International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Hybrid Geno-Fuzzy Classifier for Breast Cancer Diagnostics

Julaika Begum K.[1], Dr. Sindhu J. Kumaar[2*]

Submitted: 16/04/2024 **Revised**: 13/05/2024 **Accepted**:18/06/2024

Abstract: Although it is more common in women, breast cancer may also strike males, and it has overtaken lung cancer as the second most lethal form of disease in India. Breast cancer has been diagnosed in a growing number of individuals throughout the world for over two decades. Early detection of breast cancer not only lowers the likelihood of disease-causing mortality but also brings the total cost of treatment for the disease down dramatically. The purpose of the Adaptive Geno-Fuzzy Logic Algorithm model is to make an accurate prediction of breast cancer, which will assist medical professionals in correctly detecting and categorising the lumps that are felt in the breast. The model is made up of two different modules: one that selects breast cancer features using rough sets, and another that classifies patients based on fuzzy rules. The application of the adaptive genetic algorithm allows for the optimization of the rules that are created by fuzzy classifiers. As the initial step of the process, rough sets theory is used to determine significant factors that influence breast cancer. In the second step, breast cancer is predicted with the use of the AGFLA classifier. The experimentation is carried out using the datasets that are made accessible to the public by the Wisconsin Breast Cancer Diagnostic (WBCD). Rough set theory is helpful to the model because it reveals structural linkages within data that is imprecise and noisy. This is true for any dataset, but it is especially true for the medical dataset WBCD, which also suffers from noise. The HGFC method has been shown, via experimental study, to perform better than existing detection and classification methods.

Keywords: *Genetic Algorithm, Fuzzy Logic, Rough Set Theory, WBCD Dataset, Breast Cancer*

1. Introduction

In the recent decade, many people have lost their lives to Breast Cancer, lives that could have been saved very easily, if they were diagnosed earlier. Today the medical field has evolved a lot, especially the diagnostic practices have become highly advanced that could provide valuable insight much earlier that before. However, the medical practitioners take significant high lot of doing the manual process of identify the correct parameters to determine whether a particular patient is diagnosed with breast cancer or not. From a mathematical standpoint, utilizing Fuzzy logic is the go-to solution for these types of optimization problems.

[1]*Research Scholar, Dept. of Mathematics & AS, B.S. Abdur Rahman Crescent Institute Of Science And Technology, Chennai, Tamil Nadu, India. julaika_maths@crescent.education*
ORCID ID : 0009-0000-0563-5548
[2*]*Corresponding Author; Professor, Dept. of Mathematics & AS, B.S. Abdur Rahman Crescent Institute Of Science And Technology, Chennai, Tamil Nadu, India. sindhu@crescent.education*
ORCID ID : 0009-0004-4868-9705
**Corresponding Author Email: sindhu@crescent.education*

A combination of fully logic along with Genetic Algorithm can even more improve the efficiency by restricting the number of variables that go into the fuzzy classifier as inputs. Genetic Algorithms make use of fitness function to chose from a really large number of parameters, the most contributing parameters.

The work has been structure as detailed, an overview of the related works in the field are cited in section 2, section 3 gives a detailed description on the background, followed by the proposed HGFC classifier and experimental results in section 4 and 5 respectively, and at last section 6 presents conclusion along with possible future studies.

2. Related Works

Dimensionality reduction is one very important factor influencing the effectiveness of the classifier. Long et al [1] in their work have successfully demonstrated the use of rough set theory-based attribute / dimension reduction and fuzzy logic to identify heart diseases. There has been a spike in the number of mathematicians and data scientists along with computer scientists are increasingly making use of nature inspired algorithms to solve problems in the field of medical diagnostics, one such problem is

detailed in Ibrahim et al`s work [2] uses Dragon Fly feature selection. Support Vector Machines are also well utilized in the field of computer aided diagnostics, Anuradha`s work [3] and Egwom et al`s [4] work use SVM to demonstrate the classification efficiency of SVMs. To forecast cardiac illness, Santhanam et al [5] presented a hybrid Geno-Fuzzy model. To pick features, genetic algorithms were utilised. Using fuzzy inference, a categorization model was constructed, with the help of the selected features. The proposed model is designed based on analysis of recent works and considering their shortcoming.

