



# Expert system based on fuzzy rules for diagnosing breast cancer

Isha Thani<sup>1</sup> · Tanmay Kasbe<sup>2</sup>

Received: 3 August 2021 / Accepted: 5 February 2022 / Published online: 27 February 2022

© The Author(s) under exclusive licence to International Union for Physical and Engineering Sciences in Medicine (IUPESM) 2022

## Abstract

Present problem of every country nowadays is not to defeat other countries in a war and protect its citizens, rather to defeat the deadly diseases in their own country, which are paralyzing the strength of the countries by their adverse effects. One of the major threatening diseases among them is cancer. According to latest reports given by WHO, nearly 2.3 million women were diagnosed by breast cancer in year 2020 which resulted in 685,000 deaths worldwide and this graph was highly ascended till the end of 2020 which reported 7.8 million women (alive) tested positive with breast cancer in the past 5 years which made the Breast cancer as the most prevalent cancer among all other cancers. Cancer is the result of changes in genes known as mutations, which are responsible for the growth of cells. Breast cancer cells typically form a tumor that can be diagnosed by either of these i.e., mammography, ultrasound, MRI, biopsy or by all which are dreadful and expensive procedures. Many expert systems are designed to help the oncologists in diagnosing the disease so as to save their time and efficiency, which can be utilized in treating the patients after diagnosis and gifting them back the crown of their life from which they were almost deprived off. Still there are some problems due to which these expert systems are not fully utilized for which they are meant. First problem is that these expert systems are limited to cities because of which villagers are deprived of this facility. As we know, even after 75 years of independence, women of village area are not given importance and so is their health. Second and most prevalent problem seen is the complicated and dreadful diagnosis procedure, which threatens women the most. To solve first problem, we have taken the decision to implement this system in villages after increasing its accuracy more, which will be convenient and easy for villagers. For making diagnosis phase comfortable for women, this research has opted the way which will not frighten the women, that is just to have a blood analysis by giving blood sample which can be easily undertaken without any panic, adding up a feather of novelty to our research. This work is surely going to help the oncologist, as the work has been done under the guidance of oncologist and rules of fuzzy logic are designed as per the information shared by oncologist. This paper intends to talk about the expert system that diagnosis the breast cancer using the values of nine parameters taken from Coimbra breast cancer dataset of UCI machine learning repository such as age, BMI, glucose, insulin, Homa-IR, leptin, adiponectin, Mcp-1 and resistin and gives the output as benign or malignant. This expert system is based on fuzzy logic rules which use mamdani interface system having fuzzifier for fuzzifying the data values of parameters received from simple blood analysis, memory is used to store rules which predict the output and defuzzifier gives the output as benign or malignant with promising accuracy, sensitivity and specificity as 90.3%, 87.3% and 95% respectively. To our knowledge till now no such expert system is designed with all these attributes giving such high accuracy. It is easy to go process of diagnosing the disease with promising accuracy which will attract the women with symptoms to get it diagnosed at early stages thus help in reducing the mortality rates to a large fag end.

**Keywords** Metastasize · Biopsy · Fuzzy logic rules · Mamdani Interface System · Fuzzifier · Defuzzifier

## 1 Introduction

Cancer is the result of changes in genes known as mutations, which are responsible for the growth of cells. These mutations boost up the mitosis process in invasive cells and may

spread to adjoining parts of body called metastasized, which become the major cause of death from cancer. Breast cancer, the second foremost reason of deaths in women, is the disease, women fear the most even during its complex diagnosis phase. Breast cancer may hit the women at any age; in fact, its prevalence seems to increase significantly at the late age.

Breast cancer develops as a consequence of unusual growth of cells in breast. This unusual growth in breast is generally initiated either in glands that make the milk

✉ Tanmay Kasbe  
tanmay.kasbe@gmail.com

Extended author information available on the last page of the article

(lobular carcinoma) or in ducts that cart the milk to nipple (called ductal carcinoma). These may march into other component of the body all the way through blood via lymph nodes. Breast cancer is seen in men and women both but it is predominant in men with less than 1% cases [1–4].

According to latest reports given by WHO, nearly 2.3 million women were diagnosed by breast cancer in year 2020 which resulted in 685,000 deaths worldwide and this graph was highly ascended till the end of 2020 which reported 7.8 million women (alive) tested positive with breast cancer in the past 5 years which made the Breast cancer as the most prevalent cancer among all other cancers [5]. According to 2015 census given by WHO the most common types of cancer which slaughter women are breast, lung colorectal, cervical and stomach (given in increasing level of frequency) where breast cancer holding the first position. Breast cancer contributes to more lost Disability adjusted life years than any other cancers in women [5, 6]. DALYs (Disability life adjusted years) is the sum of potential life lost due to early death and the years of dynamic life vanished due to disability [7].

Some really threatening facts came out in the year of 2018, in which breast cancer accounted for 27.7% among all new detected cancer cases which means one in four detected women cancer patients was of breast cancer and the same accounted for 23.5% among all cancer deaths in India which means one patient died from every four Breast cancer diagnosed patients. These facts can be represented graphically (see Figs. 1 and 2).

The above two figures show the breast cancer incidence and mortality in the year 2018, where incidence means number of women newly detected while mortality means number of

women died due to breast cancer. 162,468 were total women detected with breast cancer in India and 87,090 died due to it.

Latest WHO reports update the statistics on breast cancer percentage in different age groups as shown (Figs. 3, 4, 5):

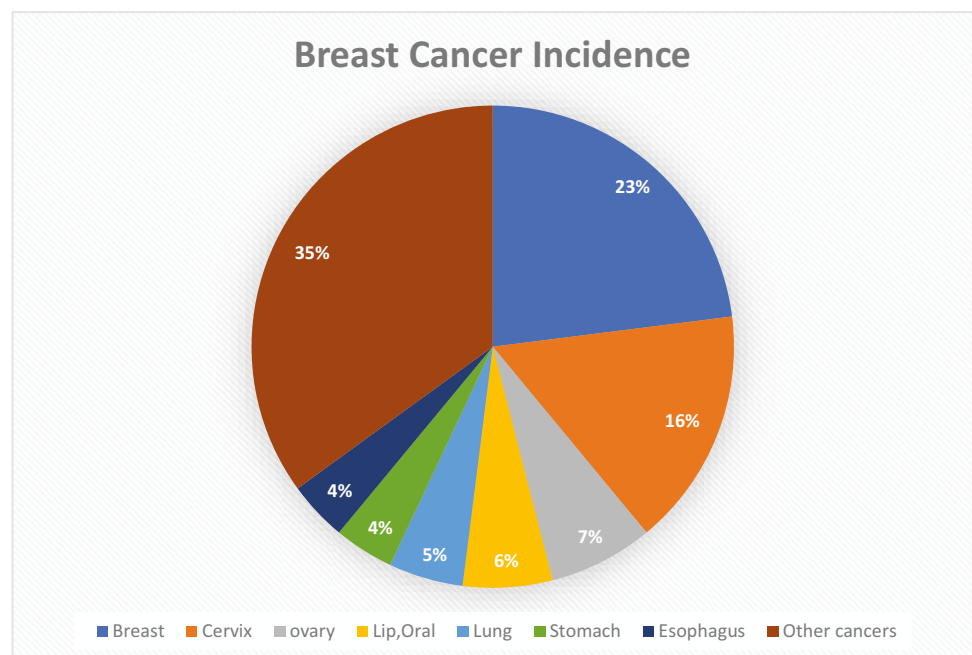
According to the Latest updates, WHO estimated the number of incident and death cases in US and India and they can be compared and roughly calculated with the given graphs of WHO [8].

The mortality ratios of India are far high then US as compared from the given graphs, the major reason behind this is late presentation due to lack of awareness, shyness, social shame, financial problems, lack of facilities or cancer centers or afraid of fearful detection procedures or tests [9–13]. To muddle through with this problem and to reduce the death rate graph of India as well work should be done upon the above points and working on the above points can be achieved by making women more comfortable with the test procedures and for the same this research was conducted to help the government in reducing the death rates of breast cancer patients.

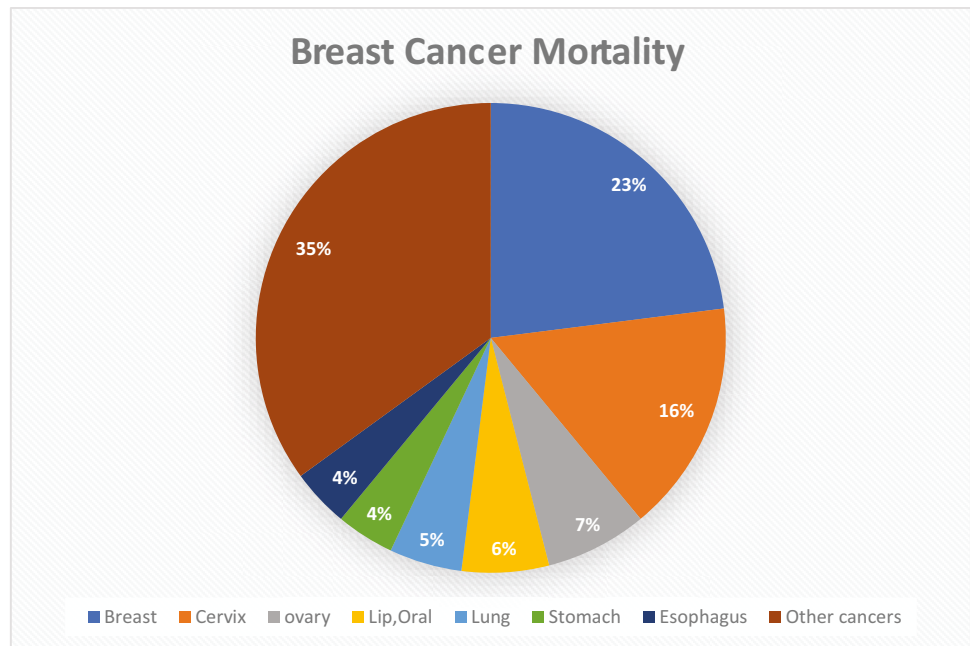
In today's scientific era cure of any deadly disease is possible only if it is detected beforehand which can shrink the transience ratio to a large degree. Breast cancer, which was considered as a fatal disease once upon a time, can be treated successfully with 100% surety in today's era but for that timely diagnosis is necessary and for that, timely tests are required which will help the physicians to take proper decisions [1–4, 7, 14].

