Since 1996

TechCon® NORTH AMERICA

OT & IT, Balancing CapEx/OpEx using AI

Thierry JUNG, Austin BYRNE,Sihui CHEN

GE VERNOVA

OT & IT, Balancing CapEx/OpEx using AI

Thierry JUNG, Austin BYRNE, Sihui CHEN

GE VERNOVA

Abstract

The impact of Operational Technology (OT) in the transmission and distribution industry has been growing in great parts due to its role in the support for Renewable Energy Integration as it enables operators to manage the variability and uncertainty associated with these sources. Coupled with the growing penetration of On-Line Condition Monitoring systems for traditional primary assets (Power Transformers, Circuit Breakers, GIS, Underground cables), it leads to a growing amount of data being circulated inside substations, between substations and operational centers, dispatch centers, monitoring centers and applications at enterprise-level (Asset Performance Management , Computerized Maintenance Management Systems, etc .). As a consequence, Information Technology (IT) is taking a center stage in the operation and planification of modern networks.
With data exponentially increasing and IT systems multiplying, utilities and industrial customers are looking towards enhanced features offered by Machine Learning (ML) and Artificial Intelligence (AI) as a way to solve the conundrum caused by an exponentially growing amount of data about their network and a decreasing amount of human expertise. The race is on to find the best ways to turn data into actionable, synthetic indices. Obvious applications are Energy load prediction, market price prediction, Distributed Energy Resources predictions. This is where benefits are immediate, data abundant and well documented, so Machine learning algorithms can be put at work fast. Failure prediction, location, power restoration (including extreme weather event preparation) come close after as choice solution for distribution networks.
The commercial range of applications has extended these last years into T&D asset health diagnostic, prediction and Estimated Residual Life (ERL). New research avenues are being opened where big data coming from proxy sources based on usage data (SCADA: load and load history), inventory data (age, manufacturer, technology) and maintenance data (failure history, service history) yield predictions with an accuracy challenging those provided by the traditional approach of sensor and inspection data coupled to FMEA models at a fraction of the total cost.
This paper will explore the potential benefits of such new solutions, based on developments of new algorithms for ERL and DGA fault detection. Discussion will clarify benefits and pitfalls of Machine Learning solutions. Specific emphasis will be given to data set preparation and acceptability of complex models which internal representations are too complex for human understanding, so that diagnostic process is sometimes not transparent, and results lack interpretability. Workarounds that accelerate acceptance of such models within utilities will be discussed.

Authors contacts: thierry.jung@ge.com,austin.byrne@ge.com,sihui.chen@ge.com

1. Introduction

Artificial Intelligence and Machine Learning (AI/ML) is the hottest topic in the technology world, including the industrial scene these last years. From staggering market size expectations to becoming a preferential subject in Cigre Conference these applications are everywhere, and T&D applications gets marketed every week. With research topics going back as far as 1996 [ref 14] the reason for such a recent surge in interest is due to worldwide attention in the advancements in generative AI, which have exceeded expectations in terms of its ability to understand and produce human-like text, generate creative content, and solve complex problems across various industries. Most utilities are now well engaged in their digital transitions, so digital data becomes now available in quantities large and varied enough to fruitfully feed AI/ML solutions, going well beyond the analysis capabilities of traditional tools. Online Condition Monitoring tools have been installed now for a few decades providing relevant timeseries data with attached expert diagnosis. Maybe last but not least, an influx of AI/ML experts are now being produced by universities in large numbers which are being hired into companies. Bringing with them new tools, a unique approach to data analysis and a mindset that shifts workforce skillsets making it easier to adopt these solutions. Add to that the recent trend to simplifies the use of AI/ML by designing no-code algorithms and you get a wide democratization of these solutions.

2. Optimizing system architecture & data sources

Utilities have started rethinking and rearchitecting how data is captured, transmitted and utilized across their organizations, with modern power grids undergoing significant transformation. The integration of Intelligent Electronic Devices (IEDs), advanced Online Condition Monitoring (OCM) systems, and sensors, coupled with a hybrid IT and OT infrastructure, is at the forefront of this digital evolution. These advancements enable utilities to optimize data management, enhance the reliability of grid operations, and improve asset condition assessment.

