An Interval Type-2 Fuzzy Set Approach to Breast Cancer Dataset Analysis

Olufunke Oladipupo, Oluwole Olajide, Simisola Olawoye, Jelili Oyelade and Adubi Stephen

***Abstract*--** The analysis of medical data is frequently characterized with uncertainties which tend to attract complexity. Therefore in this paper, an Interval Type-2 fuzzy set model: Hao and Mendel Approach (HMA) is proposed to fuzzify breast cancer data in order to handle quantitative attribute sharp boundary problem and resolve inter and intra uncertainties. The HMA comprises of the data and the fuzzy part to create interval type-2 fuzzy values. The data part involves data preprocessing of the experts' intervals and the fuzzy set part establishes the structure of the FOU. The type reduction of the aggregated FOU is achieved by computing the centroid (measure of uncertainty) of the Fuzzy Set using the Enhanced Kernik-Mendel (EKM) approach. The defuzzification of the outcome which is an interval Type-2 Fuzzy set is achieved by computing the average of the interval's two endpoints; this captures and reflects the aggregate uncertainty of all the medical experts for breast cancer analysis. This will enhance interpretability of discrete intervals in medical dataset, providing a smooth transition from a fuzzy set to another in order to handle the sharp boundary interval problem and cater for inter and intra uncertainty in data interval value as the same word has diverse connotations to different people.

Index Terms-- Breast Cancer, Interval Type-2 Fuzzy Set, Hao and Mendel Approach (HMA), Medical dataset.

I INTRODUCTION

MEDICAL databases are developing in an increasingly fast way with a huge measure of quantitative attributes. Analyzing medical data is crucial for decision making and medical administration [1]. And also for the discovery of new patterns that can be mined by analysing sample collections of example cases, defined by symbolic or numeric descriptors [2]. Cancer is a dangerous disease which is inherently caused by environmental factors that transform and mutate genes encoding critical cell-regulatory proteins [3]. Breast cancers are potentially life-threatening menaces that are formed in one or both breasts. Roughly one in 26 women is at risk of being diagnosed with breast cancer in their life time [4]. With prompt detection, there will be high rate of survival. About 97% women survive for five years or more [5].

Therefore, it is important to have a proficient and adequate medical data analysis that effectively handling the uncertainties from human intervention or opinion during the process of screening and identifying abnormality of the body in order to reduce erroneous diagnosis of fatal consequences.

Fuzziness in a survey data could be due to variations in opinion of one individual over repeated survey (Intra-expert) or variations in opinion between individuals (Inter-expert). [6] proposed a fuzzy type-1 set concept that captures a level of uncertainty called intra-uncertainties in decision making process. Intra-uncertainties mean the "uncertainty that a person has about a word" [7]. Type-1 fuzzy set has been widely applied in literature because they can manage large heterogeneous data sets, cater for sharp boundary problem and results are humanly understandable [8], [9], [10], [11]. Despite the uncertainties that are being modelled by type-1 fuzzy set, it cannot accurately reflect the linguistic uncertainties of diverse opinions from different domain experts which are very important in any decision-making process.

However, [12] proposed type-2 fuzzy set. This has a capacity to model all intra-uncertainties and inter-uncertainties in the process of making decisions. Inter-uncertainty means "uncertainty that captures a group of people's intra-uncertainties about a word" [7]. Due to the computational requirements of the type-2 fuzzy set, the interval type-2 fuzzy set was suggested. The interval type-2 fuzzy set models diverse opinions from different individuals by characterizing its members as membership grades of type-1 fuzzy set and can accommodate situations where precisely defined membership function may not be feasible for a fuzzy set. This makes interval type-2 suitable for capturing linguistic uncertainties where the same word has diverse connotations to different people. Therefore, due to the quantitative nature of most of the medical dataset there is need for an interval type 2 fuzzification process in analyzing medical dataset. This is important not only to resolve sharp boundary problem but also to cater for intra and inter uncertainties among the attribute values. This will in turn enhance the intuitive representation of the dataset in preprocessing for data analysis, datamining and prediction.

The organization of the remaining part of the paper is as follows: the related works were review in section 2, followed by methodology in section 3. Section 4 captures the experimental result and discussion and conclusion in section 5.