3. Background

Pawlak introduced rough sets in 1982 [6]. Rough sets can reveal structural relationships in noisy data. Classical sets have complementary generalizations: rough and fuzzy sets. Rough set theory approximation spaces have many membership, while fuzzy sets have partial membership [7]. Lotfi A. Zadeh's "soft computing" relies on these two approaches' quick development. Soft computing covers a wide range of research areas, not limited to fuzzy logic, neural networks, and genetic algorithms. It also includes areas such as probabilistic reasoning, Artificial Neural Networks (ANN), belief networks, machine learning, and rough sets.

Rough Set solves the following data analysis issues, Attribute-based object classification, Attribute dependence, Elimination of extras, identifying key traits, decision-rule generation.

Data model information is critical for the mathematical model to be efficient, Rough Set Theory stores them in form of tables. Each row (tuple) represents a single fact or item. Frequently, the facts contradict one another. The phrase for a data table in Rough Set terminology is Information System (Input Data). Tables may include multiple objects with identical characteristics. It is also essential to reduce the size of the table to improve the efficiency, an easier and effective approach is to reduce the table size. In general, it is done by removing duplicate entries and making sure only one representative object is stored for each collection with objects that are either similar or with somewhat identical characteristics. This category of things is known as indiscernible objects or tuples.

An associated equivalence relation, as mentioned in Eq. 1, $IND(Q)$ for any Q subset A:

$$IND(Q) = \{(x, y) \in \mathbb{U}^2 | \forall a \in Q, a(x) = a(y)\} \quad \text{Eq. 1}$$

$IND(Q)$ is known as relational indiscernibility. Here, x and y cannot be distinguished by attribute Q.

The relation of indiscernibility is an equivalence relation. Indiscernible sets are known as elementary sets. It is a formal approximation of a crisp set, defined by its Upper and Lower approximations [8].

An Upper approximation refers to the collection of objects that may or may not belong to the target set.

$$\bar{R}A = \bigcup(B \in U | R: B \cap A \neq \emptyset) \quad \text{Eq. 2}$$

The collection of items that positively belong to the target set is the lower approximation.

$$\underline{R}A = \bigcup(B \in U | R: B \subseteq A) \quad \text{Eq. 3}$$

4. Proposed Classifier

This research presents a novel classifier for breast cancer diagnostics, a classification problem based on the Wisconsin-Breast Cancer (Diagnostic) Dataset [9]. The proposed classifier operates in three stages: 1. Feature / dimension reduction using rough set theory; 2. generation of rules using Fuzzy Logic Classifier (source: reduced dataset); and 3. optimization of created rules using Adaptive GA.

The AGA applies a mathematical function typically represented as $f(x)$, know as the fitness function to optimize rules generated by the Fuzzy Logic Classifier. The following are the principal contributions of the suggested model: using set theory to determine the most important characteristics as demonstrated in the results section, Rough Set theory has proven to be an effective mechanism for dealing with imprecise, uncertain and noisy information in order to select the most pertinent attribute for a decision making / decision support system and Adaptive GAs are used to optimize the classification rules in order to improve accuracy and impact time complexity positively.

4.1 Adaptive Genetic Algorithm

A popular way to use soft computing is the genetic algorithm. Many ways are suggested to improve canonical GAs. Such a technique is Adaptive GA [10]. It follows the below mentioned steps.

