Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Machine Learning Techniques' Comparison for Early Detection Model of Transformer Health Index Classification

Md Nasif Rahman[1,*], Mohd Hatta Mohammed Ariff[1]

[1] Advanced Vehicle System iKohza, Electronic Systems Engineering Department, Malaysia-Japan International Institute of Technology Universiti Teknologi Malaysia, Kampung Datuk Keramat, 54100 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 14 February 2025
Received in revised form 7 March 2025
Accepted 15 July 2025
Available online 11 August 2025

Keywords:
Machine learning; early detection model; health index; support vector machine

ABSTRACT

Timely detection of anomalies in power transformers is essential to maintain a constant, uninterrupted supply of electricity. Early Detection Model (EDM) is a Machine Learning-based approach to derive the Health Index (HI) of power transformers using routine test features, where the absence of non-routine test features is compensated by use of derived features. Latest research in this area involves the use of only Support Vector Machine (SVM) as the classifier for constructing the sub-models of EDM. This paper explores the use of two other classifiers, Random Forest and Naïve Bayesian to deduce a comparative analysis among all three classifiers. The classifiers were developed in Google Colab platform using already available datasets. The datasets were divided into three parts for training, validation and testing. The results of the simulation showed that Naïve Bayesian achieved the highest average accuracy for all the sub-models, which was 92.45%. On the contrary, Random Forest and SVM appeared to have almost similar level of performance with average accuracies of 89.73% and 89.53% respectively. Naïve Bayesian also seemed to have the lead in terms of other performance metrics as well, making it stand out as the most preferred classifier in the experiment conducted. Further research on this topic may reveal more optimized application of Machine Learning models for early detection of transformer anomalies.

1. Introduction

Power transformers are crucial components for transmission and distribution of electricity. In Malaysia, more than 30 power stations, utilizing renewable and non-renewable resources, have been built in Melaka, Negeri Sembilan, Perak, Pahang, Terengganu, Kelantan, Johor, Sabah and Sarawak [1]. Hence, it is essential to monitor the overall health status of transformers to maintain and maximize their longevity. High voltage power transformers (HVPTs) are generally designed to operate for 20 to 35 years, but with adequate maintenance, their lifespan can be extended to 50 or even 60 years [2]. One of the approaches to maintaining transformer longevity is the health index (HI), which is a unified metric that integrates operating observations, field inspections and laboratory tests to support the Transformer Asset Management (TAM) cycle [3,4]. The transformer health index has

* *Corresponding author*
E-mail address: nasif2000@graduate.utm.my

https://doi.org/10.37934/ard.142.1.152158

become a key metric for assessing the condition and resilience of transformers within a system, drawing attention from asset owners and international organizations like CIGRE and IEEE DEIS/PES [5].

As of now, several different methodologies have been explored to classify different categories of HI in different research works. One of these is the use of Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) by Zeinoddini-Meymand and Vahidi. Both models yielded satisfactory outcomes but the study did not cover all aspects of transformer health including certain environmental factors [6]. Foros and Istad developed a combination of a physical winding degradation model, a condition model and a statistics-based end-of-life model. The assessments from this work were unnecessarily uncertain and had to be used alongside underlying data, assumptions and uncertainties [7]. Another promising approach was a statistical tool called the Mahalanobis-Taguchi System (MTS). The MTS exhibits exceptional predictive reliability when applied to limited and clustered datasets [8]. Luo *et al.,* [9] proposed an evaluation method for power transformers based on the MTS. However, although it managed to eliminate problems such as fuzzy boundary definition and strong subjectivity of selecting parameters, it was built on datasets and calculations under less flexible conditions [9]. Another recent study by Aziz *et al.,* [10] utilized Feed Forward Neural Network, a category of ANN equipped with the three training techniques Levenberg–Marquardt (LM), Bayesian Regularized (BR) and Scaled Conjugate Gradient (SCG) to determine transformer HI. This study uses a dataset of only 106 transformers from a single petrochemical plant, which limits the generalizability of the results to broader populations or other types of transformers. [10].

