Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sub-Cycle Sliding Window Technique for Real-Time Intelligent Electrical Fault Diagnosis at CW&PRS

Sheetal S. Patil
Bharati Vidyapeeth Deemed to be University College of Engineering
Raghu Chandra Garimella
raghuchandra.g@cwprs.gov.in
Central Water and Power Research Station (CW&PRS), Department of Water Resources, River Development & Ganga Rejuvenation, Ministry of Jal Shakti, Government of India
Gautam Mishra
Bharati Vidyapeeth Deemed to be University College of Engineering
Suhani Kulshrestha
Bharati Vidyapeeth Deemed to be University College of Engineering
Soumyajit Khan
Bharati Vidyapeeth Deemed to be University College of Engineering

Research Article

Keywords: Smart Grid Protection, Intelligent Fault Diagnosis, Sliding Window Architecture, Random Forest, Sub-cycle Detection

Posted Date: July 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8778206/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Sub-Cycle Sliding Window Technique for Real-Time Intelligent Electrical Fault Diagnosis at CW&PRS

Sheetal S. Patil [0000-0003-4381-1228] [1], Raghu Chandra Garimella [0000-0003-3772-9965] [2*], Gautam Mishra [0009-0002-5925-6791] [3*], Suhani Kulshrestha [0009-0000-2904-227X] [4], Soumyajit Khan [0009-0002-0114-843X] [5]

[1] Bharati Vidyapeeth Deemed to be University College of Engineering, Pune-411043, Maharashtra, India, sspatil@bvucoep.edu.in

[*2] Central Water and Power Research Station (CW&PRS), Department of Water Resources, River Development & Ganga Rejuvenation, Ministry of Jal Shakti, Government of India, Pune-411024, Maharashtra, India, raghuchandra.g@cwprs.gov.in

[*3] Bharati Vidyapeeth Deemed to be University College of Engineering, Pune-411043, Maharashtra, India, mishragautammg@gmail.com

[4] Bharati Vidyapeeth Deemed to be University College of Engineering, Pune-411043, Maharashtra, India, suhani.kulshreshth@gmail.com

[5] Bharati Vidyapeeth Deemed to be University College of Engineering, Pune-411043, Maharashtra, India, thesoumyajitkha@gmail.com

Corresponding author(s): Raghu Chandra Garimella and Gautam Mishra

Abstract: *Electrical Grid at Central Water and Power Research Station (CW&PRS), Pune, India, is experiencing its most significant growth in over 100 years. Rapid development of transmission network and integration with renewable energy led to operational challenges. Due to transmission failures, earlier CW&PRS has relied on reactive protection schemes to protect their assets from damage when a cable fault occurs. Reactive protection schemes involve electro-mechanical relays that typically require full Alternating Current (AC) cycle (≈ 20 ms) to calculate root mean square values before isolating fault. Although, reactive protection scheme is proven to be safe, this method does not meet today's demands for reliability. Artificial Intelligence based Deep Learning techniques were employed in pilot studies to intelligently detect the fault at Electrical Grid of CW&PRS, which may cause power outages. High-fidelity Sliding Window technique that replaces traditional cycle-dependent detection has been developed during the studies. This allows the detection system to bypass the 20 ms full-cycle requirement and effectively reduces fault detection latency to just 3.125 ms by structuring input data into continuous and overlapping temporal sequences. Various Machine Learning classifiers under different noise conditions were considered to evaluate the best architecture. Results demonstrate how linear models like Logistic Regression fail at accuracy rate of 76.04%, while the proposed Sliding Window framework achieve 98.91% accuracy in capturing complex non-linear patterns. The present study established an intelligent Sliding Window based Architecture as the primary driver of low-latency performance for high-fidelity fault detection within a fraction of AC cycle.*

Keywords: Smart Grid Protection, Intelligent Fault Diagnosis, Sliding Window Architecture, Random Forest, Sub-cycle Detection.

1. Introduction

Central Water and Power Research Station, (CW&PRS), Pune, India, is a premier national research institution that has been established during 1916 and is currently associated with the Ministry of Jal Shakti, Government of India. The Electric grid at CW&PRS with monthly energy demand of 90,000 kWh has undergone several upgradations over 100 years. In addition, CW&PRS has installed about 700 kWp Solar Plant during 2019 as per Ministry of New and Renewable Energy (MNRE), Government of India, guidelines. More recently, with the aid of National Hydrology Project (NHP), the facilities have undergone substantial improvements like Electrical Substation Automation and Control (ESAC) Centre. ESAC centre has automated the electrical network covering 500 acres of land, which enables high-speed transfer of electrical load information over 10,000 meters using underground optical fibre cables. The ESAC centre has connections to the 14 underground gas-insulated step-down substations (22 kV/415 V) available at Electrical Grid of CW&PRS (Fig. 1). This connection provides direct access to all the substations and ESAC centre server at CW&PRS. The centre is being used to continuously monitor the current, voltage, active power, reactive power, etc. of all electrical loads operating within the research institution including solar power and distribute the power (Fig. 2). The existing automation system is being used to detect/diagnose a fault, *if any*.



Fig. 1 Power Distribution of ESAC centre Facility at CW&PRS

The process of diagnosing faults in modern grid system has many challenges related to the large volume of information produced. Two key factors contributing to these challenges include the quality of information generated and increasing complexity of physical fault processes occurring within an active distribution network.

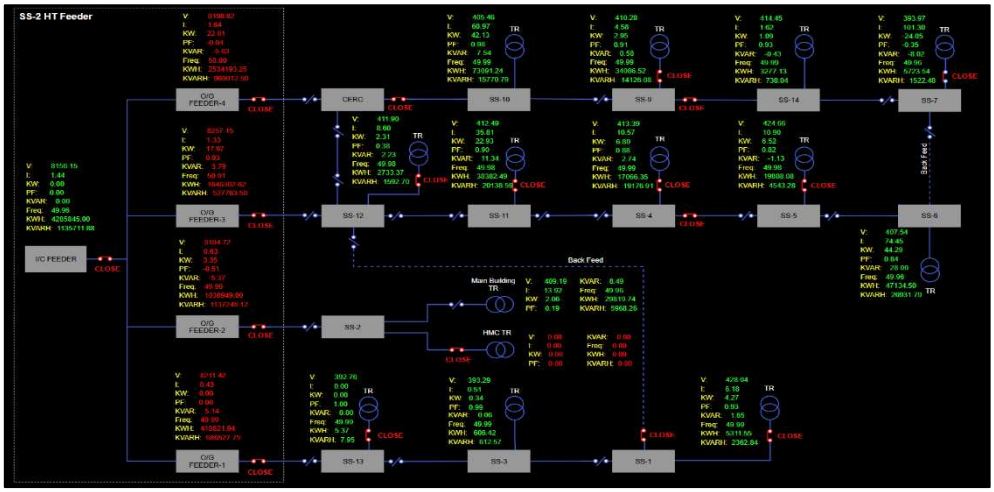


Fig. 2 Monitoring of electrical loads at ESAC centre Facility, CW&PRS

Primarily, the Supervisory Control and Data Acquisition (SCADA) systems that are the major source of system-wide information never developed to perform the fault analysis [1]. In addition, nature of the data generated by SCADA system is asynchronous and has a very low update rate (tens of seconds to minutes), and is also highly corrupted by both Gaussian and non-Gaussian noise. Furthermore, 'oscillography' provided by Phasor Measurement Units (PMUs) and Digital Fault Recorders (DFR) with very high resolution, which enables the PMUs and DFRs to offer enough information to indicate that a fault is under development [2,3]. In addition, power systems reliability data usually contains a problem known as 'class imbalance'. In a power system, normal system operation conditions usually have a higher level of reliability than with the system during a fault condition, and therefore, will generate a much larger volume of data than fault data. Consequently, amount of normal operating data will be much larger than amount of fault data, and hence, will create an imbalance.

