Tikrit Journal of Engineering Sciences (2025); 32(Sp1): 2685
DOI: http://doi.org/10.25130/tjes.sp1.2025.39

IRAQI Academic Scientific Journals

ISSN: 1813-162X (Print); 2312-7589 (Online)
Tikrit Journal of Engineering Sciences
available online at: http://www.tj-es.com

TJES Tikrit Journal of Engineering Sciences

Development and Validation of a Novel Bayesian Belief Network: A Reliable Fuzzy Weighted Diabetes Predictive Model

Shweta Kharya [a], Sunita Soni [a], Pragyan Nanda [b], Giridhar Urkudee [c], Ananya Smruti Snigdha Ojha [d], Debasish Swapnesh Kumar Nayak [e], Suprava Ranjan Laha [*f], Tripti Swarnkar [g]

a Department of Computer Science & Engineering, Bhilai Institute of Technology, Durg, India.
b Department of Information Technology, Parul University, Vadodara, Gujarat.
c Department of Computer Science & Engineering, Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India.
d Department of Computer Science and Engineering, Brainware University, Kolkata, India.
e Department of Computer Science and Engineering, Centurion University of Technology and Management, Bhubaneswar, India.
f Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be) University, Bhubaneswar, India.
g Department of Computer Applications, National Institute of Technology, Raipur, India.

Keywords:
Bayesian classifier; Fuzzy Weighted Association rule mining; Maximum likelihood estimation technique; prediction model.

Highlights:

- Introduces a Fuzzy Weighted Bayesian Belief Network for reliable diabetes prediction.
- Achieved 96.8% classification accuracy, 98.6% precision, and 97.5% recall on the Pima Indian dataset.
- Novel Fuzzy Weighted Association Rule Mining improves model interpretability.
- Validated on heart disease and breast cancer datasets, achieving up to 93.7% and 99%, respectively, of classification accuracy.

ARTICLE INFO

Article history:

Received	22 Jul. 2025
Received in revised form	25 Aug. 2025
Accepted	14 Oct. 2025
Final Proofreading	25 Dec. 2025
Available online	26 Dec. 2025

Citation: Kharya S, Soni S, Nanda P, Urkudee G, Ojha ASS, Nayak DSK, Laha SR, Swarnkar T. **Development and Validation of a Novel Bayesian Belief Network: A Reliable Fuzzy Weighted Diabetes Predictive Model**. *Tikrit Journal of Engineering Sciences* 2025; **32**(Sp1): 2685.
http://doi.org/10.25130/tjes.sp1.2025.39

*Corresponding author:
Suprava Ranjan Laha
Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be) University, Bhubaneswar, India.

Abstract: The rising burden of chronic diseases, particularly diabetes, necessitates diagnostic frameworks that can navigate the inherent ambiguity of clinical data. Conventional predictive models often struggle with the stochastic uncertainty stemming from subjective patient narratives and laboratory noise, as well as the 'black-box' lack of interpretability. To transcend these limitations, this research introduces a novel Fuzzy Weighted Bayes Association Rule Mining (FWBARM) framework. This approach integrates fuzzy logic to handle data vagueness with a weighted mechanism that implicitly learns feature importance, thereby generating robust, transparent, and clinically interpretable decision rules. The proposed system, evaluated as the 'Reliable Diabetes Prediction Model,' demonstrated superior diagnostic efficacy, achieving 96.8% accuracy, 98.6% precision, and 97.5% recall. By reconciling high predictive performance with rule transparency, this work offers a scalable solution for personalised medicine and reliable Clinical Decision Support Systems.

1.INTRODUCTION

Diabetes mellitus (DM) is a chronic disorder characterised by abnormal carbohydrate, protein, and lipid metabolism, leading to hyperglycemia [1]. It is classified into two main types: type 1 and type 2 diabetes [2]. Threats in society due to chronic diseases always exist and are ever-increasing [3]. Diabetes will become the seventh leading cause of death by 2030 [4]. However, on the positive side, a large volume of data is generated in the clinical domain [5]. Currently, in the healthcare industry, around 30% of the world's data volume is produced. However, due to the privacy policy, this clinical data is not publicly accessible. Thus, Computational Intelligence approaches focus on developing novel models, algorithms, and methods to solve real-world problems. To improve the standard of patient care, identify treatment plans and provide best-practice recommendations from standard clinical datasets available in the research world [6]. Predictive modelling plays a crucial role in the clinical domain. Modelling plays a pivotal role in clinical practice, with strong potential to improve patient care through better decision-making [7]. The Bayesian classifier is the most suitable predictive model for the clinical domain, as it captures conditional dependencies among disease symptoms [8]. Here Bayesian classifier plays crucial role is depicting and understanding the association between the symptoms In this study, to enhance the performance parameter Bayesian classifier is reconstructed to incorporate two clinical dataset characteristics, i.e. "*Fuzziness in Clinical Data,*" by eradicating the Sharp boundary in clinical data using fuzzy theory and "*Not all symptoms are equally important for prediction*" by using the weighted methods which assigns weights to the attributes according their predicting capabilities [9]. This paper proposes a novel algorithm for constructing a Fuzzy Weighted Bayesian Belief Network (FWBBN) model from medical datasets. In this study, fuzzy weight theory is integrated with a Bayesian classifier specifically for the chronic disease diabetes, yielding a novel classifier, FWBBN. The primary contributions of the proposed framework include, first, applying FL to address vagueness, sharp boundaries, and imprecision in medical attributes. Next, for each fuzzy attribute, an automated weight is assigned based on its label contribution to class-label prediction. The next step is to extract fuzzy weighted rules between two attributes, between multiple attributes, and between attributes and the class label, as Fuzzy Weighted Class Association Rules (FWAR), based on minimum threshold values for Fuzzy Weighted Support (FWS) and Fuzzy Weighted Confidence (FWC). Strong Fuzzy Weighted-class Bayesian Rules (FWBAR) rules are yielded based on Fuzzy Weighted Bayesian Confidence (FWBC) measure to build the FWBBN model that captures real-life medical situations and produces an accurate and improved classifier based on Interdependency among the medical attributes.

