Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Investigation of the Influence of Non-Routine and Derived Features in the Development of Early Detection Model for Transformer Health Index Classification

Azlan Mohmad[1,2], Mohd Hatta Mohammed Ariff[2,*], Mohd Ibrahim Shapiai[2], Mohd Solehin Shamsudin[1,2], Norulhakima Zakaria[1], Mohammad Adnan Sujan[1], Rasli Ghani[2], Ifran Bahiuddin[3]

[1] Tuah Energy Sdn Bhd, T4-2-2, Tower 4 Level 2, Maju Link, Jalan Lingkaran Tengah II (MRR2), 57100, Bandar Tasek Selatan, Kuala Lumpur, Malaysia
[2] Malaysia–Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[3] Department of Mechanical Engineering, Vocational College, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia

ARTICLE INFO

Article history:
Received 4 August 2023
Received in revised form 27 December 2023
Accepted 13 January 2024
Available online 12 February 2024

Keywords:
Transformer Health Index; non-routine test; routine test; machine learning; Transformer Asset Management

ABSTRACT

Establishing an effective HI model is challenging because it involves balancing cost, risk, and performance. The currently developed Reduced Features Model (RFM) for the transformer Health Index (HI) prediction may lead to late prediction. The RFM utilised non-routine input features to achieve a high-accuracy model where data availability is the primary concern. Hence, the common goal of Transformer Asset Management (TAM) in achieving acceptable availability and reliability of the transformer may not be achieved. In this paper, the primary objective is to investigate the performance of the HI model by considering routine test features as a baseline for developing the Early Detection Model (EDM). The development of EDM is significant, as the model shall provide a sustainable solution to the utility and plant owners in establishing their TAM strategies. Hence, this paper's case studies include performance investigation using routine, non-routine, and derived features from the routine test. Support Vector Machine (SVM) was used for the prediction modelling, and the model's performance was validated based on a 5-fold cross-validation technique to avoid biases. As a result, it was found that the average accuracy performance of 88.4% was obtained by considering only routine test features during the model validation process. However, complementing the routine test with other features, which were non-routine and derived features, increased the average performance accuracy model to 95.3%. Hence, further development of EDM is feasible and crucial for sustainable TAM solutions.

1. Introduction

Transformers are one of the critical parts of every electrical network [1]. Following IEEE C57.12.80, the transformer function transfers power by electromagnetic induction between circuits at the same frequency, usually with changed voltage and current values. Due to its nature, the

* *Corresponding author.*
E-mail address: mohdhatta.kl@utm.my

https://doi.org/10.37934/araset.39.2.141152

transformer is widely used for various applications, including electrical distribution and transmission network systems.

While the IEEE C57.91 thermal model sets the ideal life expectancy for oil-type transformers at 20.5 years, considering the hottest insulation temperature of 110°C, the transformer may experience various degradation processes throughout this operational period. These processes can stem from factors such as design safety margins, operation, and maintenance practices [2]. The transformer anticipated failure rates become unacceptable if nothing is done since the transformer needs a long time to repair [3]. Also, the statistical analysis found an increase in the failure rate for transformers aged between 30 and 40 [4]. In some cases, replacing the damaged transformer may take more than a year [5].

Given the ageing global transformer infrastructure, ongoing operational and financial constraints, and long replacement times, utility and plant owners seek strategies to optimise transformer performance while minimising costs [6]. The importance of asset management, system safety, and the adoption of advanced technology can provide valuable insights for holistic and efficiency-centred approaches in various industries [7]. The TAM program addresses these challenges, improving asset decision-making and providing greater flexibility in maintenance programs [2]. Nevertheless, establishing an effective TAM program is challenging as it requires a balance between cost, risk, and performance, especially in managing an ageing transformer asset [6]. In the context of the TAM program, the HI facilitates the evaluation of the transformer's health condition, which has been accepted as a standard technique [8].

One persuasive approach to creating a cost-effective solution for the HI assessment is the development of the RFM [9-12]. In the RFM, data preprocessing employs a feature selection method focusing on sensitivity analysis. This process selects a subset of relevant, high-quality, and non-redundant features for building learning models with improved accuracy [13]. Cost reduction is achieved using fewer features to maintain a similar performance as a full-featured model [14]. The main limitation of the RFM is its reliance on non-routine test features, hindering early predictions and consistent health assessments. The non-routine test is conducted on a case-by-case. Additionally, feature selection can be misinterpreted without expert input, complicating the model's industrial application.

