An Interval Type-2 Fuzzy Association Rule Mining Approach to Pattern Discovery in Breast Cancer Dataset

1Olufunke Oladipupo, 1Oluwole Olajide, 1Stephen Adubi, 1Jelili Oyelade and 2Zacchaeus Omogbadegun

1Department of Computer and Information Sciences, Covenant University, Nigeria
2Department of Computer Science, Chrisland University, Nigeria

Abstract: In the literature, several methods explored to analyze breast cancer dataset have failed to sufficiently handle quantitative attribute sharp boundary problem to resolve inter and intra uncertainties in breast cancer dataset analysis. In this study an Interval Type-2 fuzzy association rule mining approach is proposed for pattern discovery in breast cancer dataset. In the first part of this analysis, the interval Type-2 fuzzification of the breast cancer dataset is carried out using Hao and Mendel approach. In the second part, FP-growth algorithm is adopted for associative pattern discovery from the fuzzified dataset from the first part. To define the intuitive words for breast cancer determinant factors and expert data interval, thirty (30) medical experts from specialized hospitals were consulted through questionnaire poling method. To establish the adequacy of the linguistic word defined by the expert, Jaccard similarity measure is used. This analysis is able to discover associative rules with minimum number of symptoms at confidence values as high as 91%. It also identifies High Bare Nuclei and High Uniformity of Cell Shape as strong determinant factors for diagnosing breast cancer. The proposed approach performed better in terms of rules generated when compared with traditional quantitative association rule mining. It is able to eliminate redundant rules which reduce the number of generated rules by 39.5% and memory usage by 22.6%. The discovered rules are viable in building a comprehensive and compact expert driven knowledge-base for breast cancer decision support or expert system.

Keywords: Data Mining, Breast Cancer, Interval Type-2 Fuzzy Association Rule Mining

Introduction

In today’s information age, medical databases are increasing rapidly with a large number of quantitative attributes. Analyzing this dataset is crucial for enhancing medical decision making and management (Delgado et al., 2001). In data analysis research, medical case studies like heart disease, diabetes and breast cancer are often considered because they present a combination of imprecise causal knowledge, very large amount of information and potentially life-threatening consequences of incorrect conclusion. Breast Cancer is a dangerous disease which is inherently triggered by environmental factors that mutate genes encoding critical cell-regulatory proteins (Alison et al., 2001). Some of the procedures for detecting cancer are self-examinations, biopsy, ultrasound, magnetic resonance imaging and computed tomography. Breast cancer is one of the life-threatening diseases and the leading form of cancer with second estimated highest death rate among women (Siegel et al., 2018). In 2018, about 266,120 United States women were found to have invasive breast cancer; about 40,920 of these women were estimated to have died (Siegel et al., 2018). Early detection of breast cancer incidence has been confirmed to increase the survival rate and reduce death rate (Ed-daoudy and Maalim, 2020). According to (Yeh et al., 2009) around 97% of women can survive for 5 years or more due to earlier diagnosis and improved treatment. This fact was further buttressed by American Cancer Society’s 2018 report. The report highlighted the decline death rate of about 39% from 1989 to 2015 by breast cancer due to early detection. The statistics above shows that early detection of breast cancer is vital to decreased life crisis.
However, quick detection requires a reliable and accurate diagnosis procedure that can predict the risk even without going for surgical biopsy.

Data mining analysis is capable of extracting desirable, previously unknown and potentially useful patterns from existing medical dataset for risk determination of the disease. In healthcare field, data mining has been widely utilized to discover patterns and relationships that occur among attributes in medical dataset for the prognosis and diagnosis of diseases (Aalaei et al., 2016; Nilashi et al., 2017; Sakri et al., 2018). Breast cancer dataset is fast growing with a big number of quantitative attributes, mostly from medical laboratory tests (Delgado et al., 2001). In considering the highly increasing rate of breast cancer dataset to control the breast cancer menace, a data mining approach is found appropriate (Ehtemam et al., 2017; Ed-daoudy and Maalimi, 2020). Hence, association rule mining is a valuable technique to discover patterns for breast cancer prediction. Association Rule Mining (ARM) is used to discover the hidden relationships and potential associations among items or attributes of a large dataset (Chiu et al., 2012). ARM has been widely used for pattern discovery in the medical domain (Gupta et al., 2006; Serban and Moldovan, 2006; Umesh and Ramachandra, 2015). This is because it can manage large heterogeneous data sets and the generated rules are intuitively interpretable and humanly understandable (Karabatak and Ince, 2009). It is capable of handling unreliable diagnosis tests in medical diagnosis and presence of noise in training examples (Serban and Moldovan, 2006). However, because of the quantitative nature of breast cancer datasets and the limitation of quantitative association rule mining it is essential to identify a technique for resolving the sharp boundary problem attributed with quantitative value intervals ( Kuok et al., 1998; Gynesesi, 2001; Oladipupo et al., 2010). This is to make provision for intra and inter uncertainties among the attribute values so as to enhance data mining analysis result.

Fuzzy logic concept referred to as Type-1 fuzzy was proposed by (Zadeh, 1975) to resolve subjectivity and intra uncertainty in decision making. Beyond this level of uncertainty, uncertainty within different domain experts’ opinions in decision making is another crucial issue needed to be resolved in order to enhance medical decisions. This concern motivated the introduction of type-2 fuzzy in (Mendel, 2001), which is able to make provision for both intra and inter uncertainties in analyzing quantitative dataset and defining fuzzy attribute values interval. The inter uncertainty is attributed to diverse domain experts’ opinions in defining fuzzy attribute values interval as “words mean different things to different people” (Mendel, 2007). Most of the recent studies on breast cancer data analysis attention are more on comparative analysis of predictive algorithms (Cesar et al., 2020; Mohammed et al., 2020; Vrigazova, 2020; Ak, 2020; Ahmed et al., 2020). Most of the studies did not pay attention to uncertainties attributed with quantitative clinical dataset before they analyze the dataset (Nguyen et al., 2015a). Few studies have engaged type-1 fuzzy parameters in breast cancer data mining analysis to cater for intra uncertainty, among which are (Nilashi et al., 2017; Jain and Abraham, 2004; Alharbi and Tchier, 2017). However, type-2 fuzzy, which is able to handle both intra and inter uncertainties with association rule mining, is yet to gain proper attention in analyzing breast cancer dataset. Therefore, in this study the hybrid of type-2 fuzzy and association rule mining in analyzing breast cancer dataset is introduced with the following two distinct objectives:

(i). Construction of interval values for breast cancer determinant factors and fuzzification of breast cancer dataset based on interval type-2 fuzzy parameters

(ii). Construction of frequent pattern tree and extraction of useful patterns for breast cancer risk determination

This analysis is expected to enhance the acquired knowledge from mining breast cancer dataset with domain experts’ opinions, eliminate redundant rules, minimize determinant symptoms and possibly identify strong determinant factors for diagnosing breast cancer. The patterns could be recommended for building a less subjective, compact and comprehensive decision support system for breast cancer risk determination. The remaining part of this paper is organized as thus: Section 2, presents review of literature on the interval type-2 fuzzy algorithms and related works. Section 3 discusses the proposed approach methodology and the dataset considered for the analysis. In section 4, the data analysis, result and discussion are presented. Section 5 presents the comparative evaluation of the proposed approach: Interval type-2 fuzzy ARM and the traditional quantitative ARM. Finally, in section 6 the paper is concluded by highlighting the important findings from the proposed approach and future work highlighted.

