Int. j. inf. tecnol. (February 2023) 15(2):1117–1125
https://doi.org/10.1007/s41870-022-01153-y

ORIGINAL ARTICLE

Fuzzy weighted Bayesian belief network: a medical knowledge-driven Bayesian model using fuzzy weighted rules

Shweta Kharya[1] · Sunita Soni[1] · Tripti Swarnkar[2]

Received: 10 August 2022 / Accepted: 27 December 2022 / Published online: 12 January 2023

Abstract In this current work, Weighted Bayesian Association rules using the Fuzzy set theory are proposed with the new concept of Fuzzy Weighted Bayesian Association Rules to design and develop a Clinical Decision Support System on the Bayesian Belief Network, which is an appropriate area to work in Clinical Domain as it has a higher degree of unpredictability and causality. Weighted Bayesian Association rules to construct a Bayesian network are already proposed. A "Sharp boundary" issue related to quantitative attribute domains may cause erroneous predictions in medicine and treatment in the medical environment. So to eradicate sharp boundary problems in the medical field, the fuzzy theory is applied in attributes to deal with real-life situations. A new algorithm is designed and implemented in this paper to set up a new Bayesian belief network using the concept of Fuzzy Weighted Association rule mining under the Predictive Modeling paradigm named Fuzzy weighted Bayesian belief network using numerous clinical datasets with outshone results.

Keywords Fuzzy weighted two attributes · Multi attributes association rule · Bayesian network · Weighted Bayesian association rule · Fuzzy theory · Weighted concept

✉ Shweta Kharya
shweta.bitdurg@gmail.com

Sunita Soni
sunitasoni74@gmail.com

Tripti Swarnkar
triptiswarnakar@soa.ac.in

[1] Department of CSE, Bhilai Institute of Technology, Durg 491001, India

[2] Department of Computer Applications, S'O'A Deemed to Be University, Bhubaneshwar 751001, India

1 Introduction

The Strong Bayesian association rules are extracted using the Weighted Bayesian Association rule Mining Algorithm (WBAR) was already designed and implemented with outperforming results [1, 2]. In this paper, Predictive Modeling concepts play a crucial role in developing a new algorithm for the medical support system with enormous medical records [3]. Unfortunately, patients' records are not thoroughly mined for effective decision-making to discover hidden patterns [4]. So to analyze medical records, advanced data mining approaches show significant results in the research field, finally contributing to a more accurate and high-performance medical decision support system. Sometimes clinical and treatment decisions are taken on the ground of a doctor's experience and knowledge, despite the inside, which can be extracted from a rich, substantial medical database [5]. And also, due to redundant and interrelated symptoms in medical diagnosis, physicians may fail to diagnose it accurately. Unfortunately, at the early stage, accurate diagnosis of the disease is quite challenging due to interdependence on various features [6].

A Fuzzy clinical decision support system (CDSS) based on a Bayesian belief network (BBN) is proposed, which can support medical staff or any experts with knowledge of patient-specific information to excavate and represent the hidden information when required intelligently [7]. But uncertainty always occurs in every building phase of the decision support process. Uncertain sources are like patients lacking in describing their sufferings accurately, degree of errors in laboratory reports, doctors or nurses sometimes fail to examine precisely their detection results, and it becomes harder to determine one's prognosis. Therefore with machine learning techniques, more advanced and accurate decision support systems should be implemented to adapt to a new environment and implicitly learn from instances. So to build CDSS, various methodologies can be incorporated to predict, assess, and extract information like statistical methods, data mining techniques, Soft computing techniques, and many more can be included, and significant research should be done in academic and practical areas. But several misconceptions arise to tamper with the accuracy of CDSS in the medical field, like representation and interpretation of clinical attributes under uncertainty which need a lot of refined methodology and techniques. So to handle this uncertainty, the current work proposes a new model known as Fuzzy Weighted Bayesian Belief Network (FWBBN) CDSS with new formulas and algorithms. The main contribution of the proposed framework are as follows:

- Usage of Fuzzy Logic to deal with sharp boundaries, vagueness, and imprecision in medical attributes [8].
- Weight assignment method on medical dataset attributes [2].
- And to find the interdependence among attributes and to generate well-built rules, association rule mining is applied.
- A hybrid novel approach is anticipated, incorporating fuzzy weighted association rule mining rules to build a Bayesian belief network.

The following is the workflow of the research proposal; Sect. 2 briefly points to the related work in tabulated form. Section 3 focuses on research methodology with new formulas and the Fuzzy Weighted Bayesian Association Rule(FWBAR) algorithm; Sect. 4 covers results and discussion; Sect. 5 shows the comparative study; Sect. 6 concludes the work with future scope.

