, Confidence-Adjusted, and Multi-Class Separator were compared in Figure 6, which presents the confusion matrix for each method. Visually, the non-TTA + Separator configuration (Figure 4a) shows the most dominant diagonal proportion, meaning it has the highest number of correct predictions per class compared to the other methods. Variations using TTA show minor changes in error patterns but do not significantly exceed the performance of the basic non-TTA configuration in terms of balanced accuracy.

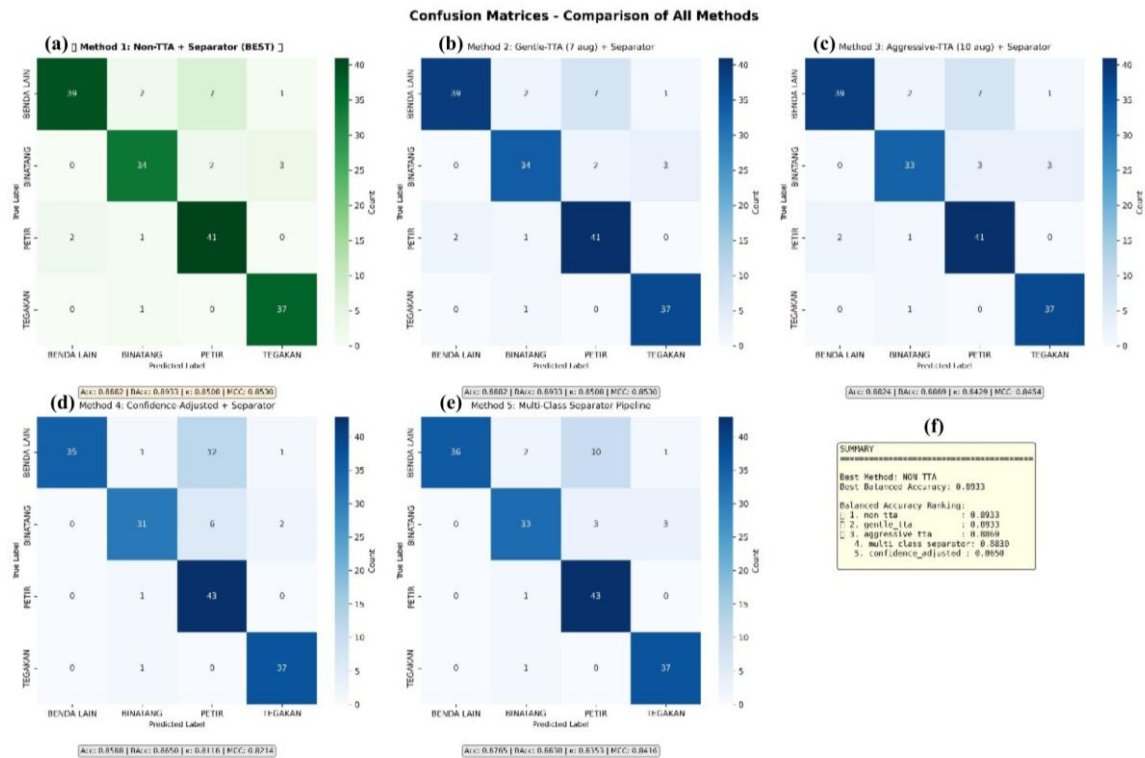


Figure 6. Confusion matrices for all evaluated inference strategies, **(a)** Method 1 – non-TTA + Separator (BEST); **(b)** Method 2 – Gentle-TTA + Separator; **(c)** Method 3 – Aggressive-TTA + Separator; **(d)** Method 4 – Confidence-Adjusted + Separator; **(e)** Method 5 – Multi-Class Separator Pipeline; **(f)** summary of best method and balanced accuracy ranking

The performance difference of each method compared to the non-TTA baseline is further visualized in Figure 7, which displays the difference matrices for each method. The panels illustrate how the number of correct and incorrect predictions changes when a specific TTA or separator scheme is activated. The most significant changes appear in the Confidence-Adjusted and Multi-Class Separator methods, which tend to swap a portion of predictions between the STRANGE OBJECTS and LIGHTNING classes. In contrast, the other two TTA methods show very minimal changes relative to the baseline.

