Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DIABETES DIAGNOSIS AND LEVEL OF CARE FUZZY RULE-BASED MODEL UTILIZING SUPERVISED MACHINE LEARNING FOR CLASSIFICATION AND PREDICTION

TEH NORANIS MOHD ARIS[1], AZURALIZA ABU BAKAR[2], NORMADIAH MAHIDDIN[3], MASLINA ZOLKEPLI[4]

[1, 4]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[2]Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi Selangor, Malaysia

[3]Computing Science Studies, College of Computing, Informatics and Media Studies, Universiti Teknologi Mara (UiTM) Pahang Branch, Raub Campus, 27600 Raub, Pahang, Malaysia

[1]nuranis@upm.edu.my (corresponding author), [2]azuraliza@ukm.edu.my, [3]normadiah@uitm.edu.my, [4]masz@upm.edu.my

ABSTRACT

A reliable medical decision-making is essential to diagnose a disease. This assists medical practitioners to detect a disease at early stage especially diabetes that causes further health complications. The diversity and availability of healthcare datasets supports medical practitioners to use computer applications in the diagnosis process. There are many medical datasets available for research usage but these datasets lacks information that allows decisions to be made accurately, which have a major impact to diagnose a disease. Fuzzy logic has contributed to handle vagueness and uncertainty issues and one of the appropriate models for the development of medical diagnostics. Most computer applications use machine learning and data mining techniques to aid classification and prediction of a disease. Therefore, a fuzzy model based on machine learning and data mining is a vital solution. In this study, ten supervised machine learning algorithms namely the J48, Logistic, NaiveBayes Updateable, RandomTree, BayesNet, AdaBoostM1, Random Forest, Multilayer Perceptron, Bagging and Stacking are applied for a simulated diabetes fuzzy dataset, verified by medical experts. The fuzzy datasets provide adequate information on the type of diabetes diagnosis and level of care related to the type of diabetes diagnosis. All algorithms were compared based on the accuracy, precision, recall, F1-Score, and confusion matrix. Experiment results for diabetes diagnosis dataset indicate 100% accuracy for the eight algorithms except AdaBoostM1 which produced 79.82% accuracy and Stacking 67.89% accuracy. In addition, level of care dataset reveals the highest accuracy of 97.15% for MLP and Bagging algorithms and the lowest accuracy of 91.66% for stacking algorithm. Overall, the proposed fuzzy rule-based diabetes diagnosis and level of care fuzzy model works well with most of the machine learning algorithms tested. Therefore, the proposed fuzzy model is a useful aid in the decision-making process, specifically in the healthcare sector.
Keywords: *Decision-making, Fuzzy, Supervised Machine Learning, Classification, Prediction*

1. INTRODUCTION

A chronic condition is a long-lasting disease that can have a significant impact on a person's quality of life, cost and even life expectancy. Diabetes is a chronic disease that occurs when the pancreas does not produce insufficient insulin with the increase of blood sugar. When the insulin is insufficient, the body cannot use the insulin effectively and the concentration of glucose in the blood increases. The high concentration of glucose in the blood is known as diabetes and can cause further complications such as heart disease, kidney failure, nerve damage and other problems related to feet, oral health, vision, hearing, and even mental health. Globally, more and more people are suffering from diabetes and is a major challenge of the twenty-first century [1]. However, early

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

detection of diabetes can save life and a trustworthy decision-making model is crucial in diabetes diagnosis [2]-[4]. Research shows that many people with diabetes are unaware that they have the disease [5]-[7].

Diabetes is identified by a set of signs and symptoms and medical practitioners diagnose diabetes based on these signs and symptoms. These signs and symptoms are captured and stored in databases which can be obtained easily by medical practitioners. This encourages the usage of computer applications which assist medical practitioners to diagnose diabetes. Furthermore, the availability of publicly accessible databases facilitates research in various fields including machine learning [8]-[12].

The nature of medical data which is exposed to vague and uncertain issues requires a way of human thinking and understanding to represent the data. Fuzzy logic is a technique that can cater these issues and improve the decision-making process. Fuzzy logic has been applied in medical diagnosis generally and diabetes diagnosis specifically and research are ongoing to improve the models [13]-[17].