A promising method for timely HI classification is the Early Detection Model (EDM), an ensemble comprising of machine learning classification models, which ideally uses routine test features to overcome data unavailability [11]. Machine learning refers to the ability of intelligent systems to utilize training data tailored to specific problems, enabling them to automate the development of analytical models and perform related tasks. Deep learning, an advanced branch of machine learning, relies on the framework of artificial neural networks [12].The effectiveness of machine learning in identifying anomalies is especially apparent in its application of neural networks, notably ANN, for diagnosing conditions such as heart disease, cancer, diabetes, and malaria [13]. In addition, Convolutional Neural Networks (CNNs) have gained significant attention due to their exceptional performance in various computer vision tasks, including visual object classification, object detection, and segmentation. CNNs enable early and accurate diagnosis of challenging conditions like Alzheimer's disease by classifying brain MRI images with remarkable precision [14]. While neural networks deliver exceptional performance, this study opts to use traditional machine learning (ML) classifiers for several reasons. Unlike neural networks, traditional algorithms excel with smaller datasets, making them ideal for scenarios where collecting extensive data is challenging. Additionally, they can often be trained using standard computing resources without relying on advanced hardware like GPUs, offering a more cost-effective and accessible solution. From a computational perspective, traditional models typically involve simpler calculations, allowing for significantly faster training compared to deep learning models. Moreover, methods like decision trees or linear regression provide greater transparency, offering clear insights into how decisions are made and simplifying the interpretation of results [15].

In power transformer applications, routine tests are prioritized for monitoring and maintenance. These tests are conducted on every manufactured unit in a production lot [16]. The benefit of using routine test features for EDM development rather than non-routine test features is that the model can allow early abnormality detection. The EDM is also flexible enough to be integrated with other TAM strategies [11]. Early detection through machine learning classification models is increasingly valued in domains beyond power systems engineering. For instance, Muthu and Palaniappan conducted a study employing Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Decision Tree (DT), and Logistic Regression for the early detection of gastric carcinoma. These algorithms were chosen particularly due to their demonstrated effectiveness in processing diverse datasets and their ability to identify complex patterns [17].

According to the experiments by Mohmad [18], an EDM comprises of multiple sub-classification models which can be constructed using SVM. The developed EDM showed improved accuracy performance in comparison to its benchmark RFM models that used routine and non-routine features. However, it did not address the concern of whether it is possible to obtain even better results if other classification algorithms were used instead of SVM. This paper aims to investigate the results of reconstructing the sub-models within the EDM using two widely used classifiers, Random Forest and Gaussian Naïve Bayesian Classifier and compare them with those of SVM. The study uses the same datasets and benchmarks outlined in [18].

2. Methodology

The datasets of oil samples used for this research, same as the ones used in [18], were taken from [11]. The data corresponded to the oil samples of a transformer referred to as Util1, rated at 66/11kV and Util2, rated at 33/11kV. The capacity of Util1 was within the range of 12.5MVA-40MVA while that of Util2 was 15MVA. The combination of two datasets was essential to overcome the limited number of data points. The general framework of the whole research methodology has been divided into the following three sub-sections.

2.1 ML Model Structure

The structure of the whole EDM along with its sub-models as developed in [18] are shown in Figure 1. The sub-blocks in the diagram other than the one labeled as HI-Model are the Early Classifier (EC), Derived Features (DF), Furan (FU) Model and the Interfacial Tension (IFT) Model. As shown in Figure 1, the EC uses the routine features to classify them as Good and a mix of Fair and Poor. For further classification of either Fair or Poor, the derived features from the remaining blocks are used. Each classification model would be executed through the phases of training, validation and testing.

The sub-models EC, FU Model, IFT Model and HI Model were redeveloped with Random Forest, Naïve Bayesian and SVM classifiers in Google Collaboratory environment. For the HI Model, 323 data samples were available. For each of the other sub-models, 1057 data samples were available.

