

Frequency-Domain Dielectric Analysis and Machine Learning Classification of Heat Transfer Oil Degradation States

Yee See Khee^{1*}, Ong Pauline², You Kok Yeow³, Fahmiruddin Esa⁴,
Fauziahanim Che Seman¹, Arslan Ahmed⁵

¹Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia

²Faculty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia

³Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia

⁴Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia

⁵Department of Electronic Engineering, The University of Larkano, 77150, Sindh, Pakistan

ARTICLE INFO

Article history:

Received 30 December 2025

Revised 22 April 2026

Accepted 06 May 2026

Published 15 September 2026

Keywords:

Heat transfer oil

Quality classification

Viscosity

Preventive maintenance

DOI:

10.24191/jmeche.v23i3.9956

ABSTRACT

Heat transfer oils (HTOs) appears as a critical fluid in industrial thermal systems because it offers high-temperature stability and operational safety compared to conventional heat transfer media. However, the efficiency and safety of operation, and equipment longevity is often compromised by the degradation of HTO. HTO suffers from oxidation, cracking, and contamination because of long term operation. Traditional laboratory analyses such as viscosity, total acid number (TAN), and Karl Fischer titration are usually performed to identify the oil quality, it is effective but time-consuming. So, it is unsuitable for *in-situ* measurement. This study presents a non-destructive dielectric-based approach for assessing HTO quality by using an open-ended coaxial probe coupled with machine learning classifiers. Dielectric constant, ϵ_r' were measured over 1–9 GHz for fresh and used samples collected from an incineration plant operating with Shell Heat Transfer Oil S2 (Shell Global). Statistical analysis identified 1.75 GHz and 1.84 GHz as optimal diagnostic frequencies, showing strong separation among good, “Medium”, and degraded oils ($p < 0.001$). This work has applied three classification models, namely Support Vector Machine (SVM), Random Forest (RF), and k -Nearest Neighbour (k NN) for training of oil classification based on dielectric data and viscosity-based labels. The dataset was partitioned into training and testing sets. The training phase utilized 85 samples (35 “Good”, 20 “Medium”, and 30 “Bad”), while 18 samples (8 “Good”, 7 “Medium”, and 3 “Bad”) were reserved for testing. Based on the analysis, it was found that SVM achieved 100% precision, sensitivity, specificity, and accuracy at 1.84 GHz. The outcome of this work highlights the potential of integrating dielectric

^{1*} Corresponding author. E-mail address: skyee@uthm.edu.my
<https://doi.org/10.24191/jmeche.v23i3.9956>

sensing with machine learning in predicting the HTO quality. It provides the proof-of-concept and theoretical framework intended for future real-time, in-situ applications.

Introduction

Heat transfer oils (HTOs) are thermal fluids that are widely applied for precise temperature control and heat management. They are a better candidate compared to water or steam because they can operate at high temperatures without compromising the system pressure. The low system pressure is very crucial to ensure safety and energy efficiency (Alharbi et al., 2025).

Generally, HTOs can be classified into mineral-based and synthetic-based formulations. Mineral-based oils, such as Shell Heat Transfer Oil S2 (Shell Global), ExxonMobil Therm 605, and Petro-Canada Purity FG Heat Transfer Fluid, are refined petroleum fractions. They are composed of paraffinic, naphthenic, and aromatic hydrocarbons. While cost-effective for large-scale use, these oils are prone to thermal cracking and oxidation when operated above 300 °C. These chemical breakdown processes lead to the formation of acidic byproducts, sludge, and coke, which ultimately result in a significant increase in oil viscosity. Such changes impair the fluid's flow characteristics and reduce the overall efficiency and lifespan of the thermal system (Sukarin et al., 2025; Nazeer, 2025). Conversely, synthetic-based HTO, including silicone-based fluids and diphenyl oxide–biphenyl eutectic mixtures, offer superior thermal stability at temperature 400 °C. Their low vapor pressure and high resistance to oxidation make them ideal for industries requiring consistent performance under extreme thermal stress (Wiltchko & Cheng, 2025).

Regardless of mineral-based or synthetic HTO, they are facing ageing-related problems after long operation. These problems could seriously impact the system performance if they are left unattended. Examples of ageing-related problems include oxidation, thermal cracking, polymerization, and contamination. When this happens, their viscosity increases, sludge, coke, and acid are formed in the oil (Pichler et al., 2025). These substances and conditions reduce the lifespan of both the system and equipment (Shen et al., 2025).

HTO degradation significantly impacts system performance and safety. In high-temperature applications such as incineration plants, the formation of sludge and acidic by-products increases corrosion in cooling circuits and elevates maintenance costs (Zhu et al., 2025). Furthermore, as oil viscosity increases due to thermal stress, heat transfer efficiency declines, leading to higher energy consumption and potential fouling of heat exchangers (Alharbi et al., 2025). These impacts have highlighted the importance of monitoring the quality of HTO for predictive maintenance. Timely replacement of HTO is critical to ensure reliable and safe industrial operation. Most of the time, the HTO is sent regularly to the laboratory for quality assurance. The common laboratory methods include viscosity measurement, water content measurement (Karl Fischer), total acid number (TAN), and flash point testing. Each of these methods provides accurate information on oil quality but comes along with practical challenges.

Viscosity measurement is usually performed by using capillary or rotational viscometers. It is a relatively inexpensive and quick method, but it can be influenced by sampling errors. As stated in ASTM D445-18, the temperature during the measurement must be well-controlled for reproducibility. Karl Fischer titration (ASTM E203-16) is a standard method that is sensitive to detect the moisture or water content in HTO. While performing the test, the skilled operator must handle the reagents and calibration carefully. The chemical reagent cost is making the testing expensive (ASTM D6304-20). TAN measurement (ASTM D664-18e2) is usually carried out via potentiometric titration. The measurement procedure is time-consuming (often requiring over an hour per sample) as the tester needs to examine the color change to quantify the acidic by-products formed during oxidation. This is strongly dependent on the experience of

the tester (ASTM D664-24). Flash point testing (ASTM D92-16B) is performed by open or closed cup methods. This test assesses the safety of the oil by detecting the presence of volatile, low-molecular-weight hydrocarbons produced through thermal cracking. A significant drop in the flash point indicates the risk of fire or explosion during high-temperature operation. However, because the flash point values typically shift only after significant degradation, it is often considered a lagging indicator with limited sensitivity for detecting the early stages of oil deterioration. To address these limitations, researchers have proposed a range of advanced spectroscopic and electrical methods. These techniques are more comprehensive in diagnosing the physicochemical changes in oils during their operation.

Spectroscopic techniques such as Fourier-transform infrared (FTIR) and nuclear magnetic resonance (NMR) spectroscopy have gained a lot of attention. FTIR is capable of identifying chemical changes due to oil oxidation, polymerization, or contamination. It can detect carbonyl, hydroxyl, and hydroperoxide groups rapidly during oil degradation. On the other hand, NMR provides structural information in detail at the molecular level. These methods are often combined with chemometric modeling and machine learning for predictive diagnostics. The outcome of the machine learning can transform complex spectral data into meaningful insights for preventive maintenance. For example, Ye & Meng (2022) demonstrated that edible oils could be classified with 100% accuracy using FTIR combined with discrimination models. While adulteration in Camellia oil blends was detected with high precision through partial least-squares models (Ye & Meng, 2022). Similarly, work presented by Bassbasi et al. (2013) showed that FTIR with multivariate regression could predict motor oil adulteration with an error margin below 3%.