1. How Chromosomes Are Made
2. Figuring out the Fitness function (Mathematical Function)
3. Crossover
4. Adaptive Mutation
5. Selection

When the fuzzy logic classifier is trying to find the

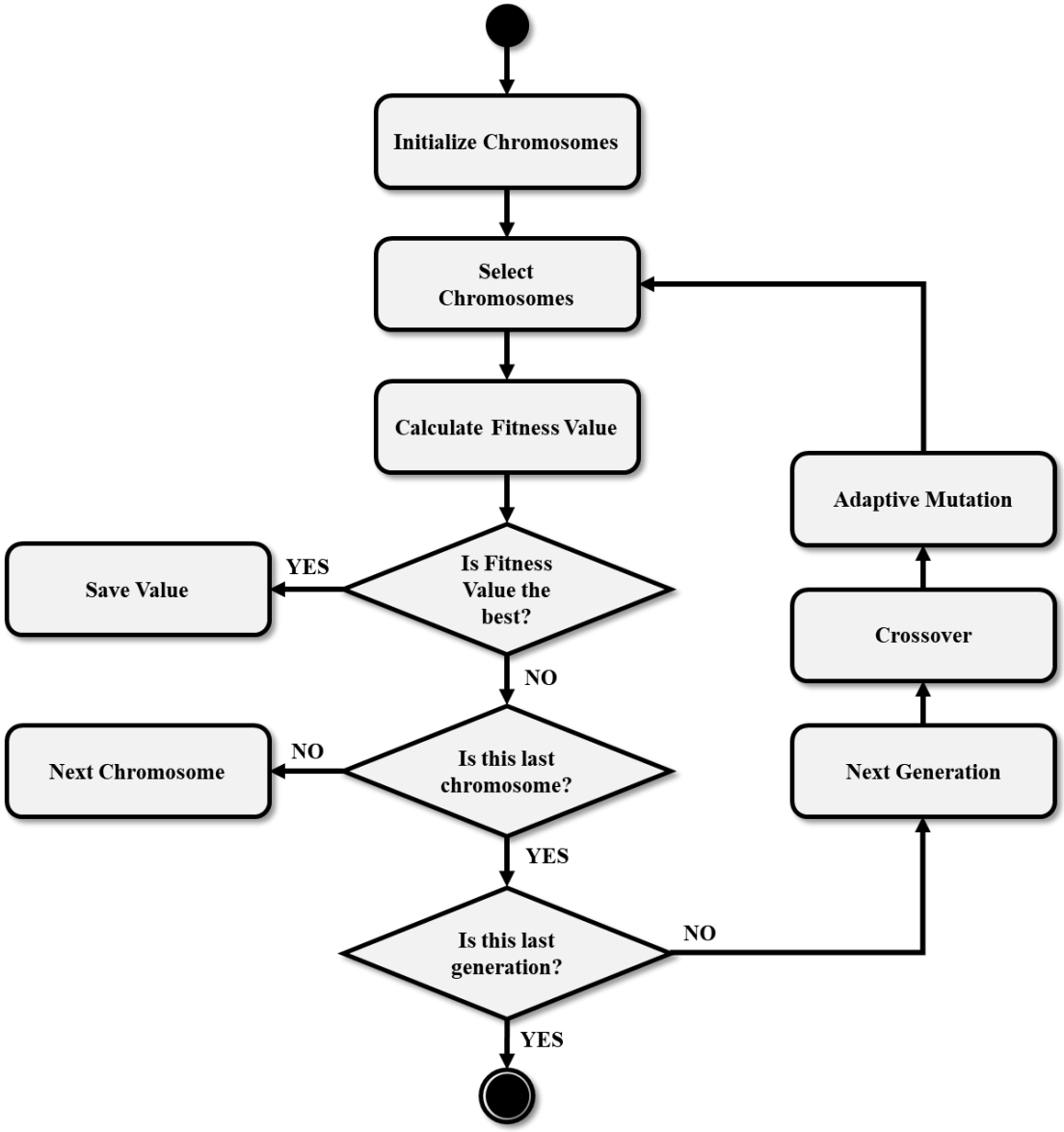

Figure 1 Adaptive GA – Flowchart

best rules, each rule is similar to a chromosome. The pool of chromosomes are made by chance, and each chromosome goes through all AGA's operations. The fitness value is used to judge the chromosomes, and then those chromosomes are sent to the output. Crossover and mutation are two of the most important steps in a genetic algorithm.

