Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

Genetic Algorithm Based Fuzzy System Optimization for Breast Cancer Detection

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

[1] *Research Scholar, Dept. of Mathematics & AS, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India*

[2] *Professor, Dept. of Mathematics & AS, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India*

Abstract:- Fuzzy logic is a branch of mathematics that focuses on the partial truth concept and approximation. The reasoning system of the theory has strong quantitative performance (accuracy) as well as language representation (interpretability). Designing a fuzzy system, on the other hand, is a difficult undertaking that necessitates the identification of numerous parameters. The performance of a fuzzy system is determined by several factors, including the fuzzy rules that are used and the membership functions. This function is responsible for mapping crisp inputs of the system to linguistic variables. In general designing these parameters is considered a difficult undertaking that necessitates the assistance of a fuzzy system expert or an effective optimization procedure. In order to overcome this problem, a genetic approach was used in this paper to create efficient fuzzy systems. The Wisconsin breast cancer diagnostic (WBCD) problem is then subjected to this optimization technique. This strategy was effective in developing fuzzy based systems which uses a smaller number of rules and provide a higher performance of over 93%.

Keywords*:* Fuzzy logic, Genetic Algorithm, Brest Cancer Detection.

1. Introduction

Fuzzy Logic was developed in the year 1965 by Lofti Zadeh [1]. The truth values of variables in this multi-valued logic can be any value between 0 and 1. The human capacity for vision and cognition served as the inspiration for the development of fuzzy logic, which makes use of linguistic variables instead of discrete values and so supports the idea of partial truth. This reasoning is at the heart of fuzzy systems, which find widespread usage in fields as diverse as fuzzy control [2], [3], medical diagnosis [4], and data classification [5]. A fuzzy system's effectiveness is determined by a number of criteria, including the fuzzy rules employed and what type of membership function (MF) used. The construction of MF is considered to be a challenging task, because defining the parameters that interpret/convert the crisp form of the input into suitable linguistic variables, requires an extensive understanding of the subject area, generally involving a subject matter expert or an appropriate optimization tool. Evolutionary algorithms have already been shown to be useful at designing effective fuzzy systems [4], [5], [6]. As a result, an implementation of evolutionary fuzzy modelling is described here, followed by its application and evaluation on the benchmarked dataset WBCD - Wisconsin Breast Cancer Diagnostic. The flow of research work is mentioned below, General overview of the Fuzzy Sets & Fuzzy Systems is presented in Section II. The optimization approach is then illustrated in Section III using a simple evolutionary algorithm. Section IV will detail the methodology utilised to solve the Wisconsin breast cancer diagnostic (WBCD) challenge, as well as the outcomes obtained. We draw some conclusions from the conducted research in Section V.

2. Fuzzy Systems

Fuzzy logic is a branch of mathematics concerned with approximate reasoning. By including the concept of degree into the condition verification process. The application of fuzzy logic provides more flexibility as well as the capacity to account for mistakes and uncertainties. Fuzzy systems use fuzzy logic to make decisions about data. To do this, they use a technique that compares each instance's occurrence with one or more entries in the rule base, a single repository for all relevant rules [7]. These rules were developed by experts in the field. The sections that follow define the fundamental ideas and concepts.

2.1. Fuzzy Logic and Sets

Similar to a normal or a crisp set, a fuzzy set in a universe of discourse can be described by requirements that allow for the identification of the elements $x \in P$, The fuzzy set is denoted by P and the universe of discourse is represented by U. These criteria are defined by a membership function.

$$\mu_P(x): U \rightarrow [0,1] \qquad \text{Eq. 1}$$

Eq. 1 expresses the degree to which X is a member of the fuzzy set P. Fuzzification operation converts a given input value x to a membership value $\lambda(x)$. Triangular, trapezoidal, or bell-shaped membership functions are frequently used because they provide a straightforward illustration of the fuzzy set they describe.

We define the fuzzy logic's primary operators using these membership functions. The AND, OR, and NOT operators of fuzzy logic can be defined in the same way as their boolean counterparts can. The operations listed are a generalisation of Boolean operators, however the Boolean operators cannot be used as is, it needs to be reduced to standard operators when fuzzy sets are transformed into crisp sets. There are many different variants of AND and OR operators event though that is the case, in this section, the focus will only be on the minimum and maximum values.