The need to better control the amount of data transmitted has led to the multiplication of IED's with edge computing capabilities. Applying IED's allows only essential data to be transmitted to central systems, minimizing latency and reducing the bandwidth required for data transfer. The use of MicroML is one technique that can be utilized to achieve the reduction of data transmission. MicroML refers to the implementation of machine learning models on microcontrollers and other resource-constrained devices. This approach is particularly valuable in environments where traditional computing resources are limited or where low power consumption is critical. By deploying machine learning models directly on edge devices, MicroML enables data processing and decision-making closer to the data source. Leading to the desired reduction in latency and bandwidth usage, as data does not need to be transmitted to centralized systems for analysis, only essential data is transmitted. Implementing machine learning on microcontrollers can be more cost-effective than using more powerful computing platforms, particularly in large-scale deployments where hardware costs are a significant factor.

The MicroML deployed on the IED can be trained on the specific parameters required to drive the evaluation and key performance indicators (KPI's) for your asset fleet. Technics using decision tree

ML libraries, can result in complex models. However it can be utilized in a MicroML setting, the model may need to be simplified or pruned to fit the limited computational resources and memory of a microcontroller. Model quantization and pruning can help reduce the size and computational requirements of these models, making them more suitable for deployment on microcontrollers.

The advancement of online condition monitoring (OCM) compute platforms has enabled vendors to take advantage of edge computation, removing the need for dedicated IED's to perform such tasks. The OCM can utilize multiple sensors and its compute platform and provide pre-processed data and information.
As well as deploying OCM systems dedicated to the task of gathering data for the condition assessment of grid assets, organizations are looking at the data generated on existing secondary grid assets such as relays to see how that data can be utilized. Relay data can be instrumental in assessing the condition of power grid assets by providing insights into the operational status and any anomalies detected within the grid. Relays are designed to detect faults in the power system, such as over currents, short circuits, and ground faults. Analyzing the frequency and type of faults can help identify deteriorating conditions in the grid, pointing to potential issues with specific assets. By examining relay operation patterns and cross correlating this information with OCM data, utilities can identify unusual behavior aligned to relay trip events, which may indicate underlying problems with grid assets, such as aging transformers or failing circuit breakers. Relays often maintain detailed logs of events, including the time, nature, and location of faults. These logs can be analyzed to understand the health and performance of grid components, aiding in predictive maintenance strategies. Relay data can provide information on load conditions and variations. Consistent overloading detected by relays may suggest that certain assets are under strain, necessitating inspection or upgrades. By reviewing relay coordination and settings, utilities can ensure that the protection schemes are functioning correctly. Miscoordination or incorrect settings might lead to unnecessary outages or indicate potential issues with protection devices.
Another critical hurdle facing power grids is the data repositories used to store the data and information gathered. In power grids, various data repositories are utilized to host data critical for assessing the condition of grid assets. These repositories store diverse types of data, such as asset design and configuration, measured parameters, operational metrics, event logs, and environmental factors, which are essential for effective condition assessment. Legacy IT and OT architectures have led to segregation of data repositories and their fragmented locations due to departmental ownership and responsibility silos. In order to fully utilize the potential of the data and information stored these silos must be bridged.
Integrating these diverse data repositories into a modern power grid's IT (Information Technology) and OT (Operational Technology) infrastructure requires a cohesive approach to ensure seamless data flow, interoperability, and enhanced grid management capabilities. Modern power grids utilize unified data platforms that aggregate data from various sources, including SCADA, PMUs, IED and OCM devices. These platforms provide a centralized repository for data storage, analysis, and visualization, facilitating integrated decision-making processes. Organizations may opt to use middleware which acts as a bridge between IT and OT systems, enabling communication and data exchange across different platforms and protocols. This integration layer helps synchronize data flows and ensures compatibility between legacy systems and newer technologies.