The significance of reducing the number of features lowers the cost of computing and improves the performance of classification. In this paper, rough sets are used to reduce the number of features, and the fuzzy classifier is employed to make a set of rules.

The Adaptive GA is used to get the best rules for predicting disease from the set of solutions. The steps involved in the proposed model are: Attribute reduction (rough set theory based), normalization, and HGFC classification.

The range [0, 1] is set for Normalizing the input dataset.

For choosing the best attributes, a rough set-based method is used. The fewer attributes will be broken up into two subsets: the testing data and the training data. HGFC is fed the training data, and the testing data is used to evaluate the proposed model. Figure 2 shows how the proposed method for predicting heart disease would work.

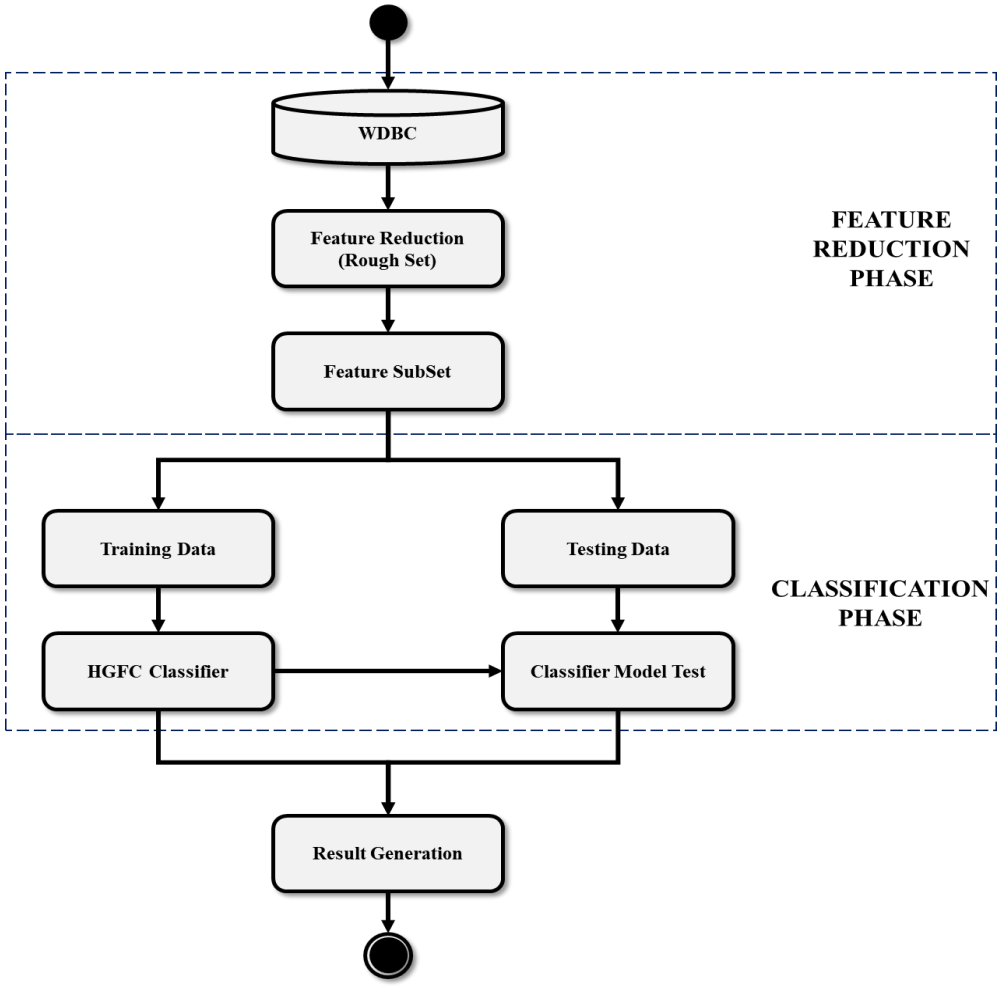

Figure 2 Workflow of Proposed HGFC

Algorithm 1: Pseudocode of the Proposed HGFC Model

INPUT: *Wisconsin Diagnostic Breast Cancer Dataset*

OUTPUT*: Optimized Fuzzy Logic Classifier with Optimized Classification Rules*

1. Use rough set theory to extract features from the input datasets.
2. Feed the retrieved features to the Fuzzy Logic Classifier in order to train the model and generate classification rules

(a) Fuzzification: Convert the crisp data into fuzzy data

(b) Generate fuzzy rules based on the fuzzy data.

If A1 is high, A2 is low, and A3 is medium, then the class is C2 class

(c)Defuzzification: Transform fuzzy rules into clear ones

3. Optimize the classification rules by applying Adaptive Genetic Algorithm to the model established in Step 2 using the model.
4. Cross-validate the model utilising test data. The model is evaluated based on its precision, specificity, and sensitivity.
5. Validate the results by employing statistical tests (specify the tests).