2.RELATED WORK

A rigorous literature review has been conducted over the past 10 years in the clinical domain, focusing on fuzzy-based classifiers that use various medical datasets from the UCI archive. Due to the Sharp Boundaries of quantitative attributes, a patient may be misclassified into a category, such as High_BloodPressure or Low_Blood_Pressure, leading to erroneous predictions in medicine and incorrect treatment in clinical practice. Fuzzy Logic (FL) is an appropriate theory for addressing the "Sharp boundary" problem. Here, Fuzzy Logic can be applied to quantitative attributes to determine the patient's partial membership in all categories. For example, consider the following discretization rule for the attribute Blood Pressure (BP), BP [80-120] →Hypertension ="Optimal", BP [84-129] →Hypertension="Normal", BP [89-139] →Hypertension="High-Normal", and BP [Above 90-Above 140] →Hypertension=" High"[10]. If, in a particular dataset record, the BP value is 141, then, according to the above discretisation rules, the patient is classified as severely hypertensive by crisp/sharp analysis, which may lead to misclassification, as the patient may also satisfy other regulations to some degree [10]. Fuzzy Logic plays an essential role in determining the extent to which the patient partially belongs to each fuzzy value of the set {Optimal, Normal, High-Normal, High}. So, to measure belongingness using the membership function, the value of the membership to the fuzzy set is like ($\mu_{(BP,"optimal")}=0.1, \mu_{(BP,"normal")}=0.2, \mu_{(BP,"high-normal")}=0.3, \mu_{(BP,"high")}=0.4$).

However, the major challenge in fuzzy association rule mining is to address the exponential growth in rules produced by fuzzy partitioning of attributes [11]. This can be resolved by incorporating attribute weight assignment into the model-building process, as the omission of attribute weights is often criticised for yielding uninteresting rules with high frequency and low importance, which lead to incorrect predictions [12]. After solving the sharp boundary problem in the data and weighting, the key is to identify the associations among the different weightings and among the attributes. Fuzzy theory can be applied to a Weighted-support significant framework that integrates quantitative and qualitative aspects of the dataset. Here, the Class Association Rule Mining method plays an important role, as it is used to discover associations between

attributes and class labels, a subgroup of Association Rule Mining. The essential findings are shown in Table 1. The literature survey offers a promising approach in the clinical domain when incorporated with various classifiers.

Table 1 A Brief Literature Review.

S.no	Refs.	Techniques	Relevant Review Findings
1	[11]	FWARM Classifier	Fuzzy theory can be used to eliminate sharp boundaries which exist in medical data.
2	[13]	Weighted Association based on classification.	Domain expert-based weight assignment.
3	[14]	Fuzzy logic	It plays a crucial role in predictive analysis and classification.
4	[15]	Modified dynamic multi-swarm optimisation.	Very efficient and adaptive prediction system. Heart Disease dataset Accuracy of 92.13%
5	[16]	Fuzzy system.	Accuracy of 96.6% on the COVID dataset.
6	[17]	Fuzzy SVM	Diabetic dataset Accuracy of 96.8%
7	[18]	Fuzzy random forest	Heart Disease dataset Accuracy of 93.4%
8	[19]	Neuro-Fuzzy classifier	Accuracy of 95.14%
9	[20]	Fuzzy Temporal Rule-Based Classifier	Breast cancer dataset. Accuracy of 99%
10	[21]	Bayesian Classifier	Diabetes UCI dataset. Accuracy of 82.5%
11	[22]	Bayesian Classifier	Diabetes's UCI dataset. Accuracy of 92.2%
12	[23]	Bayesian Belief Classifier	Breast cancer, UCI dataset. Accuracy of 97%
13	[24]	Bayesian Classifier	Breast cancer, UCI dataset. Accuracy of 96.31%
14	[25]	Bayesian Classifier	Breast cancer, UCI dataset. Accuracy of 87%
15	[26]	Bayesian Classifier	Breast cancer, UCI dataset. Accuracy of 96%

The survey also found that the most suitable classifier is the Bayesian Belief Network, which yields substantially better results when applied to clinical datasets such as Cancer, Heart disease, Diabetes, and COVID-19 data. So, an idea emerged to use the Fuzzy weighted concept with a Bayesian Belief Network to enhance the prediction rate by taking care of two important, crucial, critical, essential characteristics of clinical data, i.e. "Every symptom has a different impact rate for prediction" and "Fuzziness in medical data"studyused primarily the UCI archive's Pima Indian Diabetes Dataset (PIDD). The entire experiment is conducted primarily on medical datasets, specifically the UCI archive's Pima Indian Diabetes Dataset (PIDD), which is used to study diabetes. In this dataset, missing values are not explicitly labelled as NaN. Instead, for specific physiological measurements, a value of 0 is biologically impossible and should be treated as missing. Moreover, to handle missing values, Statistical Imputation is used. Median imputation for skewed features and mean for normally distributed ones. Additional datasets are also used for comparative analysis. The detailed distributions of the clinical datasets used in the experiments are shown in Table 2.

Table 2 Data Distribution of Various Clinical Datasets.

Data	#records	Features	Class	Positive class(%)	Negative class(%)
Breast Cancer	699	9	2	34	66
Heart Disease	303	14	2	54	46
Pima Indian Diabetes	768	9	2	35	65

The breakdown based on distribution analysis of the Pima Indians Diabetes Dataset from the UCI repository is as follows:

The features replaced with the *median* are glucose, diastolic blood pressure, skin thickness, insulin, and body mass index. Features replaced with the *mean* are diabetes pedigree function (DPF) and age, whereas pregnancies are typically left as-is (since zeros are valid values).

3.METHODOLOGY

normalisation. The method used in the proposed FWBBN model is shown in Fig. 1. It outlines the process that will be followed throughout the research for the proposed idea. Firstly, the dataset is extracted from the UCI Machine Learning Repository, which is available in normalised form and accessed via the LUCS-KDD DN software. After discretisation, the database is partitioned into fuzzy linguistic values, and the fuzzy values are computed using a trapezoidal membership function. Weights are calculated using the Maximum Likelihood Estimation Method. The next step is to generate FWAR rules using FWS and FWC. The FWBBN is built using robust rules generated as Fuzzy Weighted class Bayesian Rules (FWBAR). The detailed work is explained in the following sections. The FWBBN Hyperparameters Documentation comprises: 1. Fuzzification parameters, such as the number of fuzzy partitions per attribute, fuzzification type, and trap_a, trap_b, trap_c, and trap_d; 2. Fuzzy Rule Generation:

Hyperparameters include the minimum support threshold and the rule-pruning strategy. Fuzzy Weighting (MLE) Hyperparameters like max weight range, min weight threshold, and weight normalisation. The method of the proposed research is elaborated in the pseudocode shown in Fig. 2.