Dielectric and impedance-based techniques are another alternative for oil quality diagnosis. These methods exploit changes in electrical properties, such as dielectric constant, ϵ' , conductivity, and relaxation behavior, to predict chemical degradation, water contamination, and impurity accumulation. Alemayehu et al. (2021) demonstrated the feasibility of impedance spectroscopy for real-time transformer oil monitoring. They have highlighted its potential for cloud-connected predictive maintenance systems. Similarly, Baklanov et al. (2022) reported a resonant series-parallel impedance approach that detects spectral shifts in dielectric loss. The spectral shifts provide early indicators of oil deterioration. Even though most of this research focuses on transformer and lubricating oils, these systems are very similar to HTOs. Both HTOs and industrial oils are made of hydrocarbons. They go through the same chemical changes, like creating polar oxidation products. These changes can be easily detected by measuring their electrical properties

Most of the studies now have integrated machine learning algorithms with spectroscopic data to enhance oil quality monitoring. Models such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Gradient Boosting have been applied to detect deterioration patterns and classify oil samples with high accuracy. Karadağ et al. applied FTIR data with machine learning to distinguish between virgin and repeatedly fried olive oil. It is reported that the SVM and kNN can achieve around 94% of classification accuracy (Karadağ et al., 2024). Sinduja et al. (2022) demonstrated that support vector machine (SVM) with Bayesian optimization can classify transformer oil conditions with 99.5% recognition accuracy. This approach is far better than traditional diagnostic methods that rely on slow, destructive laboratory tests. Their work has shown that classification between oils suitable for continued use, oils requiring conditioning, and oils in degraded states is applicable (Sinduja et al., 2022).

Das et al. (2023) have developed a convolutional neural network (CNN) trained on FTIR spectral data to classify natural ester insulating oils at different ageing stages. They tested multiple architectures (ResNet50, VGG19, AlexNet) by converting 1D FTIR spectra into 2D encoded images. It was found that ResNet50 achieved 96.7% accuracy (Das et al., 2023). In his following work, Das et al. (2024) applied recurrence plot-based deep learning for transformer oil diagnosis. FTIR spectra of mineral, natural ester, and synthetic ester oils were transformed into recurrence plots, followed by CNN classification. Their approach achieved over 97% classification accuracy across all oil types. The reliability has validated its

real-world potential for in-situ monitoring of transformer insulation health (Das et al., 2024). Besides transformer applications, machine learning has also shown its success stories in food-grade edible oils. Berghian-Grosan & Magdas (2020) used Raman spectroscopy combined with ML algorithms for edible oil authentication. Their system not only detected adulteration in high-value cold-pressed oils (e.g., sesame, walnut, hemp) but also estimated the degree of adulteration (Berghian-Grosan & Magdas, 2020).

Despite the progress in using machine learning for monitoring insulating and edible oils, there is a lack of specialized research focusing on the dielectric behavior of mineral-based HTOs, particularly in high-temperature industrial environments like incineration plants. Existing diagnostic methods remain either too slow for in-situ application or have not been fully validated for the specific chemical changes, such as thermal cracking and sludge formation, which are unique to HTO systems. Consequently, there is a critical need for a diagnostic framework that identifies specific frequency-domain to enable rapid, non-destructive quality assessment of these fluids. To address this, this work aims to: (i) identify the optimal frequency ranges to assess the quality of HTO by using an open-ended coaxial probe, and (ii) develop and evaluate the performance of different machine learning classifiers (SVM, RF, and k NN) in predicting the HTO degradation based on dielectric constant, ϵ_r' and viscosity-based labels.

Material and Method

Heat transfer oil sample preparation

In this work, the heat transfer oil (HTO) samples were obtained from Medivest Sdn. Bhd., a healthcare facilities management company in Malaysia that employs oil-based thermal systems within its incineration plants. In this application, Shell Heat Transfer Oil S2 (mineral-based thermal fluid) is applied for indirect heating and cooling. A total of 20 samples of used oil were collected at four-day intervals, together with one fresh oil sample for baseline comparison. The samples were collected over a period spanning November 2024 to January 2025. The samples were drawn directly from the operating system during scheduled plant operation and transferred into clean, food-grade plastic containers when it cools down for ease of handling. To reduce the risk of oxidative or photolytic degradation, the containers were stored under dark, room-temperature ($\sim 26^\circ\text{C}$) conditions until analysis. Importantly, the samples were not filtered, as the presence of sludge, particulates, and other insoluble fractions was preserved to capture the actual degradation.

Dielectric measurement with open-ended coaxial probe

The dielectric constant, ϵ_r' of the heat transfer oil samples was measured using an in-house open-ended coaxial probe (You et al., 2013) connected to a N9925A Fieldfox vector network analyzer (VNA), manufactured by Keysight Technologies, as shown in Fig. 1. This method is widely used for liquid characterization because it is fast, non-destructive, and requires minimal sample preparation (You et al., 2013). The VNA records the reflection coefficient, $S_{11\text{meas}}$, when the aperture probe is dipped into the liquid sample.

The VNA was calibrated at the start of each measurement session using a three-standard (open-circuit, short-circuit, and water load) calibration. The probe was first exposed to air (open), then pressed against a metallic (short), and finally immersed in deionized water at 25°C (load). Air defines the open reference and provides an initial estimate of the probe capacitance in free space. The short corrects for systematic phase errors and probe end effects. Water acts as a load with a well-known permittivity. To confirm instrumental stability and calibration accuracy, verification measurements were performed on air, deionized water, and Teflon before testing the HTO samples. Recalibration was performed if any drift was detected. This protocol ensured that the observed variations in the oil's dielectric constant, ϵ_r' . These standards yield calibration constants that link the measured $S_{11\text{meas}}$ to the true relative complex permittivity, $\epsilon_r (= \epsilon_r' - j \epsilon_r'')$

of the sample. These values are then converted into relative complex permittivity, ϵ_r , using a calibration function Equation (1) as (You et al., 2013):

$$\epsilon_r = - \left(\frac{\Delta_{m_a} \Delta_{w_s} \epsilon_{water} \epsilon_{short} + \Delta_{m_s} \Delta_{a_w} \epsilon_{air} \epsilon_{water} + \Delta_{m_w} \Delta_{s_a} \epsilon_{short} \epsilon_{air}}{\Delta_{m_a} \Delta_{w_s} \epsilon_{air} + \Delta_{m_s} \Delta_{a_w} \epsilon_{short} + \Delta_{m_w} \Delta_{s_a} \epsilon_{water}} \right) \quad (1)$$

where ϵ_{air} ($= 1.0006$), ϵ_{short} ($= 1 - j10^7$), and ϵ_{water} are the relative complex permittivity of air, short-circuit, and water, respectively. While the ϵ_{water} value can be obtained from (Stogryn, 1995):

$$\epsilon_{water} = \epsilon_i + \frac{\epsilon_s - \epsilon_1}{1 + jA_1 f} + \frac{\epsilon_1 - \epsilon_i}{1 + jA_2 f} \quad (2)$$

where parameters (ϵ_s , ϵ_1 , ϵ_i , A_1 , and A_2) in Equation 2 is given as:

$$\epsilon_s = \frac{37088.6 - 82.168T}{421.854 + T} \quad (3a)$$

$$\epsilon_1 = 5.723 + 0.022379T - 0.00071237T^2 \quad (3b)$$

$$\epsilon_i = 3.6143 + 0.028841T \quad (3c)$$

$$A_1 = \frac{(5.0478 - 0.070315T + 0.00060059T^2) \times 10^{-9}}{45 + T} \quad (3d)$$

$$A_2 = \frac{(0.13652 + 0.0014825T + 0.00024166T^2) \times 10^{-9}}{45 + T} \quad (3e)$$

Here, the T is the temperature in degrees Celsius. Other parameters in Equation 1 are given as:

$$\begin{aligned} \Delta_{m_a} &= S_{11_{meas}} - S_{11_{air}} & \Delta_{m_s} &= S_{11_{meas}} - S_{11_{short}} & \Delta_{m_w} &= S_{11_{meas}} - S_{11_{water}} \\ \Delta_{a_w} &= S_{11_{air}} - S_{11_{water}} & \Delta_{s_a} &= S_{11_{short}} - S_{11_{air}} & \Delta_{w_s} &= S_{11_{water}} - S_{11_{short}} \end{aligned}$$

Symbols $S_{11_{air}}$, $S_{11_{short}}$, and $S_{11_{water}}$ are the measured reflection coefficients for air, short-circuit, and water load standards. On the other hand, $S_{11_{meas}}$ is the measured reflection coefficient of the oil sample connected at aperture probe.