After surfing on different researches done on disease diagnosis so far it was felt that it requires the extensive knowledge of the disease as well as factual and heuristic knowledge to develop such expert systems [15–17]. Mamdani system works well with medical diagnosis tool because of its intuitive nature which has easy to understand rule bases. Mamdani fuzzy

**Fig. 1** Chart depicting the percentage of Breast Cancer occurrence among all other cancers [Self-made]



**Fig. 2** Chart depicting the percentage of Breast Cancer mortality among all other cancers [Self-made]



interface system was proposed by Ebbasim Mamdani 1975. Basically it was designed for controlling the combination of steam engine and boiler by drafting the set of rules on direction given by people working on it. Many expert systems are introduced till now which use fuzzy logic concepts and high level of struggle is seen in achieving the accuracy of these expert systems but still it requires higher accuracy on this sphere [15, 18, 19].

This paper focuses on fuzzy inference engine which predicts the value of output vector after analyzing the set of rules on the input vector. Vital module of fuzzy logic system is fuzzy inference system which is well known for decision making. IF\_ THEN rules along with connectors ‘OR’ or

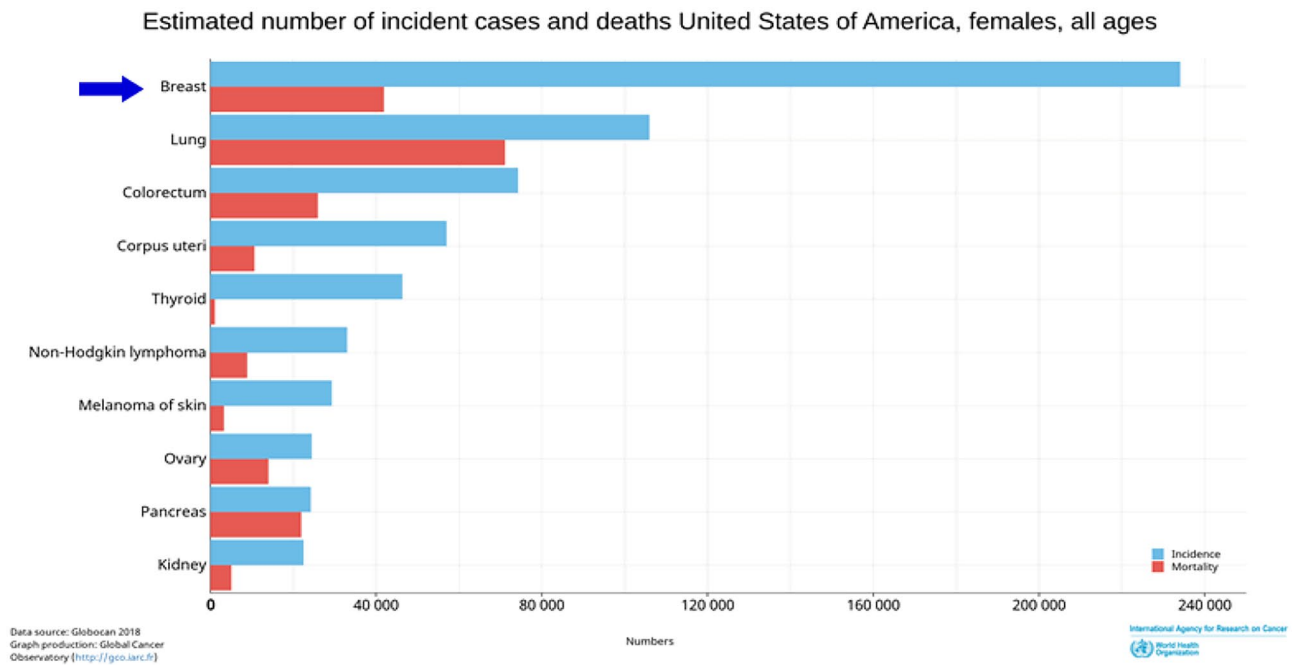
‘AND’ are used for drafting the rules. FIS works with some components such as Rule base, database, Decision making unit, Fuzzification interface unit and Defuzzification interface unit. Rule base contains IF\_ THEN rules. Database characterizes membership functions of fuzzy sets used in fuzzy rules. Decision making unit drive on rules. Fuzzification interface unit translate crisp value into fuzzy and defuzzification interface unit performs the vice versa operation [18–20].

Traditional thinking about breast cancer gives Goosebumps to almost every person in this world, as the procedure to step forward for the results are very complex, fearful and expensive [11–13].

The aim and objective of this research is to find out some other way to get to a proper result in a fearless manner with reasonable accuracy with which women will be more comfortable and will not ignore or fear the procedure of testing. With this awareness, it would be easy to diagnose the disease at early stage, as soon as the symptoms become visible or realized and can be treated early before metastatic stage and thus can get control of symptoms getting worsen which can put life at high risk. It is also designed with the goal to help oncologists and physicians in urban areas as well as in the rural areas to diagnose the disease using simple tests and thus saving their time and if utilized this time for treatment then we are not far away from the day when we will have complete control over this disease just like a simple disease [21–27]. To great extent we were able to achieve our target i.e. to find a fearless procedure which is different from uneasy and expensive methods like mammography, biopsy, MRI etc. [20]. Procedure adopted by us is based on blood analysis report which gives us the levels of leptin, adiponectin, glucose, insulin, resistin, MCP-1,



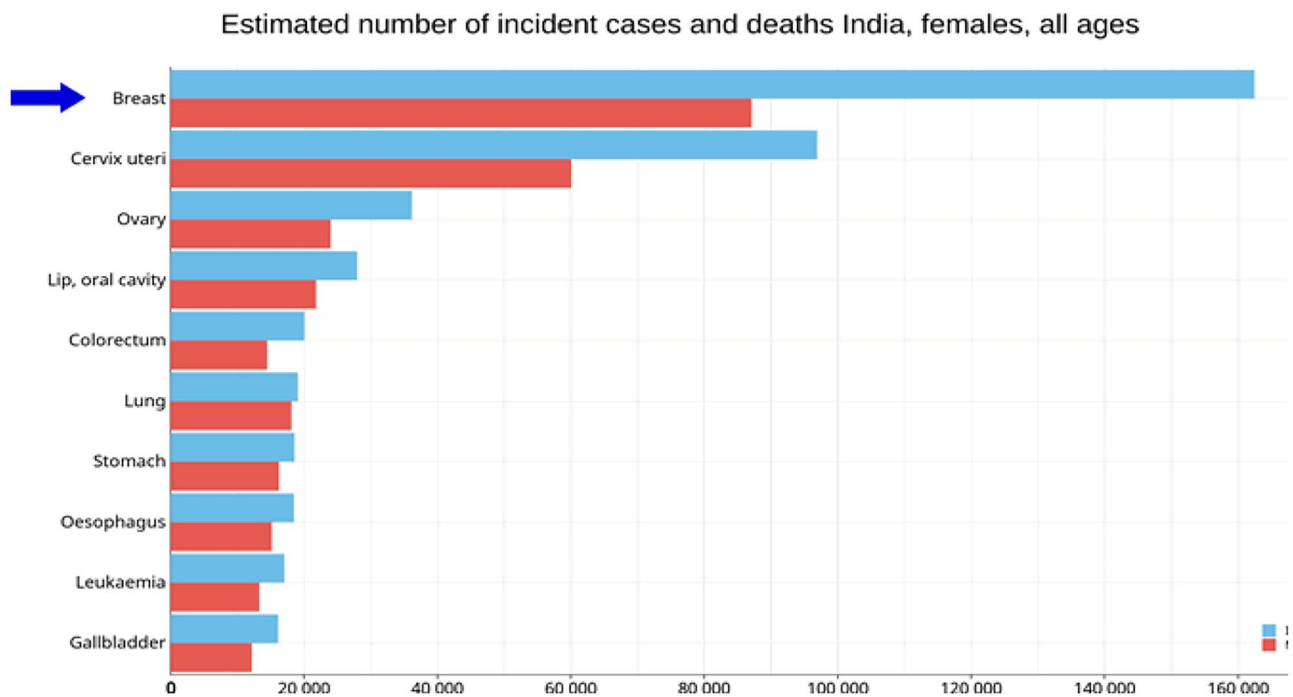
**Fig. 3** Breast Cancer occurrences in different age groups [Self-made]



**Fig. 4** Estimated number of incident cases and deaths due to Breast Cancer in US report given by Globocan 2018 [43]

Homa-IR after calculating the BMI including the age [24, 26, 28, 29]. As the patients diagnosed with a disease have different levels of these attributes in blood plasma than controls [27–35]. Considering this theory, the system was designed which can diagnose the cells as malignant or benign.

