Open Journal of Applied Sciences, 2026, 16(9), 3010-3032
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

Development and Data-Driven Internal Consistency of a Multi-Parameter Health Index for Power Transformer Condition Assessment

Kone Gbah*, Gbegbe Raymond

Laboratoire d'Ingénierie Electronique, d'Electricité et des Systèmes Embarqués (LIEESE), Institut National Polytechnique Félix Houphouët-Boigny (INPHB), Yamoussoukro, Côte d'Ivoire
Email: *gbah.kone@inphb.ci

How to cite this paper: Gbah, K. and Raymond, G. (2026) Development and Data-Driven Internal Consistency of a Multi-Parameter Health Index for Power Transformer Condition Assessment. *Open Journal of Applied Sciences*, **16**, 3010-3032.
https://doi.org/10.4236/ojapps.2026.169166

Received: August 10, 2026
Accepted: September 4, 2026
Published: September 7, 2026

Abstract

Reliable transformer condition assessment requires the combined interpretation of several diagnostic measurements, as no single test can fully describe the condition of the oil-paper insulation system. This study develops and validates a multi-parameter Health Index (HI) for power transformers using field data collected from 101 in-service units operating under different voltage levels, powers ratings and ages. The proposed method combines dissolved gas, moisture, dielectric performance, acidity, and cellulose-aging indicators to calculate a HI score ranging from 0 to 100. Based on this score, the transformers were divided into five condition groups, from very poor to very good. The results obtained show that 35 units are in very good condition, while 11 transformers fall within the most degraded range and require particular attention. In addition, the principal component analysis (PCA) was used to examine whether the different HI groups were consistent with the diagnostic patterns. The first two principal components explained 63.58% of the total variance and revealed a gradual transition from healthy units, characterized by high breakdown voltage and limited degradation, to the deteriorated transformers with increased moisture, dissolved gases, acidity, and aging products. K-means clustering revealed partial agreement with the HI groups, particularly for the extreme condition, while showing greater overlap among the intermediate categories. A case study involving two transformers also demonstrated that similar HI ranges may result from distinct degradation mechanisms. Consequently, it is necessary to interpret the HI score together with individual test results. The proposed approach provides a practical tool for utilities to assess transformer fleets, compare equipment condition, prioritize maintenance activities, and support risk-informed asset management decisions.

Keywords

Power Transformers, Health Index, Condition Assessment, Principal Component Analysis, Asset Management

1. Introduction

Electricity has become an indispensable component of modern society, supporting economic growth, industrial development, healthcare systems, transportation, communication networks, and the digital economy. As electricity demand continues to grow, electric utilities must ensure that power systems remain reliable and operate safely.

Among all the equipment in a power system, transformers are the most critical and expensive assets. Indeed, transformer failures can lead to power outages, equipment damage, and high maintenance costs. Therefore, maintaining their good operating condition is essential for reliable and secure power system operation. Hence, robust diagnostic methods are needed to accurately assess transformer condition and support maintenance decisions.

Transformer condition is commonly assessed using several diagnostic techniques. Dissolved gas analysis (DGA) helps identify thermal and electrical faults by examining gases generated within the oil-paper insulation system [1]. Moisture measurement provides information about water contamination, which can reduce dielectric strength and accelerate cellulose aging [2]. Breakdown voltage is used to evaluate dielectric performance, while acidity reflects oil oxidation and the formation of polar degradation products. Carbon oxides and furanic compounds provide additional information on cellulose degradation and insulation aging [3].

Although conventional diagnostic techniques provide useful information, they are often used independently. However, transformer degradation is a complex process driven by the combined effects of electrical, thermal, and environmental stresses on the oil-paper insulation system, leading to multiple degradation mechanisms that often occur simultaneously. Consequently, using a single diagnostic technique may not provide a complete picture of transformer condition, since each method focuses on a specific aspect of insulation aging [4] [5].

To overcome these limitations, methodologies based on the HI have attracted considerable interest in recent years [6] [7]. An HI combines several diagnostic indicators into a single numerical value that reflects the global condition of transformer. Compared to individual diagnostic tests, this approach provides a more comprehensive assessment of insulation degradation. It also facilitates comparisons among transformers within the same fleet, and simplifies maintenance planning, asset replacement decisions, and long-term investment strategies [8]. However, the reliability of an HI depends strongly on the selected indicators, scoring, and weighting factors. These elements may vary depending on equipment design, available measurements, operating practices, and environmental conditions [9].

Furthermore, many published approaches are based mainly on expert rules or on limited datasets, without sufficient field evaluation to determine whether the resulting state groups accurately reflect the measured data.

This study develops a multi-parameter health index based on diagnostic data from 101 in service power transformers operating under different service conditions. The method integrates the analysis of dissolved gases, moisture content, dielectric performance, acidity, and indicators of cellulose aging into a normalized score ranging from 0 to 100. Based on this value, transformers are classified into five condition groups, ranging from "very poor" to "very good". The proposed framework aims to preserve the diagnostic value of each test while providing a clearer representation of the equipment's total condition.

Principal component analysis (PCA) is used to assess whether the clusters formed by the proposed HI correspond to the main degradation patterns contained in the diagnostic data. The transformers are displayed on the PC1-PC2 plot according to their previously assigned HI categories. Clustering using the K-means method permits an additional comparison between the groups defined by indices and the natural structure of the dataset. Principal component analysis (PCA) and K-means clustering have been used to reduce correlated variables, identify dominant degradation patterns, and facilitate the interpretation of multidimensional measurements [10].

The main contribution of this work lies in a practical and field-based framework that combines traditional transformer diagnostic methods with multivariate statistical analysis. Unlike approaches that are limited to calculating an overall score, the proposed methodology makes it possible to assess whether the five HI groups reflect significant differences. The index obtained in this way provides a consistent basis for comparing fleets and setting maintenance priorities, while retaining the diagnostic information needed to identify major problems and select appropriate corrective actions.

2. Overview of Diagnostic Parameters

2.1. Dissolved Gas Analysis (DGA)

Dissolved Gas Analysis (DGA) is one of the most used techniques for assessing the condition of oil-filled power transformers. Electrical and thermal stresses inside the transformer decompose the oil-paper insulation system, generating gases such as hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2) [1] [11]. Each gas or combination of gases is associated with specific types of defects. Several methods are used to interpret the results of dissolved gas analysis (DGA), including the key gas method, gas ratio methods like the Rogers ratio, and graphical methods such as the Duval triangle [12] [13]. In addition, the international standards IEC 60599 [11] and IEEE C57.104 [1] provide guidelines for interpreting dissolved gas concentrations and identifying transformer faults. Although discharge gas analysis (DGA) is an effective diagnostic tool, it has certain limitations. The results can be difficult to interpret when multiple faults occur simultaneously, and periodic oil samples may not be sufficient to detect rapidly developing faults. Consequently, DGA is often combined with other diagnostic techniques to provide a more complete assessment of transformer condition [4] [14].