For each oil sample, approximately 120 mL was placed in a clean glass container. The coaxial probe was immersed carefully to avoid air gaps. Measurements were performed from 1 GHz to 9 GHz for five times to account for instrumental consistency and minor fluctuations. Instead of averaging these results, all repetitions were treated as separate samples in the subsequent statistical analysis and model training. This resulted in a total of 105 data points (21 physical samples $\times$ 5 repetitions).

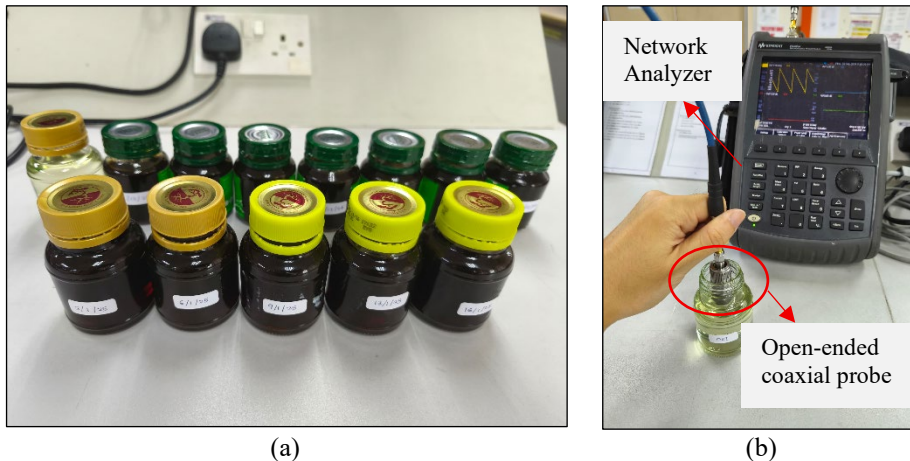


Fig. 1. (a) Heat transfer oil samples and (b) dielectric measurement setup using open-ended coaxial probe and N9925A Fieldfox network analyzer.

Viscosity measurement based on Zahn cup viscometer

Viscosity is an essential indicator of heat transfer oil (HTO) quality because it governs fluid flow, pumping energy, and overall heat transfer performance. An increase in viscosity often reflects oil oxidation, sludge formation, or thermal degradation, which can impair system efficiency and reliability (Anderson et al., 2015).

Several established techniques are available for viscosity measurement, each with distinct advantages and limitations. Capillary viscometers, which conform to ASTM D445-18, are widely regarded as the standard for kinematic viscosity determination. They provide high accuracy but require precise temperature control and relatively long measurement times (time spent cleaning the capillary), making them less convenient for frequent testing of multiple samples. Rotational viscometers offer the advantage of measuring both dynamic and kinematic viscosity across a broad range of shear rates. It is useful for non-Newtonian fluids. However, they are expensive, require careful calibration, and are more complex to operate. Due to this reason, it often limits their use to well-equipped laboratories. More advanced methods, such as an oscillatory rheometer, provide detailed rheological characterization but are far beyond the cost and complexity typically required for routine oil quality monitoring (Anderson et al., 2015).

In contrast, efflux viscometers, including Zahn, Ford, and DIN cups, provide a simple, rapid, and cost-effective alternative (ASTM D4212-16). The Zahn cup is widely applied in industrial practice due to its portability and ease of use. While less precise than capillary or rotational methods, its standardized correlation between efflux time and viscosity allows reliable comparative monitoring when measurements are conducted under controlled conditions. The reliability of the Zahn cup was confirmed by the instrument's operational standard, which defines acceptable repeatability as a difference of less than 5% between measured values. Given that the "Medium" and "Bad" oil categories in this study were defined by viscosity shifts of 3–13% and >13% respectively, the instrument's precision was well within the requirements to ensure distinct class separation. All five repeated measurements for each sample were verified against this 5% tolerance to ensure data integrity before being used for machine learning training.

In this work, a Zahn Cup Size 2 was employed for all viscosity measurements as shown in Fig. 2. This specific size was selected because its calibrated operating range (20–250 cSt) perfectly encompasses the kinematic viscosity of both the fresh (25.23 cSt) and degraded (up to 29.65 cSt) Shell Heat Transfer Oil S2 samples used in this work.



Fig. 2. Viscosity measurement using Zahn cup size 2.

During each test, the Zahn cup was filled until overflow, and timing commenced as soon as the liquid began to flow from the orifice. The stopwatch was stopped at the first break of the continuous liquid stream. The recorded efflux time was then converted to kinematic viscosity, V , using the standard calibration equation (Equation 4), where t represents the measured flow time, and A and C are calibration coefficients specific to each cup size (ASTM D4212-16). The corresponding coefficient values are provided in Table 1. To minimize variability, all measurements were performed under controlled laboratory conditions at room temperature. The Zahn cup was thoroughly cleaned after each measurement to prevent residue effects. Each sample was tested five times, and the repeated measurements helped to account for overflow consistency and potential operator timing differences.

$$V = A(t - c) \quad (4)$$

Table 1. Information of Zahn Cup models for viscosity calculation

Size	Orifice diameter (mm)	Viscosity, V (cSt)	A	C
1	2.2	60	1.1	29
2	2.74	20-250	3.5	14
3	3.76	100-800	11.7	7.5
4	4.28	200-1200	14.8	5
5	5.28	400-1800	23	0

Spectra variable selection

In order to identify the most informative frequency for heat transfer oil classification, the dielectric constant, ϵ_r of the heat transfer oil is grouped into three categories, namely “Good”, “Medium” and “Bad” oil, corresponding to their oil condition (viscosity). Each category consisted of multiple measurement sets, ranging from 1 GHz to 9 GHz. Preliminary exploration focuses on computing their mean and standard deviation at each frequency across the three oil conditions. This technique enables an overview of the spectral response and rapid identification of the frequency ranges that have pronounced separation. One-

way ANOVA tests are conducted at these frequency points (1.74 GHz, 1.75 GHz, 1.76 GHz, 1.84 GHz, 1.85 GHz, and 1.86 GHz) individually using the dataset of 103 observations (after the removal of two outliers from the original 105 samples). These frequencies are selected based on their statistical significance during initial data analysis.

Model training and evaluation protocol

To ensure a fair and unbiased evaluation of the classification models, the dataset was partitioned into training and testing subsets. The dielectric constant, ϵ_r' at the selected frequencies (1.75 GHz and 1.84 GHz) was first aggregated according to their corresponding oil condition labels (“Good”, “Medium”, and “Bad”), resulting in a total of 103 samples (“Good” = 45, “Medium” = 39, “Bad” = 19) respectively.

The dataset was partitioned using a stratified hold-out approach, with 80% of samples allocated for training ($n = 85$; “Good” = 37, “Medium” = 32, “Bad” = 16) and 20% for testing ($n = 18$; “Good” = 8, “Medium” = 7, “Bad” = 3). This stratification preserves the class distribution across both subsets.

Hyperparameter optimization for all models was performed using Bayesian optimization on the training set, which efficiently explores the parameter space by balancing exploration and exploitation. For the Support Vector Machine (SVM), the penalty parameter C and the kernel scale (γ) of the radial basis function (RBF) kernel were tuned within the range $C \in [10^{-2}, 10^3]$. For the Random Forest (RF), the number of trees (50–300) and the maximum number of splits were optimized to control model complexity. For the k -Nearest Neighbour (k NN), the number of neighbours $k \in [1, 15]$ and the distance metric were tuned, with Euclidean distance adopted due to the low-dimensional feature space.

