
DEVELOPMENT OF FAULT DIAGNOSIS FRAMEWORK FOR 132 KV CAPACITOR VOLTAGE TRANSFORMER USING MACHINE LEARNING FRAMEWORK

Moses Uchendu^{1,*}

¹Ahmadu Bello University Zaria

Email: moseslife500@gmail.com*

*Corresponding author: Moses Uchendu

DOI: <https://doi.org/10.67203/sp.ICESID.2026>

ABSTRACT

Capacitor Voltage Transformers (CVTs) play a critical role in high-voltage power systems by providing voltage measurements for protection, metering, and monitoring applications. However, their performance can be adversely affected by faults such as capacitor drift, ferroresonance, overvoltage, undervoltage, and insulation degradation, leading to inaccurate measurements and reduced system reliability. The development of intelligent fault diagnosis systems for CVTs is often constrained by the limited availability of comprehensive fault datasets and the severe class imbalance associated with rare fault events. This study proposes an integrated framework for fault diagnosis and reliability assessment of a 132 kV CVT using synthetic data generation and machine learning techniques. A high-fidelity CVT model was developed to simulate both normal and fault operating conditions while incorporating temperature variations, harmonic distortion, and measurement uncertainties. A dataset comprising 20,000 samples was generated, containing electrical and operational features including primary voltage, secondary voltage, ratio error, phase error, temperature, and total harmonic distortion (THD). The dataset was subsequently used to develop and evaluate Decision Tree, Support Vector Machine (SVM), k-Nearest Neighbor (kNN), and Ensemble Machine Learning classifiers for multi-class fault prediction. Comparative results demonstrated that the Ensemble classifier achieved superior performance with an accuracy of 99.98%, macro precision of 1.00, macro recall of 0.999, and macro F1-score of 0.999, significantly outperforming the conventional machine learning models. The proposed framework provides an effective solution for accurate fault prediction, improved fault representation, and enhanced condition-based assessment of CVTs, thereby supporting predictive maintenance and improving the operational efficiency of high-voltage power system assets.

Keywords: Capacitor Voltage Transformer; System Modeling; Machine Learning; Fault Prediction

1. INTRODUCTION

The study is on the development of a fault diagnostic framework for 132 KV capacitor voltage transformer. With the increasing expansion of power systems, evaluating the reliability of the systems is imperative for improved and sustained performance efficiency. For metering, protection, and measuring equipment, the input signal needed to perform their protection functions depends on the instrument transformer. A capacitor voltage transformer is used as an instrument or transformer to step down high transmission line voltages (66 kV and above) to the nominal voltage levels needed for the operation of metering, protection, and measuring equipment (Haghgoo & Damchi, 2022). Faults in CVTs arise from aging and system subjection to abnormal operating conditions, which could lead to malfunction in system operation and consequently to protection systems. Therefore, an effective fault modeling and classification framework is needed to evaluate the operational behavior of the CVT under normal and faulty conditions and classify faults in CVTs, thus providing an early preventive maintenance approach to ensure the safe and reliable operation of CVTs and protection systems. Traditional fault detection techniques often rely on predetermined thresholds and expert knowledge; however, these methods can be inadequate for addressing the increasing complexity and dynamic nature of modern power systems Arouche Freire *et al.*, (2019). Machine learning (ML) techniques have shown great promise in enhancing fault diagnosis and classification (FDC) in power and protection systems. ML models can automatically learn the underlying patterns and characteristics of various fault types

without the need for explicit programming by leveraging vast amounts of real-time sensor data and historical fault records (Sappati & Bhadra, 2018). Despite these advantage, ML techniques often exhibit poor performance when dealing with imbalanced fault data and do not provide a direct linkage between fault predictions. The growing adoption of ML in fault detection on transmission lines offers several advantages, including higher accuracy, the ability to detect faults under various operational conditions, and the potential for real-time detection and classification.

Aim of Study: The aim of the study is to develop a condition -aware simulation-driven fault diagnosis framework for 132 kV capacitor voltage transformer

Objectives of Study:

1. To develop a high-fidelity modeling framework for a 132 kV capacitor voltage transformer (CVT) capable of generating realistic operational and fault-condition datasets for advanced fault detection and diagnostic applications
2. To develop an ensemble machine learning-based fault prediction model using the generated 132 kV dataset and comparatively evaluate its performance against conventional machine learning algorithms, including Decision Tree, Support Vector Machine (SVM), and k-Nearest Neighbor (kNN), using accuracy, precision, recall, and F1-score as performance metrics.

2. METHODOLOGY

In this section, the materials required for actualizing the research work and the detailed step by step methods to be followed in achieving the set objectives is presented.

