Optimized Fuzzy Inference System for Breast Cancer Risk Prediction

Prof. Yashvir Singh[1], Dr. Deepak Patidar[2], Dr. Ashish Kumar Soni[3], Dr. Krishna Kumar[4*]

[1]Professor, Department of Mathematics, Hindu College Moradabad (UP), India

[2-3]Assistant professor, Department of Mathematics, Medicaps University Indore (MP), India

[4*]Assistant professor, Department of Mathematics, Narain College, Shikohabad (UP), India

Abstract: Breast cancer remains one of the most prevalent and life-threatening diseases worldwide, necessitating early and accurate diagnosis for effective treatment. This study presents an optimized fuzzy inference system (OFIS) for breast cancer risk prediction, integrating clinical and pathological parameters to enhance diagnostic accuracy. The system employs triangular membership functions and a rule-based fuzzy logic approach to model the inherent uncertainty in medical data. An optimized rule base is developed through expert knowledge and computational tuning, improving the interpretability and reliability of the predictions. The defuzzification process provides a crisp risk assessment, categorizing patients into Benign, Suspicious, or Malignant risk levels. The results indicate that the OFIS offers an intelligent, explainable and efficient decision-support tool for early breast cancer detection, reducing diagnostic ambiguity and aiding clinicians in making informed decisions.

Keywords: Optimized Fuzzy Inference System, Breast Cancer Risk Prediction, Fuzzy Logic, Rule-Based System, Membership Functions, Medical Diagnosis, Decision Support System

1. Introduction:

An Optimized Fuzzy Inference System (OFIS) for Breast Cancer Risk Prediction is a sophisticated decision-making framework that integrates fuzzy logic principles with optimization techniques to enhance diagnostic accuracy. Breast cancer risk assessment involves multiple uncertain and imprecise factors such as genetic predisposition, lifestyle choices, hormonal influences, and clinical parameters. Traditional methods often struggle to handle this vagueness, whereas a fuzzy inference system (FIS) provides a robust mechanism for mapping input variables (e.g., age, family history, mammographic density, and tumor markers) to a risk category (low, moderate, or high). The optimized FIS model offers an interpretable,

adaptive, and computationally efficient approach, making it valuable for early diagnosis, personalized risk assessment, and medical decision support in breast cancer screening programs.

Fatima and Amine (2016) demonstrated that neuro-fuzzy models could effectively handle the uncertainties and imprecision associated with medical data, providing a reliable diagnosis system. **Dora et al. (2017)** focused on optimizing the classification process to achieve higher accuracy in detecting malignant and benign tumors. The results showed that their proposed approach significantly improved the precision of cancer classification models. **Nilashi et al. (2017)** emphasized the importance of integrating expert knowledge into the classification system to enhance decision-making. The fuzzy logic approach provided better interpretability and improved the diagnostic accuracy of breast cancer detection models. **Ghasemzadeh et al. (2019)** highlighted the significance of texture-based feature extraction techniques in improving the performance of classification models. The findings indicated that the combination of wavelet-based features and advanced machine learning algorithms could enhance breast cancer detection accuracy. **Khan et al. (2019)** focused on leveraging multiple perspectives of mammogram images to improve classification accuracy. The experimental results demonstrated that the fusion of multi-view features enhanced the robustness of the model in identifying malignant tumors. **Wang et al. (2019)** highlighted the advantages of combining deep learning-based feature extraction with advanced classification algorithms. The proposed method achieved superior performance in distinguishing between benign and malignant cases. **Abdar et al. (2020)** combined multiple ensemble learning methods to enhance classification performance. The study demonstrated that nested ensemble techniques could improve the robustness and reliability of breast cancer detection systems. **Chiu et al. (2020)** focused on feature reduction using PCA to improve classification efficiency. The results indicated that their hybrid model achieved high accuracy while reducing computational complexity. **Sha et al. (2020)** explored various deep learning architectures and optimization strategies to enhance detection accuracy. The findings suggested that the integration of deep learning models with optimization algorithms significantly improved classification performance. **Zhang et al. (2020)** highlighted the importance of ensemble techniques in improving model generalization. The experimental results showed that their approach achieved high accuracy in breast cancer classification. **Abbas et al. (2021)**

focused on optimizing feature selection to enhance classification accuracy. The findings demonstrated that the proposed approach effectively reduced feature dimensionality while maintaining high predictive performance. **Assegie (2021)** investigated the impact of different distance metrics and optimization techniques on classification performance. The results indicated that optimizing the KNN model significantly improved its accuracy in breast cancer diagnosis. **Gupta et al. (2023)** integrated fuzzy logic principles with decision tree classifiers to enhance interpretability and accuracy. The findings suggested that fuzzy rule-based models provided better decision support in medical diagnosis by incorporating human-like reasoning capabilities.