A hybrid cloud and on-premises IT and OT infrastructure offers a balanced solution that combines the flexibility and scalability of cloud computing with the control and security of on-premises systems. This approach provides several advantages for modern power grids. Hybrid infrastructure

allows power grid operators to scale resources according to demand. Cloud services can handle spikes in data processing and storage needs without requiring significant on-premises investments, while critical operations can remain on-site. By leveraging cloud resources for non-critical workloads and data storage, utilities can reduce operational costs associated with maintaining and upgrading on-premises hardware. This allows for more efficient allocation of capital expenditures. Sensitive data and mission-critical operations can be kept on-premises, where operators have full control over security measures and compliance with industry regulations. Meanwhile, less sensitive data and applications can benefit from the cloud's advanced security features. Data from various sources can be effectively managed and analyzed using hybrid infrastructure. Cloud-based APM solutions can process and derive insights from large datasets for asset fleet condition assessment, while real-time and sensitive data processing can remain on-premises to reduce latency.

Adopting modern approaches to data management described above opens to door for utilities to utilizing AI and ML. By integrating diverse data sources such as IEDs, OCM systems, and sensors, utilities have access to high-quality, comprehensive datasets. This data is essential for training AI and ML models to accurately assess the condition of grid assets. The use of edge computing and MicroML allows for real-time data processing, critical data transmission and decision-making at the source. This ensures that larger AI and ML models running centrally can provide timely insights into asset conditions, enabling proactive maintenance and reducing the likelihood of unexpected failures.

AI and ML models can analyze the historical and real-time data generated to predict potential asset failures before they occur. This predictive capability allows utilities to implement condition-based maintenance strategies, optimizing maintenance schedules and extending asset lifespans. AI algorithms are adept at identifying unusual patterns or anomalies in data. By continuously monitoring asset performance data, AI can alert operators to deviations from normal operating conditions, indicating potential issues requiring attention. By aggregating and analyzing data from multiple sources, AI and ML provide utilities with a holistic view of asset conditions. This comprehensive insight supports informed decision-making, enabling utilities to prioritize maintenance activities and allocate resources efficiently. Some typical applications of AI and ML models will be described in chapters below.

3. What if scenarios to anticipate and predict future behaviors based on past history

Surprisingly enough, few industrial applications have been put to the test for solutions proposing an alternative to diagnosis and prognostics of specific asset health and failure prediction although it has been actively explored since 1996 (ref [14]). Dissolved Gas Analysis (DGA) data from Power Transformers stands out as a prime candidate for Machine Learning solutions because 1/ it is the most common form of PTR data collected by utilities as it represent a good compromise between failure mode detection ,even indirectly and costs 2/ it has been extensively documented by standardisation bodies , and interpretations techniques constantly refined over the last 30 years 3/ DGA monitoring systems have been installed for the last 20 years , thereby building a large database of long timeseries .(ref [1] to ref [5]). These incentives make DGA timeseries the ideal inputs for ML solutions that will integrate past history of PTR into the diagnostic of current condition and extrapolation of future condition, a significant step ahead of current DGA analysis techniques that mostly focus on snapshots of gas content.

GE Vernova started such a project to develop Machine Learning solution to diagnose incipient faults in Power Transformer (PT) together with the University of Laval in Quebec, based on Time series collected mostly from OCM. DGA data from Transformers across various utilities and with their corresponding fault types (see Table 1). Even with the large database collected, failure modes confirmed by experts were unevenly distributed across the samples, with a higher number of samples related to minor faults. Simple oversampling had to be used to achieve balance, as number of faults with fewer observations needs to closely match the ones with higher observations. (See Table 1). Efforts were also undertaken to retain maximal information from the dataset: for instance, addressing missing values in the timeseries was accomplished utilizing the K-Nearest Neighbours (KNN) algorithm. KNN imputation operates on the principle that similar instances tend to have similar values for their features. Additionally, data underwent normalization employing the Min-Max normalization method.