2.1 Materials

The hardware and software materials to be deployed in actualizing this research work are presented in this section. These materials include:

1. Personal Computer
2. MATLAB Software (2026a)

2.2 Research Methodology

This section presents the methodological framework adopted for the fault diagnosis and reliability assessment of the Capacitor Voltage Transformer (CVT). This chapter outlines the systematic procedures used to model, analyze, and evaluate the operational behavior of the CVT under both normal and faulty conditions. It describes the data acquisition process, feature extraction techniques, and the implementation of diagnostic algorithms employed to detect and classify faults. The parameters used in this study were obtained from standard power system theory, typical CVT design practices, and reasonable engineering approximations. Where manufacturer data was unavailable, values were selected to ensure physically meaningful and realistic system behaviour. Environmental effects, such as temperature variation and harmonic distortion, were incorporated using statistically distributed variables to reflect real-world operating conditions.

2.3 System Description

The CVT considered in this study is rated at 132 kV on the primary side (V_p) with a nominal secondary output of 110 V. The phase voltage is shown in equation (1):

$$V_p = \frac{V_{line}}{\sqrt{3}} \quad (1)$$

The capacitive divider consists of two series capacitors with nominal values of C_1 and C_2 to be 8000 pF and 72000 pF respectively. These parameters establish the nominal transformation ratio and form the basis for the CVT mathematical model as shown in Table 1.

Table 1: CVT Parameters

Parameter	Value	Unit
V_{line_nom}	132,000	V
V_{P_nom}	76,210	V
V_{Sec_nom}	110	V
C_{1_nom}	8000e-12	pF
C_{2_nom}	72000e-12	pF
$Temp_{Coeff}$	-0.00035	1/°C
$Temp_{ref}$	25	°C

2.2.1 Temperature Effect Modelling

The influence of temperature on capacitor characteristics is incorporated using a linear temperature coefficient model as shown in equation (3).

$$C(T) = C_{nom} [1 + \alpha(T - T_{ref})] \quad (3)$$

Where $\alpha = -0.00035$ and $T_{ref} = 25^\circ C$. This enables the simulation of environmental effects on CVT performance, particularly under varying field conditions.

2.2.2 Secondary Voltage Computation and Normalization

The secondary voltage is computed by incorporating the divider ratio, temperature effects, and system normalization as equation (4).

$$V_s = (V_p \times Divider\ Ratio) \times \frac{V_{sec,nom}}{V_{p,nom} \times \left(\frac{C_1}{C_1 + C_2} \right)} \quad (4)$$

This formulation ensures that the modelled CVT output aligns with the expected nominal transformation characteristics while allowing deviations under fault conditions.

2.2.3 Noise and Measurement Uncertainty Modelling

To emulate real-world measurement conditions, stochastic noise is introduced into the secondary voltage signal. Gaussian noise is applied with a small variance proportional to the signal magnitude, representing sensor inaccuracies, electromagnetic interference, and switching disturbances. This enhances the realism of the generated dataset and improves the robustness of subsequent fault diagnosis models.

2.3 Fault Modelling and Classification Scheme

A comprehensive fault modelling framework is developed by introducing parameter perturbations corresponding to different CVT fault conditions. Seven operating states are considered like Healthy condition, C_1 capacitance drift, C_2 capacitance drift, Ferroresonance condition, Overvoltage condition, Undervoltage condition and Insulation degradation. Capacitor drift faults are simulated by modifying the capacitance values, while ferroresonance is modelled through amplitude amplification and phase deviation. Voltage-related faults are introduced by scaling the primary voltage, and insulation degradation is represented through combined parameter variations and increased harmonic distortion. This classification provides a diverse dataset for fault diagnosis.

2.4 Ensemble Machine Learning Fault Prediction

1. Decision Tree (DT): Utilized for its interpretability and ability to model non-linear feature boundaries (Meng et al., 2023).
2. Support Vector Machine (SVM): Employed with kernel tricks to map complex fault features into higher-dimensional spaces (Omar et al., 2022).
3. k-Nearest Neighbors (kNN): Used as a distance-based baseline to highlight the challenges of high-dimensional imbalanced data.

2.4.1 Performance Evaluation Metrics

Because standard accuracy is misleading in highly imbalanced datasets (where a model predicting "Healthy" 100% of the time would achieve 90% accuracy), the models are rigorously evaluated using Macro-Averaged Metrics:

- i. Macro Precision, Recall, and F1-Score: These metrics treat all fault classes equally, penalizing models that fail to detect rare faults.
- ii. Confusion Matrix Analysis: Utilized to identify specific inter-class confusions (e.g., distinguishing C1 drift from Insulation Degradation) and validate the ensemble's ability to isolate distinct fault signatures like Ferroresonance and Overvoltage.

3. RESULTS AND DISCUSSION

Five representative waveform scenarios are generated such as Normal operation, C_1 capacitor drift, C_2 capacitor drift, Ferroresonance condition and Insulation degradation. Each waveform exhibits unique characteristics, enabling clear differentiation between fault conditions and supporting feature extraction for diagnostic purposes. Furthermore, the Root Mean Square (RMS) value of each waveform is computed to quantify signal magnitude and energy content.