The remaining section of paper organized is as follows: Section II covers the related work of researchers with both datasets i.e., Wisconsin and Coimbra datasets of UCI machine learning repository, to understand the pros and cons of each and select the better one for our research. Section III gives the detail



**Fig. 5** Estimated number of incident cases and deaths due to Breast Cancer in India report given by Globocan 2018 [43]



description of proposed methodology which gives the overview of dataset, attributes with their ranges and membership functions for input and output attributes. Next in sequence is fuzzy rule-based system which shows some of the rules which are followed by performance analysis which shows the result along with comparison of similar works done so far and conclusion and future works and lastly the references.

## 2 Related work

In the year 2012, Saritas [14] developed an artificial neural network to detect breast cancer and its type using ANN and BI-RADS using the parameters as age, mass shape, mass border and mass density. Data of 800 patients was used, disease calculation rate was 90.5%. ANN model was prompt, consistent and threat free therefore can be used by physicians. Punitha et al. [1] planned for IABC- EMBOT layout as an effort for pull off a well again exactness pace during for the detection of this syndrome. Average accuracy attained was 97.5%, sensitivity up to 96.5% and specificity up to 97%. Srinivasan- Gopalakrishnan [2] proposed research which outfits a novel flexible antenna for recognition of breast cancer in the early hours. With reference applications, expenditure and time to the designed antenna can be used.

Diz J et al. [3] studied and compared two Portuguese breast cancer datasets and predicted malignancy and dense breast. With the paper Results from previous works were improved by grouping classes. Adam et al. [7] used the logic of Sparse Auto Encoders and Softmax Regression for feature ensemble learning to categorize of syndrome into benevolent and malignant and for this purpose used UCI WDBC dataset. Veracity achieved was 98.2% via tenfold Cross validation. Sensitivity and specificity shown by the projected method was 97.19% and 99.71% respectively.

Krishnan et al. [36] aimed at designing a SVM based classifier for breast cancer detection by considering two datasets from UCI machine learning database. SVM shows high classification accuracy as well as high sensitivity and specificity, as a result biopsy can be avoided and disease can be diagnosed. Hsu et al. [37] designed a model which can serve as unmask the disease in low cost using the dataset collected from taipei city hospital consisting of 3976 records. It attained 14.87% specificity, 2.9% precision and 100% sensitivity. Zheng et al. [9] objected the research which was based on the sample of tumor part extracted using a combination of K means and support vector machine on WBCD from UCI machine repository. This combination improves the accuracy to 97.38%. Karabatak M-Ince [11] worked on automated identification system which trusted in association rules (AR) and neural networks (NN) for exposure of patients. It used threefold cross validation method to WBCD and found that correct classification rate by this method was 95.6%.

Ramya Devi-Anandhamala [12] done the audit mainly to target the segmented ROI to check the raise in heat as weigh against the neighboring and contra lateral sides in breast

thermograms to become aware of oddity including cancer, SVM and prime element investigation were tested on dataset and results show that proposed methodology is capable of detecting breast abnormalities with 95% efficiency and 92.3% precision, Result was acquired from a set of sixty images with 35 controls and 25 uncontrolled thermograms using SVM-RBF classifier.

Eltoukhy et al. [13] exercised curvelet transform to predict the disease from digital mammograms. Supervised classifier is created using the Euclidean distance which resulted in 98.59% cataloging accuracy rate. Houby- Enas [38] a proposal of diagnosing medical images for cancer diagnosis was projected by using 5ttttt of CAD systems. It used Ant Colony Optimization (ACO) and classifiers like Decision Tree, K-nearest neighbor, Naïve Bayes and SVM and gave the results as 96.25% accuracy 97.3% sensitivity and 95.35% specificity.

In the year 2008, Neshat et al. [39] devised a fuzzy system for erudition, investigation and identification of liver 5ldschaos. Statistics from UCI were used which utilized 6 markers as intake constraints and 345 records were tested which resulted in 91% accuracy. The system is more reliable, accurate, faster and cheaper than other traditional diagnostic systems. Sayed et al. [40]. proposed a study in 2017, that uses Breast cancer Wisconsin (Prognostic) Dataset for training and testing purpose by implementing holo entropy enable decision tree (HDT) by categorizing each case of breast cancer expansion as persistent or non-persistent. Accuracy of the model changes as the new data is arrived. Chowdhury e al [15]. helped the medical diagnosis by presenting the digital fuzzy logic circuit on a chip. The design proposed the 8 T full adder to play down the setback of the course and achieved the accuracy of 97.5% by applying Bayesian analysis. Lahsasna et al. [16] reflected on accuracy and transparency to discover CHD through fuzzy rule-based system. Accuracy is improved by employing an Ensemble of Classifiers Strategy (ECS).

Ubeyli [17] proposed Adaptive Neuro Fuzzy Inference System model for revealing the disease. The model pooled the neural network adaptive potential and the fuzzy logic qualitative practice on WBCD from UCI machine learning repository based on 9 parameters and achieved a high accuracy which can be fruitful for using this model in diagnosing the disease.

In the year 2017, Kasbe -Pippal [18] adopted the MATLAB tool to analyze risk of abnormalities related to heart by employing a fuzzy expert system with 13 input parameters and 1 output parameter. Proposed system achieved the accuracy of 93.3% using the database of V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. Badid—Baba Ahmed [20] studied and declared that the breast cancer patients are seen with higher levels of insulin, glucose, leptin and triglycerides. Reduction in amount of plasma total antioxidant status (ORAC) was seen, whereas plasma hydro peroxide and carbonyl protein levels increased in patients than in controls. In the year of 2020, Thani -Kasbe [19] worked on different literatures and finally specified the causes, symptoms, diagnosis and types of breast

cancer so as to help the fresher to gain the complete and reliable knowledge on this deadly disease under a single umbrella.

Maria Del -Jose de [41] done a cross survey on 156 women after taking their written consent, divided them in two groups 78 with obesity and suffering from breast cancer another 78 obese without breast cancer. Biochemical variables of both the groups such as glucose, insulin, leptin, triglycerides-, high- and low-density lipoproteins, cholesterol and HOMA –IR were measured and found that Serum leptin levels and leptin/BMI ratio was more in first group i.e., Obese with breast cancer than in second group. Da Chung -Yueh Fang [42] analyzed the companionship between serum levels of adiponectin and leptin in two groups of 100 recently diagnosed with malignant tumors and 100 with benign and found that serum levels of adiponectin in patients with malignant tumors was less than in controls while serum level of leptin was more in patients than in controls. Also, it showed that high L/A (leptin/adiponectin) ratio be a sign of the charisma of destructive breast cancer.

Grossmann et al. [43] analyzed that high serum levels of leptin promote the growth of cell explosion and adiponectin show evidence of anti-explosion actions. Obese women with low adiponectin levels and high L/A ratio are most prone to breast cancer. Crisostomo et al. [21] investigated the blood adipokine levels of Portuguese women divided in 4 groups depending on their BMI and presence or absence of breast cancer and a major increase was seen in glucose, leptin, Homa, insulin, resistin and MCP 1 levels of obese women with more BMI while adiponectin levels were decreased in the women of same group.

Ray [44] related the obesity with breast cancer by giving the fact that obesity helps in the development of the disease. It was seen that obese women with breast cancer have less chances of recovery. Saxena—Sharma [22] realized that adipokine leptin aid tumor cells which are shaped by adipocytes and by intermingling with peril cells it relieves in tumor maturity fruition directly. Indirectly it influences various components of tumor microenvironment resulting in fast growth of cells, low self-death, attainment of mesenchymal phenotype, vast spreading of cells and attack of tumor cells.

Engin [24] in 2017, discovered that estrogen and progesterone hormones are secreted in large amount in obese women which become the cause of breast cancer other factors which encourage the growth of cancer cells due to high BMI are aromatization activity of the adipose tissue, overexpression of pro-inflammatory cytokines, insulin resistance, hyperactivation of insulin-like growth factors (IGFs) pathways, adipocyte-derived adipokines, hypercholesterolemia and excessive oxidative stress. High risk factors are seen in postmenopausal women with less BMI. furthermore, it was seen that high cholesterol levels in plasma exaggerates the tumor cells.

Healy et al. [25] investigated the data of newly diagnosed postmenopausal breast cancer patients of St James's Hospital, Dublin and found that these patients have in common the obesity and metabolic syndrome which makes the cells more aggressive.

Ines et al. [26] opine those high levels of leptin are assembled by adipocytes which inversely help in blooming of the tissue. In fact, it triggers multiple signaling pathways by chaining itself with its own receptor's leptin binds to its own receptor exhibited by neoplastic cells, stromal components, immune cells, endothelial cells and cancer-associated fibroblasts, which successively show the way to an enlarged cell voyage and make them more unpleasant. Locomotion of persistent cells and their metastatic migration to remote organs such as liver, lung, bone and brain is accelerated by Epithelial-Mesenchymal Transition (EMT), Matrix Metalloproteinase (MMP) activity, Breast Cancer Stem Cell (BCSC) formation and maintenance, angiogenesis and recruitment of immune cells. Pichard et al. [27] suggested that decreased body weight lowers the postmenopausal breast cancer risk. Women should be encouraged to lose weight so as to reduce the menace.

Alokail et al. [30] A study on 101 subjects was done, ELISA was used to quantify the serum levels of IL-6, TNF- $\alpha$ , C reactive protein, leptin, TGF- $\alpha$ , adiponectin and insulin. Obese group was detected with considerably increased levels of IL-6, C reactive protein, leptin and extensively decreased levels of adiponectin, while the levels of TNF- $\alpha$  and TGF- $\alpha$  were untouched. The same group was also bound with a strong connection between waist circumference and IL-6. Similarly, BMI and C reactive protein, waist and hip circumferences were also linked in the same group. Breast cancer hazard was also exposed by various predictors given by stepwise multiple linear regressions.