Literature Review

In this section Interval type-2 fuzzy set (IT2 FS) models approaches and related works are discussed.

Interval Type-2 Fuzzy Set (IT2 FS) Models Approaches

In capturing the uncertainties (both intra and inter) which limited the capability of Type-1 Fuzzy Set (T1 FS) the methodology of Computing With Word (CWW) must be established, because words mean different things to different people (Hao and Mendel, 2015). In the literature there are few methods that define an IT2 FS model for a word from the data collected about the word,
especially from a set of domain experts. Liu and Mendel (2008), an Interval Approach (IA) is proposed. This approach is divided into two sections: the data process section and the fuzzy process section. In the data process section statistics is used to reduce the data intervals while in the second section (Fuzzy process) each data interval that was not removed by the statistics is mapped into a Type 1 FS. After this the Footprint Of Uncertainty (FOU) is gotten for the word by bounding all of the Type 1 FSs both from below and above. This approach is limited by too fat and wide FOUs and the Lower Membership Functions (LMFs) of the interior FOUs usually have very small height (Wu et al., 2012).

In order to improve on the IA approach, Enhanced Interval Approach (EIA) is proposed in (Wu et al., 2012). In EIA there are more steps and conditions for data interval reduction compared to IA, which improve the effectiveness of data intervals reduction process. Also, EIA incorporates and improves procedures to compute the LMF. In comparing this approach with IA, it was reported in (Wu et al., 2012) that EIA converges in a mean-squared sense to a stable model and more data intervals are collected. Also in (Tahayori and Sadeghian, 2012) another method is proposed that calculates the median boundaries range of the membership function associated with the word. This method does not directly map the data interval uncertainties into Membership Function (MF) uncertainties, as do the IA and EIA and it leads to normal IT2 FSs. Other methods are also proposed in (Pagola et al., 2013; Bilgin et al., 2012; Moharrer et al., 2013). These approaches introduce another level of uncertainties called methodological uncertainty (Hao and Mendel, 2015). It is not possible to separate this uncertainty from linguistic uncertainties; therefore, any resulting FS model that uses such data does not represent a linguistic-uncertainty model (Hao and Mendel, 2015).

Hao and Mendel (2015) proposed another approach based on Computing With Words (CWW) to enhance the performance of EIA. This new approach makes more use of the information from the data intervals, for example, the use of the overlap shared by all intervals. This approach (HMA) also has two stages; the first stage identifies a group of data intervals while the second stage determines the common overlap for this group of data intervals. An Interval Type-2 FS is determined from a set of shorter remaining data intervals that exclude the overlap. This approach is unique among all the Interval Type-2 FS word models because the resulting Interval Type-2 FSs-Shoulders and Interior FOUs are normal. This approach is thereby adopted in this study for modeling breast cancer Interval Type-2 FS.

Related Works

Nowadays, different methodologies which include data mining methods have been richly explored in the field of Medicine to extract useful patterns for diagnosis, prediction and decision making. This exploration does not exclude the area of breast cancer diagnosis and prediction for early detection of the disease (Ed-daoudy and Malmi, 2020; Chaurasia et al., 2018). He et al. (2012), association rule mining was used to explore patterns for Chinese medicine formulae to treat and prevent breast cancer. Medjahed et al. (2013), K-nearest neighbour with different distance and classification rules were engaged for breast cancer diagnosis. Fuzzy analysis of breast cancer disease with fuzzy c-means and pattern recognition was carried out in (Muhic, 2013). Nguyen et al. (2015b), classification of breast cancer was done using the fuzzy C-means clustering algorithm. Breast cancer classification using deep belief network was carried out in (Abdel-Zaheer and Eldeib, 2016). Umesh and Ramachandra (2015), association rule mining was used to predict recurrence of breast cancer on SEER breast cancer data. Aalaei et al. (2016), feature selection using genetic algorithm for breast cancer diagnosis was proposed and experimented with artificial neural network, PS-classifier and generic algorithm classifier. Vijaylakshmi and Priyadarshini (2016), Particle Swarm Optimization technique was used to analyse breast cancer dataset. A knowledge-based system for breast cancer prediction was built with expectation maximization clustering algorithm and Classification and regression tree in (Nilashi et al., 2017). Ehtemam et al. (2017) used 64 data mining models in WEKA and MATLAB to build models for prognosis and early diagnosis of Ductal and Lobular Type in Breast cancer patient. Hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence using optimized ensemble learning was presented in (Mohesian et al., 2017). Sakri et al. (2018) implemented Naïve Bayes and K-nearest neighbor techniques with Particle Swarm Optimization feature selection method for predicting Breast cancer recurrence. Many studies have captured comparative analysis of different machine leaning and data mining algorithms for predicting breast cancer (Cesar et al., 2020; Mohammed et al., 2020; Vrigazova, 2020; Ak, 2020; Ahmed et al., 2020).

Furthermore, researchers have identified the use of fuzzy logic as a potent computational approach for handling uncertainty in clinical data and also for providing intuitive results through linguistic rules (Nguyen et al., 2015a). With respect to superiority in enhancing interpretability of discrete intervals in medical data sets, fuzzy logic is the go-to approach (Delgado et al., 2001). It also provides a smooth transition from one fuzzy set to another in handling the sharp boundary interval problem. The deficiency of quantitative association rule mining led to the introduction of Fuzzy Association Rule Mining (FARM) (Kuok et al., 1998; Gyenesei, 2001). This kind of rules generated by FARM not only addresses the “sharp boundary problem”, it
provides better expression of the association rules in terms of intuition and understandability and also minimizes the number of rules generated (Oladipupo et al., 2010). In the literature, there have been instances of Type-1 fuzzy association rule mining in different domains (Chiu et al., 2012; Oladipupo et al., 2012; Ho et al., 2012; Arafah and Mukhlash, 2015; Kim et al., 2016). The implication of this is that, the fuzzy membership functions were modelled base on an opinion from an individual over a repeated survey (Intra-expert) (Mendel, 2007). Therefore, type-1 fuzzy set impact on the quantitative value could only handle the intra uncertainty which make provision for a minimal level of subjectivity. Also, a change in environmental and operating conditions can render type-1 fuzzy set sub-optimal and research has shown how Interval type-2 fuzzy logic systems outperform type-1 fuzzy logic systems. Hence, type-2 fuzzy is found more appropriate and extensive in any decision-making process to cater for both intra and inter uncertainties (Mendel, 2001). This is very crucial in a life-threatening consequence domain like medical, where high level of accuracy is expected in building decision support or expert system.

Type-2 fuzzy system has been widely used in different fields in solving real life problems with good report, among others (Nguyen et al., 2015a; Hosseini et al., 2010; Pimenta and Camargo, 2010; Sanz et al., 2010; Wu and Mendel, 2009; Chumklin et al., 2010; Abiye, 2011; Castillo and Amador-Angulo, 2018); Castillo et al., 2016; Bobillo and Straccia, 2017; Rubio et al., 2017; Celik and Akvuz, 2018; Greenfield and Chiclana, 2018; Torres-Blanc et al., 2018). Chen et al., (2015), fuzzy association rule mining with Type-2 membership functions was investigated with synthetic dataset, which was reported to perform better in relations to number of rules. In spite of the efforts of researchers there is still a gap in the effective use of Interval type-2 models with Frequent Pattern Growth (FP-Growth) to analyse breast cancer dataset.

