E3S Web of Conferences **729**, 07005 (2026) https://doi.org/10.1051/e3sconf/202672907005
SPETM 2026

Explainable machine learning for health assessment of PILC distribution cables

Ajibola Oyedeji [1*], and *Peter* Olukanmi[2]

[1]Graduate Student, School of Electrical and Electronics Engineering, University of Johannesburg, South Africa
[2]Senior Lecturer, School of Electrical and Electronics Engineering, University of Johannesburg, South Africa

Abstract. While Paper Insulated Lead Covered (PILC) cables are a legacy technology, they remain a critical fixture in medium-voltage power grids. However, as these cables are continuously used, they present growing risks to grid stability and public safety due to electrical, thermal, and environmental factors. As replacing entire networks is cost-prohibitive, utility providers need a smarter way to predict when and where a cable might fail. This paper presents an explainable, data-driven framework for multi-class health assessment of 20 kV PILC cables, aligning with the industry 4.0 standards of autonomous asset management. Using a real-world dataset of 999 inspection records from European utilities, four supervised learning models; Random Forest, AdaBoost, XGBoost, and a Multilayer Perceptron deep neural network were trained. The models categorize cable health into five standard IEC/IEEE health bands, achieving high classification accuracies between 97.0% and 99.5%. To ensure transparency in the decision-making process, SHapley Additive exPlanations (SHAP) were employed, identifying Partial Discharge (PD) and Thermal Difference (TD) Stability as the primary predictors of insulation degradation. This framework provides an interpretable tool for power utilities to transition from reactive to predictive maintenance scheduling.

1 Introduction

Power infrastructure reliability is dependent on the condition of various power infrastructure, one of which are Paper-Insulated Lead-Covered (PILC) cables that have been widely deployed for decades [1]. PILC cables remain deeply embedded in medium-voltage (MV) distribution systems globally despite the introduction of Cross-Linked Polyethylene (XLPE) cables [2]. These old cables are still used extensively in many utilities' underground networks, some of which are long past their expected lifespan as a complete network replacement is neither technically straightforward nor economically feasible [3].

The degradation of PILC cable insulation is governed by a complex interplay of electrical, thermal, and environmental factors [1]. Distribution system operators therefore face the dual burden of aging assets and intensifying operational demands. Conventional infrastructure management typically relies on time-based maintenance which schedules interventions at regular intervals irrespective of the state of the cable [2].

The emergence of Fourth Industrial Revolution (4IR) data-driven techniques such as machine learning (ML) and deep learning (DL), offer a compelling alternative to time-based maintenance [4]. Predictive maintenance plays a key role by introducing digital machine maintenance, enabling the minimisation of downtime and associated costs while maximising asset life cycle [5]. ML models can learn complex non-linear relationships between diagnostic parameters and cable health status, for autonomous and scalable condition monitoring of power infrastructure. Several studies have applied ML and DL models for condition monitoring of power distribution assets including random forest (RF), support vector machines (SVM), XGBoost, Artificial Neural Networks (ANN), Multilayer Perceptron (MLP) with strong predictive performance [6].

Monitoring the health conditions of MV cables is crucial for effective power distribution, as failures can lead to outages. By using ML models for predictive modelling, the effects can be mitigated. However, the black-box nature of ML models presents a barrier to adoption in real-world applications where field engineers must be able to understand and justify automated diagnostic decisions [7].

This paper presents an explainable artificial intelligence (XAI) pipeline for PILC cable health assessment. The contributions of this paper are as follows:

i. Multi-class condition classification using four ML and DL (RF, AdaBoost, XGBoost, and MLP) models for 20 kV PILC cables.
ii. Explainability of the models using SHapley Additive exPlanations (SHAP) to identify the models' influencing predictors.

The remaining parts of this paper are sectioned as follows: Section 2 presents a brief literature study while Section 3 describes methodology and dataset used. In Section 4, the classification results were presented, and Section 5 describes the SHAP explainability analysis.

* Corresponding author: 223235694@student.uj.ac.za

The study concludes with Section 6 providing a summary of the key findings and insights into future work.

2 Related Works

2.1 PILC Cable Degradation and Diagnostic Methods

Paper-Insulated Lead-Covered (PILC) cables are a critical component of electrical distribution infrastructure [8]. However, the operational efficiency is influenced by a combination of thermal, electrical, mechanical, and environmental factors [1]. A primary degradation mechanism is the long-term exposure to thermal cycling, which gradually degrades the paper-oil insulation system. Conventional diagnostic methods for assessing PILC health status include Partial Discharge (PD) testing, dielectric analysis, thermal monitoring and statistical modelling.