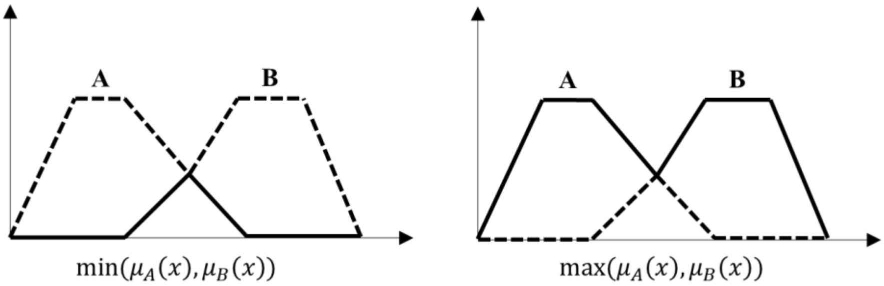

Fig. 1 Operators AND (on the left) and OR (on the right)

The NOT (complement) operator, is a widely utilized operator in fuzzy systems. To represent language variables such as Low, High, and Cold, fuzzy sets are employed. A crisp value can be subjected to a variety of membership functions. A height of 5 ft, for example, might have the membership values Tall = 0.7 and Short = 0.3, indicating that it belongs to the Tall and Short sets. Because of this, we are able to readily express some fuzzy conditions, like "His hight is tall" then apply fuzzy rules to make judgments that have significant implications.

$$\text{if (input} - \text{fuzzy} - \text{condition) then (output} - \text{fuzzy} - \text{assignment)} \qquad \text{Eq. 2}$$

where,

- *(input-fuzzy-condition) is the antecedent*
- *(output-fuzzy-assignment) is the consequent*

Eq. 2 represent the general form of fuzzy rules of a fuzzy inference system. Several rules may be present in a fuzzy inference system, and all of them are examined during the reasoning process. As a result of the fact that a clear input might potentially belong to a number of distinct fuzzy sets, it is possible to make use of a number of rules, each of which has a unique truth level associated with the output fuzzy set.

Because it is meant to come to an overall single conclusion or finding, we employ an aggregate operation to create a single fuzzy set. The OR operation (disjunction) is utilised the vast majority of the time to accomplish this task. Then, to have a crisp result, a defuzzification procedure is utilised. The Centre of Areas, often known as centroid, and the Mean of Maxima are two extensively used defuzzification procedures. These methods are depicted in fig. 2. COA is Centre of Area and MOM is Mean of Maxima

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

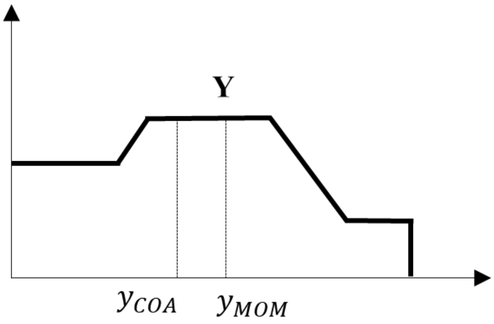

Fig. 2 Defuzzification

2.2. Fuzzy Inference Systems

Various components and sub-systems combine together to make up a fuzzy logic system, which is as follows: The following are the four components that make up a fuzzy reasoning engine: (1) a fuzzifier, (2) an inference engine, (3) a defuzzifier, and (4) a knowledge base. The fuzzy output is transformed into a clear value by the defuzzifier, and the knowledge base acts in the same capacity as a traditional database by storing rule definitions, membership functions, and permissible fuzzy operators.