2 Background work

Various soft-computing techniques, including data mining techniques, are surveyed, especially fuzzy logic, weight assignment methods, Association rule mining, and Bayesian belief network. Here Table 1 demonstrates relevant review findings of these techniques used in the clinical domain for building a predictive model are reviewed in the literature.

From the exhaustive literature survey and its relevant finding, the gap is identified to work on the dataset's attributes as attributes have extraordinary importance with sharp boundary problems and are interdependent with some association levels. So to find out the impact of attributes and their interdependencies, a novel idea is proposed in the following section.

3 Methodology

The method of the proposed research work is elaborated using the following proposed algorithm as framed in Fig. 1.

This approach incorporates fuzzy theory with the WBAR mining algorithm [1]. The previous paper discussed the basic concept of the Bayesian belief network, Association rule mining, and types of weight assignments [1]. In this paper fuzzy approach will be incorporated to enhance the accuracy. The fuzzy model is a valuable technique for discovering the presence of imprecision in data patterns and understanding data semantics [30]. The study and experiments are done using a breast cancer dataset and other clinical datasets extracted from the University of California Irvine(UCI) machine learning repository via LUCS-KDD DN software [2, 31].

3.1 Fuzzy property of quantitative attribute

Association Rule Mining (ARM) model plays a significant role in dealing with quantitative data in many applications like temperature, pressure, etc., which are very common [32]. Discretization is needed in an ARM to convert quantitative data into the nominal domain. Here to deal with this, the Apriori-type method is used. Thus, association rule P → Q gives a relationship between nominal values of data items. Consider an example like "(FamilyHistory, yes), (Obesity, severe) → (Diabetics, yes)" [9]. These mined results are affected by partitioned intervals called "Sharp Boundary", particularly for data values near interval boundaries. Numbers of quantitative parameters which suffers from sharp boundary problem are present in the medical field. Consider an attribute Smoking in a particular record of a patient where the Smoking frequency per day is 11 then according to following discretization rules, Smoking [1–3] → LungCancer = " Low", Smoking [2–5] → LungCancer = " Moderate", Smoking [4–10] → LungCancer = " High", Smoking [9-*] → LungCancer = " Severe". In this case, according to a sharp boundary, the patient falls in the severe cancerous zone, which will not give the correct result. Here comes the role of fuzzy logic, using which the patient will partially belong to the different fuzzy sets. Therefore the patient membership value to the fuzzy set should be for example (μ (LungCancer, "low") = 0.01, μ (LungCancer, "moderate") = 0.02, μ (LungCancer, "high") = 0.3) μ (LungCancer, "severe") = 0.67). Due to the impact of the sharp boundary problem on the quantitative attribute in the ARM model [4], a new idea is proposed known as the Fuzzy Weighted based ARM Algorithm. Then the redefined framework is proposed as Fuzzy Weighted Support (FWS) and Fuzzy Weighted Confidence to adapt to a Fuzzy environment. In this proposed paper fuzzy membership value of each fuzzy

Table 1 Significant review findings of various soft computing methods used in the Clinical domain predictive models