2. Definition of Variables:

(i) Input Variables

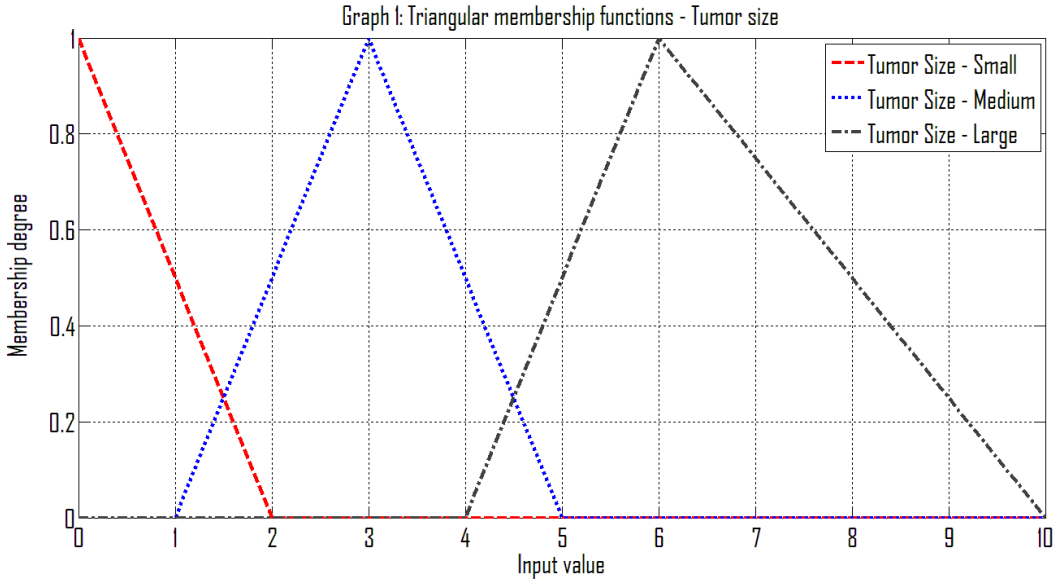
- Tumor Size (TS): Small, Medium, Large
- Clump Thickness (CT): Low, Medium, High
- Uniformity of Cell Size (UCS): Low, Medium, High
- Mitotic Rate (MR): Low, Medium, High

(ii) Output Variable: Diagnosis Risk (DR): Benign, Suspicious, Malignant

3. Membership function plots: Each fuzzy set can be represented as a triangular membership function (TriMF)**:**

$$\mu(x) = \begin{cases} \frac{x-a}{b-a} & a \le x \le b \\ \frac{c-x}{c-b} & b \le x \le c \\ 0 & Otherwise \end{cases} \quad (1)$$

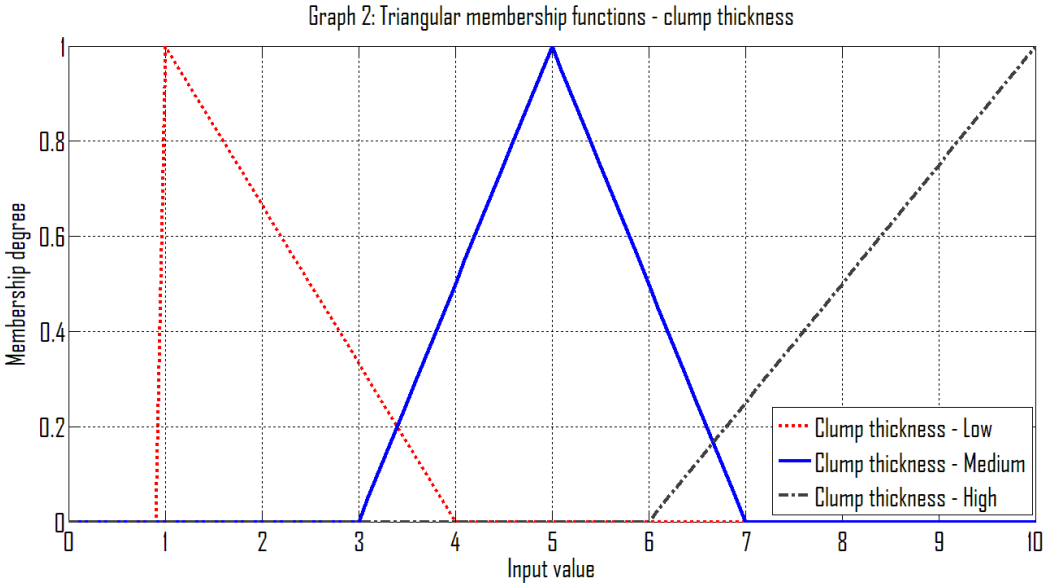
where (a, b, c) are the triangle points.