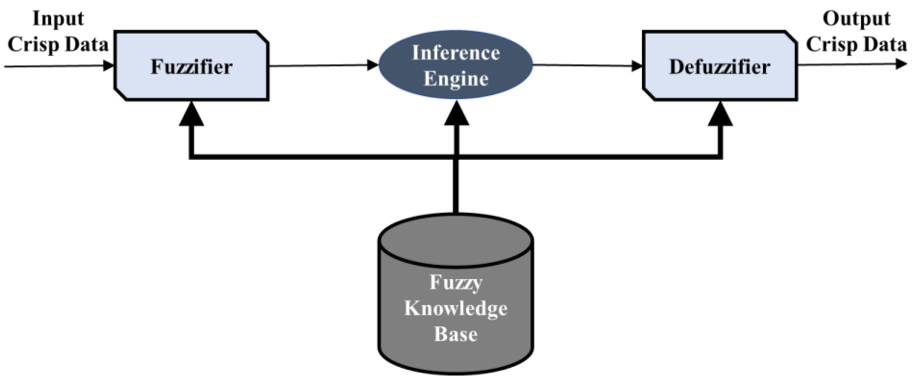

Fig. 3 The Fundamental Building Blocks of Fuzzy Systems

There are 3 different types of fuzzy systems: (1) Mamdani fuzzy system, (2) Takagi-Sugeno Kang fuzzy system, and (3) Signleton fuzzy system. The consequences of their norms distinguish them from one another. The Mamdani fuzzy system makes use of the definitions of consequents from the preceding sections, however the Takagi-Sugeno Kang uses a function of the input variable as the consequent and the Singleton fuzzy systems substitute the consequent with singletons [8]. A consequent is just a fuzzy set whose membership function represents it.

Fuzzy system parameters can be divided into four types [9] based on their function, logical parameters, structure parameters, connective parameters, and operational parameters.

3. Genetic Algorithms

EAs are algorithms that are motivated by nature and can be used to solve issues by applying Darwinian principles. They have the ability to develop increasingly better solutions to a problem by mimicking the practice of natural selection on a certain population and employing fundamental mechanisms such as selection, mutation, and crossover to do so. The performance of each person in a population was then determined by an objective function known as the fitness function.

The following steps can be used to describe a rudimentary genetic algorithm that makes use of binary numbers:

Step 1: In the first phase of the technique, an initial population is generated. Any randomization technique could be used for this generation. The size of the initial population will be represented by n

Step 2: Evaluating the overall fitness of the individuals that make up the existing population of size n.

Step 3: A mating pool is established by selecting k individuals from the initial population to serve as potential partners. For all of the decision-making that has to be done throughout this article, we shall use one of the most commonly used selection methods in Genetic Algorithms, the roulette wheel selection method. To begin, a scaling method is utilised on the fitness of every member of the population, which ultimately results in a scaled fitness value of for each person in the population, represented in Eq. 3

$$f_i' \in [0,1] \qquad \text{Eq. 3}$$

After scaling, the total is then calculated Eq. 4, and a random number r is produced as a result of this computation.

$$S = \sum_{i=1}^{n} f_i' \qquad \text{Eq. 4}$$

$$r \in [0,1] \qquad \text{Eq. 5}$$

For a member of the mating pool to be replicated it needs to satisfy the condition in Eq. 6.

$$f_k' + \ldots + f_{k+1}' \leq rS \leq f_k' + \ldots + f_{k+1}' \qquad \text{Eq. 6}$$

After that, the procedure is repeated j times. When utilising this tactic, the likelihood of getting chosen is determined by a ratio that is proportionate to a scaled fitness value.

Step 4: Perform the reproduction process (crossover procedure) in order to generate n children. Using a random selection procedure, 2 parents are chosen, and a random portion of their DNA is exchanged to produce 2 offspring.

Step 5: The freshly formed colony of bacteria was subjected to mutation. Every aspect of the individual is investigated using the probability b_p , and any of those parts may undergo modification. In the course of this research, the value of b_p was determined to be equal to 0.045.

When the desired number of generations or desired performance has been achieved, we can return to the second stage and repeat the process once more.

4. The Diagnostics Problem & Application - Wisconsin Breast Cancer Diagnosis (WBCD)

In this part, a classification system based on fuzzy logic for the detection of breast cancer has been built, with the data from the WBCD database serving as the foundation [10], [11]. This dataset is a collection of screening data collected at the University of Wisconsin Hospital using fine-needle aspiration. These efforts were conducted with the purpose of determining whether or not a patient had breast cancer (FNA). When performing this procedure, a small-gauge needle is used to drain fluid from a breast lump. Following that, the fluid is examined under a microscope.