2.2. Furan Analysis

Furan analysis is widely used to assess the aging of paper insulation in power transformers. As paper insulation ages under thermal and oxidative stress, cellulose gradually breaks down and releases furan compounds into the insulating oil. These compounds include 2-furfural (2-FAL), 5-hydroxymethyl-2-furfural (5-HMF), and other related compounds [15] [16]. Among these, 2-furfural (2-FAL) is the most used indicator for assessing the aging of paper-based insulation materials, as it is relatively stable and closely linked to cellulose degradation [17]. While DGA analysis detects electrical and thermal defects, furan analysis provides direct information on the condition of paper insulation. Since the concentration of 2-furfural (2-FAL) is closely related to the degree of polymerization (DP) of cellulose, it is commonly used to assess the aging of the insulation and estimate the transformer's remaining service life [3] [18]. Although furan analysis is primarily used to assess the long-term aging of paper insulation, it is less effective for detecting rapidly developing failures. In addition, factors such as oil replacement, operating temperature, and humidity can affect furan concentrations. For this reason, furan analysis is often combined with other diagnostic techniques, such as DGA and moisture measurement, to obtain a more comprehensive assessment of transformer condition [19].

2.3. Oil Dielectric Breakdown Voltage (BDV)

BDV test is employed to evaluate the insulating performance of transformer oil. Over time, thermal, electrical, and chemical aging gradually degrades the oil, leading to the formation of moisture, sludge, dissolved gases, and other contaminants that reduce its dielectric strength [20]. According to IEC 60156, the test consists of applying an alternating voltage that increases continuously at a rate of 2 kV/s between two electrodes immersed in a sample of oil until dielectric breakdown occurs. To ensure the reliability of the results, six consecutive breakdown tests are performed, and the average of the measured breakdown voltages is reported as the BDV value of oil [21]. The relationship between oil dielectric strength and the BDV value is described in the IEEE C57.106 standard [22]. In practice, a BDV value below 30 kV generally indicates significant oil degradation and requires additional investigations as well as corrective maintenance measures [23] [24].

2.4. Acidity

The acids present in the oil come from its decomposition or oxidation. They may also come from external sources, such as atmospheric contamination. The measure is made with Color-Indicator Titration technique according to ASTM D974.

The standard unit of measure is mg KOH/g. During transformer operation, oxidation gradually generates polar byproducts, particularly organic acids, which leads to an increase in the acid number (AN). Fresh oil insulation contains little or no acidic compounds, whereas a high AN value in service aged oil generally indicates oxidation or contamination by materials such as varnish, paint, or other foreign substances. Consequently, the acid number is widely recognized as a valuable indicator of oil aging, oxidation, nitration, and contamination, making it an important parameter for assessing the condition of transformers. According to industry recommendations, oil regeneration should be considered when the acid number reaches 0.20 mg of KOH/g [23].

2.5. Moisture Content

Moisture is one of the main factors affecting the performance and service life of transformer insulation systems. It is present in both insulating oil and paper insulation, where it reduces dielectric strength and accelerates insulation aging [25]. At high temperatures, it can also cause the formation of bubbles, which increases the risk of insulation failure [25]. The moisture content of transformer oil is generally measured in parts per million (ppm) using a Karl Fischer titration. The distribution of moisture between the insulating oil and the paper insulation depends on the temperature and the moisture equilibrium within the insulation system [26].

Although the presence of moisture does identify specific transformer faults, it significantly accelerates insulation aging and degradation. As a result, moisture is an important parameter for assessing the full condition of the transformer [19].

3. Health Index Development

Transformer insulation deteriorates under multiple interacting mechanisms. Consequently, assessing its condition based on a single diagnostic parameter can lead to incomplete conclusions. To provide a more comprehensive assessment, this study develops a multiparameter HI that integrates complementary diagnostic measurements into a single condition indicator. It was developed based on field measurements collected from 101 power transformers in service within the Ivorian transmission grid.

The methodology consists of four successive steps: data preprocessing, feature engineering, calculation of the condition index, and condition classification.

3.1. Dataset and Preprocessing

3.1.1. Dataset Description

The dataset used in this study was collected from 101 power transformers in service on the Ivorian transmission grid. Sampling of 101 transformers was conducted between September 1 and September 15, 2025. The 101 transformers were inspected as part of the public utility's regular preventive maintenance program. The measurements were taken from October 6 to October 10, 2025. A complete observation consists of 17 variables. These transformers cover a wide range of voltage levels, powers ratings, and operating conditions, providing a realistic overview of transformer performance in the field.

The fleet includes transformers that were commissioned between the mid-1960s and recent years, exhibiting varying degrees of aging and insulation conditions. Their technical characteristics range from small distribution transformers to large transmission units operating at voltages up to 225 kV. This diversity makes the dataset particularly well-suited for studying various degradation mechanisms, since transformer aging is influenced by operating history, load conditions, and environmental factors [2].

To obtain a more comprehensive assessment of the transformer's condition, diagnostic parameters from several categories were considered and grouped into three diagnostic categories, as shown in **Table 1**.

Table 1. Categories of diagnostic parameters considered in this study.

Category	Parameters	Units	Diagnostic significance
Dissolved Gas Analysis	H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, CO_2	ppm	Thermal and electrical fault detection
Oil quality	Moisture	ppm	Moisture contamination and accelerated insulation aging
	Acidity	mg KOH/g	Formation of acidic by-products
	Breakdown voltage	kV	Dielectric strength of insulation oil
Paper insulation	2-Furfural (2-FAL)	mg/L	Cellulose degradation and estimation of remain life

Each diagnostic parameter provides information about a different aspect of transformer degradation. When used together, they allow a more complete and reliable assessment of transformer condition [19].

3.1.2. Data Cleaning and Reconstruction

Field data collected from operating transformers often contains missing values, inconsistent data formats, and incomplete historical information. Consequently, the dataset has been preprocessing to improve data quality and ensure consistent results. First, all diagnostic data were organized into a structured database, in which each transformer was treated as a separate observation and each diagnostic measurement served as an input feature.

3.1.3. Missing Value Treatment

Some diagnostic measurements were missing from the dataset. To retain all records related to transformers and ensure consistent analysis, the missing values were replaced using median imputation, as shown in Equation (1):

$$x_i^{imp} = \begin{cases} x_i & x_i \neq NaN \\ median(X) & x_i = NAN \end{cases} \quad (1)$$

Median imputation was used because transformer diagnostic data often contains extreme values associated with failure conditions. Unlike the mean, the median is much less influenced by these values, making it a more robust choice for handling missing data [27].

The number of missing values for each seven variables was as follows: TDCG: (0); BDV: (0); moisture: (0); 2-FAL: (0); CO_2: (0); acidity: (0); and transformer age: (4). Median imputation was performed before calculating the HI and conducting multivariate analysis. The median for each variable was calculated based on the available observations for the fleet. The variables were then standardized using z-scores prior to PCA analysis and K-means clustering.

3.2. Additional Features

To improve diagnostic performance, two additional features (transformer age and gas aggradation indicator) were derived from the initial measurements through.

Transformer age was included as an additional parameter, since insulation degradation is largely related to cumulative operating time and thermal aging [27]. It was calculated using Equation (2):

$$Age_i = Y_{ref} - Y_{comm,i} \quad (2)$$

where Age_i denotes the age of transformer *i*, Y_{ref} is the reference year used in this study, and $Y_{comm,i}$ is the commissioning year of transformer *i*.

The Total concentration of Dissolved Combustible Gas (TDCG) was used as an complete indicator of gas production associated with transformer failures. In accordance with the definition given in Equation (3), the TDCG value was calculated as the sum of the main combustible gases dissolved in the insulating oil. To further improve the interpretation of failures, gas aggregation indicators were also considered [28] [29].

$$TDCG = \sum_{i=1}^{n} G_i \quad (3)$$

where $TDCG$ is the total dissolved combustible gas concentration (ppm), G_i denotes the concentration of the i^{th} combustible gas (ppm), and $n = 6$ (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, and CO) is the number of combustible gases considered.