The RFM predominantly focuses on DGA, OQA, and Furan as features or parameters for predicting the transformer HI. However, it overlooks the potential of derived features such as incipient faults that can give useful information to enhance the model's predictive reach. Incorporating these derived inputs could substantially refine the HI model's predictions performance. The increasing trend of research into Unsupervised Fault Detection (UFD) techniques was discussed in a systematic review conducted by Sobran and Ismail [15], also highlights the importance of early fault detection systems which useful for monitoring system in engineering fields.

Focusing on risk reduction, opportunity identification, and early process improvement is more effective in achieving a cost-effective solution than relying solely on the RFM. Developing the EDM addresses these RFM weaknesses, using routine test features to ensure data availability, facilitate continuous health assessments, and enable integration with other TAM activities. Unlike RFM, the EDM's routine features allow seamless integration into monitoring systems and workflows, benefiting from existing data sources and test protocols.

This paper investigates the EDM's feasibility by considering input features from routine tests as a baseline. The SVM ML model is developed based on the established case studies. The benchmark datasets from Alqudsi and El-Hag [13] are utilised for the performance validation of the developed SVM models. The findings are expected to be valuable tools in establishing the TAM programs and may provide long-term impacts on sustainability for the utility and plant owner.

The paper is structured as follows: Section 2 discusses relevant research within the topic that led to the problem statement. Then, in Section 3, the proposed methodology is described and followed by Section 4, with the results and discussion. Finally, the conclusion is provided in Section 5.

2. Related Work

The HI represent an effective and practical tool that combines the results of operating observations, field inspections, and site and laboratory testing into an objective and quantitative index, providing the overall health of the asset [16]. According to Rediansyah *et al.,* [17], most of the input features considered as part of HI are based on the insulation system testing, consisting of Dissolved Gas Analysis (DGA), Oil Quality Assessment (OQA), and Furan. The DGA gases include Hydrogen (H2), Methane (CH4), Ethane (C2H6), Ethylene (C2H4), and Acetylene (C2H2), while OQA consists of Water Content, Breakdown Voltage (BDV), Interfacial Tension (IFT) and Acidity as a minimum.

Various techniques have been proposed for selecting the most influential input features for the HI model development. In the study by Ghoneim and Taha [11], the RFM reveals the efficiency of the adopted feature reduction technique. The results show a slight performance difference between the full and reduced-feature input. Employing the Ensemble (EN) method, the obtained HI performance accuracy is 95.9% for the full features and 95.8% for the reducing feature. Three (3) feature selection techniques are proposed: filters, wrapper, and embedded method. The final features chosen for the reduced features were CO2, C2H2, C2H6, C2H4, colour, BDV, IFT, water content, and Furan. This suggests that even simpler models can yield comparable results, making them more resource-friendly.

In another study conducted by Alqudsi and El-Hag [13], a stepwise regression method is proposed for the feature reduction technique. This method led to the selection of six (6) primary input features for determining a transformer's Health Index. These include 1) Furan, 2) IFT, 3) H2, 4) C2H6, 5) C2H2, and 6) Acidity. For this study, eight (8) different ML classifiers were used for the performance validation, and the obtained Mean Accuracy Rate (MAR) was above 90%, where Decision Tree (J48) and Random Forest (RForest) classifiers have the best performance among the eight (8) classifier models. Finally, in Benhmed *et al.,* [10], various reduced-feature approaches are developed to find the most influential input features. Three (3) filter techniques have been proposed consisting of Info-gain (IG), ReliefF, and Correlation Based Feature Selection (CRS). Based on the investigation, it is found that water content, acidity, BDV, and Furan are the most dominant features.

Selecting input features is critical for every HI model. Unfortunately, as researchers relied on the technique and aimed to achieve high performance, it is lacking in addressing the industry practice and recommendations and less concerned about the data availability, as highlighted in CIGRE WG 761. For example, The IFT test is classified into three groups per IEC 60422 standards and is part of the complementary test category. A Furan analysis is recommended by IEC 60599 only when excessive degradation of cellulose-based insulation is suspected. This analysis, which is not part of the standard inspection routine but rather performed on an as-needed basis, aims to supplement the interpretation of DGA and confirm insulation faults. Therefore, due to these two tests' occasional nature and data unavailability, integrating IFT and Furan as part of the HI model is not practically feasible.

Also, the industry recommendation does not support selecting a few input features proposed by Benhmed *et al.,* [10] and Ghoneim and Taha [11], for their RFM, such as oil colour and BDV. Based on IEEE C57.152, no direct correlation exists between the insulating oil colour change and any failure issues. Besides, according to IEC 60422, a high BDV value does not indicate the unavailability of contaminants in the mineral oil. Instead, the BDV indicates its capacity to withstand electrical stress. Therefore, it is supposed that the selected features for the RFM correlate with deterioration factors that cumulatively contribute to the transformer's loss of life.