**Methodology**

In this study, pattern discovery in breast cancer dataset using Interval Type-2 fuzzy association rule mining is proposed and clearly covered in two major sequential stages. The proposed IT2 fuzzy association rule mining framework for Breast Cancer Data Analysis is shown in Fig. 1: The first stage is termed data fuzzification and the second stage is pattern discovery.

**Data Description**

In this study the Wisconsin Breast Cancer dataset from UCI machine learning data repository was utilized because of its effective utilization in the literature. The Dataset comprises 10 attributes and 699 instances used to differentiate malignant (cancerous) from benign (non-cancerous) samples. The first 9 attributes are the determinant factors for the breast cancer and the last attribute describes the patient breast cancer status. The dataset attribute is shown in Table 1. These determinant factors are often chosen as the optimum set of factors because: (1) They represent effective features which reduce redundancy of features space; and (2) They provide significant estimation of a high dimensional classification function with the available finite number of training data (Salama et al., 2012).

**Data Fuzzification**

In this section a detail description of IT2 fuzzification process for breast cancer dataset is discussed. This section is crucial to this study. Even though the process has been briefly examined in our previous study with results (Oladipupo et al., 2019), it is important that the process and the algorithms utilized are stated in this study for more clarification of the proposed approach.

**Expert Data Interval**

In this study, linguistic terms (words) were defined for eliciting the view of medical experts 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 word for the factors and the data intervals. The experts’ data intervals were collected using a form as shown in Fig. 2. Words such as {High, Medium, Low} were defined to express each determinant factor. These words were used in approximate reasoning by medical experts for gathering diverse opinions with respect to each determinant factor. In order to ascertain the adequacy of the linguistic word defined by the expert, Jaccard similarity measure was used to determine similarities among the three words. This was chosen because of the drawbacks attributed to the other five similarity measures for IT2 FSs (Wu and Mendel, 2008; Mendel and Wu, 2010). Afterward, the data intervals for each word defined were collected from the medical experts using a questionnaire polling method. Experts were required to describe an interval or a “range” that falls somewhere between 0 and 10 for intuitive words defined for all determinant factors.

**Table 1: Dataset description**

| S/N | Attribute               | Set of values               |
|-----|-------------------------|-----------------------------|
| 1   | ClumpThickness          | {1, 2, ..., 10}             |
| 2   | CellSizeUniformity      | {1, 2, ..., 10}             |
| 3   | CellShapeUniformity     | {1, 2, ..., 10}             |
| 4   | MarginalAdhesion        | {1, 2, ..., 10}             |
| 5   | SingleEpiCellSize       | {1, 2, ..., 10}             |
| 6   | BareNuclei              | {1, 2, ..., 10}             |
| 7   | BlandChromatin          | {1, 2, ..., 10}             |
| 8   | NormalNucleoli          | {1, 2, ..., 10}             |
| 9   | Mitoses                 | {1, 2, ..., 10}             |
| 10  | Class                   | {Malignant, benign}         |
Interval Type-2 Fuzzification Process

The interval type-2 fuzzification process was carried out using the Hao and Mendel Approach (HMA). The HMA mathematical model was simulated using Matlab. This is to take advantage of the efficiency and effectiveness of the HMA over other approaches in order to enhance the quality of the fuzzification output (Hao and Mendel, 2015) the HMA is achieved in two sections which are the data process section and the fuzzy process section.
COLLECTION OF WORD INTERVALS

Below are a number of labels that describe an interval or a range that falls somewhere between 0 and 10. For each label, please tell us where this range would start and where it would end. (In other words, please tell us how much of the distance from 0 to 10 this range for that word would cover). For example, the range for word “not so much” might start at 3.5 and end at 6.4 or start at 2 and end at 5. It is important to note that not all ranges are the same, and ranges can overlap.

1. Considering **CLUMP THICKNESS** as a sole determinant factor for breast cancer diagnosis, give your interval ranges between 0 and 10.

   | Low | Medium | High |
   |-----|--------|------|
   | From the range to |
   | From the range to |
   | From the range to |

2. Considering **UNIFORMITY OF CELL SIZE** as a sole determinant factor for breast cancer diagnosis, give your interval ranges between 0 and 10.

   | Low | Medium | High |
   |-----|--------|------|
   | From the range to |
   | From the range to |
   | From the range to |

---

**Section 1: Data Process Section**

The data part takes data intervals from the experts as the input (Wu et al. et al., 2013) and processed in 4 stages. The medical experts and processed in 4 stages.

**Bad Data Processing**

During this process, only intervals that satisfy the Eq. (1) are accepted; others are rejected because some subjects (the experts) might not take the survey serious:

\[
0 \leq a^{(0)} < b^{(0)} \leq 10 \text{ and } b^{(0)} - a^{(0)} < 10
\]

where, \( a^{(0)} \) is the low range of the interval and \( b^{(0)} \) is the upper range of the interval to be supplied by the medical experts. This step decreases \( n \) interval endpoints to \( n' \) interval endpoints.

**Processing of Outlier**

To remove outliers, Box and Whiskers tests (Walpole et al., 2013) are carried out on the remaining \( n' \) interval, then, only intervals that satisfy Eq. (2) and (3) are kept:

\[
a^{(0)} \in [Q_1(0.25)-1.5IQR_a, Q_3(0.75)+1.5IQR_a]
\]

(2)

\[
b^{(0)} \in [Q_1(0.25)-1.5IQR_b, Q_3(0.75)+1.5IQR_b]
\]

(3)

This process also decreases the \( n' \) interval endpoints to \( n'' \) interval endpoints.

Subsequently, \( Q_1(0.25), Q_3(0.75) \) and \( IQR \) are calculated using the remaining \( n \) intervals and only intervals that satisfy Eq. (4) are kept:

\[
L^{(0)} \in [Q_1(0.25)-1.5IQR_a, Q_3(0.75)+1.5IQR_a]
\]

(4)

This process decreases the \( n'' \) interval endpoints to \( m' \) interval endpoints.

