International Journal of Applied Science and Engineering

Improving the accuracy of an ANN model for transformer condition assessment using the SMOTE-R2E method

Kristoko Dwi Hartomo [1], Yessica Nataliani [1*], Denny Indrajaya [2], Nur Haliza Abdul Wahab [3], Untung Rahardja [4], Christine Dewi [5]

[1] *Department of Information System, Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia*
[2] *Master of Data Science, Faculty of Science and Mathematics, Satya Wacana Christian University, Salatiga, Indonesia*
[3] *Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia*
[4] *Faculty of Economics and Business, University of Raharja, Tangerang, Indonesia*
[5] *Department of Informatics, Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia*

ABSTRACT

A transformer is a device used for electricity-related purposes, one of which is to distribute electricity from a supplier, in this case the State Electricity Company (PLN), to customers or the community. Considering the transformer's essential role, it is crucial to conduct research to minimize damage to this device, which can be caused by a variety of factors, where gas and electrical conditions of the transformer, among other things, can be used as indicators of damage. Therefore, this study focused on creating models using Artificial Neural Network (ANN) algorithms to assess transformer conditions based on data obtained from transformers. In this study, correlation analysis was used to determine six features that served as leading indicators in evaluating the condition of the transformer. The six features were dibenzyl disulfide (DBDS), interfacial voltage, hydrogen, methane, ethylene, and water content. In modelling and testing, 80% of the data was distributed for the training dataset and 20% was for the testing dataset, with a total of 470 data points. This study also applied the Synthetic Minority Oversampling Technique-Rechecked, Reused, and Edited (SMOTE-R2E) method, which is an improvement of the SMOTE method. SMOTE-R2E is proposed in this study to overcome the limitations of unbalanced transformer data. In this study, model training was carried out using three approaches, i.e., model training using a training dataset obtained without the SMOTE method, model training using a training dataset obtained with the SMOTE method, and model training using a training dataset obtained with the SMOTE-R2E method. Each training approach was performed 100 times. Based on model testing on the testing dataset, the best model was the model obtained by applying the SMOTE-R2E method, with average accuracy and F1-score obtained from 100 iterations of 83.04% and 81.96%, respectively.

Keywords: Artificial Neural Network, SMOTE, SMOTE-R2E, Transformer

OPEN ACCESS

Received: April 06, 2024
Revised: May 14, 2024
Accepted: June 14, 2024

Corresponding Author:
Yessica Nataliani
yessica.nataliani@uksw.edu

Publisher:
Chaoyang University of Technology
ISSN: 1727-2394 (Print)
ISSN: 1727-7841 (Online)

1. INTRODUCTION

Machine learning can be applied in many fields, such as image detection, fraud detection, recommendation systems, etc. For image detection, especially in the healthcare field, some methods have been proposed, including wheat disease recognition by combining the VGG-16 CNN model and WeChat (Wang et al., 2023), acute lymphoblastic leukemia disease identification by using a CNN model (Saeed et al., 2024),

atrial fibrillation detection using the combination of one-dimensional dense residual network and bidirectional RNN (Laghari et al., 2023), and patient medical image collection and interpretation (Laghari et al., 2024).

Nowadays, machine learning is also combined with IoT. For example, this combination has been utilized to develop smart systems to help the elderly and the disabled with their daily activities (Das et al., 2023), detect anomalies using deep auto-encoder and capsule graph convolution via the sparrow search algorithm in 6G IoT (Yin et al., 2024), and track the accuracy of the train to the target speed in automatic train operations (Liu et al., 2023). Machine learning can also be applied in online video streaming to optimize the streaming quality and increase the revenue for service providers (Laghari et al., 2023). In addition, machine learning can be integrated with augmented reality to develop games. Integrating machine learning and augmented reality will expand the range of AR experiences and add a more personal touch (Laghari et al., 2024).

Machine learning can be used to detect transformer condition. A transformer is a device with magnetic coupling, which channels electric energy from one electric circuit to another, based on the principle of electromagnetic induction while maintaining the frequency (Syahfitra et al., 2017). This device is one of the crucial components in the distribution of electricity from the State Electricity Company (PLN) to the community, with electric power distribution structures located at several points, including power substations (El-Harbawi, 2022). Considering the importance of the transformer and its continuous use, it is necessary to pay more attention to the condition of the transformer, as damage to this device results in losses for both the customers and PLN. Previous research has shown that various things can damage the transformer, for example, thermal disturbances caused by transformer overload, poor bolt connections with cables, poor oil flow in the transformer, water content in the oil, and sludge in oil (Hoole et al., 2017). In addition, several gases are produced on the transformer, some of which are hydrocarbon gases such as hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2) (Nanfak et al., 2021). Seeing the impact and danger of these flammable gases, in addition to electrical data on the transformer, a device is needed to provide information about the gases on the transformer in real-time to determine the condition of the transformer. One of the parameters used to see the condition of the transformer is the Health Index (HI), which is the value deriving from analyzing various data based on test data in the field or the laboratory, considering the time of use of the tool (Zhengwei et al., 2018).

Previous studies have predicted HI values, such as a study Abdullah et al. (2021) that performed HI value prediction using Condition Situation Monitoring (CSM) diagnostic techniques. One study Alqudsi et al. (2019) also classified transformer conditions into three classes (based on HI values), with ten and four features, using eight classification methods: Random Forest (RF), Decision Tree (J48), Support Vector Machines (SVMs), k-Nearest Neighbor (k-NN), OneR, Multinomial Logistic Regression (MLR), Naïve Bayes (NB), and Artificial Neural Networks (ANNs). However, some others classified HI into four (Bohatyrewicz and Banaszak, 2022) and five categories (Hernanda et al., 2014).

Keeping in mind the losses caused when a transformer is damaged and the importance of checking the condition of the transformer based on HI in real-time, data to be used as indicators of transformer condition classification is needed. However, such data is not easily available. Because the price of the transformer is very high, the company will constantly keep the transformer in good condition, and therefore, data indicating the transformer's good condition that is necessary for developing the classification model will be much less than data indicating its poor condition. In addition, it will take a long time to obtain a large amount of data on the transformer's poor conditions. Therefore, a method for developing a transformer condition classification model with existing unbalanced transformer data is proposed.

This research used existing data, where the data obtained for each category in this study was unbalanced, with the amount of data on good transformer conditions being much less than the amount of data on poor transformer conditions. Therefore, to create a classification model, it was deemed necessary to first generate synthetic data. In previous studies, no synthetic data were generated to overcome the problem of unbalanced transformer condition data. Therefore, this study explored the classification of transformer conditions, which were grouped into four categories, with six features using the ANN method, where a model was trained by applying the Synthetic Minority Oversampling Technique (SMOTE) method. SMOTE has been proven to improve classification accuracy on unbalanced data (Pribadi et al., 2022, Lopo and Hartomo, 2023). Another model was trained by applying the Synthetic Minority Oversampling Technique-Rechecked, Reused, and Edited (SMOTE-R2E) method, an improvement of the SMOTE method proposed in this study. We also trained another model with the existing data without generating synthetic data to observe the differences.

2. MATERIALS AND METHODS

The data used in this study was not privately derived from the transformer site. Instead, it was public data with limited detailed information. Therefore, a literature study regarding data use was conducted. However, we found no research that used data for classification, including classification with the SMOTE method. Fig. 1 outlines the research method employed in this study to achieve the following research objectives:

- To demonstrate a new synthetic data generation method named SMOTE-R2E to address unbalanced data in ANN model creation.
- To demonstrate the use of the SMOTE method to address unbalanced data in ANN model creation.

- To compare the ANN classification model generated with SMOTE, the ANN model generated with SMOTE-R2E, and the ANN model generated without SMOTE.

The details of each research step are provided in the following sections.