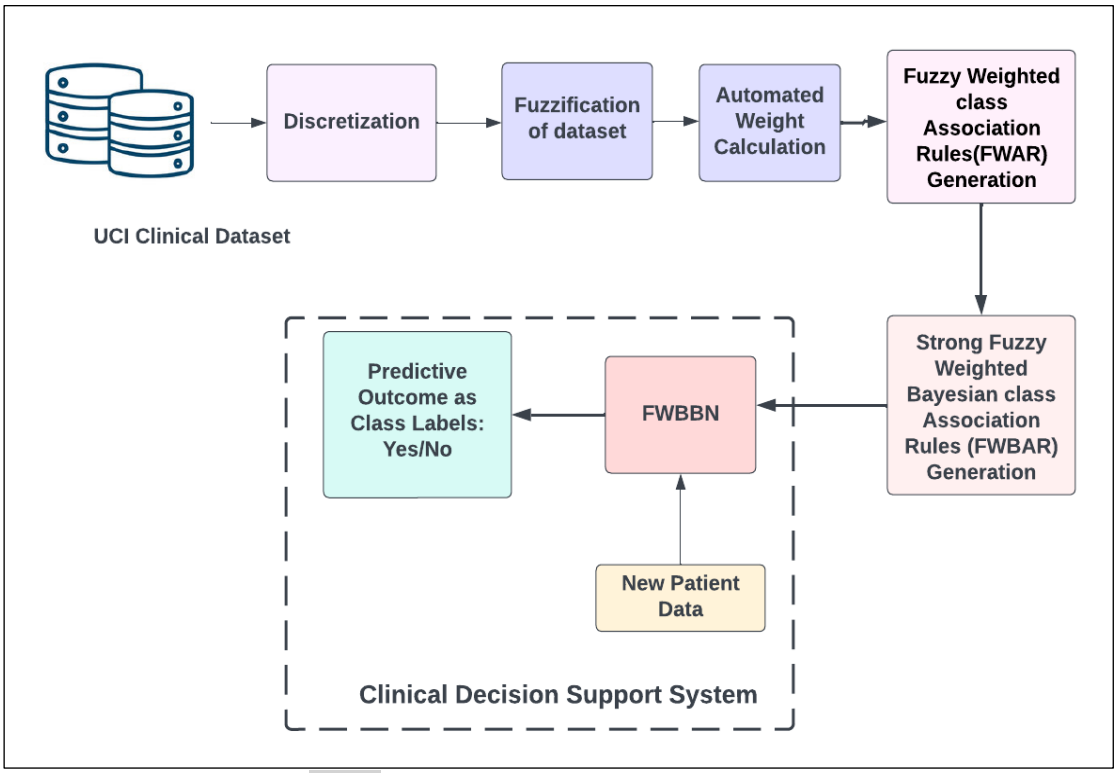

Fig. 1 Methodology of the FWBBN Model.

Algorithm: FWBAR
Input to System: The Database consists of records and attributes.
Output from System: Fuzzy Weighted Bayesian Association Strong Rules.
1. Discretise the variables of the data records given in Database D.
2 Transform Database with Fuzzy values using the Trapezoidal membership function as D1.
3. Assign weights to fuzzy attributes of Database D1.
4. Generate Fuzzy Attribute Set Weight for Database D1.
5. Calculate Fuzzy Weighted Support for two-item sets, multi-item sets, and class labels.
6. Again, Calculate Fuzzy Weighted Confidence for two items, ets, multi-item sets, and class labels.
7. Generation of strict rules.
8. For every rule, calculate Fuzzy Weighted Bayesian confidence (FWBC).
9. Construct a Bayesian Belief Network using the output rules with the highest FWBC.

Fig. 2 Fuzzy WBAR Algorithm.

3.1.ProposedModel: Fuzzy Weighted Bayesian Belief Network

Consider a Fuzzy database FD = {p_1, p_2, p_3.... p_i...p_n} with attributes set AS = (b_1, b_2, ...b_m}; each b_K relates to linguistic labels set LS={l_1, l_2, ...l_L } for example LS={high, low, moderate}. Suppose that each b_k is correlated with fuzzy set FS_k = {(b_k,l_1), (b_k,l_2), (b_k,l_3), ...(b_k,l_L)}. Consider a tuple in which each attribute bi possesses a certain degree of fuzziness. Consider any fuzzy attribute b_i of fuzzy set $l_j$in record r_k; the degree of membership will be denoted as $r_k[\mu(I_i, l_j)]$ of dataset D1. Here, FL is employed to divide the quantitative attributes into fuzzy intervals and to design a relevant set of linguistic labels, which are represented as fuzzy sets and then used as a new domain [28]. The Pima Indian Diabetic dataset is reprocessed using the approach, and each attribute is replaced with its corresponding fuzzy set. For example, the attribute *"Glucose"* is replaced with its fuzzy set *(low, medium, high, very high), and a trapezoidal membership function is used to compute* the membership value for each value. In this new domain, the weighted concept is also incorporated, as ignorance of attribute weights is always denounced because it will yield a dull rule with high frequency and low importance. To calculate the weights, a statistical approach based on Maximum Likelihood Estimation (MLE) is proposed [29]. The MLE is a statistical method in which parameter estimation is based on a probability distribution over the observed data [30]. After fuzzifying the attributes, the next step is to assign automated weights to each

fuzzified value. These weights are determined using the Maximum Likelihood Estimation (MLE) method [33]. MLE is a statistical technique that estimates parameters for estimating the probability distribution of observed data. When applied to a dataset, it identifies parameter values that maximise the likelihood of observing the given data. The likelihood function is expressed in Eq. (1):

$$L(P|x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|P) \quad \textbf{(1)}$$

where:

- P is the initial probability of the occurrence of an event,
- L(P) represents the likelihood corresponding to probability PPP,
- $x_1, x_2, \ldots, x_n$ are the n observed instances of a sample.

The process begins by calculating the prior probability of a class label (e.g., "yes") from the training dataset. MLE is then evaluated across a range of probabilities around this prior probability. By slightly varying these values, the method identifies the probability that produces the maximum likelihood of the observed data. The probability value that maximises the possibility is finally assigned as the weight for the corresponding attribute.

Example: Consider an instance of 11 records with the occurrence of "Diabetes=YES" due to the diagnostic attribute "high glucose" in the *diabetes* database, with a value of 3.

1. Find the prior (initial) probability value of the attribute in the training dataset. The initial probability (p) is (1)=3/11=0.27
2. Calculate Likelihood value L(.) for p.
   $p = 0.2727 \approx 0.30$
   L(.)= 0.00158(maximum)
3. Calculate the Likelihood value for the nearby probability value.
   ie =0.4, 0.2 and 0.1.
   L(.4)= 0.001293
   L(.2)= 0.001207
   L(.1)= 0.00035
4. The probability value, for which the Likelihood Estimation is maximum, is assigned as the weight to that symptom.
5. Here it is observed that maximum{(L(.), L(.4), L(.2), L(.1))}=0.00158, which is for the probability value =0.27≈ 0.3

By assigning higher weights to statistically significant fuzzy partitions and lower weights to irrelevant or noisy partitions, FWBBN reduces the influence of rare or spurious fuzzy sets. This acts as a regularisation effect, preventing the model from relying too heavily on unreliable partitions and thereby avoiding overfitting.