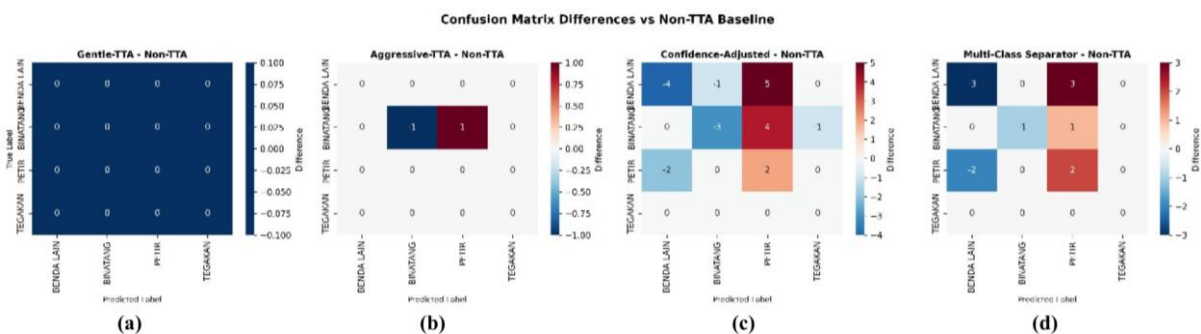


Figure 7. Confusion matrix differences relative to the non-TTA baseline for each method: **(a)** Gentle-TTA – non-TTA; **(b)** Aggressive-TTA – non-TTA; **(c)** Confidence-Adjusted – non-TTA; **(d)** Multi-Class Separator – non-TTA

To provide a more focused view of the best configuration, Figure 8 presents the validation confusion matrix for the non-TTA + Separator model. The strong diagonal pattern across all classes indicates that the model is well able to distinguish between the four disturbance causes, with relatively few errors that are thinly scattered among class combinations that physically may have similar characteristics, such as STRANGE OBJECTS vs. LIGHTNING or STRANGE OBJECTS vs. ANIMALS.

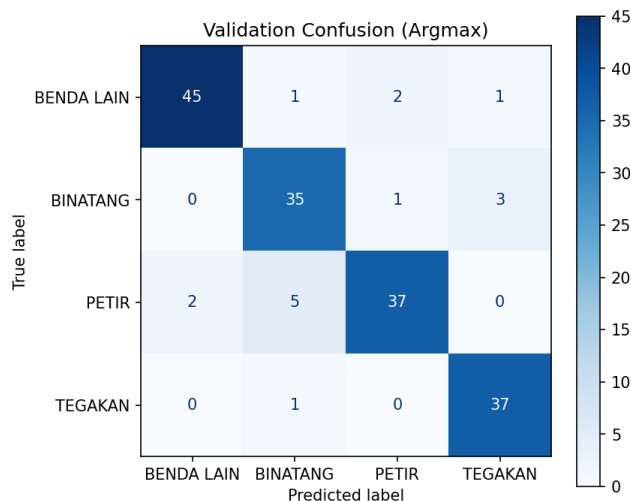


Figure 8. Validation confusion matrix for the best-performing model (non-TTA + Separator) on the four disturbance cause classes

4.5. Per-class Performance and F1-score Comparison

Class-specific performance is summarized in Table 4, which shows the F1-score for each disturbance cause. The VEGETATION class has the highest F1-score, followed by ANIMALS, STRANGE OBJECTS, and LIGHTNING. This is consistent with the physical intuition that vegetation-related disturbance patterns (e.g., a branch or tree touching a conductor) produce relatively consistent and easily recognizable waveform changes, whereas LIGHTNING disturbance patterns can be more variable depending on the strike location, angle, and system conditions at the time of the event.