Machine Learning (ML) and data mining technology has a significant potential in supporting medical decision-making and automating numerous tedious tasks. These technologies provide classification, clustering, association, and regression algorithms that have been widely used to predict various diseases, which is very important to make reliable predictions of a disease to perform appropriate treatments. Moreover, machine learning is not bounded to any comprehensive framework that provides more room for researchers to expand and improve previous research works [18]. Fuzzy classification and prediction which can explain how results were derived in a way that is interpretable and compatible to human perception is a challenging research area. This research area has contributed to various studies such as medical [19]-[21] and phrase similarity [22]. In addition, research has proved that fuzzy and data mining techniques are efficient techniques to diagnose diabetes [23].

This research work is motivated due to the significance of diabetes diagnosis and level of care classification and prediction in decision-making and the fuzzy logic approach that provides human perception understanding for an effective decision-making. Medical data especially diabetes data are subject to vagueness and the numeric values are uncertain, which leads to the lack of interpretable facts. The objective of this paper is to embed fuzzy method into an interrelated decision-making model proposed by Normadiah et al. [2] to handle the vagueness and uncertain issues. The proposed fuzzy decision-making model on the diagnosis of diabetes is crucial because diabetes consists of several categories and these categories are related to the level of care which needs to be given to the patients. Additionally, supervised machine learning techniques are applied to the proposed diabetes diagnosis and level of care fuzzy model, then the accuracy, precision, recall, F1-Score, and confusion matrix of the model is evaluated.

Most of the decision-making models discussed lack sufficient information on the relation between the types of diabetes diagnosis and level of care based on the type of diabetes diagnosis, which is the nature of decision-making in the real world. Furthermore, vagueness and uncertainty are presents in the diabetes datasets. Therefore, our proposed research work closes the research gap by the creation of linguistic labels and fuzzy rules to handle the vagueness and uncertainty issues. In addition, the relation is represented in a human understanding way that leads to an accurate decision-making.

This paper is organized into five sections. Section II reviews previous research works which is related to our research work. Section III describes the methodology applied in this research work. Section IV presents the produced results. Finally, section V is the conclusion and future works.

2. RELATED RESEARCH

This section explains about the diabetes datasets used in existing research works and provide details on previous fuzzy diabetes models. In addition, this section provides explanation about fuzzy database, fuzzy rule-base and supervise machine learning techniques which are related to our research work.

2.1 Diabetes Datasets

Patients' datasets are vital and valuable source in the research of diabetes management. Most diabetes datasets consist of predictor attributes which represent the signs and symptoms of diabetes. In addition, the dataset consists of a target attribute that indicate whether a patient has diabetes or not. Diabetes consists of four types namely type 1, type 2, gestational and autosomal inherited type of diabetes mellitus [24]-[25]. Diabetes type 2 or diabetes mellitus type 2 (T2DM) is the most common type of diabetes and is the focus of our research work. Diabetes type 2 accounts for nearly 90% of the approximately 537 million cases of diabetes worldwide [6]. The Pima Indians Diabetes Dataset (PIDD) which is developed by the National Institute of Diabetes and Digestive and Kidney Diseases [26] has been widely used in most

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

research to diagnose diabetes type 2 [8]-[12], [35]. PIDD contains eight predictor attributes, one target attribute and 768 records. The predictor attributes are pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, diabetes pedigree function and age. The target attribute in PIDD is the outcome to determine whether the patient has diabetes, or not. Besides PIDD, other publicly diabetes datasets available at Kaggle [26] are Diabetes Dataset, Diabetic Retinopathy and Diabetes Health Indicators datasets. University of California Irvine (UCI) machine learning repository also provides Diabetes dataset [27], which can be accessed publicly. In addition, researches utilized publicly available medical datasets namely Appendicitis, Australian, Banana, Bands, Diabetes, Haberman, Ionosphere, Liver, Ring, Letter Recognition to produce an interpretable fuzzy classifier framework to extract linguistically rules from the datasets [22].

This research work is based on a simulated diabetes datasets that were validated by medical experts [2]. Detailed description on the datasets will be explained in sections 3.1. These datasets are initially converted to a fuzzy inference model [38]. Among the predictor attributes for these datasets are Acanthosis Nigricans (Acan. Nig.), A1c, FPG, RPG, OGTT, HDL, TG and History of Cardiovascular Disease (CVD). Table 1 shows example of the diabetes diagnosis and level of care datasets from medical experts.