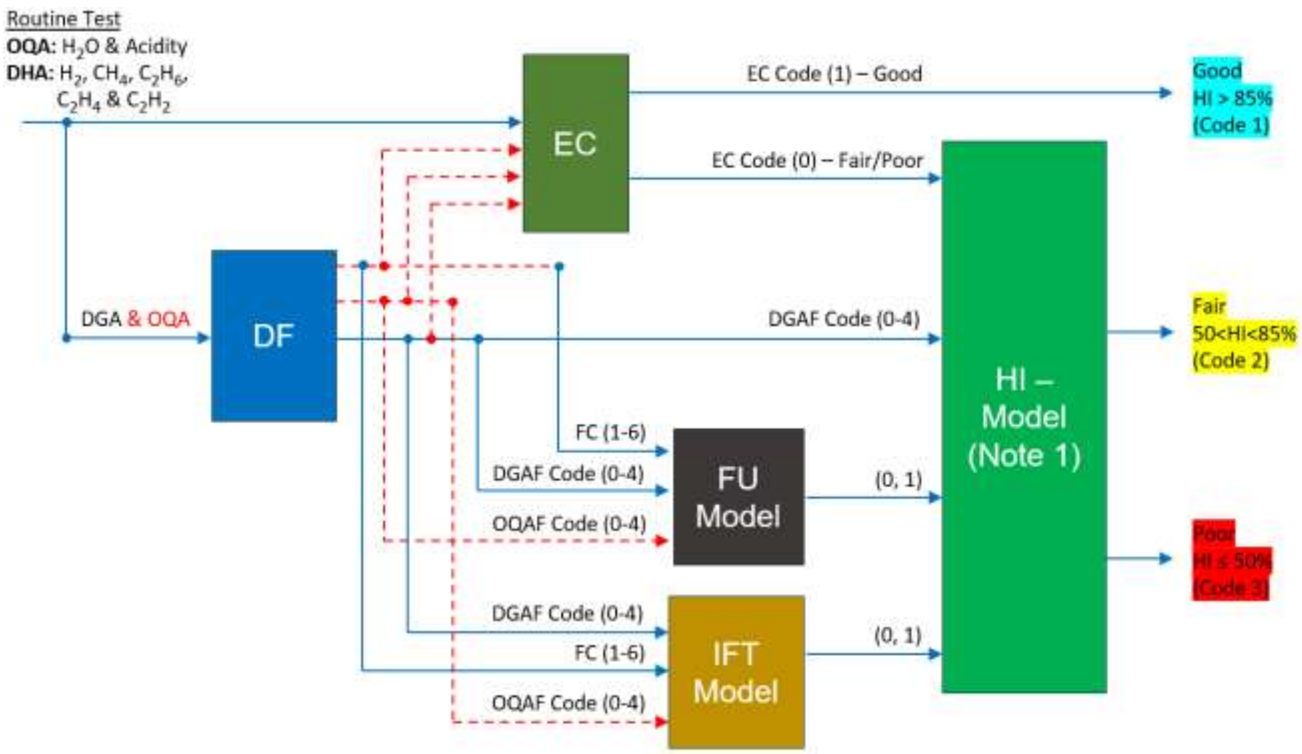

Fig. 1. Model structure for EDM [18]

2.2 ML Model Setup

The relevant hyperparameters were tuned by modifying the code at multiple occasions. Testing the impact of various hyperparameter values across different classifiers is quite essential. Ahamad *et al.,* [19] demonstrated that hyperparameter tuning can lead to improved, diminished or unchanged prediction accuracy for a given classifier [19].

The hyperparameters tested for this research were Random_state, N_estimators, Priors and Var_smoothing. Random_state was used in both Random Forest and Gaussian Naïve Bayesian classifiers. It ensures the model's output can be consistently replicated. When a specific random_state value is set, the ML model produces identical results as long as the same hyperparameters and training data are used [20]. Throughout this experiment, the random state was arbitrarily set to 7 as the value itself does not hold any particular significance. Then estimator hyperparameter, used for Random Forest, represents the number of trees constructed before performing maximum voting or averaging predictions [20].

For this experiment, it was set to 100, which was considered balanced enough to ensure stable predictions without affecting the computation time. The latter hyperparameters were used for Naïve Bayesian. The smoothing parameter was set to its default value 1×10^{-9} as the training datasets did not have any feature with zero variance. Moreover, the Priors parameter was set to 'None' because it was desired that the classifier itself calculates the prior probabilities for each class from the training dataset. For the SVM classifier, the linear kernel was used. Besides Random state, it also include the tol parameter representing tolerance, which was set to 1×10^{-5}. A small tolerance value like this ensures that the optimization process converges to a more precise solution.

