

Breast cancer disease classification using fuzzy-ID3 algorithm based on association function

Nur Farahaina Idris¹, Mohd Arfian Ismail¹, Mohd Saberi Mohamad², Shahreen Kasim³,
Zalmiyah Zakaria⁴, Tole Sutikno⁵

¹Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang, Malaysia

²Health Data Science Lab, Department of Genetics and Genomics, College of Medicine and Health Sciences, United Arab Emirates University, Abu Dhabi, United Arab Emirates

³Faculty of Computing Science and Information Technology, Universiti Tun Hussein Onn, Johor, Malaysia

⁴School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor, Malaysia

⁵Department of Electrical Engineering, Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Article Info

Article history:

Received Sep 6, 2021

Revised Jan 5, 2022

Accepted Feb 10, 2022

Keywords:

Association function

Breast cancer

Fuzzy ID3

Fuzzy system

ID3 algorithm

ABSTRACT

Breast cancer is the second leading cause of mortality among female cancer patients worldwide. Early detection of breast cancer is considered as one of the most effective ways to prevent the disease from spreading and enable human can make correct decision on the next process. Automatic diagnostic methods were frequently used to conduct breast cancer diagnoses in order to increase the accuracy and speed of detection. The fuzzy-ID3 algorithm with association function implementation (FID3-AF) is proposed as a classification technique for breast cancer detection. The FID3-AF algorithm is a hybridisation of the fuzzy system, the iterative dichotomizer 3 (ID3) algorithm, and the association function. The fuzzy-neural dynamic-bottleneck-detection (FUZZYDBD) is considered as an automatic fuzzy database definition method, would aid in the development of the fuzzy database for the data fuzzification process in FID3-AF. The FID3-AF overcame ID3's issue of being unable to handle continuous data. The association function is implemented to minimise overfitting and enhance generalisation ability. The results indicated that FID3-AF is robust in breast cancer classification. A thorough comparison of FID3-AF to numerous existing methods was conducted to validate the proposed method's competency. This study established that the FID3-AF performed well and outperform other methods in breast cancer classification.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Mohd Arfian Ismail

Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang

26600 Pekan Pahang, Malaysia

Email: arfian@ump.edu.my

1. INTRODUCTION

Breast cancer is one of the most lethal types of cancer for women worldwide, ranking as the second-largest cause of death for female cancer patients [1], [2]. Over 250,000 women are diagnosed with invasive breast cancer each year in the United States [3]. Breast cancer is primarily caused by family history, as patients inherit genetic abnormalities in their genes [4]. The disease has a rapid metastasis process, enabling the primary tumour to rapidly spread breast cancer cells to other organs [5]. Breast cancer's metastatic features are a significant factor in its high incurability rate [6], [7]. Breast cancer therapy advances have resulted in a decrease in breast cancer mortality rates across all age groups [8]–[10]. Younger women, on the other hand, had a poorer

survival rate and a greater risk of local recurrence [1]. Apart from establishing prophylactic methods, early detection of breast cancer is critical to halt the disease's rapid progression [6].

While a clinician can identify breast cancer manually, it would take an extended length of time and much effort to conduct the diagnosis or classification for breast cancer [11]–[16]. Thus, early diagnosis of breast cancer using an intelligent system, such as automatic diagnostic methods, is critical in the medical field [17]–[22]. Classification can be accomplished using well-known machine learning techniques such as deep neural networks, support vector machines, k-nearest neighbour (KNN), random forest, and decision trees [11], [14], [18]–[43]. This article presents the fuzzy-ID3 algorithm with an association function (FID3-AF) to increase the efficiency of breast cancer classification. The advantages of this strategy include the decision tree's excellent intelligibility, interpretability, and ability to deal with erroneous and uncertain data in a fuzzy representation [13], [14], [16], [44]–[48]. However, with FID3-AF, which does data fuzzification, it is required to define the fuzzy database.

The automatic definition of fuzzy databases using formal methods such as genetic algorithms and fuzzy clustering algorithms are computationally costly and do not always produce the best results because there are no general rules for selecting the most appropriate method for each specific domain [44], [49]–[54]. Thus, this article implemented the fuzzy-neural dynamic-bottleneck-detection (FUZZYDBD), an automatic fuzzy database definition method proposed by Cintra [44], to create the fuzzy database in FID3-AF. The FID3-AF approach was proposed to create a classification system that is both rapid and effective. FID3-AF fuzzified data and substituted numerical values with linguistic variables from fuzzy sets with the highest degree of compatibility. Thus, it overcomes the classic ID3 algorithm's problem of being unable to classify continuous data. The association function is implemented in the decision tree to address the overfitting and poor generalisation issues associated with the standard ID3 [55], [56]. Testing and verification processes are used to validate the method's performance. The remainder of this work is structured as: the materials and methods, the results, the discussion, and the conclusion.

2. METHOD

2.1. Fuzzy system

Zadeh [57] proposed the fuzzy logic in 1965. It is a precise imprecision logic based on the degree of truth. The fuzzy system, which is based on fuzzy logic, operates through the use of logical variables with values ranging from 0 to 1. It is typically used to address data imprecision issues through the use of fuzzy set theory. The fuzzy system is composed of three key processes: fuzzification, inference engine, and defuzzification. Fuzzy systems need granulation of the domain's features, which are represented as fuzzy sets and partitions, in order to do fuzzification using a fuzzifier [58]. Fuzzy sets and partitions would be used to create the membership functions. Each fuzzy set is individually described by a single uniquely stated membership function [59]. As a result, the specific membership functions are usually identified by the labels of the associated fuzzy sets. For each of the fuzzy sets, each input factor is allocated a range of values [60]. Membership functions are used to record the degree of membership in fuzzy sets and to grade their membership.

The rule-based fuzzy system consists of a knowledge base and an inference engine [61]. The knowledge base was divided into two parts: a fuzzy rule base (FRB) holding a set of fuzzy rules for the dataset and a fuzzy database containing the definitions of the fuzzy sets, including the linguistic variables used in the FRB [62]. Meanwhile, the inference process generated the system's outputs via fuzzy reasoning. The inference engine, which also can be referred to as the inference mechanism, would use fuzzy rules to translate the input to the output. The inference engine generates the most relevant consequents for each rule. When building a fuzzy inference system (FIS), the connection between fuzzy parameters such as rules and membership functions is essential [63]. The following is how the fuzzy rule is expressed:

$$\text{IF } (V1 \ i \ \text{IS } A1i) \ \text{AND } (V2 \ i \ \text{IS } A2 \ i) \ \text{AND} \dots (Vj \ i \ \text{IS } A_j \ i) \ \text{THEN } (\text{CLASS}i \ \text{IS } C \ i)$$

The linguistic variable V has been used to represent the linguistic variable of the attribute, whereas A is the linguistic value of the data. Following that, j denotes the number of data attributes, whereas i denotes the number of rules. The primary distinguishing characteristics of a fuzzy system are incorporating linguistic variables, the interdependence of variables via conditional rules, and applying the fuzzy technique to validate complicated relationships [64]. Incorporating linguistic variables implies that the variables can be interpreted in natural language using the fuzzy technique. The interdependence of variables via conditional rules indicates that linguistic variables in the antecedent portion of fuzzy rules express as characteristics, but those in the consequent part express as classes. Validation of complicated dependency via fuzzy approaches allows for the justification of the link between a class and a linguistic variable via fuzzy logic. Finally, defuzzification is used when precise values are necessary, and the resulting values are aggregated to create

precise output. Defuzzification is the process of converting a fuzzy output to a crisp value by utilising the fuzzy set and membership degree [60], [65]. The most well-known defuzzification methods are the centre of gravity (COG), mean of maximum (MOM), and centre average [66]. The fuzzy system is a popular method in machine learning since it is easier than other methods and can achieve high accuracy rates simultaneously [67].

According to [68], the adoption of the fuzzy approach is effective at detecting breast cancer, with an accuracy of greater than 98%. Furthermore, according to [69], when used to a histopathology image dataset, the fuzzy model attained an accuracy of 94.26%. Meanwhile, Schaefer *et al.* [70] shown that when applied a fuzzy technique, the experimental findings of breast cancer thermogram classification reached an accuracy of 80%. The superior accuracy results of prior researches have demonstrated the efficiency of the fuzzy technique in classifying data.