2.2 Machine Learning for Cable Health Assessment

The move from reactive to proactive maintenance has been facilitated by the advent of 4IR technologies with ML at its forefront. The application of supervised ML algorithms for power distribution infrastructure health assessment has seen significant growth. A recent study applied SVM, ANN, K-nearest neighbour, and Naive Bayes to classify the health index of XLPE cable insulation, with ANN achieving 94.5% accuracy [6]. A further study designed a Dempster–Shafer evidence theory-based multi-model fusion method for distribution line fault identification which achieved an accuracy of 74.6% [9]. However, these models are difficult to interpret and understand how the decisions have been made due to their inherent black-box nature.

2.3 Explainable AI in Power Asset Diagnostics

Explainable AI (XAI) addresses the lack of transparency in model decision-making. In the power systems domain, SHAP offers high trustworthiness and is suitable for post-hoc explanation of complex ML models used in condition monitoring, but its real-time applicability is limited due to high computational cost while Local Interpretable Model-agnostic Explanations (LIME) offers a less trustworthy explanation but computationally effective [7]. XAI tools like SHAP and LIME have been used for solar PV forecasting and wind turbine generators maintenance [10].

Despite the growing use of XAI in power systems, its application to PILC cable health assessment has not been reported. This paper addressed this gap by applying SHAP methods to classifiers to get feature influence that can be validated against established PILC degradation physics.

3 Methodology

This section details the methodology employed to achieve the contributions of this study, which includes exploratory data analysis and data preprocessing, model building, evaluation and model explainability.

3.1 Dataset Description and Preprocessing

The selected dataset contains data of 20kV PILC Cable Inspection records taken between the year 2007 to 2019 from a European utility company. The dataset includes 999 samples and 8 features which are Age, PD, Thermal Difference (TD) Stability, Delta TD, Mean TD, Visual Condition, Neutral Corrosion, Health Index [11].

PD measures the localized electrical discharges within insulation, while the age describes the number of years of use. Delta TD and mean TD are thermal features which describe the difference in temperature over time and the average operating temperature of the cable respectively. TD stability is the degree of stability of the temperature and neutral corrosion measures the degree of corrosion. The integration of these electrical, thermal, environmental, and physical features provides a comprehensive representation of cable health [12].

To transform the Health Index column into a multi-class health condition class, the IEC/IEEE standard 5-class health bands was applied. The distribution of the PILC cable health condition classes is presented in Fig. 1 with the class Poor having the most instances with 316 (31.6%) followed by Very Poor (286, 28.6%), while Fair had the least number of instances (110, 11.0%).

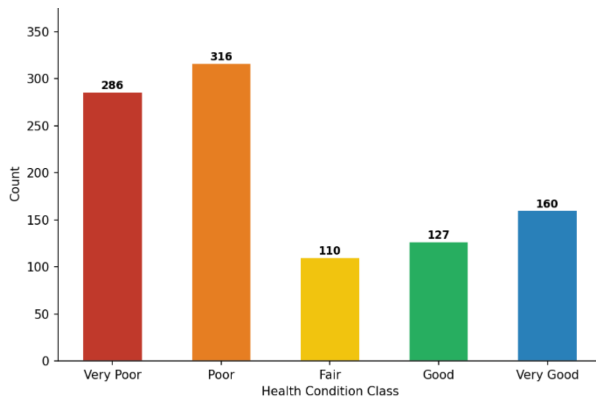

Fig. 1. Health Condition Class Distribution for the Dataset Showing the Number of Records Across the Five Classes.

The boxplot analysis of feature distributions across the five IEC/IEEE health classes presented in Fig. 2 reveals several significant patterns. Across the seven diagnostic features, there is a clear separation observed between the health classes, with the most degraded cables (Very Poor) having the highest values for electrical and corrosion-related parameters, and the healthiest cables (Very Good) showing the lowest values.

After preliminary data analysis, the data was divided into the train and test set using the stratified 80:20 split to preserve class proportions. The health index column was also dropped to avoid data leakage. Thereafter, a standard scaler was employed to normalise the data.