3.1 Capacitor Voltage Transformer Under Different Condition

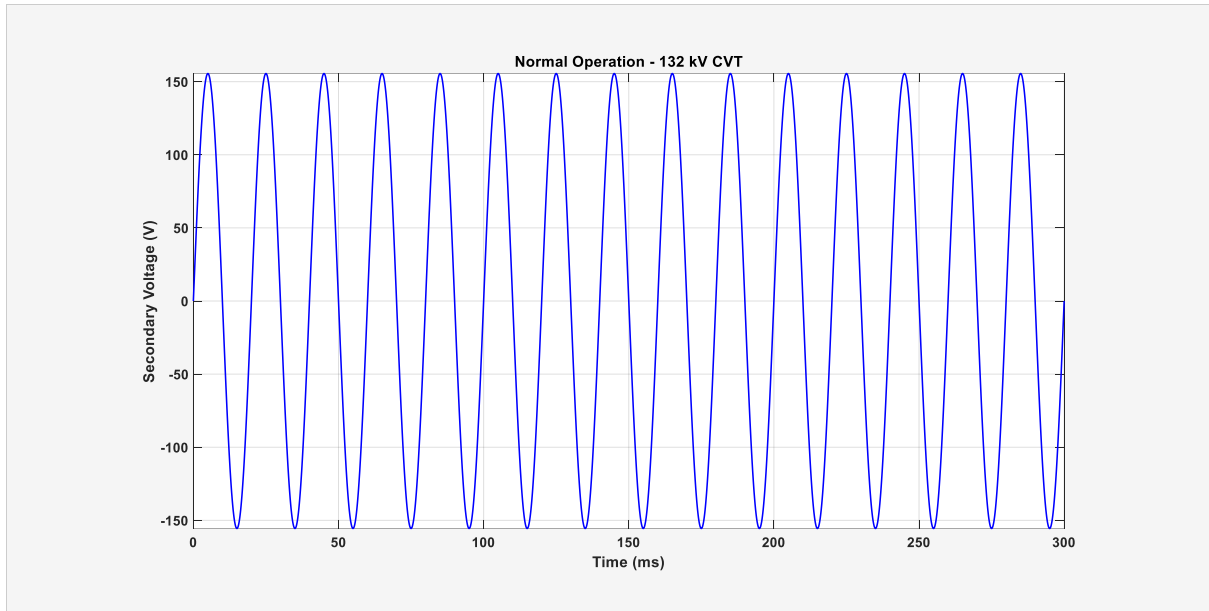


Figure 1: Normal operation

Figure 1 illustrates the time-domain waveform of the secondary voltage output of the 132 kV Capacitor Voltage Transformer (CVT) under normal operating conditions. The waveform exhibits a clean and stable sinusoidal profile, which is characteristic of a properly functioning CVT operating at the fundamental power system frequency of 50 Hz. The peak amplitude of the waveform is ± 155 V, corresponding to an RMS value close to the nominal secondary voltage of 110 V. This indicates that the CVT is accurately stepping down the high primary voltage (132 kV system) to a standardized measurable level, maintaining the expected transformation ratio. under the following condition $C_1 = C_1^{nom}$ $C_2 = C_2^{nom}$, Harmonics: none ($k_h = 0$) and Ferroresonance: none ($k_f = 0$).