3.2.Definition, Formulae, and Algorithm

attributes. Novel formulas are formulated to generate strict rules between attributes and among attributes. To show the concept. A small sample dataset, D, comprising a few patient records, is presented in Table 3.

Table 3 Sample Dataset D of the Pima Indian Dataset.

Glucose	Blood Pressure	Insulin	Body Mass Index	Class Label
10	8	7	7	Yes
9	8	7	6	Yes
9	7	4	4	Yes
1	3	3	3	No
6	8	7	4	Yes

Using Table 3 (Database D), a new Table D1 is generated using the trapezoidal fuzzy membership function to show the partial belongingness of the actual attribute to each of the new corresponding fuzzy sets. Here every single attribute is classified into four fuzzy sets using linguistic tags *[low (**L**), moderate (**M**), high (**H**), very high (**VH**)]* with discretization rules applied to the dataset as *{1-3} belongs to low,{3-5} belongs to moderate,{5-7} belongs to high and {7-10} belongs to very high fuzzy sets*. For example, consider the attribute of Table 4, which is transformed into four new fuzzy attributes.Fore.g:***Glucose,Low),(Glucose, moderate),(Glucose,high),(Glucose,ver yhigh).***

Table 4 illustrates an example of a new database, D1, with fuzzy values. Table 4 contains fuzzy values for each attribute. Here, the Fuzzy weighted concept in the dataset is introduced and incorporated in Sections 3.2.1 and 3.2.2.

3.2.1.Weight Fuzzy Attribute

Table 5 presents the fuzzy attributes of the diabetic data with random weights. This approach is used to give weight $W(I_i, l_j)$ to each fuzzy Item $I(I_i, l_j)$ where ($1 \le i \le n$), ($1 \le j \le L$), and ($0 \le w \le 1$).

Table 4 Sample Dataset D1 with Fuzzy Values.

Glucose				Blood_Pressure				Insulin				Body_Mass_Index				Class Label
L	M	H	VH	L	M	H	VH	L	M	H	VH	L	M	H	VH	
.1	.1	.2	.6	.0	.1	.2	.7	0	.1	.2	.7	.1	.1	0	.8	Yes
.1	.1	.6	.2	.1	.1	.2	.6	0	.1	.4	.5	0	.2	.1	.7	Yes
.1	.1	.2	.6	0	0	.2	.8	.6	.3	.1	0	.7	.3	0	0	Yes
.7	.2	.1	0	.4	.6	0	0	.2	.6	.2	0	0	.1	.7	.2	No
0	.4	.6	0	.1	.2	.5	.2	0	0	.8	.2	.7	.3	0	0	Yes

Table 5 Random Weights for the Pima Indian Diabetic Dataset.

Attribute Name	L	M	H	VH
Glucose	.1	.3	.7	.9
Blood_Pressure	.1	.4	.6	.9
Insulin	.1	.3	.7	1
Body_Mass_Index	.1	.5	.7	.9

3.2.2.Weight Fuzzy Attribute Set Record

r_k[FASRW(X)] is computed as the product of the membership degree of an attribute in each fuzzy set (Table 4) and the weight of the fuzzy attribute of the set (Table 5) as formulated using Eq. (2).

$$r_k\left[FASRW(X)\right]=\prod_{i=1}^{|X|}\left(\forall(I_i,l_j)\in X\right)\left[r_k\left[\mu(I_i,l_j)\times W(I_i,l_j)\right]\right] \quad \textbf{(2)}$$

Example: Considering the 2-attribute set (glucose very high), (insulin high) of the first record in database D1 (Table 4), the FASRW is calculated as:

FASRW (glucose high), (insulin high) = (.6*.9) (.2*.7) = .54 *.14 = .0756

Here, 0.9 and 0.7 are weights, and 0.6 and 0.2 are the membership values for (glucose very high) and (insulin high), respectively.

3.2.3.Weight of Fuzzy Attribute Set

FA SW (X)**:** The sum of FASRW of all clinical records is computed as FA_SW(X), and the formula is as follows: Eq. (3) and Eq. (4).

$$FA\_SW(X)=\sum_{k=1}^{D_1}r_k\left[FASRW(X)\right] \quad \textbf{(3)}$$

$$FA\_SW(X)=\prod_{i=1}^{|X|}\left(\forall(I_i,l_j)\in X\right)\left[r_k\left[\mu(I_i,l_j)\times W(I_i,l_j)\right]\right] \quad \textbf{(4)}$$

Example: Consider the two attribute sets (glucose very high) and (insulin high).

FA SW (glucose very high), (insulin high)
=[(.6*.9)(.2*1)+(.2*.9)(.4*.1)+(.6*.9)(.1*.1)+(0*.9)(.2*.1)+(0*.9)(.8*1)]
=(.54*.2)+(.18*.04)+(.54*.01)+0+0
=.108+.0072+.0054
=.1206

The following definitions and formulas are proposed to calculate the FWS and FWC for two attributes, multiple attributes, and with class labels, to build Fuzzy Weighted class association rules. Further, similar calculations are performed.

3.2.4.Support Fuzzy Weighted Concept

Support Fuzzy Weight of any rule X→Y is calculated as the sum of weights of all records in which the given Y is true, divided by the total number of records where X and Y are sets of non-empty subsets of fuzzy weighted attributes, denoted by Support of Fuzzy Weight (X→Y) provided by Eq. (5).

$$=\frac{\sum\forall r_k\,having\ \prod_{i=1}^{|X|}\left(\forall(I_i,l_j)\in X\right)\left[r_k\left[\mu(I_i,l_j)\times W(I_i,l_j)\right]\right]given\,Y}{n} \quad \textbf{(5)}$$

where r_k is all records for which the given class label/descendant attribute is true.

3.2.5.Confidence Fuzzy Weight Concept

A generalised formula is created for Fuzzy weighted Confidence of 2 attributes, Multi attributes and with the given class label.