S.no	Author	Year	Techniques	Relevant review findings
1	Ibrahim, D	2016	Soft Computing	Deals with imprecision, partial truth, and uncertainty among data Paper is based on fuzzy logic, genetic algorithms, ANN, machine learning and expert system [8]
2	Gambhir, S et al	2016	CDSS based on the classification model, rule-based expert system, fuzzy system, case-based system	Data mining techniques can work with most required, challenging and chronic sub-areas of medical research with improved accuracy [9]
3	Prakash, M et al	2018	Fuzzy Logic	The emphasis on fuzzy logic in medical data like text, signal, and image Fuzzy logic plays a vital role in research and development, especially in prediction analysis, classification, pattern recognition, and feature extraction [10]
4	Mokeddem, S.A	2018	Fuzzy theory, Random Forest algorithm, C5.0 decision tree	Random forest is used for feature ranking, C5.0 technique is used for crisp rule generation A fuzzy inference system is built with an accuracy of 90.50% on UCI heart disease datasets [11]
5	Zarandi, M.F et al	2017	Bayesian belief network with fuzzy probability	K2 method is used to construct the Bayesian network, for heart disease detection using the UCI heart disease dataset. Proposed fuzzy-based BBN claims more accuracy than BBN, multi SVM, RBF, and KNN [12]
6	Fan,C.Y et al	2011	Hybrid model using case-based data clustering method and fuzzy decision tree for medical data classification	Tthe UCI datasets, such as the Liver disorder and breast cancer Wisconsin datasets, are used Case-based clustering method is used to attain homogeneity in a dataset of each cluster To construct a decision-making system by incorporating a Fuzzy decision tree and genetic algorithm are applied [13]
7	Paul, A.K et al	2018	Fuzzy logic, Genetic algorithm And modified dynamic multi-swarm particle swarm optimization	A collaborative work using fuzzy Logic, GA and modified DMSPSO yields a more efficient and adaptive heart disease prediction system [14]
8	Adeli, A	2010	Fuzzy model	Build a fuzzy-model-based diagnostic system for heart disease Mamdani inference method is used to devise a fuzzy expert system [15]
9	Soni, S et al	2013	Weighted Associative Classifier	Weights of each attribute of heart and breast UCI datasets are calculated using the Maximum Likelihood theory WAC outperforms all datasets compared with CBA, CMAR, and CPAR [16]
10	Alwidian, J et al	2018	Weighted classification on association rules	Domain expert knowledge assigns weights to attributes of the breast cancer UCI dataset WCBA outperforms all other classifiers like CBA, CMAR, MCAR, FACA, and ECBA [17]
11	Ramasamy, S et al	2017	Keyword-based clustering algorithm	Impactful feature extraction using a Keyword-based weighting scheme [18]
12	Horný, M	2014	Bayesian Network	Genie software is used to build a Bayesian network [19]
13	Xie, Y et al	2017	Bayesian Network	K2 algorithm and Bayes net toolbox, MATLAB [20]
14	Topuz, K et al	2018	Bayesian Belief Network	BBN can model the complex non-linear relationships among different variables and provide reasoning under uncertainty [21]
15	Agrahari, R et al	2018	Co-expression networks and Bayesian networks	Results outperform on diseases like acute myeloid leukaemia and myelodysplastic syndrome [22]

Table 1 (continued)

S.no	Author	Year	Techniques	Relevant review findings
16	Ershadi, M. M et al	2020	Bayesian Belief Network	Bayesian networks were constructed based on domain experts' knowledge to acquire accuracy of 87% using 10 sample datasets Matlab R2015a, 64-bit software, is a classification method based on the k-fold cross-validation technique Bayesian Network with experts' knowledge has an accuracy of 87% [23]
17	Setiawan, N. A et al	2020	Fuzzy decision support system	Rough Set theory (RST) discovers the inferences from the UCI heart disease data The proposed fuzzy weighing method is based on supporting selected RST rules applied to Neural networks to build a Fuzzy Decision Support System. And it outperforms compared with different classifiers and other datasets [24]
18	Azar, A et al	2019	Fuzzy cognitive map (FCM) and Bayesian belief network	A combination of FCM and BBN is built for modelling operational risk to resolve data scarcity In this proposal, FCM is used for the problem structuring method to increase the capability of BBN BBN approaches are applicable for solving complex problems containing insufficient data [25]
19	Ukaoha, K. C et al	2020	Adaptive Neuro-Fuzzy inference system	ANFIS is designed using the Gaussian membership function in MATLAB COVID-19,600 datasets taken from Kaggle open source dataset repository The model reported an accuracy of 96.6% [26]
20	Amadin, F. I et al	2019	Bayesian Belief Network	Bayes network was designed for predicting neonatal jaundices, especially kernicterus The dataset of 25 patient cases was used from the University of Benin Teaching Hospital The BBN model was implemented using Bayes Server 7.5 using expert knowledge The BBN classifier consists of 15 nodes with 97% and 94% accuracy in classifying neonatal jaundice and kernicterus, respectively [27]
21	Simsek, S et al	2021	Tree Augmented Naive Bayes (TAN Bayes) Model	A probabilistic data-driven methodology is developed using the TAN Bayes model to determine no-show patient categories Variable selection is made using Extreme Gradient Boosting, Particle Swarm Optimization and Genetic Algorithms Conditional interrelationships among variables are obtained using TAN, and Bayesian belief concepts in this ROC score of .828 were achieved [28]
22	Shweta Kharya et al	2022	Weighted Bayesian Belief Network (WBBN)	Strong rules are generated using weighted Bayesian confidence and weighted Bayesian lift to build WBBN [7]

set is calculated using the trapezoidal membership function as shown in Eq. 1.

$$\mathrm{F}(\mathrm{x} : \mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}) = \begin{cases} 0, x \le a \\ (x-a)/b-a, a \le x \le b \\ (d-x)/d-c, c \le x \le d \\ 0, x \ge d \end{cases}. \tag{1}$$

Table 2. shows the fuzzy values obtained for attributes using the trapezoidal membership function named D1. Here tabulation is done for a few attributes, and only five records are populated.