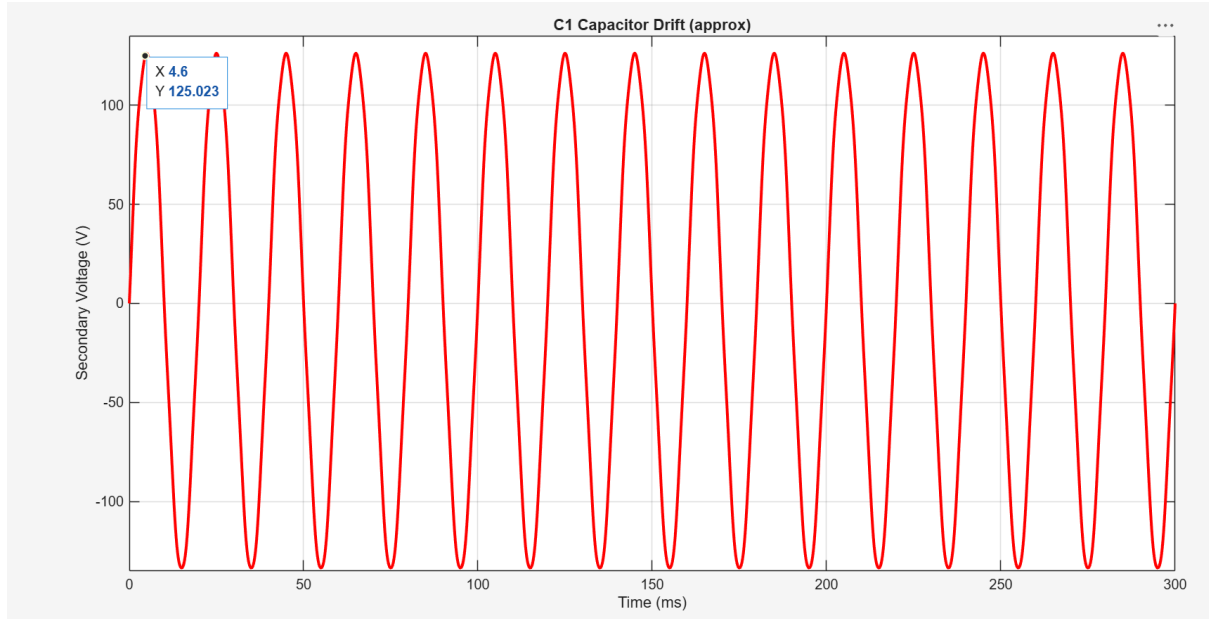


Figure 2: C_1 capacitor drift

Figure 2 presents the secondary voltage waveform of the 132 kV CVT under a fault condition characterized by an 18% reduction in the capacitance of C_1 . This condition simulates capacitor aging, dielectric degradation, or partial failure within the capacitive divider network. The waveform retains its fundamental sinusoidal shape at 50 Hz, indicating that the system frequency remains unaffected. However, a noticeable reduction in the amplitude of the secondary voltage is observed when compared to the normal operating condition. This behavior is attributed to the altered capacitive divider ratio, where a decrease in C_1 reduces its contribution to the voltage division process, under the following conditions, $C_1 = 0.82 \times C_1^{nom}$, $C_2 = C_2^{nom}$ and Moderate harmonics ($k_h \approx 0.4$).

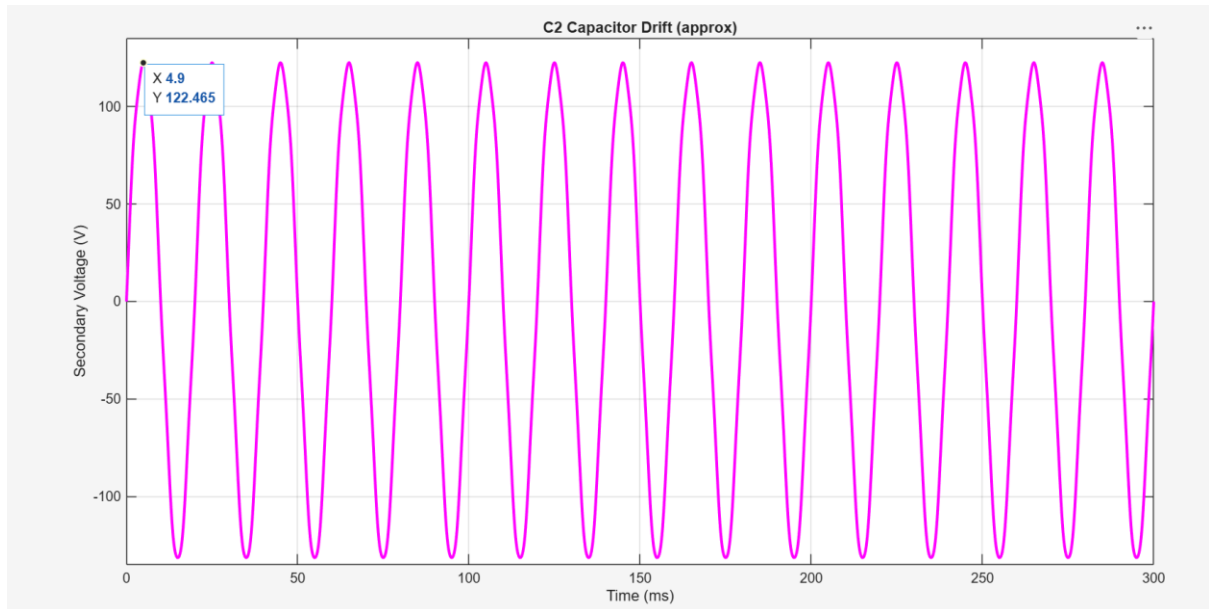


Figure 3: C_2 capacitor drift

Figure 3 illustrates the CVT secondary voltage waveform under a fault condition involving drift in the capacitance of C_2 . In this case, C_2 is increased relative to its nominal value, representing insulation changes, dielectric swelling, or manufacturing tolerances. Similar to Figure 2, the waveform maintains a sinusoidal shape at the fundamental frequency, confirming that no oscillatory instability or resonance is introduced. However, the amplitude of the waveform is altered due to the change in the capacitive

divider ratio, following this condition, $C_1 = C_1^{nom}$, $C_2 = 1.25 \times C_2^{nom}$ and moderate-to-high harmonics ($k_h \approx 0.5$).