II RELATED WORKS

Data analysis commonly involves intricacy, vulnerability and uncertainty. The utilization of fuzzy logic has been advocated to deal with uncertainty and give clinicians intuitive outcomes through linguistic rules [13]. Also, fuzzy logic has been shown to be a superior system to improve understanding of discrete intervals [1] and provide a smooth

Manuscript received July 24, 2017; revised August7, 2017. This work was supported and sponsored by Covenant University, Ota. Nigeria. All authors are with Department of Computer and Information Sciences, Covenant University
*Corresponding Author: Olufunke Oladipupo.
O. Oladipupo (+2348033319370
funke.oladipupo@covenantuninversity.edu.ng)
O. Olajide (Oluwole.olajide@stu.cu.edu.ng)
S.Olawoye (Simisola.olawoye@stu.cu.edu.ng)
J. Oyelade (jelili.oyelade@covenantuniversity.edu.ng)
S. Adubi (stephen.adubi@covenantuniversity.edu.ng)

transition from a fuzzy set to another to handle the sharp boundary interval problem.

In recent times, some fuzzy logic approaches from different researchers had been presented. [10] used fuzzy C-Means to define membership functions from quantitative dataset to enhance prediction accuracy and performance. In [14] fuzzy logic and text-mining was applied into fuzzy cognitive map to mine scenario concepts from futuristic data in unstructured document. [15] presented an approach fusing triangular and trapezoidal waveform of fuzzification on ubiquitous data streams. [16] used fuzzy set theory using membership function based on clustering approach (fuzzy k-means clustering) on quantitative data to mine rules using mutual information. [9] developed a fuzzy logic approach to find correlation relationship among a large set of data items on gene expression data. [11] applied fuzzy logic on co-movement analysis of Indonesian stock price using fuzzy parameters on categorical data and using triangular curve for the process of fuzzification.

In other computational intelligence methods apart from fuzzy logic, [17] collected breast cancer dataset from a regional teaching hospital in central Taiwan between 2002 and 2009. Classification results of SVM are slightly better as compared to ANN and Bayesian classifier and the paper shows that from a relatively low variance, SVM will be the best prognosis in clinical practice. [18] showed how decision trees are used to model actual diagnosis of Breast cancer for local and systematic treatment. The results showed that J48 classifiers with feature selection is a superior technique that can be applied on breast cancer diagnosis and can further be developed with more training data to accurately predict the same. The limitation to these methods is that the modeling of diverse opinions from different medical experts and the aggregation of the uncertainties in the opinion of all the medical experts is worth an exploration as it contributes to the intuitiveness and human perception of the breast cancer domain.

However, from the different fuzzy logic research studies reviewed, the fuzzy set engaged in the medical data analysis was basically type-1 fuzzy set which uses precise real numbers to represent fuzziness measures. The effect of this is that, the fuzzy membership functions are model based on an opinion from one individual over a repeated survey i.e. Intra-expert [7]. Therefore, fuzzy set impact could only resolve the intra uncertainty which caters for a low level of subjectivity without catering for the linguistic uncertainties of diverse opinions from different domain experts. These are very important in any decision-making process [19]. Furthermore, a change in environmental and operating conditions can render type-1 fuzzy set sub-optimal. In order to cater for a high level of subjectivity and resolve both intra and inter uncertainties, an extension to the idea of fuzzy sets has been developed. The different fuzzy logic approaches include general type-2 fuzzy logic, interval type-2 fuzzy logic and the type-n fuzzy logic. The type-2 membership grades are typically type one fuzzy sets. Type-2 fuzzy set can handle both inter- and intra-uncertainties i.e. it can effectively model diverse opinions. There are different Type-2 fuzzy set approaches that have been proposed in literature such as enhanced interval type-2 (IT2) fuzzy set [20], Interval Approach, [19] Enhanced Interval Approach [21] and Hao and Mendel Approach [22]. Research has shown how Interval type-2 fuzzy logic system outperforms type-1 fuzzy logic system [23], [24], [19], [25], [26], [27], [28],[29]. In addition, there is no effective utilization of IT2 models for breast cancer dataset analysis.

To this effect, in this study, an IT2 Fuzzy Set approach, Hao and Mendel Approach (HMA) is introduced to fuzzify breast cancer dataset in preprocessing for data mining process. It is simpler using the HMA and requires less probability suppositions and assumptions about the intervals. Therefore, in order analyse breast cancer data using HMA, six general objectives have been identified as shown in Fig. 1.