Fault Types	**No. of Samples Before Augmentation**	**No. of Samples After Augmentation**
Energy discharge	9	411
Low thermal faults $T < 300$ ℃	1135	396
Medium thermal faults $T = 300$ ℃	156	405
High thermal faults 300 ℃ $< T < 700$ ℃	252	400
Partial discharge	220	396
Mix of faults	76	409
Normal	26	364

Table 1. DGA fault types and dataset distribution

The Machine Learning algorithm was then trained using the augmented data set (see fig.1). The overall dataset was divided into 70% for training, 10% for validation & 20% for tests

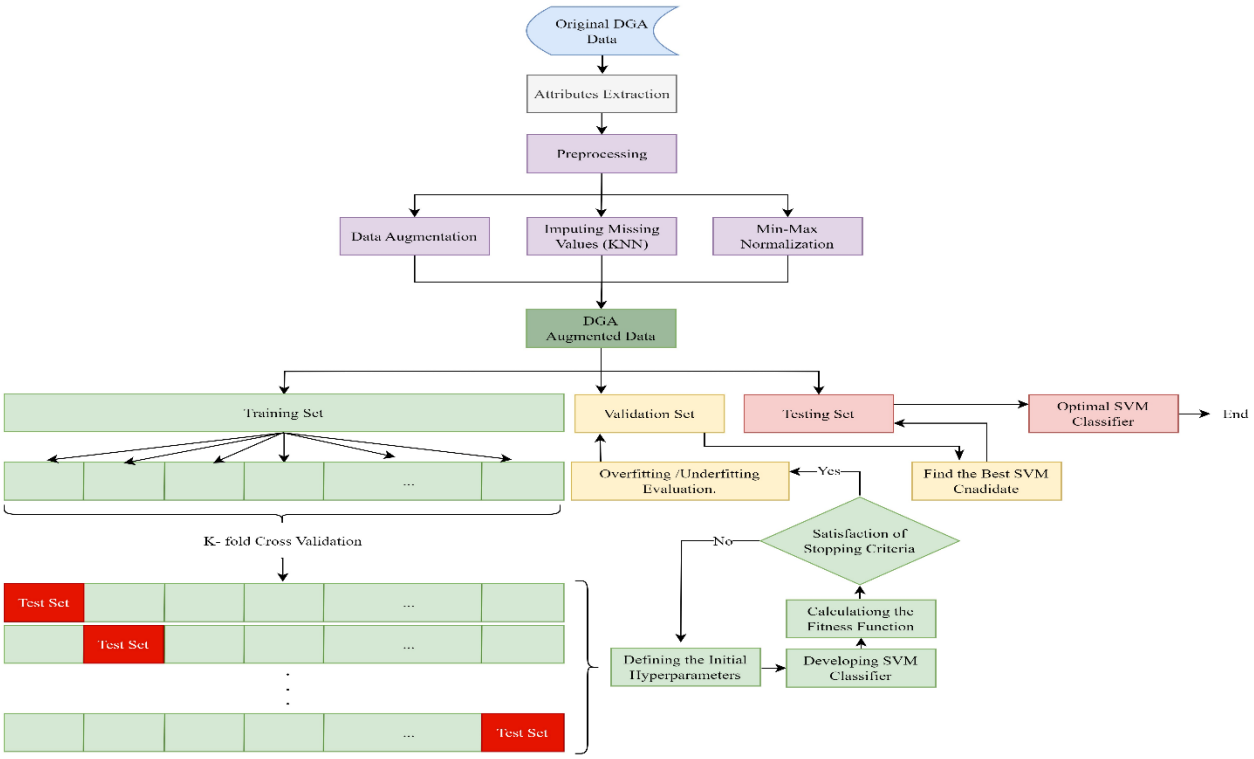

Figure 1. Exemple of Machine Learning training and validation process

Several different ML algorithms including Random Forest (RF),), Support Vector Machine (SVM), LSTM long short-term memory (LSTM) were tested on this set of data. Experimental results are shown below.