The final models were then retrained using the full training dataset with these optimized parameters and evaluated on the unseen test set. All reported performance metrics (precision, sensitivity, specificity, and accuracy) were calculated based strictly on this independent test set

Classification models

In this sub-section, the principles of three different classification models, namely k -Nearest Neighbour (k NN), random forest (RF), and support vector machine (SVM), are discussed.

k -Nearest Neighbour (k NN)

The k -Nearest Neighbors (k NN) algorithm is one of the most widely used non-parametric methods for classification and regression tasks. Its principle assumes that data points that are close to each other in feature space are likely to belong to the same class (Cover & Hart, 1967). Unlike parametric models, k NN does not build an explicit mathematical model during training. Instead, all labeled training samples are stored, and classification is performed only when a new sample is introduced.

When an unlabeled sample is tested, its distance to every training sample is calculated using a chosen similarity measure, the Euclidean distance is given in Equation 5:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

where x and y represent two samples described by n features. The algorithm then identifies the k nearest neighbors, and the new sample is assigned to the class that is most frequent among these neighbors. The decision rule can be expressed in Equation 6:

$$\operatorname{argmax} C_k \sum_{i=1}^k I(y_i = C_k) \quad (6)$$

where $\hat{c}$ is the predicted class, C_k denotes the candidate classes and I is an indicator function equal to 1 if the neighbor belongs to class C_k .

In recent research, k NN has been applied to oil quality monitoring and related fields. For instance, Ordukaya & Karlik (2017) employed k NN in combination with electronic nose data for olive oil quality control, demonstrating that k NN could effectively classify oils into different categories. Similarly, Puspitasari et al. (2022) compared classifiers, including Naïve Bayes and k NN, for the classification of palm oil varieties, showing that while NB was efficient with smaller datasets, k NN provided competitive results when sufficient training data were available. These examples highlight the flexibility of k NN in handling complex, real-world data where feature distributions may not follow simple parametric assumptions.

In this work, the heat transfer oils are classified into three categories: “Good”, “Medium” and “Bad” based on their viscosity and dielectric constant, ε_r' . It relies on the principle that samples with similar dielectric constants, ε_r' are likely to belong to the same class. Classification is achieved by calculating the distance between a new sample and previously labelled data points. Since only a single feature (single frequency) was used in this work, the Euclidean distance in Equation 5 reduces to the absolute difference, where x represents the dielectric constant, ε_r' of the new sample and y corresponds to that of a training sample.

The k parameter in Equation 6 defines how many neighbors are considered during classification. Small values of k (typically 1 or 3) make the algorithm sensitive to local variations and noise, while larger values provide smoother decision boundaries but may obscure subtle differences between classes. In this work, k is auto tuned using Bayesian optimization.

Random forest (RF)

Random Forest (RF) is an ensemble learning algorithm that builds upon the concept of decision trees to improve accuracy and reduce overfitting. Instead of relying on a single tree, RF constructs a large number of decision trees, each trained on a randomly selected subset of the data and features. The final classification is determined through majority voting across all trees, which makes the model more stable and less sensitive to noise in the dataset (Breiman, 2001).

The strength of RF lies in two key mechanisms: bootstrap aggregation (bagging) and feature randomness. Bagging reduces variance by training trees on different subsets of the data, while feature randomness ensures that each split in the tree considers only a random subset of features, which improves generalization. Mathematically, if $T_1, T_2, \dots, T_n$ represents the decision trees, the predicted class is obtained as shown in Equation 7.

$$\operatorname{argmax} C_k \sum_{i=1}^n I(T_i(x) = C_k) \quad (7)$$

where I is an indicator function that equals 1 when tree T_i assigns the sample x to class C_k , and 0 otherwise.

Recent studies have demonstrated the effectiveness of RF in oil condition monitoring. For example, Senoussaoui et al. (2021) applied RF to transformer oil quality assessment and reported superior classification accuracy compared to a single decision tree, particularly when combined with feature engineering and dimensionality reduction techniques. Similarly, Wu et al. (2022) integrated RF with data balancing methods such as SMOTE to enhance transformer fault diagnosis, showing that RF achieved higher diagnostic accuracy than alternative models.

In this work, the dielectric constant, ϵ_r' values at selected frequency are used as input features. Each decision tree in the forest splits the dielectric constant, ϵ_r' values at certain thresholds (such as values below 2.5 may tend to fall into the "Good" category, while values above 2.7 may lean toward "Bad"). By combining many such trees, the RF model can learn subtle patterns between the dielectric constant, ϵ_r' and oil quality categories.

For example, suppose a new measured dielectric constant, ϵ_r' is 2.622. Some trees may classify it as "Medium", while others may classify it as "Good" or "Bad", depending on their thresholds. The Random Forest aggregates these outputs, and the final decision is the class with the majority vote. This mechanism makes RF robust to noise in the data and more accurate than a single decision tree.

Support vector machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for classification and regression tasks. The main principle of SVM is to find an optimal decision boundary, called a hyperplane, that separates data points from different classes with the maximum possible margin. The data points that lie closest to this hyperplane are called support vectors, and they play a key role in defining the boundary (Cortes & Vapnik, 1995).

Mathematically, for a binary classification problem, SVM attempts to solve the following optimization problem in Equation (8):

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad \text{subject to} \quad y_i(w \cdot x_i + b) \geq 1, \quad \forall i \quad (8)$$

where w is the weight vector, b is the bias, x_i are the input features, and y_i are the class labels (+1 or -1). This ensures that the chosen hyperplane separates the classes with maximum margin while minimizing misclassification.

When the data is not linearly separable, SVM uses kernel functions (such as polynomial, sigmoid, or radial basis function (RBF)) to map the data into a higher-dimensional space where separation becomes possible (Decoste & Schölkopf, 2002).

SVM has been successfully applied in oil quality monitoring and dielectric analysis. For example, Sinduja et al. (2022) used kernelized SVM models, including the radial basis function (RBF) kernel, to assess the quality of high-voltage transformer oil. Their study demonstrated that SVM, combined with Bayesian optimization, achieved a recognition accuracy of up to 99.5%, showing its effectiveness in distinguishing between oil samples in good condition, those requiring conditioning, and degraded oils. This demonstrates the strong capability of SVM in handling nonlinear and complex classification problems in material diagnostics.

In this work, the dielectric constant, ϵ_r' at selected frequencies was used as the feature for classification of oil quality into three categories: "Good", "Medium", and "Bad". The Support Vector Machine (SVM) constructs decision boundaries that separate these classes with the maximum margin. Unlike simpler threshold-based methods, SVM determines an optimal hyperplane in feature space that reduces misclassification even when class boundaries overlap.

The classification function in SVM can be expressed as Equation (9):

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \right) \quad (9)$$

where α_i are the support vector weights, y_i are the class labels, $K(x, x_i)$ is the kernel function, and b is the bias. By applying a radial basis function (RBF) kernel, the dielectric constant, ϵ_r values are projected into a higher-dimensional space, allowing non-linear separation of “Good”, “Medium”, and “Bad” oil categories.

For example, if a new oil sample has $\epsilon_r=2.63$, a simple thresholding rule may misclassify it depending on the cut-off chosen. However, SVM considers the overall distribution of support vectors and the kernel transformation to correctly classify the sample as “Medium” with high confidence.

Performance assessment

The classification model’s performance is evaluated by their precision, sensitivity, specificity, and accuracy. These parameters are calculated by Equations 10 to 13.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (12)$$

$$\text{Accuracy} = \frac{\text{Number of samples classified correctly}}{\text{Total number of samples}} \quad (13)$$

whereby, TP, TN, FP, and FN represent the true positive, true negative, false positive and false negative, respectively.

Results and Discussion

Viscosity measurements

The viscosity of the fresh heat transfer oil (HTO) was measured at 25.23 cSt, which falls within the expected range for new mineral-based thermal fluids (Seeton, 2008; Shell Oil Products US, 2019). To ensure accuracy, each reported value is the average of five repeated measurements.