It is believed that the following nine properties of a FNA sample are important in the diagnosis of breast cancer:

- Clump thickness – *v1*
- Cell size uniformity – *v2*
- Cell shape uniformity – *v3*
- Marginal adhesion – *v4*
- Single epithelial cell size – *v5*
- Bare nuclei – *v6*
- Bland chromatin – *v7*

- Normal nucleoli – *v8*
- Mitosis – *v9*

In order to facilitate the diagnosis, an integer value ranging from 1 to 10 has been ascribed to each characteristic. The WBCD database has 699 cases, each of which has been assigned a diagnosis. 16 of these items had an uncertain value for one of the attributes listed above, and as a result, they were eliminated from the database used in this study.

Table 1 Sample from Wisconsin Breast Cancer Dataset

Sl. No.	Patient ID	v1	v2	v3	v4	v5	v6	v7	v8	v9	Diagnosis
206	1218105	5	10	10	9	6	10	7	10	5	4
300	63375	9	1	2	6	4	10	7	7	2	4
90	1155546	2	1	1	2	3	1	2	1	1	2
628	1190546	2	1	1	1	2	5	1	1	1	2
248	145447	8	4	4	1	2	9	3	3	1	4

In this fuzzy system, the nine properties of FNA samples are employed as inputs, and the system subsequently calculates a continuous appraisal value of said malignancy of a particular case based on those inputs. The output of the fuzzy system is then used to generate a diagnostic, which can be either benign or malignant. A benign diagnosis is assigned a value of 2 and a malignant diagnosis is assigned a value of 4. The outcome of a fuzzy system having a value that is less than 3 is classed by the threshold as a benign diagnostic, whereas the outcome of a fuzzy system having a value that is larger than 3 is categorized as the malignant diagnosis. The degree of confidence that may be inferred from the system's prognosis is shown by the degree to which the outcome of the fuzzy system deviates from a value of either 2 or 4.

4.1. Parameters of a fuzzy system

The architecture of our fuzzy system must have two features to be a useful computerised diagnosis tool: it should be capable of high performance while also being straightforward to read. Those two features are frequently at odds with one another. A high-performance system should be able to make a right diagnosis in most situations, whereas an interpretable system has to be able to give a detailed explanation of how the evaluation was carried out, in addition to supplying a confidence rating for the diagnosis it produces [12]. It's possible that a fuzzy logic-based system with a considerable number of difficult and strong fuzzy rules and sophisticated membership functions may achieve tremendous performance, but due to the complexity of the system, it would have poor interpretability.

Several research has been conducted using the WBCD database. Bennet et al`s work [13] used a linear programming model to achieve a 99.6 per cent classification efficiency on a total 487 examples, which they reported in their paper. But their diagnosis determinations are opaque, with little explanation as to how they arrived at their conclusions. Because of this, we have determined that when designing our fuzzy system, we would take the interpretability requirement into consideration. As a result, the fuzzy system must have significantly fewer rules and a minimum possible number of variables for each rule, as described above. The following describes the configuration of the fuzzy system:

Input and output membership functions (Input: High or Low; Output: Benign or Malignant) are defined by the structural parameters. The user determines the number of rules to be used.

In the case of connective parameters, the evolutionary algorithm is to be utilised in order to uncover a antecedents of criteria for making a diagnosis benign. A default-rule having malignant diagnosis has been designed since at least one rule must be utilised for each collection of inputs regardless of the inputs. The evolutionary algorithm is responsible for determining the weight and the antecedent to be used.

Operational parameters: the GA searches for the input membership-function values using the operational parameters. P = 2.7 and d = 0.6 are the fixed values for the output membership values.

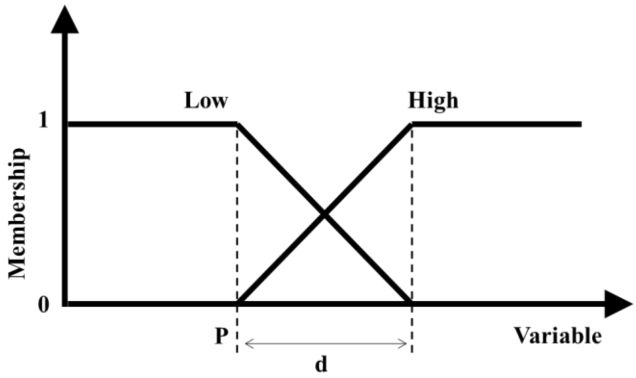

Fig. 4 The input membership functions are determined by P and d, the beginning point and edge length.