Most critically, current fault detection and protection mechanisms are limited by temporal latency. Standard fault detection algorithms used in conventional automation systems rely on Root Mean Square (RMS) or impedance estimation and require a full fundamental Alternating Current (AC) cycle (≈20 milliseconds (ms) at 50 Hertz (Hz)) to confirm a fault [4]. However, in the increasingly dynamic nature

of power infrastructures, the 20 ms temporal window may produce a sufficient probability for irreparable damage to critical infrastructure, such as larger motors/pumps at CW&PRS. In order to attain sub-cycle reaction speeds, a temporal framework that can avoid this full-cycle dependency is therefore desperately needed. Therefore, few pilot studies have been carried out using various artificial intelligence based deep learning models that were trained on normal operating condition data to detect/diagnose an electrical fault.

Most of the current research focuses on using high-frequency PMU feeds or laboratory-quality PMU data, which contain no noise in order to perform diagnostic analyses on the various deep learning models. During the research studies, an extensive analysis has been carried out based on the ability of analytical models to process low-frequency and noisy data. In addition, the research studies deliver insight into the current automation systems that may occur within the power utilities where the widespread utilization of PMU systems may not be a reality today. Comparison has been made between carious deep learning models and the raw performance of each model. To clarify the decision boundaries set by the algorithms, a feature importance metric has also been developed.

2. Literature Review

The power distribution grid is arguably the most sophisticated and extensive mechanism created by humans (Fig. 3). This is because power distribution grid includes a multitude of businesses essential for society to progress in the present day. However, at the present moment, topology of a power distribution grid continues to change [5,6]. It is in the process of being restructured from a mothership approach to a distributed multi-tiered grid, i.e. from a primarily thermal and hydroelectric plant based system to a more diverse grid that uses both traditional sources (renewable energy) and emerging technology [7]. The entire approach of transitioning to a new way of doing things relative to lowering greenhouse gas emissions on an international level relies on the success of these transformation projects [8–10]. The rapid changes occurring in these sectors, such as transportation and the growing number of renewable energy sources (solar and wind), are causing a significant burden on today's infrastructure and causing many of the physical challenges with fault failures to be extremely complicated compared to their earlier history.

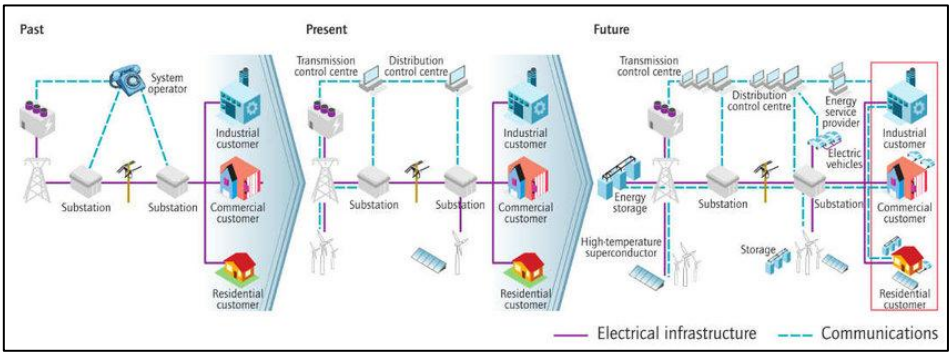


Fig. 3 Evolution electrical grid with distributed generation and bidirectional power flow [11]

Incorporation of Distributed Generation (DG) into the existing distribution network will dramatically alter the fault current characteristics of that system [12,13]. In conventional radial distribution networks, there is only one type of directionality of fault current flow from substation to the point of fault [14]. However, in a network with Distributed Generation sources, there are numerous sources of fault current, which makes it more challenging to determine the exact source of fault current [15,16]. Further, problems (Protection Blindness) caused by DGs are a direct result of complex and varied sources of fault current introduced by addition of DG to the network [17]. Specifically, level of fault current generated by DG can oversaturate the fault and sympathetic tripping [18,19]. Moreover, Impedance-Based Fault Location schemes currently in use are not applicable in networks with DGs. This is because the 'infeed' from DG introduces a non-linear relationship that contradicts principle of proportionality between fault location and distance [20]. Therefore, current study identifies the faults caused by addition

of DGs to distribution network and need for an effective deep learning-based classification technique that can accurately identify, classify, and diagnose the faults in noisy data environments.

With the aim of identifying fault precursors close to real-time or without the need for full-phase voltage estimation, this study proposes the idea of a sub-cycle based fault detection model based on raw data received directly from PMUs. Such a framework diminishes the response time significantly because the proposed approach reforms the conventional process of fault detection with the help of conventional RMS or Impedance Estimation Method with the help of a ten-sample size filtering process. This has a theoretical response time that comes close to 3.125 ms, acting like a digital reflex that comes in handy because of the highly dynamic and fluctuating environment. Such a practice creates a new avenue that pushes forward the concept of real-time fault detection infrastructures in the framework of Active Distribution Networks while providing a fundamental shift with respect to conventional fault detection philosophies [21,22].

2.1 Mathematical model based approaches

Recent studies have demonstrated the use of impedance based fault location algorithms on single source radial topologies [23]. The primary strength of these algorithms is the accuracy that is visible during extensive testing. Researchers have also indicated that 'protection blindness' and 'sympathetic tripping' are prevalent causes for failure when implemented with Distributed Generation (DG) on a topology [24]. DG impacts the distance to impedance relationship through modifications to current ratio and corresponding voltage ratios between relay and point of reference. While traveling wave techniques provide a high level of accuracy, concerns have arisen with the intricate, branched nature of distribution grids and the need for a sampling frequency in MHz or greater, which far exceed the sampling abilities of standard SCADA systems. Therefore, mathematical models dependent on physical data are inadequate to describe the non-linearities associated with smart grids today [25].

2.2 Signal based approaches

Earlier, practices such as Discrete Wavelet Transformation, Short-time Fourier Transformation, Continuous Wavelet Transformation, and Wavelet Packet Transformation have been exercised to create a link from the analysis of Raw Waveform signals to diagnostic algorithms, i.e. the extraction of fault signature characteristics from non-stationary fault signals through Time-Frequency Representation via a Time-Frequency Decomposition (albeit through multiple time-scale factor interpretations) [26,27]. Application of a Hybrid Wavelet Transform with Support Vector Machines (SVMs) is the best illustration. Similarly, Wavelet-Based Fuzzy Logic architectures (merging Interpretable Logic Rules with Signal Localization) have been demonstrated to classify Signals with nearly 100% accuracy in less than 1/2 cycle of a Power Frequency [28]. However, there are major impediments to utility distribution systems using these techniques at a scale. These impediments prevent electric utility distribution systems from implementing these analytical techniques on a large scale, like Limitations of the SCADA System, Noisy Signals, and Need for PMUs/DFRs to accurately detect fault precursors. However, at present, most PMUs/DFRs are not properly aligned for fault transitional detection within the electric utility distribution system, but rather are installed arbitrarily on different electric utility distribution system feeders [12].