Based on the provided perspective, the currently developed RFM may not be suitable yet for the actual and sustainable industry application. Furthermore, as highlighted in CIGRE WG A2.18, there is pressure to make savings by reducing maintenance from traditional time-based to condition-based action. Hence, better solutions to meet the economic and financial constraints are necessary as the current RFM leads to late predictions affecting the electrical network's availability and reliability.

3. Methodology

3.1 Framework

The proposed research methodology framework is shown in Figure 1. Five (5) main tasks are conducted to complete this work. Detailed descriptions of each block (i.e., conducted task) are described in the following sub-sections.

Fig. 1. Research methodology framework

3.2 Obtained Dataset and Label

The oil sample datasets, including HI labels, are obtained from Alqudsi and El-Hag [13], and are referred to as Util1 and Util2. Both obtained oil samples are categorised as medium voltage distribution transformers [13]. The transformer for Util1 is rated at 66/11kV with a capacity ranging between 12.5MVA to 40MVA. On the other hand, the transformer for Util2 is rated at 33/11kV with a capacity of 15MVA. Due to a limited dataset, particularly for the "Fair" and "Poor" categories, oil sample populations of both Util1 and Util2 were combined in this work. The summary of the dataset distribution is shown in Table 1.

Table 1
Summary of transformer data per HI categories

HI Status	Nos of sample		Total	
	Util 1	Util2	Nos	%
Good	496	238	734	70 %
Fair	206	84	290	27 %
Poor	28	5	33	3 %
Total	730	327	1057	100%

3.3 Dataset Preparation

The oil sampling dataset from Alqudsi and El-Hag [13], consists of DGA, OQA, and Furan. Five (5) gases are available for DGA that are H2, CH4, C2H6, C2H4, and C2H2. On the other hand, six (6) measured oil parameters are available for OQA: Water Content, Acidity, BDV, and IFT. The obtained dataset's input features are further segmented into routine and non-routine tests, and derived features are also established from the available routine dataset. The new derived features consist of the DGA Factor (DGAF), OQA Factor (OQAF), and Fault Code (FC).

Based on the established background defined in Section 2.0, although the BDV is part of the OQA routine test, this feature is not considered part of the model development as the industry does not recommend this feature. Figure 2 summarises the input features block diagram considering routine, non-routine, and relevant derived features.

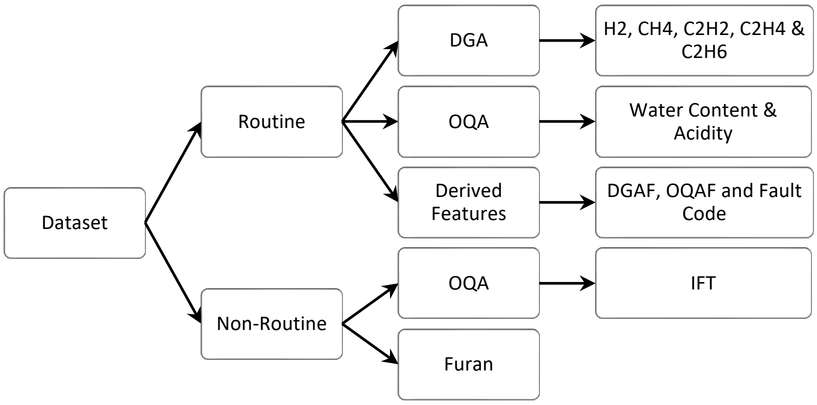

Fig. 2. Block diagram on input features

3.3.1 Derived features for fault code

IEEE C57.104 recommended a few fault identification methods, such as Roger Ratios, Doernenburg ratios, Key Gases, Duval Triangle (DT), and Duval Pentagon. This paper computes the fault code from the DT One method. Also, the standard, theoretically, the DT One utilises three gases to correlate faults' increasing energy content or temperature. This method identifies the six basic types of faults, plus mixtures of electrical/ thermal faults in zone DT. The basic type of fault consists of Partial Discharge (PD), Discharge of Low Energy (D1), Discharge of High Energy (D2), Thermal Fault below 300°C (T1), Thermal Fault between 300°C and 700°C (T2) and Thermal Fault above 700°C (T3).

The predicted incipient fault from the DT One method is later segmented according to its FC. The assigned fault codes are 1) Fault Code "0" for PD, 2) Fault Code "1" for T1, 3) Fault Code "2" for T2, 4) Fault Code "3" for T3, 5) Fault Code "4" for DT, 6) Fault Code "5" for D1 and 7) Fault Code "6" for D2. Figure 3 illustrates the DT One method.

