A Data Analytic Approach for Assessing XLPE Cable Insulation Condition via Resistance Measurements

Huajie Yi, Xi Wang, Changyou Suo, Amer M. Y. M. Ghias, *Senior Member, IEEE*, Hoay Ben Gooi, *Fellow*, *IEEE*, Cheng Tian Wee, Wen Kwang Chern, and Abdulkadir C. Yucel, *Senior Member, IEEE*

***Abstract*— The insulation resistance (IR) test has been widely conducted by electricity companies to assess cable insulation health status due to its ease of applicability. However, there exist cases in which medium-voltage power cables pass the IR test but fail in service after re-energization. So far, historical IR data obtained via measurements have been recorded, but have never been systematically studied to improve the accuracy of detecting unhealthy cables. This paper proposes a data analytic approach using historical IR data to identify patterns for distinguishing between healthy and unhealthy cables and assess the cable insulation condition. The proposed approach first leverages a two-parameter Weibull analysis to link the failure probability of the cables to their ages. Such analysis sheds light on the classification of the cables with respect to their age and material. Next, a diminishing method (DM) is used to set the critical IR values and provide maximum detection of the unhealthy cables with minimum misclassification of the healthy cables as unhealthy. Finally, a self-organizing-map-based support vector machine (SOM-SVM) is used to classify the cables as healthy or unhealthy. The hybrid DM-SOM-SVM approach is applied to the historical IR data of 22kV and 6.6kV cross-linked polyethylene (XLPE) cables. Compared to current industrial IR criteria for insulation condition diagnosis, the proposed approach allows detecting 18.5x and 1.8x more unhealthy 22kV and 6.6kV XLPE distribution cables, respectively.**

***Index Terms*—Data analysis, degradation, distribution cables, failure analysis, power cable insulation, support vector machine (SVM).**

I. Introduction

THE cables are the key components of the electricity network. The reliability and resilience of electricity delivery through distribution and transmission networks rely on the stable operation of power cables and the status of their insulations [1]. Various techniques have been proposed to assess the health of cables [2]. Partial discharge (PD), applied both online and offline, detects local defects within the cables insulation [3], but it is sensitive to noise [4]. Dielectric loss (*Tan δ*) measurement is another well-known insulation health indicator, yet its application is constrained by high power requirements for long cables under power frequency [5], [6] and temperature-dependent measurement variations [7]. An improved frequency-domain reflection (FDR) based method is introduced and validated using artificial cable defects in [8]; however, its accuracy in real-world scenarios may be affected by cable accessories and multiple joints.

Insulation resistance (IR) measurement remains widely used in the industry to assess the status of cable insulation. The measurement is usually carried out via a portable, easy-to-operate, and cost-effective IR tester [9]. The IR measurement data can also be obtained by dielectric absorption ratio and polarization index, which can be calculated via the ratio of IR measured at various time instants (i.e., 30s, 60s and 600s) [9]. Additionally, polarisation and depolarisation currents has been used to further assess cable health conditions [11]. Electricity companies commonly evaluate cable insulation health based on IR criteria before re-energizing the cable. However, failures still occur in some medium-voltage (MV) power cables that initially pass the IR test, highlighting the need for a more accurate and reliable IR-based health assessment technique.

In the past, various methods have been explored to improve cable insulation diagnostics. While machine learning (ML) techniques have been widely applied, they have primarily focused on alternative measurement data rather than historical IR values. For example, models such as tree ensembles and support vector machines (SVM) have been used for PD type classification in transformers [12]. Similarly, the researchers used SVM to recognize the PD type based on SF6 and its decomposition products [13]. Fuzzy support vector machine (FSVM) classification algorithm hs been employed to enhance cable health evaluation by weighting key aging factors, (*i.e.*, PD, cable corrosion, annual load, and operating life) [14]. Compared to the SVM, FSVM can allow for assessing the health status of cables more accurately. Deep learning approaches, such as convolutional neural networks (CNN), have been leveraged to classify cable aging states based on dielectric parameter measured under various frequencies [15]. Additionally, feature fusion techniques have been applied to analyze space charge properties in cross-linked polyethylene (XLPE) cables, with radial basis function (RBF)-based SVM demonstrating superior classification performance [16]. Some studies have also focused on post-fault classification rather than pre-fault diagnostics [17]. While these methods highlight the effectiveness of ML in cable health assessment, none have leveraged historical IR data for long-term insulation condition monitoring and predictive health assessment.

Only one study [18] has attempted to incorporate historical IR dataset to cable health assessment. It employed a simple thresholding method using 5,700 historical IR values to establish a maintenance strategy for 11kV distribution cables, identifying an imbalance degree among three phases and a critical IR threshold of 1500MΩ. However, this heuristical approach relies on a fixed threshold that does not account for

key factors such as cable age, type, and length, limiting its applicability to 22kV and 6.6kV distribution networks, MV cables considered in our study.

This paper proposes a hybrid data analytic technique using historical IR data to assess the insulation health of MV (22kV and 6.6kV) distribution cables. As shown in Fig.1, the technique first uses a two-parameter Weibull method to to establish the relationship between cable failure rates and cable age. Based on this relationship, cables are grouped with respect to their ages or types. Next, the diminishing method (DM), leveraging the principle of diminishing returns, is used to obtain high (Secnario I) or low (Secnario II) critical IR values for each group, classifying cables as healthy or unhealthy. While DM acts as a linear classifier, it alone is insufficient for maximizing true detection rates of unhealthy cables while minimizing false detections. To further exploit IR patterns, a self-organizing-map-based support vector machine (SOM-SVM) is leveraged after DM [Fig. 1]. The SOM-SVM first performs re-sampling on data to reduce the imbalance ratio of healthy and unhealthy classes without discarding any useful information. It then classifies cables into healthy or unhealthy states for failure prediction [Fig. 1]. The proposed DM-SOM-SVM approach allows for detecting 18.5x and 1.8x more unhealthy 22kV and 6.6kV cables compared to the current IR guideline utilized by the electricity company, respectively.

A preliminary version of this framework was presented in [19], where rationale for applying the two-parameter Weibull model in cable failure assessment was discussed. In addition, the work in [20] established the link between cable failures, cable type and cable age. However, this study presents a systematic hybrid data analytic method using historical IR measurement values, introducing the DM-SOM-SVM method to assess the health status of cables. For each subgroup, threshold IR values are derived, and ML methodologies are implemented to employed to improve predictive accuracy.

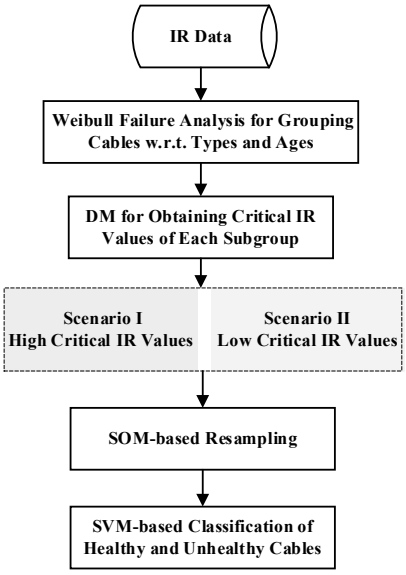


Fig. 1. Flowchart of the proposed data analytic technique.

The main contributions of this work are twofold:

1) A data analytic approach is proposed for the first time to assess the health status of power cables using the IR measurement values. Utility companies can implement the proposed data analytical approach and use it with their own historical IR measurement values to assess the health status of MV cables. The proposed approach demonstrates superior performance compared to the heuristic techniques currently employed by utility companies.

2) Threshold IR values for MV power cables, indicating the cable health status, are systematically studied and presented. Utility companies can use these threshold IR values to quickly screen the health status of their MV cables with a certain accuracy.