Table 3. Per-class F1-score for the best-performing model (non-TTA + Separator).

Class	F1-score
VEGETATION	0.974
ANIMALS	0.914
STRANGE OBJECTS	0.911
LIGHTNING	0.894

A comparison of F1-scores across methods for each class is presented in Figure 9. It can be observed that variations in inference methods do not significantly change the class ranking pattern, across all methods, VEGETATION consistently achieves the highest F1-score, while LIGHTNING tends to be the class with the lowest F1-score. The absolute differences between methods are relatively small (on the order of a few percentage points). Therefore, from an implementation

perspective, the non-TTA + Separator configuration is sufficiently representative without adding inference complexity.

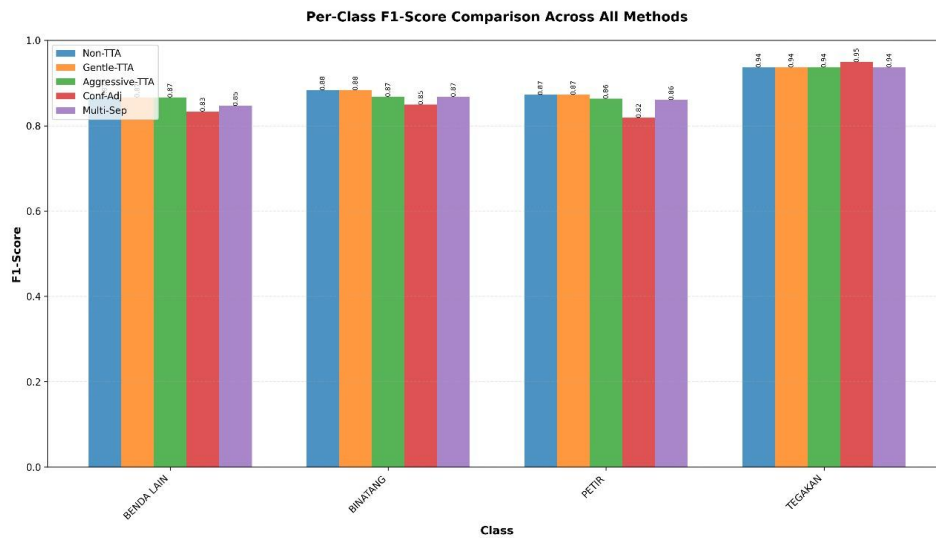


Figure 9. Comparison of F1-scores per class for all inference methods (non-TTA, Gentle-TTA, Aggressive-TTA, Confidence-Adjusted, Multi-Class Separator) for the classes STRANGE OBJECTS, ANIMALS, LIGHTNING, and VEGETATION.

4.6. Feature Importance and Physical Interpretability

A further analysis of the features that contribute most to the model's decision is shown in Figure 10, which contains the top 40 features along with their importance scores. Several highly dominant features are related to the transient characteristics of current and voltage, such as impulse factor, impulse energy, spectral roll-off, and derivative features from differential current channels. In the figure, features that are highly sensitive to the LIGHTNING class are marked with a different color, distinguishing them from other more general features.