As a reminder, common performance metrics to evaluate a ML algorithm are explained as follows.

- *Precision:* The percentage of predicted fault records that are actual anomalies.
- *Recall*: The percentage of the total number of correctly classified anomalous records to the total number of positive records. High recall value specifies the class is correctly recognized with a small number of false negatives.
- *F-Score*: This is the harmonic mean (in percentile) of precision and recall and always has a value near the smaller value of precision or recall. Thus, it provides a more realistic measure of a test's accuracy by using precision and recall. If the costs of false positives and false negatives are very different, F-Score works the best. This is the case for Power Transformer where the cost of one catastrophic failure is huge compared to one unnecessary inspection/maintenance.

Table 2 shows the best results of the evaluation metrics for SVM classifier. It has achieved a considerable overall accuracy (>98%).

Fault type	Precision	Recall	F-Score
Energy Discharge	0.99	1.00	0.99
High Thermal Fault	1.00	0.97	0.98
Low Thermal Fault	0.98	0.98	0.98
Medium Thermal Fault	0.98	0.92	0.95
Mix of Faults	0.92	0.99	0.96
Normal	1.00	1.00	1.00
Partial Discharge	1.00	1.00	1.00
Accuracy	0.98		
macro avg	0.98	0.98	0.98
weighted avg	0.98	0.98	0.98

Table 2: Results of the evaluation metrics

Of particular note, the precision corresponding to all designated class labels in this evaluative context consistently exceed 92%. This implies a limited prevalence of false negatives, thereby signifying the model's ability to accurately recognize positive instances.

Based on these conclusions, the scope of this research is now expanding to predict and simulate DGA status of transformers in the future based on past timeseries.

4. Extracting more value from existing data

When an asset gives causes for concern, additional inspections, analysis or even monitoring systems are used to keep a close eye on the developing situation. Here Machine Learning algorithms can help optimize CAPEX/OPEX spendings on these assets by handling the large number of inputs collected and finding out which are the most relevant ones likely to track catastrophic failure modes. Savings on CAPEX derive from a monitoring device optimized for the most relevant predictors of failures, OPEX saving are derived from reducing field maintenance activities. (see fig 2) . This opens the way to even more optimization of asset maintenance budgets, finetuning inspections at the level of individual assets instead of asset types.