2.1 Transformer

A transformer is an electrical device that uses the principle of electromagnetic induction to transfer energy from one electrical circuit to another (Abdullah et al., 2021). There are several components in the transformer, including iron core, transformer coils, transformer oil, insulation, and conservator tank. Transformer oil in particular is vital in the transformer as a cooling medium (Hoole et al., 2017). An electric field or thermal loads originating from the windings or transformer core can cause various dissolved gases to be present in transformer oil at a certain level (Bustamante et al., 2019). Therefore, these gases can be used as an indicator to assess the condition of the transformer.

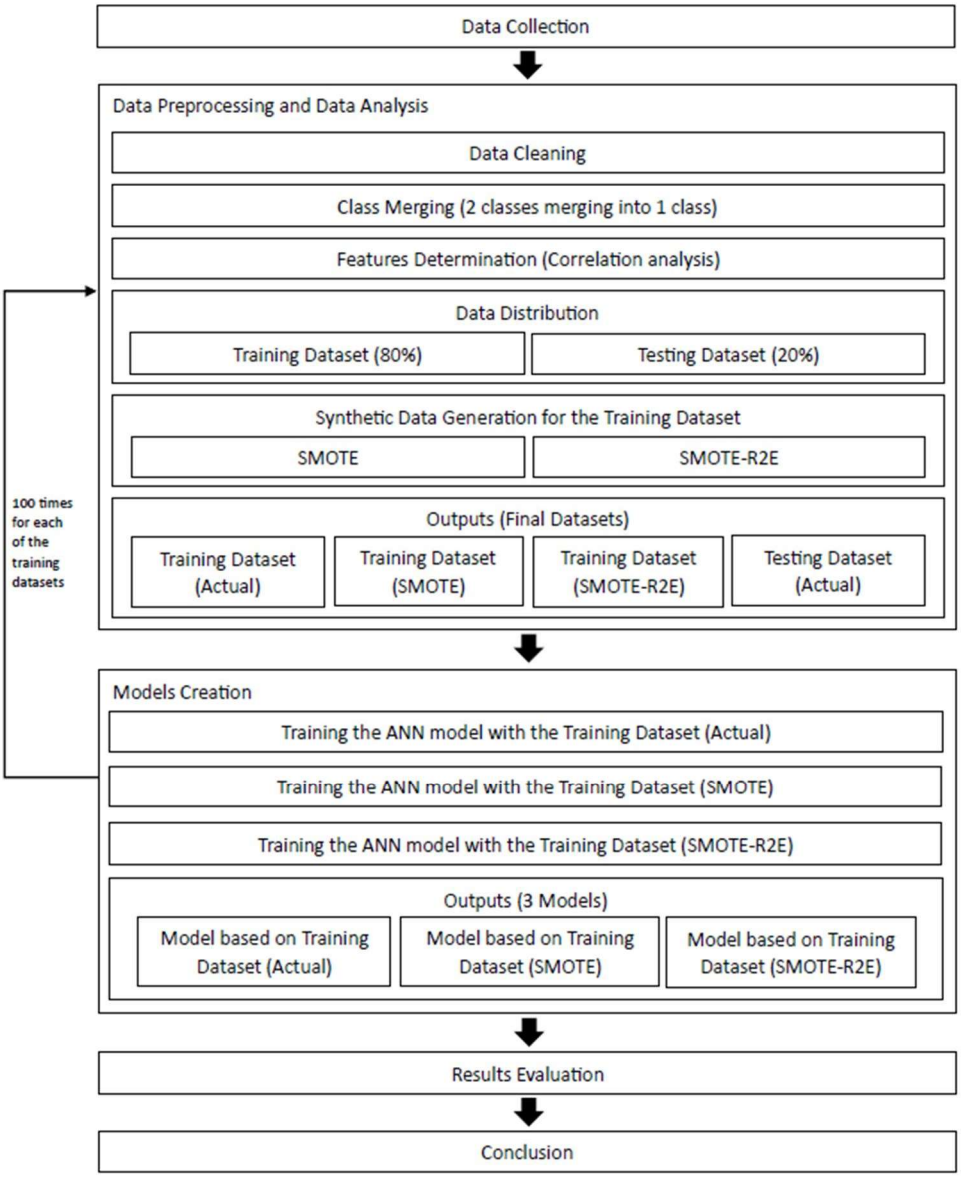

Fig. 1. The outline of the research method

2.2 Artificial Neural Networks (ANNs)

An ANN is an algorithm that works like a human neural network (Stangierski et al., 2019). In an ANN, several parameters need to be found through a training process, namely weight and bias. The ANN algorithm is generally described in Fig. 2 and mathematically represented as Equation (1) (Wang et al., 2020).

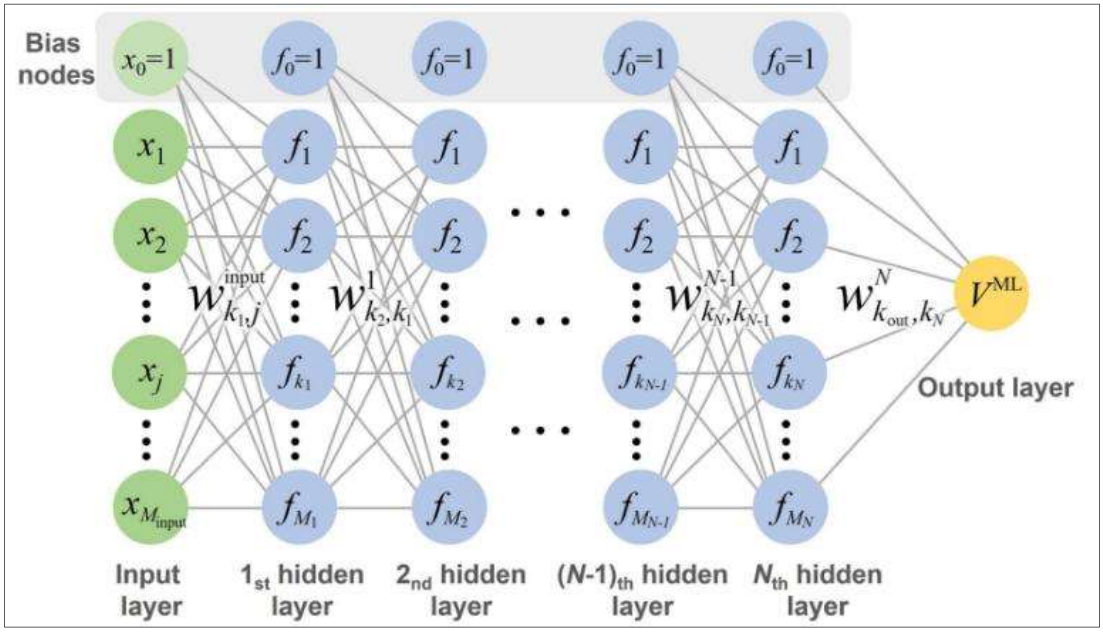

Fig. 2. ANN scheme (Wang et al., 2020)

$$V^{ML}(X) = b_N + \sum_{k_N=0}^{M_N} w_{k_{out},k_N}^{N} \cdot f\left(b_{N-1} + \sum_{k_{N-1}=0}^{M_{N-1}} w_{k_N,k_{N-1}}^{N-1}\left(\cdots f\left(b_1 + \sum_{k_1=0}^{M_1} w_{k_2,k_1}^{1} \cdot f\left(b_0 + \sum_{j=0}^{M_{input}} w_{k_1,j}^{input} \cdot x_j\right)\right)\right)\right) \quad (1)$$

where V^{ML} represents the prediction results with N hidden layers, M_n represents the number of nodes on each hidden layer with $n \in \{1, 2, .. N\}$, M_{input} represents the number of nodes on the input layer, and $w_{i,j}^{n}$ is the weight parameter connecting two layers with $j \in \{1, 2, .., M\}$. By looking at the number of nodes and biases present in the architecture, the total parameters can be calculated with Equation (2).

$$\text{The sum of all parameters} = \sum_{n=0}^{N-1} (M_n + 1) \times M_{n+1} \quad (2)$$

where the value of 1 expresses the sum of the biases, $M_0 = M_{input}$, and $M_N = M_{output}$.