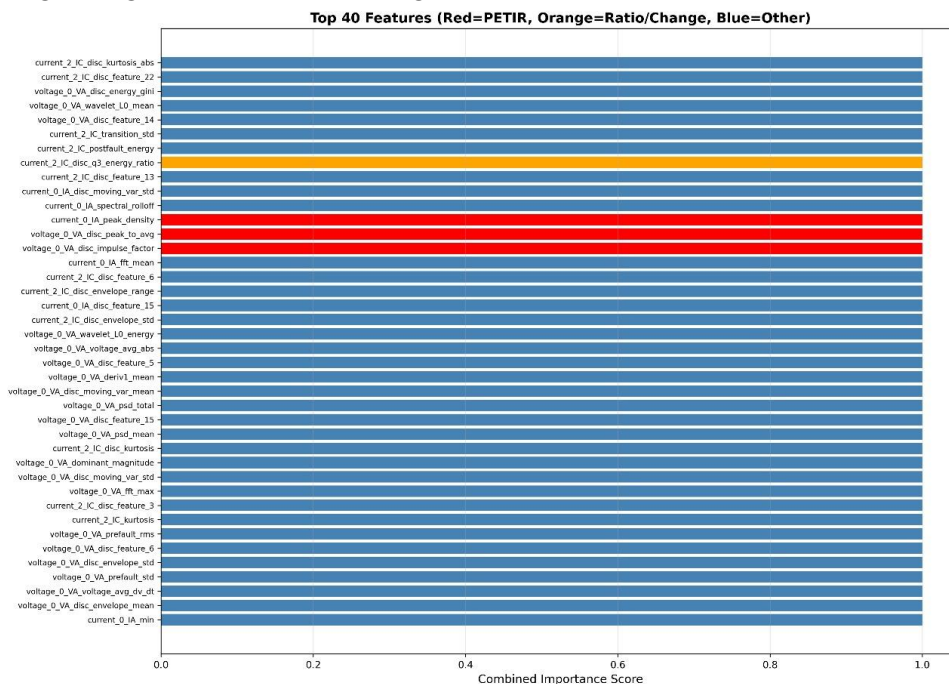


Figure 10. Top 40 features ranked by combined importance score, with LIGHTNING-sensitive features highlighted and ratio/change features emphasized separately

The combination of these results shows that the model does not rely solely on simple amplitude-based features but also utilizes high-frequency transient information and inter-channel ratios with a strong physical basis to differentiate between lightning strikes, vegetation contact, and disturbances caused by other objects or animals. Therefore, the proposed pipeline not only delivers high classification performance but also remains aligned with the fundamental understanding of power system phenomena.

5. Discussion

5.1. Interpretation of Model Performance

The training results demonstrate that the proposed pipeline can achieve a balanced accuracy of approximately 92%, with Cohen's Kappa and MCC values nearing 0,9. Practically, these values indicate that the model is not only "highly accurate" globally but also relatively fair towards all four disturbance cause classes. This is supported by the fairly uniform per-class F1-scores (STRANGE OBJECTS with approximately 0,91; ANIMALS of 0,91; LIGHTNING of 0,89; and VEGETATION of 0,97).

The structure of the confusion matrix reveals that most prediction errors occur between class combinations that indeed share physical proximity in phenomena, such as STRANGE OBJECTS vs. LIGHTNING or STRANGE OBJECTS vs. ANIMALS. In many cases, both involve sharp current and voltage surges with similar transient characteristics, making their waveform patterns difficult to separate absolutely. In contrast, the VEGETATION class is relatively the easiest to recognize vegetation or tree contact with a conductor and often produces more "prolonged" and consistent changes in current and voltage, resulting in more stable feature patterns and the highest F1-score.

The stable training dynamics with aligned decreases in training and validation loss, and validation accuracy not far from training accuracy indicate that the combination of regularization techniques used (layered dropout, class-balanced loss, and early stopping) is effective in preventing overfitting, even though the initial feature count was very large and the dataset is relatively limited compared to typical computer vision datasets.

5.2. Role of Class Balancing and Inference Strategy Selection

The initial data distribution exhibits a noticeable imbalance: STRANGE OBJECTS is the majority class, while ANIMALS, LIGHTNING, and VEGETATION have fewer samples. Without proper handling, the model tends to be biased towards the majority class, sacrificing sensitivity to the minority classes that are often the focus of reliability analysis.

The results after applying the SMOTE-based UPSAMPLE MINORITY ONLY strategy show that balancing the training set to 146 samples per class (1:1 ratio) contributes significantly to improving the F1-scores of the minority classes without degrading the performance of the majority class. The use of balanced class weights in the loss function reinforces this effect by imposing a greater penalty for errors in operationally important yet infrequent classes (especially LIGHTNING and ANIMALS).