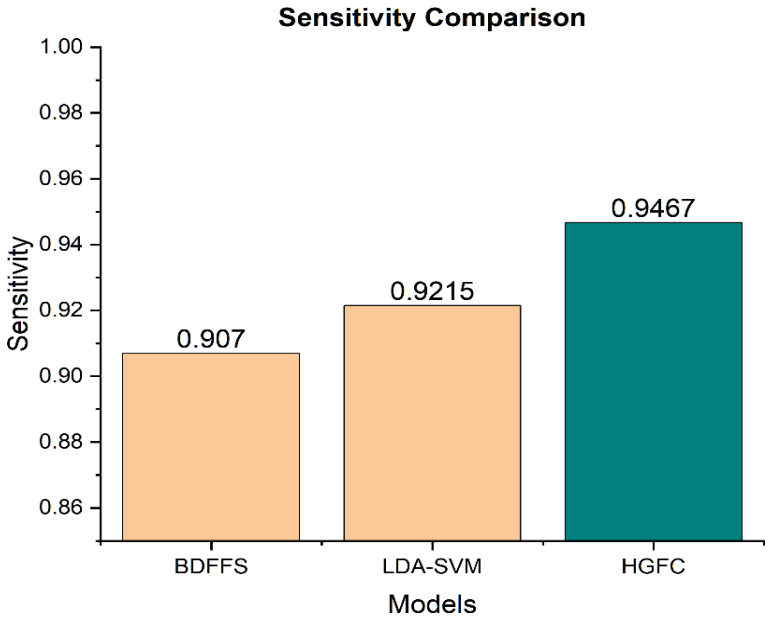

Figure 3 Sensitivity – BDFFS Vs. LDA-SVM Vs. HGFC

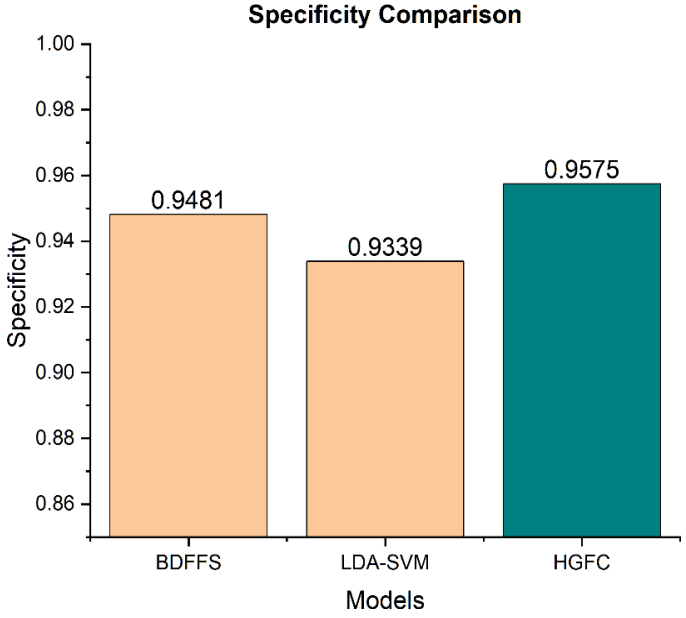

Figure 4 Specificity – BDFFS Vs. LDA-SVM Vs. HGFC

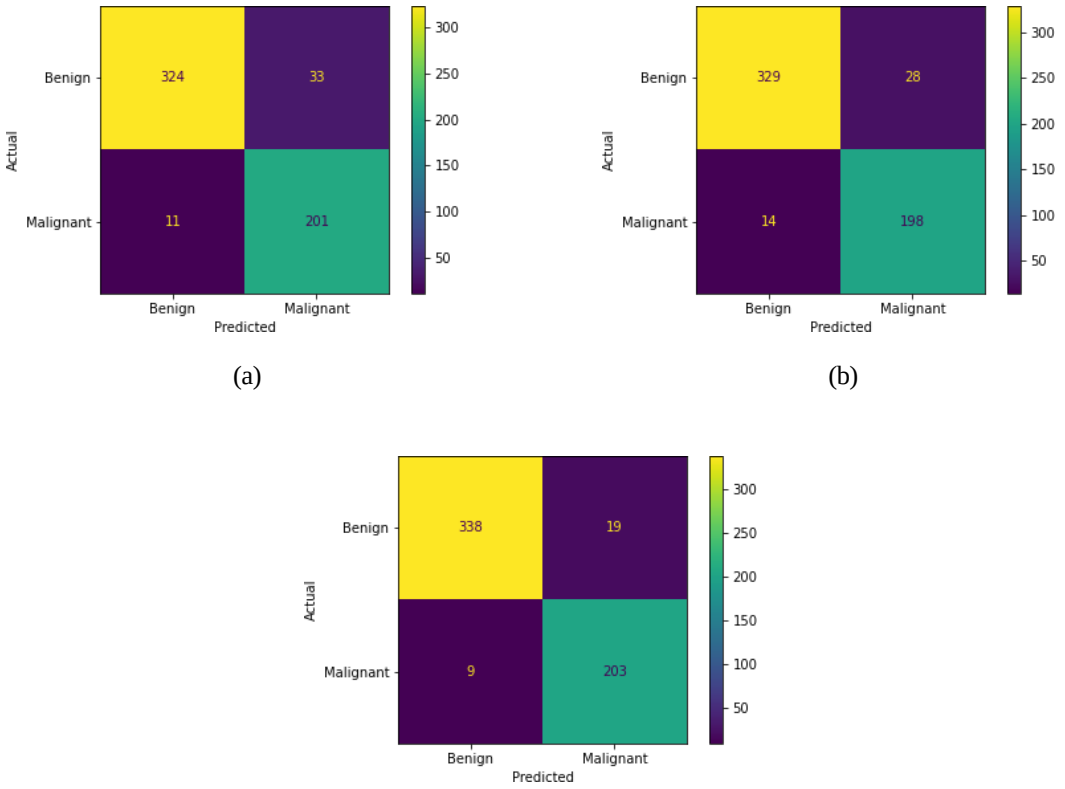

Figure 5 Confusion matrix (a) BDFFS (b) LDA-SVM and (c) HGFC

5. Experimental Results

The experiments were carried out and compared with the performance of BDFFS model proposed by Ibrahim et al [2] and LDA-SVM model proposed by Egwom et al [4]. The comparison indicated the superior performance of the proposed HGFC model. The parameters considered for comparison are as follows,

Sensitivity: The ratio of number of true positives and all positives Figure 3, Specificity: The ratio of number of true negatives and all negatives Figure 4, and Accuracy: The number of correct assessments and all assessments Figure 5 [11].