III METHODOLOGY

In this work, breast cancer analysis using IT2 Fuzzy Set, HMA is distinctly covered as shown in Fig. 1. The HMA was used to model the inevitability of uncertainties of the medical experts that gave their respective ranges on each of the breast cancer determinant factors present in the dataset.

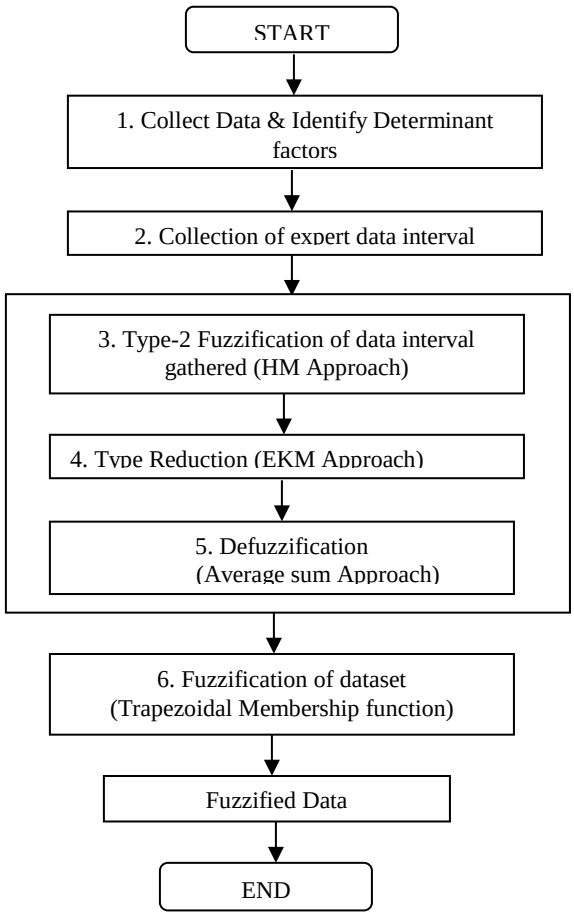

Fig. 1 Interval Type-2 Fuzzy Set framework

1. Data collection

To implement the method for this study, dataset which includes the determinant factors for pattern discovery was used. Breast cancer data was collected from the Wisconsin Breast Cancer dataset derived from the repository of the UCI machine learning data. Dataset consist of 10 attributes and 699 instances with 241 malignant (cancerous) and 458

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

benign (non-cancerous) cases. The attributes are Clump Thickness, Uniformity of Cell Size,, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli with the Class attribute distinguishing a malignant sample from a benign sample. These determinant factors are often chosen as the optimum set of factors because: 1) They represent effective features which reduce redundancy of features space;2) They give significant size of a large proportional grouping capacity with available limited number of training data [30]

The dataset has 16 missing values under the Bare Nuclei attribute. In order to ascertain that the breast cancer dataset contain only the attributes that we intend to process, attribute selection was carried out based on the determinant factors. Also, in making provision for the missing values, the mean of the other values under the affected attribute was taken and used to replace the missing attributes

.

2. Expert Data Interval

In this work, linguistic terms (words) were defined for eliciting the perception of each medical expert as regards the determinant factors and how the factors can lead to a malignant case of breast cancer. Thirty (30) medical experts from specialized hospitals were consulted for the expert data interval to define the intuitive words for the determinant factors and the data intervals. Words such as {High, Medium, Low} were defined to express each determinant factor. Subsequently, each word's interval data defined were collected from the medical experts using questionnaire approach. Experts were required to give a "range" that is between 0 and 10 for all word defined relative to each determinant factor.

3. Interval Type-2 Fuzzy Set: HM Approach

Fuzzy set was constructed from dataset using Interval Type-2 Fuzzy Set HMA which comprises of the data part and the fuzzy part [22]. The data part takes data intervals from the experts as the input [21]. This part which acts on the interval endpoints starting with the *n* intervals collected from all subjects are processed in 4 steps: (1) Bad data processing, (2) Outlier processing (3) Tolerance Limit Processing (4) Reasonable- interval processing [21**]**.