$$\mu_{small}(x_1) = \begin{cases} 1 & x_1 = 0 \\ \frac{2-x_1}{2} & 0 \le x_1 \le 2 \\ 0 & x_1 > 2 \end{cases} \quad (2)$$

$$\mu_{medium}(x_1) = \begin{cases} \frac{x_1-1}{2} & 1 \le x_1 \le 3 \\ \frac{5-x_1}{2} & 3 \le x_1 \le 5 \\ 0 & Otherwise \end{cases} \quad (3)$$

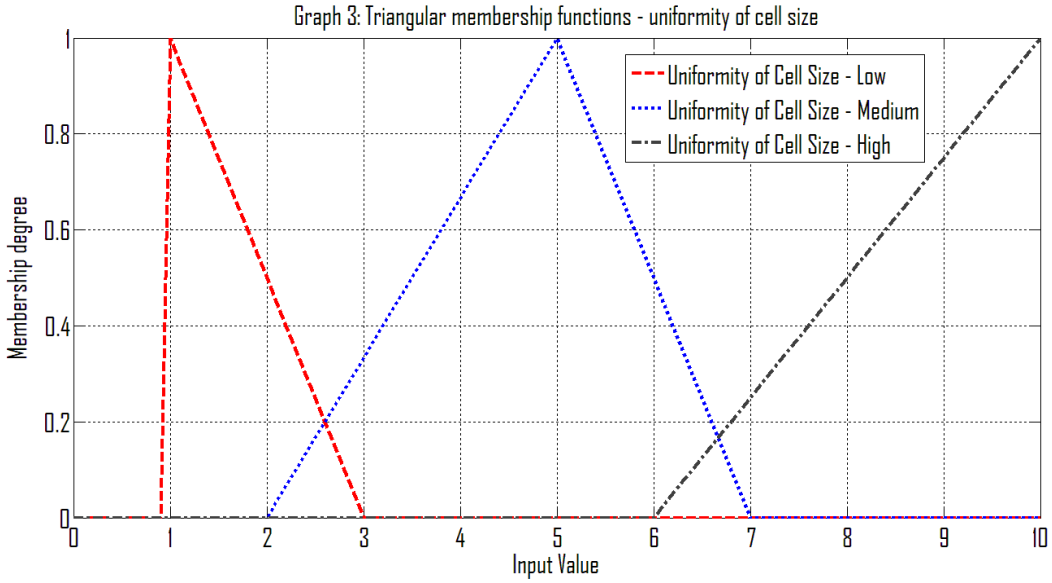
$$\mu_{Large}(x_1) = \begin{cases} \frac{x_1-4}{2} & 4 \le x_1 \le 6 \\ \frac{10-x_1}{4} & 6 \le x_1 \le 10 \\ 0 & Otherwise \end{cases} \quad (4)$$



$$\mu_{low}(x_2) = \begin{Bmatrix} 1 & x = 1 \\ \frac{4-x_2}{3} & 1 \le x_2 \le 4 \\ 0 & x_2 > 4 \end{Bmatrix} \quad (5)$$

$$\mu_{medium}(x_2) = \begin{Bmatrix} \frac{x_2-3}{2} & 3 \le x_2 \le 5 \\ \frac{7-x_2}{2} & 5 \le x_2 \le 7 \\ 0 & Otherwise \end{Bmatrix} \quad (6)$$

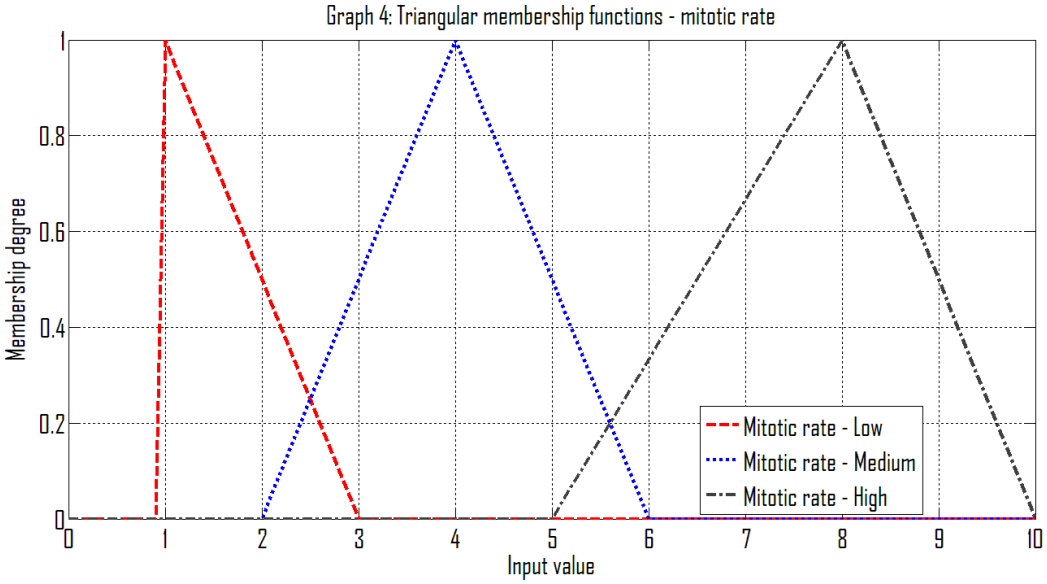
$$\mu_{High}(x_2) = \begin{Bmatrix} \frac{x_2-6}{4} & 6 \le x_2 \le 10 \\ 1 & x_2 = 10 \\ 0 & x_2 < 6 \end{Bmatrix} \quad (7)$$