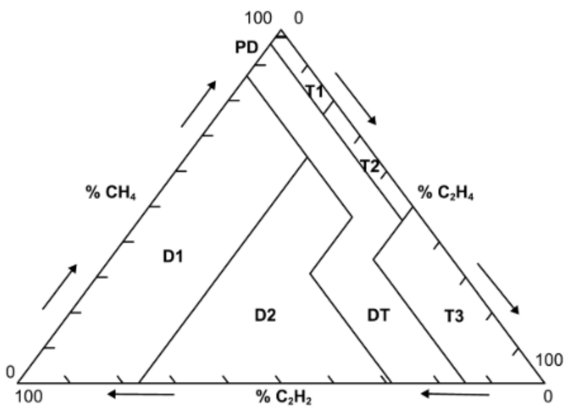

Fig. 3. Duval Triangle One method

3.3.2 Derived features for DGAF

In Jahromi *et al.,* [16], a ranking method for the DGA gases known as DGAF has been developed using the available DGA data (H2, CH4, C2H2, C2H4 & C2H6). The DGAF is computed based on Eq. (1)

$$DGAF = \frac{\sum_{n=1}^{5} S_i \times W_i}{\sum_{n=1}^{5} W_i} \quad (1)$$

Where S_i = 1,2,3,4,5 or 6 is the scoring value, W_i is the assigned weighting factor, and n is defined as the number of features (i.e. H2, CH4, C2H2, C2H4 & C2H6). S_i and W_i for the gas levels (ppm) are summarised in Table 2, while the rating code based on the obtained DGAF score is summarised in Table 3. This was also a similar approach utilised by Alqudsi and El-Hag [13].

Table 2
Scoring and weight factor for gas levels (ppm) [16]

Gas	Score (S_i)						Weight (W_i)
	1	2	3	4	5	6	
H2	< 155	< 225	< 365	< 585	< 700	> 700	2
CH4	< 103	< 145	< 240	< 400	< 600	> 600	3
C2H6	< 92.5	< 95.5	< 96.5	< 97.5	< 100	> 100	3
C2H4	< 75	< 85	< 95	< 105	< 130	> 130	3
C2H2	< 5	< 15	< 25	< 35	< 60	> 60	5

Table 3
DGAF code summary [16]

Rating Code	Condition	Score
4	Good	DGAF < 1.2
3	Acceptable	1.2 ≤ DGAF < 1.5
2	Need caution	1.5 ≤ DGAF < 2
1	Poor	2 ≤ DGAF < 3
0	Very poor	DGAF ≥ 3

3.3.3 Derived features For OQAF

The ranking method for the oil test parameters, known as OQAF, has also been developed by Jahromi *et al.,* [16]. For this paper, the computation of the OQAF was performed based on water content and acidity only. The OQAF is derived based on Eq. (2)

$$OQAF = \frac{\sum_{n=1}^{2} S_i \times W_i}{\sum_{n=1}^{2} W_i} \quad (2)$$

where S_i = 1,2,3 and 4 is the scoring value, W_i is the assigned weighting factor, and n is defined as the numbers of features (i.e. acid and moisture). The grading (i.e., scoring) method for the oil quality test parameter, as adopted by Jahromi *et al.,* [16], is shown in Table 4. The grading method applies to transformers rated U < 69 kV. Finally, the OQAF rating code, similar to DGAF, is assigned based on the OQAF scoring and summarised in Table 3. Alqudsi and El-Hag [13] also utilise a similar approach.

Table 4

Oil quality grading For U < 69 kV [16]

Oil quality	Score (Si)				Weight (W_i)
	1	2	3	4	
Acidity	≤ 0.05	0.05 – 0.1	0.1- 0.2	≥ 0.2	1
Moisture (ppm)	≤ 30	30 - 35	35 - 40	≥ 40	4

3.4 Case Study Development

A few case studies were developed to investigate the feasibility of developing the EDM by utilising routine features as a baseline. These proposed case studies were intended to evaluate the model's effectiveness, including investigating each input's strengths and weaknesses, seeking patterns, and trending the data. Also, the input features considered for each case study are summarised in Table 5. For Case Studies 2 and 3, as both IFT and acidity have the same knowledge, only IFT was considered part of the input features.

(i) Case Study 1 (CS-1): Investigate the effectiveness of input features considering transformer routine tests.
(ii) Case Study 2 (CS-2): Investigate the effectiveness of input features considering transformer routine and non-routine test features.
(iii) Case Study 3 (CS-3): Investigate the effectiveness of input features considering transformer routine, non-routine, and derived features from the available routine test data.