Also, the fuzzy set part established the FOU structure by making computations on the overlap of the intervals, removing the overlap from each of the original intervals and mapping the smaller interims to corresponding FOU.The part is follows in four steps according to [22].

4&5. Type Reduction and Defuzzification

The concurrent FOU is type-reduced by computing the centroid (uncertainty measure) of the IT2 FS using the Enhanced Kernik-Mendel (EKM) approach [31]. The result is defuzzified by obtaining the interval endpoints' average.

6. Fuzzification of Dataset

For the fuzzification process the trapezoidal membership function is applied. The function model is shown in eq. 1. It is represented by 4 variables {a',b',c',d'} deciding the x coordinates of the four angles.

$$trapezoid(x;a,b,c,d) = \max\left(\min\left(\frac{x-a\prime}{b\prime-a\prime}, 1, \frac{d\prime-x}{d\prime-c\prime}\right), 0\right) \quad (1)$$

IV EXPERIMENTAL RESULT AND DISCUSSION

The medical experts defined data intervals for each of the determinant factors. An interval or a range was used to describe a linguistic term corresponding to the determinant factors. The screenshot of the data intervals described by the medical experts for linguistic terms with regards to one of the determinant factors "Clump Thickness" is shown Figure 2.

As depicted in Fig. 2, the first medical expert defined the interval of [0, 3] for the word: Low; an interval of [2, 4] for the word Medium; High between [4, 10]. Meanwhile, the second medical expert defined the interval [0, 4.5] for the same word: Low; for Medium, an interval of [1, 3]; High between [3.5, 10]. Also the third decision maker defined the interval [0, 4] for the same word: Low; for Medium, an interval of [2.2, 5]; High between [4.5, 10]. These shows, there are different interpretations of the same word to the different decision makers.

	A	B	C	D	E	F
1	LOW		MEDIUM		HIGH	
2	0	3	2	4	4	10
3	0	4.5	1	3	3.5	10
4	0	4	2.2	5	4.5	10
5	0	4	1	5	3.3	10
6	0	3.5	1	5	2	10
7	0	3.5	2	4	3	10
8	0	3	3	5	1	10
9	0	3	2	4	4	10
10	1	3	4	2.1	7	6.5
11	0	3.5	1	3	4.5	10
12	0	3	3	5	4	10
13	0	4	3	4	4	9
14	0	4.5	3	5	1	10
15	0	4	2.5	4.9	4.5	10
16	0	4.5	2	4	3	10
17	0	5	2.5	4.5	4	10
18	0	6	4	7	3.5	10
19	0	3	2.5	4.5	4.5	10
20	0	3.5	2	4.5	3.5	10
21	0	4	2.5	4	3.5	10
22	0	3	3	5	5	10
23	0	4	4	6	6	10
24	0	3.5	3	5	4	10
25	0	3.5	3	6	5	10

Sheet1

Fig. 2 Screenshot of the data intervals described by medical experts for the linguistic terms

From IT2 data pre-processing Table I shows the remaining interval data after each stage in the Data part and Fuzzy Set part of the HMA Interval Type 2 Algorithm for each determinant factor. After applying the HM Approach to the data intervals given by 30 medical experts about each word. The last column for each row shows the credible interval data remaining, that was utilized finally for constructing footprint of uncertainty for that word. This established the maxim that "each word now means similar things to different people (medical experts)" from the initial maxim of "words mean different things to different people".

The type-2 fuzzy set derived for each word after the processing above and the values obtained after type-reduction using the Enhanced Karnik-Mendel (EKM)

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Approach are also shown in Table II. Also, the Jaccard similarity measure for the linguistic term define in relation to each determinant factor is shown in Table III. The monotonically decreases in the results ascertains the sufficiency of the linguistic words defined by the expert. The UMF i.e. Upper Membership Function and LMF i.e. Lower Membership Function parameters for each word of the determinant factors using the MATLAB are represented in Fig. 3 and 4

The crisp dataset collected are fuzzified based on the IT2 fuzzy set membership expression generated from the HMA algorithm. The snapshot representing an instance of the fuzzified dataset is shown in Figure 5. The fuzzified dataset now captures the intra-uncertainties and inter uncertainties of medical experts on the determinant factors. Unlike Type-1 fuzzification process, in IT2, individual linguistic term is fuzzified separately based on the expert's opinion's data intervals of the term relative to the linguistic variable. This is to cater for both intra uncertainty and inter uncertainties.