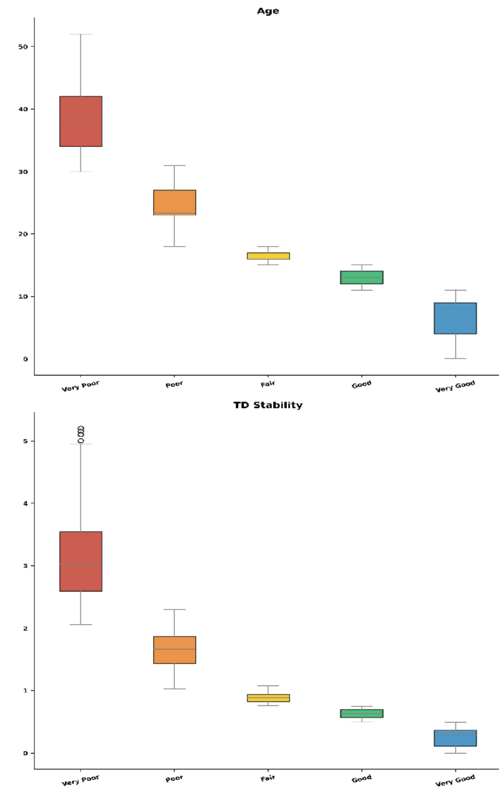

Fig. 2. Boxplot Analysis of the Age and TD Stability Distributions.

3.2 Model Development

Supervised learning methods were chosen to predict the cable health based on descriptive attributes. To perform the classification tasks, four ML and DL models are employed, namely random forest (RF), AdaBoost, Extreme Gradient Boosting (XGBoost), and a multilayer perceptron (MLP). Fig. 3 depicts the steps involved in the study.

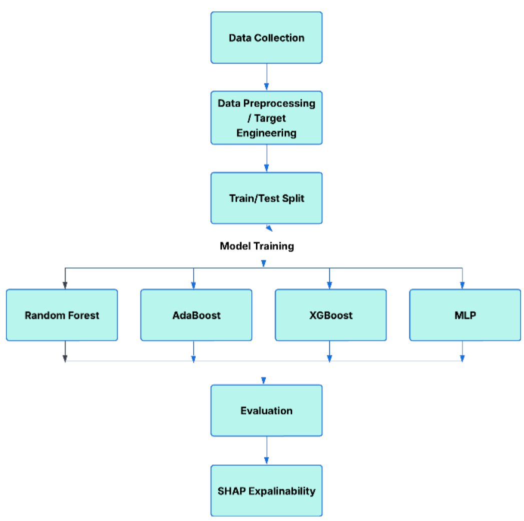

Fig. 3. The System Flowchart illustrating the Processes Involved.

3.2.1 Random Forest

Random Forest is a supervised ensemble machine learning model comprising of multiple decision tree models. Each tree is trained on a bootstrap sample a random subset of the training data drawn with replacement known as bagging [4]. RF is well-suited to the PILC cable diagnostic dataset as it handles non-linear, non-monotonic relationships between diagnostic parameters and health class without requiring feature engineering.

3.2.2 AdaBoost

Combining several weak learner models that only marginally outperform random guessing into a single strong learner with enhanced prediction performance is the basic notion behind boosting [13]. This is accomplished by an iterative process where each subsequent weak learner concentrates on fixing the mistakes made by its predecessors.

$$F_T(x) = \sum_{t=1}^{T} f_t\,(x) \tag{1}$$

where f_t is the weak learner and x is the input, T is the number of weak learners and $F_T(x)$ is the model output.

3.2.3 XGBoost

Because of its excellent performance, scalability, and effectiveness in managing big and complicated datasets, XGBoost, an improved gradient boosting solution, has become quite popular. It is an ensemble learning technique that sequentially combines several weak learners, usually decision trees, to create a powerful prediction model [4]. The accuracy of the model is improved as each new tree is trained to fix the mistakes caused by the earlier ones.

3.2.4 MLP

The DL model deployed in this study is a Multilayer Perceptron (MLP), a fully connected feedforward neural network augmented [14]. The model architecture consists of an input dense layer, two hidden dense layers with batch normalization and dropout (128 and 64 units, ReLU), a linear projection skip connection forming a residual addition, followed by a 32-unit ReLU head and a 5-class SoftMax output as shown in Fig. 4.