Saxena and Sharma [28] inspected the participation of leptin in breast cancer. It is promising dominant aspirant which couples' obesity and breast cancer. The impact of leptin can be seen on all stages of breast tumor from stage 0 to stage 4 in one or the other way. Transposition of tumor's microenvironment is seen because of movement of endothelial cells, neo-angiogenesis, recruitment of macrophage and monocytes. Studies have shown that hyperactive leptin results in enhanced proliferation, decreased apoptosis, acquisition of mesenchymal phenotype, potentiated migration and enhanced invasion potential of tumor cells.

Nyante et al. [32] investigated the companionship of basal-like category of breast cancer with adiponectin (ADIPOQ) detected with single nucleotide polymorphisms (SNPs), leptin (LEP), and the leptin receptor (LEPR). These results recommend that subtype-specific effects of breast cancer may be associated with genetic variation in LEPR.

Flores-Lopez and Martínez-Hernandez et al. [29] worked on different data and wrap up that breast cancer cell proliferation, migration and invasion can be promoted by hyperglycemia and hyperinsulinemia. Elevated mesenchymal markers vimentin, fibronectin, uPA expression were spotted in these features and Procedure was stimulated by ROS.

In the year 2011, La Vecchia et al. [34] evaluated that differentiated adipocytes or their precursor cells effect the growth of breast cancer by treating upon adipocyte conditioned media. In obese individuals, cancer cell growth can be controlled by

IGF-1 which is released by adipocytes and these adipocytes are regulated by glucose and fatty acids. This theory was the deducted from the use of formulation and clemency of cytokines and growth factors in real time RT-PCR and ELISA.

Viedma et al. [35] in year 2018 worked on high glucose levels treated with insulin on proliferation, migration, invasion, epithelial mesenchymal transition (EMT), and plasminogen activation system and then studied the effect of e-aminocaproic acid (EACA) in metastatic MDA-MB-231 cells.

Creation and chain of tumors by hyperglycemia and hyperinsulinemia can be demolished by controlling the tethering of plasminogen to the cell surface.

Micheal O.A et al. [45] proposed two hybrid methods to take the edge off dimensions for malaria vector. First with combination of genetic algorithm with principal component analysis which uses KNN classifier (GAO + PCA + KNN) and another with combination of genetic algorithm with independent component analysis using KNN classifier (GAO + ICA + KNN). Both giving different results in some criteria and similar results in other. Accuracy given first and second methods is 88.3% and 90% respectively. Sensitivity and specificity is predicted same by both the models such as 100% and 52.4% respectively. Precision noted by first model is 79.6% whereas by second model it is 86.7%. Marion et al. [46] carried out a study for diagnosis and prognosis of malaria ailment in humans by using SVM classification of GA filtering functional model followed by model and classification procedure of machine learning to lessen the dimensionality using *Anopheles gambiae* dataset which achieved the success up to 93 and 96 percentage.

Micheal O.A. et al. [47] used the KNN and decision tree algorithm to evaluate the result after feature selection process to summon up the pertinent information from RNA-seq mosquito *Anopheles gambiae* malaria vector dataset and achieved the exactness of 88.3% and 98.3%.

Micheal O.A. et al. [48] used ICA (Independent component analysis) for feature extraction and to cut off the dimensionality from RNA-seq which further uses ensemble classification for prognosis and diagnosis of malaria. The experiment concludes with 93.3% accuracy tested upon ICA and ensemble classification algorithm.

### 3 Materials

#### 3.1 Database and its attributes

Coimbra breast cancer dataset is used for the proposed methodology from UCI machine learning repository which was recorded by university Hospital, Coimbra from year 2009–2013 [49]. It was made available publicly on UCI

**Table 1** Input and output attributes with their ranges [21, 49, 50]

Input Attributes	Ranges	Fuzzy set values
Age	20–35	Young
	36–50	Middle
	> 50	Old
BMI	18.5–24.9	Normal
	25–29.9	Overweight
	> = 30	Obesity
Glucose	75–100	Normal
	100–125.9	Boundary
	> 126	High
Insulin	0–9.9	Normal
	10–19.9	Boundary
	> 20	High
Homa-IR	0.5–1.4	Normal
	1–1.9	Sensitive
	> 1.9	Resistance
Leptin	5–20	Normal
	> 20	High
Adiponectin	0.1–4.9	Low
	5–37	Normal
	> 37	High
MCP-1	15.6–1000	Low
	1001–2000	High
Resistin	7–22	Normal
	23–50	High
Output Attribute	0–1	Benign
BreastCancerDiagnosis	1–2	Malignant

website later on, which provides nine predictors whose presence or absence can help in predicting breast cancer at early stages. These can be collected easily from routine blood analysis [21, 50]. Tables 1 and 2 presents all the predictors with their difference ranges and fuzzy set values.

### 4 Methods

Fuzzy logic rules were used for prediction and confusion matrix algorithm was used to determine the performance of system. The latter part of this section covers the *Methodology* adopted as well as the flowchart to demonstrate the steps involved in 4.1, *Simulation work* which gives light on which tool and toolbox, used in implementation as well as the detail description of membership functions used for input and output attributes from toolbox and *Fuzzy data rule-based system*.

**Table 2** Sample data tested of Coimbra breast cancer dataset

Age	BMI	Glucose	Insulin	Homa	Leptin	Adiponectin	MCP1	Resistin	Our Result	Database Result
48	23.5	70	2.707	0.467409	8.8071	9.7024	417.114	7.99585	1	1
34	24.24242	92	21.699	4.924226	16.7353	21.823745	481.949	12.06534	1.5	2
64	22.22222	98	5.7	1.37788	12.1905	4.783985	395.976	13.91245	1.5	2
49	22.85446	92	3.226	0.732087	6.8317	13.67975	530.41	10.3176	0.5	1
74	28.65014	88	3.012	0.653805	31.1233	7.65222	572.401	18.35574	1.5	2
38	23.34	75	5.782	1.06967	15.26	17.95	165.02	9.35	0.5	1
36	34.17489	80	6.59	1.300427	10.2809	5.065915	581.313	15.72187	0.5	1
36	28.57668	86	4.345	0.921719	15.1248	8.6	534.224	9.1539	0.5	1
69	21.51386	112	6.683	1.84629	32.58	4.138025	713.239	15.69876	1.5	2
45	29.38476	90	4.713	1.046286	23.8479	6.644245	621.273	15.55625	2	2

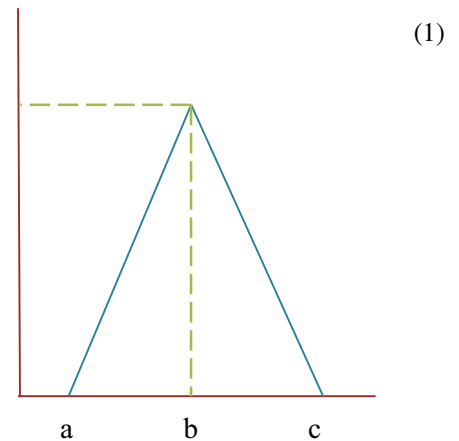
### 4.1 Methodology

Proposed methodology completed its task in different phases namely selecting the database from UCI machine learning repository, selecting the attributes and finalizing the value of attributes from authentic sources, drawing the ranges of attributes using membership functions, designing the rules, analyzing the performing and checking the result. This can be easily understood with the help of flowchart given below (Fig. 6).

$$\mu_A(x) = \begin{cases} 0 & x \leq a & (1) \\ \frac{x-a}{m-a} & a < x \leq m & (2) \\ \frac{b-x}{b-m} & m < x < b & (3) \\ 0 & x \geq b & (4) \end{cases}$$

### 4.2 Simulation work

For practical implementation of simulation work we have used MATLAB (Matrix laboratory) tool with fuzzy logic toolbox. Fuzzy Logic having the objective of approximate reasoning is the expansion of multivalued logic which gives exact solution. In fuzzy logic, output may have any value between 0 and 1 rather than the exact 0 or 1 which represents either yes or no in binary logic. True result in fuzzy logic describes its extent of truthfulness, instead of describing absolute yes or no, where false is represented by 0 and true is represented by 1 and the value between these gives the amount of its truthfulness. For our work we have used triangular and trapezoidal functions provided by fuzzy logic toolbox. We used Triangular and Trapezoidal functions for drawing the ranges of input attributes.