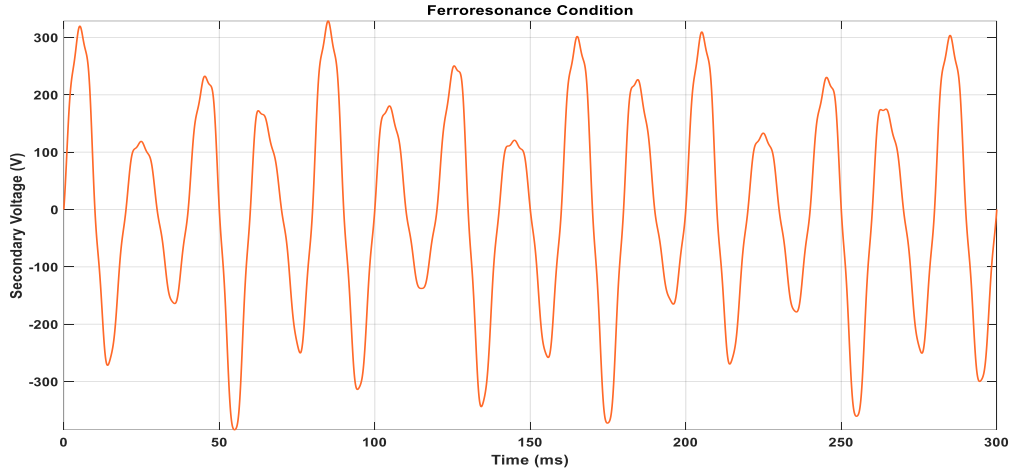


Figure 4: Ferroresonance condition

Figure 4 illustrates the secondary voltage waveform of the CVT under ferroresonance conditions. Ferroresonance is a nonlinear phenomenon that occurs due to the interaction between the CVT's capacitive elements and the nonlinear inductance of the electromagnetic unit (transformer core). This interaction leads to unstable oscillations, which are evident in the waveform as: severe deviation from a pure sinusoidal shape, irregular peaks and troughs with varying amplitudes, presence of subharmonic and harmonic components, and Intermittent oscillatory behavior. The waveform shows abnormally high voltage magnitudes, with peaks exceeding ± 300 V, which is significantly higher than the nominal secondary voltage. This indicates overvoltage conditions that can stress insulation and damage connected protection and metering equipment. Additionally, the oscillations are non-periodic, reflecting the chaotic nature of ferroresonance, following this condition, $C_1 = C_1^{nom}$, $C_2 = C_2^{nom}$, amplitude increased by 1.55, strong harmonics ($k_h \approx 1.1$) and strong ferroresonance term ($k_f \approx 1.2$).

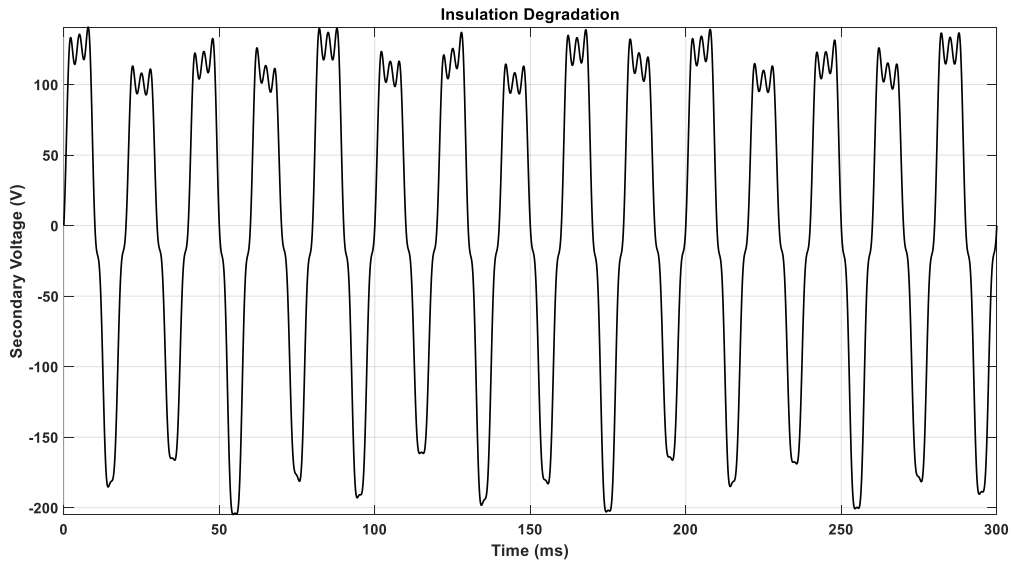


Figure 5: Insulation degradation

Figure 5 presents the secondary voltage waveform under insulation degradation conditions within the CVT. This fault is typically caused by aging, moisture ingress, thermal stress, or dielectric breakdown within the capacitive elements or internal insulation layers. The amplitude of the waveform shows moderate variation, but not as extreme as in ferroresonance. However, the presence of high-