Over the monitoring period from November 2024 to January 2025, the viscosity showed a gradual increase, reflecting the natural aging of the oil under thermal and oxidative stress, as tabulated in Table 2. In early November, values remained close to the fresh baseline (25.7–26.0 cSt), with changes of less than 3%. At this stage, the oil can be classified as in good condition, with minimal degradation. By early December, viscosity had risen by 3–12% (up to 28.3 cSt), indicating the onset of oxidation and polymerization, processes that generate heavier molecular species and reduce flowability. Oils in this range

are categorized as in “Medium” condition, where performance begins to deteriorate. By mid-January, the viscosity exceeded 13% above baseline (around 29.6 cSt), a clear sign of substantial degradation. These samples represent “Bad” condition, as elevated viscosity is likely to impair pumping efficiency and reduce heat transfer performance. The thresholds used here ($<3\%$ = “Good”, $3\text{--}12\%$ = “Medium”, $>13\%$ = “Bad”) were established based on the observed distribution of viscosity changes in this study. While not standardized across industry, these ranges provide a practical framework to distinguish oil health levels and to serve as reference labels for the subsequent dielectric-based classification.

Some fluctuations were observed in the data, such as the temporary decrease to 26.55 cSt on 19/12/2024. These deviations are relatively minor compared with the overall increasing trend and may be explained by operational factors such as intermittent cooling cycles, oil replenishment, or variations in sampling conditions. Although direct confirmation through operational records was not available, these factors provide plausible explanations for the non-linear behavior.

To verify that the viscosity increase was statistically significant across the defined quality categories, a One-Way ANOVA was performed to verify the viscosity increment was statistically significant across the categories. The analysis confirmed a highly significant difference between the “Good”, “Medium”, and “Bad” groups, $F = 79.12$, $p < 0.001$. In the statistical summary, the mean viscosity increased significantly from 25.63 cSt (“Good”) to 29.22 cSt (“Bad”). This statistical evidence confirms that the observed viscosity changes represent distinct physical states of oil degradation rather than instrumental variance.

It is also important to note that the number of samples in each category was not evenly distributed. Fewer “Bad” samples were available because the degraded oil is replaced quickly to avoid operational risks. Hence, there is limited opportunity to collect severely deteriorated samples.

In this study, viscosity serves only as a reference indicator to define the oil quality categories of “Good”, “Medium”, and “Bad”. These categories provide the baseline labels for subsequent analysis, while the machine learning framework focuses on dielectric constant, ϵ_r' . By mapping dielectric responses to viscosity-based categories, the machine learning model can classify oil condition automatically without relying on viscosity measurements.

Table 2. Viscosity of heat transfer oil and their categories

Date of Heat Transfer Oil Collected	Viscosity (cSt)	Percentage of viscosity difference from the Fresh Heat Transfer oil	Oil Categories
		$\frac{ Viscosity - Viscosity_{fresh\ HTO} }{Viscosity_{fresh\ HTO}} \times 100$	
Fresh Heat Transfer Oil	25.23	0.00	Good
11/11/2024	25.73	1.96	Good
14/11/2024	25.64	1.63	Good
18/11/2024	25.55	1.27	Good
21/11/2024	25.83	2.38	Good
24/11/2024	25.96	2.89	Good
28/11/2024	25.66	1.70	Good
02/12/2024	25.44	0.83	Good
05/12/2024	26.63	5.55	Medium
09/12/2024	27.51	9.04	Medium
12/12/2024	28.18	11.69	Medium
16/12/2024	28.11	11.41	Medium
19/12/2024	26.55	5.23	Medium
23/12/2024	27.23	7.93	Medium

27/12/2024	26.80	6.22	Medium
30/12/2024	28.32	12.25	Medium
02/01/2025	28.93	14.67	Bad
06/01/2025	29.05	15.14	Bad
09/01/2025	28.87	14.43	Bad
13/01/2025	29.65	17.53	Bad
16/01/2025	29.61	17.36	Bad

Correlation between dielectric constant and quality of heat transfer oil at different frequencies

In this work, the dielectric constant, ϵ_r' of heat transfer oils is analyzed to differentiate their quality at different stages: Good, “Medium” and “Bad”. The heat transfer oil investigated in this work is collected from incinerators that are operating at high temperature (1000 °C). Substances such as moisture, organic acids, sludge, and carbon residual contribute to dipole polarization, relaxation of water molecules, and ionic conduction, which tend to increase their dielectric constant, ϵ_r' .

The dielectric constant, ϵ_r' spectra (1–9 GHz) displayed complex oscillatory behavior with varying degrees of overlap across the oil categories as shown in Fig. 3. To improve the robustness of the machine learning classification, visual inspection was implemented to identify frequency regions with the highest diagnostic potential. Four specific frequency bands (1.7–1.9 GHz, 3.75–3.84 GHz, 4.5–4.6 GHz, and 8.3–8.4 GHz) were isolated for further statistical evaluation. These regions were chosen because they demonstrated the most pronounced separation between the Good, “Medium”, and “Bad” oil states.

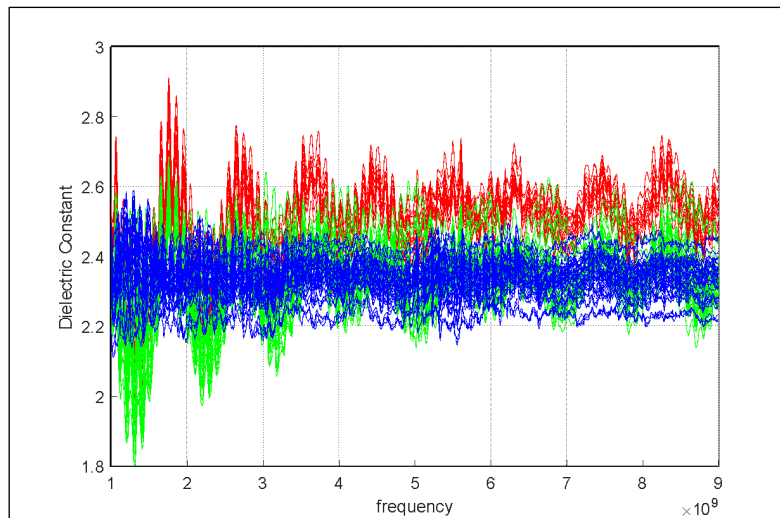


Fig. 3. Dielectric constant, ϵ_r' of heat transfer oil with different quality, “Good” (blue lines), “Medium” (green lines), and “Bad” (red lines).

Feature selection

Fig. 4 presents the mean and standard deviation of the dielectric constant, ϵ_r' of heat transfer oil across four frequency bands: 1.7–1.9 GHz, 3.75–3.84 GHz, 4.5–4.6 GHz, and 8.3–8.4 GHz. The statistical behavior of each band was summarized in Table 3 to determine the most suitable operating frequency, f for machine learning. From the table, the 1.7–1.9 GHz band stands out as the best frequency band as in this range, “Good”, “Medium”, and “Bad” oil samples are distinctly separated, especially around 1.74–

1.78 GHz and 1.84–1.88 GHz. The differences in the mean values are larger than the variation within each class, and the standard deviations are relatively small, making this band particularly reliable for classification. The 3.75–3.84 GHz band, on the other hand, shows less consistent separation. The standard deviations of “Good” and “Medium” samples overlap considerably, particularly between 3.76 and 3.84 GHz, which could reduce the accuracy of a classifier. While the 4.5–4.6 GHz band is narrow and shows smooth trends, the overlap in standard deviation between “Good” and “Medium” samples are still significant. A similar situation occurs in the 8.3–8.4 GHz band.

Based on this preliminary observation, the dielectric constant, ϵ_r' of heat transfer oil is analyzed at six different frequencies within the ranges 1.74–1.76 GHz and 1.84–1.86 GHz to identify the most suitable frequencies for machine learning classification. For each frequency band, the samples were categorized into three quality groups: “Good” ($n = 40$), “Medium” ($n = 39$), and “Bad” ($n = 15$). One-way ANOVA was applied to examine whether the mean dielectric constant, ϵ_r' differed significantly among the quality groups at each frequency.