2.3 Data-driven and machine learning approaches

The lack of performance of model-driven and signal-driven approaches has served as a catalyst for a predominant trend in AI and ML research. Instead of focusing on understanding the underlying physics, the trend is now to develop complex, nonlinear functions that map the behaviour of physical objects from operational data [29]. Traditional machine learning techniques like Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbour (kNN) are still valuable tools for many protective devices due to their low computational latency, interpretability, and robustness (Table-1).

Table-1 Classical machine learning Vs. Deep learning paradigms for power grid fault diagnosis

Aspect	*Classical ML*	*Deep Learning*
Feature extraction	Manual	Automatic
Data requirement	Low–Medium	High

Interpretability	High	Low
Computation	Low	High
Noise robustness	Moderate	High
Deployment readiness	High	Medium

For overhead line faults, SVMs achieve over 95% classification accuracy [30]. In addition to their robust high-dimensional feature set classification capabilities, SVMs can also be effective in binary classification tasks (fault/no fault) even when only small datasets are available. This is because SVMs have the ability to produce a hyperplane that separates the two classes as effectively as possible. Further, Random Forest and XGBoost algorithms became popular for classifying noisy, imbalanced datasets due to their robust performance [31]. Ensemble methods produce lower variance when levelled against individual classifiers on multi-class classification datasets [32].

3. Methodology

During the studies, it is understood that relying only on basic voltage/current magnitude is not adequate and the full spectrum of PMU data is required for deeper insights in fault analysis. Using advanced machine learning (i.e. Ensemble Learning) and Explainable AI (XAI) with rich PMU data (voltage, current, frequency, phase angle), transmission line faults can be accurately classified [32]. The primary data for this study was acquired from the Electrical Substation Automation and Control (ESAC) centre at the Central Water and Power Research Station (CW&PRS), Pune, India. Even though a complete repository of all relevant operational parameters was available from the ESAC centre, the initial dataset was found to lack structural organization, necessitating preprocessing to attain a systematic flow. A major challenge that has been identified with this dataset was a large class imbalance.

Due to the inherent characteristics of the power system, the number of instances related to faulty transmission lines was significantly high compared to non-faulty instances. This serves to illustrate the importance of special care taken in the algorithm to ensure the accuracy and reliability of the developed fault diagnosis model. The given dataset entails the use of high-resolution time series data that assumes primary importance with respect to the accurate analysis and monitoring of the power system in the presence of faults. The data includes instantaneous current (I_a, I_b, I_c) and voltage (V_a, V_b, V_c) with three-phase measurements, R, Y, and B, related to the state of the power system. Moreover, the dataset takes into consideration the real-time fluctuations in the measurement of system frequency, making use of binary values with 0 related to a healthy state and 1 related to a fault state.

In order to devise a strong end-to-end architecture that can efficiently operate in the dynamic real-world conditions of the electric grid available at CW&PRS, the proposed deep learning framework comprises a multi-stage analytical pipeline. This helps the model to access high-fidelity inputs and accurately achieve high accuracy levels in fault detection and diagnosis. Figure 4 shows the proposed deep learning frameworks that can accurately operate in real-world conditions of the electric grid available at CW&PRS.

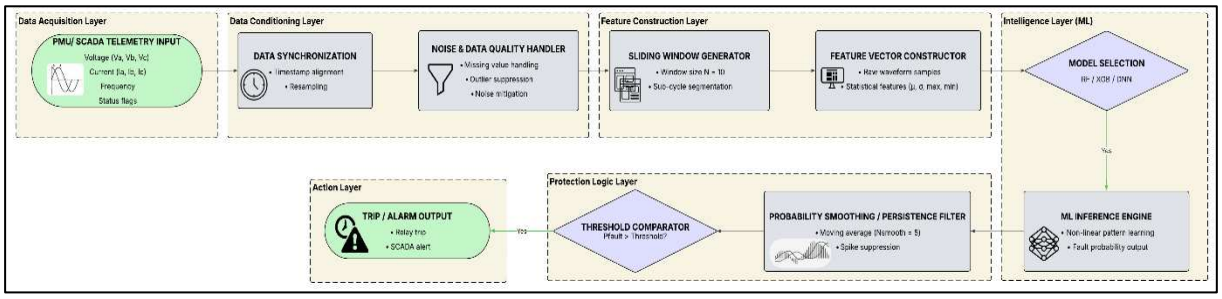


Fig. 4 Proposed end-to-end architecture for electrical power grid fault detection at CW&PRS

3.1 ETL Pipeline and Preprocessing

In order to make the raw data suitable for algorithmic use an Extract, Transform, and Load (ETL) pipeline was created to handle the high-velocity data streams from the ESAC Centre. Preprocessing

techniques are applied to synchronize asynchronous time-series signals and normalize the three-phase current and voltage magnitudes. Given the disorderly nature of the initial data collection, this stage has become critical for noise reduction and handling missing values, ensuring the deep learning model receives clean, consistent data that reflects true state of the grid. Moreover, preprocessing of the data was done using Python tools like NumPy, Pandas, and Scikit-Learn.

3.2 Exploratory Data Analysis (EDA)

Following the data acquisition from the ESAC Centre at CW&PRS, the behaviour of current and voltage waveforms was analysed across different conditions. Fig. 5 shows the transition between normal operating states and fault states for both R-Phase Current (I_a) and R-Phase Voltage (V_a). This represents the dynamic behaviour of the power system as it moves from a steady-state equilibrium to a disturbed state due to a fault occurrence. By plotting both Telemetry Parameters: Current and Voltage, this framework also shows the inverse relationship and characteristics used for Deep Learning classification. The comparative study between R-Phase Current and R-Phase Voltage can help identify normal operating signatures that transition from normal states to fault-induced states in the CW&PRS Electric Power System Network. During normal operating states, also referred to as Fault Type 0, both R-Phase Current and R-Phase Voltage have fixed and consistent values and patterns. During these states, the voltage always shows a strong match and maintained peak-to-peak normal values, indicating a relatively steady and normal Current pattern, which shows minor variations reflecting a balanced and stable system without fault circumstances. A transient fault response of Fault Type 1 enters a rapid and severe state change at around $t = 0.5\ s$, with a significant spike of current magnitude, I_a, from a nominal level of around ±7,500 A to values greater than ±30,000 A this acting as one of the main signatures for short-circuit behaviour.

At the same time, there is a large attenuation in voltage magnitude, which manifests itself as a severe sag or collapse. The decrease in peak voltages from ± 300 V to about ±100 V, or even less at most intense points, reflects system stress and the impedance changes intrinsically associated with the fault. Further, critical portions of signal instability and fluctuation over a range from 2 s to 4 s are captured in the data, directly reflecting what really happens in the real world, i.e. ‘disorderly’ nature of the data collection at the ESAC centre.