### 4.3 Triangular membership function

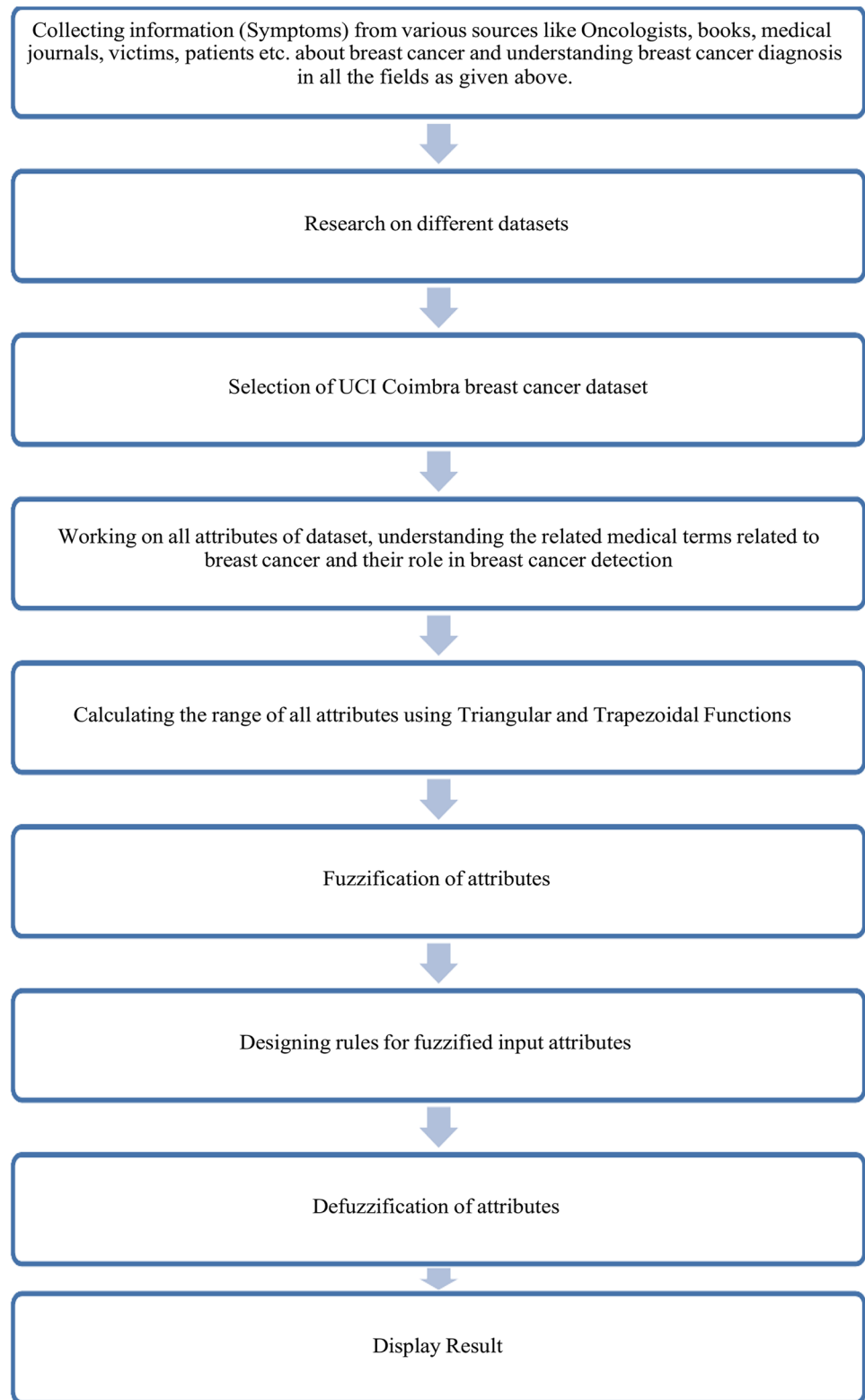
Defined by a lower limit **a**, an upper limit **b**, and a value **m**, where **a < m < b**.

### 4.4 Trapezoidal membership function

Defined by its lower limit **a**, its upper limit **d**, and the lower and upper limits of its nucleus, **b** and **c** respectively.

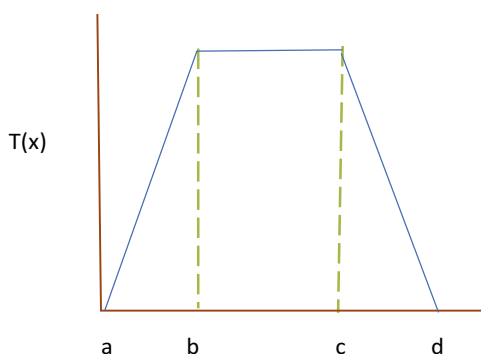


**Fig. 6** Flowchart depicting the steps involved in methodology [Self-made]



$$T(x) = \begin{cases} 0 & x \leq a \text{ or } d \leq x & (5) \\ x-a/b-a & a \leq x \leq b & (6) \\ 1 & b \leq x \leq c & (7) \\ d-x/d-c & c \leq x \leq d & (8) \end{cases}$$

(2)



### 4.5 Membership functions for input attributes

Using Triangular and trapezoidal functions, membership function ranges were plotted for input and output attributes on to membership function editor of MATLAB, selected from dataset provided by UCI have nine input attributes and one output attribute “breastCancerDiagnosis” which predicts benign if the output is in the range of 0–1 and malignancy in the range of 1–2. With the screenshots of functions from MATLAB, ranges of all the input and output attributes are shown (Figs. 7, 8, 9, 10, 11, 12):

### 4.6 Fuzzy data rule-based system

This is very important part of our proposed system as the overall performance of system depends on this part. Rules are designed to diagnose the disease as positive or negative after evaluating the impact of different attributes in patients as well as in controls. A thorough and accurate study is required from reliable sources which will help in designing the correct rules. Some of the rules from the editor window of the proposed system are shown (Fig. 13):

## 5 Result and discussion

### 5.1 Rule viewer and tested data

After designing rules, it is the time to check the correctness of rules, by supplying the values of input attributes from Coimbra breast cancer dataset. We have used 103 instances of UCI Coimbra breast cancer dataset to check the correctness by testing the values on rule viewer. The system will give the value to output attribute after evaluating all the rules. As the system is performing well, it can be used by physicians or oncologist to predict the existence or nonexistence of disease after giving the values of input attributes of patients. Figure of rule viewer on MATLAB is shown next to understand it better, followed by table with some of the tested data (Fig. 14).

### 5.2 Confusion matrix technique

Confusion matrix is an approach that encapsulates the performance of classification algorithm. It not helps in assessing the proficiency, rather it helps to focus on, at which stage and under which circumstance, the model is not performing well and becomes the torch bearer in showing the path to

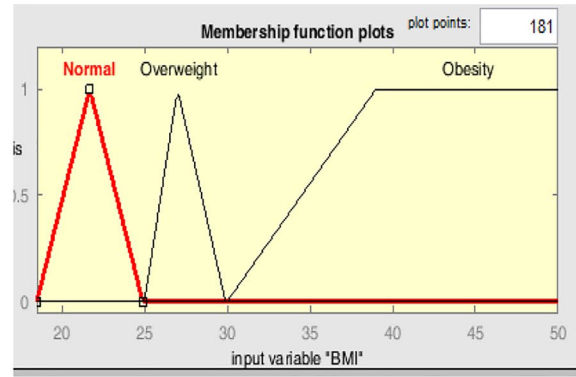
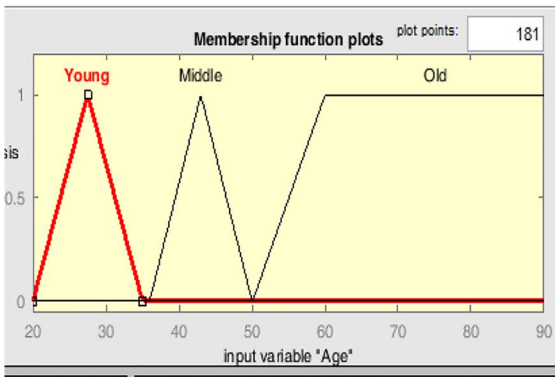


Fig. 7 Membership function for input attributes Age and BMI

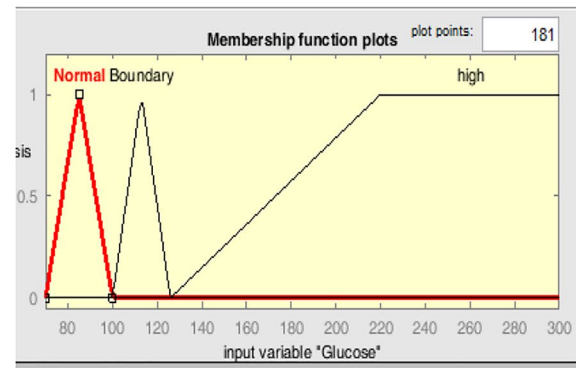
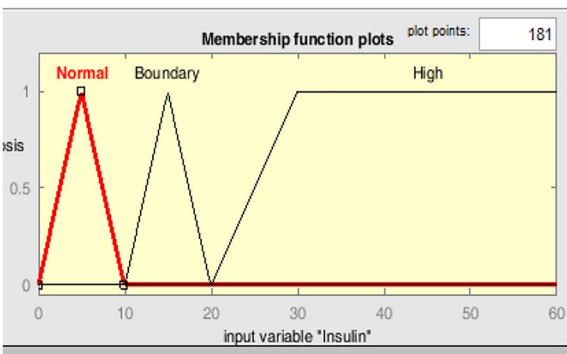


Fig. 8 Membership function for input attributes glucose and insulin

correct it. It gives the output in four forms i.e., True Positive (TP), True Negative (TN), False Positive (FP), False negative (FN) as shown in figure (Fig. 15):

True Positive (TP)-This is the case when our prediction matches with the actual value i.e., the disease exists.

True Negative (TN)-This is the case when our calculation of (disease does not exist) matches with the actual value.

False Positive (FP)-This is the case when our prediction does not match with the actual result i.e., prediction says that disease exist which is contrary to actual value.

False Negative (FN)-This is again the case when our prediction does not match with the actual result i.e., prediction says that disease does not exist which is contrary to actual values.

