MEDIUM VOLTAGE CABLE FAULT AND HEALTH ASSESSMENT THROUGH ADVANCED ANALYTICS

Akshat Kulkarni[1], Ricardo Prata[2], Pratik Bajaria[1], Sanjeev Kumar[1]

[1]*OrxaGrid Ltd, London, UK,*
[2]*ENOWA, Saudi Arabia*
akulkarni@orxa.io, ricardo.prata@ieee.org, pbajaria@orxa.io, skumar@orxa.io

Keywords: UNDERGROUND CABLES, GRID RELIABILITY, MACHINE LEARNING, GIS, UTILITIES MANAGEMENT

Abstract

This paper elaborates on innovative approaches by OrxaGrid, a leader in IoT, T&M, and analytics for power grids, focusing on enhancing the health and reliability of underground cables. These critical assets, vital for robust infrastructure, face increasing demands from emerging technologies such as electric vehicles. By integrating advanced analytics and machine learning with traditional monitoring systems, OrxaGrid provides actionable insights, significantly improving operational efficiency and reducing outage incidents. A case study with a European Distribution System Operator highlights the predictive success of our methodologies, achieving key performance metrics with precision, recall, and F1-score above 0.79.

1 Introduction

Underground cables are crucial for uninterrupted power distribution, but aging infrastructure and the integration of new technologies increase the risk of failures. This paper explores how data science can be used to enhance the resilience of these critical assets. OrxaGrid's technology leverages the Internet of Things (IoT) [1] and Artificial Intelligence (AI) [2] to monitor underground cables. By deploying sensors to collect real-time data on temperature, voltage, current, and other parameters, OrxaGrid can analyse this information using advanced data science techniques. This enables the identification of patterns and anomalies, allowing for the prediction of potential failures before they occur. By proactively addressing these issues, OrxaGrid helps ensure the reliability of the power grid.

The modern power grid faces numerous challenges, including the increasing integration of renewable energy sources and electric vehicles. These factors, coupled with aging infrastructure and environmental stresses, significantly impact the reliability of underground cable systems. These cables are susceptible to thermal aging, moisture ingress, mechanical stress, and electrochemical corrosion, which can lead to insulation breakdown and ultimately, power outages. To address these challenges, OrxaGrid's approach involves, **c**ollecting real-time data from sensors deployed along the cable route, efficiently collecting, storing, and analysing the sensor data, utilizing machine learning algorithms to identify patterns and predict potential failures, and developing actionable insights and recommendations to guide maintenance activities and optimize resource allocation.

By leveraging these technologies, OrxaGrid aims to, gain a comprehensive understanding of the health and performance of underground cable systems, minimize the need for costly emergency repairs by proactively addressing potential issues, minimize power outages and ensure a more reliable power supply for consumers, and improve grid efficiency and utilization by identifying and addressing bottlenecks.

However, existing research often focuses on specific aspects of grid management, such as fault detection or condition assessment, without a comprehensive approach that integrates all relevant data sources and stakeholders. Moreover, many studies rely on limited datasets or simplified models, which may not accurately reflect the complexities of real-world grid operations. Recent research has explored various approaches for predicting the health of underground cables. Machine learning techniques, such as support vector machines (SVM) [3], artificial neural networks (ANN) [4, 5], and random forests [6], have shown promise in identifying patterns and anomalies in sensor data to predict potential failures. Deep learning models, including convolutional neural networks (CNN) [7] and recurrent neural networks (RNN) [8], have been applied to analyse complex time-series data from distributed fiber optic sensors (DFOS) [9], enabling more accurate and timely predictions. Other studies have investigated the use of physics-based models [10] and hybrid approaches that combine physics-based models with data-driven techniques [11, 12] to improve the accuracy and interpretability of predictions. These advancements in cable health prediction contribute to proactive maintenance strategies, enhancing grid reliability and reducing operational costs.

This paper aims to address these research gaps by:

- **Developing a holistic framework for underground cable management:** Integrating data from multiple sources, including sensor data, historical records, and external factors, to provide a comprehensive view of cable health.
- **Utilising advanced machine learning algorithms:** Exploring the potential of deep learning and other cutting-edge techniques for improved fault prediction and anomaly detection.
- **Conducting real-world field trials:** Validating the proposed approach through rigorous field trials on operational underground cable systems.

2 Methodology

OrxaGrid recognizes the diverse technological landscape within the utility sector. To ensure the applicability of our solutions across various levels of technological adoption, we employ a flexible and adaptable approach.