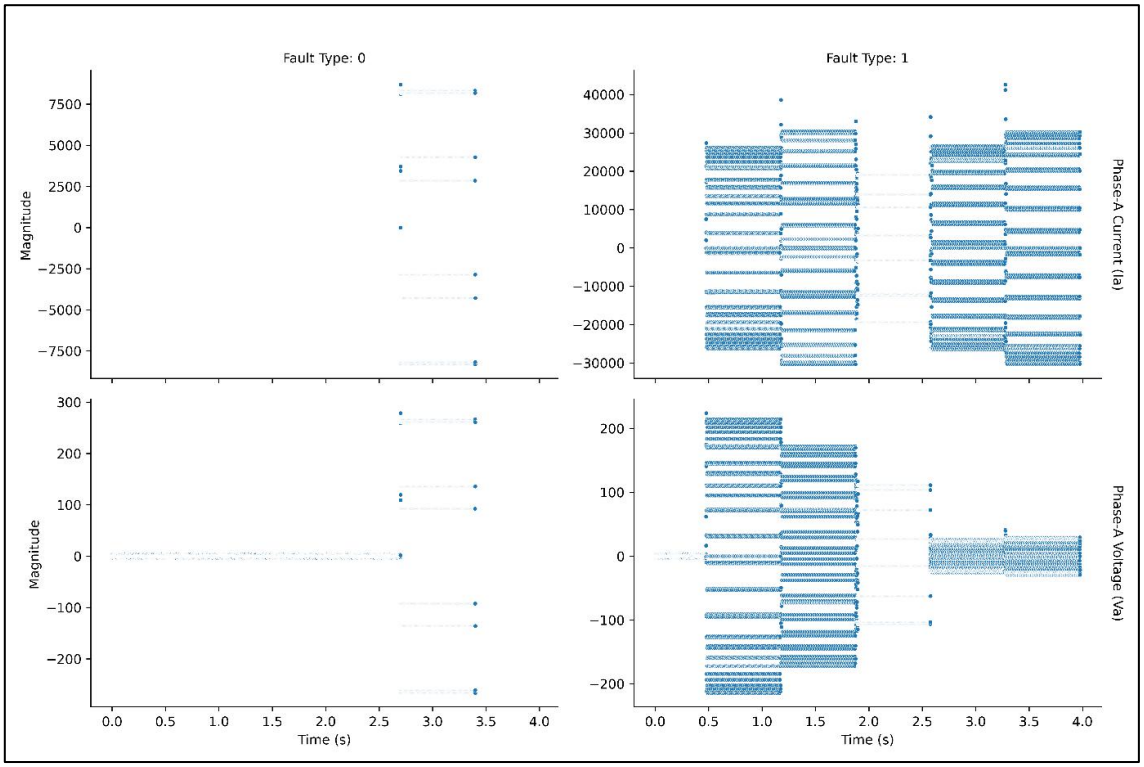


Fig. 5 Behaviour of R-phase current and voltage waveforms during steady-state and fault conditions

The key periods of fault discontinuity and recovery are seen to play an important role in facilitating this proposed framework for deep learning, as it ensures that accurate differentiation from external sensor noise takes place. This ensures that the model adopted remains effective even in non-ideal conditions, as experienced in high-voltage transmission lines. The evaluation has also enabled a clear comprehension regarding specific characteristics or exclusive patterns associated with different classes of faults. For every class of faults in the system, it has been calculated that the average, standard deviations, maximum, and minimum values pertaining to phase voltages as well as phase currents fall within Equations (1) to (4).

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2} \tag{2}$$

$$x_{max} = \max(x_1, \dots, x_N) \tag{3}$$

$$x_{min} = \min(x_1, \dots, x_N) \tag{4}$$

These generated statistical features and their intrinsic properties represent the feature set that machine learning algorithms will use in order to come up with predictive models.

3.3 Correlation Analysis

This need for correlation analysis arises because, in a multidimensional telemetry data, there is a necessity to find and quantify dependencies among data points so that the feature set for deep learning is efficient and mathematically appropriate. In this way, relationships will be examined in such detail that identification of redundant variables would take place, thus establishing priorities for those variables that are significant predictors of grid instability in the end-to-end architecture. Correlation heatmaps using the Seaborn library in Python were generated to graphically represent statistical associations across variables related to electricity. The correlation analysis results provide a quantitative estimation of the interdependency relationships existing between different parameters related to electricity signalling within the power grid dataset. The relationships between different numerical parameters such as the voltage levels in each phase, i.e. V_a, V_b, V_c, and current levels in each phase, i.e. I_a, I_b, and I_c, for each R, Y, or B phase, are visualized using a heatmap obtained using a Seaborn library. The relationships are visualized on a scale from -1 to 1. The correlation matrix is very important for indicating how a qualitative correlation can occur between two different parameters during a phase, i.e., positive or negative correlations. The heatmap shows how strong physical couplings occur due to relationships between other phases as well. For example, strong correlations are visible between voltage levels and current levels in some phases that indicate a high inverse correlation seen during fault transients in a system, i.e., when a spike in current levels corresponds to a sudden fall in voltage. Moreover, the almost negligible values of correlation observed for the primary electrical values and identifier 'source_id', 'Class_label', etc., in some of the sectors point towards the notion that though the individual values have a physical connection, the classification of fault occurs based on non-linear patterns. This grounding of statistics ensures that the engineered feature set remains optimized in terms of reducing redundancy and retaining the most influencing values in real-time fault diagnosis in the CW&PRS substation.

As shown in Figure 6, the heatmap provides a quantitative representation of these correlations on a scale from 1, representing a perfect positive correlation, to -1, representing a perfect negative correlation. This matrix provides a comprehensive insight into the interplay of various features of power grid on the operational environment of ESAC centre, CW&PRS, facilitating the extraction of prominent patterns, including the inverse relationship between phase current surges and voltage sags during fault conditions. The heatmap uses a color-coded scale where a coefficient of 1.00 (dark red) represents a perfect positive correlation, and negative values (blue) represent an inverse correlation. One of the most prominent features is the positive correlation between V_{a_Mag} and I_{a_Mag} (0.54), which indicates that under prevailing load conditions captured at ESAC centre, two R-phase components tend to increase proportionally

during a specific operational phase. However, prominent negative correlations are observed between various phases, including V_{b_Mag} and V_{c_Mag} (-0.67), which indicates interdependent and counteractive changes in phase magnitudes during unbalanced steady-state conditions of the system. The matrix also shows that 'Class_label', which marks the presence of a fault, has a near-zero linear correlation with raw magnitudes of the data. This is an important result for proposed deep learning architecture, as it shows that data cannot be classified using simple linear thresholds. Rather, it shows that complex, non-linear feature extraction and deep learning components of end-to-end architecture are necessary for fault classification. Finally, high correlation value of 0.68 between 'Class_label' and 'source_id' is indicative of particular data structure in ESAC centre repository, in which some sources of data are more strongly linked to the presence of fault events.

4. Results and Discussion

To assess the efficacy of proposed end-to-end solution, different machine learning and deep learning models were considered, such as Decision Trees, Logistic Regression, Support Vector Machines (SVM), Random Forest, XGBoost, and Artificial Neural Networks (ANN). The effectiveness of models was measured based on key parameters for an early warning system in an electric power utility, such as overall accuracy, precision, recall, F-1 score, and False Positives and False Negatives.