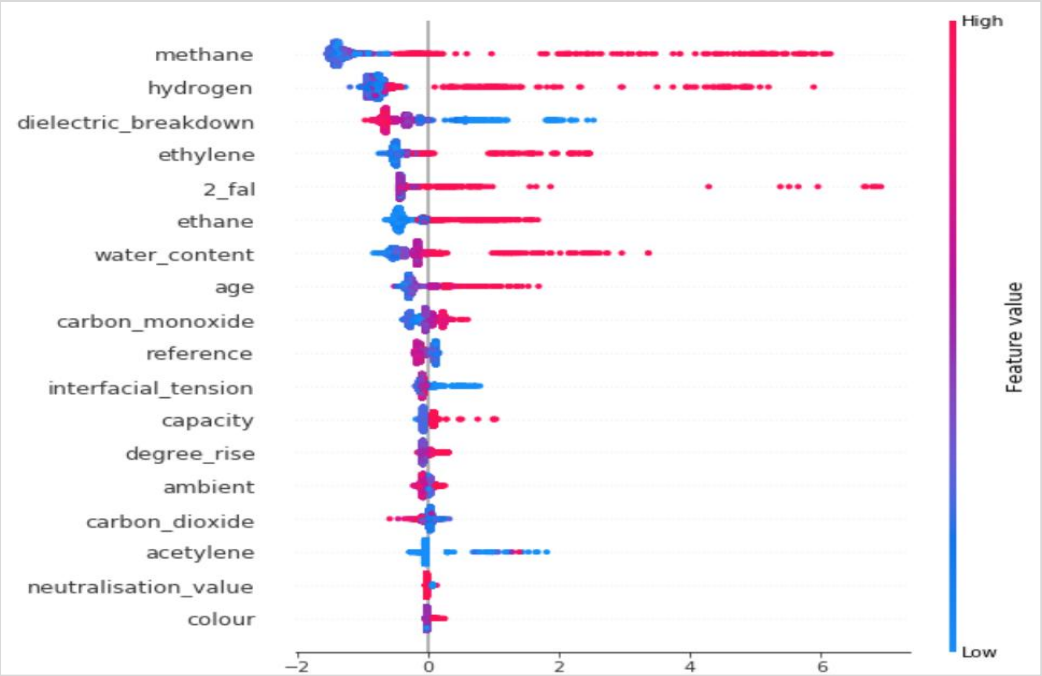

Figure 2. input discrimination

Capex savings can also come from some postmortem analysis. By combing through a vast number of condition parameters and event records, failure modes, maintenance records , and also static attributes (like for instance ,Transformer Cooler type, Transformer Cooling mode ,Transformer substation type, Transformer Power Rating (kVa),Manufacturer, manufacturing factory, vintage, etc,) Machine learning algorithms are able to find out populations and sub populations that will (rep. have behave) last longer than the expected Useful Life . This kind of information is precious for system planning and CAPEX forecasting, but also for sourcing departments in order to steer the sourcing policy towards better products. The survival curves reconstructed and plotted in figure 3 show clearly that the assets attached to the orange plot show way better prospects for long useful life than the assets related the blue plot ….