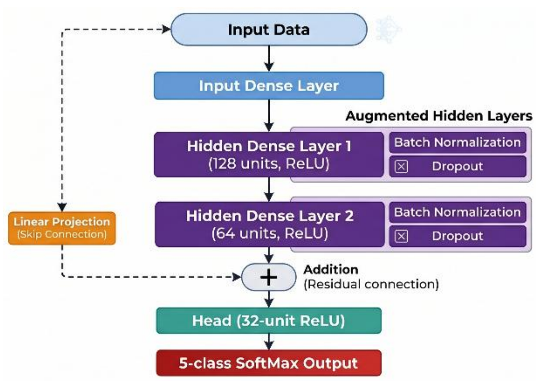

Fig. 4. MLP Model Architecture with an Input Dense Layer, 2 Hidden Dense Layers of Sizes 128 and 68 Respectively, 32 ReLU, and Linear Project Skip.

3.3 SHAP Explainability

A critical limitation of many existing studies on ML-based power systems condition assessment is the development of black-box models whose decision-making process cannot be described in a physically interpretable way. For critical utility systems, this presents a practical barrier to real-world engineering environment adoption. To address this limitation, this study integrates Shapley Additive exPlanations (SHAP) as a post-hoc explainability layer applied to all models.

SHAP is a solution grounded in game theory which provides a theoretically consistent attribution framework [7]. SHAP provides multiple explainer implementations optimised for different model architectures. In this study, The TreeExplainer is applied to both the RF and XGBoost, while the KernelExplainer is used for AdaBoost and MLP [15]. For each model in this study, SHAP values for Global feature importance, Per-class Beeswarm plots and normalised Class-level importance heatmap are computed.

3.4 Evaluation

To determine the performance and reliability of the models, standard performance metrics were used to evaluate. These metrics include accuracy, precision, recall and F1-macro average for classification [4].

Accuracy is the overall correctness of the prediction defined in equation 2. Precision calculates the percentage of correctly predicted positive cases out of all cases classified as positive, defined in equation 3. Recall measures the percentage of real positive examples that were correctly detected by the model. The F1-score which is the harmonic mean of precision and recall.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{5}$$

where TP= true positives, TN=true negatives, FP= false positive and FN=false negative.

4 Results and Discussion

This section presents the results of the classification experiment. The study used four ML and DL models namely random forest, XGBoost, AdaBoost, and MLP for the multiclassification of the health condition of the PILC cables in power distribution systems.

The confusion matrix of the performance of the RF model on the test data set is presented in Fig. 5. The results reveal that all models performed well, with all perfectly classifying the extreme classes, that is, 'Very Poor' and 'Very Good'. The 'Fair' class serves as the primary transition zone in the dataset, and it is where the few classification errors occur.

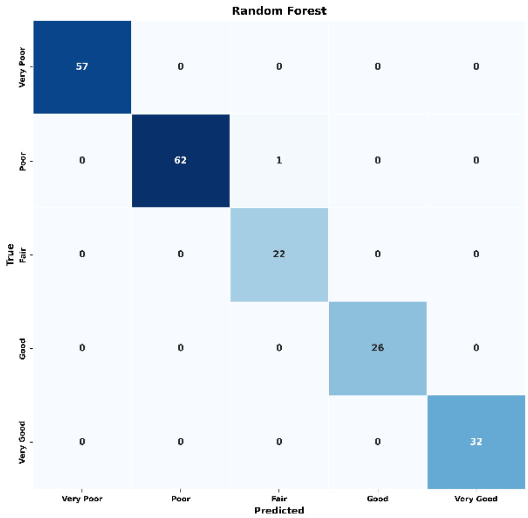

Fig. 5. Confusion Matrix Showing the 5-Class Health Status Classification by the Random Forest Model.

Table 1 below presents the result of the classification models using the evaluation metrics described in Section 3.4. The accuracy ranges from 97% to 99.50% with random forest achieving the best value. RF also had the best F1-score which is essential for an imbalanced dataset. With F1-score of 99.40%, RF had the best performance, followed by AdaBoost and XGBoost with 98.77% and 98.55%. MLP was the least performing model with 97% accuracy and 95.66% F1-score. Although, AdaBoost and XGBoost performed identically in terms of accuracy (99.00%), AdaBoost showed a slightly higher F1-score, suggesting better performance on minority classes.

Table 1. Classification Evaluation.