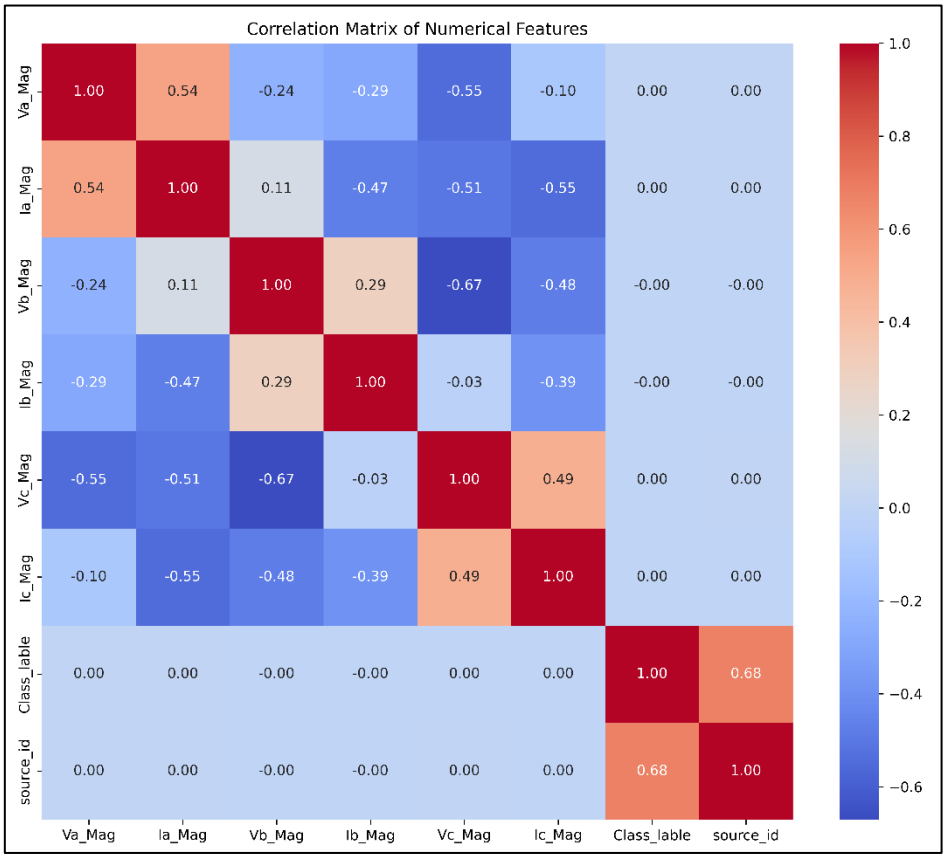


Fig. 6 Correlation matrix of grid telemetry features and fault classes

4.1 Comparative Analysis of Classifier Performance

As highlighted in Table 2, ensemble learning models (Random Forest) provided a better trade-off between fault sensitivity and reliability. Random Forest model resulted in an accuracy of 98.91% and a fault recall of 95.81% with zero false alarms. This is highly desirable for practical grid implementation, where false alarms can cause unnecessary operational expenses and actual faults may cause system collapse. Additionally, the combination of sliding window strategy was successful in modelling the fast transients during fault occurrence in high-resolution telemetry time-series data. Although, Decision Tree classifier resulted in a slightly higher fault recall of 97.51%, it resulted in 59 false alarms, suggesting a propensity to classify grid noise as faults. SVM (Linear) classifier performed well with zero false alarms

and an accuracy of 98.74%, although it had a higher computational complexity, which may be a drawback for ultra-fast real-time applications.

Table-2 Performance of Classifiers

Model	*Accuracy (%)*	*Fault Precision*	*Fault Recall*	*Fault F1-Score*	*False Alarms (FP)*	*Missed Faults (FN)*
XGBoost	100.00	1.000	1.000	1.000	0	0
Random Forest	98.91	1.000	0.958	0.979	0	91
SVM (Linear)	98.74	1.000	0.952	0.975	0	105
Decision Tree	98.64	0.973	0.975	0.974	59	54
ANN	96.28	1.000	0.860	0.925	0	316
Logistic Regression	76.04	0.530	0.726	0.613	1396	596

4.2 Linear vs. Non-Linear Modelling Challenges

There was a notable difference between the linear and non-linear modelling methods. Logistic Regression model was not effective, achieving only 76.04% accuracy with a high number of false alarms (1396). This is because the complexity of power system data is inherently high. The voltage-current feature space shows strongly non-linear, overlapping boundaries between normal and faulty regions, as shown in Fig. 7. Linear classifiers are mathematically incapable of capturing such complex decision boundaries, while non-linear classifiers are able to successfully traverse the complex overlaps of features. Despite the XGBoost Classifier achieving perfect results (100% on all metrics), there appears to be a possible case of overfitting. Perfect classification can be attributed to the synthetic data used in the initial training samples, possibly not fully capturing the stochastic 'disorderly' noise inherent in raw ESAC centre data.

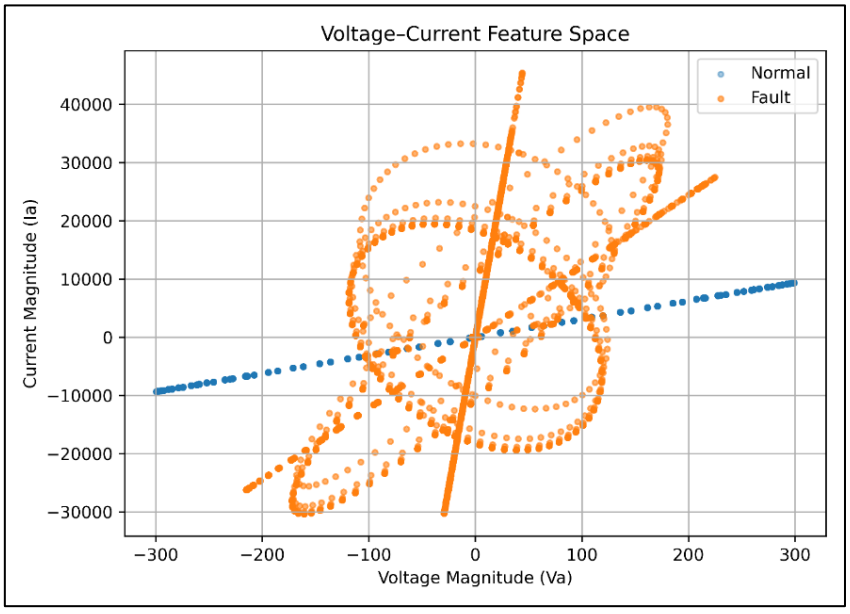


Fig. 7 Non-linear separation between normal and fault operating conditions in the voltage–current feature space, illustrating the limitations of linear classifiers

4.3 Neural Network and Optimal Model Selection

Artificial Neural Network (ANN) showed a high level of precision with no false positives. Nevertheless, the lower recall value of 86.05% shows a relative insensitivity to the initial stages of occurrence of faults compared to ensemble techniques. Considering the importance of early warning systems to identify faults as soon as they occur, Random Forest algorithm is identified as the most effective solution to be implemented in PMU based fault detection systems. This is because it offers required sensitivity to identify faults while maintaining the required stability for high-voltage grid monitoring.