The ANN model architecture also uses activation functions to determine the number of hidden layers and the number of nodes on every hidden layer. The activation functions used in this study were ReLU and Softmax. The ReLU function is represented as Equation (3), and Softmax function is represented as Equation (4) (Sharma et al., 2020).

$$f(x) = \max(0, x) \quad (3)$$

$$f_j(x) = \frac{e^{x_j}}{\sum_{i=1}^{K} e^{x_i}} \quad (4)$$

with $j = 1, \ldots, K$, where K is the number of classes used.

2.3 Data Collection

This study used transformer and HI data from the website https://www.kaggle.com with the format *.csv (Velásquez and Lara, 2020). Based on the information on the page, the HI can be divided into five conditions. Samples of the data obtained from the site are shown in the Tables 1A and 1B (Alqudsi and El-Hag, 2019).

Table 1A. Transformer data samples

Index data	Hydrogen	Oxygen	Nitrogen	Methane	CO	CO_2	Ethylene	Ethane
1	14	514	70600	11	674	11700	13	5
2	13	12300	43400	0	309	908	0	0
3	38	3127	16464	11	177	727	28	3
4	36	2950	63300	101	525	4080	193	35
5	488	11861	48353	13	85	1957	29	23
6	4	26100	60600	1	206	1440	13	0
7	13500	343	36500	3150	113	984	5	1230
8	16	2470	59600	8	520	2660	5	8
9	2845	5860	27842	7406	32	1344	16684	5467
10	3210	3570	47900	160	360	2130	4	43

Table 1B. Transformer data samples (cont.)

Index Data	Acetylene	Dibenzyl disulfide	Power factor	Interfacial voltage	Dielectric rigidity	Water content	Health index	Condition
1	1	0.0	0.62	42	45	21	28.1	0
2	0	0.0	1.00	42	47	26	26.7	0
3	19	0.0	1.00	47	75	38	49.9	1
4	0	127.0	0.37	40	53	6	49.2	1
5	0	164.0	0.27	37	72	10	68.0	2
6	13	5.0	1.32	40	56	4	63.4	2
7	1	1.0	4.93	37	52	6	75.6	3
8	2	164.0	0.29	44	56	4	72.8	3
9	7	19.0	1.00	45	55	0	95.2	4
10	4	1.0	0.77	44	55	3	85.2	4

2.4 Data Cleaning and Class Merging

Data cleaning was carried out first in data processing, where missing data samples were deleted. After cleaning the data, we analyzed the amount of data available for each class. As described in the Introduction section, many previous studies have applied various classification methods to transformer data. One study used the ANN method to classify three classes using ten and four features.

However, to control transformer conditions routinely, the use of more classes would be more favorable, as shown in the study by Hernanda et al. (2014) and Table 2A. Since the amount of transformer data for the "good" and "very good" conditions was small, this study combined both conditions into one condition (see Table 2B).

Table 2A. Transformer health index categorization

Health index	Condition
85–100	Very good
70–85	Good
50–70	Fair
30–50	Poor
0–30	Very poor

Table 2B. New transformer health index categorization

Health index	Condition
70–100	Good
50–70	Fair
30–50	Poor
0–30	Very poor

2.4 Features Determination

The selection of features was performed based on correlation analysis according to Indrajaya et al. (2022). The study showed that selecting features using correlation analysis could provide high accuracy with only a few features, applying the k-Nearest Neighbors and Naïve Bayes methods for classification. The correlation coefficient r from correlation analysis expresses the relationship between two variables, ranging from -1 up to 1, where the correlation coefficient $r = 0$ means there is no relationship, and $r = \pm 1$ shows a perfect relationship (Schober and Schwarte, 2018). Equation (5) gives Pearson's correlation coefficient formula.

$$r_{xy} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \tag{5}$$

Using Pearson's correlation formula, the correlation coefficient between transformer and HI data was obtained (see Table 3). The purpose of creating this classification model was to monitor the condition of the transformer in real-time. Therefore, computational speed was deemed important. One way to enhance computational speed is to utilize relatively few features. As a result, only six features were used in this study because the correlation values obtained from the correlation analysis were relatively low. The six features were Dibenzyl Disulfide (DBDS), Interfacial Voltage, Hydrogen, Methane, Ethylene, and Water Content.

Table 3. Correlation coefficients of sensor data and transformer health index

Data name	Correlation coefficient (r)
DBDS	0.47
Interfacial V	0.40
Hydrogen	0.38
Methane	0.36
Ethylene	0.27
Acetylene	0.24
Ethane	0.24
CO_2	0.17
Oxygen	0.12
CO	0.11
Power factor	0.09
Nitrogen	0.09
Dielectric rigidity	- 0.10
Water content	- 0.28

2.5 Data Normalization

In addition to data cleaning and selection, data normalization was also performed at this stage. Various formulas can be used in data normalization, one of which is provided as Equation (6), which has been applied in many previous studies on machine learning (Jamal et al., 2014).

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (6)$$

2.6 Data Distribution

The data used in this study was distributed as 80% for the training dataset (376 data points) and 20% for the testing dataset (94 data points). By distributing data into training and testing datasets, each class had at least one sample. The amount of data for each class in the training and testing datasets is shown in Table 4.

Table 4. Dataset distribution summary

Class	Class code	Training dataset (80%)	Testing dataset (20%)
Very poor	0	8	1
Poor	1	32	10
Fair	2	104	30
Good	3	232	53

2.7 Synthetic Data Generation

Synthetic data generation could improve the goodness of fit of the model by modelling the classification of unbalanced transformer conditions for each class. One method to create synthetic data is called SMOTE, which works by applying k-nearest neighbors. Here are the steps for implementing SMOTE.

1. For each data point X_0 of the minority class, select one of its closest neighbors X (which also belongs to the minority class);
2. Create a new data point Z at a random point on a line segment connecting the selected pattern and the neighbor with the formula shown in Equation (7).

$$Z = X_0 + w(X - X_0) \quad (7)$$

where w is a uniformly distributed random value having a limit $0 \leq w \leq 1$.

Applying SMOTE will generate new data points, each being between two data points, as shown by Equation (7), where the two data points are X_0 and $X_1 \in X$.

The synthetic data obtained will have a poor distribution in this process, especially if the data with minor classes used is too small. In addition, SMOTE also makes it possible to generate data with incorrect classes. Therefore, a new synthetic data generation approach was developed by applying an improvement of SMOTE. The new method is called Synthetic Minority Oversampling Technique-Rechecked, Reused, and Edited (SMOTE-R2E). The following are steps to apply SMOTE-R2E, and the idea behind SMOTE-R2E is outlined in Fig. 3.

1. Prepare the data that needs up-sampling by using SMOTE-R2E;
2. Define the targeted amount of data to be created for each class;
3. For each data point X_0 of the minority class, whose amount of data does not reach the targeted amount of data, randomly select one neighbor that also belongs to the minority class;
4. Create a new data point Z at a random point on a line segment connecting the selected data point and its neighbor with the formula shown in Equation (7);
5. Check whether the new data point Z is from among k neighbors in the same class as the new data point Z. If appropriate, the data point is used; otherwise, it is not used;
6. Repeat steps 1 to 5 until the amount of synthetic data meets the target or all data points from the synthetic data generation are used;
7. If all data has been used and the amount of synthetic data has not met the targeted amount of data, combine the obtained synthetic data with the previous data used for the generation of further synthetic data;
8. Repeat steps 1 to 7 until the amount of data for each class reaches the targeted amount of data;
9. Output data meeting the targeted amount is obtained.