These tabulated fuzzy values of attributes remove the sharp boundary problems present in the medical world. They can further be used to assign different weights using the automatic weight assignment method.

3.2 Weight assignment using maximum likelihood estimation method

After the fuzzification of attributes, the next step is calculating automated weights for each fuzzified value. Here weights are computed using the Maximum Likelihood Estimation (MLE) method [33]. MLE is a statistical method in which parameter estimation is done using probability distribution on the observed data. When enforced with a data set, MLE estimates the model's parameters. This technique discovers the estimate of a parameter which maximizes the probability of a particular observed value for a given training data model. The likelihood function is defined as Eq. 2:

$$L(\mathrm{P}|x_1, x_2, \ldots ..x_n) = \prod_{i=1}^{n} f\left(\frac{xi}{P}\right), \tag{2}$$

where P is the initial probability of occurrence of a particular event.

L(P) is the likelihood value for probability value P.

$x_1,x_2,\ldots x_n$ is the n instance of a given sample.

Here the calculation starts by finding a prior probability of a class label "yes" value using the training data set. The MLE is measured upon divergent probability values in the neighbouring locality of this prior probability, varying in slight offset amounts to compute the likelihood of the observed data with the highest value, i.e. the probability value for which the Likelihood estimation is maximum is assigned as the weight to that particular attributes. All the weights are calculated using the MLE technique, as shown in Table 3.

In this proposal, novel modifications are done in the medical domain to construct BBN with improved prediction accuracy by fuzzyfing quantitative medical attributes and

Fig. 1 Fuzzy WBAR Algorithm

Algorithm: FWBAR
Input to System: The Database consists of records and attributes.
Output from System: FuzzyWeighted Bayesian Association Strong Rules.
1. Discretize 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 Itemset, Multi-Item set, and Class label.
6. Again, Calculate Fuzzy Weighted Confidence for Two Itemset, Multi-Item set and Class label.
7. Generation of strong 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.

Table 2 Fuzzy values using the trapezoidal membership functions

Clump thickness				Uniformity of CELLSIze				Uniformity of cellshape				Marginal adhesion			
Low	Medium	High	Very high	Low	Medium	High	Very high	Low	Medium	High	Very high	Low	Medium	High	Very high
0	0	0.75	0.25	0	0	1	0	0	0	0	1	0	0	0	1
0	0	0.75	0.25	0	0	1	0	0	0	1	0	0	0	0.75	0.25
0.9	0.1	0	0	0	0	0	1	0	0	1	0	0	0	1	0
0	0.35	0.5	0.15	0	0	0.6	0.4	0	0	0.6	0.4	0	0	0	1
0	0	1	0	0	0	1	0	0	0	0	1	0.9	0.1	0	0

Table 3 Computation of Weights using MLE

Attribute type	Low	Medium	High	Very high
Clump_Thickness	3.08E−139	1.27E−47	1.81E−182	2.04E−101
Uniformity_ofCellSize	1.34E−97	3.27E−32	9.63E−109	4.70E−61
Uniformity_ofCellShape	6.31E−107	1.36E−35	5.90E−119	6.80E−71
Marginal_Adhesion	7.80E−108	1.31E−07	1.16E−93	2.20E−50
SingleEpithelial_CellSize	3.69E−200	1.08E−44	2.38E−125	6.62E−56
Bare_Nuclei	4.10E−66	6.74E−21	2.77E−73	1.08E−44
Bland_Chromatin	4.10E−192	1.31E−08	1.18E−115	1.21E−96
Normal_Nucleoli	1.94E−85	7.84E−29	1.48E−75	1.88E−20
Mitoses	2.14E−62	9.35E−14	0.00071375	2.13E−15

then applying weights. Hence the core problem is to define the terms and new concepts to build Fuzzy Weighted BBN.

3.3 Fuzzy weighted approach

Consider a dataset comprised of fuzzy relational Database D = { t_1, t_2, t_3.... t_i...t_n} with a set of attributes A = (a_1, a_2, ...a_m}; each a_K is related with a linguistic labels set L = {l_1, l_2, ...l_L} for example L = {high, low, moderate}. Consider that each a_k is associated with fuzzy set F_k = {(a_k,l_1), (a_k,l_2), (a_k,l_3), ...(a_k,l_L)}. In the given record r_k, each attribute a_i is associated with some degree of fuzzy sets. A membership degree in the range [0.0.1] is produced by some degree of association. Consider any fuzzy attribute a_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 to generate association rules and strong rules between attributes following definitions and formulas are offered.