Confidence Fuzzy_ Weight of a rule X→Y, where X is a non-empty set of attributes, and Y is considered as an attribute or a class label. Eq. (6) and Eq. (7) are used to calculate the confidence values.

$$\text{Confidenceoffuzzy\_weight}=\frac{\text{supportoffuzzy\_weight}(X\cup Y)}{\text{supportoffuzzy\_weight}(X)} \quad \textbf{(6)}$$

$$=\frac{\sum\forall r_k\,having\ \prod_{i=1}^{|X|}\left(\forall(I_i,l_j)\in X\right)\left[r_k\left[\mu(I_i,l_j)\times W(I_i,l_j)\right]\right]given\,Y}{\sum_{k=1}^{|D_1|}\prod_{i=1}^{|X|}\left(\forall(I_i,l_j)\in X\right)\left[r_k\left[\mu(I_i,l_j)\times W(I_i,l_j)\right]\right]} \quad \textbf{(7)}$$

Using the above formulas, FWAR rules are generated. Next, the following formulas are designed to create strict rules for building the predictive model based on the Fuzzy Weighted-class Bayesian Theory. Lastly, a new concept, fuzzy_weighted_bayesian_confidence, is proposed to generate Fuzzy Weighted-class Bayesian Rules and construct a FWBBN model. For every Fuzzy weighted class rule, the joint probability distribution is calculated using Eq. (8).

$$P(X_1,X_2,\ldots,X_N)=\prod_{i=1}^{N}P(X_i\,|\,Parents(X_i)) \quad \textbf{(8)}$$

3.2.6.Fuzzy Weighted Bayesian Confidence (FWBC)

Consider a rule X→Y, where X is a set of predictors, Y is the class label, and the value of FWBC is calculated using Eq. 9.

$$FW\_BC(X \rightarrow Y) = \frac{supportoffuzzy\_weight(X,Y)}{supportoffuzzy\_weight(X)} \quad \textbf{(9)}$$

Here, the support of fuzzy weight (X, Y) is the value of the joint probability distribution calculated from Eq. 8. The Algorithm is shown in Fig. 3. It illustrates the steps to generate FWBAR to build the proposed model based on the Apriori Algorithm. The following algorithm is proposed, based on the above formulas and concepts, to generate a strong (FWBAR).

Algorithm

Our proposed algorithm is designed using the concepts and formulas for incorporating the fuzzy-weighted concept into the dataset shown in Fig. 3 below. Here, *the algorithm FWBAR generates partially fuzzy weighted rules to discover fuzzy weighted associations among* two attributes, multiple attributes, and Class labels, yielding FWAR. Calculated. Thereafter, the joint probability distributions for all FWARs are computed. The value of each rule is used to calculate FWBC.FWBC depicts the reliability or strength of FWBAR rules used to build the predictive model. Ultimately, the model is mature using FWBARs with the highest FWBC.

Algorithm: FWBAR

Function:GenerationOf_Partially_FuzzyWeightedRule (m,DB,B[Optional])
[Over a given dataset D
B this function to mine m attribute partial rules with high FWCand where B is the highly associated attribute sets of cardinalities m-1]

1. Highly frequent m-attribute sets are mined named as FREQ_ITEMS with provided Min_FWS Threshold value.
2.Given each member M∈ fFREQ_ITEM Iterate 2.1 & 2.2
2.1 Generation of non-empty subsets of M termed as S1
2.2 For each X∈ S1 Produce the FWAR X→M-X and append it to rule_SET1.

3 .For each rule of rule_SET1, Compute FWC.
4. Using the given Min_FWC_Threshold, Partial rules are mined and appended it to rule_SET1_ PARTIAL.

5.For each rule of rule_SET1_ PARTIAL like E→F , add the R.H.S attribute F to L.H.S attribute set E to form set EF and append it to the set n- HAFA.

6. Return n-HAFA.

Algorithm: FWBAR
[This algorithm mines strong FWBARs on the medical data DB with m attributes]
Data provided as Input: DB , database with m attributes with ClassLabel .
Result:FWBARules.

1. Discretization process is applied on the datanase DB.
2.Transform the DBas D1as Fuzzy values using Trapezodial membership function.
3. Apply MLEto compute weights of fuzzy attributes of D1.
4.Generation of weighted 2- Highly_Associated_Fuzzy AttributeSet termed as HAFA.
X[2]=Partially_FuzzyWeightedrule_generator (2,D1)
5. Repeat step 5.1 For l=3,4,....., m
5.1. X[l]=Generator Of_Partially _FuzzyWeightedRule_ (l,D1,X[l-1]) to extract Fuzzyweighted l – Highly_Associated_AttributeSet .
6. Compute corelation of m-Highly_Associated_AttributesSet with ClassLabel
FWR= Partially _FuzzyWeightedRule_generator (m+1, D1,X[m])
7. Each v∈ FWR Iterate steps 7.1 and 7.2
7.1. Compute the joint probability distribution of v.
7.2 Compute the Fuzzy Weighted Bayesian confidence (FWBC) for v.

8. Strong FWBAR rules with highest FWBC are generated to build the model.

Fig. 3 Algorithm for the Generation of FWBAR.

4.EXPERIMENTAL RESULTS

When different threshold values are used, the FWBBN model is empirically tested to achieve the highest accuracy, as shown in Table 3. First, FWARs are generated using formulas and algorithms that operate on attributes, multiple attributes, and class labels, by setting minimum threshold values for FWS and FWC. Then, based on Fuzzy Weighted Bayesian confidence (FWBC), strict rules FWBAR are extracted to form the FWBBN. Finally, the FWBBN model is trained using these strict rules. To assess the accuracy of the FWBBN model, the model is trained on test data, and the resulting accuracy is reported in Table 6. To build the model, the front end is implemented in Java 1.8, and the back end uses MySQL 8. Here, working with different minimum threshold values is significant: lowering the minimum threshold adds more rules to the rule base, whereas raising it may exclude some relevant rules. The experiment demonstrates that when the model is trained with 80% of the data, seven FWBAR strict rules are generated on setting a minimum threshold value of 40% for FWS and 80% for FWC. To verify the model's accuracy, it is tested with the remaining 20% of the data, achieving a highest accuracy of 96.8% for the Pima Indian dataset. The graphical representation of these results is projected in Fig. 4. The FWBBN classifier is now ready for predictive analysis. To the model, when new patient data is fed, it analyses against the strict rules and predicts the appropriate outcomes. The proposed model is further applied to additional clinical datasets, specifically the heart disease and Breast Cancer datasets from the UCI repository.

Table 6 Experimental Results of the Model FWBBN with Achieved Accuracies.

Setting Min_Thres fuzzyweightedValue	Train_ data	Test_ data	FWAR rules based on FWS and FWC	FWBAR strongrules basedon FWBC	Accuracy
Support=36% Confidence=70%	80%	20%	12	8	94%
	70%	30%	10	5	95%
	60%	40%	11	8	93.5%
Support=40% Confidence=80%	80%	20%	13	7	96.8%
	70%	30%	11	5	88%
	60%	40%	28	10	94%
Support=26% Confidence=60%	80%	20%	22	12	95%
	70%	30%	23	12	95.7%
	60%	40%	11	9	93%
Support=10% Confidence=50%	80%	20%	20	10	93.7%
	70%	30%	19	10	94%
	60%	40%	18	14	92%