**Processing Tolerance Limit**

This is done to keep on the data intervals that are within a satisfactory two-sided tolerant limit. This is done on \( a^{(0)} \) and \( b^{(0)} \) firstly and then on \( L^{(0)} = b^{(0)} - a^{(0)} \). Subsequently, only intervals that satisfy Equation (5) and (6) are kept:

\[
a^{(0)} \in \left[m_a - k\sigma_a, m_a + k\sigma_a\right]
\]

(5)

\[
b^{(0)} \in \left[m_b - k\sigma_b, m_b + k\sigma_b\right]
\]

(6)

**Reasonable-Interval Processing**

This is done to eliminate data intervals that have no overlap or too small overlap with other data intervals, so as to establish the maxim “word must mean similar things to different people”. To do this, one finds one of the values:

\[
\bar{\xi} = \frac{\left(\frac{m_a - \xi^*_a}{\sigma_a} - \xi^*_a\right) + \sigma_a}{\left(\frac{m_a - \xi^*_a}{\sigma_a} - \sigma_a\right) + 2(\sigma_a - \sigma_b)\ln\left(\frac{\sigma_b}{\sigma_a}\right)^2}}
\]

(7)

such that \( m_a \leq \bar{\xi} \leq m_b \),

where, \( m_a \) and \( m_b \) are the mean values of the left and right endpoints of the surviving \( m'' \) intervals. Only the intervals \([a^{(0)}, b^{(0)}]\) are kept such that:

\[
2m_a - \bar{\xi} < \xi^* < \bar{\xi} \leq 2m_b - \bar{\xi}
\]

(8)

At the end of the Data process section, the original \( n \) data intervals have been reduced to a set of \( m \) data intervals \( \left([a^{(0)}, b^{(0)}]\right)^m \), where \( m \leq n \).
Section 2: Fuzzy Set Process Section

The fuzzy set process section established the nature of the FOU as either a Left- or Right-shoulder or an Interior FOU, by making computations on the overlap of the intervals, removing the overlap from each of the original intervals and mapping the set(s) of smaller intervals into the two parameters that define the respective FOU. The fuzzy set part of the HMA is as follows in four steps (Hao and Mendel, 2015):

1. Classify the FOU as one of the following: (1) Left-shoulder FOU, (2) Right-shoulder FOU and (3) Interior FOU. We achieve this by (1) computing the one-sided tolerance intervals for the end-points and (2) classifying the words as follows:
   - When \( a \leq 0 \), \( W \rightarrow \text{Left-shoulder FOU} \); when \( b \geq 10 \), \( W \rightarrow \text{Right-shoulder FOU} \); otherwise, \( W \rightarrow \text{Interior FOU} \)

2. Calculate the overlap \([o_i, o_j]\) of the \( m \) intervals. For a Left-shoulder FOU, \([o_i, o_j] = \left[0, \text{min}, b^{(i)}\right]_i\) or an Interior FOU, \([o_i, o_j] = \left[\text{max}, a^{(i)}\text{min}, b^{(i)}\right]_i\) for a Right-shoulder FOU, \([o_i, o_j] = \left[\text{max}, a^{(i)}\text{max}, 10\right]_i\)

3. Remove the overlap \([o_i, o_j]\) from each of the original \( m \) intervals, \( \left[a^{(i)}, b^{(i)}\right]_m \). A Left-shoulder FOU leaves a new set of smaller intervals, \( [o_i = \text{min}, b^{(i)}]_m \); a Right-shoulder FOU leaves a new set of smaller intervals, \( [a^{(i)}\text{max}, a^{(i)}]_m \); and an Interior FOU leaves two new sets of smaller intervals, \( [a^{(i)}\text{max}, a^{(i)}]_m \) and \( [o_i = \text{min}, b^{(i)}\text{max}, 10]_m \).

4. The set(s) of smaller intervals are mapped into the two parameters that define the respective FOU. For a Left- or Right-shoulder FOU, there are exactly two parameters, \( b_i \) and \( b_r \) or \( a_i \) and \( a_r \); but, for an Interior FOU, there are four parameters \( a_i \), \( a_r \), \( b_i \) and \( b_r \). For the Interior FOU, \( [a^{(i)}\text{max}, a^{(i)}]_m \) are mapped into \( a_i \) and \( a_r \), and \( [o_i = \text{min}, b^{(i)}\text{max}, 10]_m \) are mapped into \( b_i \) and \( b_r \).

The mappings are done so that two measures of uncertainty about the \( m \) smaller length data intervals are mapped into two comparable measures of uncertainty about the FOU.

Type Reduction and Defuzzification

The combined FOU is type reduced by computing the centroid (measure of uncertainty) of the IT2 FS with the Enhanced Kernik-Mendel (EKM) approach (Algorithm 1 and 2) (Wu and Mendel, 2009). The result is an interval valued set, which is defuzzified by taking the average of the interval’s two endpoints.

Algorithm 1: The EKM algorithm for computing \( y_i \) is:

1. Sort \( x_i \) (\( i = 1, 2, \ldots, N \)) in increasing order and call the sorted \( x_i \) by the same name, but now \( x_1 \leq x_2 \leq \ldots \leq x_N \). Match the weights \( w_i \) with their respective \( x_i \) and renumber them so that their index corresponds to the renumbered \( x_i \).
2. Set \( k = \text{[N/2.4]} \) (the nearest integer to \( \text{N/2.4} \)) and compute
   \[
   a = \sum_{i=1}^{N} x_i w_i + \sum_{i=N+1}^{N} x_i w_i
   \]
   \[
   b = \sum_{i=1}^{N} x_i w_i + \sum_{i=N+1}^{N} x_i w_i
   \]

3. Find \( k' \in [1, N-1] \) such that \( x_{k'} \leq y \leq x_{k'+1} \).
4. Check if \( k' = k \). If yes, stop, set \( y_i = y \) and call \( k \). If no, continue.
5. Compute \( s = \text{sign}(k' - k) \)

Algorithm 2: The EKM algorithm for computing \( y_i \) is:

1. Sort \( x_i \) (\( i = 1, 2, \ldots, N \)) in increasing order and call the sorted \( x_i \) by the same name, but now \( x_1 \leq x_2 \leq \ldots \leq x_N \). Match the weights \( w_i \) with their respective \( x_i \) and renumber them so that their index corresponds to the renumbered \( x_i \).
2. Set \( k = \text{[N/1.7]} \) (the nearest integer to \( \text{N/1.7} \)) and compute
   \[
   a = \sum_{i=1}^{k} x_i w_i + \sum_{i=k+1}^{N} x_i w_i
   \]
   \[
   b = \sum_{i=1}^{k} x_i w_i + \sum_{i=k+1}^{N} x_i w_i
   \]

3. Find \( k' \in [1, N-1] \) such that \( x_{k'} \leq y \leq x_{k'+1} \).
4. Check if \( k' = k \). If yes, stop, set \( y_i = y \) and call \( k \). If no, continue.
5. Compute \( s = \text{sign}(k' - k) \) and
\[ a' = a - s \sum_{i \in \text{min}(k, k')} (\bar{w}_i - \bar{w}) \]
\[ b' = b - s \sum_{i \in \text{max}(k, k')} (\bar{w}_i - \bar{w}) \]
\[ y' = a'/b' \]

6) Set \( y = y' \), \( a = a' \), \( b = b' \) and \( k = k' \). Go to Step 3.

Pattern Discovery

Interval Type-2 fuzzy association rule mining was used to achieve pattern discovery in breast cancer dataset. The FP Growth algorithm was adopted to discover all frequent patterns. The algorithm is in two stages: the first stage is the generation of frequent itemsets by constructing the FP-Tree compact structure. The second stage is the pattern discovery from the frequent item set. The input into the Fuzzy FP-growth algorithm is the IT2 fuzzified dataset from the data pre-processing phase (Zhao and Bhownick, 2003) for FP Growth algorithm details. The algorithm was implemented with Python Programming language.