The rest of this paper is organized as follows. Section 2 introduces the two-parameter Weibull method to analyze historical failure events and classify cables by age/type. Section 3 introduces the DM method and demonstrates the improvement in detection capability achieved via DM. Section 4 presents the SOM-SVM technique, where datasets are re-sampled after applying DM. Section 5 demonstrates the results of the hybrid DM-SOM-SVM technique, followed by Section 6, which discusses the overall performance of the proposed method. Finally, Section 7 concludes the work.

II. Preliminary Analysis of Failure Events via Weibull Model

In recent years, the Singapore utility company has experienced failures on distribution cables insulated with XLPE. In addition, there were cases of cable failures when the cable circuit passed the IR screening. The health status was then initially determined by the Singapore power company via the critical IR values and the IR unbalance ratio, heuristically derived via the company's years of experience, as shown in Table I. The IR unbalance ratio is the ratio of the maximum IR value to the minimum IR value among the three phases. Generally, cable was classified as unhealthy if the cable failed during operation (Case 1), cable could not pass the IR screening (Case 2 & 3), or cable failed during re-energizing after the IR measurement (Case 4).

The existing IR guidelines assumed can produce fair and reasonable classification. However, it felt short as it was determined heuristically with experience and did not take into account the cables' type, age, and length. Furthermore, it was not derived from a systematic data analytics study on historical data. In theory, cable IR degrades over its operational period due to the growing exposure to environmental, thermal, and electrical stresses. It is inversely proportional to the cable's length. In short, the current approach based solely on analyzing IR without considering the cables' age, type, and length is biased and inaccurate [21].

TABLE I
INITIAL IR GUIDELINES FOR CABLE HEALTHY LABELLING

Case	Minimum IR among 3 phases	IR Ratio	Re-energising	Power Utility Labelling	Labelling in this study
1	Fail during operation			Unhealthy	Unhealthy
2	< 1000 MΩ	N/A	N/A	Unhealthy	Unhealthy
3	≥ 1000 MΩ	> 15	N/A	Unhealthy	Unhealthy
4	≥ 1000 MΩ	≤ 15	Fault	Healthy	Unhealthy
5	≥ 1000 MΩ	≤ 15	Pass	Healthy	Healthy

A. Data Preparation

Apart from IR measurements carried out on-field, the historical IR measurement values obtained since April 2013 are also used in this study. In total, two databases were provided by the utility company: the first database contains operational information such as cable circuit ID, voltage level, cable age, cable type, cable length, and up-to-date IR measurement values; the other database has the records of failed circuits including the information of cable circuit ID, the last IR measurement value, and date of failure. Comprehensive circuit information with IR values was obtained via the following steps:

1) The circuit with missing cable age/type/length information or IR measurement values was discarded.

2) The cable circuit ID was matched up between two databases to find the operational information of the cables. After data clean-up, the IR values of 1642 22kV cables and 1680 6.6kV cables have been included in the dataset.

3) By applying the initial IR guidelines, 96 unhealthy 22kV cable circuits and 72 6.6kV cable circuits have been identified. The large imbalance ratio between the 'Health' and 'Unhealthy' classes of the 22kV and 6.6kV was noted as 16.1 and 22.3, respectively.

B. Specifications and IR Measurement Setup

In this study, 22kV and 6.6kV three-core XLPE cables owned by the Singapore utility company are considered. The structure of these medium voltage power cables is provided in Fig 2. Generally, each cable phase has five layers: conductor core, conductor screen, insulation, insulation screen, and metallic screen. The thicknesses of the cable layers are provided in Table II. Specifically, nominal thicknesses of the insulation for 22kV and 6.6kV power cables are 5.5mm and 2.8mm, respectively.

The IR values of cables were measured under DC excitation by using Megger S1-1568 at the power cable terminal located in a substation. Table III provides the specifications of the Megger S1-1568. The IR measurement accuracy depends on the test voltage and resistance range. The large resistance value at the orders of tera-ohms leads to an extremely low current value, which consequently reduces the measurement accuracy. Megger S1-1568 has a sufficiently broad resistance range and is suitable for IR measurements conducted in this study.

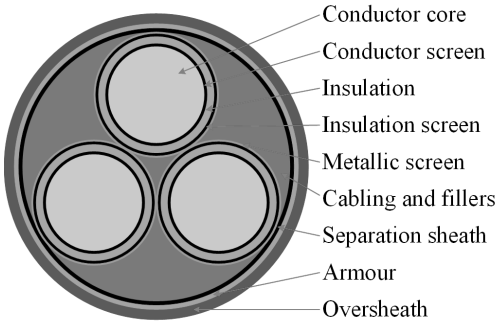


Fig. 2. The structure of MV power cables.

TABLE II
SPECIFICATION OF 22KV AND 6.6KV POWER CABLES

Parameter	22kV	6.6kV
Nominal conductor cross-section area (mm^2)	300	300
Nominal thickness of conductor screen (mm)	0.8	0.5
Nominal thickness of insulation (mm)	5.5	2.8
Nominal thickness of insulation screen (mm)	1.1	1.1
Thickness of metallic screen (mm)	0.4	0.4
Nominal thickness of separation sheath (mm)	2.1	1.9
Nominal thickness of armour (mm)	1.6	1.0
Nominal thickness of PVC oversheath (mm)	3.9	3.4

TABLE III
SPECIFICATIONS OF MEGGER S1-1568

Parameter	Details
Accuracy @10 kV	±5% to 2TΩ ±20% to 20TΩ
Accuracy @ 5 kV	±5% to 1TΩ ±20% to 10TΩ
Resistance range	10 k to 35 TΩ @ 10 kV 10 k to 15 TΩ @ 5 kV
Current accuracy	±5% or ±0.2 nA at all voltages (20 °C)
Current range	0.01 nA to 6 mA
Display range	10 kΩ to 35 TΩ
Real-time output (V, I, R)	Readings at a rate of 1 Hz
Operating temperature range	20 °C to 50 °C
Remote control	Via USB only
Weight	6.5 kg
Dimensions	305 mm × 194 mm × 360 mm

The connection of the test instrument is shown in Fig. 3. During on-field IR measurements, the tests were remotely controlled by the PowerDB Lite software provided by Megger instrument. When the indicator is lit green, the remote control becomes activated. Two test leads are required to make the connection to the test circuit. In detail, one lead, marked as high voltage electrode, was connected to the cable connector lug, while the other lead, called measuring electrode, was clipped on the cable earthing braid. The data was recorded at 1 Hz where the data was transferred to the laptop via a USB cable. With consideration of instrument response time in practice, it usually takes ten seconds to generate a steady current. Therefore, the power was supplied until the 70^{th} seconds and the IR measured at 70^{th} seconds was used in the database. Uncertainties like human error, inconsistent test methods, and missing humidity or temperature data are not considered due to a lack of records.

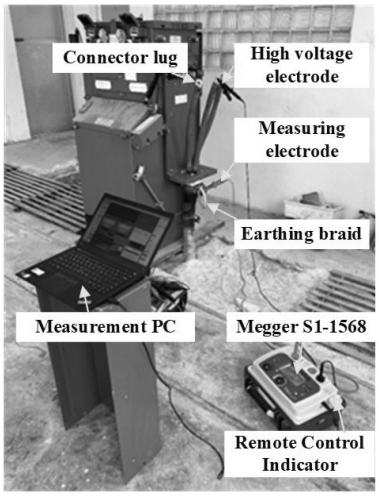


Fig. 3. The connection of the test instrument for on-field IR measurement.