3.3. Health Index Formulation (HI)

Transformer degradation is influenced by several electrical, thermal, and chemical stresses that occur simultaneously. To capture these various aspects into a single indicator, a multiparameter HI was developed by [30] [31]. This approach estimates the HI of transformers through a three-step evaluation process. First, the dissolved gas analysis factor (DGAF) is calculated with the concentrations of the seven key dissolved gases using Equation (4):

$$DGAF = \frac{\sum_{i=1}^{7} S_i W_i}{\sum_{i=1}^{7} W_i} \tag{4}$$

where S_i is the condition score assigned to the ith dissolved gas according to the applicable diagnostic standard, W_i is the weighting factor reflecting its relative importance, and *i* represents the seven characteristic gases (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, CO, and CO_2).

Next, the Oil Quality Factor (OQF) is determined by evaluating the main physicochemical properties of the insulating oil, including dielectric breakdown voltage, acidity, and moisture content. The OQF is calculated using the weighted average formulation given in Equation (5):

$$OQF = \frac{\sum_{i=1}^{3} S_i W_i}{\sum_{i=1}^{3} W_i} \tag{5}$$

where S_i denotes the condition score of the i^{th} oil quality parameter, W_i is its corresponding weighting factor, and irefers to the three oil properties considered in the assessment.

The weighting factors assigned to each diagnostic parameter and the corresponding score are presented in Table 2 and Table 3 for the DGAF and the OQF, respectively.

Table 2. DGAF parameters [30] [32].

Gas (ppm)	Score (S_i)						Weight (W_i)
	1	2	3	4	5	6	
H_2	≤100	100 - 200	200 - 300	300 - 500	500 - 700	>700	2
CH_4	≤75	75 - 125	125 - 200	200 - 400	400 - 600	>600	3
C_2H_6	≤65	65 - 80	80 - 100	100 - 120	120 - 150	>150	3
C_2H_4	≤50	50 - 80	80 - 100	100 - 150	150 - 200	>200	3
C_2H_2	≤3	3 - 7	7 - 35	35 - 50	50 - 80	>80	5
CO	≤350	350 - 700	700 - 900	900 - 1100	1100 - 1400	>1400	1
CO_2	≤2500	2500 - 3000	3000 - 4000	4000 - 5000	5000 - 7000	>7000	1

Table 3. oil quality parameters [30] [32].

Oil parameters	U < 69 kV	69 kV ≤ U < 230 kV	U ≥ 230 kV	Score (S_i)	Weight (W_i)
Breakdown voltage (kV)	≥45 kV	≥52	≥60	1	3
	35 - 45	45 - 52	50 - 60	2	
	30 - 35	35 - 45	40 - 50	3	
	≤30	≤ 35	≤40	4	
Acidity (mg KOH/g)	≤0.05	≤0.04	≤0.03	1	1
	0.05 - 0.1	0.04 - 0.1	0.03 - 0.07	2	
	0.1 - 0.2	0.1 - 0.15	0.07 - 0.1	3	
	≥0.2	≥0.15	≥0.4	4	

Continued

Water content (ppm)	≤20	1	4
	20 - 30	2	
	30 - 40	3	
	>40	4	

Similarly, the evaluation of paper insulation condition using furan is presented in Table 4. The score of DGAF and OQF calculated are also classified into five levels, as shown in Table 5.

Finally, the global HI is obtained by combining the different diagnostic indicators into a single metric, as shown in Equation (6). Details of HIFj and Kj indices for each parameter are presented in Table 6. Table 7 presents the final assessment of transformers condition based on the transformer HI methods.

Table 4. Furan factor ranking [30].

2 FAL (ppm)	Rating code	Condition
0 - 0.1	A	Good
0.1 - 0.25	B	Acceptable
0.25 - 0.5	C	Need caution
0.5 - 1	D	Poor
>1	E	Very poor

Table 5. Assement the score of DGAF and OQF [30] [32].

Rating code	Condition	Description
A	Good	x < 1.2
B	Acceptable	1.2 ≤ x < 1.5
C	Need caution	1.5 ≤ x < 2
D	Poor	2 ≤ x < 3
E	Very poor	x ≥ 3

$$HI(\%) = 100\left(\frac{\sum_{j=1}^{n} K_j HIF_j}{4\sum_{j=1}^{n} K_j}\right) \tag{6}$$

where K_j is the weight factor indicating the significance of each input to the final health index and HIF_j is the health index factor of each input, as shown in Table 6.

Table 6. Health index assessment [30] [32].

Transformer parameters	Kj	Rating code	HIFj
DGAF	10	A, B, C, D, E	4, 3, 2, 1, 0
OQF	8	A, B, C, D, E	4, 3, 2, 1, 0
Furan	5	A, B, C, D, E	4, 3, 2, 1, 0

Table 7. Health index final assessment [30] [32].

Health index	Condition	Recommended action
85 - 100	Very good	Routine monitoring
70 - 85	Good	Normal monitoring
50 - 70	Fair	Additional testing and shorter monitoring interval
30 - 50	Poor	Detailed investigation and corrective maintenance
0 - 30	Very poor	Immediate assessment and intervention planning

4. Results and Discussion

4.1. Relationship between Moisture and Dielectric Degradation

Figure 1 illustrates the relationship between water saturation and tan δ measured at 90˚C. Most transformers have a water saturation of less than 20% and exhibit relatively low tan δ values. However, tan δ generally tends to increase with water saturation. Several transformers with higher moisture levels also exhibit higher dielectric losses, indicating that moisture may contribute to the degradation of the oil-paper insulation system. Water can increase electrical conductivity and polarization losses, thereby altering the dielectric response of insulating materials. Previous studies have also shown that the dielectric behavior of transformer insulation is strongly influenced by humidity and temperature [33]. Nevertheless, this relationship is not consistent across all transformers. Some units with water saturation exceeding 20% maintain low tan δ values, while a transformer with water saturation of approximately 17% exhibits an unusually high dissipation factor. These observations indicate that moisture may contribute to an increase in dielectric losses, but it is not the only factor involved. Oil oxidation, polar contaminants, insulation aging, temperature can also affect tan δ.

From the perspective of the health index, water saturation and tan δ should be considered complementary indicators rather than interchangeable ones. Water saturation describes the moisture condition of the insulation system, while tan δ reflects the entire dielectric losses associated with moisture, aging byproducts, and contamination. Including both parameters thus provides a more complete picture of insulation degradation.

In sum, **Figure 1** shows that higher water saturation is generally associated with greater dielectric losses. However, neither of these two parameters, taken alone, can't reliably describe the condition of the transformer's insulation. Combining them with other physicochemical and electrical indicators therefore provides a more reliable assessment.

4.2. Relationship between Cellulose Aging and Gas Generation

Figure 2 illustrates the relationship between carbon monoxide (CO) concentration and 2-furfural (2-FAL) content. Some transformers with a 2-FAL content of less than 1 mg/L have CO concentrations ranging from less than 100 ppm to more than 1,400 ppm. This wide dispersion indicates that the two indicators do not exhibit a strong linear relationship.

Other units with 2-FAL concentrations exceeding 2 mg/L also show relatively high CO concentrations, suggesting advanced degradation of the cellulose insulation. However, this trend is not consistent across the entire dataset. For example, transformers with 2-FAL levels above 3 mg/L have CO concentrations from approximately 450 to 1,200 ppm. Similarly, some units exhibit high CO levels despite relatively low 2-FAL levels. These differences may be related to operating temperature, oxygen availability, oil replacement or treatment, paper type, loading history, and the migration of degradation products between the solid insulator and the oil [34].