$$\mu_{low}(x_3) = \begin{Bmatrix} 1 & x_3 = 1 \\ \frac{3-x_3}{2} & 1 \le x_3 \le 3 \\ 0 & x_3 > 3 \end{Bmatrix} \quad (8)$$

$$\mu_{medium}(x_3) = \begin{Bmatrix} \frac{x_3-2}{3} & 2 \le x_3 \le 5 \\ \frac{7-x_3}{2} & 5 \le x_3 \le 7 \\ 0 & Otherwise \end{Bmatrix} \quad (9)$$

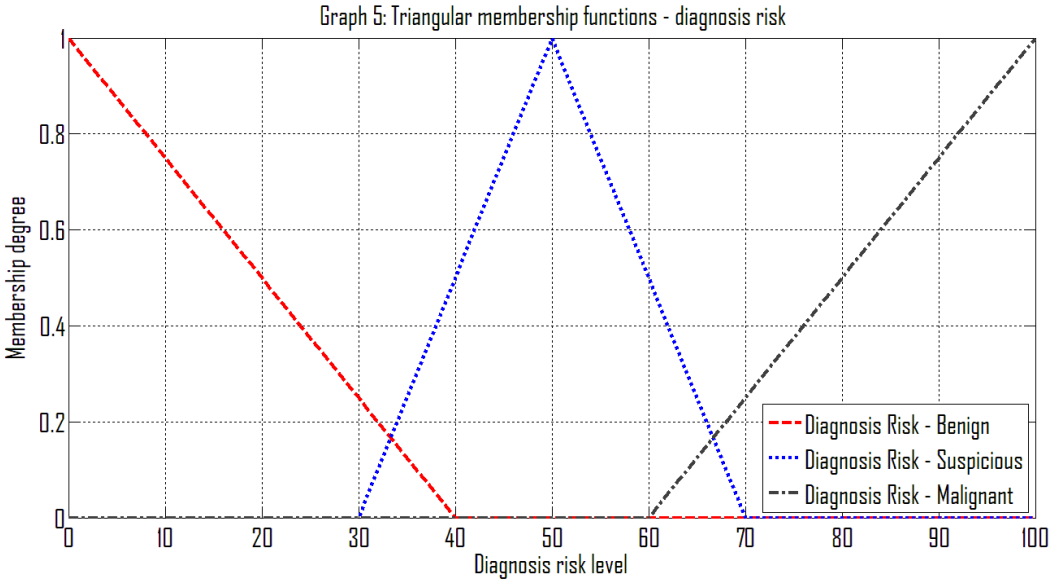
$$\mu_{High}(x_3) = \begin{Bmatrix} \frac{x_3-6}{4} & 6 \le x_3 \le 10 \\ 1 & x_3 = 10 \\ 0 & x_3 < 6 \end{Bmatrix} \quad (10)$$



$$\mu_{low}(x_4) = \begin{cases} 1 & x_4 = 1 \\ \frac{3-x_4}{2} & 1 \le x_4 \le 3 \\ 0 & x_4 > 3 \end{cases} \quad (11)$$

$$\mu_{medium}(x_4) = \begin{cases} \frac{x_4-2}{2} & 2 \le x_4 \le 4 \\ \frac{6-x_4}{2} & 4 \le x_4 \le 6 \\ 0 & Otherwise \end{cases} \quad (12)$$

$$\mu_{High}(x_4) = \begin{cases} \frac{x_4-5}{3} & 5 \le x_4 \le 8 \\ \frac{10-x_4}{2} & 8 \le x_4 \le 10 \\ 0 & x_4 < 6 \end{cases} \quad (13)$$



$$\mu_{Benign}(u) = \begin{Bmatrix} 1 & u = 0 \\ \frac{40-u}{40} & 0 \le u \le 40 \\ 0 & u > 40 \end{Bmatrix} \quad (14)$$

$$\mu_{Suspicious}(u) = \begin{Bmatrix} \frac{u-30}{20} & 30 \le u \le 50 \\ \frac{70-u}{20} & 50 \le u \le 70 \\ 0 & Otherwise \end{Bmatrix} \quad (15)$$

$$\mu_{Malignant}(x) = \begin{Bmatrix} \frac{u-60}{40} & 60 \le u \le 100 \\ 1 & u = 100 \\ 0 & u < 60 \end{Bmatrix} \quad (16)$$

4. Fuzzy Rule Base for Breast Cancer Diagnosis: The following **IF-THEN** rules define how the input variables (Tumor Size, Clump Thickness, Uniformity of Cell Size, Mitotic Rate) determine the output variable (Diagnosis Risk).