2.2. Decision tree: ID3 algorithm

In machine learning, a decision tree (DT) is a well-known classification method [71]–[76]. Numerous advantages exist for using DT in classification, including superior interpretability, scalability, and the ability to express in both graphical and textual formats [19], [77]–[79]. The three most well-known decision tree learning algorithms are ID3, C4.5, and CART. ID3 is chosen to examine decision tree learning in this research since it is the most often used learning algorithm [80]–[82]. In 1986, Quinlan created the iterative dichotomizer 3 (ID3) [83], [84]. The ID3 approach is based on recursive partitioning, in which the training data is divided into subsets, and the subsets produce the decision tree's partitions [77], [85]. The ID3 algorithm selects attributes based on Shannon's entropy and information gain [85]. The information entropy is determined using (1), where S represents sets that are classified into multiple classes and p_i denotes the proportion of data that is classified into each class. Equation (2) illustrates the computation of information gain, where A represents the current attribute, S indicates sets that are separated into various classes, and n denotes the potential values for the corresponding attribute. S_j signifies the subsets that have the same credits attribute A .

$$Entr(S) = - \sum_{i=1}^k p_i \log_2(p_i) \quad (1)$$

$$Gain(A, S) = Entr(S) - \sum_{j=1}^n \frac{|S_j|}{|S|} Entr(S_j) \quad (2)$$

The ID3 algorithm prioritises the largest information gain in order to determine the most acceptable qualities in each tree branching. The branching procedure is repeated repeatedly using a greedy top-down technique until the decision tree satisfies the termination conditions, such as all attributes in the datasets are fully utilised or all balance instances have the same class. Classification using ID3 is only possible if the dataset in question contains more than one class [83]. The ID3 algorithm creates rules for class prediction while also highlighting the properties or characteristics of the corresponding class [86]. The generated decision tree size is small with ID3 algorithm implementation due to the algorithm's usage of quality measures and logical reasoning.

Angayarkanni and Kamal [87] evaluated the ID3 algorithm's performance and capability for breast cancer domain classification using an magnetic resonance imaging (MRI) mammogram image dataset. The aforementioned dataset has three distinct classes: benign, malignant, and normal. The average accuracy of the ID3 algorithm was 99.9%, and the training time of the algorithm only takes 0.03 seconds. It illustrates the algorithm's ability to get correct classification results in a reasonable duration. According to Yang *et al.* [88], the algorithm achieved a prediction accuracy of 90.56% using the wisconsin breast cancer dataset. Additionally, Jacob and Ramani [89] demonstrated that when used with the wisconsin prognostic breast cancer dataset (WPBC), ID3 beat other classifiers in terms of classification accuracy, including naive Bayes and partial least squares for classification (C-PLS). There are several advantages to using the ID3 algorithm for DT learning, and the most significant is its short execution time [56], [90].

2.3. Fuzzy-neural dynamic-bottleneck-detection (FUZZYDBD) method

The fuzzy automated definition technique is crucial for developing a fuzzy database [49]. The three components of developing a fuzzy database are the type of the membership functions, followed by the number and distribution of fuzzy sets associated with each domain characteristic [44]. Numerous techniques, such as the evolutionary algorithm, the artificial neural network, and the KNN, may be used to construct

fuzzy databases to estimate the number of fuzzy sets. Despite the fact that numerous methods exist, it is critical to emphasise that many studies define the number of fuzzy sets empirically and then distribute them evenly for the membership function due to the high complexity of the available methods and the adaptability of fuzzy logic to achieve improved performance [49]. Additionally, there is a dearth of consensus on the most effective existing techniques for certain applications and domains [91]. The study identified the FUZZYDBD technique, proposed in [49], as a quick, simple, and effective method for defining fuzzy databases [9]. Furthermore, it had been empirically evaluated in several other fields, including breast cancer. The FUZZYDBD method collects all essential features to develop a fuzzy database and overcomes the limitations of previous techniques for developing fuzzy databases [49], [91], [92]. According to Cintra [44], the FUZZYDBD method employs a variety of techniques, including the Wang Mendel method [93] for determining the total number of fuzzy sets for all attributes, the equalised universe method for partitioning of fuzzy sets across domains and the usage of triangular membership functions.

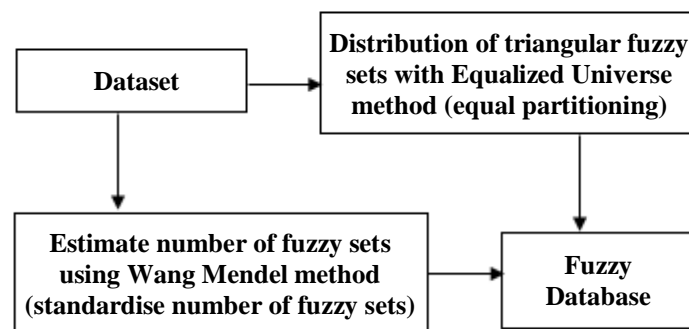


Figure 1. Approaches of the FUZZYDBD method

In the approach of the FUZZYDBD method, the number of fuzzy sets is uniform and consistent across all attributes in the dataset and can be ranging from two to ten triangular membership functions. The appropriate number of fuzzy sets between the ranges is determined empirically. Additionally, a domain expert can develop more suitable fuzzy set values for the attributes [44], [94]. The Wang-Mendel technique uses the same number of fuzzy sets, distribution, and shape of fuzzy sets to define all attributes in the obtained dataset [44]. Although the number of fuzzy sets can be defined in the range of two to ten, the FUZZYDBD method is more usually used with three, five, or seven [44], [95], [96]. Both two and three are the most conservative estimates of the number of fuzzy sets with the lowest error rate for the breast cancer dataset [49].

Chen and Wang devised the equalised universe method in 1999, and this technique was later incorporated by the FUZZYDBD method [44]. The approach is applied with the same width for each fuzzy set to ensure that the fuzzy sets in the domains' attributes are partitioned equally. The attribute's maximum value is located at the peak of the most right triangular membership function, while the attribute's minimum value is located at the peak of the most left triangular membership function. When triangular shape membership functions with half overlap are applied, no region has a membership degree greater than 0.5. This method is a frequently used technique in the literature [44]. One of this technique's primary advantages is that it generates extremely simple fuzzy databases.

2.4. Association function

The association function (AF), developed from the correlation function, emphasises the association between randomly chosen variables [86]. One of the benefits of AF implementation is that it overcomes the limitation of the conventional ID3 algorithm, which tends to choose attributes with a large number of attribute values and illustrates the link between the attributes and their corresponding class attributes concurrently [97]. The implementation begins with the computation of AF for each attribute in the dataset. For example, there are multiple attributes in dataset D , and let C be any of the dataset's class attributes. Thus, (3) depicts the degree function of the relationship between S , which represents one of the attributes, and C . Equation (3) lists the representation, where n denotes the number of attribute types in S . Both x_{i1} and x_{i2} illustrate two distinct situations, with dataset D 's attribute S representing the i th value and category C representing the j th value ($j=1,2$) [82]. Then, the computation value of the normalised relation degree function, also known as the normalised gain, is performed. Equation (4) is the equation for normalised gain for each attribute, where m is the total number of attributes in the dataset.

$$AF(S) = \frac{\sum_{i=1}^n |x_{i1} - x_{i2}|}{n} \quad (3)$$

$$V(k) = \frac{AF(k)}{AF(1) + AF(2) + \dots + AF(m)} \quad (4)$$

Finally, the initial (old) information gain has been combined with a normalised gain, as shown in (5), where $Gain'(A)$ denotes the new gain. The normalised gain value is later multiplied with the existing information gain to create a new gain that serves as the decision tree's attribute selection criterion [46]. Similar to ID3, the tree's node is determined by the largest value for the attribute selection criterion. This approach generates a new criterion for the ID3 algorithm's attribute splitting [86].

$$Gain'(A) = (I(s_1, s_2, \dots, s_m) - E(A)) * V(A) \quad (5)$$

According to [82], AF overcame the limitations of the classic ID3 algorithm involving multi-value bias because the generated decision tree chose age with three attribute values as the root. In contrast, the classic ID3 algorithm chose colour-cloth with four attribute values as the root when applied to the customer dataset. According to [97], when applied to the mutual fund application dataset, AF created a more plausible tree because the DT chose status with two attribute values as the root instead of the standard ID3 approach, which chose age with three attribute values as the root. The application of AF reduces the significance of attributes with a large number of values. To address the greedy algorithm's flaws, it enhances the relevance of characteristics with fewer values and uses the relation degree values to express the attributes' importance [82]. Sivakumar *et al.* [55] further demonstrated that when the student dropout dataset was used, the modified DT method had greater accuracy than the classic ID3 algorithm, at 97.5% versus 92.5%. Implementing AF can help alleviate the overfitting problem and increase the generalisation of the model [55]. Additionally, AF produces more productive rules for the decision tree since it generates more optimal DT [97].