C. The Two-Parameter Weibull Model

The two-parameter Weibull model, as one of the statistical approaches, has been applied to analyze cable failure events and predictions in early 1990s [22]. Based on the two Weibull parameters, the failure distribution for the cable population located in a certain district and operational condition can be evaluated. To perform survival analysis and link the cables' specifications (age and type) to their failure probability, two-parameter Weibull model is introduced as

$$f(t)=\frac{\beta}{\eta}\cdot(\frac{t}{\eta})^{\beta-1}\cdot\exp(-(\frac{t}{\eta})^{\beta}), \tag{1}$$

where t is time, β is the shape parameter, and η is the scale parameter. The parameters β and η are used to model the trend of failure rate within the lifespan. In detail, the value of β reflects the aging rate of a population, while the value of η indicates the characteristic life when 67.3% of the population is failed [20]. Based on (1), the Weibull cumulative distribution function (CDF), $F(t)$, is then provided as

$$F(t)=1-\exp(-(\frac{t}{\eta})^{\beta}). \tag{2}$$

Usually, the Weibull parameters are derived via Maximum Likelihood Estimation. Before applying this approach, the type of data should be classified via a censorship indicator, δ_i, which is

$$\delta_i=\begin{cases}1, & Uncensored\\ 0, & Censored\end{cases}. \tag{3}$$

The first type of data is uncensored data or failed/unhealthy cable data, denoted as '1'. For this type of data, the exact time of each cable failure is known. The other type of data denoted as '0', which will be involved in this paper, is called censored data or survived/healthy cable data. The cable belonging to censored data indicates it is still in operation at the time of measurement [23]. For a cable population whose survival behavior obeys Weibull distribution, the likelihood function can be expressed as

$$L(\beta,\eta)=\prod_{\delta_i} p(t_i), \tag{4}$$

where t_i is the years of service for each observation, $p(t_i)$ is the probability function and subscript i denotes the i^{th} year of the service. The $p(t_i)$ depends on the censorship of each observation and can be expressed via cumulative and probability density functions as

$$p(t_i)=\begin{cases}f(t_i), & Uncensored\\ R(t_i)=1-F(t_i), & Censored\end{cases}. \tag{5}$$

To combine the two censorships in a single expression, δ_i and $(1-\delta_i)$ are moved to the powers of $f(t_i)$ and $R(t_i)$, respectively, as

$$p(t_i)=f(t_i)^{\delta_i}\cdot R(t_i)^{1-\delta_i}. \tag{6}$$

When $\delta_i=1$, $f(t_i)$ contributes to the likelihood function and, when $\delta_i=0$, $R(t_i)$ contributes to the likelihood function. After substituting (1), (2) and (6) in (4), we obtain [24]

$$\begin{aligned}L(\beta,\eta)&=\prod_{i=1}^{n} f(t_i)^{\delta_i}\cdot R(t_i)^{1-\delta_i}\\ &=\prod_{i=1}^{n}\left\{\frac{\beta}{\eta}\cdot\left(\frac{t_i}{\eta}\right)^{\beta-1}\cdot\exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right)\right\}^{\delta_i}\\ &\cdot\prod_{i=1}^{n}\left\{\exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right)\right\}^{1-\delta_i}\end{aligned} \tag{7}$$

Then, by taking the logarithm of both sides of (7), the Weibull parameters β and η are estimated by maximizing the logarithm of the likelihood function $\ln(L(\beta,\eta))$. In this study, the estimation of Weibull parameters via the likelihood function is performed using MATLAB '*wblfit*' function.

D. Application of the Two-Parameter Weibull Model

The applications of the two-parameter Weibull model to the whole 22kV cable data and 6.6 kV cable data are shown in Fig. 4(a) and Fig. 4(b), respectively. In Fig. 4(a), the fitted Weibull distribution has R^2 value of 0.391, which indicates a weak correlation between the Weibull plot of failures and the fitted distribution. The Weibull plot of failures represents the observed unrealiabilities which can be estimated via the adjusted rank method as we presented in the preliminary version of the proposed framework [19]. The application of the model to the overall 6.6 kV cable data (Fig. 4(b)) also shows a poor fit with an R^2 value of 0.767. Because of the poor-fitted results, the three-parameter Weibull model with location parameter (γ) is tested, where the values of β, η, and γ are obtained as 163.7, 1097.9, and -1040.4 for 22kV power cables and -2989.50, 501.9, and 3045.6 for 6.6kV power cables, respectively. The values of β and η for both voltage levels are unrealistic large, while the values of γ are negative. However, the data used in our study do not include any extremely large or negative values, i.e., cable in service for thousands of years or negative 'cable age'. Besides, the concave-up shaped Weibull probability also indicates a negative value of γ if the failure data is fitted by a 3-parameter model [24].

In addition, the goodness-of-fit of other failure distributions is examined. As indicated in [25], commonly used distributions for engineering failure analysis are Gumbel, lognormal, and exponential distributions. The Gumbel distribution is a Type 1 extreme value distribution and is used to model critical and

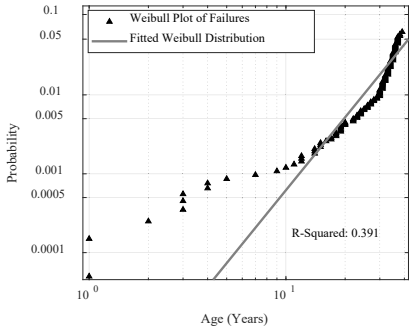
extreme events, such as earthquakes [26], natural disasters, and breakdown in liquids. Thereby, it is not suitable for this study. The lognormal distribution has a poor goodness-of-fit with R^2 value of 0.597 and 0.672 for 22kV and 6.6kV cables, respectively. To this end, it is not considered in this research. The R^2 value for the exponential distribution is 0.381 and 0.003 for 22kV and 6.6kV datasets, and thereby it is not applicable to this study.

Furthermore, fitted Weibull distribution has a corner or dogleg, indicating a mixture of failure modes [24], [25]. Similar trending has been observed for the 6.6 kV case (Fig. 4(b)). Therefore, the two-parameter Weibull model with multiple aging mechanisms or various pairs of Weibull parameters are to be determined for further analysis.

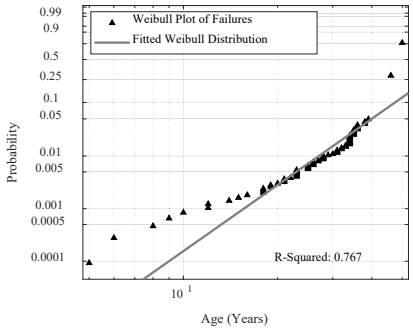
To this end, three different age subgroups for 22kV cable data have been introduced and the two-parameter Weibull model has been applied to data in each subgroup separately. While age group 1 (AG1) includes the data of the 22kV cables with ages from 1 to 9, age groups 2 and 3 (AG2 and AG3) include the data of 22kV cables with ages between 10-23 and 24-40, respectively. It is clear in Table IV that the fitted Weibull distributions have R2 values more than 0.90 for all age groups, while Fig. 4(c) shows a good match between Weibull plots of failures and fitted distributions. A similar grouping has been performed for the 6.6 kV cable data. In particular, data for 6.6kV has been grouped according to the cable types: the ones with copper cores (type 1) and aluminum cores (type 2) (Table IV). In fact, the average age for 6.6kV type 1 cables was around 20 years, whereas that for 6.6kV type 2 cables was nearly 32 years and thus greater degradation has been accumulated for the type 2 cables, which results in higher aging rate or shape parameter. Thereby, it is reasonable to group 6.6kV cables according to their types. While the R2 values for all subgroups are obtained more than 0.96, a good match between the Weibull failure data and distributions is shown in Fig. 4(d).