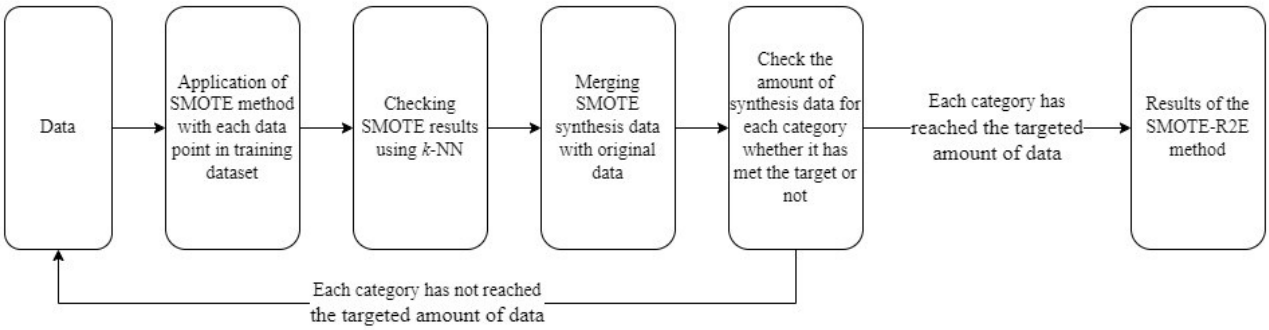

Fig. 3. The idea behind the SMOTE-R2E method

The difference in synthetic data obtained with the SMOTE and SMOTE-R2E methods can be seen more clearly in Fig. 4 and 5. The transformer data samples without the addition of synthetic data are shown in Fig. 3. These graphs show that the distribution of the synthetic data generated with SMOTE-R2E is more spread out than the distribution of the synthetic data generated with SMOTE. This is because SMOTE only creates new data points from the actual data, which means that if there are only three data points in the actual data, then the location of the synthetic data generated will be very limited to the three-line segments generated from those three data points.

Meanwhile, with SMOTE-R2E, the resulting synthetic data can be used to create new line segments, which means that the distribution of the new synthetic data generated is more spread out.

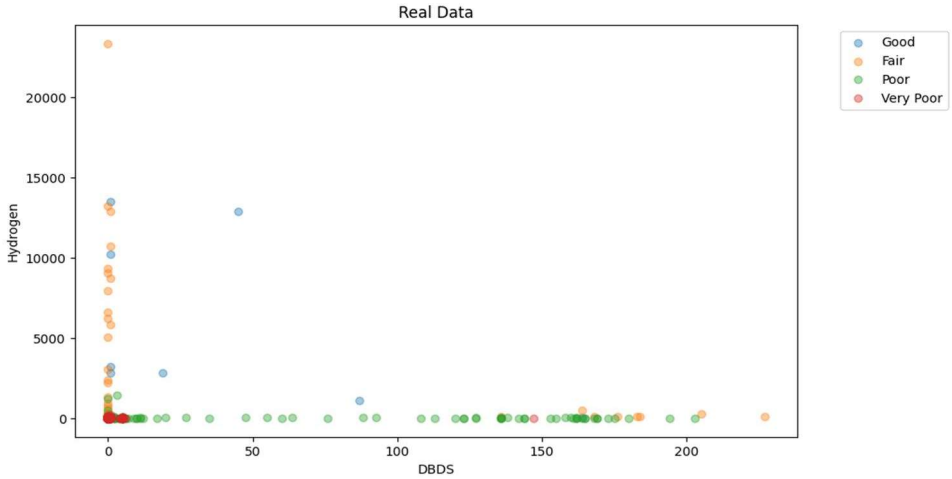

Fig. 4. Sample graph of the original hydrogen and DBDS data (features) on the transformer and its classes

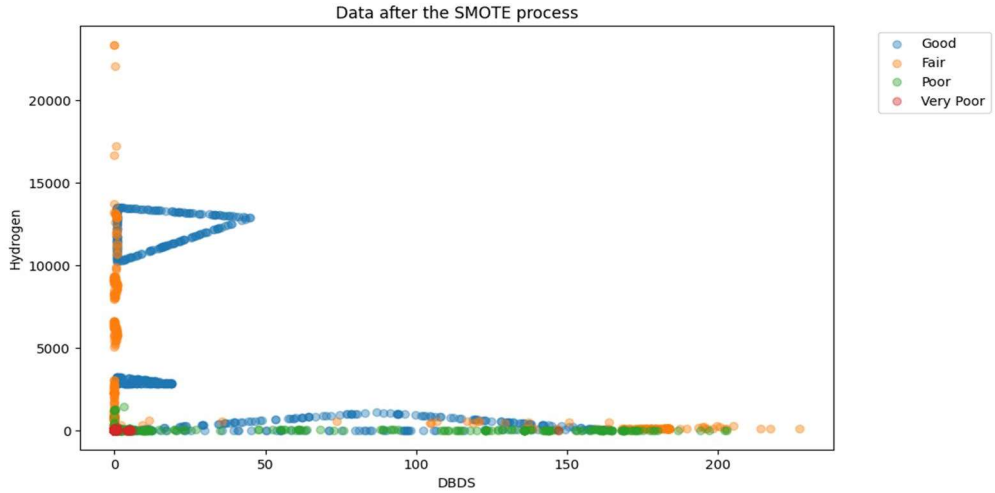

Fig. 5. Sample graph of hydrogen and DBDS data (features) from SMOTE results on the transformer and its classes

Fig. 6. Sample graph of hydrogen and DBDS data (features) from SMOTE-R2E results on the transformer and its classes

2.8 Model Creation

This study used the ANN method for classification with an architecture as shown in Table 5. There was a total of 1,236 parameters used. According to Table 5, the input layer had six nodes, indicating the number of features used, namely Dibenzyl Disulfide (DBDS), Interfacial Voltage, Hydrogen, Methane, Ethylene, and Water Content. The first, second, and third hidden layers contained 16, 32, and 16 nodes, respectively. Finally, the output layer contained four nodes, which indicate the final results of the classification (probability for each class). The "Activation Function" column contains information on the activation functions used to obtain computational results on the layers where these functions were respectively applied as shown in the Table 5.

Table 5. ANN architecture for classification

Layer type	Number of nodes	Activation function	Number of parameters
Input layer	6	-	-
1st hidden layer	16	ReLU	112
2nd hidden layer	32	ReLU	544
3rd hidden layer	16	ReLU	528
Output layer	4	Softmax	68

The training parameters used in this study for the ANN are shown in Table 6. The weight and bias parameter values were to be determined during the training process, but initialization of both parameters was required before the training process started. A better initialization value close to the expected solution would make the optimal model faster.

There are several methods to determine the initial values. In this study, the He uniform variance scaling initializer was used (Bingham and Miikkulainen, 2021). The distribution of the initial value for each weight ($w_{i,j}^{n}$) obtained by the method is formulated as Equation (8).

Table 6. Training parameter values

Parameter	ANN for transformer condition classification
Initiation value	He Uniform
Number of epochs	300
Batch size	256
Loss function	Categorical cross-entropy
Optimizer	Adam

$$\left(w_{i,j}^{n}\right) = [-b, b] \tag{8}$$

where $b = \sqrt{\frac{6}{number\ of\ input\ nodes}}$. The number of input nodes is the number of n-layer nodes used as input to obtain the value on the $(n+1)^{th}$ layer, as presented in Fig. 2. The initial value for all bias parameters used in this study was 0.

In the training process, the Categorical Cross-entropy loss function calculated the loss value of the parameter for each update, as shown in Equation (9). Categorical Cross-entropy was used in the ANN algorithm for classification, which resulted in more than two classes. The activation function used was Softmax (Ho and Wookey, 2020).

$$L_{CCE} = -\frac{1}{n}\sum_{i=1}^{n}(y_i \ln(p_i) + (1 - y_i)\ln(1 - p_i)) \tag{9}$$

where n is the number of samples in the training dataset, y_i is the actual data with a value of 0 or 1, and p_i is the probability of the Softmax activation function.

In this study, Adam was used as an optimizer in the ANN method. This optimizer is an evolution of stochastic gradient descent (Jais et al., 2019). In this optimizer, the parameter values broadly affect all weight updates. The parameters used in this optimizer are learning rate, β_1, β_2, and ε. The values for each of these parameters in this study were as follows: learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$.