Definition 1 Weight of Fuzzy Attribute: Table 3 exhibits the automated weight computed for fuzzy attributes of the breast cancer dataset [14]. 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 ($0 \le w \le 1$).

Definition 2 Weight of Fuzzy Attribute Set Record: r_k[FASRW(X)] is calculated as the product of the weight of the fuzzy attribute of the set and membership degree of an attribute in a given fuzzy set in the transaction r_k as formulated below in Eq. 3.

$$\mathrm{rk[FASRW(X)]} = \prod(\forall(\mathrm{Ii, Ij})\epsilon \mathrm{X})[\mathrm{rk}\left[\mu(\mathrm{Ii, Ij}) * \mathrm{W(Ii, Ij)}\right]. \tag{3}$$

Definition 3 Weight of Fuzzy Attribute_Set: FA_SW(X) is calculated as the sum of FASRW of all clinical records, and the formula is framed as follows Eqs. 4 and 5.

$$\mathrm{FA}_{\mathrm{SW(X)}} = \sum_{k=1}^{D1} rk[FASRW(X)], \tag{4}$$

$$FA_S W(X) = \sum_{k=1}^{D1} \prod_{i=1}^{\mathrm{X}} (\forall(\mathrm{Ii, Ij})\epsilon \mathrm{X})[\mathrm{rk}[\mu(\mathrm{Ii, Ij}) * \mathrm{W(Ii, Ij)}]. \tag{5}$$

Definition 4 Support with Fuzzy_Weighted Concept: In this concept, a generalized formula is framed for Fuzzy weighted support of two attributes, Multi attributes and class label.

SupportOfFuzzy _Weight of rule X → Y, where X and Y are set of non-empty subsets of fuzzy weighted attributes is calculated as the sum of weights of all records in which the given Y is true, divided by the total number of records, denoted by SupportOfFuzzy_Weight (X → Y) provided by Eq. 6.

$$\text{Support of Fuzzy\_Weight}(\mathrm{X} \rightarrow \mathrm{Y}) = \frac{\sum \forall r_k \text{ having } r_k[\text{FASRW}(X)] \text{ given} Y}{\text{No. of records in } D1}, \tag{6}$$

where r_k is all transactions for which the given class_label is true.

Definition 5. Confidence with Fuzzy_Weight Concept: In this concept, a generalized formula is framed for Fuzzy weighted Confidence of two attributes, Fuzzy weighted Confidence of Multi attributes and Fuzzy weighted.

Confidence in the given class label. Confidence Fuzzy_ Weight of a rule X → Y where X is non-empty set of attribute and Y is also an attribute. And it is defined as the ratio of SupportOf Fuzzy_Weight of (X ∪ Y) and SupportOfFuzzy _Weight of (X) as mentioned in Eq. 7.

$$\text{Confidence Of Fuzzy}_{\text{Weight}} = \frac{\text{Support Of Fuzzy\_Weight}(\mathrm{X} \cup \mathrm{Y})}{\text{SupportOfFuzzy\_Weight}(\mathrm{X})}. \tag{7}$$

A new concept known as fuzzy_weighted_bayes_confidence is proposed to construct a fuzzy_weighted Bayesian belief network, i.e. FWBBN.

Definition 6. To define FuzzyWeighted _BayesianConfidence (FW_BC) consider a rule X → Y which is framed as P (Y|X) as in Eq. 6 and used to assess BN as given below in Eq. 8.

$$\mathrm{FW_{BC}(X \rightarrow Y) = P(Y|X)} = \frac{\text{Support Of Fuzzy}_{\text{Weight(X,Y)}}}{\text{Support Of Fuzzy}_{\text{Weight(X)}}}. \tag{8}$$

Applying the above algorithm and formulas to various clinical datasets to achieve desired and outshone results.

4 Result and discussion

The model is developed using the proposed methodology and designed formulas in which the dataset's attributes are manipulated using a fuzzy weighted approach related to the generation of strong rules to build the Bayesian networks for the medical domain, which will be an efficient model for higher accuracy. Table 4. reveals the experimental value setup, generation of rules, and extraction of solid rules based on Fuzzy Weighted Bayesian Confidence (FWBC)