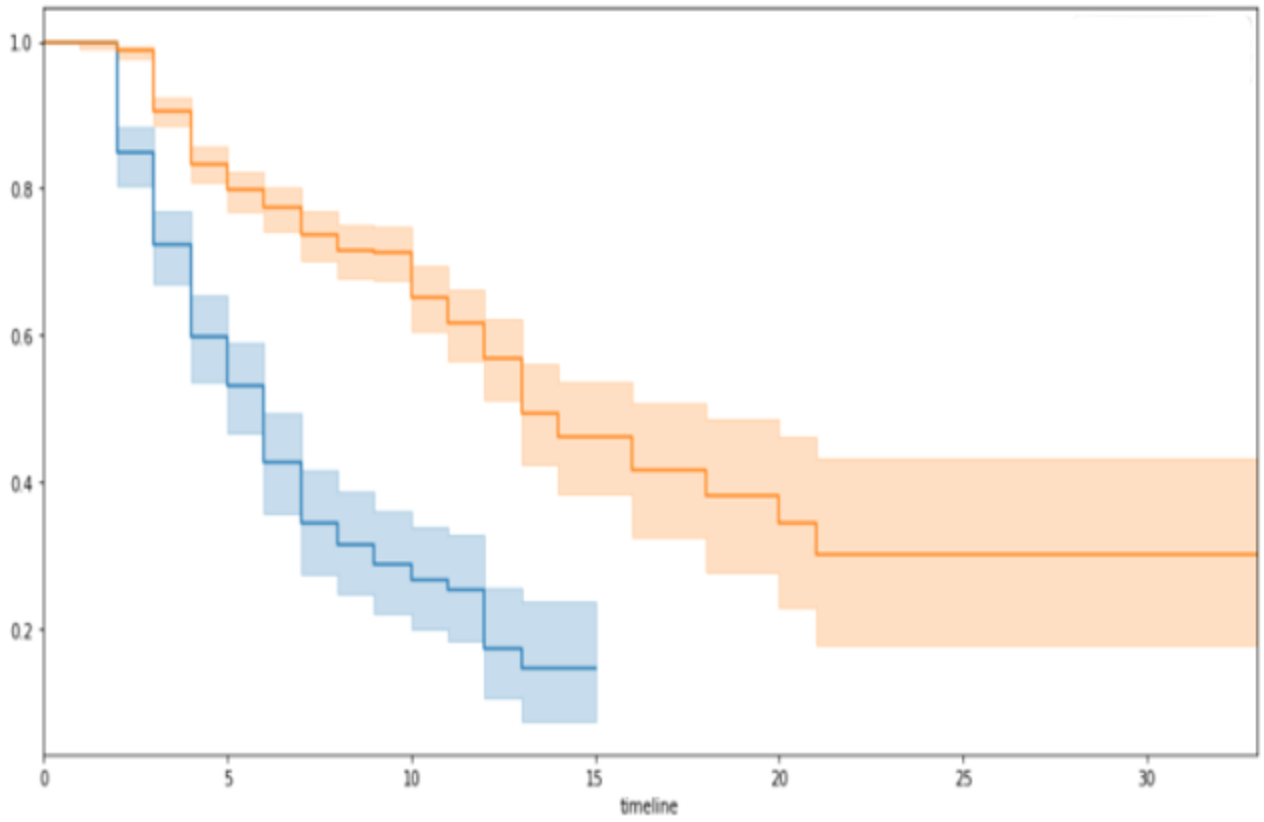

Figure 3. Survival curves

5. Integrating Machine Learning into APM solutions

Despite all the talk about algorithms, human experts are at the core of ML solutions: they will provide the expertise necessary to identify the failure modes attached to the times series, identifies irregularities and tag them to maintenance operations or outages, and then validate the diagnostics and predictions at training stage. As ML solution mature and deploy in the field, they will be most likely responsible for validating the proposed diagnostics and, hence promoting these new solutions. Comparison with established diagnostics methods, mostly FMEA models will remain critical for the acceptance of these solutions too, probably for a long time (i.e there are not that many failures of transformers …) . Therefore, APM solutions will need to propose the coexistence of several diagnostics models (i.e ML and FMEA based) in the same environment .

However successful advanced ML models described above are, as they become more complex and more accurate, they will face the hurdle of the lack of interpretability, “or black box model “. Their internal representations become too complex for human understanding.

Experts may have a hard time explaining their predictions if diagnostic process is sometimes not transparent and results lack interpretability. This could hamper acceptance of such solution in utilities or

extend the “probation period” during which results are systematically evaluated with traditional methods to several years. A possible workaround is to apply eXplainable Artificial Intelligence (XAI) and Interpretable Machine Learning (iML). Such models integrate interpretability with high accuracy for explainable fault diagnosis and failure modes evolution, characterizing the strengths and weaknesses of the decision-making process by making results more understandable by humans. Interpretability is defined here as the ability to understand the contribution of individual features in a model. In particular, such models can capture the nonlinear relationship between dissolved gases and fault types while also modelling the correlations between dissolved gases through interaction terms.

Together with human expertise, the importance of data is crucial for the success of any AI/ML project, but particularly in T&D technical domains . Recent literature shows (Ref [15]): globally one third of the AI/ML projects end in failure. Reasons can be summarized as : insufficient training data , insufficient grasp of AI/ML capabilities and so improper specification stage of the project and lack of follow up between R&D and product development .There is a strong hint that “unrealistic expectations” and “AI did not perform as promised or expected” can be related to insufficient training data sets available or misunderstanding of the efforts and resources needed from the customer side to collect enough quality data provide expected results . For instance, experience has shown that a successful image recognition program would require at least 2000 carefully curated pictures.
Therefore, the most time-consuming phase of creating effective ML algorithms is obtaining, evaluating, and preprocessing datasets for both training and testing stage. Gathering comprehensive datasets that include true positive, true negative, false positive, and false negative diagnostics for early-stage failure detection can be particularly difficult, especially when considering preventive maintenance practices are aimed at reducing critical fault occurrences. It is important to note that training datasets that lack critical diagnostics, lead to sample imbalance and affect the accuracy of the developed models. Here partnerships between utilities, industry working groups and committees, and equipment manufacturers can be the key to unlock the potential of AI and ML.