2.5. Proposed method: fuzzy-ID3 algorithm based on association function (FID3-AF)

FID3-AF will classify the fuzzified data by utilising a hybrid ID3 algorithm with AF as the classifier and configure the required fuzzy set parameters for the fuzzification process using the FUZZYDBD method. FID3-AF is a simple, comprehensible and interpretable method. FID3-AF fuzzified the dataset and estimated each numerical attribute's membership degree, including the attribute with integer or continuous values. The test data would be applied automatically using the most compatible rules with the input pattern to determine the classes. The proposed approach, which fuzzifies the entire dataset, enhances patients' privacy by concealing specific patient data. Medical secrecy is critical for patient records' security and confidentiality. The linguistic variable is beneficial, especially when a coalition of support for the linguistic words spans the whole domain, greatly improving accuracy and performance [98].

The FID3-AF implementation begins with listwise deletion to handle missing data, followed by the construction of membership functions for all continuous features in the datasets using the FUZZYDBD method. Each gathered dataset's numerical attributes with continuous values are defined using triangular equal partitioning membership functions and a predefined number of fuzzy sets (all domain attributes have the same count of fuzzy sets). Three, five, or seven fuzzy sets are the most often utilised number of fuzzy sets in the FUZZYDBD method study [95], [96]. All attributes in the datasets were assigned a value of three because this is also the best-estimated number of fuzzy sets in the breast cancer domain, along with the value of two, unless a medical expert determined that a different number of fuzzy sets should be used for the respective attribute [44], [49]. The dataset's attribute values are then fuzzified to create data with fuzzy values. All integer and continuous attribute data are replaced with linguistic labels from fuzzy sets with the maximum degree of compatibility with the input values or fuzzy values to assure the linguistic formation of the newly generated dataset. After fuzzifying the dataset, the training data and testing data are separated using a 10-fold cross-validation technique. Following that, the training data would be classified using the ID3 algorithm that had been undergone modification with the association function. FID3-AF's classification method is identical to the standard ID3 algorithm but includes AF, normalised gain, and new gain (modification of information gain) as the new computation technique. The enhanced decision tree, FID3-AF, employed new gain as an attribute selection criterion. The fuzzified data is utilised to construct an upgraded fuzzy decision tree, and the fuzzy decision tree's generated rules are used to evaluate the method's effectiveness. The most compatible rules were applied to deduce the classes of test data. Finally, the classification rates would be ascertain using the testing data classification performance. The flow of FID3-AF is seen in Figure 2.

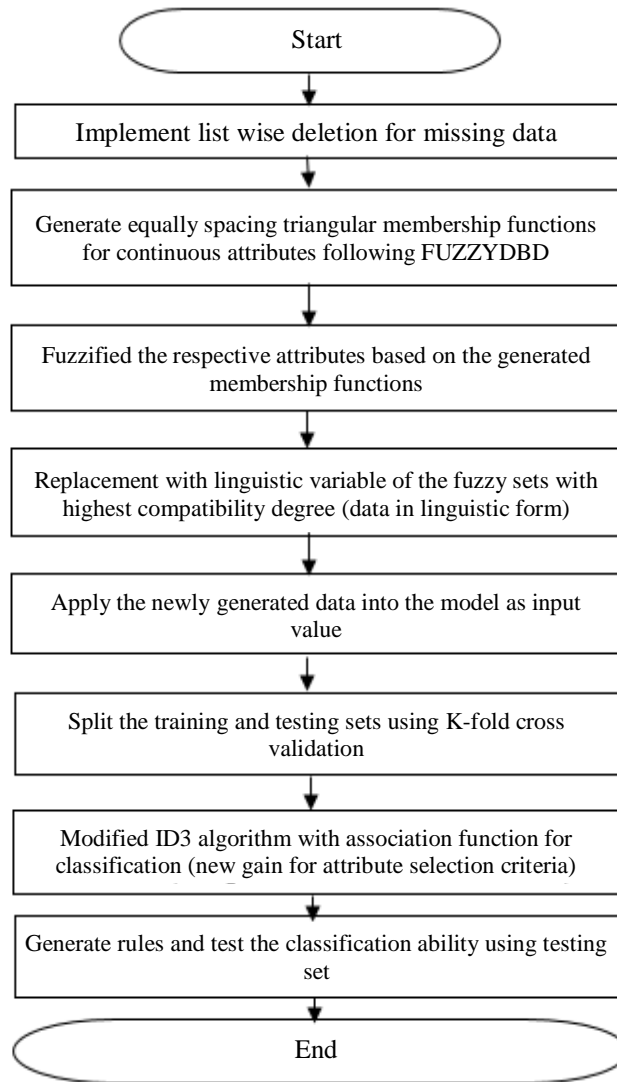


Figure 2. Flow of FID3-AF

3. RESULTS AND DISCUSSION

This study utilised four breast cancer datasets: the wisconsin breast cancer dataset (original), which can be referred to as WBCD (original), the wisconsin diagnostic breast cancer (diagnostic), which can be referred to as WDBC (diagnostic), the Coimbra dataset, and the mammographic mass dataset. All datasets were collected from the University of California Irvine machine learning repository. The WBCD (original), WDBC (diagnostic), and Mammographic Mass datasets are used to differentiate between benign and malignant samples, whilst the Coimbra dataset is used to identify healthy controls and patients. Table 1 contains descriptions of the datasets that were gathered.

The proposed method's effectiveness is determined using a 10-fold cross-validation technique, and missing data are handled via listwise deletion. The study defined all attributes in the datasets using three evenly distributed triangular fuzzy sets for fuzzification. The confusion matrix is utilised to demonstrate the proposed method's classification capability through the computation of the accuracy, precision, recall, and F1-measure. The results from ten independent runs were averaged using the macro-average approach. The precision, recall, F1-measure, and standard deviation of the accuracy (SD) are tabulated in Table 2.

The proposed method's accuracy was compared to other existing techniques in order to establish and validate the proposed model's capabilities. Tables 3 to 6 compare the proposed method's accuracy to existing techniques for each of the collected datasets. When the approach does not support the classification, a (-) is indicated. The results show that the proposed method, FID3-AF, acquired the highest accuracies compared to other existing methods.

Table 1. The number of attributes, instances and classes for the datasets

Dataset	Number of attributes	Number of instances	Number of classes
WBCD (Original)	10	683	2
WDBC (Diagnostic)	32	569	2
Coimbra	10	116	2
Mammographic Mass	6	830	2

Table 2. Experimental results of the collected datasets

Dataset	Precision (%)	Recall (%)	F1-Measure (%)	SD (decimal)
WBCD (Original)	94.425	92.992	93.641	0.0046
WDBC (Diagnostic)	94.194	93.983	94.118	0.0030
Mammographic Mass	81.423	81.232	81.267	0.0064
Coimbra	70.040	68.959	69.229	0.0167

Table 3. Accuracy of the methods using the WBCD (Original) dataset

Method	Accuracy (%)
Kwok's SVM Gaussian [99]	91.600
DT (C4.5) [100]	92.970
Linear regression [101]	79.840
ID3 algorithm	92.650
Proposed method	94.305

Table 4. Accuracy of the methods using the WDBC (Diagnostic) dataset

Method	Accuracy (%)
RBF Network [102]	93.670
Naïve Bayes [103]	90.351
Random Forest-CFS [104]	89.510
ID3 algorithm	-
Proposed method	94.517

Table 5. Accuracy of the methods using the Coimbra dataset

Method	Accuracy (%)
Average K-means + Local Outlier Rectifier V.2.0 [105]	56.900
Neural Network [106]	60.000
Random Forest-ReliefF [107]	68.000
ID3 algorithm	-
Proposed method	70.862

Table 6. Accuracy of the methods using the Mammographic Mass dataset

Method	Accuracy (%)
Random Tree-SVR [108]	80.440
RBF Network [109]	80.190
CHAID [110]	78.130
ID3 algorithm	74.542
Proposed method	81.325

The radar charts were created to evaluate and compare the FID3-AF algorithm's overall classification performance with the existing ID3 algorithm. The ID3 algorithm cannot classify the WDBC and Coimbra datasets because it cannot handle real-valued data or continuous values with floating points. Thus, performance comparisons between the two methods were undertaken solely based on the WBCD dataset, which contains integer values ranging from 1 to 10, and the Mammographic Mass dataset, which has only one continuous attribute. The radar chart for the WBCD dataset classification is shown in Figure 3, while the radar chart for the Mammographic Mass dataset classification is shown in Figure 4. FID3-AF outperforms the ID3 algorithm in terms of accuracy, precision, recall, and F1-measure for both datasets.