4.2. The evolutionary setup

In order for the evolutionary technique to work properly, it is important to search for several fuzzy-system parameters. These parameters include the inputs membership function, antecedents of rules, and weights provided to default rules. A conventional genetic algorithm has been used in conjunction with a Pittsburgh-style strategy [14]. In this method, each member of a population stands in for the entirety of a fuzzy system, and the job of the evolutionary algorithm is to keep a population of candidate fuzzy systems alive and well. Three parameters are encoded in the genome of everyone:

Parameters for the membership-function. For each of the inputs $(v_1 - v_9)$, two parameters P and d are specified, and these describe the feature of the input membership function, respectively. In the range, the values of P and d can be any number between 1 and 8, inclusive.

A benign rule is of the following structure: if $if\ (v_1\ is\ A_1^i)\ and\ \dots and\ \ (v_9\ is\ A_9^i)\ then\ (out\ is\ benign)$ where A_j^i represents the linguistic variable applicable to v_j.

A_j^i can have one of the following values: Low: 1, High: 2, or Don`t Care: (0 or 3), Don't Care means v_j will not be utilised by our rule. Once a rule has been eliminated from the list of rules because the evolutionary algorithm has assigned the value Don't Care to all its antecedents, the rule is no longer considered.

Parameters for the default rules: The algorithm selects the weight and the index of the suitable antecedent variable based on the information provided. Weight and index takes the following constraints $w \in \{0.1, \dots, 1.0\}$ and $k \in \{1, \dots, 9\}$ respectively.

The evolution first starts out with a population of fixed size, 100. The process of evolution will come to an end when the max number of generations has been achieved. The maximum size of the population is represented by G_{max}.

The value of $G_{max} = 20$ was chosen for this work to keep the processing time as low as possible, however this number might be increased if a greater level of performance is desired. The G_{max} value selection is generally a trade-off between the processing time and level of performance.

During the process of creating the mating pool, a total of 25 individuals were chosen.

There are three criteria for measuring the performance / outcome of a fuzzy logic system: (1) classification performance: F_c, (2) average quadratic error: F_e (3) number of variables (average) per active rule: F_v.

In this case, the fitness function of the designed fuzzy system is denoted by the Eq. 7

$$F = F_c - \alpha F_v - \beta F_e, \quad \text{Eq. 7}$$

where α and β are tiny empirical values. In this section, we corrected $\alpha = 0.05$ and $\beta = 0.01$.

Table 2 Parameter Encoding – Individual Genome

Parameters	Bits	Quantity	Total Bits	Values
P	3	9	27	$\{1, \ldots, 8\}$
d	3	9	27	$\{1, \ldots, 8\}$
A	2	$9N_r$	$18N_r$	$\{0,1,2,3\}$
w	4	1	4	$\{0.1, \ldots, 1.0\}$
k	4	1	4	$\{1, \ldots, 9\}$

4.3. Results

One to five harmless rules were given. The evolutionary method was trained using all 683 instances from the WBCD database; a future enhancement to this implementation might include the definition of test and training sets in order to reduce overfitting. Features exhibited by the developed fuzzy logic system is illustrated in Table 3

Table 3 Performance of the Fuzzy System

N_r	1	2	3	4	5
F_C	0.9541	0.9511	0.9401	0.9427	0.9454

N_r is the number of rules specified, while F_c denotes the proportion of correctly categorised cases.

Raising the number of rules doesn't always necessarily boost performance, as demonstrated. By doing a larger number of evolutionary runs it can be discovered how the average performance changes as the number of benign rules and generations increases.

Table 4 Parameters - Fuzzy System - One Rule

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P		4	1			1			
d		1	5			6			

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
R_1		1				1			
R_d			1						

R_d – default rules with a weight of $w = 0.3$.

The tables Table 4 and Table 5 illustrate the designed fuzzy systems using one/two benign rules. P and d are the membership functions' parameters.

When a rule does not use any of the available variables, the parameters related to it become irrelevant and are excluded from the subsequent tables.

The A_i, the rules' antecedents: $A_i = 1$ denotes a low value for v_i, whereas $A_i = 2$ denotes a high value for v_i.

Consequently, when our fuzzy systems are optimised with a genetic algorithm, they may attain high performance in a reasonable amount of generations.