Although both variables are associated with cellulose aging, they reflect different degradation processes. CO is primarily produced by the thermal decomposition and oxidation of paper insulation. In contrast, 2-FAL is a more specific marker of cellulose depolymerization, although its concentration can be affected by oil replacement, moisture, temperature, and adsorption by the paper.

In the context of HI, CO and 2-FAL should be treated as complementary indicators rather than interchangeable ones as explained earlier.

The entire graph suggests a certain correlation between high CO concentrations and high 2-FAL levels in several old transformers, but this relationship remains weak and inconsistent across the population of transformers. Presenting Spearman's correlation coefficient would allow us to quantify this association while limiting the influence of outliers.

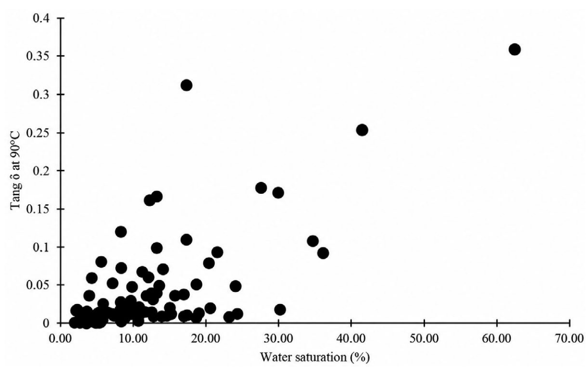

Figure 1. Relationship between water saturation and the dielectric dissipation factor (tan δ) measured at 90˚C.

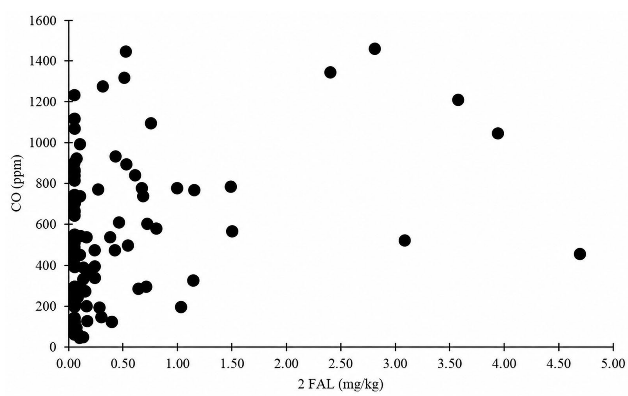

Figure 2. Cellulose aging: 2-FAL versus CO.

4.3. Correlation Analysis of Diagnostic Indicators

Figure 3 shows the Pearson correlations between the selected transformer condition indicators and the HI. The results reveal several significant relationships between oil quality, cellulose aging, gas generation, and the global condition of the transformers.

The HI index shows its strongest correlations with humidity (r = −0.73), total dissolved combustible gases (TDCG, r = −0.70), CO_2 (r = −0.67), and 2-FAL (r = −0.62). Moderate negative relationships are also observed with age (r = −0.57), acidity (r = −0.49), and CO (r = −0.47). In contrast, the BDV shows a positive correlation with the HI index (r = 0.55). These results indicate that higher moisture content, gas concentrations, acidity, and furan levels, and transformer age, are generally associated with lower HI index values, while higher dielectric strength corresponds to a better condition of the transformer. Several condition indicators are also closely related. Moisture shows a positive correlation with 2-FAL (r = 0.68), acidity (r = 0.62), CO_2 (r = 0.54), and age (r = 0.52). These relationships are consistent with the combined effects of moisture accumulation, oil oxidation, and cellulose aging. BDV shows a negative correlation with moisture (r = −0.57), CO_2 (r = −0.37), 2-FAL (r = −0.36), and TDCG (r = −0.32). This suggests that dielectric strength generally decreases as humidity and aging-related byproducts accumulate. However, these correlations remain moderate, confirming that the BDV is influenced by several factors and cannot independently describe the overall condition of the insulation system.

This correlation analysis shows that transformer aging results from several interconnected electrical, thermal, chemical, and moisture-related processes. This finding justifies the use of a multiparameter diagnostic framework.

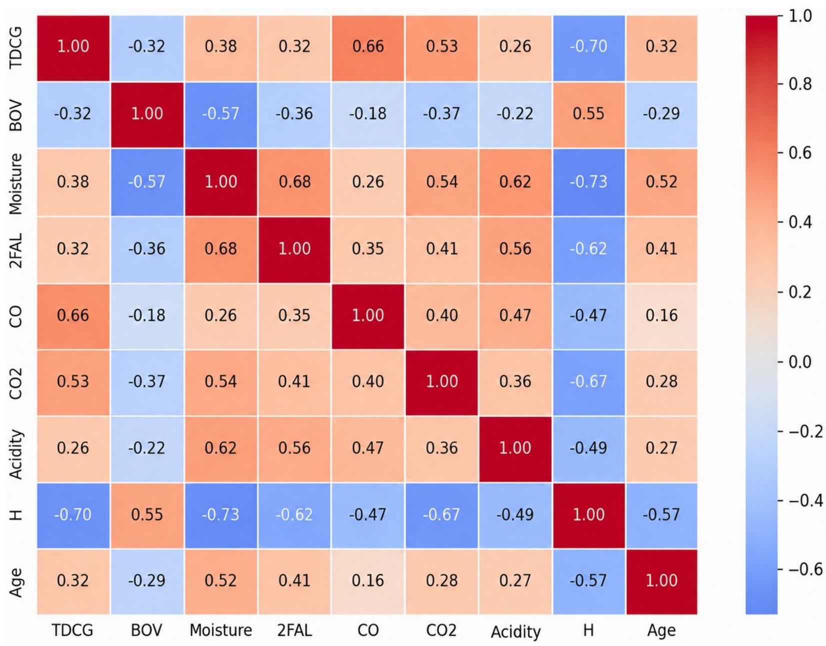

Figure 3. Pearson correlation matrix of transformer condition indicators and the HI.

4.4. Health Index-Based Condition Classification

Table 8 shows the HI scores for 20 transformers. Based on the proposed HI, the 101 transformers were grouped into five condition categories. As shown in Figure 4, the largest group consists of 35 units with HI scores ranging from 85 to 100, indicating that approximately one-third of the fleet is in very good condition. The intermediate ranges of 30 to 50 and 50 to 70 contain 20 and 21 transformers, respectively, while 14 units fall within the 70 to 85 range. Only 11 transformers have scores below 30, representing the smallest group, but also the one of greatest concern.

The distribution suggests that the global condition of the transformer fleet is generally satisfactory, although a considerable number of units show moderate to significant deterioration. Transformers with lower HI scores may be affected by several interacting factors, such as increased humidity, dielectric losses, oil degradation, dissolved gases, and cellulose aging. These units require closer monitoring and additional testing or corrective actions. By combining multiple condition indicators into a single index, the proposed HI provides a clear overview of the fleet while helping to identify the transformers that require the most attention. This classification can facilitate inspection planning, maintenance prioritization, and long-term asset management decisions.

Table 8. HI scores for 20 transformers.