2.3 Performance Evaluation and Comparison

Each classification model would generate a classification report at the end of the final testing phase. The report comprised of performance metrics such as precision, recall, F1-score, support and accuracy. Apart from these, a confusion matrix would be used to obtain a visual representation of the classification result. As part of this particular research, only the Random Forest and Naïve Bayesian classifiers were developed from scratch and the SVM classifier was modified after the one used in [11]. The results are discussed in more detail in Section 3.

3. Result

The required metrics for evaluating the performance of the classifiers for EC, Furan Model, IFT Model and HI Model are presented in Tables 1 to 4 respectively.

Table 1
Result metrics for early classifier

Metrics	Classification model (Early classifier)		
	Random Forest	Naïve Bayesian	Support Vector Machine
Accuracy	0.855	0.905	0.881
Precision	0.521	0.781	0.744
Recall	0.776	0.850	0.825
F1-Score	0.624	0.814	0.783

Table 2
Result metrics for Furan model

Metrics	Classification model (Furan model)		
	Random Forest	Naïve Bayesian	Support Vector Machine
Accuracy	0.863	0.905	0.883
Precision	0.455	0.618	0.533
Recall	0.610	0.850	0.800
F1-Score	0.522	0.716	0.640

Table 3
Result metrics for IFT model

Metrics	Classification model (IFT model)		
	Random Forest	Naïve Bayesian	Support Vector Machine
Accuracy	0.949	0.965	0.921
Precision	0.720	0.720	0.514
Recall	0.643	0.857	0.643
F1-Score	0.679	0.783	0.571

Table 4
Result metrics for HI model

Metrics	Classification model (HI model)		
	Random Forest	Naïve Bayesian	Support Vector Machine
Accuracy	0.922	0.923	0.896
Precision	0.977	0.988	0.942
Recall	0.933	0.924	0.911
F1-Score	0.954	0.955	0.926

Based on the results presented in Tables 1 to 4, Naïve Bayesian outperformed Random Forest and SVM in terms of accuracy and had the highest value of precision and F1-score for all the sub-models. Considering that each sub-model had their own datasets, dataset changes have influenced the performance metrics of all three classifiers, especially in precision and recall. In terms of accuracy and F1-score, SVM managed to outperform Random Forest for Early Classifier and Furan Model, whereas Random Forest outperformed SVM for IFT Model and HI Model. So, it can be said that while Random Forest and SVM were generally on par with each other, Naïve Bayesian stood out as the best performing classifier out of the three. These findings underscore the robustness and reliability of Naïve Bayesian classifiers in varied conditions, highlighting their potential for broader application in transformer health assessments.

4. Conclusion

The results of the experiments under this project reveal that while Random Forest and SVM had almost similar capabilities, Naïve Bayesian is the most promising medium of taking full advantage of the EDM structure for classifying the health indices of power transformers. Nevertheless, the possibility of getting different results with a different number of data samples and method of tuning the classifiers still remain. Future work on this topic could be carried out using deep learning models on a comprehensive and varied dataset, unlike the one used for this project. Furthermore, the development of a user-friendly interface integrated with the EDM can also be considered to make it convenient for industries to apply it more effectively to their transformers.

Acknowledgement
All praises to Almighty Allah for enabling me to work towards the successful completion of this paper. I express my deepest gratitude to Ir. Ts. Dr. Mohd Hatta Mohammed Ariff, the supervisor for my Final Year Project, for his insightful guidance and feedback. I should also mention Mr Azlan Mohmad, whose research on the Early Detection Model has served as a guiding beacon for conducting my own research. I am also thankful to my classmates, seniors and friends for sharing their knowledge on crafting a proper project report. Finally, I extend my heartfelt thanks to my family for their constant support and love throughout my studies at university.

References
[1] Ishak, Norshahidatul Hasana, Nur Farahin Johari, Siti Nuramalina Johari, Hazrati Zaini, Abdul Hasib Sadiqin Adam Shukri, and Aqil Afifi Ahmad Senusi. "Learning and Training of Coal-Fired Energy using Augmented Reality (C-FAR)." *Journal of Advanced Research in Computing and Applications* 17, no. 1 (2019): 20-24.