Table 5 Parameters Of The Fuzzy System With Two Rules

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
$\boldsymbol{P}$	6		8	1	1	7		5	3
$\boldsymbol{d}$	1		6	3	2	5		6	7

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
$\boldsymbol{R_1}$				2	1				
$\boldsymbol{R_2}$	1		1			1		1	1
$\boldsymbol{R_d}$						1			

R_d – default rules with a weight of $w = 0.7$.

5. Conclusion

This paper aims to demonstrate how a genetic algorithm optimization method may be used to create high-performance fuzzy systems. This procedure was then applied to benchmarked dataset WBCD. The research work considers a population of size 100 different individuals and for a total of 20 generations, using this a fuzzy system was constructed. This fuzzy system was able to attain a performance rate of 93%

Future research may involve estimating the average performance of fuzzy systems throughout several evolutionary runs. Incorporating a training dataset into the evolutionary algorithms may also be advantageous.

Refrences

[1] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.

[2] S. P, S. D.N, and P. B, "Temperature Control using Fuzzy Logic," *Int. J. Instrum. Control Syst.*, vol. 4, no. 1, pp. 1–10, Jan. 2014, doi: 10.5121/ijics.2014.4101.

[3] J. Y. Zhang and Y. D. Li, "Application of genetic algorithm in optimization of fuzzy control rules," *Proc. - ISDA 2006 Sixth Int. Conf. Intell. Syst. Des. Appl.*, vol. 1, pp. 529–532, 2006, doi: 10.1109/ISDA.2006.102.

[4] J. Zhang, Q. Gao, Z. G. Huang, and Z. Huang, "Optimization of fuzzy rules by Muilti-objective genetic algorithm in avionic fault diagnosis system," *Proc. - 2009 IEEE Int. Conf. Intell. Comput. Intell. Syst. ICIS 2009*, vol. 1, pp. 433–437, 2009, doi: 10.1109/ICICISYS.2009.5357809.

[5] G. Saggio *et al.*, "Comparison of two different classifiers for mental tasks-based brain-computer interface: MLP Neural Networks vs. Fuzzy logic," *2009 IEEE Int. Symp. World Wirel. Mob. Multimed. Netw. Workshop WOWMOM 2009*, 2009, doi: 10.1109/WOWMOM.2009.5282406.

[6] P. A. D. Castro and H. A. Camargo, "Learning and optimization of fuzzy rule base by means of self-adaptive genetic algorithm," *IEEE Int. Conf. Fuzzy Syst.*, vol. 2, pp. 1037–1042, 2004, doi: 10.1109/FUZZY.2004.1375552.

[7] Julaika Begum K, Dr. Sindhu J Kumaar, "Computational Utilization of Numerical Equivalences: Fuzzy systems and Neural Networks – A comparative Survey," *Des. Eng.*, vol. 2021, no. 07, pp. 14971–14981, Sep. 2021.

[8] S. Kaur, "A Comprehensive Study of New Fuzzy Logic Metric Based on Interval Type-2 Fuzzy Sets Using Real Application," *Tuijin JishuJournal Propuls. Technol.*, vol. 44, no. 3, Art. no. 3, Sep. 2023, doi: 10.52783/tjjpt.v44.i3.726.

[9] C. A. Peña Reyes, "Coevolutionary fuzzy modeling," Mar. 2005, doi: 10.5075/EPFL-THESIS-2634.

[10] O. L. Mangasarian and W. H. Wolberg, "Cancer diagnosis via linear programming," University of Wisconsin-Madison Department of Computer Sciences, 1990.

[11] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, no. 23, p. 9193, 1990, doi: 10.1073/PNAS.87.23.9193.

[12] R. Boadh *et al.*, "Study of fuzzy expert system for the diagnosis of various types of cancer," *Mater. Today Proc.*, vol. 56, pp. 298–307, Jan. 2022, doi: 10.1016/j.matpr.2022.01.161.

[13] K. P. Bennett and O. L. Mangasarian, "Neural network training via linear programming," University of Wisconsin-Madison Department of Computer Sciences, 1990.

[14] A. K. Varshney and V. Torra, "Literature Review of the Recent Trends and Applications in Various Fuzzy Rule-Based Systems," *Int. J. Fuzzy Syst.*, vol. 25, no. 6, pp. 2163–2186, Sep. 2023, doi: 10.1007/s40815-023-01534-w.