Table 5
Summary of input features considered for each case study

No	Case Study	Routine test					Non-Routine Test	
		OQA			DGA	Derived	OQA	Paper insulation
		BDV	Acidity	Water	H2, CH4, C2H6, C2H4, C2H2	OQF, DGAF	IFT	Furan
1	CS-1	No	Yes	Yes	Yes	No	No	No
2	CS-2	No	No	Yes	Yes	No	Yes	Yes
3	CS-3	No	No	Yes	Yes	Yes	Yes	Yes

3.5 ML Model Development

As defined in Section 1.0, the SVM method is selected as a tool for the performance comparison. Moreover, SVM is a supervised ML Algorithm that can be used for classification problems. Besides, the proposed method was also utilised by Ghonemi and Taha [11], Alqudsi and El-Hag [13], and Ashkezi *et al.,* [18], where this technique requires minimum parameter setting. Hence, suitable for this paper's study, where the main objective is to investigate the input features selection for developing the EDM.

As shown in Figure 4, SVM finds the best-separating hyperplane to maximise the margin between data samples. The filled circles represent the support vectors, while the unfilled circles represent the support vectors. SVM creates a hyperplane to separate samples from different classes. The hyperplane is constructed using the training datasets and used as a classifier for a new sample to determine the true class of each tested sample. The kernel function is popular for identifying the hyperplane [11]. The structure of the SVM classification model with input-output consideration and SVM model parameter setup are shown in Figure 5 and Table 6, respectively.

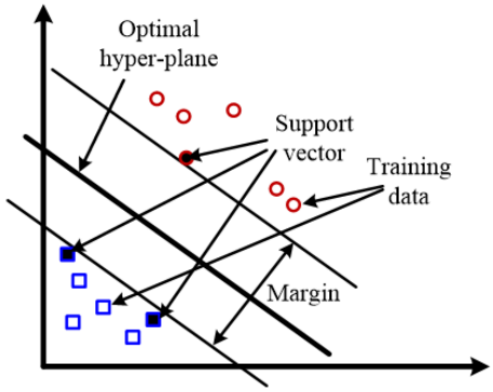

Fig. 4. Separation between two classes by SVM [11]

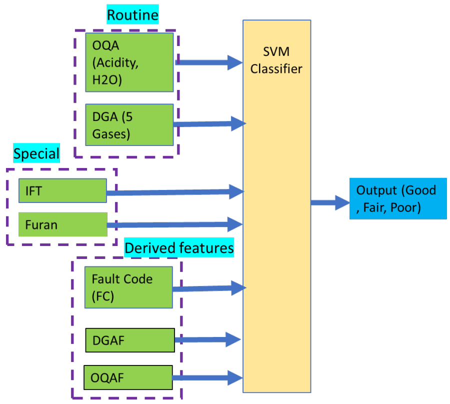

Fig. 5. Model structure for SVM model for various case studies

Table 6
SVM model parameter setup

No	Classifier	Parameter/s
1	SVM	Regularization = 12 norm, Loss = Square hinge

3.6 Training and Validation of the SVM Model

The developed SVM model is trained using 80% of the available dataset, and the remaining 20% is set for validation purposes. Zeinoddini-Meymand and Vahidi [19] also utilise a similar configuration. Besides that, the testing parameter of 20% is selected due to the limited data distribution, especially for the "Poor" categories, which govern only 3% of the total data distribution.

The cross-validation technique was implemented to deter overfitting. Cross-validation is an ML model evaluation technique that encompasses training multiple machine learning models on subsets of the available input data and assessing them on the complementary subset. Therefore, a 5-Fold cross-validation technique is implemented, similar to the selection in Kari *et al.,* [20].

3.7 Performance Evaluation

Evaluating the SVM model is a critical step. This study uses classification accuracy and confusion matrix to evaluate the model performance. Classification accuracy is the ratio of the number of correct predictions to the total number of input samples. At the same time, the confusion matrix gives a matrix as output and describes the complete performance of the model. In addition, the confusion matrix provides insight into the errors produced by the classifier and the types of mistakes produced.

4. Result and Discussion
4.1 Result

Table 7 summarises the obtained result for all three studied cases compared to the benchmark in terms of model accuracy performance. The benchmark accuracy performance published by Alqudsi and El-Hag [13], also obtained from the SVM method, was adopted for performance comparison purposes. The average performance accuracy from the benchmark result is 88.4%. The benchmark result utilises all input features from the available dataset. In CS-1, by exploiting only routine features, the average accuracy is slightly lower than the base case, 87.7%. Interestingly, the obtained minimum accuracy for CS-1 was slightly better than the benchmark result. In CS-2, the average accuracy was increased by applying routine and non-routine tests to 93.3%. Finally, in CS-3, further model improvement is observed, where the average accuracy was increased to 95.3%.