A t-test was performed to find out if the classification accuracy of the ID3 algorithm and FID3-AF is statistically different. The t-value and p-value were determined by analysing the accuracy of ten runs per sample with a total degree of freedom of 18. The significance level (α) was adjusted to 0.05, and a two-tailed test was used. If the p-value is greater than α , the null hypothesis, H_0 , is adopted, stating that there is no significant

difference between ID3 and FID3-AF. Otherwise, if the p-value is less than α , the alternative hypothesis, H1, that asserts a substantial difference exists is adopted. The t-test result for the WBCD dataset classification returns a t-value of 7.28533 and a p-value less than 0.00001. Thus, the results between the two methods for WBCD dataset classification are significantly different at p less than 0.05. Then, the t-test result for the Mammographic Mass dataset classification is 20.27207, with a p-value less than 0.00001. Therefore, the Mammographic Mass dataset classification result is also significantly different at p less than 0.05. According to this research, all p-values were lower than the level of significance. Thus, it shows that the results of ID3 and FID3-AF are statistically and significantly different at a p-value less than 0.05, where the null hypothesis, H0, is rejected and the alternative hypothesis, H1, is accepted.

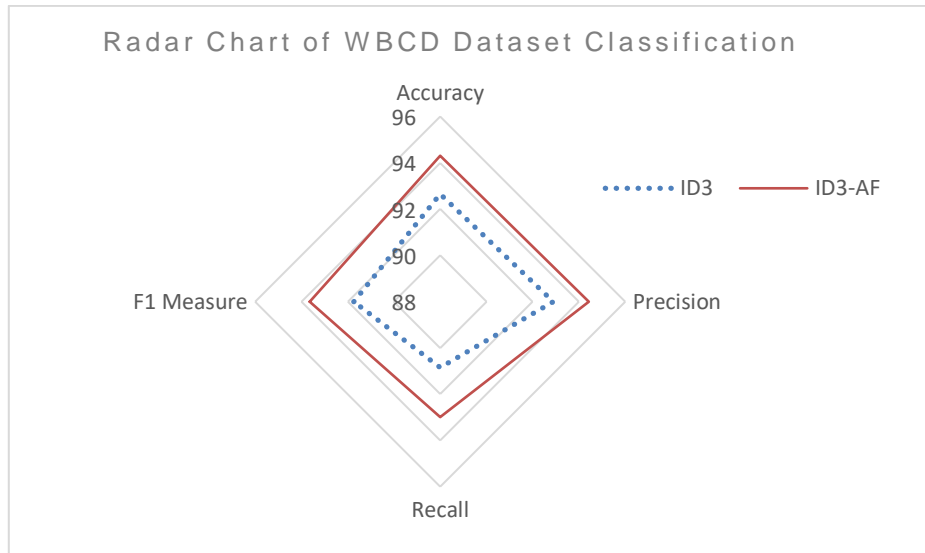


Figure 3. Radar chart of WBCD dataset classification

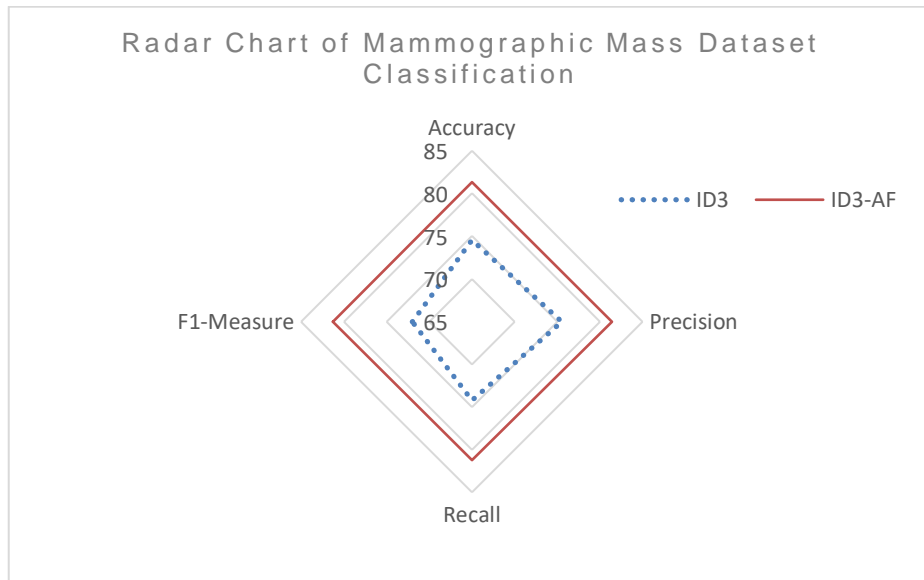


Figure 4. Radar chart of mammographic mass dataset classification

FID3-AF achieved an accuracy of 94.305% when tested against the WBCD dataset, 94.517% when tested against the WDBC dataset, 70.862% when tested against the Coimbra dataset, and 81.325% when tested against the Mammographic Mass dataset. The total results indicated that FID3-AF outperformed and was comparable to a large number of previously published methods. The standard ID3 algorithm is incapable

of classifying datasets that contain continuous values with floating points, such as the WDBC and Coimbra datasets. FID3-AF is more practical than classic ID3 in terms of solving breast cancer classification problems since it has demonstrated superior classification results and has the ability to handle continuous data. The FUZZYDBD method was widely used for data fuzzification in the literature and had shown exceptional performance when evaluated on a breast cancer dataset. In FID3-AF, the FUZZYDBD technique was used to configure the fuzzy sets parameters in order to increase the efficacy of data fuzzification, notably in the breast cancer field. Additionally, FID3-AF performed a replacement operation with the most compatible fuzzy set's linguistic variable, converting the dataset's continuous attributes to linguistic forms. Thus, any gathered datasets can be classified, including those with continuous values. The use of data fuzzification and the linguistic variable replacement process in FID3-AF enabled the decision tree to choose the optimal rules for new instances based on logical reasoning. The implementation of AF addressed the shortcomings of the ID3 algorithm, which included high variance, poor generalisation ability, and a proclivity for overfitting. Furthermore, FID3-AF also proved to be more robust than ID3. In ID3, even discrepancies of 0.01 continuous values in the data would result in a tree with different routes and classes, whereas FID3-AF evaluates the membership degree of each input and has a high tolerance for data uncertainty. In future research, the feature selection approach may be utilised to improve classification performance.

4. CONCLUSION

FID3-AF is efficient and produces superior classification results for breast cancer data. The FUZZYDBD method is simple and effective, allowing for a rapid fuzzification process to occur. The proposed FID3-AF method is simple to comprehend, performs well in classification, and alleviates the overfitting and generalisation issues associated with the classic ID3. Moreover, FID3-AF overcomes the drawback of the classic ID3 algorithm, which is unable to classify data with continuous values. In general, the FID3-AF is beneficial and productive for data classification.

ACKNOWLEDGEMENTS

The authors thank to Universiti Malaysia Pahang, United Arab Emirates University, Universiti Tun Hussein Onn, Universiti Teknologi Malaysia and Universitas Ahmad Dahlan for supporting this collaborative research in the present work. This work was funded by the Universiti Malaysia Pahang's Postgraduate Research Scheme (PGRS) vot No PGRS200397 and by Universiti Teknologi Malaysia under UTMER FASA 1/2020 (VOT NO 19J59).