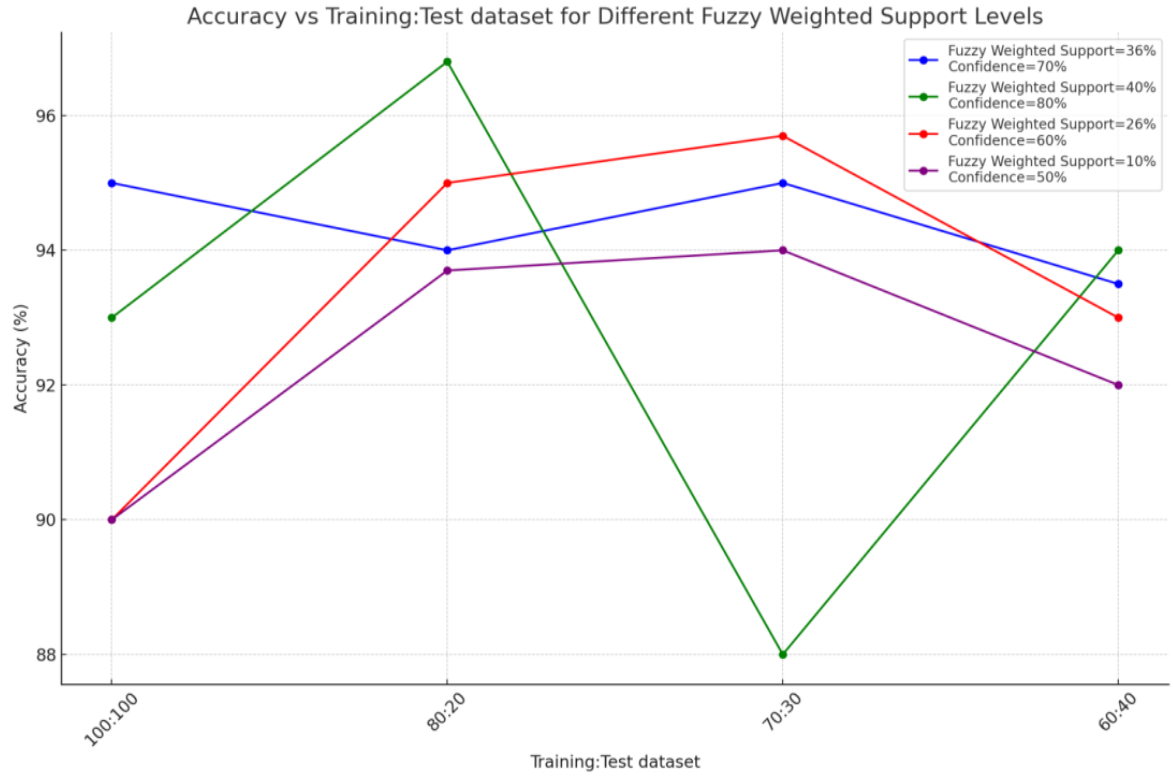

Fig. 4 Empirical Analysis of FWBBN Using the Pima Indian Diabetic Dataset on different Parameters.

Here, the FWBBN model is built using strict rules which are self-interpretable and self-explainable. These rules can be helpful to Clinicians for diagnosis, prognosis, and treatment plans for specific patients. Furthermore, because FWBBN is a probabilistic model, it is straightforward to quantify uncertainty and perform well with limited clinical data.

5.COMPARATIVE STUDIES

Next, the results of the FWBBN model are further evaluated based on the number of strict rules and their accuracy. Using three clinical datasets, Table 7 depicts the highest accuracy attained by the FWBBN. Firstly, with the provided minimum threshold, the proposed model FWBBNis built, when trained on 70% of the dataset, which yields five strict rules of the breast cancer dataset to construct the model and achieved the highest accuracy of 99% when tested on 30% of the dataset. Similarly, with the provided minimum threshold, the FWBBN model is trained on 70% of the data using seven strict rules on the heart disease dataset, achieving an accuracy of 93.7% on the remaining 30%. Again, with the provided minimum threshold, the FWBBN model, when trained on 80% of the data using seven strict rules of the PIDD dataset, achieves an accuracy of 96.8% when tested on the remaining 20% data. The graphical representation of these results is shown in Fig. 5.

Table 7 Accuracy of FWBBN on different Clinical Datasets.

Datasets	Min_Thres	Train: Test Split	Strong rules FWBAR	Accuracy (%)
Breast Cancer	FWS=36% FWC=70%	70%:30%	5	99%
Heart Disease	FWS=36% FWC=70%	70%:30%	7	93.7%
PimaIndian Diabetic Dataset	FWS=40% FWC=80%	80%:20%	7	96.8%

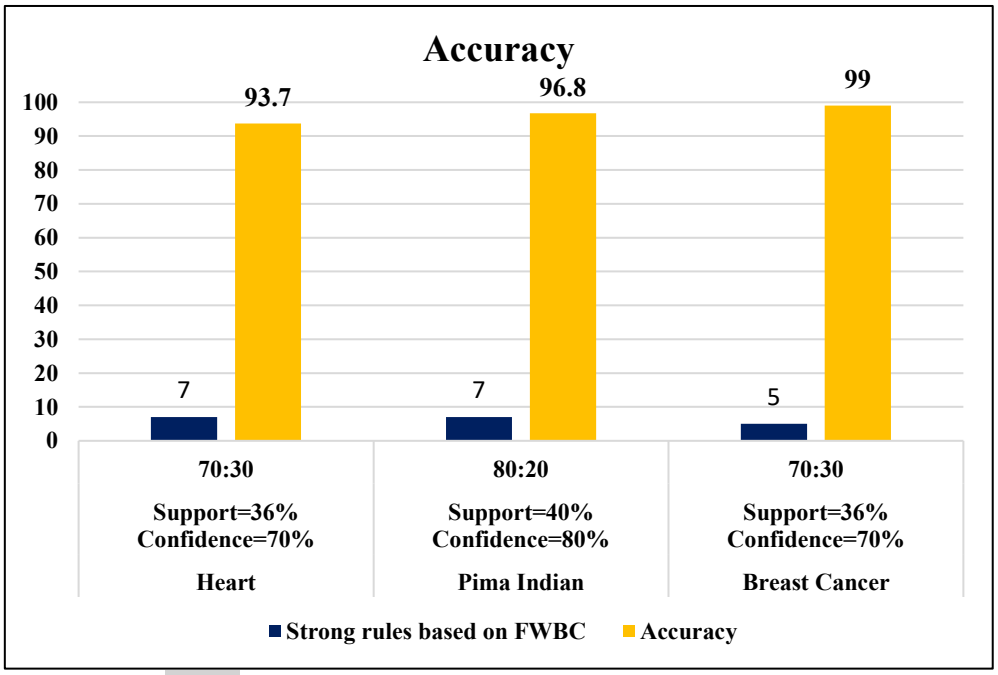

Fig. 5 Accuracy of FWBBN on Various Clinical Datasets.

As the prediction is done on the clinical domain, only the performance metric 'accuracy is insufficient; two more metrics, known as precision and recall, are also required to analyse the performance of the proposed model.