TABLE IV
CATEGORIZATION OF 22KV AND 6.6KV CABLE DATA VIA WEIBULL ANALYSIS

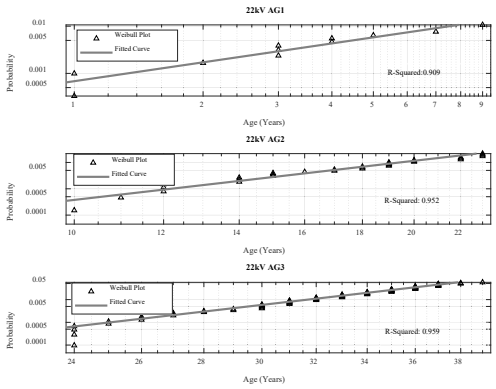
Voltage	Subgroup Information		R^2
	Age Group (AG)	Cable Type	
22kV	Young (AG1)	Copper Core(Type 1)	0.909
	Middle Aged (AG2)		0.952
	Old (AG3)		0.959
6.6kV	All Age	Copper Core (Type 1)	0.972
		Aluminum Core (Type 2)	0.962



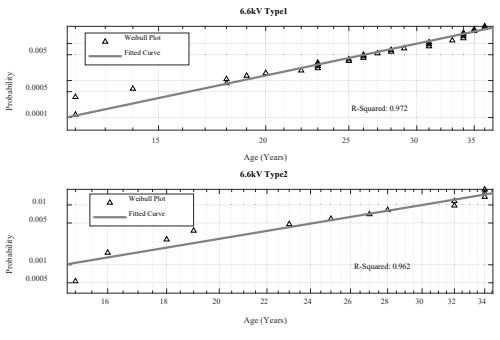
(a)



(b)



(c)



(d)

Fig. 4. Weibull distribution fitting to the Weibull plot of failures when whole (a) 22kV cable data and (b) 6.6kV cable data are considered, and when (c) 22kV cable data and (d) 6.6kV cable data are considered after grouping with respect to age and type, respectively.

III Enhanced IR Criteria Based on DM

The data in each subgroup determined via two-parameter Weibull analysis for 6.6kV and 22kV cable assets is further processed by DM. In specific, a critical IR value used to classify a cable as healthy or unhealthy is determined by the DM applied to the data of each subgroup. Note that these IR values can serve as enhanced IR guidelines for utility companies, yet we use them as auxiliary knowledge in an intermediate step of our proposed data analytic framework. The current guideline used by the utility company is based on a critical IR value (e.g., 1,000MΩ) for all cables at all ages determined by users' experience and heuristics. The cables are classified as unhealthy if their IR values are smaller than the critical IR value. However, the enhanced critical IR values obtained by the diminishing method are determined by considering the historical data for certain cables at certain ages or with certain types. Thereby, the enhanced IR guidelines yield more accurate health prediction/classification than those achieved by electricity companies' current IR guidelines.

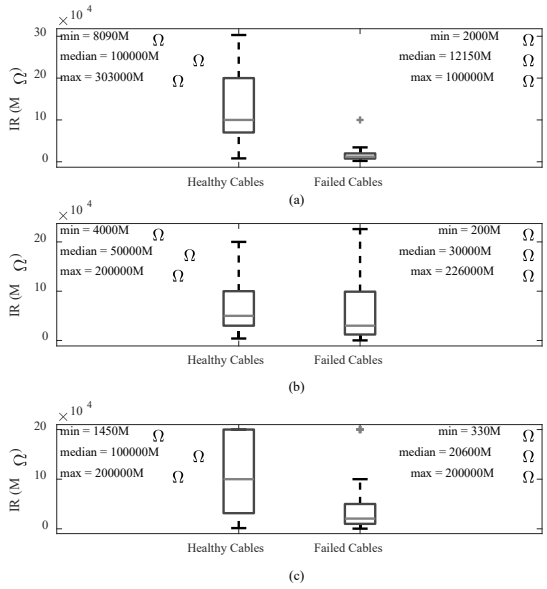


Fig. 5. IR distributions of the healthy and failed 22kV cables for (a) AG1, (b) AG2, and (c) AG3 subgroups.

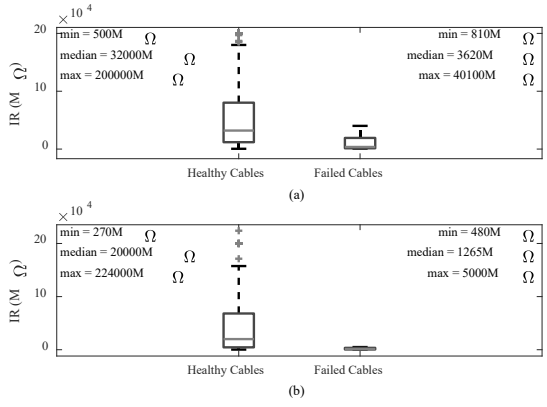


Fig. 6. IR distributions of the healthy and failed 6.6kV cables for (a) type 1 and (b) type 2 subgroups.

Before applying the diminishing method, distributions of historical IR values of all subgroups are analyzed to understand better the IR behaviour of the cables of each subgroup. For each subgroup, IR values are classified as the ones belonging to healthy cables and unhealthy/failed cables. While the IR values of the healthy cables are measured when the cables are operational and healthy, the IR values of the failed cables are the ones measured just before cable failures. Boxplots of the IR distributions of the healthy and unhealthy/failed 22kV cables are provided in Fig. 5. For the AG1 (Fig. 5(a)), after removing the one outlier (the red cross) in failed cables, two classes of data nearly become separable via a prescribed IR threshold. The IR values of failed cables are extremely small compared to those of the healthy ones since the failure at this age period (AG1) is primarily due to the poor quality of manufacturing or wrong installation practices [27]. For the AG2 (Fig. 5(b)), the IR distributions of healthy and failed cables are hard to be separated via a threshold, however, the median of the healthy class is 1.76x higher compared to that of the unhealthy/failed class. For the AG3 (Fig. 5(c)), the majority of the IR data of both classes is separable via a threshold, especially for the healthy cables, no outliers is found. In this wear-out period, the bulk dielectric strength of insulation degrades because of long time exposure to electrical, thermal, environmental, and mechanical stresses. For 6.6kV cable type 1 and type 2 subgroups, boxplots of the IR distributions of the healthy and unhealthy/failed were similar to the 22kV AG1 cables. As illustrated in Fig. 6, after removing the outlier in failed cables, two classes of data nearly become separable via a prescribed IR threshold.

As seen in the previous investigation of IR for each subgroup, a cut-off value for each subgroup can be determined so that cables can be classified as healthy when the IR is higher than the cut-off value. For this purpose, the DM has been used. The DM aims to find a cut-off value for maximizing the detection of unhealthy cables (true detection) while minimizing the misclassification of healthy cables as unhealthy (false detection). As the problem at hand is a classification/detection problem, the metrics used in this classification are as follows: The majority of samples belonging to the healthy cables are counted in the negative class while the remaining minority of samples belonging to unhealthy/failed cables are included in the positive class. Two metrics are defined to assess the true positive detection (of unhealthy/failed) and true negative detection (of healthy cables) as [28]

$$Acc+ \text{ (}True\ Positive\ Rate\text{)} = TP / (TP+FN), \quad (8)$$

$$Acc- \text{ (}True\ Negative\ Rate\text{)} = TP / (TP+FN), \quad (9)$$

where the definitions of true positive (*TP*), false negative (*FN*), true negative (*TN*), and false positive (*FP*) are provided in Table