REFERENCES

- [1] H.-B. B. Lee and W. Han, "Unique Features of Young Age Breast Cancer and Its Management," *J. Breast Cancer*, vol. 17, no. 4, p. 301, 2014, doi: 10.4048/jbc.2014.17.4.301.
- [2] T. Samec and R. Seetan, "Modeling and analyzing predictive monthly survival in females diagnosed with gynecological cancers," *International Journal of Public Health Science*, vol. 10, no. 4. Intelektual Pustaka Media Utama, pp. 888–897, 2021, doi: 10.11591/IJPHS.V10I4.20936.
- [3] E. J. Watkins, "Overview of breast cancer," *J. Am. Acad. PAs*, vol. 32, no. 10, pp. 13–17, 2019, doi: 10.1097/01.JAA.0000580524.95733.3d.
- [4] W. Majeed *et al.*, "Breast Cancer: Major Risk Factors and Recent Developments in Treatment," *Asian Pacific J. Cancer Prev.*, vol. 15, pp. 3353–3358, 2014, doi: <https://doi.org/10.7314/APJCP.2014.15.8.3353>.
- [5] S. Sree Kumar, A. K. Radhakrishnan, and S. K. Cheong, "Rapid Metastasis of Breast Cancer Cells from Primary Tumour to Liver," *Pakistan J. Biol. Sci.*, vol. 13, no. 7, pp. 303–315, 2010, doi: <https://doi.org/10.3923/pjbs.2010.303.315>.
- [6] Y. S. Sun *et al.*, "Risk factors and preventions of breast cancer," *Int. J. Biol. Sci.*, vol. 13, no. 11, pp. 1387–1397, 2017, doi: 10.7150/ijbs.21635.
- [7] S. Solikhah, K. N. A. Setyawati, and M. Sangruangake, "Lifestyle breast cancer patients among indonesian women: A nationwide survey," *International Journal of Public Health Science*, vol. 10, no. 4. Intelektual Pustaka Media Utama, pp. 730–734, 2021, doi: 10.11591/ijphs.v10i4.20913.
- [8] Akrom and R. Capritasari, "Clinical pharmacoepidemiology of antibiotics usage in intensive care unit of cancer special hospital," *International Journal of Public Health Science*, vol. 9, no. 3. Universitas Ahmad Dahlan - Faculty of Public Health, pp. 169–175, 2020, doi: 10.11591/ijphs.v9i3.20447.
- [9] N. P. W. P. Sari, "Social wellbeing among women living with cancer," *International Journal of Public Health Science*, vol. 9, no. 1. Universitas Ahmad Dahlan - Faculty of Public Health, pp. 62–70, 2020, doi: 10.11591/ijphs.v9i1.20414.
- [10] N. P. W. P. Sari, "Stress and fear in women living with cancer: An argumentation towards the adaptation theory," *International Journal of Public Health Science*, vol. 9, no. 3. Universitas Ahmad Dahlan - Faculty of Public Health, pp. 272–280, 2020, doi: 10.11591/ijphs.v9i3.20413.
- [11] N. Khuriwal and N. Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm," in *Proc. - IEEE 2018 Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2018*, 2018, pp. 98–103, doi: 10.1109/ICACCCN.2018.8748777.
- [12] T. S. Lim, K. G. Tay, A. Huong, and X. Y. Lim, "Breast cancer diagnosis system using hybrid support vector machine-artificial neural network," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4. Institute of Advanced Engineering and Science, pp. 3059–3069, 2021, doi: 10.11591/ijece.v11i4.pp3059-3069.

- [13] A. Ridok, N. Widodo, W. F. Mahmudy, and M. Rifa'i, "A hybrid feature selection on AIRS method for identifying breast cancer diseases," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1. Institute of Advanced Engineering and Science, pp. 728–735, 2021, doi: 10.11591/ijece.v11i1.pp728-735.
- [14] M. Y. Kamil, "Computer-aided diagnosis system for breast cancer based on the Gabor filter technique," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5. Institute of Advanced Engineering and Science, pp. 5235–5242, 2020, doi: 10.11591/IJECE.V10I5.PP5235-5242.
- [15] M. S. Croock, S. D. Khuder, A. E. Korial, and S. S. Mahmood, "Early detection of breast cancer using mammography images and software engineering process," *Telkonnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 4. Universitas Ahmad Dahlan, pp. 1784–1794, 2020, doi: 10.12928/TELKOMNIKA.V18I4.14718.
- [16] M. A. Kahya, "Classification enhancement of breast cancer histopathological image using penalized logistic regression," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 1. Institute of Advanced Engineering and Science, pp. 405–410, 2019, doi: 10.11591/ijeecs.v13.i1.pp405-410.
- [17] S. Harish and G. F. Ali Ahammed, "Integrated modelling approach for enhancing brain MRI with flexible pre-processing capability," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4. pp. 2416–2424, 2019, doi: 10.11591/ijece.v9i4.pp2416-2424.
- [18] J. A. Jupin, T. Sutikno, M. A. Ismail, M. S. Mohamad, and S. Kasim, "Review of the machine learning methods in the classification of phishing attack," *Bull. Electr. Eng. Informatics*, vol. 8, no. 4, 2019, doi: 10.11591/eei.v8i4.1922.
- [19] A. S. I. Hilaiwah, H. A. A. Abed Allah, B. A. Abbas, and T. Sutikno, "Live to learn: learning rules-based artificial neural network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 1, 2021, doi: 10.11591/ijeecs.v21.i1.pp558-565.
- [20] E. J. Baker *et al.*, "User identification system for inked fingerprint pattern based on central moments," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 2. Institute of Advanced Engineering and Science, pp. 1149–1160, 2021, doi: 10.11591/ijeecs.v24.i2.pp1149-1160.
- [21] I. S. Masad, A. Alqudah, A. M. Alqudah, and S. Almashaqbeh, "A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest x-ray images," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6. Institute of Advanced Engineering and Science, pp. 5530–5540, 2021, doi: 10.11591/ijece.v11i6.pp5530-5540.
- [22] N. H. Ali, A. R. Abdullah, N. M. Saad, A. S. Muda, T. Sutikno, and M. H. Jopri, "Brain stroke computed tomography images analysis using image processing: A review," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 4. Institute of Advanced Engineering and Science, pp. 1048–1059, 2021, doi: 10.11591/IJAI.V10.I4.PP1048-1059.
- [23] A. M. Abdulazeez, D. M. Hajy, D. Q. Zeebaree, and D. A. Zebari, "Robust watermarking scheme based LWT and SVD using artificial bee colony optimization," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 2, pp. 1218–1229, 2020, doi: 10.11591/ijeecs.v21.i2.pp1218-1229.
- [24] N. S. Nordin *et al.*, "A comparative analysis of metaheuristic algorithms in fuzzy modelling for phishing attack detection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 1146–1158, 2021, doi: 10.11591/ijeecs.v23.i2.pp1146-1158.
- [25] A. H. Ali, M. N. Abbod, M. K. Khaleel, M. A. Mohammed, and T. Sutikno, "Large scale data analysis using MLLib," *Telkonnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 5. pp. 1735–1746, 2021, doi: 10.12928/TELKOMNIKA.v19i5.21059.
- [26] R. Sahak, W. Mansor, K. Y. Lee, and A. Zabidi, "Performance of principal component analysis and orthogonal least square on optimized feature set in classifying asphyxiated infant cry using support vector machine," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 1. Institute of Advanced Engineering and Science, pp. 139–145, 2018, doi: 10.11591/ijeecs.v9.i1.pp139-145.
- [27] E. G. Nihad, E.-N. E. Mokhtar, Z. Abdelhamid, and A. A. Mohammed, "Hybrid approach of the fuzzy C-means and the K-nearest neighbors methods during the retrieve phase of dynamic case based reasoning for personalized follow-up of learners in real time," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 6. Institute of Advanced Engineering and Science, pp. 4939–4950, 2019, doi: 10.11591/ijece.v9i6.pp4939-4950.
- [28] L. Machap, A. Abdullah, and Z. A. Shah, "Co-clustering algorithm for the identification of cancer subtypes from gene expression data," *Telkonnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 4. Universitas Ahmad Dahlan, pp. 2017–2024, 2019, doi: 10.12928/TELKOMNIKA.V17I4.12773.
- [29] A. Selwal and I. Raof, "A Multi-layer perceptron based intelligent thyroid disease prediction system," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1. Institute of Advanced Engineering and Science, pp. 524–532, 2020, doi: 10.11591/ijeecs.v17.i1.pp524-532.
- [30] M. Alzaqebah, N. Alrefai, E. A. E. Ahmed, S. Jawarneh, and M. K. Alsmadi, "Neighborhood search methods with moth optimization algorithm as a wrapper method for feature selection problems," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4. Institute of Advanced Engineering and Science, pp. 3672–3684, 2020, doi: 10.11591/ijece.v10i4.pp3672-3684.
- [31] G. Saranya and A. Pravin, "A comprehensive study on disease risk predictions in machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4. Institute of Advanced Engineering and Science, pp. 4217–4225, 2020, doi: 10.11591/ijece.v10i4.pp4217-4225.
- [32] N. Omar, A. M. Abdulazeez, A. Sengur, and S. G. S. Al-Ali, "Fused faster RCNNs for efficient detection of the license plates," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 2. Institute of Advanced Engineering and Science, pp. 974–982, 2020, doi: 10.11591/ijeecs.v19.i2.pp974-982.
- [33] M. A. Basir, M. S. Hussin, and Y. Yusof, "Integrated bio-search approaches with multi-objective algorithms for optimization and classification problem," *Telkonnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 5. Universitas Ahmad Dahlan, pp. 2421–2431, 2020, doi: 10.12928/TELKOMNIKA.V18I5.15141.
- [34] R. Malhotra and A. Sharma, "Threshold benchmarking for feature ranking techniques," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2. Institute of Advanced Engineering and Science, pp. 1063–1070, 2021, doi: 10.11591/eei.v10i2.2752.
- [35] O. AlShorman, B. AlShorman, and F. Alkahtani, "A review of wearable sensors based monitoring with daily physical activity to manage type 2 diabetes," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1. Institute of Advanced Engineering and Science, pp. 646–653, 2021, doi: 10.11591/ijece.v11i1.pp646-653.
- [36] B. Mouaz, C. Walid, B.-H. Abderrahim, and E. Abdelmajid, "A new framework based on KNN and DT for speech identification through emphatic letters in Moroccan dialect," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3. Institute of Advanced Engineering and Science, pp. 1417–1423, 2021, doi: 10.11591/ijeecs.v21.i3.pp1417-1423.
- [37] A. A. Abbod, Q. M. Shallal, and M. A. Fadhel, "Automated brain tumor classification using various deep learning models: A comparative study," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 1. Institute of Advanced Engineering and Science, pp. 252–259, 2021, doi: 10.11591/ijeecs.v22.i1.pp252-259.
- [38] J. Jasmir, S. Nurmaini, R. F. Malik, and B. Tutuko, "Bigram feature extraction and conditional random fields model to improve