4.4 Comprehensive Performance Validation via Confusion Matrices

Performance of the tested classifiers is quantitatively visualized using confusion matrices in Fig. 8, which give a detailed breakdown of the true positives, true negatives, false alarms, and missed detections. Among the tested classifiers, Random Forest and XGBoost classifiers show the best performance, with Random Forest classifier achieving an outstanding trade-off between accuracy and zero false alarms, correctly detecting 2080 instances of faults. The high accuracy of classifier is critical for the reliability of electric grid at ESAC centre, CW&PRS, as it ensures that protection systems are not activated by transient noise or non-fault disturbances. On the other hand, the Logistic Regression model has the highest degradation in performance because it fails to identify the non-linear boundaries of telemetry data set.

This is evident by its high-rate of 1396 false alarms and 596 missed faults, further proving that linear modelling is not sufficient for the complex and swirled interactions that are characteristic of voltage and current feature space during fault transients. Moreover, relative performance of other models, namely Decision Tree, SVM, and ANN, highlights trade-off between sensitivity and accuracy. While Decision Tree model is able to identify 2117 fault instances, its 59 false alarms indicate a vulnerability to misinterpretation of grid fluctuations. Conversely, while Artificial Neural Network (ANN) model has zero false alarms, its 316 missed faults indicate a relatively lower sensitivity to initial stages of fault development compared to ensemble models. In any case, aforementioned confusion matrices have been used to further validate the Random Forest model as the most accurate architecture for real-time early warning systems, where concurrent minimization of missed and false alarms is of utmost importance.

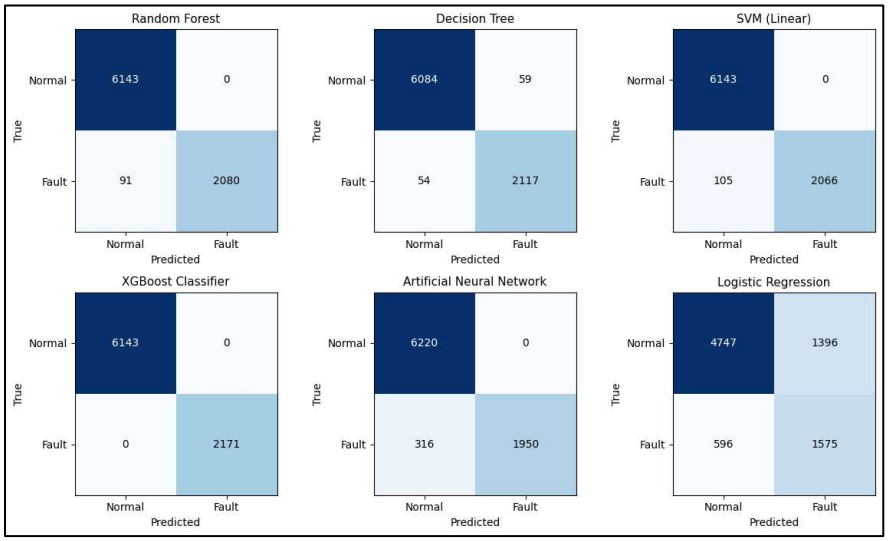


Fig. 8 Confusion matrices for the binary classification of power grid states at CW&PRS

4.5 Early Prediction Capabilities and 'Digital Reflex' Architecture

To improve the responsiveness of grid monitoring at the ESAC centre, CW&PRS, this research deviates from traditional data categorization approaches in favour of a solution that reflects the chaotic and real-world environment of an electrical grid. The aim is to develop an early warning system that can deliver information faster than conventional utility relays through 'digital reflex' response to rapid changes in vital power infrastructure. Real-world electrical faults happen randomly in time, not in a systematic 'label' at regular times. To incorporate randomness, an Artificial Stream Generator was developed to extend beyond conventional static data set, which often fails to properly reflect data sparsity and continuity in power system dynamics. By simulating a 'virtual timeline' that inserts punctuated fault events (LG, LL, LLL, and so on) with random intervals of steady-state Normal (No Fault (NF)) operation, the system is tasked to separate fast and subtle waveform interruptions from steady sinusoidal pattern of a correct pump function.

This 'stress testing' technique not only tests accuracy of individual classification but also detection-to-response time in a continuous data stream. Further, the technique uses a flattened input vector of 60 raw

telemetry data points, which describe three electrical phases (I_{abc}, V_{abc}) representing R, Y, and B phases at 10 different time points. This is a high-dimensional input, which allows calculation of a Fault Probability Score (P_{fault}). Moreover, this technique leverages autonomous feature extraction capability of Deep Neural Networks (DNNs), which allows the system to 'learn' how to detect jagged edges of a fault and magnitude differences between a fault along with a normal load change. In a conventional protection relay, a complete AC cycle (20 ms for a 50 Hz system) is required to calculate the RMS values of current and voltage before a trip can be initiated as described in Equations (5) and (6). Nonetheless, in Variable Frequency Drives (VFDs) or motors where windings are subjected to extreme stress, this 20 ms delay could be disastrous.

$$I_{rms} = \sqrt{\frac{1}{T}\int_0^T i^2(t)dt} \tag{5}$$

$$V_{rms} = \sqrt{\frac{1}{T}\int_0^T v^2(t)dt} \tag{6}$$

To address this, a Sliding Window Architecture was implemented to enable a reactive 'reflex' action, as illustrated in Fig. 9. Rather than waiting for a full cycle, the system utilizes a sub-cycle observation window (10 samples). As shown, the sliding window continuously moves across the PMU signal, when a fault inception occurs, the architecture identifies the anomaly within a fraction of a cycle. This sliding window approach allows the model to detect the 'onset' of a fault (i.e. the moment waveform deviates from its sinusoidal path), enabling a trip response much faster than traditional calculation-based methods. This sub-cycle granularity is essential for protecting sensitive equipment at the CW&PRS facility from high-magnitude transient damage.

By restricting the window size (N=10), the framework treats fault detection as a shape recognition task rather than a traditional phasor estimation task. The sliding window, defined by Equation (7), allows the model to identify when the trajectory of the waveform deviates from the normal sinusoidal curve (e.g., a steep non-linear slope) without waiting for a full cycle of data. As shown in Figure 9, the response time of proposed system is inherently time-related.

$$x_t = [I_a(t-k), I_b(t-k), I_c(t-k), V_a(t-k), V_b(t-k), V_c(t-k)]_{k=0}^{N-1} \tag{7}$$

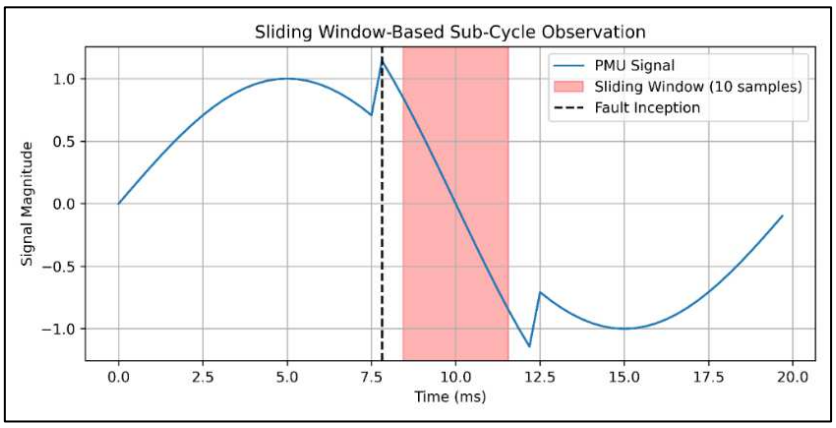


Fig. 9 Sub-cycle fault detection using a sliding window architecture for real-time PMU signal monitoring

After the occurrence of a fault event, probability of detection jumps dramatically within a window of only 1 to 2 sample periods, indicating capability to detect the fault within a few microseconds of its occurrence.