6. Conclusion

This paper proposes a unique method for classifying breast cancer utilising Rough Set theory and Fuzzy logic-based classification using an Adaptive GA. This paper proposes a categorization model, namely HGFC as follows: Initially, feature / dimension reduction is performed using rough sets theory. HGFC then performs disease forecasting. Using Adaptive Genetic Algorithm, the created rules are optimised. The experiment is conducted using the WDBC dataset. In terms of sensitivity, specificity and accuracy, HGFC fared successful than other models, according to the total experimental analysis. Major characteristics of the proposed model include its ability to handle noisy data efficiently and its ability to function effectively even with many attributes. In addition, the suggested model prevents local optimum capture.

References

[1] N. C. Long, P. Meesad, and H. Unger, "A highly accurate firefly based algorithm for heart disease prediction," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8221–8231, Nov. 2015, doi: 10.1016/j.eswa.2015.06.024.

[2] M. M. Ibrahim, D. A. Salem, and R. A. A. A. A. Seoud, "Deep Learning Hybrid with Binary Dragonfly Feature Selection for the Wisconsin Breast Cancer Dataset," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 12, no. 3, Art. no. 3, Jul. 2021, doi: 10.14569/IJACSA.2021.0120314.

[3] R. Anuradha, "Support Vector Machine Classifier for Prediction of Breast Malignancy using Wisconsin Breast Cancer Dataset," *Asian J. Converg. Technol. AJCT ISSN -2350-1146*, vol. 7,

no. 3, Art. no. 3, Dec. 2021, doi: 10.33130/AJCT.2021v07i03.010.

[4] O. J. Egwom, M. Hassan, J. J. Tanimu, M. Hamada, and O. M. Ogar, "An LDA–SVM Machine Learning Model for Breast Cancer Classification," *BioMedInformatics*, vol. 2, no. 3, Art. no. 3, Sep. 2022, doi: 10.3390/biomedinformatics2030022.

[5] T. Santhanam and E. Ep, "Heart Disease Prediction Using Hybrid Genetic Fuzzy Model," *Indian J. Sci. Technol.*, vol. 8, p. 797, May 2015, doi: 10.17485/ijst/2015/v8i9/52930.

[6] Z. Pawlak and R. Sowinski, "Rough set approach to multi-attribute decision analysis," *Eur. J. Oper. Res.*, vol. 72, no. 3, pp. 443–459, Feb. 1994, doi: 10.1016/0377-2217(94)90415-4.

[7] "Semantics of Fuzzy Sets in Rough Set Theory | SpringerLink." https://link.springer.com/chapter/10.1007/978-3-540-27778-1_15 (accessed Dec. 26, 2022).

[8] Q. Zhang, Q. Xie, and G. Wang, "A survey on rough set theory and its applications," *CAAI Trans. Intell. Technol.*, vol. 1, no. 4, pp. 323–333, Oct. 2016, doi: 10.1016/j.trit.2016.11.001.

[9] Dr. William H. Wolberg, W. Nick Street, and Olvi L. Mangasarian, "UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set." https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29 (accessed Dec. 25, 2022).

[10] C. Lin, "An Adaptive Genetic Algorithm Based on Population Diversity Strategy," in *2009 Third International Conference on Genetic and Evolutionary Computing*, Oct. 2009, pp. 93–96. doi: 10.1109/WGEC.2009.67.

[11] A. Tharwat, "Classification assessment methods," *Appl. Comput. Inform.*, vol. 17, no. 1, pp. 168–192, Jan. 2020, doi: 10.1016/j.aci.2018.08.003.