Transformer	Age (years)	TDCG (ppm)	CO_2 (ppm)	BDV (kV)	Moisture (mg/kg)	Acidity (mg KOH/g)	2-FAL (mg/L)	HI (%)
T1	47	2105	10,172	51.8	10	0.0099	0.16	42.39
T2	11	436	3187	66.5	6	0.0099	0.0499	100
T3	37	6.5	3468	66.9	11	0.023	0.0499	78.26
T4	10	199	695	66.9	8	0.0099	0.0499	100
T5	16	774	19,835	37.7	27	0.019	0.15	35.87
T6	52	153	1616	69.8	21	0.0099	0.0499	82.61
T7	33	1836	1554	55.6	29	0.113	0.16	33.7
T8	50	172	1432	75.5	17	0.0099	0.0499	100
T9	41	1038	12,258	95.7	11	0.059	0.53	51.09
T10	54	508	6389	52.3	26	0.206	0.42	60.87
T11	47	2240	16,787	71.8	27	0.055	0.24	44.57
T12	41	226	3133	66.2	15	0.0099	0.28	89.13
T13	46	1311	6481	63	9	0.036	0.72	40.22
T14	13	992	1630	78.3	7	0.0099	0.0499	89.13
T15	46	577	1489	66.9	9	0.0099	0.08	67.39
T16	13	504	6582	63.5	6	0.0099	0.0499	89.13
T17	11	305	1106	72.4	5	0.0099	0.0499	100

Continued

T18	66	1245	4850	31.4	60	0.103	4.69	10.87
T19	31	1463	8539	46.2	45	0.411	0.51	27.17
T20	50	386	5965	63.5	19	0.139	0.24	75

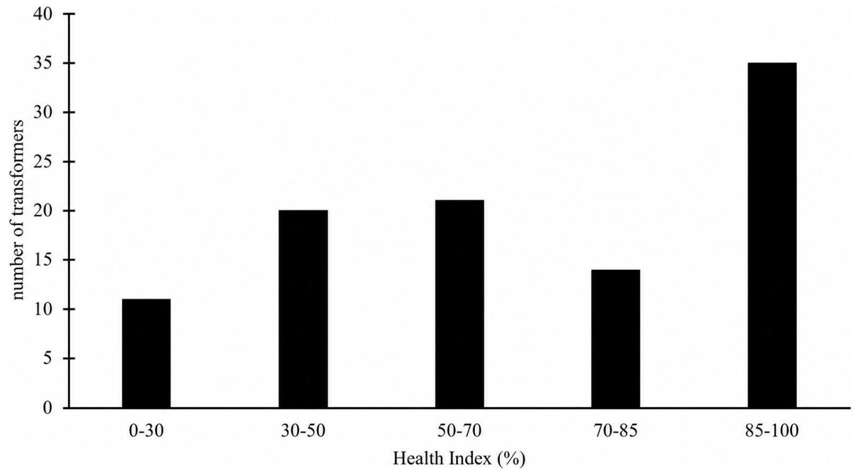

Figure 4. Distribution of transformers across health index condition ranges.

4.5. PCA- and K-Means-Based Assessment of Internal Consistency of the HI Groups

Principal component analysis (PCA) and the K-means clustering method were used to assess whether the five condition categories defined by the proposed HI corresponded to the multivariate diagnostic structure of the transformer fleet. Given that the variables used in PCA and the K-means algorithm largely overlap with the diagnostic data from the HI index, this analysis was not intended to serve as an independent validation. Rather, it aimed to determine whether the multivariate structure of the underlying measures corresponded to the categories of conditions established by the index.

PCA was performed on seven diagnostic variables including total dissolved combustible gas (TDCG) concentration, breakdown voltage (BDV), moisture, 2-furfural (2-FAL), carbon dioxide (CO_2), acidity, and transformer age. Before analysis, the variables were standardized using z-scores, as they are expressed in different units and have widely varying numerical ranges. This transformation gives each variable equal initial weight and prevents high amplitude measurements to dominate the principal components.

The first two principal components explain 63.58% of the total variance, with PC1 accounting for 50.19% and PC2 for 13.39%. PC1 reflects the entire condition of transformer insulation system as shown in Figure 5. Moisture, 2-FAL, acidity, CO_2, TDCG, and age all have a positive effect on this parameter, while BDV has a negative effect. PC1 can therefore be interpreted as a measure of general degradation axis. The PC2 component primarily displays differences between indicators of gas and oil acidity. TDCG and CO_2 have positive loading coefficients, while acidity has a strongly negative loading coefficient. This component makes it possible to distinguish transformers characterized by gas production from those exhibiting oil oxidation.

Figure 6 shows the transformers scores on the PC1-PC2 plane, with observations colored according to their HI category. It reveals a clear transition in the condition of the transformers along PC1. Units classified as "very good" are concentrated on the negative side, while most of those classified as "very poor" appear at high positive values. Good, fair, and poor transformers fall mainly between these two extremes. This gradual progression supports the HI classification, since a deterioration in condition is generally associated with movement from left to right. PC2 highlights the differences among units with a similar general condition. Its greater dispersion within the "Poor" and "Very Poor" groups suggests that severely damaged transformers do not exhibit the same degradation profile. Some may be primarily affected by gas production, while others may indicate more pronounced effects related to moisture, acidity, or paper aging.

The adjacent categories overlap particularly Good, Fair and Poor, because the deterioration of transformers occurs gradually rather than in distinct stages. The extreme groups are more clearly distinguished, while the intermediate categories reflect transitional stages.

In general, the graph provides significant statistical support for the health index groups, without suggesting that they constitute entirely distinct populations.

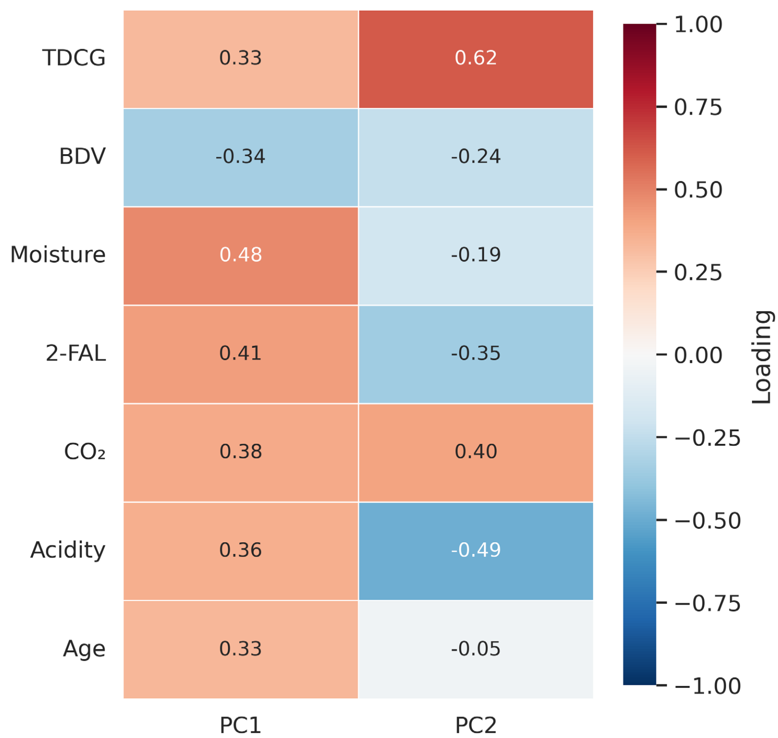

Figure 5. Variable loadings on PC1 and PC2.

The K-means clustering method was applied directly to the seven normalized diagnostic variables: TDCG, BDV, moisture, 2-FAL, CO_2, acidity, and transformer age. The analysis was performed in Python using scikit-learn version 1.8.0. The algorithm used k-means++ initialization, 50 independent initializations, a maximum of 300 iterations, a convergence tolerance of 10−4, and a fixed random seed of 42 to ensure reproducibility.