Table 7
Results summary in percentage

	Performance accuracy (%)			
	Min	Max	Avg.	Std. Dev.
Benchmark (with full features) [13]	85.6	92.6	88.4	3.0
CS-1	86.3	89.2	87.7	1.2
CS-2	92.0	94.8	93.3	1.3
CS-3	94.8	96.7	95.3	0.8

Besides performance accuracy obtained from the developed case studies, a confusion matrix was also established. The confusion matrix information is essential to provide the performance of the developed model. Table 8-10 describe confusion matrix results obtained from CS-1, CS-2, and CS-3. Here, the efficiency per class for each confusion matrix class was also calculated for in-depth analysis of the obtained model performance, particularly in classifying the targeted HI label.

Table 8
Confusion matrix for CS-1

	Predicted label			Efficiency per class
	Good	Fair	Poor	(%)
Good	139	8	0	94.6
Fair	15	40	3	69.0
Poor	0	3	4	57.1

Table 9
Confusion matrix for CS-2

	Predicted label			Efficiency per class
	Good	Fair	Poor	(%)
Good	143	4	0	97.3
Fair	8	47	3	81.0
Poor	0	2	5	71.4

Table 10
Confusion matrix for CS-3

	Predicted label			Efficiency per class
	Good	Fair	Poor	(%)
Good	145	2	0	98.6
Fair	4	52	2	89.7
Poor	0	3	4	57.1

4.2 Analysis and Discussion

Three (3) case studies have investigated the effects of routine test features as a baseline for developing the EDM. In CS-1, the average accuracy was 87.7%, compared with the benchmark result of 88.4%. However, no details on the model setting parameters, such as training and validation data distribution, were mentioned by Alqudsi and El-Hag [13] for the benchmark result. Hence, a fair comparison might be arguable as far as optimised model selection is concerned. Nevertheless, since the main objective of this work is to investigate the influence of input features selection for HI determination, the model tuning task was not considered, and targeted comparison only focused on the selection of numbers and nature (i.e., routine, non-routine, and derivation) of input features.

By adopting a default setting of SVM model parameters setup (i.e., as in Table 6), the obtained accuracy from CS-1 demonstrated that the model that utilises only routine features has promising performance compared with the benchmark model, which utilises all the available features. Therefore, it can be concluded that features such as BDV and colour, which are not recommended by industry practice or using features with similar knowledge, such as acidity and IFT, may not necessarily improve the model. Utilising these features could introduce extraneous or 'noise' data that may distort the true pattern in the dataset. Nevertheless, the obtained confusion matrix from CS-1, Table 8, shows that further improvements in the "Fair" and "Poor" classes are required as the obtained efficiency is only 69.0% and 57.1%, respectively. Hence, consideration for the non-routine test is further investigated.

In the second case study (i.e. CS-2), the average accuracy improved from 87.7% to 93.3%. It shows that the non-routine input features, i.e., Furan and IFT are essential for predicting the transformer HI. The confusion matrix from CS-2, Table 9, also shows further improvement in the "Fair" and "Poor" classes as the obtained efficiency increases from 69.0% to 81.0% and 57.1% to 71.4%, respectively. Unfortunately, as the features of IFT and Furan are part of a non-routine test, the data is unavailable, preventing the utility and plant owner from providing continuous health assessments. Besides that, conducting the test are also expensive. As per Prasojo *et al.,* [14], the normalised price for conducting the Furan test from five (5) providers is 0.89. Also, based on this Alqudsi and El-Hag [21], the estimated cost to conduct IFT per sample is US$115. However, a few researchers have successfully predicted the IFT and Furan [17,22]. Hence, a similar approach may be considered for developing an early detection model, especially for the IFT and Furan condition.

Meanwhile, in CS-3, the average accuracy improved from 93.3% to 95.3%. It shows that the derived features from the routine test, consisting of DGAF, OQAF, and Fault Code, can provide additional knowledge and complement other input features. The confusion matrix from CS-3, Table 10, also shows further improvement in the "Good" and "Fair" classes as the obtained efficiency increases from 97.3% to 98.6% and 81.0% to 89.7%, respectively. Only the "Poor" class shows a drop in efficiency from 71.4% to 57.1%. Although the "Fair" class shows improvement in the model prediction, there is a possibility for the model to be misclassified the "Fair" as "Good." For developing a robust early prediction model, misclassifying a superior class for an instant, "Fair" as "Good" or "Poor" as "Fair" should be avoided. Misclassification in the dataset may be due to its distribution. For instance, in Table 1, the "Fair" category represents around 27% of the data, while the "Poor" category is about 3%. This imbalance can favour predicting the overrepresented class, leading to more misclassification of the minority class. Therefore, balancing the dataset can help alleviate these issues. Despite this, the technique could be improved – exploring ensemble ML methodologies might offer further enhancements due to their inherent flexibility.