4.6 The Latency: Sub-Cycle Response Speed

This design leads to a theoretical value of detection latency ($t_{latency}$) that is significantly faster than that of traditional impedance relays. With a limited observation window of N=10, expected value of the latency is approximately 3.125 ms, as follows in Equation (8),

$$t_{latency} = N \times \Delta t \tag{8}$$

This sub-cycle response offers a vital safety margin for the sensitive grid infrastructure at ESAC centre, CW&PRS, to ensure protection is activated well before the implication of a standard 20 ms cycle.

4.7 Temporal Analysis and System Response Performance

The temporal efficacy of the proposed early warning system is shown in Fig. 10, where a real-time verification of 'digital reflex' design is achieved by correlating the model's transient response to physical fault injections. As shown, model's prediction probability (denoted by red/black line) remains at a stable baseline of approximately 0.0 until the occurrence of a physical fault, denoted by the black step function. At this point, the prediction probability output (P_{fault}) increases to a value above 0.9 almost instantaneously, specifically within 1-2 time steps of the fault injection. This indicates that system is capable of detecting the onset of grid instability within the sub-cycle timescale, thus providing a critical temporal advantage over traditional methods that are limited by full-cycle RMS calculations. By formulating fault detection problem as a shape recognition task, the model is able to detect deviation from normal sinusoidal trajectory without requiring integration of data over a full 20 ms cycle. This is a critical capability for high-stress assets such as VFDs and motors, where immediate detection is required to prevent transient damage.

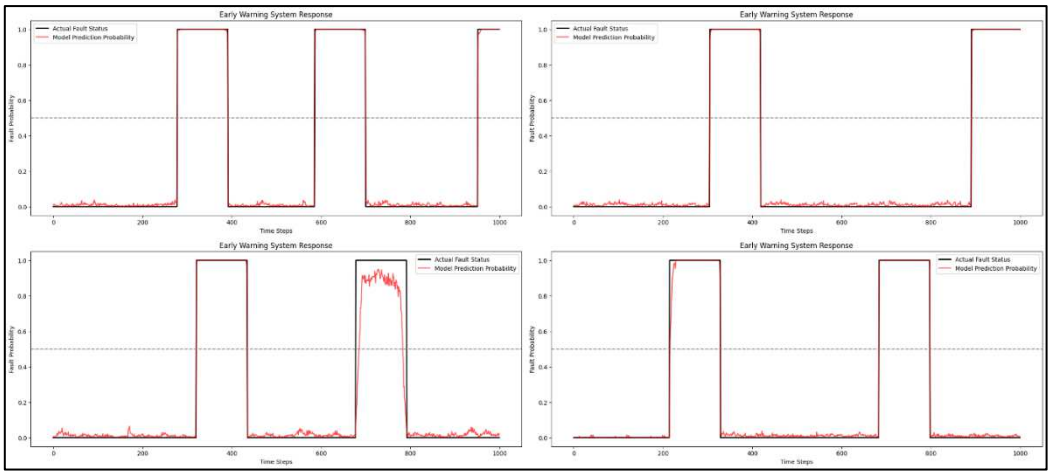


Fig. 10 Temporal response of the early warning system for real-time validation of fault probability scores (P_{fault}) illustrating sub-cycle detection against physical fault injection timestamps

4.8 "Safety Filter" (Post-Processing Logic)

One of the difficulties in fast fault detection is trade-off between sensitivity and noise immunity. In the early stages of development, model sometimes produced 'nervous' responses, brief peaks in fault probability ($P_{fault} > 0.8$) that lasted only 1-2 samples. These artifacts, due to stochastic noise in the data stream, could cause spurious system shutdowns unless removed. To suppress these false alarms, a Persistence Filter with a moving average was introduced, as specified in Equation (9).

$$\overline{P_{fault}}(t) = \frac{1}{N_s} \sum_{i=0}^{N_s-1} P_{fault}(t-i) \tag{9}$$

With the setting of $N_{smooth} = 5$, the system smoothed out the isolated noise while preserving the sharp increase in probability due to the actual fault occurrence. Further, protection logic, acting as a 'sanity check' for the grid, is expressed in Equation (10).

$$\text{Alarm} = \begin{cases} 1, \overline{P_{\text{fault}}} > 0.5 \\ 0, \text{otherwise} \end{cases} \quad (10)$$

The effectiveness of this post-processing logic is shown in Fig. 11, where raw model output is contrasted with smoothed and trustworthy output. Further, raw output (red curve) shows some fluctuations and spikes while the system is running normally. In addition, the smoothed output (blue curve) is held down during noise events but follows true fault status (black curve) very closely once a stable pattern is established. This logic also ensures that alarm is raised only if model has been indicating a fault pattern for a period of about 1.5 ms. Thus, the system, through this logic, ignores noise spikes and maintains continuity of operation for water resources at ESAC centre, CW&PRS. Even with this delay due to filtering, total response time is less than 3.125 ms, which is much faster than 20 ms response time of conventional protective systems, which become aware of faults only after they have been fully developed. This 'digital reflex' fills the gap between randomness of real-world grid dynamics and structured representation of diagnostic data.

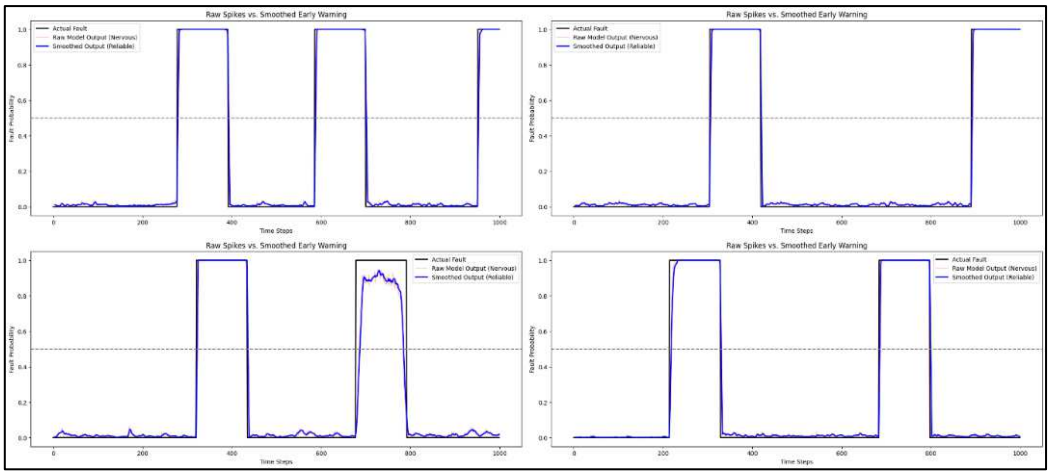


Fig. 11 Effect of probability smoothing in suppressing transient false positives to ensure operational reliability