Data Analysis, Result and Discussion

In achieving this study objectives, the breast cancer dataset was fuzzified with interval type-2 fuzzy concept and analyzed with association rule mining. The superiority of this analysis comes from the handling of inter and intra uncertainty in breast cancer dataset by exploring interval type-2 fuzzify approach. This approach enables different medical experts with diverse opinions to be involved in the analysis. The medical experts engaged in this study defined data intervals for each of the three intuitive words (Low, Medium and High) established for each determinant factor as it was carried out in our previous study (Oladipupo et al., 2019). An instance (Twenty-four records) of the data intervals described by the medical experts for Clump Thickness’ linguistic terms is shown in Table 2.

As depicted in Table 2, the first medical expert well-defined the interval of \([0, 3] \) for the word Low; an interval of \([2, 4] \) for the word Medium and High as \([4, 10] \). Meanwhile, the second medical expert defined the interval \([0, 4.5] \) for the same word Low, for Medium an interval of \([1, 3] \) and High as \([3.5, 10] \). This culminated to 30 medical experts to define the intervals of what each word interprets to them. The disparity in data intervals shows there are different understandings of the same word to different medical experts in diagnosing breast cancer. This supports the maxim that word means different things to different people (Hao and Mendel, 2015).

| S/N | Low   | MEDIUM | HIGH   |
|-----|-------|--------|--------|
|     | Lower-range | Upper-range | Lower-range | Upper-range | Lower-range | Upper-range |
| 1   | 0     | 3.0    | 2.0    | 4.0    | 4.0    | 10.0    |
| 2   | 0     | 4.5    | 1.0    | 3.0    | 3.5    | 10.0    |
| 3   | 0     | 4.0    | 2.2    | 5.0    | 4.5    | 10.0    |
| 4   | 0     | 4.0    | 1.0    | 5.0    | 3.3    | 10.0    |
| 5   | 0     | 3.5    | 1.0    | 5.0    | 2.0    | 10.0    |
| 6   | 0     | 3.5    | 2.0    | 4.0    | 3.0    | 10.0    |
| 7   | 0     | 3.0    | 3.0    | 5.0    | 1.0    | 10.0    |
| 8   | 0     | 3.0    | 2.0    | 4.0    | 4.0    | 10.0    |
| 9   | 0     | 3.0    | 4.0    | 2.1    | 7.0    | 6.5     |
| 10  | 0     | 3.5    | 1.0    | 3.0    | 4.5    | 10.0    |
| 11  | 0     | 3.0    | 3.0    | 5.0    | 4.0    | 10.0    |
| 12  | 0     | 4.0    | 3.0    | 4.0    | 4.0    | 9.0     |
| 13  | 0     | 4.5    | 3.0    | 5.0    | 1.0    | 10.0    |
| 14  | 0     | 4.0    | 2.5    | 4.9    | 4.5    | 10.0    |
| 15  | 0     | 4.5    | 2.0    | 4.0    | 3.0    | 10.0    |
| 16  | 0     | 5.0    | 2.5    | 4.5    | 4.0    | 10.0    |
| 17  | 0     | 6.0    | 4.0    | 7.0    | 3.5    | 10.0    |
| 18  | 0     | 3.0    | 2.5    | 4.5    | 4.5    | 10.0    |
| 19  | 0     | 3.5    | 2.0    | 4.5    | 3.5    | 10.0    |
| 20  | 0     | 4.0    | 2.5    | 4.0    | 3.5    | 10.0    |
| 21  | 0     | 3.0    | 3.0    | 5.0    | 5.0    | 10.0    |
| 22  | 0     | 4.0    | 4.0    | 6.0    | 6.0    | 10.0    |
| 23  | 0     | 3.5    | 3.0    | 5.0    | 4.0    | 10.0    |
| 24  | 0     | 3.5    | 3.0    | 6.0    | 5.0    | 10.0    |
Results Obtained from Breast Cancer Dataset Fuzzification and Discussion

The interval type-2 fuzzy set: HM approach 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. Some instance of the results in this section have been presented in our previous work (Oladipupo et al., 2019). However, the whole results are presented in this study systematically. After executing the HM interval type-2 algorithm (Data and Fuzzy part) the output is shown in Table 3. For each row, the last column in Table 3 shows the number of reliable data intervals remaining that were finally utilised to construct the footprint of uncertainty for each 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 similarities among the words, computed with the Jaccard similarity, are summarized in Table 4. The Jaccard similarity provides a realistic result because it can be observed that the similarity declines monotonically as two words get further away from each other. This ascertains the adequacy of the linguistic word defined by the expert. The Upper Membership Function (UMF) and Lower Membership Function (LMF) parameters for each word of the determinant factors using the MATLAB are represented in Fig. 3a-3i.
Fig. 3: The Fuzzy Sets plot for each word used in eliciting the determinant factor; (a) “Clump thickness”; (b) “uniformity of cell size”; (c) “uniformity of cell shape”; (d) “marginal adhesion”; (e) “single epithelial cell size”; (f) “bare nuclei”; (g) “bland chromatin”; (h) “normal nucleoli”; (i) “mitoses”
Table 3: The output from the HMA process

| Processing | Data part | Fuzzy part |
|------------|-----------|------------|
| Word       | Step1 n'  | Step2 m'   | Step3 m''  | Step4 m''' | FinalStep m* |
| “Clump Thickness” |          |            |            |            |             |
| Low        | 30        | 28         | 25         | 25         | 25          |
| Medium     | 30        | 28         | 24         | 23         | 19          |
| High       | 30        | 26         | 26         | 23         | 23          |
| “Uniformity of Cell Size” |          |            |            |            |             |
| Low        | 30        | 29         | 29         | 28         | 28          |
| Medium     | 30        | 27         | 27         | 26         | 14          |
| High       | 30        | 29         | 26         | 25         | 25          |
| “Uniformity of Cell Shape” |          |            |            |            |             |
| Low        | 30        | 26         | 25         | 25         | 25          |
| Medium     | 30        | 28         | 27         | 27         | 17          |
| High       | 30        | 29         | 29         | 29         | 29          |
| “Marginal Adhesion” |          |            |            |            |             |
| Low        | 30        | 26         | 25         | 25         | 25          |
| Medium     | 30        | 28         | 27         | 27         | 17          |
| High       | 30        | 29         | 29         | 29         | 29          |
| “Single Epithelial Cell Size” |          |            |            |            |             |
| Low        | 30        | 29         | 28         | 27         | 27          |
| Medium     | 30        | 27         | 26         | 25         | 20          |
| High       | 30        | 30         | 30         | 29         | 29          |
| “Bare Nuclei” |          |            |            |            |             |
| Low        | 30        | 28         | 28         | 28         | 28          |
| Medium     | 30        | 24         | 24         | 24         | 17          |
| High       | 30        | 28         | 28         | 28         | 28          |
| “Bland Chromatin” |          |            |            |            |             |
| Low        | 30        | 27         | 27         | 27         | 27          |
| Medium     | 30        | 28         | 28         | 27         | 20          |
| High       | 30        | 28         | 29         | 29         | 28          |
| “Normal Nucleoli” |          |            |            |            |             |
| Low        | 30        | 27         | 27         | 27         | 27          |
| Medium     | 30        | 28         | 28         | 25         | 22          |
| High       | 30        | 29         | 29         | 28         | 28          |
| “Mitoses” |          |            |            |            |             |
| Low        | 30        | 28         | 28         | 28         | 28          |
| Medium     | 30        | 21         | 21         | 21         | 7           |
| High       | 30        | 29         | 29         | 29         | 29          |