Table 8 Performance Metrics of FWBBN on Different Clinical Datasets.

Datasets	Accuracy	Precision	Recall
Breast Cancer	99%	99.36%	97%
Heart Disease	93.7%	95%	94%
Pima Indian Diabetes Dataset	96.8%	98.6%	97.5%

Table 8 presents the overall performance metrics of FWBBN across three datasets. For Diabetes, the achieved accuracy is 96.8%, Precision is 98.6%, and recall is 97.5%. For Breast Cancer, accuracy is 99%, precision is 99.3%, and recall is 97%. For Heart disease, accuracy is 93.7%, precision is 95%, and recall is 94%. The results show that FWBBN, when used in the clinical domain, yields auspicious results. In a subsequent evaluation, the proposed FWBBN model is compared with traditional Bayesian models. Table 9 shows a rigorous comparison in which the proposed FWBBN model is compared based on accuracy, and the final results show that FWBBN outperforms and gives very promising and better results in the clinical domain.

Table 9 Comparative Analysis of FWBBN with Existing Traditional Bayesian Networks.

Models	Breast Cancer Dataset	Heart Disease Dataset	Pima Indian diabetic Dataset
Existing Bayesian Networks	97.13%[23] 96.49%[32] 87%[25] 97.1%[26] 96.31%[24] 97.18% [8] 98% ([36] 96%([37] 96.66%([38]	84%[31] 85%[33] 92.7%[9] 83%([34]	82.48%[21] 92.2%[22] 95.8%[9] 90%([35]
Proposed Model-FWBBN	99%	93.7%	96.8%

The performance evaluation of the FWBBN model was conducted utilising the validation performance on the three diseases and comparisons with existing Bayesian Models. The results show that precision and recall are high. The substantialpredictive performance is good, indicating that the FWBBN model has substantial practical value and is reliable in the field. Predictive performance is strong, indicating that the FWBBN model has significant practical value and is reliable in the context of chronic diseases such as Diabetes, Heart Disease, and Breast cancer.

6.CONCLUSIONSAND FUTURE SCOPE

- The developed Reliable Diabetes Prediction model, FWBBN, integrates the characteristics of the clinical data, particularly the principles like "all symptoms are not equally important for prediction" and “fuzziness in clinical data”.
- The study and experiment of the novel proposed modelin the clinical domain as CDSS has shown promising results with three different chronic disease datasets. The combination of fuzzy logic and weighted concepts, using Bayesian networks, provides a robust framework for handling uncertainty and imprecision in weighted clinical data.
- The FWBBN model faces significant scalability concerns with large datasets, such as high fuzzification cost, Exponential growth in fuzzy partitions, and Expensive MLE-based weight computation. However, these issues can be mitigated using Dimensionality reduction and adaptive fuzzy partitioning.
- The current model provides enhanced decision-making, increased accuracy, and more transparency and interpretability as it is a rule-based model, which can surely help in real-time monitoring of patients.
- In the future, FWBBN may be utilised to create personalised treatment plans. This can lead to more effective and tailored healthcare solutions. Continued research and development in this domain will further unlock its capabilities and applications, paving the way for a more intelligent and efficient healthcare system.

Code Availability Statement:

The code used to analyse the data is not publicly available due to confidentiality agreements. However, the dataset used in this study is publicly available at [https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes].

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Shweta Kharya: Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation; Data curation; Writing – original draft; Visualization
Sunita Soni: Methodology; Software; Investigation; Writing – review & editing
Giridhar Urkude: Methodology; Investigation; Data curation
Ananya Smruti Snigdha Ojha: Investigation; Data curation
Debasish Swapnesh Kumar Nayak: Resources; Supervision; Writing – review & editing
Suprava Ranjan Laha: Project administration; Review & editing
Tripti Swarnkar: Writing – review & editing; Supervision

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

REFERENCES

[1] Chang V, Bailey J, Xu QA, Sun Z. **Pima Indians Diabetes Mellitus Classification Based on Machine Learning (ML) Algorithms.** *Neural Computing and Applications* 2022; **35**: 1-17.

[2] Kazi M, Khalid M, Akhter A, Moni MA. **Towards Reliable Diabetes Prediction: Innovations in Data Engineering and Machine Learning Applications.** *Journal of Data Science and Engineering* 2024; **9**: 144-158.

[3] Balma M, Sardanelli F, Trimboli P, et al. **The Role of Theragnostics in Breast Cancer: A Systematic Review of the Last 12 Years.** *Current Medical Imaging*

Formerly Current Medical Imaging Reviews 2023; **19**(8): 824-844.

[4] Sturgeon LP, Bragg-Underwood D, Tonya M, Blankenship. **Practice Matters: Prevention and Care of Individuals with Type 2 Diabetes.** *International Journal of Faith Community Nursing* 2016; **2**(2): 32–40.

[5] Dash S, Shakyawar SK, Sharma M, Kaushik S. **Big Data in Healthcare: Management, Analysis and Future Prospects.** *Journal of Big Data* 2019; **6**(1): 1-25.

[6] Kalantari A, Kamsin A, Shamshirband S, Gani A, Alinejad-Rokny H, Chronopoulos AT. **Computational Intelligence Approaches for Classification of Medical Data: State-of-the-Art, Future Challenges and Research Directions.** *Neurocomputing* 2018; **276**: 2–22.

[7] Kaur D, Sanyal S, Ray D, Ghosh R, Ray S. **Application of Bayesian Networks to Generate Synthetic Health Data.** *Journal of the American Medical Informatics Association* 2021; **28**(4): 801-811.

[8] Kharya S, Soni S, Swarnkar T. **Weighted Bayesian Belief Network: A Computational Intelligence Approach for Predictive Modeling in Clinical Datasets.** *Computational Intelligence and Neuroscience* 2022; **2022**: 1–8.

[9] Kharya S, Soni S, Swarnkar T. **Weighted Bayesian Belief Network for Diabetics: A Predictive Model.** *Frontiers in Artificial Intelligence* 2024; **7**: 1–11.

[10] Soni S, Vyas OP. **Fuzzy Weighted Associative Classifier: A Predictive Technique for Health Care Data.** *International Journal of Computer Science Engineering and Information Technology* 2012; **2**(1): 11–22.

[11] Nithya NS, Duraiswamy K. **Gain Ratio Based Fuzzy Weighted Association Rule Mining Classifier for Medical Diagnostic Interface.** *Sadhana - Academy Proceedings in Engineering Sciences* 2014; **39**(1): 39–52.

[12] Kang X, Porter CS, Bohemia E. **Using the Fuzzy Weighted Association Rule Mining Approach to Develop a Customer Satisfaction Product Form.** *Journal of Intelligent and Fuzzy Systems* 2020; **38**(4): 4343–4357.