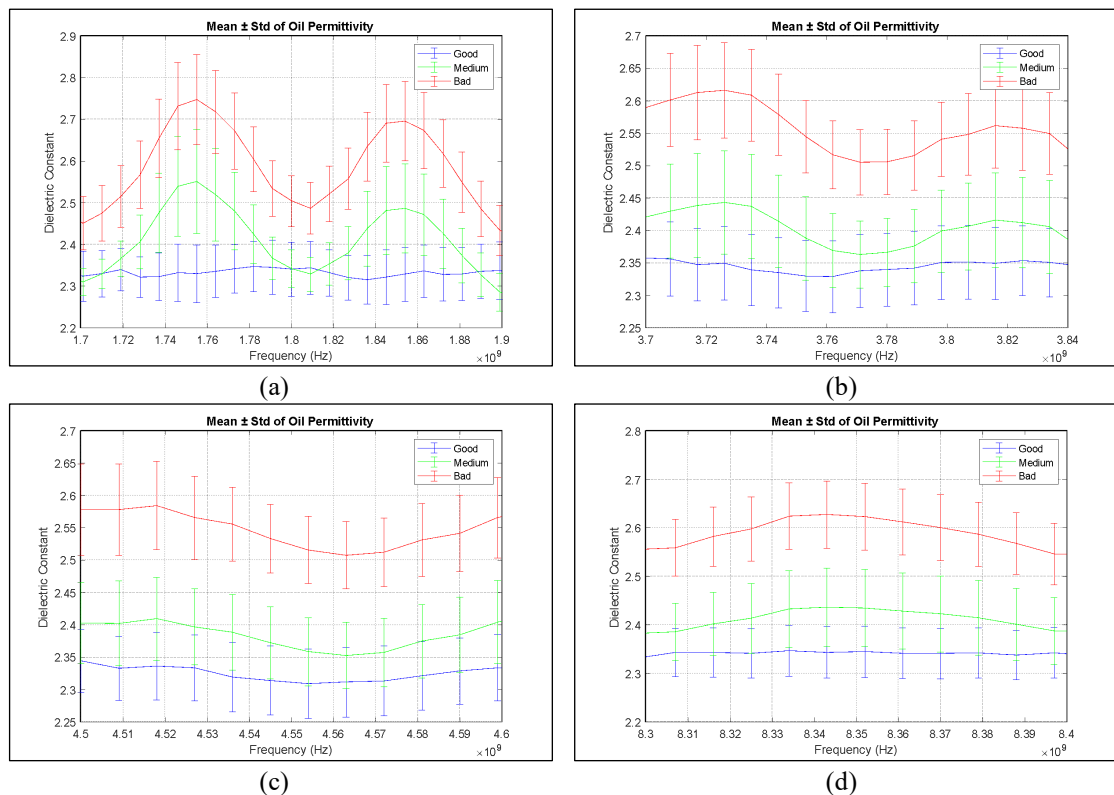


Fig. 4. Mean $\pm$ standard deviation of dielectric constant, ϵ_r' , for heat transfer oil quality groups across 4 selected frequency regions: (a) 1.7–1.9 GHz, (b) 3.75–3.84 GHz, (c) 4.5–4.6 GHz, and (d) 8.3–8.4 GHz.

Table 3. Comparative assessment of selected frequency regions for heat transfer oil-quality classification viscosity of heat transfer oil and their categories

Frequency Band (GHz)	Separation Between Classes	Standard Deviation Overlap	Trend Characteristic	Suitability for Machine Learning Classification
1.7 – 1.9	Moderate-High	Low especially from 1.74 - 1.77 GHz and 1.84 - 1.86 GHz	Clear monotonic trend with distinct peaks for Medium and “Bad” oil. Good oil does not have peak.	Strong (Best)
3.75 – 3.9	Moderate-High	Moderate-High. Medium and Good Oil standard deviation overlap in the whole frequency band	Visible separation with oscillation (“Bad” and Medium)	Low
4.5 - 4.6	Moderate-High	Clear cut in between “Bad” and Medium oil. Medium and Good oil overlapping	Smooth and stable curves, “Bad” oil consistently above others.	Medium
8.3 – 8.4	Moderate-High	Clear distinction between “Bad” and Medium oil. Medium and Good oil overlapping	“Bad” Oil is clearly above Medium and Good oil.	Medium

The ANOVA results are summarized in Table 4. In general, the results indicate there is a statistically significant difference between “Good”, “Medium” and “Bad” oil for all frequencies ($p < 0.001$ for hypothesis testing). The F -statistic, which ranges from 144.412 to 155.072, reflects a substantial separation between the quality groups relative to the within-group variance. The between-group variance (MSB) is much larger than the within-group variance (MSW). This shows that the measurements within each quality class were highly consistent and the differences between classes of oil are pronounced.

Table 4. F-statistic and p-values result from ANOVA test

Frequency, GHz	MSB	MSW	F -statistic = MSB/MSW	p -value	Effect Size ($\eta^2 = \frac{SSB}{SST}$)
1.84	1.132	0.007	155.072	<0.001	0.773
1.75	1.391	0.009	154.141	<0.001	0.772
1.76	1.010	0.006	154.023	<0.001	0.772
1.74	1.269	0.008	152.358	<0.001	0.770
1.85	1.004	0.006	149.406	<0.001	0.766
1.86	0.844	0.006	144.412	<0.001	0.760

1.84 GHz appears as the frequency with highest F-statistic, 155.072, followed closely by 1.75 GHz (154.141) and 1.76 (154.023). These values indicate that these frequencies are strong in providing separation among the oil quality and suitable for classification purposes.

The low MSW demonstrates that the dielectric constant, ϵ_r ’ measurements are highly stable within each class. When this is combined with the high F-statistics, it suggests that these frequencies can produce reliable prediction for “Good”, “Medium” and “Bad” oil samples.

To further evaluate the magnitude of these differences, the effect size was calculated using η^2 . The values ranged from 0.760 to 0.773, indicating an exceptionally large effect size. Specifically, at 1.84 GHz, the effect size of 0.773 implies that 77.3% of the dielectric variance is directly attributable to the oil quality degradation. This high degree of categorical separation provides a clear statistical rationale for selecting these frequencies as input features. The large effect size minimizes the overlap between classes, which directly contributed to the robust 100% classification accuracy observed in the SVM and RF models.

Although all the frequencies displayed significant F -statistics, a feature reduction strategy was employed to select the most representative variables for the machine learning models. 1.75 GHz and 1.84 GHz were selected as the primary inputs as these two points represent the local maxima for both the F -statistic and effect size (η^2) within their respective bands which provide the most distinct separation between oil quality categories.

Classification analysis

Tables 5 to 7 present the performance of SVM, RF and k NN classifiers when using the two selected frequencies, 1.75 GHz and 1.84 GHz corresponds to the independent test set obtained from the stratified hold-out procedure described earlier.

A clear trend can be seen across all three models. 1.84 GHz consistently provides the most reliable classification between “Good”, “Medium” and “Bad” oil samples. At this frequency, every classifier achieved perfect precision, sensitivity, specificity and accuracy. This result matches the earlier spectral analysis, where 1.84 GHz showed the largest separation in dielectric constant, ϵ_r' among the three oil categories.