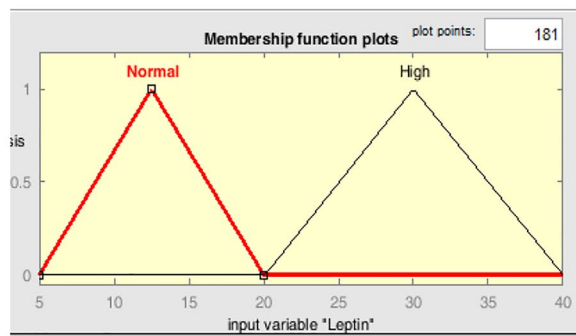
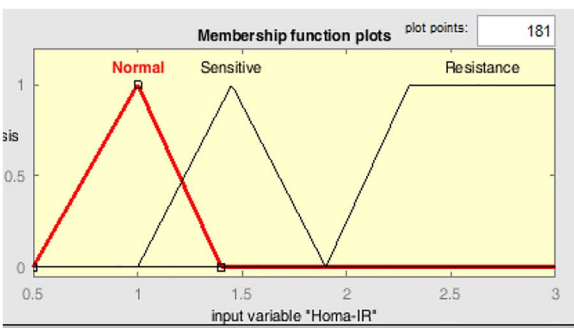


Fig. 9 Membership function for input attributes Homa-IR and Leptin

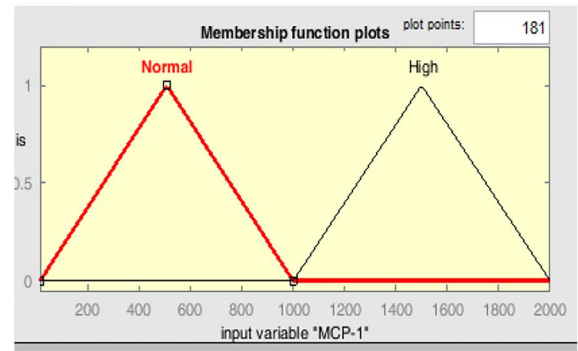
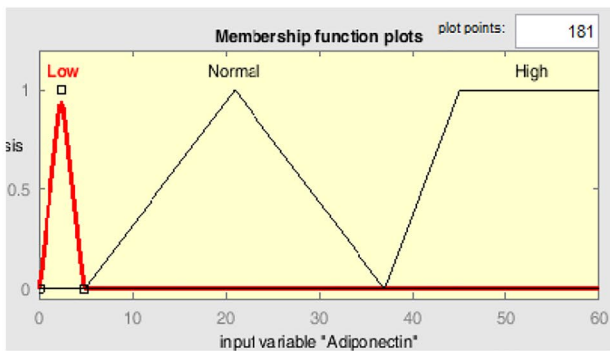
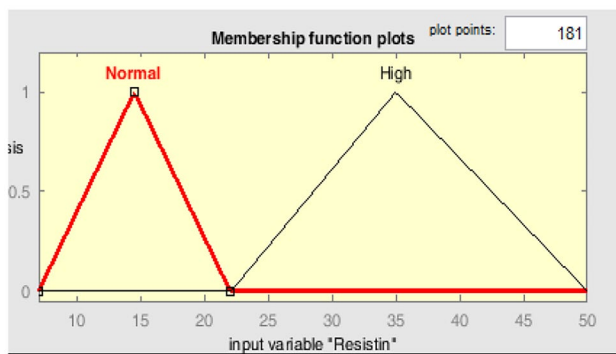


Fig. 10 Membership function for input attributes Adiponectin and MCP-1



$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (9)$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \times 100\% \quad (10)$$

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \times 100\% \quad (11)$$

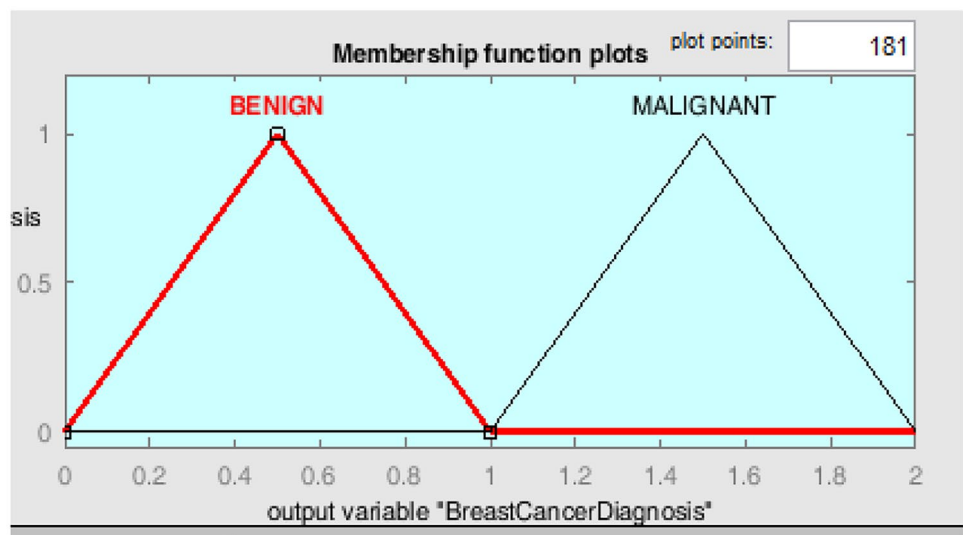
(3)

Fig. 11 Membership function for input attributes Resistin

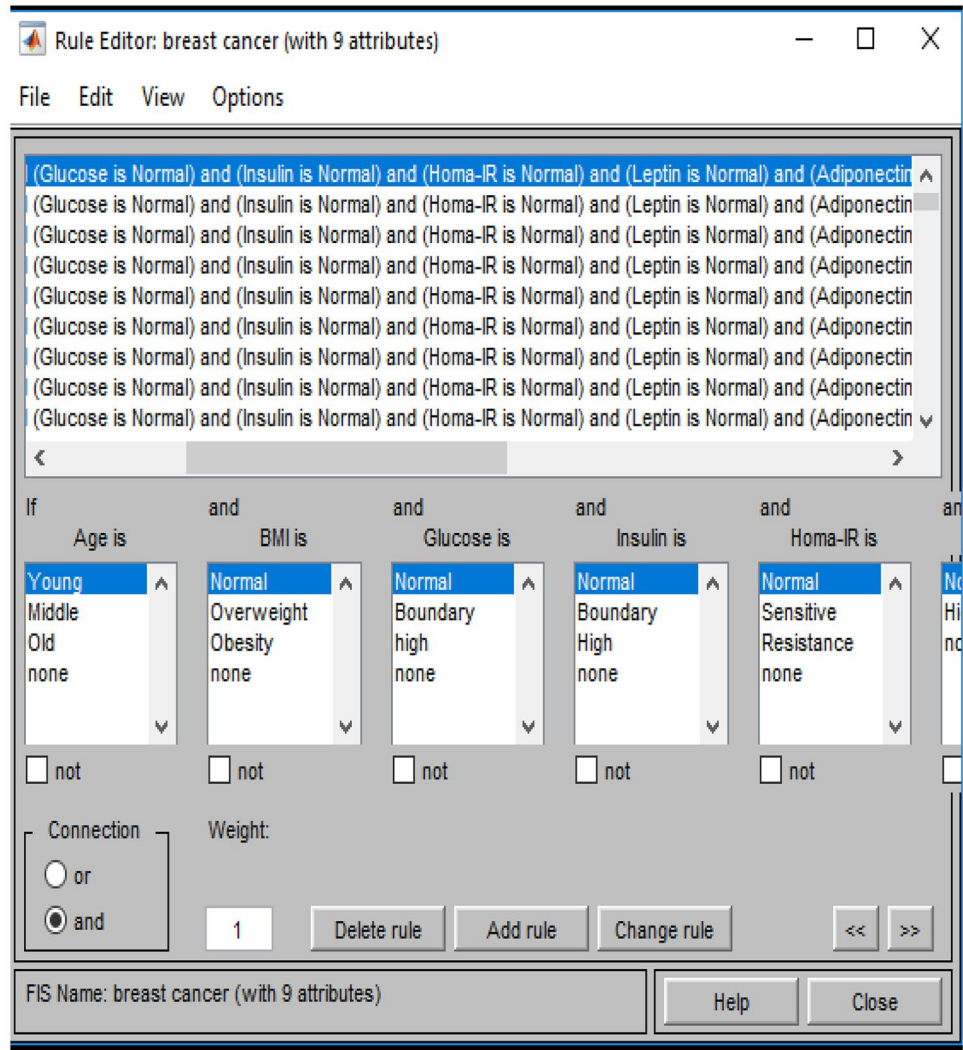
Formulas to calculate accuracy, specificity and sensitivity which aid in grading the model are given below.

Performance of our model can be scrutinized by forecasting the outcomes on application of confusion matrix technique shown by Table 3 and by applying formulas given in Eqs. 1, 2, 3, 4.

Fig. 12 Membership function for output attributes Breast Cancer Diagnosis



**Fig. 13** Rule editor window



$$\text{Accuracy} = \frac{(55+38)}{(55+38+2+8)} \times 100\% = 90.3\% \quad (9.1)$$

$$\text{Specificity} = \frac{38}{(38+2)} \times 100\% = 95\% \quad (10.1)$$

$$\text{Sensitivity} = \frac{55}{(55+8)} \times 100\% = 87.3\% \quad (11.1)$$

(4)

The result of this research shows output as benign or malignant with the promising accuracy, specificity and sensitivity as 90.3%, 87.3% and 95% respectively.

The purpose of this work is to stand in awe of women who feel very distressed during their complex diagnosis process

which can be understood only by another women. The goal of this groundwork was to design an expert system using fuzzy logic rule base which can be used as biosignature of breast cancer, based on anthropometric data and parameters obtained from routine blood analysis which will make women more comfortable with this non-threatening test. This is a step ahead of all the steps taken till now to diagnose the disease. It will help the medical experts to diagnose the disease in the initial stage thus saving their time which can be utilized in treatment procedure and control the mortality rates.