2.9 Analysis and Evaluation

The obtained results were analyzed to see the correlation between the feature selection and the evaluation of the obtained models. The values that became a reference in the evaluation of the models were the accuracy and F1-score for the transformer condition resulting from the classification models. The evaluation of the classification models was based on the numbers of correctly and incorrectly classified objects in the testing and training datasets. The numbers were tabulated in a matrix called Confusion Matrix (Gorunescu, 2011). A Table 7 is an example of the confusion matrix that contains True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) values. These values are used to obtain the accuracy value that indicates the performance of the classifier method. Equation (10) is the formula for obtaining accuracy, and Equation (11) is the formula for obtaining F1-score.

Table 7. Example of a confusion matrix

Actual	Predictions	
	Category 1	Category 2
Category 1	TP	FN
Category 2	FP	TN

$$Accuracy = \frac{TP + TN}{n} \tag{10}$$

$$F1 - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{11}$$

3. RESULTS AND DISCUSSION

This section will elaborate on the models created in this study using different approaches. The models were designed to classify transformer conditions into four classes using

ANNs. One model was trained using the dataset generated by SMOTE, one model was trained using the dataset generated by SMOTE-R2E, and one model was trained using the initial dataset. The model architecture designed was the same across the three models. Each model was trained 100 times with random data distribution on the training and testing datasets. This section also presents the results of the analysis of the models' classification abilities.

3.1 Classification Results Using ANN Method Only

In the model that used the initial dataset without SMOTE data generation, 80% of the total data was used as a training dataset (see Table 4). The training process was performed using only the existing transformer data without creating synthesis data based on the training dataset. In light of the varying accuracy values and F1-scores yielded by the ANN method in each training process, this study diluted the training process 100 times with the architecture and hyperparameters shown in Tables 5 and 6. Histograms of the accuracy values and F1-scores obtained from testing the training and testing datasets are shown in Fig. 7.

Histograms of accuracy values and F1-scores from the testing on the training dataset, with a tendency towards left skewness, are depicted in Fig. 7(A). These histograms show that models generated with various data distributions tend to yield high results within the distribution of the accuracy values and F1-scores obtained. Compared to the histograms depicted in Fig. 7(A), the histograms depicted in Fig. 7(B), which represent the results of the testing on the testing dataset, appear to be more symmetric, indicating that many models provide average accuracy values and F1-scores within the distribution of the accuracy values and F1-scores obtained. It should also be noted that the ranges between minimum and maximum values shown in the histograms in Fig. 7(B) are still quite large and that the majority of accuracy values and F1-scores are relatively low in the distribution of the accuracy values and F1-scores obtained.

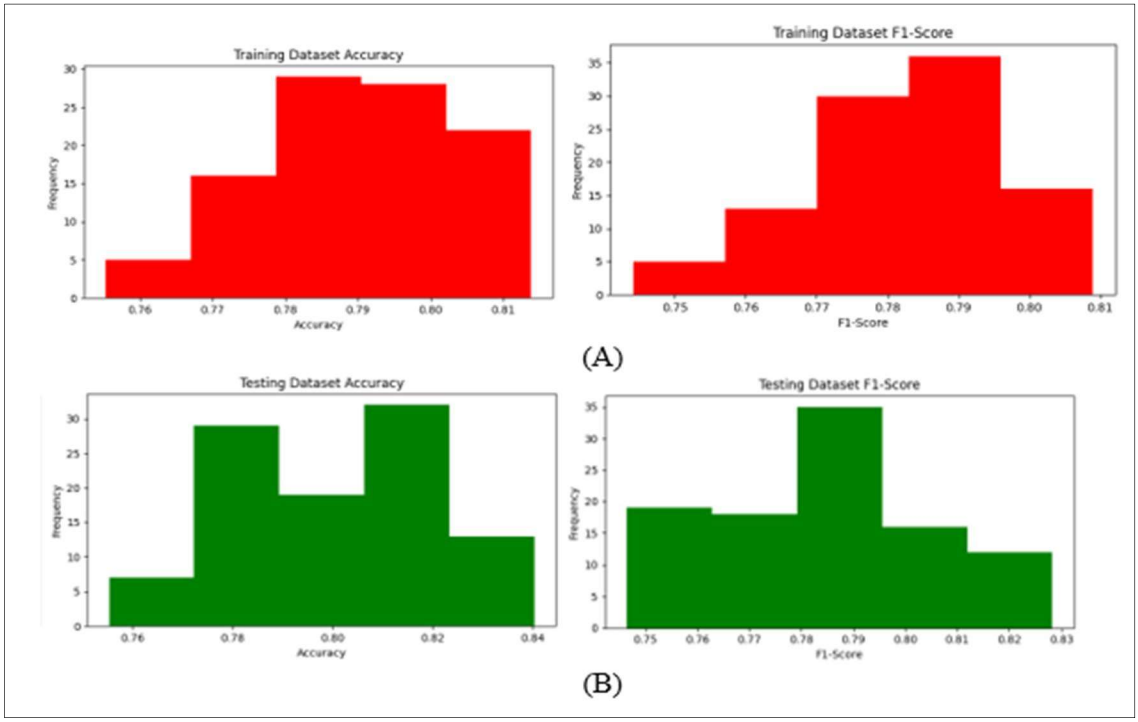

Fig. 7. Accuracy and F1-score graphs from 100 iterations of training on original data only; (A) Testing on the training dataset; (B) Testing on the testing dataset

Based on the results of the 100 iterations of training shown in Fig. 7, the mean, minimum, and maximum testing values obtained from the training and testing datasets were obtained, as summarized in Table 8. The average accuracy obtained from the 100 iterations of testing using the testing dataset was 80.03%, while the average F1-score was 78.46%. The highest accuracy and F1-score were 84.04% and 82.83%, respectively. However, it is worth noting that the values from testing the model on the training dataset were relatively lower than those obtained from testing the model on the testing dataset. The data used, which was randomly distributed, was assumed to have a role in this result.

Table 8. Confusion matrix for the testing dataset

Parameter	Training dataset		Testing dataset	
	Accuracy	F1-Score	Accuracy	F1-Score
Mean	79.01%	78.20%	80.03%	78.46%
Min	75.53%	74.42%	75.53%	74.65%
Max	81.38%	80.89%	84.04%	82.83%

Table 9. Example classification results for the testing dataset

Actual	Predictions			
	Very poor	Poor	Fair	Good
Very poor	56	3	0	0
Poor	11	17	0	0
Fair	1	1	3	0
Good	0	0	2	0

The results of data classification on the testing dataset in one iteration are shown in Table 9. The classified conditions ranged in level from "very poor" to "good". The results were quite good, since the majority of errors were found between two adjacent levels. For example, a "very poor" condition was mistakenly predicted as "poor".

3.2 Classification Results Using the ANN and SMOTE methods

ANN modelling in this section was combined with the SMOTE method, which generated synthetic data until the data for each class in the training dataset reached 500 data points. Using the architecture and hyperparameters shown in Tables 5 and 6, histograms of accuracy values and F1-scores from the testing on the initial training dataset, testing dataset, and training dataset obtained from the SMOTE process were created.