Table 4 Strong rules based on FWBC and its accuracy

Setting Minimum Threshold fuzzyweightedValue	Training data (%)	Testing data (%)	Rules-based on fuzzy weighted support and confidence	Strong rules based on FWBC	Accuracy (%)
Support = 36% Confidence = 70%	100	100	22	10	97.08
	80	20	11	7	95.7
	70	30	11	5	**99**
	60	40	11	7	92.5
	100	100	11	7	89.53
Support = 40% Confidence = 80%	80	20	11	7	95.74
	70	30	11	7	86
	60	40	28	12	92.55
Support = 26% Confidence = 60%	100	100	23	11	89.53
	80	20	22	11	95.74
	70	30	23	12	97.18
	60	40	11	9	92.55
	100	100	23	12	89.53
Support = 10% Confidence = 50%	80	20	23	12	95.74
	70	30	23	12	97
	60	40	23	12	92.5

The FWAR mining algorithm is applied to generate strong rules based on FWBC to design a Bayesian model termed FWBBN with the highest accuracy achieved 99% when the model is built using 5 strong rules on a dataset with a ratio of 70% training dataset and 30 % testing dataset

using a minimum threshold value of fuzzy weighted support and fuzzy weighted confidence to eradicate overfitting and underfitting problem [34]. FWAR mining is applied to generate strong rules to design a Bayesian model termed FWBBN with an efficient and more accurate predictive model in the form of a clinical decision support system.

The experiment shows that the model developed using training data = 70% and test data = 30% with strong rules based on fuzzy weighted Bayes confidence gives the accuracy of 99% for the breast cancer dataset particularly.

5 Comparative analysis

This model is enforced to numerous clinical datasets from the UCI repository for rigorous comparative analysis. The LUCS KDD DATASETS in.num format are downloaded of Heart disease, Pima Indian diabetic, Hepatitis and liver disorder datasets [31]. The results are excellent as FWBBN perform with noteworthy accuracy, proving that the proposed model FWBBN executes efficiently with diverse clinical datasets, as shown in Table 5. This analysis reveals the highest accuracy by setting different minimum threshold values for fuzzy weighted support and fuzzy weighted confidence with varying training and testing datasets ratios. Thus, the

Table 5 FWBBN Results on other Clinical Datasets

Datasets	Class labels	Fuzzy weighted minimum threshold	Training data (%)	Testing data (%)	Strong rules based on FWBC	Accuracy (%)
Heart	5	Support = 36% Confidence = 70%	70	30	7	93.7
Pima Indian	2	Support = 40% Confidence = 80%	80	20	7	96.8
Hepatitis	2	Support = 36% Confidence = 70%	70	30	6	95
Liver Disorder	2	Support = 40% Confidence = 80%	70	30	5	95.3

Table 6 Comparison of FWBNN with existing fuzzy-based classification models

Datasets	Models	Accuracy (%)
Wisconsin Breast Cancer Dataset	FWBBN	99
Pima Indian Diabetic Dataset	Fine Tuning Fuzzy KNN	90.63
Cord-19 dataset	FDT	99
Cleveland dataset	FRF	93.4
UCI Breast Cancer Dataset	Neuro-Fuzzy Classifier	95.14
UCI Breast Cancer Dataset	Fuzzy Temporal rule-based classifier	99

proposed model outshone its performance in varieties of the clinical dataset, proving that Bayesian Networks is best suited to work in the clinical world.

The put forward model FWBBN is analyzed with existing fuzzy classification models using various medical datasets in the clinical world. Table 6. manifest the comparisons of the proposed model with other already available state-of-the-art systems like Fine Tuning Fuzzy KNN classifier [35], Spare Bayesian Randon Weight Fuzzy Neural Network (RWFNN) [36], Fuzzy Decision Tree (FDT) [37], Fuzzy Random Forest (FRF)Technique [38], Neuro-Fuzzy Classifier [39], Fuzzy Temporal rule-based classification model [40].

Through rigorous comparisons of the proposed model with existing fuzzy models, it seems FWBBN outperforms when compared with some models and is at par for some. And the experimental results confirmed that the FWBBN is more bonafide and justifiable than other existing models and can be used for various disease diagnoses and refinements.

6 Conclusions and future scope

A new methodology and algorithm for improving WBAR are proposed and termed FWBAR, an efficient algorithm for constructing CDSS using BBN as FWBBN. This proposed algorithm with new formulas and concepts is implemented using the UCI machine learning repository, especially with the breast cancer data, Heart disease data, and many more benchmark datasets to be worked with. The fuzzy approach is applied to reduce the sharp boundary problem in WBAR. Thus, stronger rules will be yielded to datasets using a weighted and fuzzy method. For prediction, FWBBN-CDSS can be utilized very effectively and accurately in terms of high performance, minor error, and low time complexity compared to the conventional Bayesian model. In future work, fuzzy weighted Bayesian rules can be used to generate synthetic datasets most demanded in the clinical world for research and deep analysis, which will be validated using the FWBBN model.