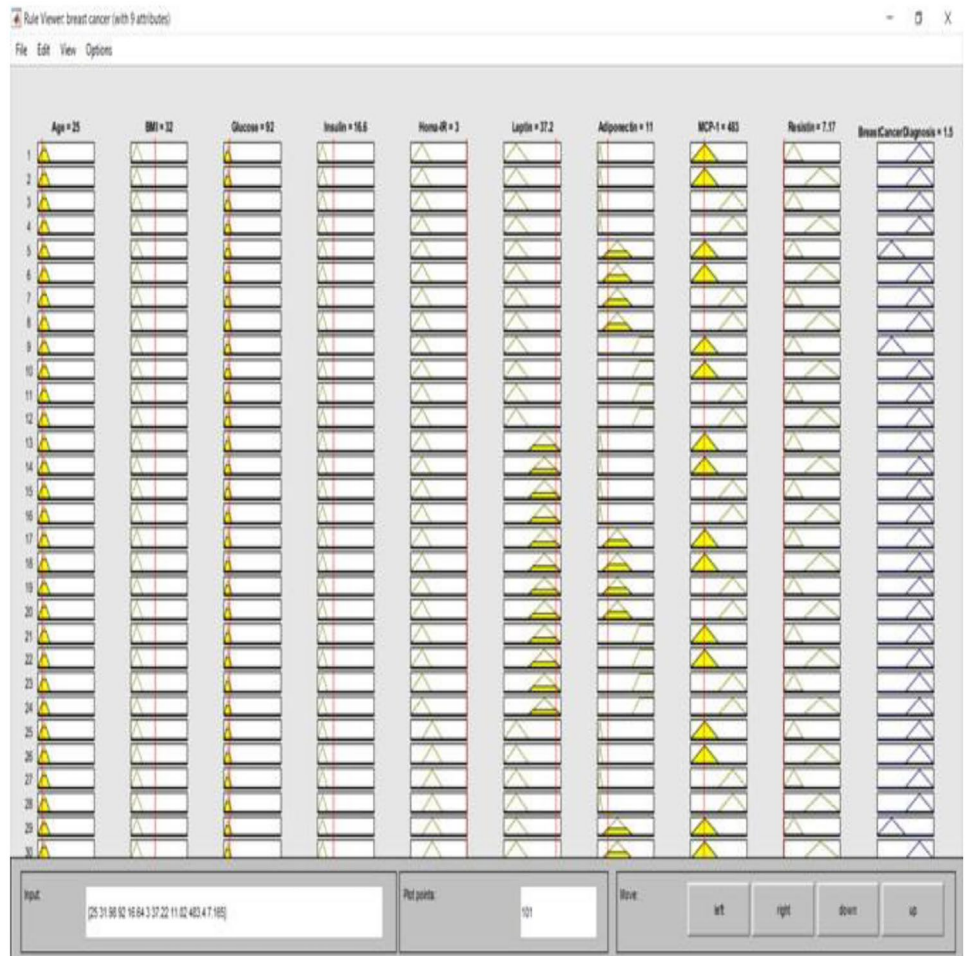
### 5.3 Performance comparison

Performance of our model is compared with different models introduced so far with the Coimbra breast cancer dataset and impressive results were exhibited shown in Table 4.

Many researchers have worked upon this domain but the novelty of our work is we have used Coimbra breast cancer Dataset which includes anthropometric data and parameters



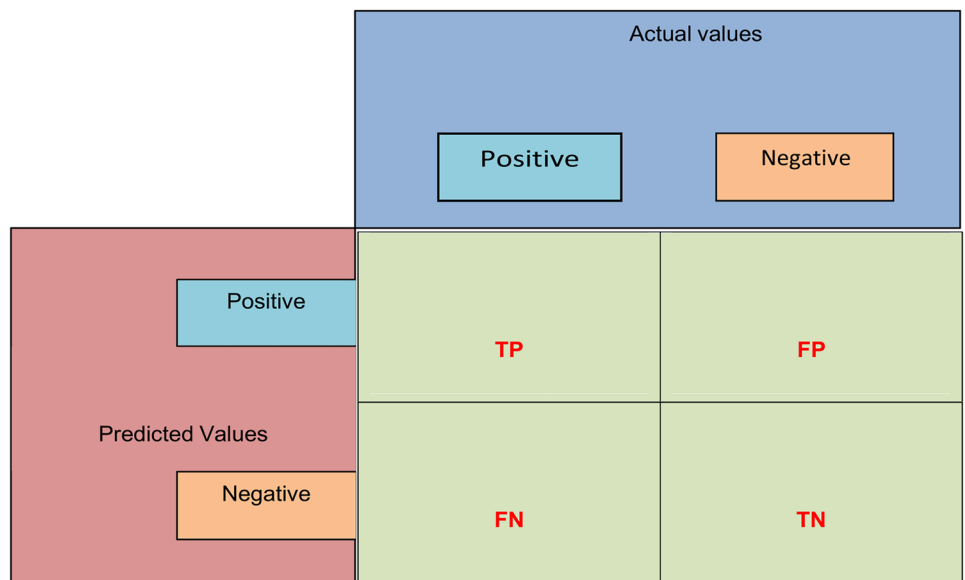
**Fig. 14** Rule viewer window to check performance evaluation



that can be obtained from routine blood analysis which will make women more comfortable with non-threatening, inexpensive test. Our expert system is based on rules given by

oncologist to diagnose the disease from the parameters of routine blood analysis making the expert system more practical to be used including very easily available parameters.

**Fig. 15** Confusion matrix representation [Self-made]



**Table 3** Confusion Matrix Measurement on Coimbra Breast Cancer dataset

Total Cases: 103	Predicted value: Yes	Predicted value: No
Actual value: Yes	55	2
Actual value: No	8	38

the disease. It will help the medical experts to diagnose the disease in the initial stage thus saving their time which can be utilized in treatment procedure and control the mortality rates.

Proposed work is done for standing in awe of women who feel very distressed during their complex diagnosis process which can be understood only by another women. The research work uses 9 inputs namely age, BMI, glucose, insulin, Homa-

**Table 4** Comparing the performance of similar models with our model

Author	Method	Accuracy (In %)	Sensitivity (In %)	Specificity (In %)
Muhammet Fatih Aslan	ELM Machine Learning	80	-	-
Akben and S.B	Decision tree Algorithm	90.5	92.2	90.4
B.K.Singh	Medium Gaussian SVM	83.6	84.7	52.9
M.M.Rahman	Machine learning tenfold	93.9	95.1	94
Patricio et al	Machine Learning Algorithm	Between 87 and 91	Between 82 and 88	Between 85 to 90
B.K.Singh	Medium KNN	92.1	95.2	88.2
<b>Proposed Study</b>	<b>Mamdani Interface System</b>	<b>90.3</b>	<b>87.3</b>	<b>95</b>

Till now no work has been done with this view to collect the parameters from blood analysis and to apply the human generated rules on that to diagnose the disease. This research is surely going to help the oncologist, as the work has been done under the guidance of oncologist and rules of fuzzy logic are designed as per the information shared by oncologist. Thus, adding a new feather in the field of research.

## 6 Conclusion and future work

Breast cancer being the deadly disease can reduce the mortality rates if diagnosed earlier. As women fear the test procedures to be carried out for diagnosing this disease, they try to ignore the symptoms at early stage. This system is developed with the aim to make women comfortable with the diagnosing procedures so that they can go for the test at early seen symptoms and get treated well.

The goal of this groundwork was to design an expert system using fuzzy logic rule base which can be used as biosignature of breast cancer, based on anthropometric data and parameters obtained from routine blood analysis which will make women more comfortable with this non-threatening test. This is a step ahead of all the steps taken till now to diagnose

IR, leptin, adiponectin, Mcp-1 and resistin and 1 output attribute which is collected from breast cancer Coimbra dataset from UCI machine learning repository. Rules are designed which identify whether the output attribute gives the result as benign or malignant with the promising accuracy, specificity and sensitivity as 90.3%, 87.3% and 95% respectively.

Future work will be to increase the accuracy and work with more datasets so as to make this system ready to be implemented and use.

**Authors' contributions** (optional—please review the submission guidelines from the journal whether statements are mandatory).