TABLE V
Confusion Matrix for Classification

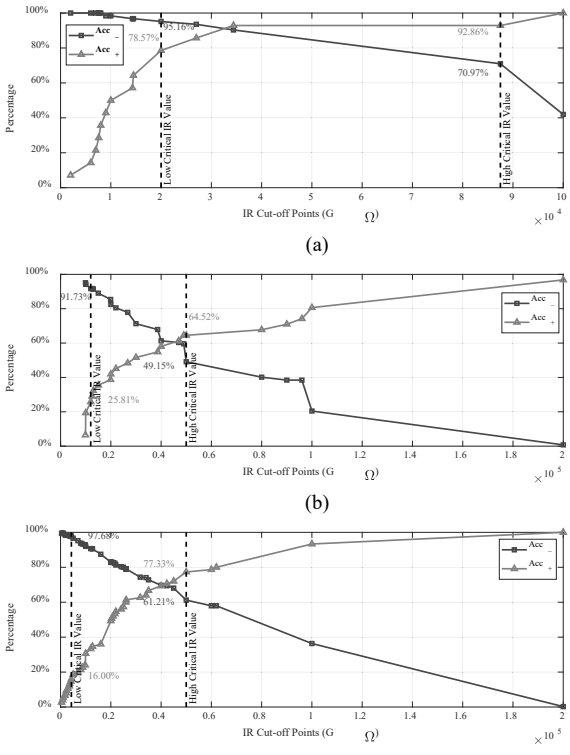
	Predicted Positive (Unhealthy/Failed)	Predicted Negative (Healthy)
Real Positive (Unhealthy/Failed)	True Positive (TP)	False Negative (FN)
Real Negative (Healthy)	False Positive (FP)	True Negative (TN)

V [29]. *Acc*+ and *Acc*– reflect the proportion of correctly identified cables in both failed/unhealthy and healthy cable populations, respectively. In the procedure of DM, first, the IR values of failed cables are sorted in ascending so that the unique values of IR are recorded as useful cut-off points, while the IR values of healthy cables are compared with the useful cut-off points. Then the *Acc*+ and *Acc*– are computed for each useful cut-off point (and plotted in Fig. 7 and Fig. 8). Finally, the optimal cut-off point is determined as the improved critical IR value. It is clear in Figs. 7 and 8 that *Acc*+ and *Acc*– are directly and inversely proportional to the values at cut-off points, respectively.

Following the DM procedure, two different critical IR values, high and low critical IR values, are determined and indicated in Figs. 7 and 8. Selecting a high critical IR value yields detecting more unhealthy cables (at least 70% of the unhealthy cables (*Acc*+ ≥ 70%)), in exchange of misclassifying more healthy cables as unhealthy. (Note: Misclassification of health cables as unhealthy results in a heavy workload of mitigation measures or unnecessary and costly measurements performed after the IR test, which the electricity companies do not desire.) On the other hand, selecting a low critical IR value yields slight improvement in detecting unhealthy cables while ensuring the minimum misclassification of health cables as unhealthy (at least detecting one more unhealthy cable while keeping *Acc*– around 90%). These facts can also be observed in Table VI, which shows the true positive and negative detection rates, *Acc*+ and *Acc*– , achieved after setting low and high critical IR values. Clearly, the accurate detection of the positive (unhealthy/failed) class, *Acc*+, increases with increasing critical IR value while the accurate detection of the negative (healthy) class, *Acc*–, decreases. (Note: *Acc*+ and *Acc*– in Table VI are different from those in Figs. 7 and 8 as those are calculated for the whole cable populations after setting the low and high critical IR values provided in Table VI.) This study's target is achieving more than 70% accuracy in detecting unhealthy/failed cables while minimizing the misclassification of healthy cables as unhealthy, i.e., increasing both *Acc*+ and *Acc*–. Therefore, further effort via an ML algorithm is devoted to increasing the classification accuracy, as explained in the following section.

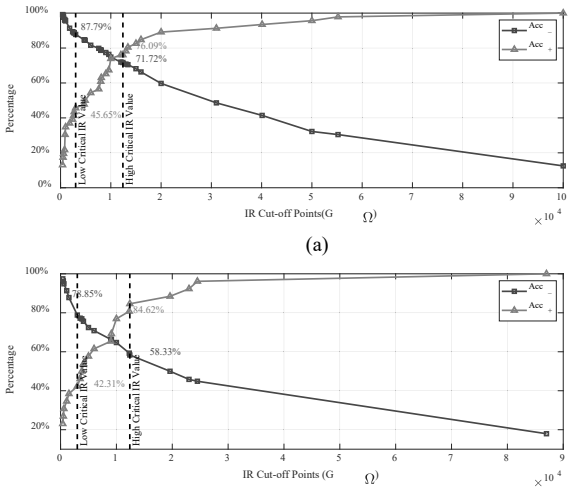
TABLE VI
CLASSIFICATION ACCURACY W.R.T. HIGH AND LOW CRITICAL IR VALUES OBTAINED BY DM

	Group Name	High			Low		
		IR (GΩ)	*Acc*+	*Acc*–	IR (GΩ)	*Acc*+	*Acc*–
22kV	AG1	87.5	59%	70%	20.0	96%	22%
	AG2	50.0			12.0		
	AG3	50.0			4.3		
6.6kV	Type1	12.4	69%	79%	3.0	86%	46%
	Type 2	12.4			3.0		



(a) (b) (c)

Fig. 7. Results of DM procedure for (a) AG1, (b) AG2, and (c) AG3 of 22kV Cables



(a) (b)

Fig. 8. Results of DM procedure for (a) Type 1 and (b) Type 2 of 6.6kV Cables

IV. SELECTION OF CLASSIFIER

For linear separable data, the classification problem can be solved by SVM via finding a hyperplane and maximizing the distance between the support vectors and the hyperplane. For an example of a two-dimensional space, the hyperplane is expressed as $\mathbf{W}^T\mathbf{X} + b = 0$. Here $\mathbf{W}$ denotes the weight vector, $\mathbf{X}$ represents the input vector, and b is the bias. The predicted label of each entry of $\mathbf{X}$ is set to -1 or 1 (belonging to negative or positive classes) if its value is below or above the hyperplane, respectively. When solving the real-world classification

problems, the classes in datasets oftentimes are not linearly separable. To this end, a kernel transformation function is leveraged to create new features based on the features from the input space [30]. Then the hyperplane is optimized in the high dimensional space instead of the original low dimensional input space. A Gaussian RBF is selected as the kernel transformation function in this study. The Gaussian RBF allows SVM to model non-linear relationships between variables, which can result in better classification performance. Furthermore, it requires the adjustment of only two parameters, which can be tuned to fit the specific needs of the real-world problem at hand. In addition, the effectiveness of the Gaussian RBF has been demonstrated in a wide range of applications. It was shown that the Gaussian RBF outperforms other kernel functions [31].

The metrics used to estimate the performance of the classifier are [30]

$$Accuracy = (TP+TN) / (TP+TN+FP+FN), \quad (10)$$

$$Precision = TP / (TP+FP), \quad (11)$$

$$Recall = TP / (TP+FN), \quad (12)$$

Accuracy provides the general ability of the model to predict the correct class. However, when the datasets are unbalanced in nature, *Accuracy* can still be high, but the model fails to correctly predict the class of interest (minority class in our case). *Precision* or *Recall* is recommended to acquire more information on the model's performance. *Precision* is preferred when the cost of FP is higher, and *Recall* is selected when the cost of FN is higher [32]. In our case, the misclassification of unhealthy cable as healthy cable (FN) is more costly compared to the misclassification of healthy cable as unhealthy cable (FP). While FN results in higher cost of power outages, loss of revenue, and the costs from the customers' side, the FP just requires an additional monitoring of the cable with another technique (pre-maintenance). To this end, since the cost of FN is higher than the cost of FP, *Recall* is more significant for our study.