Data availability The datasets used in this proposal are extracted from the University of California Irvine machine learning repository. Like UCI machine learning breast cancer dataset is extracted from "https://csc.liv.ac.uk/~frans/KDD/software/LUCS-KDDDN/datasets/dataSet.html".

References

1. Kharya S, Soni S, Swarnkar T (2019) Weighted Bayesian association rule mining algorithm to construct Bayesian belief network. In: Proceedings - 2019 International Conference on Applied Machine Learning, ICAML 2019, pp 27–33. https://doi.org/10.1109/ICAML48257.2019.00013
2. Kharya S et al (2022) Weighted Bayesian belief network : a computational intelligence approach for predictive modeling in clinical datasets. Comput Intell Neurosci 2022:1–8. https://doi.org/10.1155/2022/3813705
3. Jameel R, Ashish MS, Mourya K (2022) Predictive modeling and cognition to cardio-vascular reactivity through machine learning in Indian adults with sedentary and physically active lifestyle. Int J Inf Technol 14(4):2129–2140. https://doi.org/10.1007/s41870-021-00721-y
4. Tech GSM (2011) Decision support in heart disease prediction system using Naive Bayes 2(2):170–176
5. Yadav DC, Pal S (2022) Thyroid prediction using ensemble data mining techniques. Int J Inf Technol 14(3):1273–1283. https://doi.org/10.1007/s41870-019-00395-7
6. Anooj PK (2012) Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules. J King Saud Univ Comput Inf Sci 24(1):27–40. https://doi.org/10.1016/j.jksuci.2011.09.002
7. Sharma A (2022) Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. Int J Inf Technol 14(4):1949–1960. https://doi.org/10.1007/s41870-021-00671-5
8. Dhyani M, Singh G (2022) A novel intuitionistic fuzzy inference system for sentiment analysis. Int J Inf Technol. https://doi.org/10.1007/s41870-022-01014-8
9. Ibrahim D (2016) An overview of soft computing. Procedia Comput Sci 102:34–38. https://doi.org/10.1016/j.procs.2016.09.366
10. Gambhir S, Malik SK, Kumar Y (2016) Role of soft computing approaches in healthcare domain: a mini review. J Med Syst. https://doi.org/10.1007/s10916-016-0651-x
11. Susmita Mishra MP (2018) Study of fuzzy logic in medical data analytics. Int J Pure Appl Math 119(12): 16321–16342. https://acadpubl.eu/hub/2018-119-12/articles/6/1515.pdf
12. Mokeddem SA (2018) A fuzzy classification model for myocardial infarction risk assessment. Appl Intell 48(5):1233–1250. https://doi.org/10.1007/s10489-017-1102-1
13. Fazel Zarandi MH, Seifi A, Ershadi MM, Esmaeeli H (2018) An expert system based on fuzzy bayesian network for heart disease diagnosis. Adv Intell Syst Comput 648:191–201. https://doi.org/10.1007/978-3-319-67137-6-21