2.1 Statistical Risk Modelling for Utilities with Limited Monitoring

For utilities without sophisticated monitoring systems such as SCADA or ADMS, we develop robust statistical risk models. These models leverage readily available data points, including: **Cable Age and Lifespan:** By analysing the age of the cable and considering its expected lifespan based on manufacturer specifications and operating conditions, we can estimate the level of degradation due to aging. **Usage Patterns and Fault History:** Examining historical load data provides insights into the stress levels experienced by the cable. Analysing past fault occurrences within the network helps identify areas of higher risk and potential failure modes.**Diagnostic Test Results:** Integrating the results of periodic diagnostic tests, such as sheath current measurements and partial discharge testing, provides valuable information about the cable's internal condition.

Based on these parameters, we develop a "Cable Health Score" that quantifies the overall risk of failure for each cable segment. This score serves as a valuable tool for prioritizing maintenance activities and allocating resources effectively.

2.2 Advanced Diagnostics with Machine Learning for Utilities with Advanced Monitoring

For utilities with access to real-time data through SCADA or ADMS systems at medium voltage level, OrxaGrid leverages advanced data analytics techniques to enhance the accuracy and granularity of our predictions. This involves:

- **Data Integration and Preprocessing:** We integrate diverse datasets, including real-time sensor data (temperature, current, voltage), historical fault records, cable characteristics (type, size, installation date), network topology data (connectivity, loading), and event tags (maintenance activities, weather events). Rigorous data preprocessing steps are crucial to ensure data quality, consistency, and accurate mapping of events to specific cable segments.
- **Machine Learning Model Development and Validation:** We employ a range of advanced machine learning algorithms, including: **XGBoost:** A powerful ensemble learning method known for its high accuracy and efficiency in handling high-dimensional datasets. **Neural Networks:** Deep learning models capable of capturing complex non-linear relationships within the data and identifying subtle patterns that may not be apparent to traditional statistical methods.
- **Model Validation and Refinement:** We rigorously validate the developed models using historical data to assess their predictive accuracy and robustness. This iterative process involves continuous refinement of model parameters and feature selection to optimize performance and minimize prediction errors.
- **Geographic Information System (GIS) Integration:** The integration of GIS is crucial for enhancing the precision and spatial accuracy of our analyses. GIS enables us to: **Visualize and analyse network topology:** Accurately map underground cable routes, identify critical connections, and assess the impact of potential failures on network operations. **Incorporate environmental factors:** Model the impact of

environmental factors, such as soil conditions, groundwater levels, and proximity to heat sources, on cable health and failure risk. **Optimise maintenance planning:** Visualize maintenance schedules, prioritize critical areas, and optimize resource allocation for maintenance crews.

2.3 Case Study POC Analysis

Our models were applied to data obtained from a European Distribution System Operator (DSO), beginning with foundational analyses and advancing to more intricate diagnostics across several substations. This process included: **Exploratory Data Analysis:** We charted current and voltage over time to pinpoint patterns indicative of faults. **Data Segmentation and Feature Engineering:** We improved model accuracy by selectively incorporating features essential for detecting faults.

The initial analysis focused on data from a single feeder, plotting current (I) and voltage (V) over time, along with current against a target variable indicating fault presence (1 for faults, 0 for no faults). This preliminary exploration was crucial for understanding the data's characteristics and identifying potential fault patterns.

- **Expansion of Analysis:** The analysis was subsequently broadened to include a single substation, and eventually all substations linked to second-generation cables. This comprehensive examination provided a full overview of the network's performance and the occurrences of faults.
- **Dataset Splitting:** The data was divided into training (80%) and testing (20%) sets randomly. This division was not based on a time series approach, which may affect the model's predictive accuracy for future trends.
- **Feature Selection and Engineering:** We identified ninety-three variables as features for our machine learning model, particularly focusing on those related to substations connected to second-generation cables. Voltage and current metrics were especially prioritized due to their significant impact on fault prediction accuracy.

Leveraging Data for Network Topology
We incorporated GIS data into the model to enhance the representation of cable sections and network topology. Additionally, we assessed how proximity to water bodies might affect cable health.