5. Conclusion

In conclusion, this paper has successfully investigated the influence of routine, non-routine, and derived features in developing the EDM for the HI classification.

From the dataset acquired by Alqudsi and El-Hag [13], three (3) case studies have been simulated. The SVM ML technique was selected as a tool for the performance comparison, and a 5-fold validation technique has been used for diagnosing model overfitting. The obtained result from this paper shows that the EDM for HI classification by considering routine features as a baseline is feasible, provided the listed consideration is carried out:

i. Avoid using input features with similar knowledge or input features not recommended by the industry, as the features do not improve the model and might cause noise in the dataset.
ii. Developed prediction model for IFT and Furan as it provides essential knowledge to the model. The prediction model shall utilise input from the routine test to avoid late predictions. Hence, it allowed for continuous assessment by the utility and plant owner.
iii. Mitigate the unbalance dataset issues, especially for "Fair" and "Poor" categories, to avoid any bias in the model prediction. If the unbalanced dataset could not mitigate for "Fair" and "Poor" classes, the early prediction model is proposed to predict the worst class. This allowed the model to predict accurately, especially for the minority class.
iv. Explore other techniques, as a single ML classifier may not be suitable for predicting the "Fair" and "Poor" categories.

Acknowledgement
This work was supported by Industry/International Incentive Grant (IIIG) Universiti Teknologi Malaysia (No Vote: Q.K130000.3643.03M70), Malaysia-Japan International Institute of Technology (MJJIT) Universiti Teknologi Malaysia under MJJIC-ICATAS 2022 Conference Fees Assistance and Centre of Artificial Intelligent and Robotic (CAIRO) Universiti Teknologi Malaysia.

References
[1] Bhatia, Neha Kamalraj, Ayman H. El-Hag, and Khaled Bashir Shaban. "Machine learning-based regression and classification models for oil assessment of power transformers." In *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pp. 400-403. IEEE, 2020. https://doi.org/10.1109/ICIoT48696.2020.9089647
[2] Singh, A., and A. G. Swanson. "Development of a plant health and risk index for distribution power transformers in South Africa." *SAIEE Africa Research Journal* 109, no. 3 (2018): 159-170. https://doi.org/10.23919/SAIEE.2018.8532192
[3] Diwyacitta, K., Rahman A. Prasojo, and H. Gumilang. "Effects of lifetime and loading factor on dissolved gases in power transformers." In *2017 International Conference on Electrical Engineering and Computer Science (ICECOS)*, pp. 243-247. IEEE, 2017. https://doi.org/10.1109/ICECOS.2017.8167142
[4] Tenbohlen, S., J. Jagers, and F. Vahidi. "Standardized survey of transformer reliability." In *2017 International Symposium on Electrical Insulating Materials (ISEIM)*, vol. 2, pp. 593-596. IEEE, 2017. https://doi.org/10.23919/ISEIM.2017.8166559
[5] Murugan, Raji, and Raju Ramasamy. "Understanding the power transformer component failures for health index-based maintenance planning in electric utilities." *Engineering Failure Analysis* 96 (2019): 274-288. https://doi.org/10.1016/j.engfailanal.2018.10.011
[6] Guy, Paul, and Brian Sparling. "Transformer asset life extension-When, why and how to apply continuous condition monitoring systems." In *2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia)*, pp. 779-784. IEEE, 2019. https://doi.org/10.1109/GTDAsia.2019.8715957
[7] Khattak, Muhammad Adil, Tuan Mohamad Hakimi Tuan Yahya, Muhammad Wazif Mohd Sallehhudin, Muhammad Izzuan Mohd Ghazali, Nur Awal Abdullah, Nur Athirah Nordin, and Nur Ainida Mohd Mohsein Dzakurnain. "Energy Security Policy Shift of North America and Ontario, Canada Following 2003 Power Blackout: A Review." *Progress in Energy and Environment* 4 (2018): 14-24.
[8] Kittan, Stefan, Stefan Kornhuber, Peter Kästel, Gunnar Nitsche, Gerd Valtin, and Michael Weise. "Review and implementation of transformer health index methods in line with the development of a condition assessment tool." In *2018 International Conference on Diagnostics in Electrical Engineering (Diagnostika)*, pp. 1-4. IEEE, 2018. https://doi.org/10.1109/DIAGNOSTIKA.2018.8526034
[9] Alqudsi, Alhaytham Y., and Ayman H. ElHag. "A cost effective artificial intelligence based transformer insulation health index." In *2017 3rd International Conference on Condition Assessment Techniques in Electrical Systems (CATCON)*, pp. 108-111. IEEE, 2017. https://doi.org/10.1109/CATCON.2017.8280194
[10] Benhmed, Kamel, Abdelniser Mooman, Abdunnaser Younes, Khaled Shaban, and Ayman El-Hag. "Feature selection for effective health index diagnoses of power transformers." *IEEE Transactions on Power Delivery* 33, no. 6 (2017): 3223-3226. https://doi.org/10.1109/TPWRD.2017.2762920