The performance of the SVM with Gaussian kernel classifier is compared with those of the classifiers, including Logistic Regression, Naïve Bayes, KNN, and Decision Tree (DT). The performance metrics of all classifiers are tabulated in Table VII and Table VIII. In this comparison, the ML techniques were trained with the whole datasets of 22kV and 6.6kV. Because of extremely large number of samples in the majority class (healthy class), the *Accuracy* values are noticeably higher than *Recall* values, but *Recall* is still considered the primary metric in our study. For 22kV cable IR dataset, the comparison of the *Recall* among the various classifiers shows that the SVM classifier with Gaussian kernel outperformed the rest of the classifiers. For 6.6kV cable IR dataset, although the *Recall* of Naïve Bayes classifier is the largest, the *Accuracy* of this classifier is the worst, resulting from 108 cases of FPs. The substantial number of FPs results in excessive workload and additional tests for the power utility company. To this end, the SVM classifier is still the optimal option.

The performance of the SVM classifier combined with DM and reduced imbalance ratio will be investigated next.

TABLE VII
PERFORMANCE OF CLASSIFIERS ON THE OVERALL 22KV IR DATA

Name of Classifier	Accuracy	Precision	Recall
SVM with Gaussian kernel	80.1%	78.1%	31.0%
Logistic Regression	74.4%	N/A	0.0%
Naive Bayes	76.6%	76.9%	13.0%
KNN	73.7%	47.6%	25.0%
DT	76.6%	88.9%	10.0%

TABLE VIII
PERFORMANCE OF CLASSIFIERS ON THE OVERALL 6.6KV IR DATA

Name of Classifier	Accuracy	Precision	Recall
SVM with Gaussian kernel	87.4%	83.3%	37.7%
Logistic Regression	82.0%	N/A	0.0%
Naive Bayes	60.9%	29.9%	44.4%
KNN	82.7%	55.0%	20.8%
DT	83.7%	64.7%	20.8%

V. THE SOM-SVM TECHNIQUE

Enhanced IR criteria allow detecting healthy and unhealthy/failed cables more accurately compared to the original IR criteria utilized by the electricity company, yet these criteria are insufficient to achieve high classification accuracy. To improve the accuracy, SVM is applied to the data classes obtained after DM. The SVM is chosen for the binary classification problem being considered as it is accurate and highly robust [33] and it has better generalization capability compared to other classification algorithms [34]. Furthermore, since the solution of the classification problem is obtained by the determined support vectors, SVM provides the globally optimal solution [35]. However, the performance of the SVM, when quantified by *Recall*, deteriorated after the implementation of DM. This is because there exists an imbalance between the numbers of samples in positive (unhealthy/failed) and negative (healthy) classes: The number of samples for the negative (healthy) set is much larger than that of the positive (unhealthy/failed) set. This imbalance gives rise to inaccurate calculation of the training and testing errors of the ML algorithm and it is alleviated by the SOM (self-organizing-maps) in this study, as explained below.

The ratios between the numbers of samples in negative (healthy) and positive (unhealthy/failed) classes are as high as 16.1 and 22.3 for 6.6kV and 22kV cables, respectively. These high imbalance ratios of imbalanced data sets lead to poor classification accuracy. As the performance of SVM depends on the selected support vectors, the removal of samples far away from the decision boundary in the majority class (negative class in our case) will not affect the predictions [29]. To this end, a judicious under-sampling of the majority class should be performed to prevent the misguidance of the ML algorithm and a large number of false negatives. Random under-sampling is the simplest method to re-sample the majority class. However, it randomly eliminates samples and useful information can be discarded, which consequently results in inaccuracy in predictions. On the other hand, SOM avoids eliminating the useful information of the majority class while performing under-sampling. It is an unsupervised clustering technique based on neural networks [36].

The general process of SOM clustering can be described as follows. First, the map size or the number of neurons (cluster centers) is determined. Next, the vectors from the input space are assigned to the neurons with the smallest distance. The

distance is usually evaluated via Euclidean distance. As a result, the vectors associated with the same neuron are from the same clustering on the output layer. The results before and after SOM under-sampling are illustrated for an example in Fig. 9. The SOM algorithm assigns all samples into several clusters. The clusters composed of all samples of the negative class (Cluster 1 and Cluster 2) are removed before applying the SVM.

During the application of SOM, map sizes of 10 and 20 were initially evaluated with three inputs, namely age (in years), length (in meters), and IR values (in MΩ). However, the imbalance ratio after the re-sampling was still as high as 8.1 and 2.6 for 22kV and 6.6kV data, respectively. Then, the map size of 30, corresponding to a grid with 30 rows and 30 columns, was implemented and yielded an acceptable imbalance ratio of 1.1. A similar process has been performed for the 6.6kV dataset. The map size of 30 reduced the imbalance ratio from 22.3 to 1.4. The SOM technique effectively decreased the imbalance ratio and is applicable in this study. The effectiveness of SOM after DM is also examined and will be introduced in the next section.

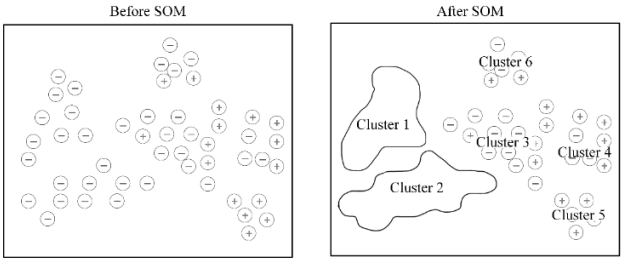


Fig. 9. An example illustration of the SOM under-sampling

VI. The Application of the DM-SOM-SVM Technique

SOM-SVM is applied to data classified by the DM with high critical IR values (Scenario I) and low critical IR values (Scenario II). In Scenario I, a large number of healthy cables are misclassified as unhealthy/failed since the cut-off IR value is high. For such case, the objective of SOM-SVM is to shift some FPs (misclassified healthy cables as unhealthy) to TPs (correctly classified unhealthy cables as unhealthy). Thus, SOM-SVM is applied to the provisional unhealthy class obtained after applying DM. In Scenario II, low critical IR values give rise to the failure to detect unhealthy cables. Hence, SOM-SVM is applied to the provisional healthy class obtained after applying DM.

Based on the original datasets provided by the electricity company, the age (in years), length (in meters) and IR values (in Ohms) of cables are selected as the features for SOM-SVM. Table IX demonstrates the datasets used by SOM-SVM for Scenarios I and II. The imbalance ratios after DM are large, especially when low critical IR values are used.

TABLE IX
Datasets Used By SVM Before and After SOM Under-Sampling

Voltage Level	Scenario	No. of Real Unhealthy/ Failed Cables	No. of Real Healthy Cables		Imbalance Ratio	
			Before	After	Before	After
6.6 kV	I	57	497	69	8.7	1.2
	II	39	1,382	42	35.4	1.1
22 kV	I	74	635	83	8.6	1.1
	II	73	1,489	80	20.4	1.1

Table IX shows the results after applying SOM under-sampling. It can be seen that by using the SOM, the imbalance ratio decreased remarkably and became close to 1, which means that the two classes tentatively have a similar number of samples. The samples of the majority (negative) class are in the proximity of the samples of the positive class in the three-dimensional input space and these samples cannot be linearly separable, which requires Gaussian RBF for the transformation.

A ten-fold cross-validation is applied to the dataset. In particular, the dataset is partitioned into ten parts and the learning process is repeated ten times. For each learning process based on the under-sampling dataset, nine parts of the dataset were used for training and the remaining one part of the dataset was used for the testing process. Before any SVM algorithm is executed, the data is normalized. For each dataset, the model with minimum classification error is used for making predictions with the optimal value of kernel scale for the original dataset.