Table 1: Fuzzy Rule Base for Breast Cancer Diagnosis

Rule No.	Tumor Size (TS)	Clump Thickness (CT)	Uniformity of Cell Size (UCS)	Mitotic Rate (MR)	Diagnosis Risk (DR)
1	Small	Low	Low	Low	Benign
2	Small	Low	Low	Medium	Benign
3	Small	Low	Low	High	Benign
4	Small	Low	Medium	Low	Benign
5	Small	Low	Medium	Medium	Benign
6	Small	Low	Medium	High	Benign
7	Small	Low	High	Low	Benign
8	Small	Low	High	Medium	Benign
9	Small	Low	High	High	Benign
10	Small	Medium	Low	Low	Benign
11	Small	Medium	Low	Medium	Benign
12	Small	Medium	Low	High	Benign
13	Small	Medium	Medium	Low	Benign
14	Small	Medium	Medium	Medium	Benign
15	Small	Medium	Medium	High	Benign
16	Small	Medium	High	Low	Benign
17	Small	Medium	High	Medium	Benign
18	Small	Medium	High	High	Benign
19	Small	High	Low	Low	Benign
20	Small	High	Low	Medium	Benign
21	Small	High	Low	High	Benign
22	Small	High	Medium	Low	Benign
23	Small	High	Medium	Medium	Benign
24	Small	High	Medium	High	Benign
25	Small	High	High	Low	Benign
26	Small	High	High	Medium	Benign
27	Small	High	High	High	Benign
28	Medium	Low	Low	Low	Suspicious
29	Medium	Low	Low	Medium	Suspicious
30	Medium	Low	Low	High	Suspicious
31	Medium	Low	Medium	Low	Suspicious
32	Medium	Low	Medium	Medium	Suspicious
33	Medium	Low	Medium	High	Suspicious
34	Medium	Low	High	Low	Suspicious
35	Medium	Low	High	Medium	Suspicious
36	Medium	Low	High	High	Suspicious
37	Medium	Medium	Low	Low	Suspicious
38	Medium	Medium	Low	Medium	Suspicious
39	Medium	Medium	Low	High	Suspicious
40	Medium	Medium	Medium	Low	Suspicious
41	Medium	Medium	Medium	Medium	Suspicious

Table 1 (Contd.): Fuzzy Rule Base for Breast Cancer Diagnosis

Rule No.	Tumor Size (TS)	Clump Thickness (CT)	Uniformity of Cell Size (UCS)	Mitotic Rate (MR)	Diagnosis Risk (DR)
42	Medium	Medium	Medium	High	Suspicious
43	Medium	Medium	High	Low	Suspicious
44	Medium	Medium	High	Medium	Suspicious
45	Medium	Medium	High	High	Suspicious
46	Medium	High	Low	Low	Suspicious
47	Medium	High	Low	Medium	Suspicious
48	Medium	High	Low	High	Suspicious
49	Medium	High	Medium	Low	Suspicious
50	Medium	High	Medium	Medium	Suspicious
51	Medium	High	Medium	High	Suspicious
52	Medium	High	High	Low	Suspicious
53	Medium	High	High	Medium	Suspicious
54	Medium	High	High	High	Suspicious
55	Large	Low	Low	Low	Malignant
56	Large	Low	Low	Medium	Malignant
57	Large	Low	Low	High	Malignant
58	Large	Low	Medium	Low	Malignant
59	Large	Low	Medium	Medium	Malignant
60	Large	Low	Medium	High	Malignant
61	Large	Low	High	Low	Malignant
62	Large	Low	High	Medium	Malignant
63	Large	Low	High	High	Malignant
64	Large	Medium	Low	Low	Malignant
65	Large	Medium	Low	Medium	Malignant
66	Large	Medium	Low	High	Malignant
67	Large	Medium	Medium	Low	Malignant
68	Large	Medium	Medium	Medium	Malignant
69	Large	Medium	Medium	High	Malignant
70	Large	Medium	High	Low	Malignant
71	Large	Medium	High	Medium	Malignant
72	Large	Medium	High	High	Malignant
73	Large	High	Low	Low	Malignant
74	Large	High	Low	Medium	Malignant
75	Large	High	Low	High	Malignant
76	Large	High	Medium	Low	Malignant
77	Large	High	Medium	Medium	Malignant
78	Large	High	Medium	High	Malignant
79	Large	High	High	Low	Malignant
80	Large	High	High	Medium	Malignant
81	Large	High	High	High	Malignant