Table 4: Similarities among the three words using the Jaccard similarity measure

| Word                  | Low     | Medium | High  |
|-----------------------|---------|--------|-------|
| “Clump Thickness”     | 1.00    | 0.23   | 0.08  |
| Low                   | 0.23    | 1.00   | 0.11  |
| Medium                | 0.08    | 0.11   | 1.0   |
| “Uniformity of Cell Size” | 1.00    | 0.12   | 0.07  |
| Low                   | 0.12    | 1.00   | 0.26  |
| Medium                | 0.07    | 0.26   | 1.00  |
| “Uniformity of Cell Shape” | 1.00    | 0.15   | 0.09  |
| Low                   | 0.15    | 1.00   | 0.25  |
| Medium                | 0.09    | 0.25   | 1.00  |
| “Marginal Adhesion”   | 1.00    | 0.07   | 0.08  |
| Low                   | 0.07    | 1.00   | 0.23  |
| Medium                | 0.08    | 0.23   | 1.00  |
| “Single Epithelial Cell Size” | 1.00    | 0.16   | 0.09  |
| Low                   | 0.16    | 1.00   | 0.20  |
| Medium                | 0.09    | 0.20   | 1.00  |
The type-2 fuzzy set derived for each word after the processes and the values obtained after type-reduction using the Enhanced Karnik-Mendel (EKM) Approach are also shown in Table 5. Finally, the Crips breast cancer dataset collected are fuzzified based on the IT2 fuzzy set membership expression generated from the HM approach. The fuzzified dataset now captures the intra and inter uncertainties of medical experts on the determinant factors. The snapshot for the generated fuzzified dataset is represented in Fig. 4.
Results Obtained from Pattern Discovery and Discussion

After generating IT2 fuzzy dataset, the FP Growth algorithm accepts the dataset as input to construct FP-Tree compact structure and saves it in the system memory. The system interface snapshot for frequent pattern tree representation is shown in Fig. 4. The next step is rule generation from the pattern tree. The association rule mining FP Growth algorithm was applied to uncover potentially useful patterns from the breast cancer IT2 fuzzy dataset. Firstly, the association rule mining algorithm generated 350 pattern rules when the minimum support was set to 20% and confidence value of 0.2. The system snapshot for rule generation is shown in Fig. 5. To determine the interesting rules, the rules that determine malignant breast cancer with minimum confidence greater than or equal to 0.8 were considered. Out of the 350 rules 35 determine malignant breast cancer while just 18 of them meet the minimum confidence of 0.8 (80%). The final interesting rules highlighted are 18 in number as shown in Table 6.
Evaluation of the Proposed Approach

An Interval Type-2 Fuzzy Association Rule Mining (IT2FARM) was explored for patterns’ discovery in breast cancer dataset in this study. The breast cancer dataset from Wisconsin in UCI data repository was used for the mining process. Specifically, a test dataset consisting of 10 attributes and 699 instances used to distinguish malignant (cancerous) from benign (non-cancerous) samples was used.

A comparison between the proposed IT2FARM and quantitative Association Rule Mining (ARM) (Gyenes, 2001) implementation using Rapid miner on the same breast dataset was carried out. The comparison was based on the number of rule generated, size on disk and maximum condition in rule premise as shown in Table 7. The confidence value of 0.2 was used as the threshold. The snapshot for quantitative association rule mining using Rapid miner analytical tool is shown in Fig. 6.

Fig. 6: Snapshot for rules generation using the FP growth algorithm
Table 7: Tabular report of evaluation of IT2FARM and ARM

| Measures                          | IT2FARM          | ARM on Rapid miner |
|-----------------------------------|------------------|--------------------|
| Linguistic terms                  | 28               | 28                 |
| Number of rules (Min supp 0.2)    | 350              | 887                |
| Size on disk                      | 84 Kb            | 372 Kb             |
| Max conditions in rule premise    | 8                | 9                  |

Fig. 6: Snapshot of the rules obtained using quantitative ARM in Rapid miner

One major observation from this evaluation result is the number of rules generated. In building a comprehensive expert system, there is need for a compact knowledge base with minimized number of rule to avoid knowledge based unwieldiness (Meesad, 2001; Oladiypuo et al., 2010). The proposed approach was able to reduce the amount of rule generated by 39.5%. This is an advantage of fuzzy parameter over the quantitative mining approach. This is supported by the result in (Oladiypuo et al., 2010) where fuzzy was used to enhance the knowledge discovery in coronary heart disease. Another advantage is the disk size usage, which is well minimized by IT2FARM. The observation shows that IT2FARM outperformed quantitative ARM based on the above criteria in analyzing breast cancer dataset, which makes it recommendable for evolving a comprehensive knowledge base for breast cancer decision support system.

Conclusion

In this study, an Interval Type-2 fuzzy association rule mining was proposed to extract interesting patterns from breast cancer that correspond more intuitively to expert’s perception. The analysis in this study was divided into two stages; the data fuzzification process and pattern discovery. Hao and Mendel approach was adopted for interval Type-2 data fuzzification while FP-growth algorithm was explored for pattern discovery in breast cancer. The breast cancer dataset was acquired (699 instances and 11 attributed) from Wisconsin in UCI machine learning repository. Thirty (30) medical experts from specialized hospitals were consulted using a questionnaire poling method to define the intuitive words for each attribute and the data intervals. This study is able to generate rule with fewer symptoms for diagnosing breast cancer with confidence greater than 80%. It also identified High Uniformity of Cell Shape (10/18) and High Bare Nuclei (9/18) as two strong determinant factors for determining breast cancer. Benchmarking, the IT2FARM with traditional Quantitative ARM, the former performed better in terms of rules with 39.5 and 22.6% reduction in generated rules and memory respectively. This supports the effect of type-2 fuzzy in association rule mining as reported in (Chen et al., 2015) due to its ability to resolve intra and inter uncertainty within the dataset. The discovered rules in this study are viable in building a comprehensive and compact expert driven knowledge base for breast cancer decision support or expert system Thus, for future improvements on this research, a hybridization of Interval type 2 fuzzy logic with other data mining techniques like rough sets, artificial neural and decision tree is recommended for prediction...
purposes. Also, building an expert system with discovered patterns with IT2 fuzzy Association rule mining could be considered in future work.

**Acknowledgment**

We acknowledge Covenant University for providing a conducive environment where this study was successfully carried out by the second author as his MSc. Research work, under the supervision of the first author.

**Author’s Contributions**

Olufunke Oladipupo: Originated and prepared the conceptual framework for the work, contributed to the analysis and interpretation of the result. Final approval of the version to be submitted and any revised version. She supervised the work.

Oluwole Olajide: Acquisition of data, implementation and data analysis. He wrote the first draft of the paper.

Stephen Adubi: Reviewing the manuscript critically for significant intellectual content and formatting of the submitted draft.

Jelili Oyelade: Contributed to the material structure and Reviewing the manuscript critically for significant intellectual content

Zacchaeus Omobhadegun: Reviewing the manuscript critically for significant intellectual content.