Table 5. The evaluation metrics of heat transfer oil quality using SVM classifier for prediction set at 1.75 GHz and 1.85 GHz F-statistic and p-values result from ANOVA test

Variable Selection Frequency	Group	Good	Medium	Bad	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
1.84 GHz	Good	8	0	0	100.00	100.00	100.00	100.00
	Medium	0	7	0	100.00	100.00	100.00	
	Bad	0	0	3	100.00	100.00	100.00	
	Average				100.00	100.00	100.00	
1.75 GHz	Good	8	0	0	100.00	100.00	100.00	100.00
	Medium	0	7	0	100.00	100.00	100.00	
	Bad	0	0	3	100.00	100.00	100.00	
	Average				100.00	100.00	100.00	

Table 6. The evaluation metrics of heat transfer oil quality using RF classifier for prediction set at 1.75 GHz and 1.85 GHz F-statistic and p-values result from ANOVA test

Variable Selection Frequency	Group	Good	Medium	Bad	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
1.84 GHz	Good	8	0	0	100.00	100.00	100.00	100.00
	Medium	0	7	0	100.00	100.00	100.00	
	Bad	0	0	3	100.00	100.00	100.00	
	Average				100.00	100.00	100.00	
1.75 GHz	Good	8	0	0	88.89	100.00	90.00	94.44
	Medium	1	6	0	100.00	85.71	100.00	
	Bad	0	0	3	100.00	100.00	100.00	
	Average				96.30	95.24	96.67	

Table 7. The evaluation metrics of heat transfer oil quality using k-Nearest Neighbour classifier for prediction set at 1.75 GHz and 1.85 GHz F-statistic and p-values result from ANOVA test

Variable Selection Frequency	Group	Good	Medium	Bad	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
1.84 GHz	Good	8	0	0	100.00	100.00	100.00	100.00
	Medium	0	7	0	100.00	100.00	100.00	
	Bad	0	0	3	100.00	100.00	100.00	
	Average				100.00	100.00	100.00	
1.75 GHz	Good	8	0	0	80.00	100.00	80.00	88.89
	Medium	2	5	0	100.00	100.00	100.00	
	Bad	0	0	3	100.00	71.43	100.00	
	Average				93.33	90.48	93.33	

At 1.75 GHz, the classifiers behaved differently. SVM managed to maintain perfect scores, showing that it could still form a clean boundary between the three groups, even when the separation dielectric value was slightly weaker. By using a Radial Basis Function (RBF) kernel, SVM effectively handled the feature overlap by mapping the data into a higher-dimensional space.

The superior performance of the SVM model in this study is consistent with other high-precision diagnostics in oil monitoring. For instance, Sinduja et al. (2022) reported a recognition accuracy of 99.5% for transformer oil using SVM with Bayesian optimization. Similarly, Ye & Meng (2022) achieved 100% accuracy in edible oil classification using discrimination models. These findings proof the capability of SVM in handling material diagnostics.

RF and k NN, however, showed small performance drop which can be attributed to a combination of model limitations and small sample size. For RF, the overall accuracy decreased to 94.44%, mainly because some “Medium” samples were misclassified as “Good”. This is likely because RF uses hierarchical decision splits that can become less precise when the class boundaries are not strictly orthogonal. k NN showed a bigger performance decline, with accuracy at 88.89%, again caused by “Medium” samples being predicted as “Good”. The performance of the k NN classifier at 1.75 GHz was significantly influenced by the hyperparameter k , which was optimized to $k=3$ via Bayesian search. While $k=3$ provided perfect accuracy at 1.84 GHz, it resulted in a drop to 88.89% at 1.75 GHz. This performance decline is attributed to the local sensitivity of the k NN algorithm. At 1.75 GHz, the dielectric constant values for the “Good” and “Medium” categories exhibit more overlap. Two “Medium” samples were misclassified because their three nearest neighbors were predominantly from the “Good” group. While a larger k could theoretically provide a smoother decision boundary, it often leads to 'majority-class bias' in small, imbalanced datasets, potentially masking the critical “Bad” oil category.

The strong performance of SVM across both frequencies reflects its strength in creating stable decision margins, which works well when the dataset has only one or two key input features. RF's excellent results at 1.84 GHz also make sense, as the ensemble approach helps it remain robust even if individual trees encounter noisy patterns. On the other hand, k NN naturally becomes less reliable when the feature space is tight, because it heavily depends on the local distribution of samples.

Another interesting observation is that all models were able to correctly identify the “Bad” oil samples, even though this category had fewer samples compared to “Good” and “Medium” oils. This suggests that the dielectric signature of degraded oil is strong enough to stand out, especially around 1.84 GHz. “Bad” oil consistently shows a larger deviation in dielectric constant, ϵ_r' , making it easier for the models to recognize despite the smaller sample count.

In summary, these results highlight that 1.84 GHz is the most informative and reliable frequency for oil quality classification. While 1.75 GHz can still be used, especially with SVM, it is more sensitive to small variations between “Good” and “Medium” samples. The consistency and stability of the classification at 1.84 GHz strongly support its selection as the primary variable for the machine learning development.

For the incineration plant operation, the sensitivity of 100% for “Bad” samples represents a 'safety-first' diagnostic capability. It guarantees that oil reaching the end of its useful life is always identified, mitigating risks of thermal cracking or equipment downtime. Conversely, the 100% Specificity for 'Good' samples has direct economic benefits, as it prevents the premature replacement of healthy oil, thereby optimizing resource consumption. Finally, the successful identification of the 'Medium' category provides a strategic 'early warning' window. This allows maintenance teams to transition from reactive repairs to planned, condition-based maintenance (CBM), providing ample lead time for oil procurement and scheduling during non-peak operational hours.

Conclusion

This study demonstrates that the integration of dielectric characterization with machine learning (ML) provides a robust, non-destructive alternative for monitoring Heat Transfer Oil (HTO) quality. Analysis of the dielectric spectra revealed a clear categorical separation between “Good,” “Medium,” and “Bad” oil samples, particularly within the 1.74–1.86 GHz range. Among the traditional ML models evaluated, the Support Vector Machine (SVM) achieved a perfect 100% classification accuracy, outperforming Random Forest (RF) and k-Nearest Neighbour (kNN) at the lower frequency bands.

A primary finding of this work is the identification of 1.84 GHz as the optimal diagnostic frequency. Although the SVM model achieved perfect scores at both 1.75 GHz and 1.84 GHz, the latter is preferred for industrial deployment due to its significantly larger F-statistic and effect size (η^2). This increased statistical margin provides a vital "buffer zone" that ensures the diagnostic signal remains robust against industrial noise and minor chemical variations. Furthermore, the 100% consensus across all tested classifiers at 1.84 GHz confirms its reliability as a primary feature for future hardware development.

From a practical engineering perspective, the high accuracy achieved at a single frequency point simplifies the requirements for real-time monitoring. Instead of relying on expensive, bulky wideband Vector Network Analyzers (VNAs), future industrial sensors can be designed as compact, low-cost single-frequency microwave resonators tuned specifically to 1.84 GHz. This paves the way for a miniaturized device that can be permanently installed on HTO pipelines for continuous, autonomous health monitoring.

Despite these promising results, certain limitations remain. The current dataset ($N = 103$) is relatively small, which favored the use of traditional ML over deep learning architectures. Additionally, the inherent class imbalance “Bad” samples ($n = 15$) require further validation. Future research will focus on external validation using diverse oil formulations and samples from various incineration plants to confirm the global generalizability of the 1.84 GHz marker across different industrial environments.

Acknowledgements/ Funding

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (vot Q870).

Conflict of Interest Statement

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

Authors' Contributions

Study conception and design: Yee See Khoo; Funding acquisition: Yee See Khoo; Data collection: You Kok Yeow, Fahmiruddin Esa; Software and model development: Ong Pauline; Data analysis and interpretation: Yee See Khoo, Arslan Ahmed, Ong Pauline, You Kok Yeow; Draft manuscript preparation: Yee See Khoo; Manuscript revision: Fauziahanim Che Seman. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Generative AI in the Writing Process

During the preparation of this work, the authors used Gemini to improve the readability, grammatical accuracy, and language quality of the manuscript text during the peer-review revision phase. After using this service, the authors thoroughly reviewed and edited the content as needed and take full responsibility for the final content of the publication.

Data Availability/ Supplementary Materials

All data generated or analyzed during this study are included in this published article.

Ethics Statement

The authors declare that this research did not involve human or animal subjects. All experimental procedures were performed following the institutional Safety, Health, and Environmental (HSE) protocols of Universiti Tun Hussein Onn Malaysia.