The classification results of the DM-SOM-SVM scheme are shown for Scenarios I and II in Fig. 10. For 22kV cables, the proposed scheme achieves a higher classification accuracy in Scenario I compared to that in Scenario II. The proposed scheme can detect more unhealthy cables while misclassifying a small number of healthy cables as unhealthy. For 6.6kV cables, although the proposed scheme can achieve an accuracy as high as 94%, it can only detect 7% of unhealthy cables in Scenario I. This is because the size of the dataset used for classification after DM was significantly reduced. Additionally, the IR values of the healthy and failed cables exhibited highly similar distributions, which resulted in the reduced performance of the proposed scheme for this case. On the other hand, the scheme can detect around 60% of the unhealthy cables in Scenario II.

It should be noted that deep learning (DL) methods were not adopted in this study due to the following reasons: 1) The size of the datasets studied in this work is limited. DL models require a larger dataset to effectively learn patterns and prevent overfitting, whereas machine learning (ML) models can better extract insights from smaller datasets. 2) The data in this work were all obtained via on-field IR measurements, which are influenced by test conditions, such as temperature, humidity, and test manners, introducing unstructured noises into the IR data. Unlike DL models, ML models with proper preprocessing (DM & SOM in our study), tend to handle small and noisy datasets more effectively.

VII. Discussion

Based on the previous investigation on cable classification using DM-SOM-SVM scheme, Table X lists the metrics used to evaluate the classification performance. By comparing the values of *Recall* provided with and without SOM, it can be concluded that SOM can significantly improve the performance of SVM while the values of *Recall* can be maintained over 70%. The table also includes classification accuracy achieved by using the electricity company's initial IR guidelines, stating that if the cable's IR is lower than 1,000MΩ, it is unhealthy; otherwise, it is healthy. The confusion charts obtained after applying the initial IR guidelines to the historical data are provided in Fig. 11.

TABLE X
METRICS USED TO EVALUATE THE CLASSIFIERS

Data	Method	TP	FP	*Recall*	*Accuracy*
6.6kV Scenario II	Initial IR guidelines	24	83	33.3%	92.2%
	DM-SVM	33	226	12.7%	82.3%
	DM-SOM-SVM	43	430	59.7%	72.7%
22kV Scenario II	Initial IR guidelines	4	4	4.2%	94.2%
	DM-SVM	13	52	20%	91.0%
	DM-SOM-SVM	74	362	77.1%	76.6%

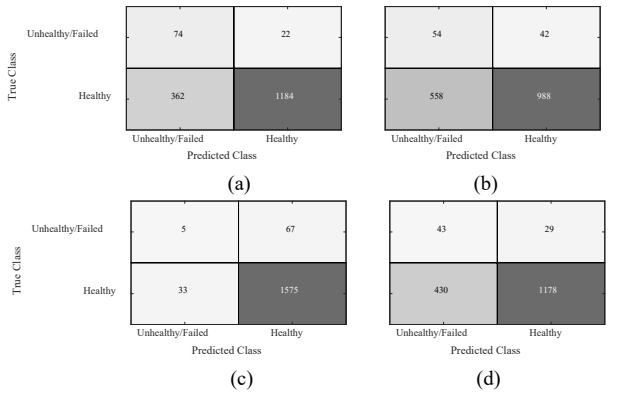


Fig. 10. The results of the classification by DM-SOM-SVM approach for (a) Scenario I and (b) Scenario II of 22 kV cables and for (c) Scenario I and (d) Scenario II of 6.6 kV cables.

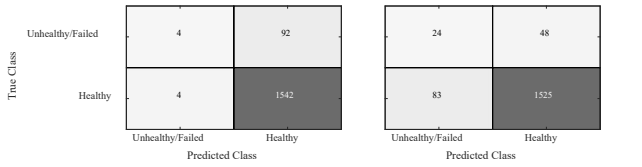


Fig. 11. The classification results obtained via the initial IR guidelines used by the electricity company for (a) 22 kV and (b) 6.6 kV cables.

As clearly seen from the first row ('TP') of Table V, the proposed approach DM-SOM-SVM can detect 18.5x and 1.8x more unhealthy cables compared to the initial IR guidelines for 22kV and 6.6kV cables, respectively. However, as seen from the second row ("FP"), the proposed approach requires follow-up maintenance actions for 362 and 430 of 22kV and 6.6kV healthy cables, respectively. The proposed approach doubled the value of *Recall* for 6.6kV cables and reached approximately 18x higher *Recall* for 22kV cables compared to the initial approach. Due to the large imbalance ratios, the classification performance of the initial approach is misleading, as discussed in Section 4. For distribution power cables, as the cable failure costs many times compared to that of planned maintenance outages, the cost of FNs is greater than the cost of FPs. In such a scenario, Recall's improvement is more significant compared to the other metrics.

Usually, IR values are used as quick diagnostic metrics to assess insulation health. No international standard has been published so far to diagnose insulation health using cable IR values. The value of IR depends on the voltage level and cable length, while it fluctuates with the temperature and humidity of the soil environment. The value of IR can also be significantly affected by the measurement errors. Besides, the IR values of the failed/unhealthy cable circuits used in this study are the last IR readings from on-site measurements before the failures occurred. The other limitation is that this study can be only based on the information that has been provided, other information that can be used as input features, such as polarisation index, dielectric absorption ratio, dielectric loss etc. were not available. Hence, although the datasets investigated in this paper are not as comprehensive as desired, the proposed DM-SOM-SVM method can still differentiate the majority of the unhealthy cables from the healthy ones. The methodology and findings of this study will form the basis for consecutive studies on the health assessment of the distribution cable insulation via IR values.

VIII. CONCLUSION AND FUTURE WORK

This paper proposed a data analytical approach for diagnosing the health condition of the distribution cables via IR values. The approach made use of distinct techniques for judiciously processing the data and performing classification. First, a two-parameter Weibull method was used to link the failure rates to the cables' ages and types and to partition the data into subgroups. Next, DM, a linear classification scheme, was applied to data to obtain enhanced IR cut-off values. The SOM clustering method was used to alleviate the misbalance between the numbers of samples of positive (unhealthy/failed) and negative (healthy) classes. Finally, the SVM classifier was leveraged to diagnose the health of cable insulation. The key input features for the proposed method were the ages, lengths and IR values of cables. The performance of the final classification results has been evaluated based on Recall. Compared to the initial IR guidelines of the electricity company, which uses 1,000MΩ as the critical IR value to estimate cable health status, the enhanced method could detect more unhealthy cables. Although more false positives were generated, the Recall was significantly increased by the proposed approach. In the future, the monetized cost for TP, FP, TN and FN will be considered to improve the method further and optimize the results from a cost-risk perspective. Cable IR, as bulk solid insulation, can only reflect the entire insulation condition. Data on dielectric loss and partial discharge magnitude will be used as additional input features to reflect water moisture and local defects in the cable; Such data will increase the accuracy of the proposed hybrid DM-SOM-SVM model. To address uncertainties arising from data collection procedures, adaptive estimation techniques, as discussed in [37], based on raw data (applied test voltage and measure current) will be further investigated in future work.

ACKNOWLEDGEMENT

This research is supported by SP Group, the National Research Foundation, Singapore, the Energy Market Authority under its Energy Programme (EP Award <EMA-EP010-SNJL-005>) and Nanyang Technological University.