Solutions ranging from k = 2 to k = 7 were evaluated using the silhouette coefficient, the Davies-Bouldin index, the Calinski-Harabasz index, and inertia as shown in Figure 7. The highest silhouette coefficient (0.347) and Calinski-Harabasz index (48.104) were obtained for k = 2, indicating that two clusters provide the clearest natural separation in the dataset. However, the proposed HI classifies the condition of transformers into five operational categories. Consequently, the solution k = 5 was selected to assess the consistency between the clusters resulting from the data analysis and these predefined categories. For k = 5, the silhouette coefficient was 0.297 and the Davies-Bouldin index was 1.315, indicating moderate separation in the seven-dimensional feature space.

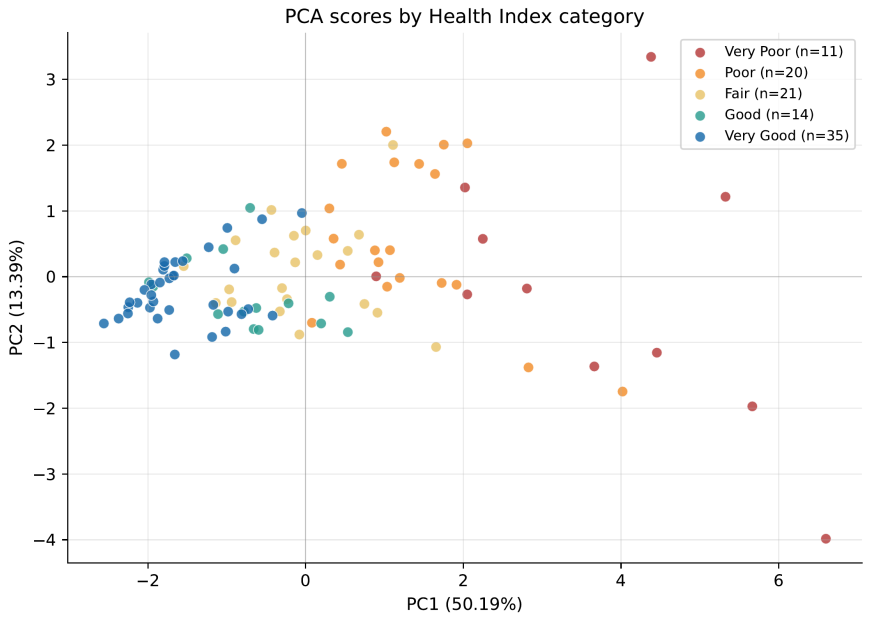

Figure 6. PCA score plot colored by health index category.

Figure 8 presents the K-means seven variables projection. The five clusters contained 31, 31, 30, 3, and 6 transformers, respectively. Their average HI values were 96.6%, 76.3%, 45.1%, 31.2%, and 15%, respectively, indicating a clear overall progression in the condition of transformers, from best to worst. The agreement with the five categories of the health index was partial, as shown in Table 9, with an adjusted Rand index of 0.312, a normalized mutual information score of 0.407, and an optimal agreement rate of 60.4%. The strongest agreement was observed for the “Very Good” and “Very Poor” categories, while greater overlap was found among the intermediate categories.

For visualization, the clusters obtained from the seven variables were projected onto the PC1-PC2 plane to illustrate the cluster structure; the K-means analysis, meanwhile, was performed in the full seven-dimensional space.

On the whole, principal component analysis (PCA) and the K-means algorithm provide data-driven evidence that the proposed health status categories follow a technically significant status gradient within the transformer fleet. The results support the consistency of the healthiest groups and most deteriorate, while revealing gradual transitions and multiple degradation profiles among the intermediate categories. Consequently, this analysis confirms the internal statistical consistency of the HI classification.

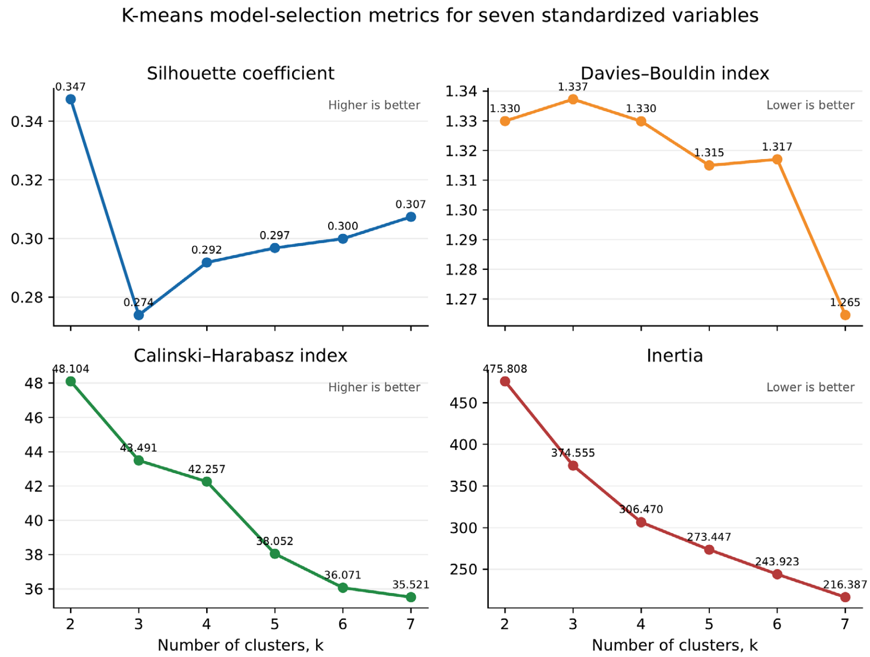

Figure 7. K-means model-selection metrics for seven standardized variables.

Table 9. Agreement between Health Index Categories and K-Means Cluster Assign-ments.

Category HI	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	total
Good	0	0	2	3	6	11
Acceptable	0	2	6	11	1	20
Need caution	1	5	13	2	0	21
Poor	5	8	1	0	0	14
Very poor	23	8	4	0	0	35
Total	29	23	26	16	7	101

4.6. Case Study

To illustrate the practical application of the proposed HI, two transformers with different operating conditions were examined. Table 10 presents the values determined based on the complete (electrical and physicochemical) analysis of oil.

Transformer T1 shows high concentrations of dissolved gases, related to thermal activity and possible localized overheating. However, its moisture and acidity levels remain low, indicating limited aging of the insulation. Nevertheless, its breakdown voltage and oil leaks should continue to be closely monitored. Transformer T7 is in poorer condition. Its test results indicate the possible presence of electrical discharges, high moisture content, oil degradation, particle contamination, and leaks. The condition index confirms this difference: T1 scored 42.39% and T7 scored 33.70%, placing both units in the "Poor" category, with T7 requiring special attention.

Although conventional techniques are essential for identifying individual defects, interpreting their results separately may not provide a clear picture of the transformer's entire condition. The proposed HI overcomes this limitation by integrating the available diagnostic information into a single condition score. The index thus preserves the diagnostic value of each test while making it easier to identify differences in severity level. This case study highlights its usefulness as a complementary decision-making tool for comparing units, establishing maintenance priorities, and selecting transformers that require further examination.