**Ethics**

In this study the dataset was acquired from Wisconsin in UCI machine learning repository where ethics procedures have been properly certified. We have also run the manuscript through a plagiarism/similarity check (Turnitin) to ensure the originality and uniqueness of the research work. We have ensured every item in the References List has been properly cited and vice versa.

**References**

Aalaei, S., Shahraki, H., Rowghanimanesh, A., & Eslami, S. (2016). Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. Iranian Journal of Basic Medical Sciences, 19(5), 476.

Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. Expert Systems with Applications, 46, 139-144. https://doi.org/10.1016/j.eswa.2015.10.015

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. https://doi.org/10.1007/s00521-010-0414-4

Ahmed, M. T., Imtiaz, M. N., & Karmakar, A. (2020). Analysis of Wisconsin Breast Cancer original dataset using data mining and machine learning algorithms for breast cancer prediction. Journal of Science, Technology and Environment Informatics, 9(02), 665-672. https://doi.org/10.18801/jstei.090220.67

Ak, M. F. (2020, June). A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. In Healthcare (Vol. 8, No. 2, p. 111). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/healthcare8020111

Alharbi, A., & Tchier, F. (2017). Using a genetic-fuzzy algorithm as a computer aided diagnosis tool on Saudi Arabian breast cancer database. Mathematical Biosciences, 286, 39-48. https://doi.org/10.1016/j.mbs.2017.02.002

Alison, M. R., Poulsom, R., & Forbes, S. J. (2001). Update on hepatic stem cells. Liver: Review Article, 21(6), 367-373. https://doi.org/10.1034/j.1600-0676.2001.210601.x

Arafah, A. A., & Mukhlash, I. (2015). The Application of Fuzzy Association Rule on Co-Movement Analyze of Indonesian Stock Price. Procedia Computer Science, 59, 235-243. https://doi.org/10.1016/j.procs.2015.07.541

Bilgin, A., Hagras, H., Malibari, A., Alhaddad, M. J., & Alghazzawi, D. (2012, June). Towards a general type-2 fuzzy logic approach for computing with words using linear adjectives. In 2012 IEEE International Conference on Fuzzy Systems (pp. 1-8). IEEE. https://doi.org/10.1109/FUZZ-IEEE.2012.6251220

Bobillo, F., & Straccia, U. (2017). Generalizing type-2 fuzzy ontologies and type-2 fuzzy description logics. International Journal of Approximate Reasoning, 87, 40-66. https://doi.org/10.1016/j.ijar.2017.04.012

Castillo, O., & Amador-Angulo, L. (2018). A generalized type-2 fuzzy logic approach for dynamic parameter adaptation in bee colony optimization applied to fuzzy controller design. Information Sciences, 460-461, 476-496. https://doi.org/10.1016/j.ins.2017.10.032

Castillo, O., Cervantes, L., Soria, J., Sanchez, M., & Castro, J. R. (2016). A generalized type-2 fuzzy granular approach with applications to aerospace. Information Sciences, 354, 165-177. https://doi.org/10.1016/j.ins.2016.03.001

Celik, E., & Akyuz, E. (2018). An interval type-2 fuzzy AHP and TOPSIS methods for decision-making problems in maritime transportation engineering: The case of ship loader. Ocean Engineering, 155, 371-381. https://doi.org/10.1016/j.oceaneng.2018.01.039
Chaurasia, V., Pal, S., & Tiwari, B. B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, 12(2), 119-126. https://doi.org/10.1177/1748301818756225

Chen, C. H., Hong, T. P., & Li, Y. (2015, March). Fuzzy association rule mining with type-2 membership functions. In Asian Conference on Intelligent Information and Database Systems (pp. 128-134). Springer, Cham. https://doi.org/10.1007/978-3-319-15705-4_13

Chiu, H. P., Tang, Y. T., & Hsieh, K. L. (2012). Applying cluster-based fuzzy association rules mining framework into EC environment. Applied Soft Computing, 12(8), 2114-2122. https://doi.org/10.1016/j.asoc.2011.08.010

Chunklk, S., Auephanwiriyakul, S., & Theera-Umporn, N. (2010, July). Microcalcification detection in mammograms using interval type-2 fuzzy logic system with automatic membership function generation. In International Conference on Fuzzy Systems (pp. 1-7). IEEE. https://doi.org/10.1109/FUZZY.2010.5858496

Delgado, M., Sánchez, D., Martín-Bautista, M. J., & Vila, M. A. (2001). Mining association rules with improved semantics in medical databases. Artificial Intelligence in Medicine, 21(1), 241-245. https://doi.org/10.1016/S0933-3657(00)00092-0

Ed-daoudy, A., & Maalmi, K. (2020). Breast cancer classification with reduced feature set using association rules and support vector machine. NetMAHIB, 9(1), 34. https://doi.org/10.1007/s13721-020-00237-8

Ehtemam, H., Montazeri, M., Khajouei, R., Hosseini, R., Nemati, A., & Maazed, V. (2017). Prognosis and Early Diagnosis of Ductal and Lobular Type in Breast Cancer Patient. Iranian Journal of Public Health, 46(11), 1563-1571.

Greenfield, S., & Chilclana, F. (2018). The Collapsing Defuzzifier for discretised generalised type-2 fuzzy sets. International Journal of Approximate Reasoning, 102, 21-40. https://doi.org/10.1016/j.ijar.2018.07.008

Gupta, N., Mangal, N., Tiwari, K., & Mitra, P. (2006). Mining quantitative association rules in protein sequences. In Data Mining (pp. 273-281). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11677437_21

Gyeniesvi, A. (2001). A fuzzy approach for mining quantitative association rules. Acta Cybern., 15(2), 305-320.

Hao, M., & Mendel, J. M. (2015). Encoding words into normal interval type-2 fuzzy sets: HM approach. IEEE Transactions on Fuzzy Systems, 24(4), 865-879. https://doi.org/10.1109/TFUZZ.2015.2486814

He, Y., Zheng, X., Sit, C., Loo, W. T., Wang, Z., Xie, T., & Chen, J. (2012, September). Using association rules mining to explore pattern of Chinese medicinal formulae (prescription) in treating and preventing breast cancer recurrence and metastasis. In Journal of Translational Medicine (Vol. 10, No. S1, p. S12). BioMed Central. https://doi.org/10.1186/1479-5876-10-S1-S12

Ho, G. T., Ip, W. H., Wu, C. H., & Tse, Y. K. (2012). Using a fuzzy association rule mining approach to identify the financial data association. Expert Systems with Applications, 39(10), 9054-9063. https://doi.org/10.1016/j.eswa.2012.02.047

Hosseini, S. M. S., Maleki, A., & 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. https://doi.org/10.1016/j.eswa.2009.12.070

Jain, R., & Abraham, A. (2004). A comparative study of fuzzy classification methods on breast cancer data. Australasian Physics & Engineering Sciences in Medicine, 27(4), 213-218. https://doi.org/10.1007/BF03178651

Karabatak, M., & 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. https://doi.org/10.1016/j.eswa.2008.02.064

Kim, J., Han, M., Lee, Y., & 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. https://doi.org/10.1016/j.eswa.2016.03.043

Kuok, C. M., Fu, A., & Wong, M. H. (1998). Mining fuzzy association rules in databases. ACM Sigmod Record, 27(1), 41-46. https://doi.org/10.1145/273244.273257

Liu, F., & Mendel, J. M. (2008). Encoding words into interval type-2 fuzzy sets using an interval approach. IEEE transactions on Fuzzy Systems, 16, 1503-1521. https://doi.org/10.1109/TFUZZ.2008.2005002

Medjahed, S. A., Saadi, T. A., & Benyettou, A. (2013). Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. International Journal of Computer Applications, 62(1) 1-5. https://doi.org/10.5120/10041-4635
Meesad, P. (2001). Quantitative measures of a fuzzy expert system. Expert Systems, 1-6.