14. Fan CY, Chang PC, Lin JJ, Hsieh JC (2011) A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Appl Soft Comput J 11(1):632–644. https://doi.org/10.1016/j.asoc.2009.12.023
15. Paul AK, Shill PC, Rabin MRI, Murase K (2018) Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. Appl Intell 48(7):1739–1756. https://doi.org/10.1007/s10489-017-1037-6
16. Adeli A, Neshat M (2010) A fuzzy expert system for heart disease diagnosis. In: Proc. Int. MultiConference Eng. Comput. Sci. 2010, IMECS 2010, pp 134–139
17. Soni S, Vyas OP (2013) Building weighted associative classifiers using maximum likelihood estimation to improve prediction accuracy in health care data mining. J Inf Knowl Manag. https://doi.org/10.1142/S0219649213500081
18. Alwidian J, Hammo BH, Obeid N (2018) WCBA: Weighted classification based on association rules algorithm for breast cancer disease. Appl Soft Comput J 62:536–549. https://doi.org/10.1016/j.asoc.2017.11.013
19. Ramasamy S, Nirmala K (2017) Disease prediction in data mining using association rule mining and keyword based clustering algorithms. Int J Comput Appl 7074:1–8. https://doi.org/10.1080/1206212X.2017.1396415
20. Horný M (2014) Bayesian networks: A Technical report. Commun ACM 53(5):15. http://www.bu.edu/sph/files/2014/05/bayesian-networks-final.pdf%0Ahttp://portal.acm.org/citation.cfm?doid=1859204.1859227
21. Xie J, Liu Y, Zeng X, Zhang W, Mei Z (2017) A Bayesian network model for predicting type 2 diabetes risk based on electronic health records. Mod Phys Lett B 31(19–21):1–6. https://doi.org/10.1142/S0217984917400553
22. Topuz K, Zengul FD, Dag A, Almehmi A, Yildirim MB (2018) Predicting graft survival among kidney transplant recipients: a Bayesian decision support model. Decis Support Syst 106:97–109. https://doi.org/10.1016/j.dss.2017.12.004
23. Agrahari R et al (2018) Applications of Bayesian network models in predicting types of hematological malignancies. Sci Rep 8(1):1–12. https://doi.org/10.1038/s41598-018-24758-5
24. Ershadi MM, Seifi A (2020) An efficient Bayesian network for differential diagnosis using experts' knowledge. Int J Intell Comput Cybern 13(1):103–126. https://doi.org/10.1108/IJICC-10-2019-0112
25. Setiawan NA, Venkatachalam PA, Hani AFM (2009) Diagnosis of coronary artery disease using artificial intelligence based decision support system. In: Proceedings of the International Conference on Man-Machine Systems (ICoMMS), October, pp 11–13
26. AdelAzar KD (2019) A method for modelling operational risk with fuzzy cognitive maps and Bayesian belief networks. Expert Syst Appl 115:607–617. https://doi.org/10.1016/j.eswa.2018.08.043
27. Kingsley C (2020) Adaptive neuro fuzzy inference system for diagnosing coronavirus disease 2019 (COVID-19). Int J Intell Comput Inf Sci 20(2):1–31. https://doi.org/10.21608/ijicis.2020.40518.1027
28. Amadin FI, Bello ME (2019) A Bayesian belief network approach for predicting kernicterus. Niger J Technol 38(2):416. https://doi.org/10.4314/njt.v38i2.18
29. Simsek S, Dag A, Tiahrt T, Oztekin A (2020) A Bayesian belief network-based probabilistic mechanism to determine patient no-show risk categories. Omega. https://doi.org/10.1016/j.omega.2020.102296
30. Sunita Soni OPV (2012) Fuzzy weighted associative classifier : a predictive technique for health care data. Int J Comput Sci Eng Inf Technol 2(1):11–22, 2012. https://doi.org/10.5121/ijcseit.2012.2102.
31. UCI machine learning breast cancer dataset. http://csc.liv.ac.uk/~frans/KDD/software/LUCS-KDDDN/datasets/dataSet.html
32. Dutta P (2022) ORIGINAL RESEARCH A new association coefficient measure for the conflict management and its application in medical diagnosis. Int J Inf Technol. https://doi.org/10.1007/s41870-022-01000-0
33. Kaur I, Kumar V, Kavitha NT, Mohan P (2022) Maximum likelihood based estimation with quasi oppositional chemical reaction optimization algorithm for speech signal enhancement. Int J Inf Technol. https://doi.org/10.1007/s41870-022-01032-6
34. Manogaran G, Varatharajan R (2018) Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. Multimed Tools Appl, pp. 4379–4399
35. Salem H, Shams MY, Elzeki OM, Elfattah MA, Al-amri JF, Elnazer S (2022) Fine-tuning fuzzy KNN classifier based on uncertainty membership for the medical diagnosis of diabetes. Appl Sci 12(3):1–26. https://doi.org/10.3390/app12030950
36. Altilio R, Rosato A, Panella M (2018) A sparse Bayesian model for random weight fuzzy neural networks. IEEE Int Conf Fuzzy Syst 2018:1–7. https://doi.org/10.1109/FUZZ-IEEE.2018.8491645
37. Maheshwari V et al (2021) Nanotechnology-based sensitive biosensors for COVID-19 prediction using fuzzy logic control. J Nanomater. https://doi.org/10.1155/2021/3383146
38. Zeinulla E, Bekbayeva K, Yazici A (2020) Effective diagnosis of heart disease imposed by incomplete data based on fuzzy random forest. Conf Fuzzy Syst IEEE Int. https://doi.org/10.1109/FUZZ48607.2020.9177531
39. Tarle B, Akkalaksmi M (2019) Improving classification performance of neuro-fuzzy classifier by imputing missing data. Int J Comput 18(4):495–501. https://doi.org/10.47839/ijc.18.4.1619
40. Kanimozhi U, Ganapathy S, Manjula D, Kannan A (2019) An intelligent risk prediction system for breast cancer using fuzzy temporal rules. Natl Acad Sci Lett 42(3):227–232. https://doi.org/10.1007/s40009-018-0732-0