Table I The Output from the HMA process

Processing		Data Part				Fuzzy Part
Word	n	Step 1 *n'*	Step 2 *m'*	Step 3 *m''*	Step 4 *m''*	Final Step *m**
"Clump Thickness"						
Low	30	30	28	25	25	25
Medium	30	29	28	24	23	19
High	30	29	26	26	23	23
"Uniformity of Cell Size"						
Low	30	30	29	29	28	28
Medium	30	30	27	27	26	14
High	30	30	29	26	25	25
"Uniformity of Cell Shape"						
Low	30	30	29	29	28	28
Medium	30	29	28	28	26	21
High	30	30	29	29	29	29
"Marginal Adhesion"						
Low	30	30	26	25	25	25
Medium	30	30	30	28	27	17
High	30	30	29	29	29	29
"Single Epithelial Cell Size"						
Low	30	30	29	28	27	27
Medium	30	30	27	26	25	20
High	30	30	30	30	29	29
"Bare Nuclei"						
Low	30	30	28	28	28	28
Medium	30	30	29	24	24	17
High	30	30	28	28	28	28
"Bland Chromatin"						
Low	30	30	27	27	27	27
Medium	30	29	29	28	27	20
High	30	30	28	28	28	28
"Normal Nucleoli"						
Low	30	30	27	27	27	27
Medium	30	20	28	28	25	22
High	30	30	29	29	28	28
"Mitoses"						
Low	30	30	28	28	28	28
Medium	30	27	27	21	21	7
High	30	29	27	27	27	27

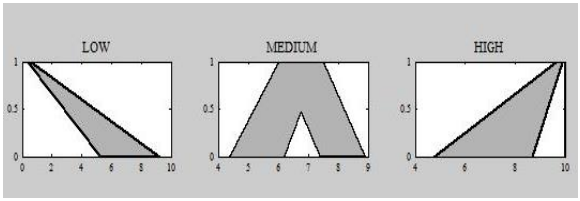

Fig. 3 "Clump thickness"

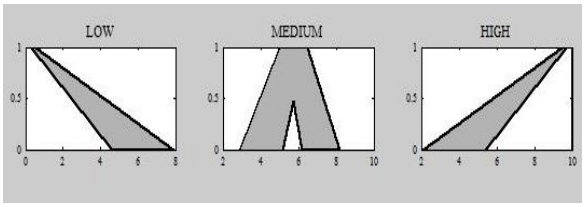

Fig. 4 "Uniformity of cell size"

Table II Output from Type reduction using EKM algorithm and Deffuzification process

Word	Low	Medium	High
"Clump Thickness"			
Low	1.00	0.23	0.08
Medium	0.23	1.00	0.11
High	0.08	0.11	1.0
"Uniformity of Cell Size"			
Low	1.00	0.12	0.07
Medium	0.12	1.00	0.26
High	0.07	0.26	1.00
"Uniformity of Cell Shape"			
Low	1.00	0.15	0.09
Medium	0.15	1.00	0.25
High	0.09	0.25	1.00
"Single Epithelial Cell Size"			
Low	1.00	0.16	0.09
Medium	0.16	1.00	0.20
High	0.09	0.20	1.00

V. CONCLUSION AND FUTURE WORK

In this work, an Interval Type-2 Fuzzy Set using the HM Approach is proposed to Fuzzify breast cancer dataset. The breast cancer dataset from Wisconsin Breast cancer dataset in UCI machine learning data repository was used. The breast cancer fuzzy dataset generated can be used for data mining process. The fuzzy models constructed with the IT2 fuzzy captured different expert opinion in order to cater for sharp boundary problem and capture domain experts intra and inter uncertainty that limit the traditional fuzzy type 1 process. This will in turn enhance the rules generated or models constructed from the breast cancer dataset.

Thus, for future improvements on this study, another elicitation methodology that best covers the scope of the determinant factors considered can be proposed in establishing the variables of the interval type-2 fuzzy sets, construction of the FOU of linguistic terms/words defined and incorporated with Mining algorithm.