[2] Islam, Naimul, Riaz Khan, Sajal K. Das, Subrata K. Sarker, Md Manirul Islam, Masuma Akter, and S. M. Muyeen. "Power transformer health condition evaluation: A deep generative model aided intelligent framework." *Electric Power Systems Research* 218 (2023): 109201. https://doi.org/10.1016/j.epsr.2023.109201

[3] 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

[4] Naderian, A., S. Cress, R. Piercy, F. Wang, and J. Service. "An approach to determine the health index of power transformers." In *Conference Record of the 2008 IEEE International Symposium on Electrical Insulation*, pp. 192-196. IEEE, 2008. https://doi.org/10.1109/ELINSL.2008.4570308

[5] Li, Shuaibing, Xinchen Li, Yi Cui, and Hongwei Li. "Review of transformer health index from the perspective of survivability and condition assessment." *Electronics* 12, no. 11 (2023): 2407. https://doi.org/10.3390/electronics12112407

[6] 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

[7] Foros, Jorn, and Maren Istad. "Health index, risk and remaining lifetime estimation of power transformers." *IEEE Transactions on Power Delivery* 35, no. 6 (2020): 2612-2620. https://doi.org/10.1109/TPWRD.2020.2972976
[8] Shanmuganathan, Tejasvini, and Faizir Ramlie. "Production Line Monitoring using Mahalanobis-Taguchi System in Rubber-Based Product Industry." *Semarak Engineering Journal* 4, no. 1 (2024): 1-17.
[9] Luo, Yunhe, Xiaosong Zou, Wei Xiong, Xufeng Yuan, Kui Xu, Yu Xin, and Ruoyu Zhang. "Dynamic State Evaluation Method of Power Transformer Based on Mahalanobis–Taguchi System and Health Index." *energies* 16, no. 6 (2023): 2765. https://doi.org/10.3390/en16062765
[10] Abd Aziz, Azmi Murad, Mohd Aizam Talib, Ahmad Farid Abidin, and Syed Abdul Mutalib Al Junid. "Development of power transformer health index assessment using feedforward neural network." *J. Adv. Res. Appl. Sci. Eng. Technol.* 30, no. 3 (2023): 276-289. https://doi.org/10.37934/araset.30.3.276289
[11] 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
[12] Janiesch, Christian, Patrick Zschech, and Kai Heinrich. "Machine learning and deep learning." *Electronic markets* 31, no. 3 (2021): 685-695. https://doi.org/10.1007/s12525-021-00475-2
[13] Parveen, R., M. Nabi, F. A. Memon, S. Zaman, and M. Ali. "A review and survey of artificial neural network in medical science." *Journal of Advanced Research in Computing and Applications* 3, no. 1 (2016): 7-16.
[14] Khaw, Li Wen, and Shahrum Shah Abdullah. "Mri Brain Image Classification Using Convolutional Neural Networks and Transfer Learning." *Journal of Advanced Research in Computing and Applications* 31, no. 1 (2023): 20-26. https://doi.org/10.37934/arca.31.1.2026
[15] Stihec, Jan. “Choose Your AI Weapon: Deep Learning or Machine Learning.” Shelf, May 12, 2024.
[16] https://www.facebook.com/electrical4u. “Routine Tests and Type Tests of Transformers | Electrical4u.” Electrical4U, May 10, 2013.
[17] Muthu, Shanmuga Pillai Murutha, and Sellappan Palaniappan. "Categorization of Early Detection Classifiers for Gastric Carcinoma through Data Mining Approaches." *Journal of Advanced Research in Computing and Applications* 32, no. 1 (2023): 1-12. https://doi.org/10.37934/arca.32.1.112
[18] A. Mohmad, "Early Detection Model for Transformer Health Classification using Ensemble Support Vector Machine," M.A. thesis, Universiti Teknologi Malaysia, 2023.
[19] Ahamad, Ghulab Nabi, Shafiullah, Hira Fatima, Imdadullah, S. M. Zakariya, Mohamed Abbas, Mohammed S. Alqahtani, and Mohammed Usman. "Influence of optimal hyperparameters on the performance of machine learning algorithms for predicting heart disease." *Processes* 11, no. 3 (2023): 734. https://doi.org/10.3390/pr11030734
[20] Donges, Niklas, M. Urwin, and S. Pierre. "Random Forest: a complete guide for machine learning." *Built In* (2023).