frequency components suggests an increase in THD, which is consistent with insulation degradation effects. This distortion arises because insulation degradation alters the effective capacitance and introduces leakage paths, resulting in nonlinear voltage distribution across the capacitive divider increased dielectric losses, and degraded signal quality. Consequently, this condition leads to increased ratio error, noticeable phase displacement, reduced measurement accuracy and gradual reliability deterioration. Figure 5 represents early-stage faults such as insulation aging or partial discharge. The parameter used are $C_2 = 0.92 \times C_2^{nom}$, Slight amplitude reduction, high harmonic content ($k_h \approx 2.8$) and Mild ferroresonance ($k_f \approx 0.3$)

3.2 Fault Classification in CVTs Using Machine Learning

This section outlines the core methodologies, data pre-processing steps, algorithm selection criteria, and validation metrics used in ML-driven fault prediction. By highlighting how these models detect early degradation patterns, the discussion aims to provide a practical foundation for building, evaluating, and deploying reliable predictive systems that reduce downtime, lower maintenance costs, and enhance operational safety. Although the original dataset consisted of 20000 samples/ observations, the confusion matrix was generated using an independent test subset only. A 70:30 training-testing partition was applied, resulting in approximately 6000 unseen samples used for the model evaluation. The confusion matrix thus represents the classification performance on the test dataset rather than the complete dataset.

Decision Tree Confusion Matrix

True Class \ Predicted Class	C1 _{Drift}	C2 _{Drift}	Ferroresonance	Healthy	Insulation _{Degradation}	Overvoltage	Undervoltage
C1 _{Drift}	125				2		
C2 _{Drift}		131		1			
Ferroresonance			189				
Healthy				5379	3		
Insulation _{Degradation}	1				58		
Overvoltage						56	
Undervoltage							55

Figure 6: Confusion Matrix of the Decision Tree Model for CVT Fault Classification

The confusion matrix for the Decision Tree classifier in Figure 6 demonstrates robust fault prediction performance across a highly imbalanced dataset, with the majority "Healthy" class (5,379 true positives) and minority fault classes (ranging from 55 to 189 samples) generally well-discriminated along the diagonal. Minor misclassifications were observed: two instances of *C1_Drift* were predicted as *Insulation Degradation* and one vice versa, suggesting overlapping feature representations between these degradation modes; additionally, one *C2_Drift* sample was incorrectly classified as *Healthy* (a false negative of operational concern), while three *Healthy* samples were falsely flagged as *Insulation Degradation*. Notably, the model achieved perfect recall for *Ferroresonance*, *Overvoltage*, and *Undervoltage* faults, indicating that the learned decision boundaries effectively isolate these distinct fault signatures.

DEVELOPMENT OF FAULT DIAGNOSIS FRAMEWORK FOR 132 KV CAPACITOR VOLTAGE TRANSFORMER USING MACHINE LEARNING FRAMEWORK

		SVM Confusion Matrix						
True Class	C1 _{Drift}	71			32	24		
	C2 _{Drift}	57			45	30		
	Ferroresonance	94			53	42		
	Healthy	2620	9	8	1522	1209	14	
	Insulation _{Degradation}	31			15	13		
	Overvoltage	23			15	18		
	Undervoltage	29			18	8		
		Predicted Class						
		C1 _{Drift}	C2 _{Drift}	Ferroresonance	Healthy	Insulation _{Degradation}	Overvoltage	Undervoltage

Figure 7: Confusion Matrix of the Support Vector Machine Model for CVT Fault Classification

The confusion matrix for the Support Vector Machine (SVM) in Figure 7 reveals significant classification challenges, particularly regarding class imbalance and the separation of minority fault signatures. The model exhibits a high false positive rate for the "Healthy" class, misclassifying 2,620 instances as "C1_Drift" and 1,209 as "Insulation Degradation," which indicates a strong prediction bias toward these specific labels. Furthermore, the SVM demonstrated zero recall for "C2_Drift," "Ferroresonance," "Overvoltage," and "Undervoltage," failing to correctly identify any true positives for these categories and instead scattering predictions primarily into "C1_Drift," "Healthy," and "Insulation Degradation." This suggests that the linear or kernel-based decision boundaries established by the SVM were insufficient to isolate the complex feature spaces of the minority faults, resulting in poor overall diagnostic reliability compared to tree-based methods.