- text classification clinical trial document,” *Telkonnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 3. Universitas Ahmad Dahlan, pp. 886–892, 2021, doi: 10.12928/TELKOMNIKA.v19i3.18357.
- [39] H. A. Al-Jubouri and S. M. Mahmmod, “A comparative analysis of automatic deep neural networks for image retrieval,” *Telkonnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 3. Universitas Ahmad Dahlan, pp. 858–871, 2021, doi: 10.12928/TELKOMNIKA.v19i3.18157.
- [40] N. N. Moon *et al.*, “Natural language processing based advanced method of unnecessary video detection,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6. Institute of Advanced Engineering and Science, pp. 5411–5419, 2021, doi: 10.11591/ijece.v11i6.pp5411-5419.
- [41] S. J. Kuo, Y. H. Hsiao, Y. L. Huang, and D. R. Chen, “Classification of benign and malignant breast tumors using neural networks and three-dimensional power Doppler ultrasound,” *Ultrasound Obs. Gynecol.*, vol. 32, no. 1, pp. 97–102, 2008, doi: 10.1002/uog.4103.
- [42] N. Abdul Malik, M. S. Othman, and L. M. Yusuf, “Comparative Analysis of Classifiers for Education Case Study,” *Int. J. Softw. Eng. Comput. Syst.*, vol. 5, no. 1, pp. 67–76, 2019, doi: 10.15282/ijsecs.5.1.2019.5.0055.
- [43] H. R. Ibraheem, Z. F. Hussain, S. M. Ali, M. Aljanabi, M. A. Mohammed, and T. Sutikno, “A new model for large dataset dimensionality reduction based on TLBO and LR,” *Telkonnika (Telecommunication Comput. Electron. Control.)*, vol. 18, no. 3, pp. 1688–1694, 2020, doi: 10.12928/TELKOMNIKA.v18i3.13764.
- [44] M. E. Cintra, “Genetic generation of fuzzy knowledge bases: new perspectives,” 2012.
- [45] R. A. Ibrahim Alhayali, M. A. Ahmed, Y. M. Mohialden, and A. H. Ali, “Efficient method for breast cancer classification based on ensemble hoeffding tree and naïve Bayes,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 2, pp. 1074–1080, 2020, doi: 10.11591/ijeecs.v18.i2.pp1074-1080.
- [46] J. J. Patel and S. K. Hadia, “An enhancement of mammogram images for breast cancer classification using artificial neural networks,” *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, pp. 332–345, 2021, doi: 10.11591/ijai.v10.i2.pp332-345.
- [47] C. Aroef, Y. Rivan, and Z. Rustam, “Comparing random forest and support vector machines for breast cancer classification,” *Telkonnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 815–821, 2020, doi: 10.12928/TELKOMNIKA.V18I2.14785.
- [48] N. M. Ali, N. A. A. Aziz, and R. Besar, “Comparison of microarray breast cancer classification using support vector machine and logistic regression with LASSO and boruta feature selection,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 2. Institute of Advanced Engineering and Science, pp. 712–719, 2020, doi: 10.11591/ijeecs.v20.i2.pp712-719.
- [49] M. E. Cintra, H. A. Camargo, and T. Martin, “Optimising the fuzzy granulation of attribute domains,” *2009 Int. Fuzzy Syst. Assoc. World Congr. 2009 Eur. Soc. Fuzzy Log. Technol. Conf. IFSA-EUSFLAT 2009 - Proc.*, no. May 2014, pp. 742–747, 2009.
- [50] A. Hadioui, Y. B. Touimi, N.-E. El Faddouli, and S. Bennani, “Intelligent machine for ontological representation of massive pedagogical knowledge based on neural networks,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1675–1688, 2021, doi: 10.11591/ijece.v11i2.pp1675-1688.
- [51] W. Khairunizam *et al.*, “Analysis of attribute domain for geometrical gesture performed by arm movements,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 2. Institute of Advanced Engineering and Science, pp. 759–766, 2019, doi: 10.11591/ijeecs.v16.i2.pp759-766.
- [52] C. Neelima and S. S. V. N. Sarma, “Blended intelligence of FCA with FLC for knowledge representation from clustered data in medical analysis,” *International Journal of Electrical and Computer Engineering*, vol. 9, no. 1. Institute of Advanced Engineering and Science, pp. 635–645, 2019, doi: 10.11591/ijece.v9i1.pp.635-645.
- [53] N. R. Hamza, R. A. Dihin, and M. H. Abdulameer, “A hybrid image similarity measure based on a new combination of different similarity techniques,” *International Journal of Electrical and Computer Engineering*, vol. 10, no. 2. Institute of Advanced Engineering and Science, pp. 1814–1822, 2020, doi: 10.11591/ijece.v10i2.pp1814-1822.
- [54] D. Azzeddine, S. Jabri, B. Youssef, and G. Taoufiq, “The selection of the relevant association rules using the electre method with multiple criteria,” *IAES International Journal of Artificial Intelligence*, vol. 9, no. 4. Institute of Advanced Engineering and Science, pp. 638–645, 2020, doi: 10.11591/ijai.v9.i4.pp638-645.
- [55] S. Sivakumar, S. Venkataraman, and R. Selvaraj, “Predictive Modeling of Student Dropout Indicators in Educational Data Mining using Improved Decision Tree,” *Indian J. Sci. Technol.*, vol. 9, no. January, 2016, doi: 10.17485/ijst/2016/v9i4/87032.
- [56] N. F. Idris and M. A. Ismail, “Attribute Related Methods for Improvement of ID3 Algorithm in Classification of Data : A Review,” *Technol. Reports Kansai Univ.*, vol. 62, no. 08, pp. 4759–4767, 2020.
- [57] L. A. Zadeh, “Is there a need for fuzzy logic?,” *Inf. Sci. (Ny)*, vol. 178, no. 13, pp. 2751–2779, 2008, doi: 10.1016/j.ins.2008.02.012.
- [58] M. E. Cintra, M. Monard, and H. D. A. Camargo, “FuzzyDT- A Fuzzy Decision Tree Algorithm Based on C4.5,” in *Conf. Proc. 2nd Brazilian Congr. Fuzzy Syst.*, 2013, no. January, pp. 200–211.
- [59] X. Dai, L. Gao, and C. Dong, “Self-Adaptive Fuzzification in Fuzzy Decision Tree Induction,” *2010 Int. Conf. Mach. Learn. Cybern.*, vol. 1, pp. 296–301, 2010, doi: 10.1109/ICMLC.2010.5581048.
- [60] S. Thaker and V. Nagori, “Analysis of Fuzzification Process in Fuzzy Expert System,” in *Int. Conf. Comput. Intell. Data Sci. (ICCIDS 2018)*, 2018, vol. 132, pp. 1308–1316, doi: 10.1016/j.procs.2018.05.047.
- [61] M. E. Cintra, M. C. Monard, H. A. Camargo, T. P. Martin, and A. Majidian, “On Rule Generation Approaches for Genetic Fuzzy Systems,” in *Congr. da Soc. Bras. Comput.*, 2011, no. Im.
- [62] M. E. Cintra and H. De Arruda Camargo, “Fuzzy rules generation using genetic algorithms with self-adaptive selection,” in *2007 IEEE Int. Conf. Inf. Reuse Integr. IEEE IRI-2007*, 2007, pp. 261–266, doi: 10.1109/IRI.2007.4296631.
- [63] C. A. Peña-Reyes and M. Sipper, “Fuzzy CoCo: A cooperative-coevolutionary approach to fuzzy modeling,” *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 5, pp. 727–737, 2001, doi: 10.1109/91.963759.
- [64] L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” *Syst. Man Cybern. IEEE Trans.*, no. 1, pp. 28–44, 1973, doi: 10.1109/TSMC.1973.5408575.
- [65] N. Saad and O. Wahyunggoro, *Development of Fuzzy-Logic-Based Self Tuning PI Controller for Servomotor*. United Kingdom: Intechopen, 2010.
- [66] M. A. S. Masoum and E. F. Fuchs, *Optimal Placement and Sizing of Shunt Capacitor Banks in the Presence of Harmonics*. United Kingdom: Elsevier Inc., 2008.
- [67] P. Surya Prasad, V. Tarun Varma, V. Sai, V. Harish, and K. S. R. Kumar, “Classification of Different Textures Using SVM and Fuzzy logic,” *Int. J. Adv. Comput. Res.*, no. 4, pp. 4–7, 2012.
- [68] S. K. Ali and W. K. Mutlag, “Early detection for breast cancer by using fuzzy logic,” *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 17,