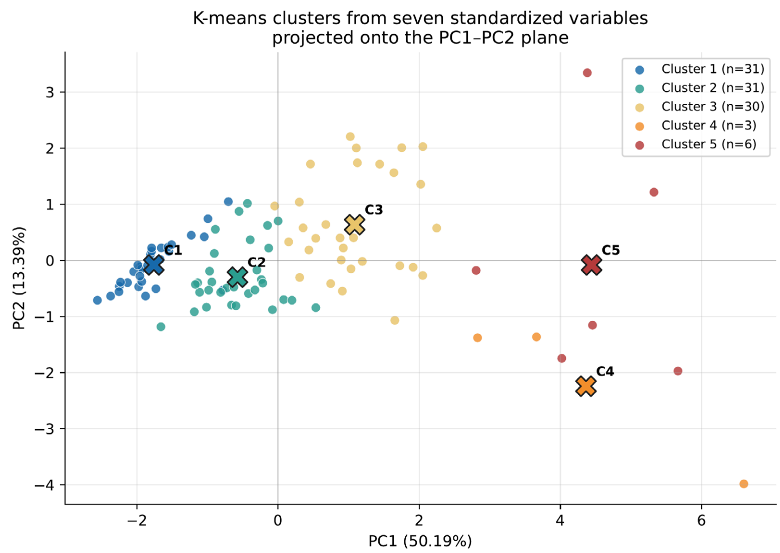

Figure 8. K-means clusters obtained from the seven standardized diagnostic variables and projected onto the PC1-PC2 plane. The X markers represent the projected cluster centroids.

4.7. Practical Implications for Asset Management

The proposed HI provides asset managers a way to assess the global condition of a transformer fleet. By combining the results of dissolved gas analysis, acidity, moisture content, dielectric performance, and insulation aging indicators into a single index, this method facilitates comparisons between transformers with different operational histories and degradation profiles. Transformers with high HI values may continue to be monitored regularly, while those in an intermediate condition may undergo additional testing or preventive maintenance. Units with low ratings should be prioritized for a thorough inspection, corrective action, or replacement planning. The HI index proves particularly useful when conventional tests yield mixed results. For example, one transformer may have high concentrations of combustible gases while still maintaining acceptable oil quality, whereas another may show moderate abnormalities across several diagnostic parameters. Evaluating each measurement separately can make it difficult to determine which unit requires priority attention. The proposed approach synthesizes these results and provides a more balanced view of the overall condition. However, the final score should not replace individual diagnostic methods, as detailed test results are still necessary to identify the primary failure and select the appropriate maintenance action. At the fleet level, this classification can facilitate risk-based maintenance, inspection planning, budget allocation, and long-term investment planning. It can also help public utilities to identify units that are deteriorating before they reach a critical condition, thereby reducing the risk of unplanned outages. When considered in addition with the transformer's load, the consequences of a failure, and its maintenance history, the proposed HI provides a practical basis for making technically asset management decisions.

Table 10. Parameters values based on the complete oil analysis for the two transformers studied.

Parameters	Units	Transformer T1	Transformer T7
Power rating	MVA	47	33
Voltage	kV	70	40
Moisture content	mg/kg	10	29
Water content at 20˚C	mg/kg	2	9
Water saturation	%	3.87	17.29
Dissipation factor (tanδ) at 90˚C	-	0.0361	0.3125
Total acide nimber (TAN)	mg KOH/g	0.0099	0.113
Breakdown voltage	kV	51.8	55.6
Carbone dioxyde (CO_2)	ppm	10,172	1554
Carbon monoxyde (CO)	ppm	362.6	536.5
Hydrogen (H_2)	ppm	156.9	60.4
Methane (CH_4)	ppm	222.3	203.2
Ethane (C_2H_6)	ppm	928.3	870.8
Ethylene (C_2H_4)	ppm	435.2	128.6
Acetylene (C_2H_2)	ppm	0.099	36.9
TDCG	ppm	2105	1836
2-FAL	mg/L	0.16	0.16
DGAF		3.611	3.722
OAF		1.375	1.75
HI	%	42.39	33.70

5. Conclusions

This study developed of a practical health index designed for assessing the condition of power transformers using supplementary diagnostic information. The proposed method combines dissolved gas analysis, moisture, dielectric properties, acidity and cellulose aging indicators into a single index. This approach provides a more accurate picture of the general condition of the equipment than interpretation of individual tests.

The results showed that most transformers in the studied fleet are still in satisfactory condition, while a smaller group exhibits moderate or advanced deterioration. Analyses using PCA and the K-means method provided data-driven evidence of the statistical consistency of the proposed classification. The results revealed an overall gradient in health status and partial agreement with the HI categories, with greater consistency among the extreme groups and more overlap in the intermediate categories. Although the strongest natural division was obtained for $k = 2$, the five HI categories were retained because they provide a more detailed and operationally useful representation of transformer condition. Their boundaries should therefore be interpreted as decision thresholds rather than as five completely distinct natural populations. The two-transformer case study also demonstrated that similar HI ranges can correspond to different failure mechanisms, confirming the need to examine the underlying measurements before selecting a maintenance action.

In sum, the proposed index offers a simple and technically reliable tool for comparing transformers, identifying vulnerable units, and setting maintenance priorities. It should be used to supplement traditional diagnostic methods rather than to replace them. Future work could help strengthen this approach by incorporating load history, the consequences of failures, real-time monitoring data, and periodic score updates. Validation on larger fleets and in different operating environments would also help confirm its applicability on a larger scale within the framework of risk-based asset management.

Author Contributions

Kone Gbah: Conceptualization; methodology; validation; formal analysis; investigation; resources; data curation; writing—original draft preparation; Writing review and editing; visualization; supervision. Gbegbe Raymond: software; resources; data curation.

All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] (2019) IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers. *IEEE Std C*57.104-2019 (*Revision of IEEE Std C*57.104-2008),

1-98. https://ieeexplore.ieee.org/document/8931178

[2] Fofana, I. (2013) 50 Years in the Development of Insulating Liquids. *IEEE Electrical Insulation Magazine*, **29**, 13-25. https://doi.org/10.1109/mei.2013.6585853

[3] Zhang, E., Liu, J., Zhang, C., Zheng, P., Nakanishi, Y. and Wu, T. (2023) State-of-art Review on Chemical Indicators for Monitoring the Aging Status of Oil-Immersed Transformer Paper Insulation. *Energies*, **16**, Article 1396. https://doi.org/10.3390/en16031396

[4] Tenbohlen, S., Coenen, S., Djamali, M., Müller, A., Samimi, M.H. and Siegel, M. (2016) Diagnostic Measurements for Power Transformers. *Energies*, **9**, Article 347. https://doi.org/10.3390/en9050347

[5] Duval, M., De Pablo, A., Atanasova-Hoehlein, I. and Grisaru, M. (2017) Significance and Detection of Very Low Degree of Polymerization of Paper in Transformers. *IEEE Electrical Insulation Magazine*, **33**, 31-38. https://doi.org/10.1109/mei.2017.7804314

[6] Badawi, M., Ibrahim, S.A., Mansour, D.A., El-Faraskoury, A.A., Ward, S.A., Mahmoud, K., *et al.* (2022) Reliable Estimation for Health Index of Transformer Oil Based on Novel Combined Predictive Maintenance Techniques. *IEEE Access*, **10**, 25954-25972. https://doi.org/10.1109/access.2022.3156102

[7] Guo, H. and Guo, L. (2022) Health Index for Power Transformer Condition Assessment Based on Operation History and Test Data. *Energy Reports*, **8**, 9038-9045. https://doi.org/10.1016/j.egyr.2022.07.041

[8] Bohatyrewicz, P., Płowucha, J. and Subocz, J. (2019) Condition Assessment of Power Transformers Based on Health Index Value. *Applied Sciences*, **9**, Article 4877. https://doi.org/10.3390/app9224877

[9] Azmi, A., Jasni, J., Azis, N. and Kadir, M.A. (2017) Evolution of Transformer Health Index in the Form of Mathematical Equation. *Renewable and Sustainable Energy Reviews*, **76**, 687-700. https://doi.org/10.1016/j.rser.2017.03.094

[10] Zeinoddini-Meymand, H., Saeid, M. and Vahidi, B. (2024) Calculation of Power Transformer Health Index Using the Combination of Principal Component Analysis and Mass-Spring-Damper Model. *Electrical Engineering*, **106**, 7399-7409. https://doi.org/10.1007/s00202-024-02459-9

[11] IEC Webstore (2022) IEC 60599:2022 Mineral Oil-Filled Electrical Equipment in Service—Guidance on the Interpretation of Dissolved and Free Gases Analysis. International Electrotechnical Commission.