		kNN Confusion Matrix						
True Class	C1 _{Drift}			1	126			
	C2 _{Drift}				132			
	Ferroresonance			3	186			
	Healthy	3	4	16	5356	3		
	Insulation _{Degradation}			2	57			
	Overvoltage				56			
	Undervoltage				55			
		Predicted Class						
		C1 _{Drift}	C2 _{Drift}	Ferroresonance	Healthy	Insulation _{Degradation}	Overvoltage	Undervoltage

Figure 8: Confusion Matrix of the k-Nearest Neighbors Model for CVT Fault Classification

The confusion matrix for the k-Nearest Neighbors (kNN) classifier in Figure 8 demonstrates a severe classification bias toward the majority "Healthy" class, resulting in a near-total failure to detect fault conditions. The model achieved zero recall for C1_Drift, C2_Drift, Insulation Degradation, Overvoltage, and Undervoltage, and correctly identified only 3 instances of Ferroresonance, effectively misclassifying the vast majority of faults as "Healthy." This outcome highlights a critical limitation of standard distance-based kNN algorithms on imbalanced datasets, where the high density of majority

class neighbors overwhelms the minority class signals, rendering the model ineffective for fault prediction without advanced resampling or cost-sensitive learning techniques.

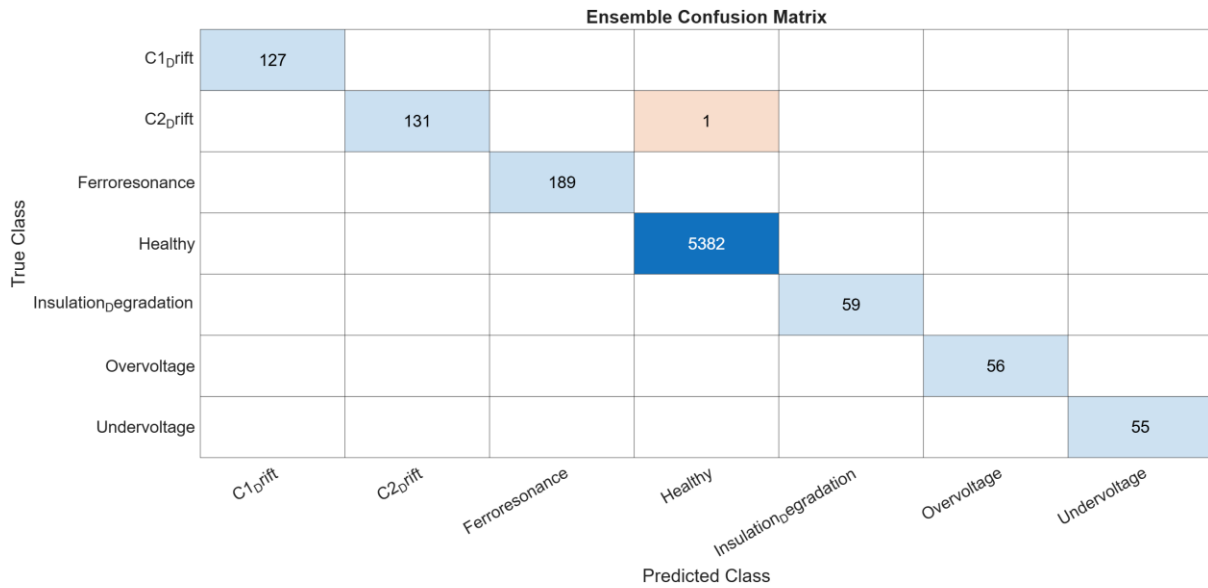


Figure 9: Confusion Matrix of Ensemble Model for CVT Fault Classification

The confusion matrix for the Ensemble classifier in Figure 9 reveals superior diagnostic performance, effectively resolving the classification errors observed in the individual base learners. The model demonstrates exceptional accuracy, with the vast majority of instances correctly classified along the diagonal, including precise identification of all "Ferroresonance," "Overvoltage," and "Undervoltage" faults, as well as 5,382 "Healthy" samples. Notably, the ensemble approach successfully eliminated the confusion between "C1_Drift" and "Insulation Degradation" seen in the Decision Tree model, achieving perfect recall for these categories. The only recorded misclassification was a single instance of "C2_Drift" falsely predicted as "Healthy," underscoring the ensemble method's robustness in handling class imbalance and its high reliability for multi-class fault diagnosis.

Based on the comparative evaluation, the Ensemble classifier emerges as the most effective model for multi-class fault prediction, significantly outperforming the standalone Decision Tree, SVM, and kNN. While the Decision Tree demonstrated solid baseline performance, the Ensemble method achieved superior diagonal alignment and completely resolved the feature-overlap confusion previously observed between C1_Drift and Insulation Degradation. Technically, this performance gain is driven by the ensemble's aggregation mechanism, which reduces the high variance and overfitting tendencies inherent to single decision trees, while implicitly balancing class distributions through diversified voting or weighted predictions. In contrast, the SVM and kNN models exhibited severe majority-class bias and near-zero minority recall, underscoring the structural limitations of margin-based and distance-dependent algorithms when applied to highly imbalanced sensor data without explicit class weighting or synthetic resampling.