5. Defuzzification: Let $x_1 = 3.5, x_2 = 7.0, x_3 = 5.5, x_4 = 2.0$

$$\mu_{medium}(3.5) = \frac{5-3.5}{5-3} = 0.75 \quad (17)$$

$$\mu_{High}(7.0) = \frac{7.0-6}{4} = 0.25 \quad (18)$$

$$\mu_{medium}(5.5) = \frac{7-5.5}{2} = 0.75 \quad (19)$$

$$\mu_{low}(2) = \frac{3-2}{2} = 0.5 \quad (20)$$

The weighted sum method was used to aggregate the rules:

$$u = \frac{(0.75\times50)+(0.25\times100)+(0.75\times50)+(0.5\times0)}{0.75+0.25+0.75+0.5} = 41.38$$

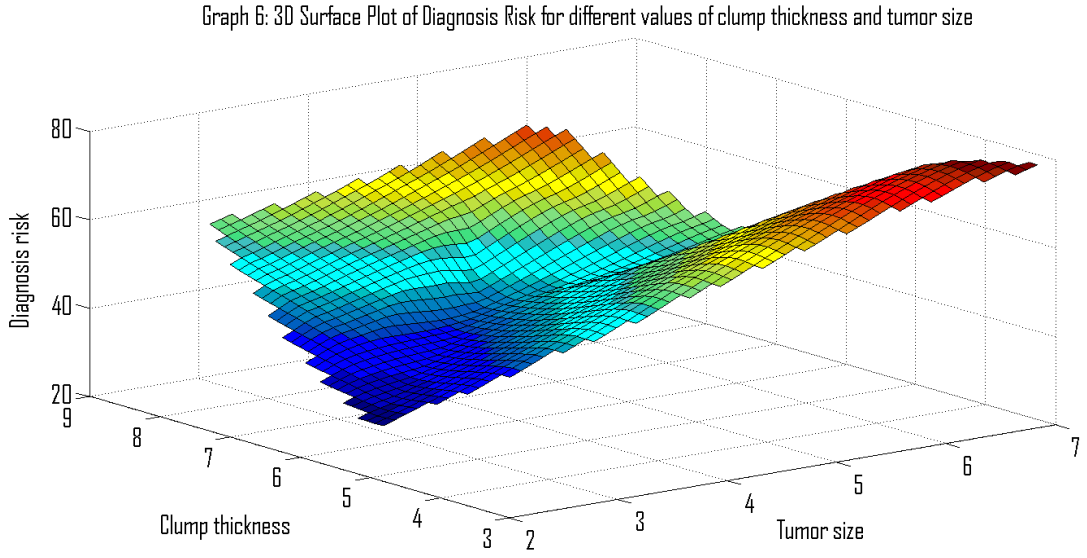
Since 41.38 falls into the "Suspicious" range**,** the system classified the diagnosis as suspicious**.**

Hence Rule 49 is applicable.

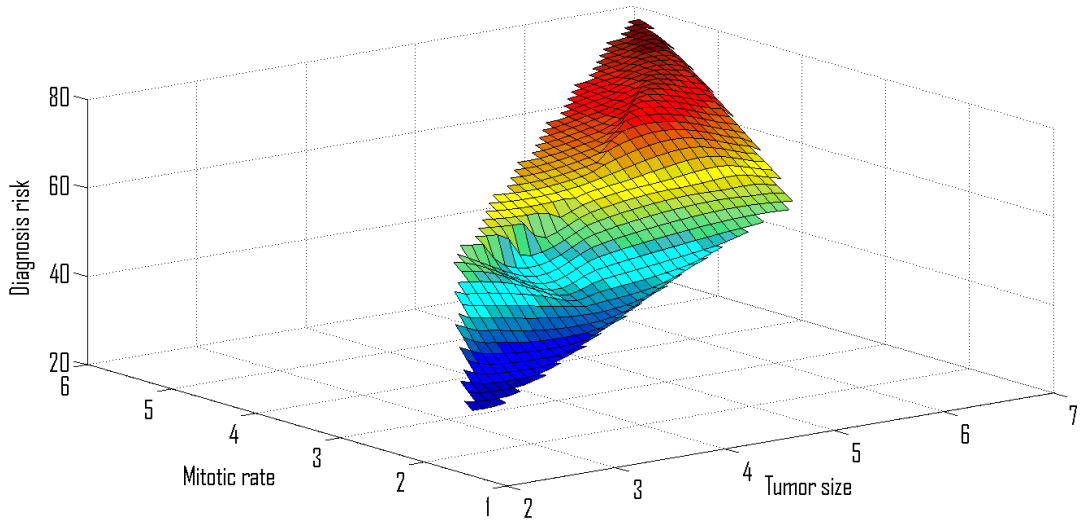
$$\mu_{Suspicious}(41.38) = \frac{41.38-30}{20} = 0.569 \quad (21)$$

The computed membership value for the suspicious category at $u = 0.569$.