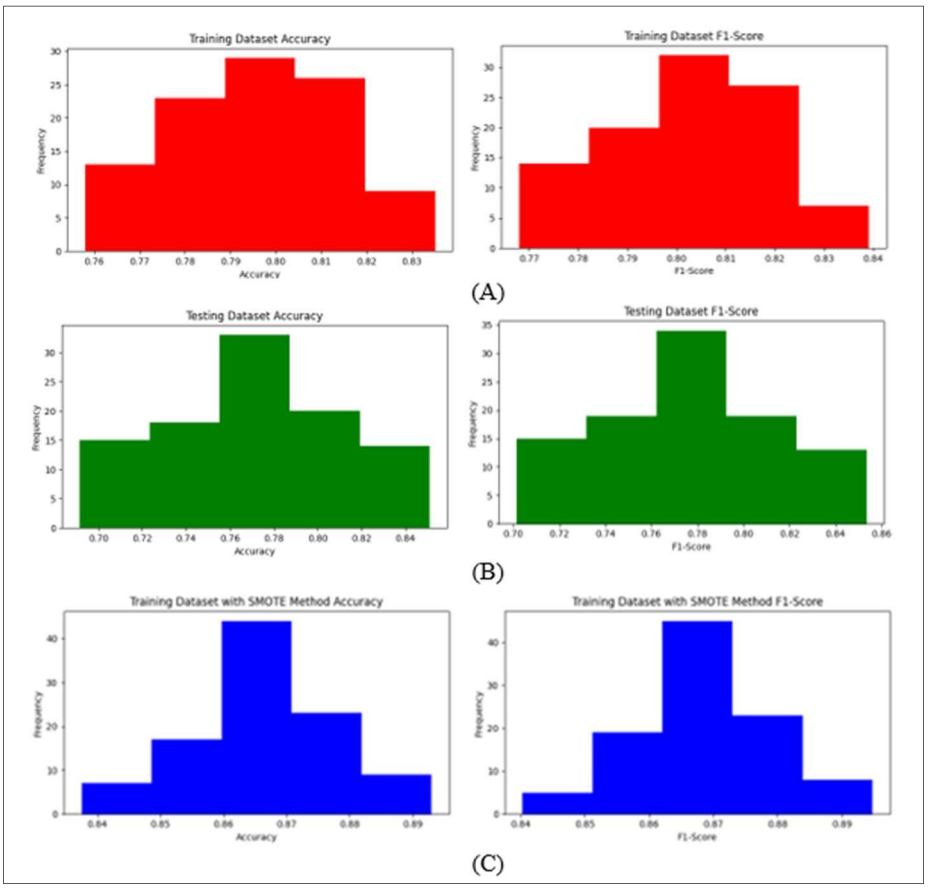

Fig. 8. Accuracy and F1-score graphs from 100 iterations of training. (A) testing on the initial training dataset; (B) testing on the testing dataset; (C) testing on the training dataset from SMOTE results

Based on the histograms of the results of the tests on various datasets shown in Fig. 8, the accuracy values and F1-scores obtained from various data distributions appear to be symmetric. This indicates that the data used to build the model was sufficiently good, in that the generated model could provide average accuracy values and F1-scores within the distribution of the accuracy values and F1-scores obtained. However, it should be noted that as shown in the histograms of results of the testing on the training and testing datasets in Fig. 8(A) and Fig. 8(B), the ranges between the minimum and maximum values obtained were large. Therefore, although the histograms have a symmetric look, the accuracy was relatively low. This was also the case with the histogram shown in Fig. 8(C).

A summary of the mean, minimum, and maximum values from the testing on the training and testing datasets with 100 iterations is provided in Table 10. The average accuracy and F1-score obtained from testing on the training dataset generated with the SMOTE method were higher than the average accuracy and F1-score values obtained from testing on the initial training dataset and the testing dataset. This was also the case with the maximum values. Notably, the testing on the testing dataset yielded a lowered minimum value of 69.15%.

Table 10. Confusion matrix for the testing dataset

Parameter	Training dataset		Testing dataset		SMOTE dataset	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
Mean	79.64%	80.34%	76.57%	77.48%	86.65%	86.88%
Min	75.80%	76.80%	69.15%	70.14%	83.75%	84.03%
Max	83.51%	83.92%	85.11%	85.36%	89.30%	89.47%

Table 11. Confusion matrix for the testing dataset

Actual	Predictions			
	Very poor	Poor	Fair	Good
Very Poor	50	9	0	0
Poor	3	22	0	3
Fair	0	1	4	0
Good	0	0	1	1

Although from 100 iterations of training the accuracy values and F1-scores obtained using the testing dataset were low, the misclassification performed by the model was not too severe (still within one-level difference). For example, one instance might be classified as "poor", while it actually was "very poor". However, a severe error (beyond a 1-level difference) was still found, as in the misclassification of "poor" conditions as "good" (shown in Table 11).

3.3 Classification Results Using the ANN and SMOTE-R2E Methods

In making models with the combination of the ANN and SMOTE-R2E methods, the same architecture and hyperparameters as those in the previous two models were used. The target number of data points generated for training was also the same, i.e., 500 data points for each class. From the testing of the model on the initial training dataset, testing dataset, and training dataset from the SMOTE-R2E process, accuracy and F1-score histograms were created, as shown in Fig. 9.

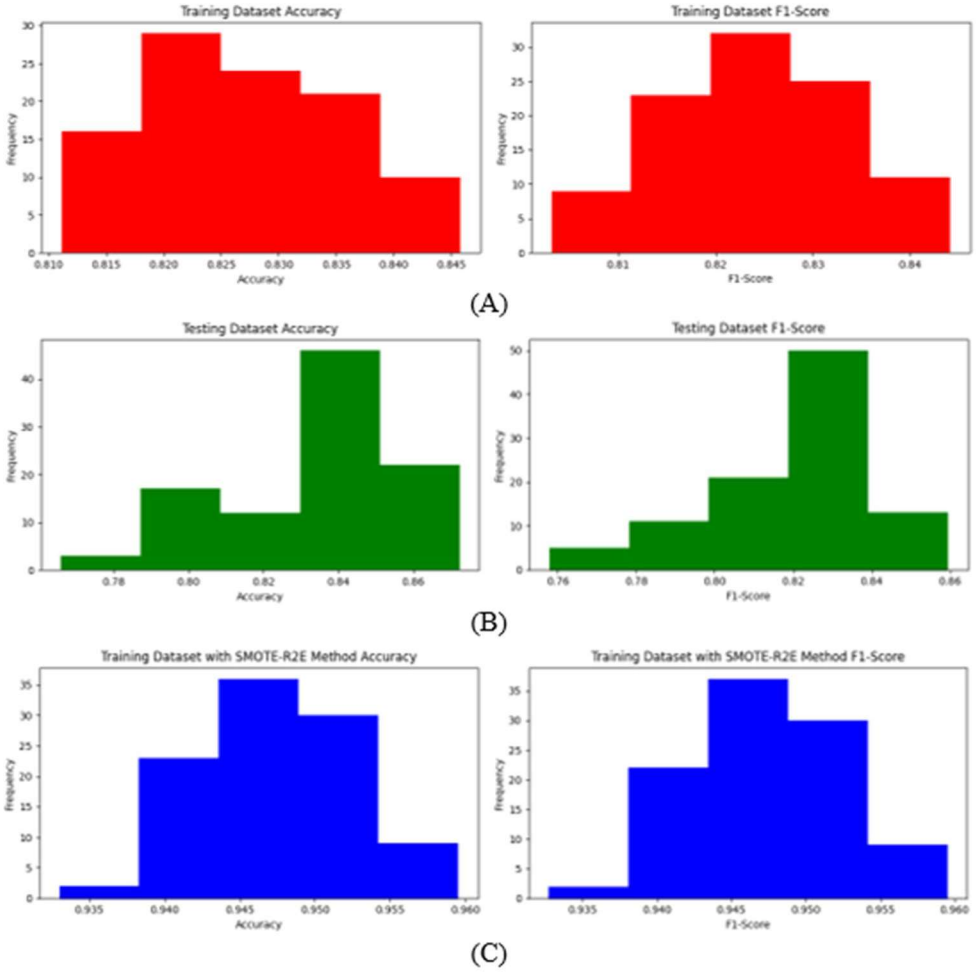

Fig. 9. Accuracy and F1-score graphs from 100 iterations of training. (A) testing on the initial training dataset; (B) testing on the testing dataset; (C) testing on the training dataset from SMOTE-R2E results

The histograms depicted in Fig. 9 vary in skewness's. In Fig. 9(A), the accuracy histogram is right-skewed, indicating that the 100 iterations of training with the training dataset yielded relatively low accuracy values in the distribution of accuracy values obtained. However, this does not necessarily imply poor performance because if we observe closely, the range between minimum and maximum values was not large, and the minimum value was above 80%. As for the F1-score, the histogram in Fig. 9(A) appears symmetric with a relatively narrow range between minimum and maximum values. Meanwhile, as shown in Fig. 9(B), the histograms of the results of the testing on the testing dataset exhibit left-skewed distributions, indicating that the accuracy values and F1-scores obtained tended to be high within the distribution of the accuracy values and F1-scores obtained. However, it is still important to note that the range between the maximum and minimum values yielded with the testing dataset was quite large, although there were fewer low accuracy values and F1-scores in that distribution. As shown in Fig. 9(C), the application of SMOTE-R2E yielded fairly good results because the various data distribution for the model training process could provide accuracy values and F1-scores that tended to be high. In addition, the range between minimum and maximum values was relatively narrow, with a minimum value above 93%.