Models	Accuracy	Precision	Recall	F1-score
RF	99.50	99.52	99.50	99.40
AdaBoost	99.00	99.00	99.00	98.77
XGBoost	99.00	99.02	99.00	98.55
MLP	97.00	97.02	97.00	95.66

5 SHAP Interpretability Analysis

To ensure that the performance of the models is rooted in physical degradation indicators rather than statistical noise, this study utilized SHAP to understand the features driving the decision-making process of each model.

5.1 Global Feature Importance

The global importance of features across all four models was analysed using the mean absolute SHAP value, which represents the average impact on the model's output magnitude.

 https://doi.org/10.1051/e3sconf/202672907005

As illustrated in Fig. 6, there is a high degree of consensus among RF, XGBoost and AdaBoost models. Partial Discharge (PD) and TD Stability consistently emerge as the most influential predictors of the PILC health condition. Although, XGBoost demonstrates a significantly higher sensitivity to these features with SHAP values compared to the other models, the relative ranking remains stable. This consistency exhibited from the tree-based models to neural networks supports the robustness of these features as reliable predictors of PILC cables' health.

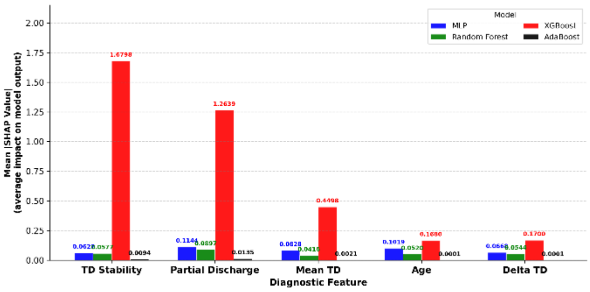

Fig. 6. SHAP Global Feature Importance with Partial Discharge (PD) and TD Stability as Most Influential Features.

5.2 Class-Level Dynamics

While global importance identifies which features drive decision-making, the SHAP Class-Level Heatmap for the best model (RF) reveals how feature roles shift across the health spectrum as presented in Fig. 7.

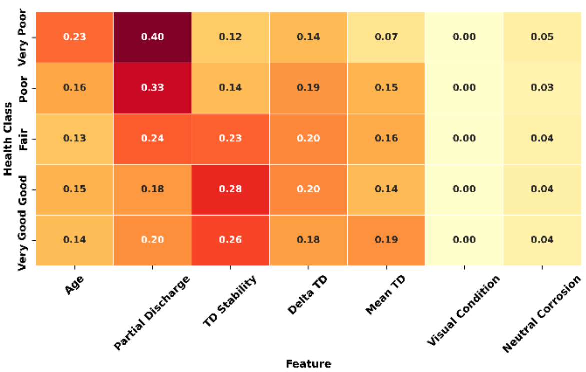

Fig. 7. SHAP Class-Level Heatmap Depicting the Feature Normalized Feature Importance by Random Forest Model.

The class-level heatmap analysis for RF indicates a clear transition in model sensitivity as asset health deteriorates from Very Good to Very Poor. For the 'Very Poor' class, the model relies on a combination of Partial Discharge (PD) (0.40) and Age (0.23). This suggests that the model identifies the highest risk states from discharge activity in older assets. As the asset condition improves to 'Very Good' the importance of PD drops to 0.20, and TD Stability (0.26) becomes the primary predictor. However, Delta TD maintains a consistent importance (0.14–0.20) across all classes, acting as a baseline indicator of insulation health regardless of the specific category, while Neutral Corrosion and Visual Condition have insignificant influence across all classes.

To provide a comprehensive view of the model's decision-making across the entire PILC degradation lifecycle, the SHAP Beeswarm plots for all five health classes of the Random Forest model are presented in Fig. 8.