Table 2: Fault prediction metrics for ML Classifiers

Method	Accuracy (%)	Macro Precision	Macro Recall	Macro F1-Score
Decision Tree	99.88	0.988	0.994	0.991
Support Vector Machine	26.77	0.133	0.152	0.071
K Nearest Neighbour	89.32	0.148	0.144	0.139
Ensemble	99.98	1.00	0.999	0.999

As summarized in Table 2, the Ensemble classifier unequivocally outperforms all competing methods, achieving 99.98% accuracy alongside near-perfect macro precision (1.00), recall (0.999), and F1-score (0.999). While the Decision Tree demonstrates strong baseline performance, the Ensemble model surpasses it by systematically reducing high-variance decision boundaries and amplifying discriminative signals for minority fault classes through aggregated voting or stacking mechanisms. Critically, the stark divergence between overall accuracy and depressed macro metrics for SVM (26.77% accuracy, 0.071 macro F1) and kNN (89.32% accuracy, 0.139 macro F1) reveals a severe majority-class bias; these margin- and distance-based algorithms default to predicting the dominant "Healthy" label, artificially inflating accuracy while failing to detect faults. Macro-averaged metrics penalize this imbalance by weighting all classes equally, confirming that only the Ensemble framework successfully learns balanced, class-invariant representations for rare fault signatures.

4. CONCLUSION

A framework for fault diagnosis framework for 132 kV capacitor voltage transformer is presented. The study presents a condition-aware simulation driven fault diagnosis framework of a 132 KV capacitor voltage transformer via the combination of CVT system modeling with operating conditions. A high-fidelity CVT model was developed to simulate both normal and fault operating conditions while incorporating temperature variations, harmonic distortion, and measurement uncertainties. A dataset comprising 20,000 samples was generated, containing electrical and operational features including primary voltage, secondary voltage, ratio error, phase error, temperature, and total harmonic distortion (THD). The generated dataset was subsequently used to develop and evaluate Decision Tree, Support Vector Machine (SVM), k-Nearest Neighbor (kNN), and Ensemble Machine Learning classifiers for multi-class fault prediction. Comparative results demonstrated that the Ensemble classifier achieved superior performance with an accuracy of 99.98%, macro precision of 1.00, macro recall of 0.999, and macro F1-score of 0.999, significantly outperforming the conventional machine learning models. The zero-recall achieved by SVM for "C2_Drift," "Ferroresonance," "Overvoltage," and "Undervoltage" are due to the fact that SVMs are sensitive to class imbalance and overlapping between classes and when such classes or subtle or underrepresented, the margin can be positioned so that the samples are pushed into other classes, giving zero recall even if the accuracy looks acceptable. The ensemble classifier reduces the weakness of SVM and KNN by combining multiple complimentary decision rules. The proposed framework provides an effective solution for accurate fault prediction, improved fault representation, and enhanced condition-based evaluations of CVTs.

5. REFERENCES

- Arouche Freire, J. C., Garcez Castro, A. R., Homci, M. S., Meiguins, B. S., & De Moraes, J. M. (2019). Transmission Line Fault Classification Using Hidden Markov Models. *IEEE Access*, 7, 113499–113510. <https://doi.org/10.1109/ACCESS.2019.2934938>
- Haghighoo, O., & Damchi, Y. (2022). Reliability modeling of capacitor voltage transformer using proposed Markov model. *Electric Power Systems Research*, 2021, 107573. <https://doi.org/10.1016/j.epsr.2021.107573>.
- Meng, L., Bai, B., Zhang, W., Liu, L., & Zhang, C. (2023). Research on a Decision Tree Classification Algorithm Based on Granular Matrices. *Electronics (Switzerland)*, 12(21), 1–14. <https://doi.org/10.3390/electronics12214470>.
- Omar, C. A., Dora, A. S., Jose, A. A. T., Ronald, F. D. L., Rufino, A. I., Lilly, R. M. C., & Lilia, R. C. (2022). Supervised learning using support vector machine applied to sentiment analysis of teacher performance satisfaction. *Indonesian Journal of Electrical Engineering and Computer Science*, 28(1), 516-524.
- Sappati, K. K., & Bhadra, S. (2018). Piezoelectric polymer and paper substrates: A review. *Sensors (Switzerland)*, 18(11). <https://doi.org/10.3390/s18113605>

Shahid, M.H.B and Azim, A. (2023). Ensemble method for fault detection and classification in Transmission lines using machine learning. *Proceedings of 2023, IEEE International Systems Conference (SysCon), Vancouver, BC, Canada, 17-20. pp. 1-7.*