A summary of the minimum, maximum, and mean values obtained is provided in Table 12. The values obtained with the model where SMOTE-R2E was applied to the training dataset demonstrated increases. The average accuracy was quite high compared to the accuracy of the two other models. In addition, there were also increments in the maximum and minimum values on all three types of datasets

Table 12. Confusion matrix for the testing dataset

Parameter	Training dataset		Testing dataset		SMOTE-R2E dataset	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Mean	82.67%	82.40%	83.04%	81.96%	94.74%	94.75%
Min	81.12%	80.31%	76.60%	75.79%	93.30%	93.27%
Max	84.57%	84.41%	87.23%	85.95%	95.95%	95.95%

The increased accuracy and F1-scores were in correspondence with reductions in classification errors. However, misclassifications that occurred still need to be considered. In this model, the misclassifications were still tolerable because they did not exceed a one-level difference. For example, the misclassifications between "very poor" and "fair" (more than one-level difference) were not as frequent as in the other two models. The results of one iteration using this model are shown in Table 13.

Table 13. Confusion matrix for the testing dataset

Actual	Predictions			
	Very poor	Poor	Fair	Good
Very Poor	56	1	2	0
Poor	8	20	0	0
Fair	0	2	3	0
Good	0	0	2	0

Based on the results, the ANN method with the architecture used in this study is viable for classifying transformer conditions with fairly good results. However, in this study, the data used was not balanced, which affected the model obtained. Therefore, SMOTE and SMOTE-R2E synthetic data generation techniques were used in this study. The evaluation results showed that the SMOTE-R2E method was able to increase the average accuracy and F1-score from testing the training and testing datasets: the average accuracy and F1-score for the training dataset were 82.67% and 82.40%, respectively, and the average accuracy and F1-score for the testing dataset were 83.04% and 81.96%, respectively. A complete comparison can be seen in Table 14 and Table 15.

Table 14. Summary of accuracy results

Dataset	The model mean accuracy		
	With		Without
	SMOTE-R2E	SMOTE	SMOTE
Training dataset (376 data points)	82.67%	79.64%	79.01%
Testing dataset (94 data points)	83.04%	76.57%	80.03%

Table 15. Summary of F1 score results

Dataset	The model mean F1 score		
	With		Without
	SMOTE-R2E	SMOTE	SMOTE
Training dataset (376 data points)	82.40%	80.34%	78.20%
Testing dataset (94 data points)	81.96%	77.48%	78.46%

Martinez-Gil et al. (2022) used the ANN method and achieved a high accuracy, but more data was required. A similar data limitation was encountered by Velásquez and Lara (2020). It can be inferred from these reports that the direct use of the ANN method with limited data will not yield good results (see Table 15). Hence, synthetic data generation is needed to overcome limited data. One of the synthetic data generation methods commonly used in previous studies is SMOTE (Pribadi et al., 2022; Lopo and

Hartomo, 2023). These studies have shown that SMOTE could improve the goodness of fit of various classification models. However, in this study, the accuracy and F1-score derived from the evaluation using the testing dataset generated with SMOTE were lower than those derived from the evaluation using the testing dataset without SMOTE, although the evaluation using the training dataset generated with SMOTE showed better results. This indicates that using SMOTE on the dataset yielded less optimal results because the generated synthetic data was not evenly distributed. The drawbacks of SMOTE are addressed with a method developed in this study, namely SMOTE-R2E. Using this novel method, better accuracy and F1-score were obtained from the testing on both the testing and training datasets (Table 14 and Table 15).

Martinez-Gil et al. (2022) have demonstrated the applicability of various methods for classification using Velásquez's dataset. They obtained an accuracy of 81.4% with the SWRL (expert rules) and Random Forest methods to detect machine failure (transformer), with a total of two classes, higher than the accuracy obtained with other methods, including the Multilayer Perceptron, Support Vector Machine, and k-Nearest Neighbors methods. However, in comparison to Martinez-Gil et al.'s accuracy, the present study showed better accuracy from the testing on the testing dataset (Table 16). Additionally, this study also used more classes than the number of classes used by Martinez-Gil et al. (four vs. two), which also means higher complexity. Therefore, it is possible that SMOTE-R2E will yield an even higher accuracy when two classes are used as in the previous study. Thus, a future study can be conducted using two classes and the same features as the previous study to see the effectiveness of the SMOTE-R2E method.

Table 16. Classification accuracy values using Velásquez's dataset

Method	Accuracy
Multilayer perceptron	76.30%
Support vector machine	78.00%
k-Nearest neighbors	79.70%
Random forest	81.40%
SWRL (expert rules)	81.40%
ANN with SMOTE-R2E	81.96%

SMOTE has had variations as it develops. For example, SMOTE-Edited Nearest Neighbors (SMOTE-ENN) deletes data obtained by SMOTE if the data allows misclassifications (Nishat et al., 2022). Then, SMOTE-Tomek Link Removal (SMOTE-Tomek) implements sample removals, in which case a pair of very close minority and majority samples are removed (Sasada et al., 2020). There is also SVM-SMOTE, which is a combination of SMOTE and SVM (Almajid, 2022). The present study did not use the aforementioned SMOTE variants but developed a novel method called SMOTE-R2E. Certainly, this novel method took an inspiration from previous development ideas since it includes data deletion and combines SMOTE with other methods. The novelty, which is the focus of the SMOTE-R2E method, lies in the distribution of more varied synthesis data, as shown in Fig. 6, involving the checking of synthesis data and the combination of synthesis and actual data, performed repeatedly until the desired amount of data is reached. In addition, selecting the correct k value for the synthesis data checking process in the SMOTE-R2E method could prevent an inappropriate data point on a line segment between two outliers or data points in inappropriate positions (data with incorrect classification).

The results of this study can be maximized by changing various parameters in the ANN architecture. In addition, the number of categories used is essential; four categories were used in this study. The categories were created based on the HI value of the transformer condition, which means that the categories can be changed. Fewer categories in previous studies provided better accuracy in classifying transformer conditions. For example, Alqudsi and El-Hag (2019) achieved a 95.1% accuracy with three categories. The better accuracy with these three categories was obtained because of more balanced data for each category. In addition, when carefully examined, similar data points obtained in this study fell into different categories: "poor" and "very poor". Merging these two categories into one could result in better accuracy.

4. CONCLUSION

Using the ANN method with the architecture designed in this study, four categories of transformer conditions were classified using six features. The test results showed that using the proposed method could overcome the problem of unbalanced data and provide better results than yielded by the SMOTE method. One hundred iterations of training with SMOTE-R2E on the testing dataset resulted in average accuracy and F1-score of 83.04% and 81.96%, respectively. The average accuracy value was 6.47% higher than the accuracy value obtained by the model trained on the dataset generated using the SMOTE method, and 3.01% higher than the accuracy value obtained by the model trained only on the existing training dataset without using the SMOTE method. Meanwhile, the F1-score was 4.48% higher than the F-score obtained by model trained on dataset generated using the SMOTE method, and 3.5% higher than the F1-score obtained by the model trained using the unbalanced training dataset that did not apply the SMOTE method.