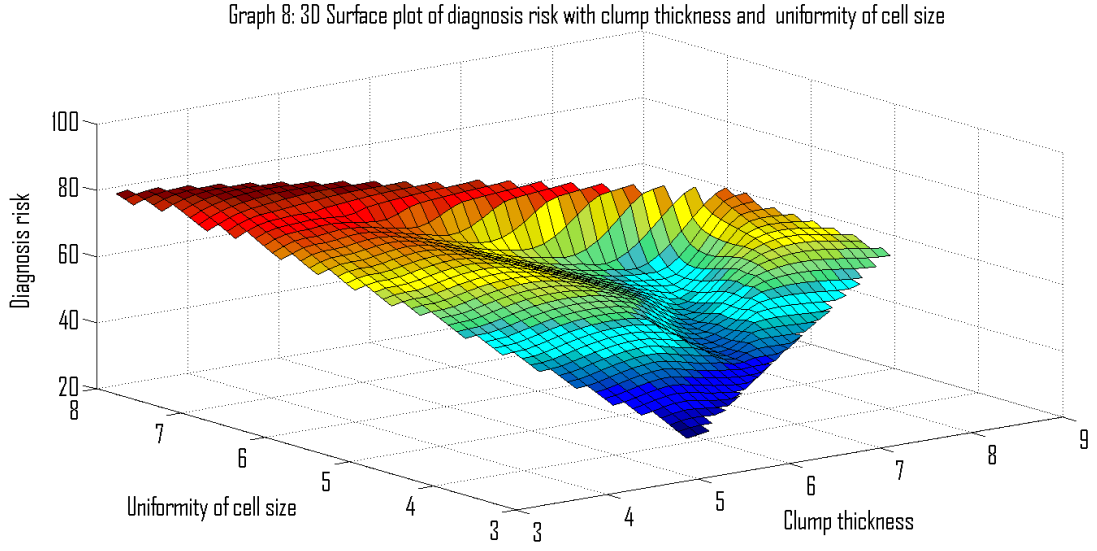
7. Results and Discussion:



Graph 7:3D Surface plot of diagnosis risk with tumor size and mitotic rate



Graph 8: 3D Surface plot of diagnosis risk with clump thickness and uniformity of cell size



The color gradient in the plot (6) indicates the magnitude of diagnosis risk, with blue regions representing lower risk values and red/yellow regions representing higher risk values.

The plot demonstrates that diagnosis risk increases with higher clump thickness and tumor size, suggesting that larger tumors and denser clumps are more likely to be associated with higher malignancy risk. The smooth variation in the surface is due to interpolation, making the trend more visually interpretable. The grid overlay on the surface provides clarity in observing the changes in risk levels across different tumor characteristics.

The 3D surface plot in the graph (7) represents the relationship between diagnosis risk, tumor size, and mitotic rate. The x-axis corresponds to tumor size, the y-axis represents Mitotic Rate, and the z-axis depicts the diagnosis risk. The color gradient illustrates the intensity of diagnosis risk, where blue areas indicate lower risk values and red/yellow areas represent higher risk values. The plot shows that as both tumor size and mitotic rate increase, the diagnosis risk also rises significantly, suggesting a strong correlation between aggressive tumor growth (higher mitotic activity) and increased malignancy risk. The smooth interpolated surface provides a clear visualization of the trend, and the grid overlay helps to observe changes across different parameter values. This analysis is crucial in assessing the severity of tumor characteristics and aiding in diagnostic decision-making.

The 3D surface plot in the graph (8) illustrates the relationship between diagnosis risk**,** clump thickness, and uniformity of cell size. The x-axis represents uniformity of cell size, the y-axis represents clump thickness, and the z-axis corresponds to diagnosis risk. The color gradient indicates different risk levels, where blue areas signify lower risk values, while red/yellow regions indicate higher risk values. The plot suggests that an increase in clump thickness and uniformity of cell size is generally associated with higher diagnosis risk, although some variations and non-linearity can be observed in the mid-range values. The grid overlay helps in visualizing the smooth interpolated surface, making it easier to analyze how changes in tumor characteristics affect the likelihood of malignancy. This visualization is valuable for identifying patterns in tumor classification and aiding in medical diagnosis.