Comprehensive Development Process and Outcomes
Our approach to fault prediction in second-generation cables utilized machine learning techniques tailored to a specific network setup (Rede Ativa LXHIOZ1 cables). Here's an outline of our strategy and its results:

- **Data Utilisation:** The project capitalized on metadata concerning cable installations, such as installation dates, manufacturers, voltage levels, the number of sections, and historical SCADA data covering a minimum of three months. This extensive dataset laid the groundwork for recognising patterns and anomalies associated with cable faults.
- **Data Preparation:** We meticulously cleaned and refined the dataset, focusing solely on second-generation cables from Rede Ativa LXHIOZ1, resulting in roughly one million records based on SCADA data from 2022 and 2023.
- **Preliminary Analysis:** Initial data examinations were conducted on data from a single feeder by plotting current and voltage over time, along with current against a fault indicator. This exploratory phase was instrumental in discerning data behaviours and identifying potential fault indicators.
- **Expanded Analysis:** After analysing the feeder level, we extended our study to encompass one substation and then all substations linked to second-generation cables, providing a holistic view of network performance and fault instances.
- **Feature Selection and Engineering:** We pinpointed ninety-three parameters as critical features for our machine learning model, particularly emphasising those related to substations connected to second-generation cables. The significance of voltage and current parameters was underscored due to their crucial role in accurate fault detection.
- **Model Training and Evaluation:** We employed various machine learning models, with an XGBoost-based model showing the highest efficacy. This model achieved impressive precision, recall, and F1 scores of 0.79 in fault scenarios, indicating a robust capability to accurately identify both fault and non-fault conditions.

3 Challenges & Results

Implementing machine learning for power cable health indexing presents several key challenges. Primarily, acquiring sufficient, high-quality data is difficult, as cable failure data is often sparse and imbalanced. Data standardization across diverse sensor types is also complex. Again, model accuracy is hindered by the intricate, non-linear degradation processes of cables. Furthermore, ensuring model robustness and interpretability for critical infrastructure decisions poses a significant hurdle. Lastly the constant evolution of cable infrastructure and sensor technologies require continual model updates. The collaborative project with the European DSO allowed overcoming the above challenges and making the data available from the live devices.

Thus, the machine learning (ML) model yielded high predictive accuracy, with key performance metrics such as precision, recall, and F1-score surpassing the 0.79 mark *Fig. 1-4*. Our new model has achieved a remarkable accuracy rate (recall >.95, AUC>0.7). This high level of accuracy enables the proactive scheduling of maintenance works, thus significantly reducing both outages and associated costs.

The machine learning models were tested on diverse cable generations, from paper-oil insulated to modern dry underground cables. The actual number of faults detected in the second-generation cables was around 60. However, in the randomized test data, 7,124 fault labels were generated based on the criteria that for one fault, a 15-day window with 15-minute interval readings was considered. This approach likely helped in amplifying the instances of fault conditions for the model to learn from.

In summary, the project effectively used historical and metadata to build a machine learning model capable of detecting faults in second-generation cables. The use of an XGBoost model, informed by carefully selected features emphasizing voltage and current parameters, led to high precision and recall in fault prediction. Further, [13, 14] portray competitiveness of XGBOOST models alongside ANN models. This case study underscores the importance of targeted feature selection and the potential of machine learning in improving fault detection processes in power networks. The model demonstrated high effectiveness in fault prediction with precision, recall, and F1 scores of 0.79 for fault cases and a perfect score for non-fault cases. This indicates a strong ability to identify both fault and non-fault conditions accurately. Recall above 50% is considered good; 79% is very good for a model predicting faults. The prediction timeframe of 15 days in advance can be adjusted to 5 or 10 days as needed.

Our models utilise Health and Criticality Index Bands to categorise cables by health and operational risk, which informs strategic maintenance and operational decisions. The integration of GIS data has been pivotal in refining these models, providing a detailed understanding of geographical influences on cable health.

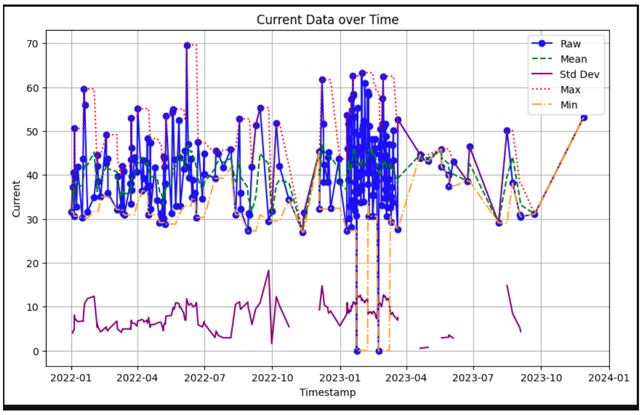

Fig. 1. SCADA current data over time for the entire duration of the model.