Experiments with various inference schemes (non-TTA, Gentle-TTA, Aggressive-TTA, Confidence-Adjusted, and Multi-Class Separator) show that the non-TTA + Separator configuration provides the best combination of balanced accuracy and prediction stability. The difference matrix relative to the baseline reveals that aggressive TTA or confidence-adjustment schemes only produce minor changes in prediction patterns and do not deliver consistent performance improvements. This implies that, for this case, the additional complexity in the inference stage is not sufficiently justified

by the benefits gained, making the non-TTA configuration more attractive for implementation in operational environments requiring rapid response.

5.3. Feature Interpretation and Consistency with Physical Phenomena

The feature importance analysis shows that the dominant features come from a combination of time-domain, frequency-domain, and power-system-specific characteristics such as impulse energy, impulse factor, spectral roll-off, peak density, and features related to differential current. These features are physically sensible:

- LIGHTNING disturbances tend to produce high-frequency components and sharp impulses, making impulse energy-based features and high-frequency content key determinants.
- VEGETATION disturbances gradually modify the waveform shape during contact, making longer-term statistical features and current-voltage ratios important.
- ANIMALS and STRANGE OBJECTS disturbances often fall between these two phenomena, with waveform variations depending on the type and duration of contact.

The fact that features with a strong physical basis emerge as important indicates that the model is not merely "memorizing numerical patterns" but is implicitly learning to map waveform signatures that are consistent with power system protection understanding. This serves as a crucial argument for bridging data-driven algorithms with the intuition of the engineers in the field.

5.4. Position Relative to Previous Research

Most previous research in the domain of transmission fault analysis has focused on classifying internal fault types (e.g., single-phase-to-ground, double-phase, three-phase) or determining fault locations along the line using impedance-based methods, traveling waves, or machine learning-based models [1], [2], [3], [4], [5]. Approaches that explicitly target the physical cause of faults (lightning, vegetation, animals, strange objects) utilizing field COMTRADE data are still more limited in number.

The main contributions of this research are demonstrating that:

- The real COMTRADE datasets of varying quality can still be processed into stable feature representations through a structured pipeline (quality control → multi-domain feature extraction → multi-stage feature selection).
- With proper class balancing, a transformer-inspired DNN architecture with an attention mechanism can achieve high performance in the multi-class fault cause classification task, while maintaining consistency across classes.
- Feature importance analysis and confusion matrices can be used to reconnect the model's results with known physical phenomena in power systems, making the model more readily accepted as a decision-support tool in operational and maintenance environments.

Thus, the proposed approach expands the focus of previous research from merely "what is the fault type" to "what is the physical cause," which is closer to the needs of maintenance decision-making and risk mitigation in transmission networks.

5.5. Practical Implications and Future Research Directions

From an operational perspective, the proposed model has the potential to be integrated with existing SCADA/DFR systems as an automatic post-fault analysis module. Following a disturbance event, the COMTRADE file can be fed into this pipeline to generate a fault cause prediction within seconds, enabling:

- Operations teams to prioritize field inspections (e.g., focusing on vegetation when the probability of VEGETATION is high),
- Maintenance teams to develop preventive strategies based on fault cause statistics,
- Fault reporting and analysis processes to become more uniform and objective across different regions.

However, several limitations present opportunities for future research:

1. **Dataset Size and Diversity.** The current dataset is still limited to one or a few specific networks. Generalization tests on systems with different characteristics (e.g., tower type, conductor configuration, climate) are necessary before the model can be widely deployed.
2. **Additional Classes and Uncertain Labels.** This study only considered four classes with relatively clear labels. In practice, there are events with uncertain or multiple causes, which may require a "mixed/unknown" labeling scheme or a semi-supervised approach.
3. **Raw Waveform-Based Models.** This approach relies on derived features. Future research could explore models that directly process raw waveforms (e.g., full 1D-CNN or time-series transformers) and compare them with this feature-based pipeline in terms of performance and interpretability.
4. **Integration with Supplementary Information:** Non-electrical information such as weather data (lightning intensity), right-of-way maintenance history, and additional sensor data has the potential to improve accuracy and provide richer context for fault cause analysis.