Mendel, J. M. (2001). Uncertain rule-based fuzzy logic systems: introduction and new directions. Upper Saddle River: Prentice Hall PTR.

Mendel, J. M. (2007). Computing with words and its relationships with fuzzistics. Information Sciences, 177(4), 988-1006. https://doi.org/10.1016/j.ins.2006.06.008

Mendel, J., & Wu, D. (2010). Perceptual computing: Aiding people in making subjective judgments (Vol. 13). John Wiley & Sons. https://doi.org/10.1002/9780470599655

Mohammed, S. A., Darrab, S., Noaman, S. A., & Saake, G. (2020, July). Analysis of breast cancer detection using different machine learning techniques. In International Conference on Data Mining and Big Data (pp. 108-117), Springer, Singapore. https://doi.org/10.1007/978-981-15-7205-0_10

Moharrer, M., Tahayori, H., & Sadeghian, A. (2013). Modeling complex concepts with type-2 fuzzy sets: The case of user satisfaction of online services. In Advances in Type-2 Fuzzy Sets and Systems (pp. 133-146). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6666-6_9

Mohradian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. Computational and Structural Biotechnology Journal, 15, 75-85. https://doi.org/10.1007/journal.csbj.2016.11.004

Muhic, I. (2013). Fuzzy analysis of breast cancer disease using fuzzy c-means and pattern recognition. Southeast Europe Journal of Soft Computing, 2(1), 1-6. https://doi.org/10.21533/ijscjournals.v2i1.45

Ngyuen, T., Khosravi, A., Creighton, D., & Nahavandi, S. (2015a). EEG signal classification for BCI applications by wavelets and interval type-2 fuzzy logic systems. Expert Systems with Applications, 42(9), 4370-4380. https://doi.org/10.1016/j.eswa.2015.01.036

Ngyuen, T., Khosravi, A., Creighton, D., & Nahavandi, S. (2015b). Medical data classification using interval type-2 fuzzy logic system and wavelets. Applied Soft Computing, 30, 812-822. https://doi.org/10.1016/j.asoc.2015.02.016

Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. Telematics and Informatics, 34(4), 133-144. https://doi.org/10.1016/j.tele.2017.01.007

Oladipupo, O. O., Oyelade, O. J., & Aborisade, D. O. (2012). Application of Fuzzy Association Rule Mining for Analysing Students Academic Performance. International Journal of Computer Science Issues, 9(3), 216-223. https://doi.org/10.1504/IJAIISC.2012.048179

Oladipupo, O., Olajide, O., Olawoye, S., Oyelade, J., & Stephen, A. (2019). An Interval Type-2 Fuzzy Set Approach to Breast Cancer Dataset Analysis. Oladipupo, O.O., Uwadia, O.C & Ayo, K.C. (2010). On sharp boundary problem in rule-based expert systems in medical domain. International Journal of Healthcare Information Systems and Informatics. 5, 14-26. https://doi.org/10.4018/jhiis.201007012

Pagola, M., Barrenechea, E., Fernández, J., Jurio, A., Galar, M., Sanz, J. A., & Bustince, H. (2013). Construction of interval type-2 fuzzy sets from fuzzy sets: methods and applications. In Advances in Type-2 Fuzzy Sets and Systems (pp. 147-163). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6666-6_10

Pimenta, A. H., & Camargo, H. A. (2010, July). Interval type-2 fuzzy classifier design using genetic algorithms. In International Conference on Fuzzy Systems (pp. 1-7). IEEE. https://doi.org/10.1109/FUZZY.2010.5584520

Rubio, E., Castillo, O., Valdez, F., Melin, P., Gonzalez, C. I., & Martinez, G. (2017). An extension of the fuzzy possibilistic clustering algorithm using type-2 fuzzy logic techniques. Advances in Fuzzy Systems, 2017. https://doi.org/10.1155/2017/7094046

Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrenc prediction. IEEE Access, 6, 29637-29647. https://doi.org/10.1109/ACCESS.2018.2843443

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

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. https://doi.org/10.1016/j.ins.2010.06.018

Serban, G., & Moldovan, G. S. (2006, September). A new k-means based clustering algorithm in aspect mining. In 2006 Eighth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (pp. 69-74). IEEE. https://doi.org/10.1109/SYNASC.2006.5

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. CA: A Cancer Journal for Clinicians, 68(1), 7–30. https://doi.org/10.3322/caac.21442

Tahayori, H., & Sadeghian, A. (2012). Median interval approach to model words with interval type-2 fuzzy sets. International Journal of Advanced Intelligence Paradigms, 4(3-4), 313-336. https://doi.org/10.1504/IJAIP.2012.052074
Torres-Blanc, C., Hernández-Varela, P., & Cubillo, S. (2018). Self-contradiction for type-2 fuzzy sets whose membership degrees are normal and convex functions. Fuzzy Sets and Systems, 352, 73-91. https://doi.org/10.1016/j.fss.2017.12.015
Umesh, D. R., & Ramachandra, B. (2015, December). Association rule mining based predicting breast cancer recurrence on SEER breast cancer data. In 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) (pp. 376-380). IEEE. https://doi.org/10.1109/ERECT.2015.7499044
Vijaylakshmi, S., & Priyadarshini, J. (2016). An Analysis of Particle Swarm Optimization Technique for Breast Cancer Dataset. International Journal of Computer Technology and Applications, 9, 279-306.
Vrigazova, B. P. (2020). Detection of Malignant and Benign Breast Cancer Using the ANOVA-BOOTSTRAP-SVM. Journal of Data and Information Science, 5(2), 62-75. https://doi.org/10.2478/jdis-2020-0012
Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2013). Probability and statistics for engineers and scientists. London: Pearson.
Wu, D., & Mendel, J. M. (2008). A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets. Information Sciences, 178(2), 381-402. https://doi.org/10.1016/j.ins.2007.04.014
Wu, D., & Mendel, J. M. (2009). Enhanced kernik-mendel algorithms. IEEE Transactions on Fuzzy Systems, 17(4), 923-934. https://doi.org/10.1109/TFUZZ.2008.924329
Wu, D., Mendel, J. M., & 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. https://doi.org/10.1109/TFUZZ.2011.2177272
Yeh, W. L., Lu, D. Y., Lee, M. J., & Fu, W. M. (2009). Leptin induces migration and invasion of glioma cells through MMP-13 production. Glia, 57(4), 454-464. https://doi.org/10.1002/glia.20773
Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-I. Information Sciences, 8(3), 199-249. https://doi.org/10.1016/0020-0255(75)90036-5
Zhao, Q., & Bhowmick, S. S. (2003). Association rule mining: A survey. Nanyang Technological University, Singapore, 135.