**Funding** No funding was received for conducting this study.

**Data availability** All data and materials support the published claims and comply with field standards.

**Code availability** Software application or custom code support the published claims and comply with field standards.

## Declarations

**Conflicts of interest** The authors have no relevant financial or non-financial interests to disclose.

## References

- Punitha S, Amuthan A, Joseph KS. Enhanced Monarchy Butterfly Optimization Technique for effective breast cancer diagnosis. *Journal of Medical System*. 2019;43:206. <https://doi.org/10.1007/s10916-019-1348-8>.
- Srinivasan D, Gopalakrishnan M. Breast Cancer Detection Using Adaptable Textile Antenna Design. *Journal of Medical System*. 2019;43:177. <https://doi.org/10.1007/s10916-019-1314-5>.
- Diz J, Marreiros G, Freitas A. Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. *Journal of Medical System*. 2016;40:203. <https://doi.org/10.1007/s10916-016-0561-y>.
- Hernandez -Capistran J., Martínez-Carballido J.F., Rosas-Romero R. False Positive Reduction by an Annular Model as a Set of Few Features for Microcalcification Detection to Assist Early Diagnosis of Breast Cancer. *Journal of Medical System*. 2018;42:134. <https://doi.org/10.1007/s10916-018-0989-3>.
- <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [https://www.breastcancerindia.net/statistics/stat\\_global.html](https://www.breastcancerindia.net/statistics/stat_global.html).
- Kadam VJ, Jadhav SM, Vijayakumar K. Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Autoencoders and Softmax Regression. *Journal of Medical System*. 2019;43:263. <https://doi.org/10.1007/s10916-019-1397-z>.
- Issac NS, Palanisamy P, Chibbar R, et al. An Expert Support System for Breast Cancer Diagnosis using Color Wavelet Features. *Journal of Medical System*. 2012;36:3091–102. <https://doi.org/10.1007/s10916-011-9788-9>.
- Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert System Application*. 2014;41(4):1476–82.
- Singh SP, Urooj S. An Improved CAD System for Breast Cancer Diagnosis Based on Generalized Pseudo-Zernike Moment and Ada-DEWNN Classifier. *Journal of Medical System*. 2016;40:105. <https://doi.org/10.1007/s10916-016-0454-0>.
- Karabatak M, Ince MC. An expert system for detection of breast cancer based on association rules and neural network. *Expert System Applications*. 2009;36(2):3465–9.
- Ramya DR, Anandhamala GS. Analysis of Breast Thermograms Using Asymmetry in Infra-Mammary Curves. *Journal of Medical System*. 2019;43:146. <https://doi.org/10.1007/s10916-019-1267-8>.
- Eltoukhy MM, Faye I, Samir BB. Breast cancer diagnosis in digital mammogram using multiscale curvelet transform. *Computerized medical imaging and graphics*. Elsevier. 2010;34(4):269–76.
- Saritas I. Prediction of breast cancer using artificial neural networks. *Journal of Medical System* 2012;36:2901–2907. <https://doi.org/10.1007/s10916-011-9768-0>.
- Roy CS, Roy A, Saha H. ASIC Design of a Digital Fuzzy System on Chip for Medical Diagnostic Applications. *Journal of Medical System*. 2011;35:221–35. <https://doi.org/10.1007/s10916-009-9359-5>.
- Lahsasna A, Ainon RN, Zainuddin R, et al. Design of a Fuzzy-based Decision Support System for Coronary Heart Disease Diagnosis. *Journal of Medical System*. 2012;36:3293–306. <https://doi.org/10.1007/s10916-012-9821-7>.
- Ubeyli ED. Adaptive Neuro-Fuzzy Inference Systems for Automatic Detection of Breast cancer. *Journal of Medical System*. 2009;33:353. <https://doi.org/10.1007/s10916-008-9197-x>.
- Kasbe T, Pippal RS. Design of heart disease diagnosis system using fuzzy logic. *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. 2017;3183–3187.
- Thani I, Kasbe T. Breast Cancer: State-of-the-art causes and diagnosis, 2nd International Conference on Data. *Eng Applications (IDEA)*. 2021;1–6.
- Badid N, Baba Ahmed FZ, Merzouk H. Oxidant/Antioxidant Status, Lipids and Hormonal Profile in Overweight Women with Breast Cancer. *Pathology and Oncology Research*. 2010;16:159–67.
- Crisostomo J, Matafome P, Santos-Silva D. Hyperresistinemia and metabolic dysregulation: a risky crosstalk in obese breast cancer. *Endocrine*. 2016;53:433–42.
- Saxena NK, Sharma D. Multifaceted Leptin Network: The Molecular Connection between Obesity and Breast Cancer. *J Mammary Gland Biol Neoplasia*. 2013;18:309–20.
- Cust AE, Stocks T, Lukanova A. the influence of overweight and insulin resistance on breast cancer risk and tumour stage at diagnosis: a prospective study. *Breast Cancer Res Treat*. 2009;113:567–76.
- Engin A. Obesity-associated Breast Cancer: Analysis of risk factors. Obesity and Lipotoxicity, *Advances in Experimental Medicine and Biology*. Springer. 2017;960:571–606.
- Healy LA, Ryan AM, Carroll P, Ennis D, Crowley V, Boyle T, Kennedy MJ, Connolly E, Reynolds JV. Metabolic Syndrome, Central Obesity and Insulin Resistance are Associated with Adverse Pathological Features in Postmenopausal Breast Cancer. *Clin Oncol*. 2010;22(4):281–8.
- Ines B, Cinzia G, Daniela B, Sebastiano A, Stefania C. The weight of obesity in breast cancer progression and metastasis: Clinical and molecular perspectives. *Seminars in Cancer Bio*. 2020;60:74–284.
- Pichard C, Plu-Bureau G, Neves-e CM, Gompel A. Insulin resistance, obesity and breast cancer risk. *Maturitas*. 2008;60(1):19–30.
- Saxena NK, Sharma D. Multifaceted Leptin Network: The Molecular Connection between Obesity and Breast Cancer. *J Mammary Gland Biology and Neoplasia*. 2013;18:309–20. <https://doi.org/10.1007/s10911-013-9308-2>.
- Flores-Lopez LA, Martinez-Hernandez MG, Viedma-Rodríguez R. High glucose and insulin enhance uPA expression, ROS formation and invasiveness in breast cancer-derived cells. *Cell Oncology*. 2016;39:365–78. <https://doi.org/10.1007/s13402-016-0282-8>.
- Alokail MS, Al-Daghri NM, Al-Attas OS. Combined effects of obesity and type 2 diabetes contribute to increased breast cancer risk in premenopausal women. *Cardiovasc Diabetol*. 2009;8:33.
- Niraula S, Dowling RJO, Ennis M. Metformin in early breast cancer: a prospective window of opportunity neoadjuvant study. *Breast Cancer Res Treat*. 2012;135:821–30.
- Nyante SJ, Gammon MD, Kaufman JS. Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. *Breast Cancer Res and Treatment*. 2011;129:593–606. <https://doi.org/10.1007/s10549-011-1517-z>.
- D'Esposito V, Passaretti F, Hammarstedt A. Adipocyte-released insulin-like growth factor-1 is regulated by glucose and fatty acids and controls breast cancer cell growth in vitro. *Diabetologia*. 2012;55:2811–22. <https://doi.org/10.1007/s00125-012-2629-7>.
- La Vecchia C, Giordano SH, Hortobagyi GN, Chabner B. Overweight, obesity, diabetes, and risk of breast cancer: interlocking pieces of the puzzle. *Oncologist*. 2011;16:726–9.
- Viedma-Rodríguez R, Martínez-Hernández MG, Flores-Lopez LA. Epsilon-aminocaproic acid prevents high glucose and insulin induced-invasiveness in MDA-MB-231 breast cancer cells, modulating the plasminogen activator system. *Mol Cell Biochem*. 2018;437:65–80.
- Krishnan MMR, Banerjee S, Chakraborty C, Chakraborty C, Ray AK. Statistical analysis of mammographic features and its classification using support vector machine. *Expert system application*. 2010;37:470–8.
- Hsu J, Hung P, Lin H. Applying Under-Sampling Techniques and Cost-Sensitive Learning Methods on Risk Assessment of Breast Cancer. *Journal of Medical System*. 2015;39:40. <https://doi.org/10.1007/s10916-015-0210-x>.

38. El Houby EMF. Framework of Computer Aided Diagnosis Systems for Cancer Classification Based on Medical Images. *Journal of Medical System*. 2018;42:157. <https://doi.org/10.1007/s10916-018-1010-x>.
39. Neshat M, Yaghobi M, Naghibi MB, Esmaelzadeh A. Fuzzy Expert System Design for Diagnosis of Liver Disorders. 2008 International Symposium on Knowledge Acquisition and Modeling. Wuhan. 2008;2008:252–6.
40. Sayed S, Ahmed S, Poonia R. Holo entropy enabled decision tree classifier for breast cancer diagnosis using Wisconsin (prognostic) data set. 7th International Conference on Communication Systems and Network Technologies (CSNT), Nagpur. 2017;172–176.
41. Maria DS, R.F., Jose de J. G. G., Jesus D.M., Guadalupe M.G., Angel G.V., Jhony De la C.V. Insulin and Leptin Levels in Obese Patients with and Without Breast Cancer. *Clin Breast Cancer*. 2013;13(6):482–5.
42. Da Chung C, Yueh FC, Yao TY, Hso CC, Fu CK, Ou YF, Hue YC, Ming FH, Shyng SFY. Serum adiponectin and leptin levels in Taiwanese breast cancer patients. *Cancer Lett*. 2006;237(1):109–14.
43. Grossmann ME, Ray A, Nkhata KJ. Obesity and breast cancer: status of leptin and adiponectin in pathological processes. *Cancer Metastasis*. 2010;29:641–53.
44. Ray A. Adipokine leptin in obesity-related pathology of breast cancer. *Journal of Bioscience*. 2012;37:289–94.
45. Micheal OA, Marion OA, Ayodele AA, et al. Optimized hybrid investigative based dimensionality reduction methods for malaria vector using KNN classifier. *J Big Data*. 2021;8:29. <https://doi.org/10.1186/s40537-021-00415-z>.
46. Marion OA, Micheal OA, et al. A genetic algorithm for prediction of RNA-seq malaria vector gene expression data classification using SVM kernels. *Bulletin of Electrical Eng Informat-ics*. 2021;10(2):1071–1079. <https://doi.org/10.11591/eei.v10i2.2769>.
47. Micheal OA, Marion OA, Ayodele AA. A genetic algorithm approach for predicting ribonucleic acid sequencing data classification using KNN and decision tree. *Telecomm Computing Electron Control*. 2021;19(1):310–316. <https://doi.org/10.12928/telkonnika.v19i1.16381>.
48. Micheal OA, Marion OA, Ayodele AA. An ICA-ensemble learning approaches for prediction of RNA-seq malaria vector gene expression data classification. *Int J Electrical Comp Eng (IJECE)*. 2021;11(2):1561–69. <https://doi.org/10.11591/ijece.v11i2.pp1561-1569>.
49. Patricio M, Pereira J, Crisostomo J, et al. Using Resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer* 2018;18(29). <https://doi.org/10.1186/s12885-017-3877-1>.
50. <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Isha Thani<sup>1</sup> · Tanmay Kasbe<sup>2</sup> 

Isha Thani  
lataagnani@gmail.com

<sup>2</sup> Associate Professor, Faculty of Computer Science, Oriental University, Madhya Pradesh, Indore, India

<sup>1</sup> Faculty of Computer Science, Research Scholar, Oriental University, Indore, Madhya Pradesh, India