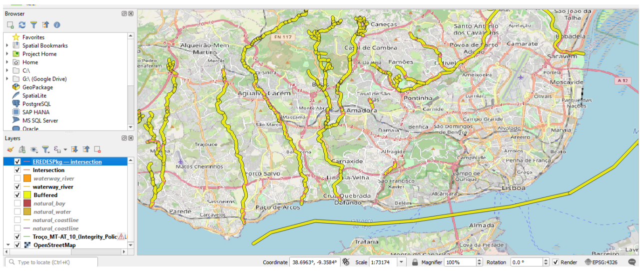

Fig. 2. GIS based analysis of cable sections that have higher probability of failures.

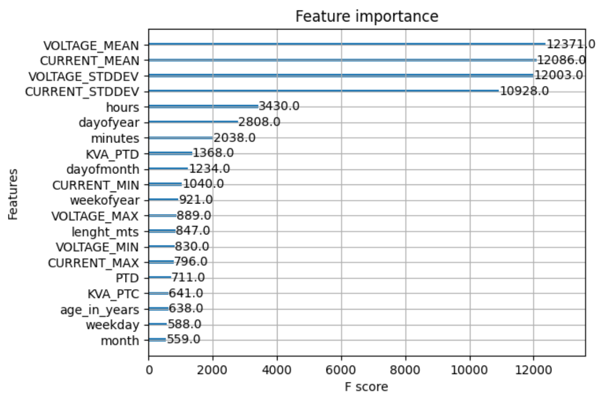

Fig. 3. Weightages of the parameters in the model.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	373062
1	0.79	0.78	0.79	7124
accuracy			0.99	380186
macro avg	0.89	0.89	0.89	380186
weighted avg	0.99	0.99	0.99	380186

Fig. 4. Results (f1 score) of the model.

4 Conclusion

Building upon our successful applications, OrxaGrid plans to expand the scope of our models to encompass a wider range of cable types, including low-voltage networks, and incorporate even more sophisticated predictive techniques, such as reinforcement learning. This evolution aims to further mitigate the impacts of aging infrastructure and effectively accommodate the evolving demands placed upon legacy systems. The real-time data (from OrxaGrid's STEM device), combined with routine test and measurement (T&M) data from diagnostic tests (e.g., partial discharge, sheath current measurements), provides a comprehensive dataset for ongoing diagnostics and predictive analytics. Further, the initiatives undertaken by OrxaGrid establish new benchmarks for the integration of machine learning and GIS within the management of traditional power systems. This technological advancement is poised to significantly transform grid management practices globally, enhancing the resilience and efficiency of power distribution networks while effectively meeting the challenges of future energy demands.

5 Acknowledgements

Gratitude is extended to the European Distribution System Operator for their collaboration on the case study, and to the research teams at OrxaGrid for their support and contributions to this research.

6 References

[1] Kulkarni, Yash, and Akshat Kulkarni. "First of its kind implementation of IOT system in Indian Power Sector." (2019).

[2] Kulkarni Akshat et. al. "A machine learning-induced cable health indexing model for utilities", CIGRE (2024).

[3] A. K. Swain, H. K. Pradhan, and L. M. Patnaik, "Support vector machine based fault diagnosis of power transformers," IEEE Transactions on Power Delivery, vol. 21, no. 1, pp. 404–411, 2006.

[4] J. Jiang, F. Luo, and C. Yang, "Application of artificial neural network in power cable fault diagnosis," IEEE Transactions on Power Delivery, vol. 16, no. 4, pp. 791–796, 2001.

[5] M. A. El-Sharkawi, R. J. Marks, M. J. Damborg, and J. J. Pau, "Artificial neural networks for short-term load forecasting and classification," IEEE Transactions on Power Systems, vol. 8, no. 2, pp. 693–700, 1993.

[6] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[8] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.

[9] S. M. Lee, S. J. Park, and J. H. Kim, "Distributed fiber optic sensor for structural health monitoring," Sensors, vol. 13, no. 11, pp. 14761–14781, 2013.

[10] J. R. Marti, "Accurate modeling of underground cables," IEEE Transactions on Power Delivery, vol. 10, no. 2, pp. 840–850, 1995.

[11] M. P. Moghaddam, A. J. Keane, and M. I. Friswell, "Structural health monitoring: A review," Shock and Vibration Digest, vol. 36, no. 3, pp. 277–287, 2004.

[12] A. Widodo and B. S. Yang, "Support vector machine in machine condition monitoring," Mechanical systems and signal processing, vol. 20, no. 6, pp. 2717–2739, 2006.

[13] Liu, Keyan, et al. "A fault location method of distribution network based on XGBoost and SVM algorithm." IET Cyber-Physical Systems: Theory & Applications 6.4 (2021): 254-264.

[14] Wang, Bo, et al. "The applications of XGBoost in fault diagnosis of power networks." 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). IEEE, 2019.