- pp. 5717–5728, 2018.
- [69] F. T. Johra and M. M. H. Shuvo, “Detection of breast cancer from histopathology image and classifying benign and malignant state using fuzzy logic,” *2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEICT 2016*, no. 1, pp. 2–6, 2017, doi: 10.1109/CEEICT.2016.7873137.
- [70] G. Schaefer, T. Nakashima, M. Závisek, Y. Yokota, A. Drastich, and H. Ishibuchi, “Breast cancer classification using statistical features and fuzzy classification of thermograms,” *IEEE Int. Conf. Fuzzy Syst.*, 2007, doi: 10.1109/FUZZY.2007.4295520.
- [71] M. M. Saad, N. Jamil, and R. Hamzah, “Evaluation of support vector machine and decision tree for emotion recognition of malay folklores,” *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 3. Institute of Advanced Engineering and Science, pp. 479–486, 2018, doi: 10.11591/eei.v7i3.1279.
- [72] F. Jauhari and A. A. Supianto, “Building student’s performance decision tree classifier using boosting algorithm,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3. Institute of Advanced Engineering and Science, pp. 1298–1304, 2019, doi: 10.11591/ijeecs.v14.i3.pp1298-1304.
- [73] T. A. Assegie and P. S. Nair, “Handwritten digits recognition with decision tree classification: A machine learning approach,” *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5. Institute of Advanced Engineering and Science, pp. 4446–4451, 2019, doi: 10.11591/ijece.v9i5.pp4446-4451.
- [74] E. B. B. Palad, M. J. F. Burden, C. R. Dela Torre, and R. B. C. Uy, “Performance evaluation of decision tree classification algorithms using fraud datasets,” *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 6. Institute of Advanced Engineering and Science, pp. 2518–2525, 2020, doi: 10.11591/eei.v9i6.2630.
- [75] N. Jahan and R. Shahariar, “Predicting fertilizer treatment of maize using decision tree algorithm,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3. Institute of Advanced Engineering and Science, pp. 1427–1434, 2020, doi: 10.11591/ijeecs.v20.i3.pp1427-1434.
- [76] M. H. B. Mohd Razali, R. B. Saian, Y. B. Wah, and K. R. Ku-Mahamud, “A class skew-insensitive ACO-based decision tree algorithm for imbalanced data sets,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 1. Institute of Advanced Engineering and Science, pp. 412–419, 2021, doi: 10.11591/ijeecs.v21.i1.pp412-419.
- [77] S. B. Begeenova and T. V. Avdeenko, “Building of fuzzy decision trees using ID3 algorithm,” *J. Phys. Conf. Ser.*, vol. 1015, no. 2, 2018, doi: 10.1088/1742-6596/1015/2/022002.
- [78] U. Sa’Adah, M. Y. Rochayani, and A. B. Astuti, “Knowledge discovery from gene expression dataset using bagging lasso decision tree,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 2. Institute of Advanced Engineering and Science, pp. 1151–1159, 2020, doi: 10.11591/ijeecs.v21.i2.pp1151-1159.
- [79] N. Rachburee and W. Punlumjeak, “Oversampling technique in student performance classification from engineering course,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4. Institute of Advanced Engineering and Science, pp. 3567–3574, 2021, doi: 10.11591/ijece.v11i4.pp3567-3574.
- [80] Q. Liu and Y. Wang, “Improved ID3 algorithm using ontology in computer forensics,” in *ICCASM 2010 - 2010 Int. Conf. Comput. Appl. Syst. Model. Proc.*, 2010, vol. 11, no. Iccasm, pp. 494–497, doi: 10.1109/ICCASM.2010.5623159.
- [81] H. Luo, Y. Chen, and W. Zhang, “An improved ID3 algorithm based on attribute importance-weighted,” in *2010 2nd Int. Work. Database Technol. Appl.*, 2010, pp. 1–4, doi: 10.1109/dbta.2010.5659010.
- [82] J. Chen, D. Luo, and F. Mu, “An Improved ID3 Decision Tree Algorithm,” in *Proc. 2009 4th Int. Conf. Comput. Sci. Educ.*, 2009, pp. 127–130, doi: 10.1109/iccse.2009.5228509.
- [83] S. Wu, L. Wu, Y. Long, and X. D. Gao, “Improved Classification Algorithm by Minsup and Minconf Based on ID3,” in *2006 Int. Conf. Manag. Sci. Eng.*, 2006, pp. 135–139, doi: 10.1109/ICMSE.2006.313896.
- [84] V. K. Nijhawan, M. Madan, and M. Dave, “The Analytical Comparison of ID3 and C4 . 5 using WEKA,” *Int. J. Comput. Appl.*, vol. 167, pp. 1–4, 2017, doi: 10.5120/ijca2017914286.
- [85] Y. Liu and N. Xie, “Improved ID3 algorithm,” in *2010 3rd Int. Conf. Comput. Sci. Inf. Technol.*, 2010, no. 4, pp. 465–468, doi: 10.1109/iccst.2010.5564765.
- [86] S. Teli and P. Kanikar, “A Survey on Decision Tree Based Approaches in Data Mining,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 4, pp. 613–617, 2015.
- [87] A. S. P. Angayarkanni and N. B. Kamal, “MRI mammogram image classification using ID3 algorithm,” *IET Conf. Publ.*, vol. 2012, no. 600 CP, pp. 1–5, 2012, doi: 10.1049/cp.2012.0464.
- [88] S. Yang, J. Z. Guo, and J. W. Jin, “An improved ID3 algorithm for medical data classification,” *Comput. Electr. Eng.*, vol. 65, pp. 474–487, 2018, doi: 10.1016/j.compeleceng.2017.08.005.
- [89] S. G. Jacob and R. Geetha Ramani, “Efficient classifier for classification of prognostic breast cancer data through data mining techniques,” *Lect. Notes Eng. Comput. Sci.*, vol. 1, pp. 493–498, 2012.
- [90] R. Chai and M. Wang, “A more efficient classification scheme for ID3,” in *2010 2nd Int. Conf. Comput. Eng. Technol.*, 2010, vol. 1, pp. 329–332.
- [91] M. E. Cintra, M. C. Monard, E. A. Cherman, and H. De Arruda Camargo, “On the estimation of the number of fuzzy sets for fuzzy rule-based classification systems,” in *Proc. 2011 11th Int. Conf. Hybrid Intell. Syst. HIS 2011*, 2011, no. December, pp. 211–216, doi: 10.1109/HIS.2011.6122107.
- [92] M. E. Cintra, T. P. Martin, M. C. Monard, and H. de A. Camargo, “An Approach for the Extraction of Classification Rules from Fuzzy Formal Contexts,” 2011. [Online]. Available: https://web.icmc.usp.br/SCATUSU/RT/BIBLIOTECA_113_RT_368.pdf.
- [93] L. X. Wang and J. M. Mendel, “Generating Fuzzy Rules by Learning from Examples,” *IEEE Trans. Syst. Man Cybern.*, vol. 22, no. 6, pp. 1414–1427, 1992, doi: 10.1109/21.199466.
- [94] M. E. Cintra, J. Francisco, F. Ribeiro, and O. A. Neves, “A Fuzzy Decision Tree Model to Support the Task of Bus Reallocation in Public Transport Systems,” *J. Inf. Data Manag.*, vol. 5, no. 1, pp. 104–113, 2014.
- [95] M. E. Cintra and O. A. Neves, “A Fuzzy Decision Tree for Bus Network Management,” in *Knowl. Discov. Min. Learn. – KDMiLe*, 2013, pp. 1–8.
- [96] M. E. Cintra, C. A. A. Meira, M. C. Monard, H. A. Camargo, and L. H. A. Rodrigues, “The use of fuzzy decision trees for coffee rust warning in Brazilian crops,” in *2011 11th Int. Conf. Intell. Syst. Des. Appl.*, 2011, pp. 1347–1352.
- [97] P. A. M. Bhadgale and M. S. Natu, “Implementation of Improved ID3 Algorithm Based on Association Function,” *Int. J. Pure Appl. Math.*, vol. 114, no. 10, pp. 1–9, 2017.
- [98] P. Kumar, K. I. Varma, and A. Sureka, “Fuzzy based clustering algorithm for privacy preserving data mining,” *Int. J. Bus. Inf. Syst.*, vol. 7, no. 1, pp. 27–40, 2011, doi: 10.1504/IJBIS.2011.037295.
- [99] A. Patri, A. Nayak, and A. Anurag, “High accuracy back-retreat diffusion-fuzzy clustering of breast cancer data for the detection of malignancy,” in *Proc. IASTED Int. Conf. Biomed. Eng. BioMed 2014*, 2014, no. June, pp. 7–15, doi: 10.2316/P.2014.818-018.
- [100] H. Saoud, A. Ghadi, M. Ghailani, and B. A. Abdelhakim, *Using Feature Selection Techniques to Improve the Accuracy of Breast Cancer Classification*, no. January. Springer International Publishing, 2019.