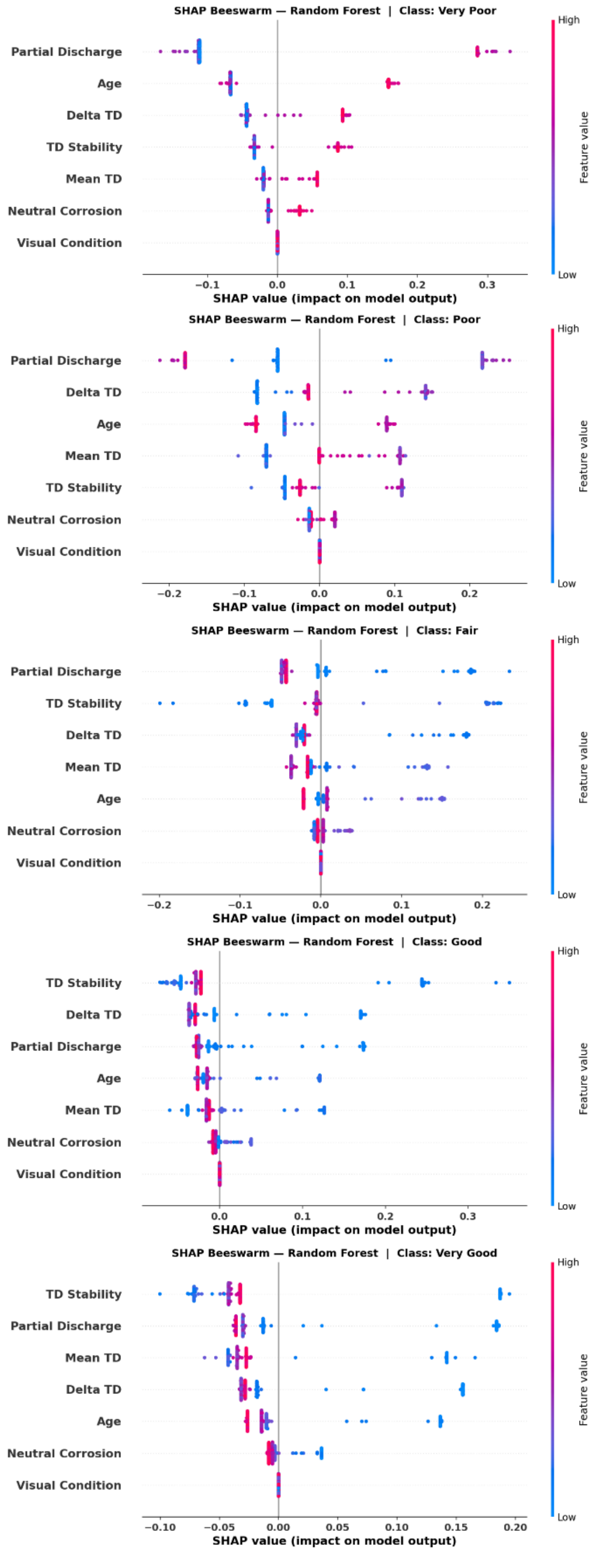

Fig. 8. SHAP Local Beeswarm for RF Showing the Impact of the Features in Decision Making across the Classes.

High PD values represented with red points result in large positive SHAP values (up to +0.3), significantly increasing the probability of a 'Very Poor' classification, while low values (blue) strongly decrease this probability. Similarly, age plays an important role, where older assets are more likely to be classified as very poor. Similarly, thermal degradation indicators such as Delta TD, TD Stability, and Mean TD all exhibit a consistent pattern with higher values increasing the likelihood of a 'Very Poor' classification, indicating that both the magnitude and instability of thermal behaviour are key contributors.

In the healthy categories ('Good' and 'Very Good'), the TD Stability moves to the top of the feature hierarchy, but with an inverted impact with low feature values (blue clusters) now providing the positive SHAP offsets required to confirm a 'Good' or 'Very Good' status.

Across all plots, Neutral Corrosion and Visual Conditions remain at the bottom with a very tight SHAP distribution close to 0. Although, they are known physical phenomenon, SHAP analysis reveals they have negligible impacts on the model's diagnostic output compared to internal dielectric signatures (PD and TD Stability). This suggest that internal paper-oil insulation breakdown is the primary driver of PILC cable failure in this study.

6 Conclusion and Recommendations

This paper presented an explainable machine learning framework for the health condition assessment of aging Paper Insulated Lead Covered (PILC) distribution cables. By evaluating four ML architectures, namely random forest, AdaBoost, XGBoost and multilayer perceptron, we demonstrated that machine learning can achieve great diagnostic accuracy, thus improving predictive maintenance. The random forest model achieved the best classification performance with 99.5% accuracy and 99.40% F1-score.

A critical contribution of this study is the integration of SHapley Additive exPlanations (SHAP) to achieve interpretability of these models. The interpretability analysis showed that across all models, Partial Discharge (PD) and TD Stability were identified as the most influential predictors of health assessment, validating that the models' decision-making processes are aligned with established high-voltage insulation theory.

A significant limitation of this work is that all experimental results are derived from a single European utility company dataset which are specific to the network conditions, climate, environmental, loading profiles, and maintenance practices of that utility, which limits the generalization of the model performance and influencing factors.