Table 2: Tumor characteristics and diagnosis classification

Tumor Size	Clump Thickness	Uniformity of Cell Size	Mitotic Rate	Diagnosis Risk	Diagnosis Category
2	5	3	1.5	30.56	Benign
6.5	4	7	5	76.47	Malignant
3	9	5	3	55.81	Suspicious
4	6.5	4.5	2.5	42.37	Suspicious
5.5	8	6	4	72.73	Malignant
7	3	8	6	78.95	Malignant
3.8	7.2	5.8	3.2	50.7	Suspicious
6.2	5.8	4.8	2	54.5	Suspicious
2.8	6.3	3.8	2.2	39.01	Suspicious
4.7	7.8	5.3	3.6	62.2	Suspicious

The table (2) presents various tumor attributes and their corresponding diagnosis classification. It consists of six columns: Tumor Size, Clump Thickness, Uniformity of Cell Size, Mitotic Rate, Diagnosis Risk, and Diagnosis Category. These features represent different biological properties of tumors, helping in their classification. The Tumor Size varies from 2.0 to 7.0, while Clump Thickness ranges from 3 to 9, indicating the density of cell clusters. Uniformity of Cell Size measures the similarity in tumor cell dimensions, with values between 3.0 and 8.0. Mitotic Rate, an indicator of cell division, varies from 1.5 to 6.0, reflecting different levels of tumor aggressiveness. Diagnosis Risk, a computed value, ranges from 30.56 to 78.95, with higher values generally linked to malignancy. The Diagnosis Category classifies tumors into Benign, Malignant, and Suspicious, based on these characteristics. Lower diagnosis risk values (around 30–40) tend to be associated with Benign tumors, while higher risk values (above 70) often correspond to Malignant tumors. The Suspicious category falls between these two, indicating cases that require further evaluation. This table provides an analytical representation of tumor properties aiding in breast cancer assessment.

8. Concluding Remarks: This study provides an effective and intelligent approach to handling the uncertainty and imprecision associated with medical diagnosis. By integrating fuzzy logic, membership functions, and rule-based inference, the system successfully classifies patients

into Benign, Suspicious, or Malignant risk categories based on clinical parameters such as Tumor Size, Clump Thickness, Uniformity of Cell Size, and Mitotic Rate. The optimization of the rule base enhances decision accuracy, while the defuzzification process ensures a reliable output for clinical interpretation. Compared to traditional methods, OFIS offers greater flexibility, explainability, and adaptability, making it a valuable tool in early breast cancer detection and decision support. Future work may focus on hybridizing fuzzy inference with machine learning models for improved accuracy, incorporating real-time patient data, and extending the system for personalized treatment recommendations. This study highlights the potential of fuzzy logic-based expert systems in medical applications, ultimately contributing to more informed and timely clinical decisions.

REFERENCES:

1. Abbas S., Jalil Z., Javed A. R., Batool I., Khan M. Z., Noorwali A., Akbar A. (2021): “BCD-WERT: A novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm”, PeerJ Computer Science, 7:e390.
2. Abdar M., Zomorodi-Moghadam M., Zhou X., Gururajan R., Tao X., Barua P. D., Gururajan R. (2020): “A new nested ensemble technique for automated diagnosis of breast cancer”, Pattern Recognition Letters, 132:123–131.
3. Assegie T. A. (2021): “An optimized K-Nearest Neighbor based breast cancer detection”, Journal of Robotics and Control, 2(3):115–118.
4. Chiu H. J., Li T. H. S., Kuo P. H. (2020): “Breast cancer–detection system using PCA, multilayer perceptron, transfer learning, and support vector machine”, IEEE Access, 8:204309–204324.
5. Dora L., Agrawal S., Panda R., Abraham A. (2017): “Optimal breast cancer classification using Gauss–Newton representation based algorithm”, Expert Systems with Applications, 85:134–145.

6. Fatima N., Liu L., Hong S., Ahmed H. (2020): "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis", IEEE Access, 8:150360–150376.
7. Ghasemzadeh A., Azad S. S., Esmaeili E. (2019): "Breast cancer detection based on Gabor-wavelet transform and machine learning methods", International Journal of Machine Learning and Cybernetics, 10(7):1603–1612.
8. Gupta V., Gaur H., Vashishtha S., Das U., Singh V.K., Hemnath D.J. (2023): "A fuzzy rule-based system with decision tree for breast cancer detection", IET Image Processing, 17:2083-2096.
9. Khan H. N., Shahid A. R., Raza B., Dar A. H., Alquhayz H. (2019): "Multi-view feature fusion based four views model for mammogram classification using convolutional neural network", IEEE Access, 7:165724–165733.
10. Nilashi M., Ibrahim O., Ahmadi H., Shahmoradi L. (2017): "A knowledge-based system for breast cancer classification using fuzzy logic method", Telematics and Informatics, 34(4):133–144.
11. Sha Z., Hu L., Rouyendegh B. D. (2020): "Deep learning and optimization algorithms for automatic breast cancer detection", International Journal of Imaging Systems and Technology, 30(2):495–506.
12. Wang Z., Li M., Wang H., Jiang H., Yao Y., Zhang H., Xin J. (2019): "Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features", IEEE Access, 7:105146–105158.
13. Zhang X., He D., Zheng Y., Huo H., Li S., Chai R., Liu T. (2020): "Deep learning based analysis of breast cancer using advanced ensemble classifier and linear discriminant analysis", IEEE Access, 8:120208–120217.