- [101] M. Kumari, V. Singh, and P. Ahlawat, "Decision Support System for Breast Cancer Prediction," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 7, no. 3, pp. 816–821, 2019, doi: 10.22214/ijraset.2019.3142.
- [102] S. Aruna, S. P. Rajagopalan, and L. V Nandakishore, "Knowledge Based Analysis of Various Statistical Tools in Detecting Breast Cancer," *Comput. Sci. Inf. Technol.*, vol. 2, pp. 37–45, 2011, doi: 10.5121/csit.2011.1205.
- [103] T. Sridevi, K. Shyamala, and A. Murugan, "An innovative algorithm for feature selecton based on rough set with fuzzy C-means clustering," *J. Theor. Appl. Inf. Technol.*, vol. 68, no. 3, pp. 514–522, 2014.
- [104] S. Jeyasingh and M. Veluchamy, "Modified bat algorithm for feature selection with the Wisconsin Diagnosis Breast Cancer (WDBC) dataset," *Asian Pacific J. Cancer Prev.*, vol. 18, no. 5, pp. 1257–1264, 2017, doi: 10.22034/APJCP.2017.18.5.1257.
- [105] R. O. Badiang, B. D. Gerardo, and R. P. Medina, "Relocating local outliers produced by K-means and K-medoids using local outlier rectifier V.2.0," in *2019 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2019*, 2019, pp. 89–94, doi: 10.1109/ICACSIS47736.2019.8979741.
- [106] Y. Li and Z. Chen, "Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction," *Appl. Comput. Math.*, vol. 7, no. 4, p. 212, 2018, doi: 10.11648/j.acm.20180704.15.
- [107] F. Kayaalp and M. S. Basarslan, "Performance Analysis Of Filter Based Feature Selection Methods On Diagnosis Of Breast Cancer And Orthopedics," *6th Int. Congr. Fundam. Appl. Sci. 2019*, no. July, pp. 1–11, 2019.
- [108] Z. Sahri, F. Arif, S. Sakinah, S. Ahmad, R. Ahmad, and R. Yusof, "Imputing Missing Values in Mammography Mass Dataset: Will it Increase Classification Performance of Machine Learning Algorithms?," pp. 8–12, 2017, doi: 10.17758/eirai.iae1017305.
- [109] B. Lairenjam and S. Krishan Wasan, "A note on analysis of Mammography Data," *Int. J. Open Probl. Compt Math.*, vol. 3, no. 5, pp. 39–48, 2010, [Online]. Available: www.i-csrs.org.
- [110] S. A. Mokhtar and A. M. Elsayad, "Predicting the Severity of Breast Masses with Data Mining Methods," 2013, [Online]. Available: <http://arxiv.org/abs/1305.7057>.

BIOGRAPHIES OF AUTHORS






Nur Farahaina Idris    received the B.SC. degree in computer science from Universiti Malaysia (UMP). She is currently pursuing a Master Degree in Universiti Malaysia Pahang (UMP). Sde can be contacted at email: farahainaidris@gmail.com.






Mohd Arfian Ismail    is a Senior Lecturer at the Faculty of Computer Systems and Software Engineering in University Malaysia Pahang, Malaysia. He received B. Sc, M. Sc. and PhD degree in Computer Science from Universiti Teknologi Malaysia (UTM) in 2008, 2011 and 2016, respectively. His current research interests are in the areas of machine learning methods and optimization method. He can be contacted at email: arfian@ump.edu.my.






Mohd Saberi Mohamad    is a Professor of Bioinformatics and Artificial Intelligence at Department of Genetics and Genomics, College of Medicine and Health Sciences, UAEU. He received the BSc and MSc degrees in computer science from the Universiti Teknologi Malaysia, in 2002 and 2005, respectively. He received the PhD degree in intelligent systems for bioinformatics from Osaka Prefecture University, in 2010. He has published more than 300 publications in the field of bioinformatics and artificial intelligence. He is Director of Health Data Science Lab of the UAEU. His research areas are cancer bioinformatics, metabolic engineering, computational genomic and proteomic, evolutionary computation, and learning systems. He can be contacted at email: saberi@uaeu.ac.ae.






Shahreen Kasim    is currently an Associate Professor in Department of Security Information and Web Technology, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia. Her areas of interest include bioinformatics, soft computing, data mining, web and mobile application. She can be contacted at email: shahreen@uthm.edu.my.



Zalmyah Zakaria    is a Senior Lecturer at School of Computing, Universiti Teknologi Malaysia. She received B.Sc., M.Sc. and PhD degrees in Computer Science from Universiti Teknologi Malaysia, in 1999, 2001 and 2009, respectively. Her research interests focus on timetabling/scheduling, cloud computing, particle swarm optimisation, quality of service, search problems, simulated annealing, Taguchi methods, ant colony optimisation, biometrics, digital signatures, face recognition, genetic algorithms, public key cryptography, resource allocation and machine learning. She can be contacted at email: zalmyah@utm.my.



Tole Sutikno    is a lecturer in Electrical Engineering Department at the Universitas Ahmad Dahlan (UAD), Yogyakarta, Indonesia. He received his B.Eng., M.Eng. and Ph.D. degrees in Electrical Engineering from Universitas Diponegoro, Universitas Gadjah Mada and Universiti Teknologi Malaysia, in 1999, 2004 and 2016, respectively. He has been an Associate Professor in UAD, Yogyakarta, Indonesia since 2008. He is currently an Editor-in-Chief of the TELKOMNIKA and the Head of the Embedded Systems and Power Electronics Research Group. His research interests include the field of digital design, industrial applications, industrial electronics, industrial informatics, power electronics, motor drives, renewable energy, FPGA applications, embedded system, artificial intelligence, intelligent control, information technology and digital library. He can be contacted at email: tole@ee.uad.ac.id.