[13] Alwidian J, Hammo BH, Obeid N. **Weighted Classification Based on Association Rules Algorithm for Breast Cancer Disease.** *Applied Soft Computing Journal* 2018; **62**: 536–549.

[14] Mishra SP. **Study of Fuzzy Logic in Medical Data Analytics.** *International Journal of Pure and Applied Mathematics* 2018; **119**(12): 16321–16342.

[15] Paul AK, Shill PC, Rabin MRI, Murase K. **Adaptive Weighted Fuzzy Rule-Based System for the Risk Level Assessment of Heart Disease.** *Applied Intelligence* 2018; **48**(7): 1739–1756.

[16] Kingsley C. **Adaptive Neuro Fuzzy Inference System for Diagnosing Coronavirus Disease 2019 (COVID-19).** *International Journal of Intelligent Computing and Information Sciences* 2020; **20**(2): 1–31.

[17] Hao J, Luo S, Pan L. **Rule Extraction from Biased Random Forest and Fuzzy Support Vector Machine for Early Diagnosis of Diabetes.** *Scientific Reports* 2022; **12**(1): 1–12.

[18] Zeinulla E, Bekbayeva K, Yazici A. **Effective Diagnosis of Heart Disease Imposed by Incomplete Data Based on Fuzzy Random Forest.** *IEEE International Conference on Fuzzy Systems* 2020; **2020**: 1-7.

[19] Tarle B, Akkalaksmi M. **Improving Classification Performance of Neuro-Fuzzy Classifier by Imputing Missing Data.** *International Journal of Computing* 2019; **18**(4): 495–501.

[20] Kanimozhi U, Ganapathy S, Manjula D, Kannan A. **An Intelligent Risk Prediction System for Breast Cancer Using Fuzzy Temporal Rules.** *National Academy Science Letters* 2019; **42**(3): 227–232.

[21] Xie J, Liu Y, Zeng X, Zhang W, Mei Z. **A Bayesian Network Model for Predicting Type 2 Diabetes Risk Based on Electronic Health Records.** *Modern Physics Letters B* 2017; **31**(19–21): 1–6.

[22] Joseph LP, Joseph EA, Prasad R. **Explainable Diabetes Classification Using Hybrid Bayesian-Optimized TabNet Architecture.** *Computers in Biology and Medicine* 2022; **151**: 106178.

[23] Kumar A, Sushil R, Tiwari AK. **Comparative Study of Classification Techniques for Breast Cancer Diagnosis.** *International Journal of Computer Science and Engineering* 2019; **7**(1): 234–240.

[24] Krishnakumar N, Abdou T. **Detection and Diagnosis of Breast Cancer Using a Bayesian Approach.** *Lecture Notes in Computer Science* 2020; **12047**: 335–341.

[25] Ershadi MM, Seifi A. **An Efficient Bayesian Network for Differential Diagnosis Using Experts'**

Knowledge. *International Journal of Intelligent Computing and Cybernetics* 2020; **13**(1): 103–126.

[26] Sengar PP, Gaikwad MJ, Nagdive AS. **Comparative Study of Machine Learning Algorithms for Breast Cancer Prediction.** *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology* 2020; **2020**: 796–801.

[27] **UCI Machine Learning Breast Cancer Dataset.** *University of California Irvine* 2022.

[28] Kharya S, Soni S, Swarnkar T. **Fuzzy Weighted Bayesian Belief Network: A Medical Knowledge-Driven Bayesian Model Using Fuzzy Weighted Rules.** *International Journal of Information Technology* 2023; **15**(2): 1117–1125.

[29] Kaur I, Kumar V, Kavitha NT, Mohan P. **Maximum Likelihood Based Estimation with Quasi Oppositional Chemical Reaction Optimization Algorithm for Speech Signal Enhancement.** *International Journal of Information Technology* 2022; **14**: 1-12.

[30] Soni S, Vyas OP. **Building Weighted Associative Classifiers Using Maximum Likelihood Estimation to Improve Prediction Accuracy in Health Care Data Mining.** *Journal of Information and Knowledge Management* 2013; **12**(1): 1-15.

[31] Fazel Zarandi MH, Seifi A, Ershadi MM, Esmaeeli H. **An Expert System Based on Fuzzy Bayesian Network for Heart Disease Diagnosis.** *Advances in Intelligent Systems and Computing* 2018; **648**: 191–201.

[32] Liu S, et al. **Quantitative Analysis of Breast Cancer Diagnosis Using a Probabilistic Modelling Approach.** *Computers in Biology and Medicine* 2018; **92**: 168–175.

[33] Muibideen M, Prasad R. **A Fast Algorithm for Heart Disease Prediction Using Bayesian Network Model.** *Journal of Electrical Engineering* 2020; 1–11.

[34] Suo X, Huang X, Zhong L, Luo Q, Ding L, Xue F. **Development and Validation of a Bayesian Network-Based Model for Predicting Coronary Heart Disease Risk from Electronic Health Records.** *Journal of the American Heart Association* 2024; **13**(1): 1–12.

[35] Liang X, Song W, Wy B, Yue Z. **Endocrine and Metabolic Science.** *Endocrine and Metabolic Science* 2025; **10**: 100155.

[36] Al Tawil A, Almazaydeh L, Alqudah B, Abualkishik AZ, Alwan AA. **Predictive Modeling for Breast Cancer Based on Machine Learning Algorithms and Features Selection Methods.** *International Journal of Electrical and Computer Engineering* 2024; **14**(2): 1937–1947.

[37] Karthiga R, Narasimhan K, Raju N, Amirtharajan R. **Automatic Approach for Breast Cancer Detection Based on Deep Belief Network Using Histopathology Images.** *Multimedia Tools and Applications* 2025; **84**(8): 4733–4750.

[38] Mohammad IB, Ahmad MM. **Bayesian Model Prediction for Breast Cancer Survival: A Retrospective Analysis.** *European Journal of Breast Health* 2025; **21**(3): 255–264.

[39] Dash L, Pattanayak BK, Laha SR, Pattnaik S, Mohanty B, Habboush AK, Al Smadi T. **Energy Efficient Localization Technique Using Multilateration for Reduction of Spatially and Temporally Correlated Data in RFID System.** *Tikrit Journal of Engineering Sciences* 2024; **31**(1): 101–112.

[40] Al-Sharo YM, Al Smadi K, Al Smadi T. **Optimization of Stable Energy PV Systems Using the Internet of Things (IoT).** *Tikrit Journal of Engineering Sciences* 2024; **31**(1): 127–137.

[41] Mahmood AT, Gaeid KS, Al Smadi TA. **Direct Torque Control Space Vector Modulation for Induction Motor Driven by Matrix Converter.** *Tikrit Journal of Engineering Sciences* 2024; **31**(4): 58–69.