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table III Similarities among the three words using the Jaccard similarity measure

Word	UMF	LMF	Centroid	Center of centroid
	"Clump Thickness"			
Low	[0, 0,0.41,5.92]	0, 0, 0.28,3.95,1]	[1.33, 2.04]	1.68
Medium	[1.48, 3.00, 4.00, 5.58]	[2.59, 3.50, 3.50, 4.31, 0.65]	2.91, 4.09]	3.50
High	[0.78, 9.36, 10, 10]	[3.42, 9.54, 10, 10, 1]	[6.84,7.79]	7.3
	"Uniformity of Cell Size"			
Low	[0, 0, 0.64, 9.22]	[0, 0, 0.37, 5.27, 1]	[1.77, 3.21]	2.49
Medium	[4.38, 6.00, 7.50, 8.91]	[6.19, 6.75, 6.75,7.41, 0.47]	[5.79, 7.64]	6.71
High	[4.73, 9.63, 10, 10]	[8.68, 9.91, 10, 10, 1]	[7.92, 9.55]	8.73
	"Uniformity of Cell Shape"			
Low	[0, 0, 0.55, 7.90]	[0, 0, 0.32, 4.61, 1	[1.55, 2.75]	2.15
Medium	[2.88, 5.00, 6.50, 8.22]	[5.19, 5.75, 5.75, 6.21, 0.47]	[4.45, 6.82]	5.64
High	[2.10, 9.45, 10, 10]	[5.39, 9.68, 10, 10, 1]	[7.25, 8.45]	7.85
	"Marginal Adhesion"			
Low	[0, 0, 0.50, 7.24]	[0, 0, 0.36, 5.13, 1]	[1.72, 2.48]	2.10
Medium	[4.09, 5.50, 6.50, 7.91]	[5.29, 5.83, 5.83, 6.21, 0.53]	[5.17, 6.68]	5.92
High	[2.10, 9.45, 10, 10]	[6.71, 9.77, 10, 10, 1]	[7.16, 8.89]	8.03
	"Single Epithelial Cell Size"			
Low	[0, 0, 0.60, 8.56]	[0, 0, 0.28, 3.95, 1]	[1.33, 3.04]	2.18
Medium	[3.38, 5.00, 6.50, 7.91]	[5.19, 5.75, 5.75, 6.21, 0.47]	[4.71, 6.67]	5.69
High	[3.42, 9.54, 10, 10]	[7.37, 9.82, 10, 10, 1]	[7.62, 9.11]	8.37
	"Bare Nuclei"			
Low	[0, 0, 0.55, 7.90]	[0, 0, 0.18, 2.63, 1]	[0.89, 2.94]	1.91
Medium	[2.38, 4.00, 5.50, 7.02]	[4.09, 4.75, 4.75, 5.31, 0.47]	[3.76, 5.67]	4.72
High	[2.10, 9.45, 10, 10]	[7.37, 9.82, 10, 10, 1]	[7.06, 9.11]	8.09
	"Bland Chromatin"			
Low	[0, 0, 0.55, 7.90]	[0, 0, 0.37, 5.27, 1]	[1.77, 2.72]	2.25
Medium	[3.38, 5.00, 6.50, 8.02]	[5.41, 5.75, 5.75, 6.21, 0.47]	[4.66, 6.81]	5.73
High	[3.42, 9.54, 10, 10]	[7.37, 9.82, 10, 10, 1]	[7.62, 9.11]	8.37
	"Normal Nucleoli"			
Low	[0, 0, 0.64, 9.22]	[0, 0, 0.28, 3.95, 1]	[1.33, 3.30]	2.31
Medium	[3.28, 5.00, 6.60, 8.72]	[5.19, 5.80, 5.80, 6.21, 0.43]	[4.64, 7.14]	5.89
High	[3.42, 9.54, 10, 10]	[7.37, 9.82, 10, 10, 1]	[7.62, 9.11]	8.37
	Mitoses"			
Low	[0, 0, 0.60, 8.56]	[0, 0, 0.28, 3.95, 1]	[1.33, 3.04]	2.18
Medium	[4.59, 5.50, 6.50, 7.91]	[5.09, 5.83, 5.83, 6.21, 0.53]	[5.40, 6.60]	6.00
High	[2.10, 9.45, 10, 10]	[8.03, 9.86, 10, 10, 1]	[6.88, 9.33]	8.10