REFERENCES

[1] C. Zhou, H. Yi, and X. Dong, "Review of recent research towards power cable life cycle management," *High Voltage*, 2(3), pp. 179-187, 2017.
[2] H. M. Hashemian, "State-of-the-Art Predictive Maintenance Techniques," *IEEE Trans Instrum Meas.*, 60(1), pp. 226-236, Jan. 2011
[3] S. Govindarajan, A. Morales, J. A. Ardila-Rey and N. Purushothaman, "A review on partial discharge diagnosis in cables: Theory, techniques, and trends," *Meas*, 216, pp. 1-19, Jul. 2023.

[4] G. C. Tone, "Partial discharge. VII. Practical techniques for measuring PD in operating equipment," *IEEE Electr. Insul. Mag.*, 7(4), pp. 9-19, July-Aug. 1991.

[5] IEEE guide for field testing of shielded power cable systems using very low frequency (VLF) (less than 1 Hz), *IEEE Std 400.2-2013*, 2013.

[6] S. Morsalin, T.B. Phung, M. Danikas and D. Mawad, "Diagnostic challenges in dielectric loss assessment and interpretation: a review," *IET Sci. Meas*, 13(6), pp.767-782, 2019.

[7] A. Ghaderi, A. Mingotti, F. Lama, L. Peretto and R. Tinarelli, "Effects of Temperature on MV Cable Joints Tan Delta Measurements," *IEEE Trans Instrum Meas.*, 68(10), pp. 3892-3898, Oct. 2019.

[8] Z. Tang, K. Zhou, P. Meng and Y. Li, "A Frequency-Domain Location Method for Defects in Cables Based on Power Spectral Density," *IEEE Trans Instrum Meas*., 71, pp. 1-10, 2022.

[9] B. Oyegoke, D. Birtwhistle and J. Lyall, "New techniques for determining condition of XLPE cable insulation from polarization and depolarization current measurements," in *Proc. IEEE Int. Conf. Dielectr. Liq.*, Winchester, UK, 2007, pp. 150-153.

[10] H. Torkaman and F. Karimi, "Measurement variations of insulation resistance/polarization index during utilizing time in HV electrical machines – A survey," *Meas*, 59, pp. 21-29, Jan. 2015.

[11] B. Oyegoke, D. Birtwhistle and J. Lyall, "Condition assessment of XLPE cable insulation using short-time polarisation and depolarisation current measurements," *IET Sci. Meas. Technol.*, 2(1), pp. 25–31, Jan. 2018.

[12] W.L. Woon, A. El-Hag and M. Harbaji, "Machine learning techniques for robust classification of partial discharges in oil–paper insulation systems," in *IET Sci. Meas. Technol.*, 10(3), pp. 221-227, May 2016.

[13] J. Tang, F. Liu, X. Zhang, X. Liang and Q.Fan, "Partial discharge recognition based on SF6 decomposition products and support vector machine," *IET Sci. Meas. Technol.*, 6(4), 198-204, Jul. 2012.

[14] X. Wu, Y. Liu, L. Wang, X. Ren and X. Tan, "XLPE cable health assessment based on Relief-F feature weighted FSVM," in *IOP Conf. Series: Earth and Environmental Science*, Xiamen, China, 2021.

[15] S. Quan, S. Gao, Q. Mu, C. Liu, L. Zhou and D. Wang, "Onboard EPR Cable Aging Evaluation by Rectangular-SPP-CNN Based on LMMGS Processing Method," *IEEE Trans Instrum Meas*., 70, pp. 1-10, Sep. 2021.

[16] S. S. Roy, A. Paramane, J. Singh, A. K. Das, S. Chatterjee and X. Chen, "Automated Space Charge Classification Inside ±500-kV HVDC Cable Insulation Using Fusion of Superpixel and Deep Features for Remote Condition Assessment," *IEEE Trans Instrum Meas*., 72, pp. 1-8, 2023.

[17] K. Zhu and P. W. T. Pong, "Fault Classification of Power Distribution Cables by Detecting Decaying DC Components With Magnetic Sensing," *IEEE Trans Instrum Meas*., 69(5), pp. 2016-2027, May 2020.

[18] K. Zhu, "Use of insulation resistance imbalance degree for the condition assessment of power distribution cables," *HKIE Transactions*, 30(4), pp. 2–12, Jan. 2023

[19] H. Yi *et al*., "Enhanced Cable Insulation Resistance Guidelines Based on Weibull Analysis and Diminishing Method," in *Proc. IEEE Int. Conf. on High Voltage Eng. and Appl. (ICHVE)*, Chongqing, China, 2022, pp. 1-4.

[20] H. Yi, I. Hancock, D. Chen and C. Zhou, "An improved Weibull model with consideration of thermal stress for analysis of cable joint failures," in *Proc. IEEE 3rd Int. Conf. Dielectr.*, 2020, pp. 609-612.

[21] Nuclear power plants – Instrumentation and control important to safety -- Electrical equipment condition monitoring methods - Part 6: Insulation resistance, *IEC/IEEE 62582-6:2019*, 2019.

[22] R. M. Bucci, R. V. Rebbapragada, A. J. McElroy, E. A. Chebli and S. Driller, "Failure prediction of underground distribution feeder cables," *IEEE Trans. on Power Deliv*., 9(4), pp. 1943-1955, Oct. 1994.

[23] L. A. Ferreira and J. L. Silva, "Parameter estimation for Weibull distribution with right censored data using EM algorithm," *Eksploatacja i Niezawodnosc – Maintenance and Reliability*, 19 (2), pp. 310–315, 2017.

[24] Abernethy, R.B. *The New Weibull Handbook*, 2nd ed, Blackwell, 1996.

[25] IEC/IEEE guide for the statistical analysis of electrical insulation breakdown data (Adoption of IEEE Std 930-2004), *IEC 62539 First Edition 2007-07 IEEE 930*, 2007.

[26] A. Grous. *Fracture Mechanics. 1, Analysis of Reliability and Quality Control*. Mechanical Engineering and Solid Mechanics Series, London: Hoboken, N.J, 2013.

[27] G. A. Klutke, P. C. Kiessler and M. A. Wortman, "A critical look at the bathtub curve," *IEEE Trans. Rel.*, 52(1), pp. 125-129, Mar 2003.

[28] X. Liu, J. Wu and Z. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern., Part B (Cybern.)*, 39(2), pp. 539-550, Apr 2009.

[29] Y. Tang, Y. Zhang, N. V. Chawla and S. Krasser, "SVMs modeling for highly imbalanced classification," in *IEEE Trans. Syst., Man, Cybern., Part B (Cybern.)*, 39(1), pp. 281-288, Feb. 2009.

[30] H. Belyadi and A. Haghighat, "Chapter 5 - Supervised learning," in *Machine Learning Guide for Oil and Gas Using Python*, H. Belyadi and A. Haghighat, Eds, Gulf Professional Publishing, 2021, pp 169-295.

[31] S. Caner, and F. Dovis, "The impact of different kernel functions on the performance of scintillation detection based on support vector machines" *Sensors*, 19(23), Nov 2019.

[32] A. Subasi, "Chapter 3 - Machine learning techniques," in *Practical Machine Learning for Data Analysis Using Python*, A. Subasi, Ed, Academic Press, 2020, pp. 91- 202.

[33] Iqbal H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN COMPUT. SCI.*, 2(160), pp. 1-21, Mar 2021.

[34] R. Adhikari and R. K. Agrawal, "An introductory study on time series modeling and forecasting," *LAP*, 2013.

[35] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Trans. Neural Netw.*, 14(6), pp. 1506-1518, Nov. 2003.

[36] T. Kohonen,"Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, 43(1), pp.59–69, Jan 1982.

[37] O. Tutsoy, "Design and Comparison Base Analysis of Adaptive Estimator for Completely Unknown Linear Systems in the Presence of OE Noise and Constant Input Time Delay," *Asian J. Control,* 18(3), pp. 1020-1029, May 2016.