Future work will focus on the validation of the proposed framework on PILC cable datasets collected from multiple utilities across diverse geographic regions and climatic conditions. Cross-utility validation is essential for establishing whether the feature importance consensus observed in this study holds across different inspection protocols, cable construction standards, and operating environments, or whether region-specific recalibration of model parameters is required.

Data availability statement: The data is publicly available.

Author contribution statement: **AO**: Conceptualization, Methodology, Visualization, Writing-Original draft; **PO**: Supervision, Writing-Review and Editing.

References

1. H. S. A. Halim, P. Ghosh, Condition Assessment of Medium Voltage Underground PILC Cables Using Partial Discharge Mapping and Polarization Index Test Results, in Conference Record of the 2008 IEEE International Symposium on Electrical Insulation, IEEE, Jun. (2008). https://doi.org/10.1109/ELINSL.2008.4570270
2. M. Zapf, T. Blenk, A. C. Müller, H. Pengg, I. Mladenovic, C. Weindl, Lifetime assessment of PILC cables with regard to thermal aging based on a medium voltage distribution network benchmark and representative load scenarios in the course of the expansion of distributed energy resources. Energies (Basel). **14**, 2 (2021). https://doi.org/10.3390/en14020494
3. A. Rafati, H. Mirshekali, H. R. Shaker, N. Bayati, Power Grid Renovation: A Comprehensive Review of Technical Challenges and Innovations for Medium Voltage Cable Replacement. Smart Cities. **7**, 6, (2024). https://doi.org/10.3390/smartcities7060144
4. A. Oyedeji, P. Olukanmi, Intelligent Condition Monitoring of Power Transformers via Machine Learning, in 2025 33rd Southern African Universities Power Engineering Conference (SAUPEC), IEEE, Jan. 30-31 (2025). https://doi.org/10.1109/SAUPEC65723.2025.10944449
5. M. Achouch et al., On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. Applied Sciences. **12**, 16, (2022). https://doi.org/10.3390/app12168081
6. A. Ansari, S. Yousef Nia, A. Afshari, S. Mishra, Estimating the insulation health index of XLPE cables using machine learning. Sci. Rep. 15, 1, (2025). https://doi.org/10.1038/s41598-025-25080-7
7. M. R. Shadi, H. Mirshekali, H. R. Shaker, Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects. Renewable and Sustainable Energy Reviews. **216**, (2025). https://doi.org/10.1016/j.rser.2025.115668
8. F. Zec, S. D. Milić, Đ. Lazarević, J. Dragosavac, Prediction of Remaining Life and Insulation Failure of High-Voltage Distribution Cable Using Statistical Methods. Mathematics. **14**, 7, (2026). https://doi.org/10.3390/math14071164
9. W. Wan, Y. Yue, W. Liu, M. Deng, J. Zhang, Z. Qin, An evidence theory based multiple model fusion method for fault diagnosis of distribution line. Sci. Rep. **16**, 1, (2026). https://doi.org/10.1038/s41598-025-30770-3

https://doi.org/10.1051/e3sconf/202672907005

10. L. Rajaoarisoa, R. Randrianandraina, G. J. Nalepa, J. Gama, Decision-making systems improvement based on explainable artificial intelligence approaches for predictive maintenance. Eng. Appl. Artif. Intell. **139**, (2025). https://doi.org/10.1016/j.engappai.2024.109601
11. R. Alden, 20kV PILC Cable Historical Inspection Data. IEEE DataPort. (2020). https://doi.org/10.21227/babt-n267
12. J. Tao, S. U. Rehman, R. Ali, S. A. Raza, Advancement and challenges: A review of power cable aging monitoring and diagnostic techniques. Renewable and Sustainable Energy Reviews. **222**, (2025). https://doi.org/10.1016/j.rser.2025.115970
13. M. Begum et al., Enhancing reliability in electrical grids: A hybrid machine learning approach for electrical faults classification. PLoS One. **21**, 2, e0341238, (2026). https://doi.org/10.1371/journal.pone.0341238
14. V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, G. Kasneci, Deep Neural Networks and Tabular Data: A Survey. IEEE Trans. Neural Netw. Learn. Syst. **35**, 6, (2024). https://doi.org/10.1109/TNNLS.2022.3229161
15. S. M. Lundberg et al., From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. **2**, 1, (2020). https://doi.org/10.1038/s42256-019-0138-9