	A	B	C	D	E	F	G	H
1	LOW CLUMP THICKNESS	MEDIUM CUMP THICKNESS	HIGH CLUMP THICKNESS	LOW UNIFORMITY OF CELL SIZE	MEDIUM UNIFORMITY OF CELL SIZE	HIGH CELL SIZE	LOW CELL SHAPE	MEDIUM CELL SHAPE
2	0.00	0.00	0.39	0.93	0.00	0.00	0.90	0.
3	0.00	0.00	0.39	0.48	0.00	0.00	0.39	0.
4	0.42	0.79	0.12	0.93	0.00	0.00	0.90	0.
5	0.00	0.00	0.53	0.00	0.16	0.42	0.00	0.
6	0.20	0.79	0.26	0.93	0.00	0.00	0.90	0.
7	0.00	0.00	0.80	0.00	0.00	1.00	0.00	0.
8	0.86	0.00	0.00	0.93	0.00	0.00	0.90	0.
9	0.64	0.00	0.00	0.93	0.00	0.00	0.73	0.
10	0.64	0.00	0.00	0.93	0.00	0.00	0.90	0.
11	0.20	0.79	0.26	0.78	0.00	0.00	0.90	0.
12	0.86	0.00	0.00	0.93	0.00	0.00	0.90	0.
13	0.64	0.00	0.00	0.93	0.00	0.00	0.90	0.
14	0.00	0.00	0.39	0.63	0.00	0.00	0.56	0.
15	0.86	0.00	0.00	0.93	0.00	0.00	0.90	0.
16	0.00	0.00	0.80	0.04	1.00	0.09	0.22	0.
17	0.00	0.00	0.67	0.48	0.00	0.00	0.04	1.
18	0.20	0.79	0.26	0.93	0.00	0.00	0.90	0.
19	0.20	0.79	0.26	0.93	0.00	0.00	0.90	0.
20	0.00	0.00	1.00	0.04	1.00	0.09	0.00	0.
21	0.00	0.00	0.53	0.93	0.00	0.00	0.90	0.
22	0.00	0.00	0.67	0.63	0.00	0.00	0.73	0.
23	0.00	0.00	1.00	0.33	0.00	0.00	0.22	0.
24	0.42	0.79	0.12	0.93	0.00	0.00	0.90	0.
25	0.00	0.00	0.80	0.48	0.00	0.00	0.22	0.

Fig. 5 Screenshot of the fuzzified dataset

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

REFERENCE

[1] Delgado, M., SáNchez, D., MartıN-Bautista, M. J., and Vila, M. A. (2001). Mining association rules with improved semantics in medical databases. Artificial Intelligence in Medicine, 21(1), 241-245.

[2] Lavrac, N., Keravnou, E., and Zupan, B. (1997). Intelligent data analysis in medicine and pharmacology. Kluwer.

[3] Alison, M. R., Poulsom, R., and Forbes, S. J. (2001). Update on hepatic stem cells. Liver International, 21(6), 367-373.

[4] Nwaneri, A., Osuala, E. O., Okpala, P. U., Emesowum, A. C., and Iheanacho, P. (2017). Knowledge and awareness of breast cancer among rural women in UmuowaOrlu Local Government Area Imo State, South East, Nigeria. Nigerian Journal of Clinical Practice, 20(4), 489.

[5] Yeh, W. L., Lu, D. Y., Lee, M. J., and Fu, W. M. (2009). Leptin induces migration and invasion of glioma cells through MMP-13 production. Glia, 57(4), 454-464.

[6] Zadeh, L. A (1965). Fuzzy sets. Journal of Information and Control. Vol. 8: 338-353.

[7] Mendel, J. M. (2007). Computing with words and its relationships with fuzzistics. Information Sciences, 177(4), 988-1006.

[8] Karabatak, M., and Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. Expert systems with Applications, 36(2), 3465-3469.

[9] Barik, S., Mishra, D., Mishra, S., Satapathy, E. K., Rath, A. K., and Acharya, M. (2010). Pattern discovery using fuzzy FP-growth algorithm from gene expression data. International Journal of Advanced Computer Science and Applications, Vol. 1, No. 5, 50-55.

[10] Sowan, B., Dahal, K., Hossain, M. A., Zhang, L., and Spencer, L. (2013). Fuzzy association rule mining approaches for enhancing prediction performance. Expert Systems with Applications, 40(17), 6928-6937.