[12] Rogers, R.R. (1978) IEEE and IEC Codes to Interpret Incipient Faults in Transformers, Using Gas in Oil Analysis. *IEEE Transactions on Electrical Insulation*, **13**, 349-354. https://doi.org/10.1109/tei.1978.298141

[13] Duval, M. (1974) Fault Gases Formed in Oil-Filled Breathing EHV Power Transformers-Interpretation of Gas-Analysis Data. In: *IEEE Transactions on Power Apparatus and Systems*, IEEE Institute Electrical Electronics Engineers Inc, 1745-1746.

[14] Thiviyanathan, V.A., Ker, P.J., Leong, Y.S., Abdullah, F., Ismail, A. and Zaini Jamaludin, M. (2022) Power Transformer Insulation System: A Review on the Reactions, Fault Detection, Challenges and Future Prospects. *Alexandria Engineering Journal*, **61**, 7697-7713. https://doi.org/10.1016/j.aej.2022.01.026

[15] Feng, D., Chen, G., Yan, X., Fu, X. and Liao, R. (2023) Furfural Analysis on Aging State Assessment of Oil-Immersed Transformers: A Review. *IEEE Transactions on Dielectrics and Electrical Insulation*, **30**, 2758-2768. https://doi.org/10.1109/tdei.2023.3314390

[16] Ali, M.S., Awalin, L.J., Abd Halim, S., Abdul Jaafar, A.S., Abu Bakar, A.H., Smadi,

I.A., *et al*. (2025) Assessing Polymerization-Based Diagnostics for Transformer Insulation via Comparative Statistics. *Energies*, **18**, Article 5990. https://doi.org/10.3390/en18225990

[17] Ghoneim, S.S.M. (2021) The Degree of Polymerization in a Prediction Model of Insulating Paper and the Remaining Life of Power Transformers. *Energies*, **14**, Article 670. https://doi.org/10.3390/en14030670

[18] N'cho, J.S., Fofana, I., Hadjadj, Y. and Beroual, A. (2016) Review of Physicochemical-Based Diagnostic Techniques for Assessing Insulation Condition in Aged Transformers. *Energies*, **9**, Article 367. https://doi.org/10.3390/en9050367

[19] Jin, L., Kim, D., Abu-Siada, A. and Kumar, S. (2022) Oil-Immersed Power Transformer Condition Monitoring Methodologies: A Review. *Energies*, **15**, Article 3379. https://doi.org/10.3390/en15093379

[20] Sun, P., Sima, W., Yang, M. and Wu, J. (2016) Influence of Thermal Aging on the Breakdown Characteristics of Transformer Oil Impregnated Paper. *IEEE Transactions on Dielectrics and Electrical Insulation*, **23**, 3373-3381. https://doi.org/10.1109/tdei.2016.005306

[21] Calcara, L., Sangiovanni, S. and Pompili, M. (2019) Standardized Methods for the Determination of Breakdown Voltages of Liquid Dielectrics. *IEEE Transactions on Dielectrics and Electrical Insulation*, **26**, 101-106. https://doi.org/10.1109/tdei.2018.007685

[22] (2016) IEEE Guide for Acceptance and Maintenance of Insulating Mineral Oil in Electrical Equipment. *IEEE Standard C*57.106-2015 (*Revision of IEEE Standard C*57.106-2006), 1-38.

[23] IEC 60422 Mineral Insulating Oils in Electrical Equipment—Supervision and Maintenance Guidance. International Electrotechnical Commission.

[24] Wang, M., Vandermaar, A.J. and Srivastava, K.D. (2002) Review of Condition Assessment of Power Transformers in Service. *IEEE Electrical Insulation Magazine*, **18**, 12-25. https://doi.org/10.1109/mei.2002.1161455

[25] IEC 60814. (1997) Insulating Liquids-Oil-Impregnated Paper and Pressboard-Determination of Water by Automatic Coulometric Karl Fischer Titration. IEC 60814.

[26] Zhou, J., Wu, Z., Guo, Y., Zhang, R., Zhu, W. and Zhang, Q. (2024) Model Moisture Transport in Oil-Paper Insulation of Transformer: Theory and Experiment. *High Voltage*, **9**, 879-887. https://doi.org/10.1049/hve2.12414

[27] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning. Springer.

[28] Syafruddin, H. and Nugroho, H.P. (2020) Dissolved Gas Analysis (DGA) for Diagnosis of Fault in Oil-Immersed Power Transformers: A Case Study. 2020 4*th International Conference on Electrical*, *Telecommunication and Computer Engineering* (*ELTICOM*), Medan, 3-4 September 2020, 57-62. https://doi.org/10.1109/elticom50775.2020.9230491

[29] El Amine Senoussaoui, M., Fofana, I. and Brahami, M. (2021) Influence of Oil Quality on the Interpretation of Dissolved Gas Analysis Data. 2021 *IEEE* 5*th International Conference on Condition Assessment Techniques in Electrical Systems* (*CATCON*), Kozhikode, 3-5 December 2021, 170-175. https://doi.org/10.1109/catcon52335.2021.9670513

[30] Hernanda, I.G.N.S., Mulyana, A.C., Asfani, D.A., Negara, I.M.Y. and Fahmi, D. (2014) Application of Health Index Method for Transformer Condition Assessment. *TENCON* 2014—2014 *IEEE Region* 10 *Conference*, Bangkok, 22-25 October 2014, 1-

6. https://doi.org/10.1109/tencon.2014.7022433

[31] Jahromi, A., Piercy, R., Cress, S., Service, J. and Fan, W. (2009) An Approach to Power Transformer Asset Management Using Health Index. *IEEE Electrical Insulation Magazine*, **25**, 20-34. https://doi.org/10.1109/mei.2009.4802595

[32] Li, S., Li, X., Cui, Y. and Li, H. (2023) Review of Transformer Health Index from the Perspective of Survivability and Condition Assessment. *Electronics*, **12**, Article 2407. https://doi.org/10.3390/electronics12112407

[33] Zukowski, P., Rogalski, P., Kierczynski, K. and Koltunowicz, T.N. (2021) Precise Measurements of the Temperature Influence on the Complex Permittivity of Power Transformers Moistened Paper-Oil Insulation. *Energies*, **14**, Article 5802. https://doi.org/10.3390/en14185802

[34] Yang, L., Lin, Y., Liao, R., Zhao, X., Sun, W. and Zhang, Y. (2016) Effects of Temperature and Aging on Furfural Partitioning in the Oil-Paper System of Power Transformers. *IEEE Transactions on Dielectrics and Electrical Insulation*, **23**, 1393-1401. https://doi.org/10.1109/tdei.2015.005530