References

- Alemayehu, B., Kota, A., Neidhard-Doll, A., Chodavarapu, V., & Subramanyam, G. (2021). Cloud-connected real-time oil condition monitoring of utility transformers using impedance spectroscopy. *Instrumentation Science and Technology*, 49(5), 509–520. <https://doi.org/10.1080/10739149.2021.1890119>
- Alharbi, A. F., Mohammad, F., Usman, M., Khan, N., & Abushiba, W. (2025). Thermophysical analysis of time-dependent magnetized Casson hybrid nanofluid flow (Cu + GO/Kerosene Oil) using Darcy-Forchheimer and thermal radiative models for industrial cooling applications. *Scientific Reports*, 15, 3398. <https://doi.org/10.1038/s41598-025-87743-9>
- Anderson, A. M., Bruno, B. A., & Smith, L. S. (2015). Viscosity measurement. *Mechanical Engineers'* <https://doi.org/10.24191/jmeche.v23i3.9956>

- Handbook (pp. 1–28). John Wiley & Sons. <https://doi.org/10.1002/9781118985960.MEH123>
- American Society for Testing and Materials. (2024). Test Method for Acid Number of Petroleum Products by Potentiometric Titration (ASTM D664). ASTM International. <https://doi.org/10.1520/D0664-24>
- American Society for Testing and Materials. (2020). Test Method for Determination of Water in Petroleum Products, Lubricating Oils, and Additives by Coulometric Karl Fischer Titration (ASTM D6304-20). ASTM International. <https://doi.org/10.1520/D6304-20>
- American Society for Testing and Materials. (2016). Test Method for Flash and Fire Points by Cleveland Open Cup Tester (ASTM D92-16B). ASTM International. <https://doi.org/10.1520/D0092-16B>
- American Society for Testing and Materials. (2018). Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity) (ASTM D445-18). ASTM International. <https://doi.org/10.1520/D0445-18>
- American Society for Testing and Materials. (2016). Test Method for Viscosity by Dip-Type Viscosity Cups (ASTM D4212-16). ASTM International. <https://doi.org/10.1520/D4212-16>
- Baklanov, A., Lankin, A., & Lankin, M. (2022). Development of a method of resonant series-parallel impedance spectroscopy for quality control of transformer oil. International Conference on Industrial Engineering, Applications and Manufacturing (pp. 1096–1101). IEEE Publisher. <https://doi.org/10.1109/ICIEAM54945.2022.9787265>
- Bassbasi, M., Hafid, A., Platikanov, S., Tauler, R., & Oussama, A. (2013). Study of motor oil adulteration by infrared spectroscopy and chemometrics methods. Fuel, 104, 798–804. <https://doi.org/10.1016/J.FUEL.2012.05.058>
- Berghian-Grosan, C., & Magdas, D. A. (2020). Raman spectroscopy and machine-learning for edible oils evaluation. Talanta, 218, 121176. <https://doi.org/10.1016/j.talanta.2020.121176>
- Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Das, S., Paramane, A., Chatterjee, S., & Mishra, P. (2024). Recurrence plot-aided deep learning framework for condition monitoring of transformer insulation using fourier transform infrared spectroscopy. IEEE Transactions on Industry Applications, 60(6), 8393–8402. <https://doi.org/10.1109/TIA.2024.3430254>
- Das, S., Paramane, A., Kumari, S., & Chatterjee, S. (2023). Deep learning aided classification of ageing condition of natural ester oils using FTIR analysis. IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation (pp. 1-6). IEEE Publisher. <https://doi.org/10.1109/SEFET57834.2023.10246005>
- Decoste, D., & Schölkopf, B. (2002). Training invariant support vector machines. Machine Learning, 46(1–3), 161–190. <https://doi.org/10.1023/A:1012454411458>
- Karadağ, K., Yucegonul, G., Kelley, S. S., & Karaogul, E. (2024). A novel chemometric learning of virgin and deep frying olive-oil by Fourier transform infrared spectroscopy (FT-IR). Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 15(2), 293–300. <https://doi.org/10.24012/dumf.1407248>

- Nazeer, M. (2025). Theoretical investigation on heat transfer rate of hydromagnetic transformer oil based multiphase fluid: comprehensive modeling and computational study [Preprint]. Research Square. <https://doi.org/10.21203/rs.3.rs-5809689/v1>
- Ordukaya, E., & Karlik, B. (2017). Quality control of olive oils using machine learning and electronic nose. *Journal of Food Quality*, 2017(1), 9272404. <https://doi.org/10.1155/2017/9272404>
- Pichler, J., Agocs, A., Pisarova, L., Minami, I., Frauscher, M., & Dörr, N. (2025). Comparative study of squalane products as sustainable alternative to polyalphaolefin: oxidation degradation products and impact on physicochemical properties. *Lubricants*, 13(2), 48. <https://doi.org/10.3390/lubricants13020048>
- Puspitasari, N., Rosmasari, R., Pratama, F. W., & Sulastri, H. (2022). Quality classification of palm oil varieties using Naive Bayes Classifier. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 13(1), 11–23. <https://doi.org/10.31849/digitalzone.v13i1.9773>
- Seeton, C. J. (2008). Viscosity-temperature correlation for liquids. *Proceedings of STLE/ASME International Joint Tribology Conference* (pp. 131–142). ASME Publisher. <https://doi.org/10.1115/ijtc2006-12139>
- Senoussaoui, M. E. A., Brahami, M., & Fofana, I. (2021). Transformer oil quality assessment using random forest with feature engineering. *Energies*, 14(7), 1809. <https://doi.org/10.3390/en14071809>
- Shell Oil Products US. (2019). Shell heat transfer oil S2 [Technical data sheet].
- Shen, Z., Li, Y., Zhong, L., & Su, W. (2025). The impact of different shaped pin-fins on the flow and heat transfer in the internal cooling structure of gas turbine blades trailing edges. *The Canadian Journal of Chemical Engineering*, 103(9), 4540–4553. <https://doi.org/10.1002/cjce.25617>
- Sinduja, M., Maheswari, R. V., & Vigneshwaran, B. (2022). Transformer oil quality assessment using machine learning techniques. *International Conference on Computer Communication and Informatics* (pp. 1-5). IEEE Publisher. <https://doi.org/10.1109/iccci54379.2022.9741071>
- Stogryn, A. P., Bull, H. T., Rubayi, K., & Iravanchy, S. (1995). The microwave permittivity of sea and fresh water. GenCorp Aerojet, Azusa, CA.
- Sukarin, S., Janthong, M., & Thongwik, S. (2025). Comparison of heat transfer from flue gas waste heat in a charcoal kiln using heat transfer oil in a helical tube heat exchanger. *Nanotechnology Perceptions*, 21(S1), 332-349. <https://doi.org/10.62441/nano-ntp.v21iS1.4784>
- Wiltschko, F., & Cheng, X. (2025). Experimental investigation of heat transfer to supercritical pressure fluid in tubes with rough inner surface. *ASME Journal of Heat and Mass Transfer*, 147(5), 051801. <https://doi.org/10.1115/1.4067608>
- Wu, M., Wang, G., & Liu, H. (2022). Research on transformer fault diagnosis based on SMOTE and random forest. *4th International Conference on Electrical Engineering and Control Technologies* (pp. 359–363). IEEE Publisher. <https://doi.org/10.1109/ceect55960.2022.10030548>
- Ye, Q., & Meng, X. (2022). Highly efficient authentication of edible oils by FTIR spectroscopy coupled with chemometrics. *Food Chemistry*, 385, 132661. <https://doi.org/10.1016/j.foodchem.2022.132661>
- You, K. Y., Mun, H. K., You, L. L., Salleh, J., & Abbas, Z. (2013). A small and slim coaxial probe for single rice grain moisture sensing. *Sensors*, 13(3), 3652–3663. <https://doi.org/10.3390/s130303652>
- Zhu, R., Pei, X., Zou, S., & Hou, L. (2025). Flow distribution of supercritical hydrocarbon fuel in parallel

regenerative cooling channels under non-uniform heat flux. *Propulsion and Energy*, 1(1), 4.
<https://doi.org/10.1007/s44270-024-00001-7>



© 2026 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-SA) license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)