ACKNOWLEDGMENTS

This research is supported by the Vice-Rector for Research, Innovation, and Entrepreneurship of Satya Wacana Christian University.

REFERENCES

Abdullah, A.M., Ali, R., Yaacob, S.B., Mansur, T., Baharudin, N.H. 2021. Prediction of transformer health index using condition situation monitoring (csm) diagnostic techniques. Journal of Physics: Conference Series, 1878(1), 1–9.

Almajid, A.S. 2022. Multilayer perceptron optimization on imbalanced data using svm-smote and one-hot encoding for credit card default prediction. Journal of Advances in Information Systems and Technology, 3(2), 67–74.

Alqudsi, A., El-Hag, A. 2019. Application of machine learning in transformer health index prediction. Energies, 12(14), 1–13.

Bingham, G., Miikkulainen, R. 2021. AutoInit: analytic signal-preserving weight initialization for Neural Networks. The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23), 6823–6833.

Bohatyrewicz, P., Banaszak, S. 2022. Assessment criteria of changes in health index values over time—A transformer population study. Energies, 15(16), 1–15.

Bustamante, S., Manana, M., Arroyo, A., Castro, P., Laso, A., Martinez, R. 2019. Dissolved gas analysis equipment for online monitoring of transformer oil: A review. Sensors, 19(19), 4057.

Das, S., Adhikary, A., Laghari, A.A., Mitra, S. 2023. Eldo-care: EEG with kinect sensor based telehealthcare for the disabled and the elderly. Neuroscience Informatics, 3(2), 1–6.

El-Harbawi, M. 2022. Fire and explosion risks and consequences in electrical substations—a transformer case study. ASME Open Journal of Engineering, 1, 1–27.

Gorunescu, F. 2011. Data Mining Concepts, Models and Techniques. Springer-Verlag Berlin Heidelberg. New York. USA.

Hernanda, I.G.N.S., Mulyana, A.C., Asfani, D.A., Negara, I. M.Y., Fahmi, D. 2014. Application of health index method for transformer condition assessment. TENCON 2014 - 2014 IEEE Region 10 Conference.

Ho, Y., Wookey, S. 2020. The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. IEEE Access, 8, 4806–4813.

Hoole, P.R.P., Anak, S., Izzati, N., Hafiez, M. 2017. Power transformer fire and explosion : causes and control. International Journal of Control Theory and Applications, 10(16), 211–220.

Indrajaya, D., Setiawan, A., Susanto, B. 2022. Comparison of k-Nearest Neighbor and Naive Bayes Methods for SNP Data Classification. Matrik: Jurnal Manajemen, Teknik Informatika, Dan Rekayasa Komputer, 22(1), 149–164.

Jais, I.K.M., Ismail, A.R., Nisa, S.Q. 2019. Adam optimization algorithm for wide and deep neural network. Knowledge Engineering and Data Science, 2(1), 41–46.

Jamal, P., Ali, M., Faraj, R.H., Ali, P.J.M., Faraj, R.H. 2014. Data normalization and standardization: A Technical Report. Machine Learning Technical Reports, 1(1), 1–6.

Laghari, A.A., Estrela, V.V., Li, H., Shoulin, Y., Khan, A.A., Anwar, M.S., Wahab, A., Bouraqia, K. 2024. Quality of experience assessment in virtual/augmented reality serious games for healthcare: A systematic literature review. Technology and Disability, Pre-press, 1–12.

Laghari, A.A., Estrela, V.V, Yin, S. 2024. How to collect and interpret medical pictures captured in highly challenging environments that range from nanoscale to hyperspectral imaging. Current Medical Imaging, 20, 1–10.

Laghari, A.A., Shahid, S., Yadav, R., Karim, S., Khan, A., Li, H., Shoulin, Y. 2023. The state of art and review on video streaming. Journal of High Speed Networks, 29(3), 211–236.

Laghari, A.A., Sun, Y., Alhussein, M., Aurangzeb, K., Anwar, M.S., Rashid, M. 2023. Deep residual-dense network based on bidirectional recurrent neural network for atrial fibrillation detection. Scientific Reports, 13, 1–12.

Liu, D., Chen, X., Guo, Z., Yuan, J., Yin, S. 2023. Research on the online parameter identification method of train driving dynamic model. International Journal of Computational Vision and Robotics, 13(5), 497–509.

Lopo, J.A., Hartomo, K.D. 2023. Evaluating sampling techniques for healthcare insurance fraud detection in imbalanced dataset. JITEKI, 9(2), 223–238.

Martinez-Gil, J., Buchgeher, G., Gabauer, D., Freudenthaler, B., Filipiak, D., Fensel, A. 2022. Root cause analysis in the industrial domain using knowledge graphs: A Case study on power transformers. Procedia Computer Science, 200, 944–953.

Muntasir Nishat, M., Faisal, F., Jahan Ratul, I., Al-Monsur, A., Ar-Rafi, A.M., Nasrullah, S.M., Reza, M.T., Khan, M. R.H. 2022. A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset. Scientific Programming, 2022 (Cvd).

Nanfak, A., Eke, S., Kom, C.H., Mouangue, R., Fofana, I. 2021. Interpreting dissolved gases in transformer oil: A new method based on the analysis of labelled fault data. IET Generation, Transmission and Distribution, 15(21), 3032–3047.

Pribadi, M.R., Purnomo, H.D., Hendry, Hartomo, K. D., Sembiring, I., Iriani, A. 2022. Improving the accuracy of text classification using the over sampling technique in the case of Sinovac Vaccine. 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI).

Saeed, U., Kumar, K., Khuhro, M.A., Laghari, A.A., Shaikh, A.A., and Rai, A. 2024. DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification. Multimedia Tools and Applications, 83, 21019–21043.

Sasada, T., Liu, Z., Baba, T., Hatano, K., Kimura, Y. 2020. A resampling method for imbalanced datasets considering noise and overlap. Procedia Computer Science, 176, 420–429.

Schober, P., Schwarte, L.A. 2018. Correlation coefficients:

appropriate use and interpretation. Anesthesia and Analgesia, 126(5), 1763–1768.

Sharma, S., Sharma, S., Athaiya, A. 2020. Activation Functions in Neural Networks. International Journal of Engineering Applied Sciences and Technology, 4(12), 310–316.

Stangierski, J., Weiss, D., Kaczmarek, A. 2019. Multiple regression models and Artificial Neural Network (ANN) as prediction tools of changes in overall quality during the storage of spreadable processed gouda cheese. European Food Research and Technology, 245, 2539–2547.

Syahfitra, F.D., Syahputra, R., Putra, K.T. 2017. Implementation of backpropagation Artificial Neural Network as a forecasting system of power transformer peak load at bumiayu substation. Journal of Electrical Technology, 1(3), 118–125.

Velásquez, R.M.A., Lara, J.V.M. 2020. Data for: root cause analysis improved with machine learning for failure analysis in power transformers. Engineering Failure Analysis, 115, 104684.

Wang, C.I., Joanito, I., Lan, C.F., Hsu, C.P. 2020. Artificial neural networks for predicting charge transfer coupling. Journal of Chemical Physics, 153(21), 1–15.

Wang, J., Fan, Y., Li, H., Yin, S. 2023. WeChat mini program for wheat diseases recognition based on VGG-16 convolutional neural network. International Journal of Applied Science and Engineering, 20(3), 1–9.

Yin, S., Li, H., Laghari, A. A., Gadekallu, T. R., Sampedro, G. A., Almadhor, A. 2024. An anomaly detection model based on deep auto-encoder and capsule graph convolution via Sparrow Search Algorithm in 6G internet-of-everything. IEEE Internet of Things Journal, Early Access.

Zhengwei, G., Xinju, G., Xiangli, L., Min, T., Dong, L. 2018. Study on Decomposition gas explosion of transformer oil under Arc. 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), 1–5.