[11] Ghoneim, Sherif S. M., and Ibrahim B. M. Taha. "Comparative study of full and reduced feature scenarios for health index computation of power transformers." *IEEE Access* 8 (2020): 181326-181339. https://doi.org/10.1109/ACCESS.2020.3028689

[12] Abu-Elanien, Ahmed E. B., and M. M. A. Salama. "Evaluation of transformer health condition using reduced number of tests." *Electrical Engineering* 101, no. 2 (2019): 357-368. https://doi.org/10.1007/s00202-019-00785-x

[13] Alqudsi, Alhaytham, and Ayman El-Hag. "Application of machine learning in transformer health index prediction." *Energies* 12, no. 14 (2019): 2694. https://doi.org/10.3390/en12142694

[14] Prasojo, Rahman Azis, Heri Sutikno, Nur Ulfa Maulidevi, and Soedjarno P. Bambang Anggoro. "Power Transformer Oil Measurement Prioritization based on Technical and Economic Criteria using Analytic Hierarchy Process." In *2021 3rd International Conference on High Voltage Engineering and Power Systems (ICHVEPS)*, pp. 547-551. IEEE, 2021. https://doi.org/10.1109/ICHVEPS53178.2021.9601037

[15] Sobran, Nur Maisarah Mohd, and Zool Hilmi Ismail. "A Systematic Literature Review of Unsupervised Fault Detection Approach for Complex Engineering System." *Journal of Advanced Research in Applied Mechanics* 103, no. 1 (2023): 43-60. https://doi.org/10.37934/aram.103.1.4360

[16] Jahromi, A., Ray Piercy, Stephen Cress, J. Service, and Wang Fan. "An approach to power transformer asset management using health index." *IEEE Electrical Insulation Magazine* 25, no. 2 (2009): 20-34. https://doi.org/10.1109/MEI.2009.4802595

[17] Rediansyah, Dhanu, Rahman Azis Prasojo, Suwarno Suwarno, and A. Abu-Siada. "Artificial intelligence-based power transformer health index for handling data uncertainty." *IEEE Access* 9 (2021): 150637-150648. https://doi.org/10.1109/ACCESS.2021.3125379

[18] Ashkezari, Atefeh Dehghani, Hui Ma, Tapan K. Saha, and Chandima Ekanayake. "Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers." *IEEE Transactions on Dielectrics and Electrical Insulation* 20, no. 3 (2013): 965-973. https://doi.org/10.1109/TDEI.2013.6518966

[19] Zeinoddini-Meymand, Hamed, and Behrooz Vahidi. "Health index calculation for power transformers using technical and economical parameters." *IET Science, Measurement & Technology* 10, no. 7 (2016): 823-830. https://doi.org/10.1049/iet-smt.2016.0184

[20] Kari, Tusongjiang, Wensheng Gao, Dongbo Zhao, Kaherjiang Abiderexiti, Wenxiong Mo, Yong Wang, and Le Luan. "Hybrid feature selection approach for power transformer fault diagnosis based on support vector machine and genetic algorithm." *IET Generation, Transmission & Distribution* 12, no. 21 (2018): 5672-5680. https://doi.org/10.1049/iet-gtd.2018.5482

[21] Alqudsi, Alhaytham, and Ayman El-Hag. "Assessing the power transformer insulation health condition using a feature-reduced predictor mode." *IEEE Transactions on Dielectrics and Electrical Insulation* 25, no. 3 (2018): 853-862. https://doi.org/10.1109/TDEI.2018.006630

[22] Shaban, Khaled Bashir, Ayman H. El-Hag, and Kamel Benhmed. "Prediction of transformer furan levels." *IEEE Transactions on Power Delivery* 31, no. 4 (2016): 1778-1779. https://doi.org/10.1109/TPWRD.2016.2521320