[11] Arafah, A. A., and Mukhlash, I. (2015). The Application of Fuzzy Association Rule on Co-movement Analyze of Indonesian Stock Price. Procedia Computer Science, 59, 235-243.

[12] Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. Information sciences, 8(3), 199-249

[13] Nguyen, T., Khosravi, A., Creighton, D., and Nahavandi, S. (2015). EEG signal classification for BCI applications by wavelets and interval type-2 fuzzy logic systems. Expert Systems with Applications, 42(9), 4370-4380.

[14] Kim, J., Han, M., Lee, Y., and Park, Y. (2016). Futuristic data-driven scenario building: incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. Expert Systems with Applications, 57, 311-323.

[15] Moustafa, A., Abuelnasr, B., and Abougabal, M. S. (2015). Efficient mining fuzzy association rules from ubiquitous data streams. Alexandria Engineering Journal, 54(2), 163-174.

[16] Lotfi, S., and Sadreddini, M. H. (2009). Mining fuzzy association rules using mutual information. In Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 1). IMECS 2009, March 18 - 20, 2009, Hong Kong.

[17] Yu, C., Chao, C., and Cheng, B. (2014) Prediction Survival in Patients with Breast Cancer using three artificial intelligence Techniques. Journal of Theoretical and Applied Information Technology . Vol. 60 No.1 pp 179-184

[18] Sumbaly, R., Vishnusri N. and Jeyalatha, S. (2014) Diagnosis of Breast Canser using Decision Tree Data Mining Technigue. International Journal of Computer Applications (0975 – 8887) Volume 98– No.10, July 2014.pp 16-24

[19] Mendel. J.M and Wu, D. (2010) Perceptual Computing: Aiding People in Making Subjective Judgement . Willey. IEEE Press ISBN: 9780470478769. Online ISBN: 9780470599655. DOI: 10.1002/9780470599655

[20] Mendel, J. M. (2003, May). Fuzzy sets for words: a new beginning. In Fuzzy Systems, 2003. FUZZ'03. The 12th IEEE International Conference on (Vol. 1, pp. 37-42). IEEE.

[21] Wu, D., Mendel, J. M., and Coupland, S. (2012). Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis. IEEE Transactions on Fuzzy Systems, 20(3), 499-513.

[22] Hao, M., and Mendel, J. M. (2016). Encoding words into normal interval type-2 fuzzy sets: HM approach. IEEE Transactions on Fuzzy Systems, 24(4), 865-879.

[23] Hosseini, S. M. S., Maleki, A., and Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Systems with Applications, 37(7), 5259-5264.

[24] Sanz, J. A., Fernández, A., Bustince, H., & Herrera, F. (2010). Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning. Information Sciences, 180(19), 3674-3685.

[25] Herman, P., Prasad, G., McGinnity, T. M., & Coyle, D. (2008). Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 16(4), 317-326.

[26] Chua, T. W., and Tan, W. W. (2008, December). Genetically evolved fuzzy rule-based classifiers and application to automotive classification. In Asia-Pacific Conference on Simulated Evolution and Learning (pp. 101-110). Springer Berlin Heidelberg.

[27] Chumklin, S., Auephanwiriyakul, S., and Theera-Umpon, N. (2010, July). Microcalcification detection in mammograms using interval type-2 fuzzy logic system with automatic membership function generation. In Fuzzy Systems (FUZZ), 2010 IEEE International Conference on (pp. 1-7). IEEE.

[28] Abiyev, R. H. (2011). Fuzzy wavelet neural network based on fuzzy clustering and gradient techniques for time series prediction. Neural Computing and Applications, 20(2), 249-259.

[29] Albarracin, L. F., and Melgarejo, M. A. (2010, July). An approach for channel equalization using quasi type-2 fuzzy systems. In Fuzzy Information Processing Society (NAFIPS), 2010 Annual Meeting of the North American (pp. 1-5). IEEE.

[30] Salama, G. I., Abdelhalim, M., and Zeid, M. A. E. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. Breast Cancer (WDBC), 32(569), 2.

[31] Wu, D., & Mendel, J. M. (2009). Enhanced karnik--mendel algorithms. IEEE Transactions on Fuzzy Systems, 17(4), 923-934.