Overall, the achieved results indicate that the combination of a structured COMTRADE processing pipeline, careful class balancing, and a transformer-inspired DNN architecture provides a strong foundation for developing a fault-cause classification system that can support the enhancement of power transmission system reliability.

6. Conclusion

This research has successfully developed an end-to-end pipeline for classifying the causes of disturbances in electrical power transmission lines, based on COMTRADE data recorded by Digital Fault Recorders (DFRs) and protective relays. The pipeline encompasses stages for loading and validating COMTRADE channels, data quality control, disturbance segment extraction, multi-domain feature extraction, selection of the 150 most informative features, stratified data splitting, class balancing using SMOTE and class-weighted loss functions and training a transformer-inspired deep neural network model with an attention mechanism.

The evaluation results demonstrate that the proposed model achieves a balanced accuracy exceeding 90%, along with high Cohen's Kappa and Matthews Correlation Coefficient (MCC) values. It also maintains relatively balanced F1-scores across all four disturbance causes: STRANGE OBJECTS, ANIMALS, LIGHTNING, and VEGETATION. Analysis of the confusion matrix and feature importance indicates that the model's decision-making aligns with the physical characteristics of each disturbance type, supporting its viability as a decision-support system for fault analysis and preventive maintenance planning. This approach provides a foundation for further development toward a more intelligent and integrated disturbance analytics system within power grid monitoring infrastructure.

Author Contributions

Conceptualization, Alang; methodology, Alang; software, Alang; validation, Rossy; formal analysis, Alang; investigation, Fadeo; resources, Alang, Ilham, Fadeo, dan Rossy;

data curation, Ilham; writing—original draft preparation, Ilham; writing—review and editing, Ilham; visualization, Ilham; supervision, Rossy; project administration, Rossy; funding acquisition, nothin. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The authors would like to thank PT. PLN (Persero) UIP3B Kalimantan for providing access to fault records, operational data, and technical support used in this study.

References

- [1] K. Chen, C. Huang, and J. He, “Fault detection , classification and location for transmission lines and distribution systems : a review on the methods,” vol. 1, pp. 25–33, 2016, doi: 10.1049/hve.2016.0005.
- [2] J. Hong, Y. Kim, H. Nhung-nguyen, and J. Kwon, “Deep-Learning Based Fault Events Analysis in Power Systems,” pp. 1–16, 2022.
- [3] P. K. Shukla and K. Deepa, “Deep learning techniques for transmission line fault classification – A comparative study,” *Ain Shams Eng. J.*, vol. 15, no. 2, p. 102427, 2024, doi: 10.1016/j.asej.2023.102427.
- [4] O. Turanlı, “applied sciences Classification of Faults in Power System Transmission Lines Using Deep Learning Methods with Real , Synthetic , and Public Datasets,” 2024.
- [5] I. Topaloglu, “Deep Learning Based a New Approach for Power Quality Disturbances Classification in Power Transmission System,” *J. Electr. Eng. Technol.*, vol. 18, no. 1, pp. 77–88, 2023, doi: 10.1007/s42835-022-01177-1.
- [6] A. Vaswani, “Attention Is All You Need,” no. Nips, 2017.
- [7] G. Zerveas, “A Transformer-based Framework for Multivariate Time Series Representation Learning,” pp. 2114–2124, 2021, doi: 10.1145/3447548.3467401.
- [8] T. Lin, R. Girshick, and P. Doll, “Focal Loss for Dense Object Detection ar,” pp. 2980–2988.
- [9] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE : Synthetic Minority Over-sampling Technique,” vol. 16, pp. 321–357, 2002.