5. Conclusions

This work has thus provided a strong, intelligent framework for real-time fault prediction and classification in smart transmission networks, designed specifically for the stochastic operational conditions of the ESAC centre, CW&PRS. Through a comprehensive assessment of a wide variety of machine learning models, this work has concluded that the capacity to handle non-linear data patterns is the decisive criterion for fault diagnosis in noisy grid conditions. Comparison results of various Deep Neural Network (DNN) models showed a large performance gap, while linear models such as Logistic Regression were inadequate for handling complex fault patterns and achieved only 76.04% accuracy. Moreover, Random Forest model performed outstandingly with a maximum accuracy of 98.91%. This clearly supports that ensemble learning can efficiently fill the gap between normal load operation and severe faults such as LG, LL, and LLL. Although conventional electro-mechanical and numerical relays take a complete AC cycle (approx. 20 ms) to compute RMS values for fault selection, the proposed system employs a 10 sample sliding window structure to detect grid instability in a sub-cycle time of only 3.125 ms. This 84% reduction in detection latency represents a transformative advancement over standard impedance based protection, enabling isolation of transients before they result in permanent thermal damage to sensitive infrastructure like Variable Frequency Drives and oversized motors.

Acknowledgements

The authors would like to thank Dr. Prabhat Chandra, Director, and Shri T.K. Swain, Scientist 'D' (Electrical), CW&PRS, Government of India, and Dr. Rajesh Prasad, Principal, Bharati Vidyapeeth Deemed to be University, Pune, Maharashtra, who has provided a greater support during the research studies.

Competing Interests
The authors have no relevant financial or non-financial interests to disclose.

Conflicts of Interests
The authors have no conflicts of interest to declare.

Funding
The authors gratefully acknowledge the financial support provided by Bharati Vidyapeeth Deemed to be University, Pune-411043, Maharashtra, India and Central Water and Power Research Station,

References
1. M. P. Ward, An Architectural Framework for Describing Supervisory Control and Data Acquisition (SCADA) Systems, PhD Thesis, Monterey, CA; Naval Postgraduate School, 2004.

2. J. D. Follum, R. Hovsapian, N. M. Stenvig, Y. Agalgaonkar, K. Chatterjee, K. Mahapatra, A. Riepnieks, A. J. Wilson, S. Chanda, and S. Granda, *Advanced Measurements for Resilient Integration of Inverter-Based Resources: PROGRESS MATRIX Year-1 Report* (Pacific Northwest National Laboratory (PNNL), Richland, WA (United States), 2023).

3. J. D. Follum, E. J. Ellwein, P. V. Etingov, X. Fan, H. Kirkham, L. E. Miller, and A. Riepnieks, *Advanced Power Systems Measurements: A Literature Review* (Pacific Northwest National Laboratory (PNNL), Richland, WA (United States), 2020).

4. J. Arrillaga, N. R. Watson, and S. Chen, *Power System Quality Assessment* (John Wiley & Sons, 2000).

5. T. Gönen, C.-W. Ten, and A. Mehrizi-Sani, *Electric Power Distribution Engineering* (CRC press, 2024).

6. M. Amin and J. Stringer, MRS Bull. **33**, 399 (2008).

7. A. M. Amani and M. Jalili, IEEE Access **9**, 119010 (2021).

8. N. R. Council, G. Affairs, P. Division, and B. on S. Development, *Our Common Journey: A Transition toward Sustainability* (National Academies Press, 1999).

9. N. R. Council, D. on Earth, L. Studies, B. on A. Sciences, and C. on M. for E. G. G. Emissions, *Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements* (National Academies Press, 2010).

10. A. Rojey, *Energy and Climate: How to Achieve a Successful Energy Transition* (John Wiley & Sons, 2009).

11. G. Prinsloo, R. Dobson, and A. Mammoli, Curr. Altern. Energy **2**, 2 (2018).

12. M. J. B. B. Davi, M. Oleskovicz, and F. V. Lopes, IEEE Access **12**, 179516 (2024).

13. S.-E. Razavi, E. Rahimi, M. S. Javadi, A. E. Nezhad, M. Lotfi, M. Shafie-khah, and J. P. Catalão, Renew. Sustain. Energy Rev. **105**, 157 (2019).

14. N. Cho, S. Yun, and J. Jung, Renew. Sustain. Energy Rev. **109**, 102 (2019).

15. A. M. El-Zonkoly, Electr. Power Syst. Res. **81**, 1482 (2011).

16. S. S. Gururajapathy, H. Mokhlis, and H. A. Illias, Renew. Sustain. Energy Rev. **74**, 949 (2017).

17. A. A. Majeed, A. S. Altaie, M. Abderrahim, and A. Alkhazraji, Energies **16**, 7587 (2023).

18. B. O. Agili, M. M. A. AL-Ziz, and H. K. Yossif, in *2008 IEEEPES Transm. Distrib. Conf. Expo.* (IEEE, 2008), pp. 1–14.

19. H. Sabra, D. K. Ibrahim, and M. Gilany, J. Eng. **2018**, 1181 (2018).

20. K. Jia, T. Bi, Z. Ren, D. W. Thomas, and M. Sumner, IEEE Trans. Smart Grid **9**, 807 (2016).

21. D. Leite, E. Andrade, D. Rativa, and A. M. Maciel, Sensors **25**, 60 (2024).

22. A. Haleem Medattil Ibrahim, S. K. Sadanandan, T. Ghaoud, V. Subramaniam Rajkumar, and M. Sharma, IEEE Access **12**, 112822 (2024).

23. S. Das, S. Santoso, A. Gaikwad, and M. Patel, IEEE Access **2**, 537 (2014).

24. J. P. Desai, J. Inst. Eng. India Ser. B **105**, 297 (2024).

25. D. K. Panda, Data Analytics and Control for Stability and Resilience of Future Smart Grids, University of Exeter, 2022.

26. C. M. Akujuobi, *Wavelets and Wavelet Transform Systems and Their Applications: A Digital Signal Processing Approach* (Springer International Publishing, Cham, 2022).

27. M. S. Priyadarshini, M. Bajaj, L. Prokop, and M. Berhanu, Sci. Rep. **14**, 3443 (2024).

28. O. A. S. Youssef, IEEE Trans. Power Deliv. **19**, 582 (2004).

29. B. Cai, L. Huang, and M. Xie, IEEE Trans. Ind. Inform. **13**, 2227 (2017).

30. S. F. Nasim, S. Aziz, A. Qaiser, U. Kulsoom, and S. Ahmed, Sir Syed Univ. Res. J. Eng. Technol. **14**, 27 (2024).

31. T.-T.-H. Le, Y. E. Oktian, and H. Kim, Sustainability **14**, 8707 (2022).

32. A. Ayodele, World J. Adv. Res. Rev. **19**, 1633 (2023).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SourceCodeandData.zip