6. Conclusion

As outlined above, in order to evolve into modern power grids utilities are rethinking their data strategies. The integration of Intelligent Electronic Devices (IEDs), advanced Online Condition Monitoring (OCM) systems, and sensors, alongside a hybrid IT and OT infrastructure, is leading this change. By using IEDs with edge computing, utilities can efficiently process data at the source, reducing latency and bandwidth usage. A key challenge is integrating diverse, fragmented data repositories. Bridging these silos is crucial for unlocking data potential, ensuring seamless data flow, and enhancing grid management. Unified data platforms and middleware are used to aggregate data, ensuring system interoperability. Hybrid cloud and on-premises infrastructure provide a balanced solution, combining cloud scalability with on-premises control and security. These advancements are essential for creating resilient, efficient, and intelligent power grids. As this digital transformation is gathering steam and new data automatically added to all-encompassing digital repositories, sometimes even automatically curated (filtered ,cleaned ,enhanced) by machine learning algorithms ,the amount of ML-ready data will grow exponentially .Together with "No-code ML algorithms" progress (ML algorithms that do not require any coding skills, utilizing visual interfaces and drag-and-drop tools so that a layman can create and deploy ML solutions through a user-friendly platform), it is expected that such an abundance of digital data will lead to a boom in AI and ML application in the T&D industry in the near future .

7. References

[1] Duval, M.; Lamarre, L. *The Duval Pentagon—A new complementary tool for the interpretation of dissolved gas analysis in Transformers*. IEEE Electr. Insul. Mag. 2014, 30, 9–12.

[2] Carlos Parra,Adolfo Crespo Marquez,Vicente et alii,"*Integration of Asset Management Standard ISO55000 with a Maintenance Management Model* " 14th (WCEAM), Singapore 2019

[3] CIGRE, "*Technical Brochure 227 - Life Management Techniques for Power Transformers*".2003

[4] CIGRE, "*Technical Brochure 445 - Guide for Transformer Maintenance*," 2011.

[5] CIGRE, "*Technical Brochure 761 - Condition assessment of power transformers*," 2019.

[6] A. Jahromi, R. Piercy, J. Service, S. Cress and W. FAN, 2009 "*An approach to power transformer asset management using health index*," IEEE Electrical Insulation Magazine, vol. 25, pp. 20-34,2009.

[7] British Standard Institute. 2008. *PAS 55 asset management - Part 1: Specification for the optimized management of physical assets,* British Standard Institute.

[8] ISO/CD 55000.2. 2012. *Asset management – Overview, principles and terminology*.

[9] 55001.2, ISO/CD. 2012. *Asset management – Management systems – Requirements, ISO.*

[10] 55002.2, ISO/CD. 2012. *Asset management – Management systems – Guidelines for the application of ISO 55001*, ISO.

[11] https//www.ofgem.gov.uk/sites/default/files/docs/2021/04/ d*no_common_network_asset_indices_methodology*_v2.1_final_01-04-2021.pdf

[12] *Localising the British CNAIM Standard for Risk Modelling and Risk-based Decision-Makin*g Fin Jennrich, Copperleaf and Claus Lund, CIRED 2021 Conference.

[13] Navin Rego and Nilesh Kulkarni. 2016, *"Condition based Asset Management for Distribution Substations – A Kahramaa Experience*." GCC POWER 2016, Doha, Qatar

[14] "*An artificial neural network approach to transformer fault diagnosis*," IEEE Transactions on Power Delivery, vol. 11, no. 4, pp. 1836 - 1841, 1996.

[15] *Create More Business Value from Your Organizational Data:* Ritu Jyoti Raghunandan Kuppuswamy, Dataiku february 2023.

[16] F*ault diagnosis of Power Transformer based on machine learning approach* ,32nd European Safety and Reliability Conference (ESREL 2022) Dublin, Ireland 28th August - 1st September 2022.

[17] *DGA Fault Prediction in Transformers based on Support Vector Machine* CMDM 2023 1-3 november 2023 Bucharest.