Table 1: Example of Diabetes Diagnosis and Level of Care Datasets

Physical Exam	Lab Report									
Acan. Nig.	A1c (mmol/ mol)	FPG	RPG	OGTT	HDL	TG	History of CVD	Diagnosis		Level of Care
Y	36	5.2	8.7	Nil	1.4	1.2	T	Healthy	Prediabetes	Primary Care
Y	40	6.1	10	6.3 (FPG)	1.3	2.8	T	IFG	Diabetes	Primary Care
Y	48	7.3	12.3	Nil	0.9	2.8	T	T2DM	Diabetes	Primary Care
T	60	8.8	16	Nil	0.9	3	Y	T2DM	Diabetes	Secondary Care
T	55	7.5	13	Nil	1	2.8	Y	T2DM	Diabetes	Secondary Care
Y	43	6.4	9.8	9.8 (2-hr PPG)	1.2	2	T	IGT	Prediabetes	Primary Care

Several datasets were developed to facilitate the machine learning research. Saudi Arabian dataset [18] was constructed to classify and predict three types of diabetes: pre-diabetes, type 1 diabetes, and type 2 diabetes. Other datasets are ShanghaiT1DM and ShanghaiT2DM which contribute to the development of data-driven algorithms/models and diabetes monitoring/managing technologies [28]. Secondary dataset from medical database record review were used to classify and predict type-2 diabetes in public hospitals in Afar regional state, Ethiopia [7]. Insulin-dependent diabetes patients' dataset were developed to create an automated closed-loop advice system for intense insulin therapy in clinical practice [34]. Furthermore, researchers have developed a heart disease dataset to propose an enhanced genetic algorithm based on fuzzy weight updating Support Vector Machine (SVM) algorithm [36].

2.2 Supervised Machine Learning Classification Techniques

Supervised machine learning is an algorithm used to learn the mapping function from the input (1) to the output (O), where O = *f*(1), with the purpose to determine the mapping function accurately for the output (O) to be predicted when a new data (1) occurs [29]. Supervised learning consists of two categories: classification and regression. This research work focuses on classification.

Classification is a form of data analysis which can efficiently be used to divide data models for predictions to enable the identification of trends in datasets [30]. Classification is an essential technique with a wide area of applications. Classification is also regarded as a machine learning algorithm to identify and predict categories among data points; categories are then assigned with corresponding groupings to enable greater prediction accuracy. Following sections are the supervised machine learning classification techniques used in this research work [2], [7], [18]-[19], [32], [37].

2.2.1 Decision Trees

The general notion of a decision tree is to construct a tree that represents the whole dataset. Decision trees classify a given population into branch-like sections which build an inverted tree consisting of a root node, internal nodes, and leaf nodes. Decision tree algorithm exploits a type of tree branching methodology to explore available outcome of a decision influenced by certain conditions. The structure of decision trees comprise of an internal node with a decision rule represented by a branch and the outcome is represented by each individual leaf node. The topmost node is the root node that partitions the tree based on a feature value. It adopts recursive partitioning which partitions into an easy to interpret and comprehend structure diagram. J48, RandomTree and Random Forest decision trees are utilized in this research work. J48 algorithm is based on the extended Iterative

Dichotomiser 3 (ID3) that classifies a new instance by creating a decision tree from the attribute values of a given training dataset. RandomTree are ensemble tree predictors called as forest, which takes the input feature vector, classifies every tree in the forest and output the class label which obtained majority of votes. Random Forest creates a cluster of decision trees using a random subset of the sample data for building and merging numerous decision trees to achieve better prediction accuracy.

2.2.2 Activation Functions

Activation functions are the process of mapping the summed weights into a neuron output. The activation functions handle the level of neuron activation combined with the signal strength of the output. The original input from the dataset is passed to the visible layer, which uses a neuron to assign the input value and pass the value to the next layer. This is followed by the hidden layers that have no direct interaction with the input layer. A hidden layer uses a single neuron to output the corresponding value, plus complex networks consisting of massive amounts of hidden layers. The output layer is the final layer, which provides the output value that is related to the specified problem. Logistic and Multilayer Perceptron (MLP) is the activation function chosen in this research work which can execute binary class and multi class.

2.2.3 Bayes Theorem

Bayes theorem is based on the concept of conditional probability. Probability is set as the hypothesis and the alternative event is set as the evidence. Bayes theorem applies previous knowledge and implements possibility to measure the probability of a hypothesis. Bayes theorem consists of Naïve Bayesian classifiers and multi class Naïve Bayes classifiers. Naïve Bayesian classifiers are type of statistical classifiers which can predict class membership probabilities. Naïve Bayes assumes that the predictors are fully independent and equal, therefore applying no direct or indirect influence on any existing predictors. As a result, the probability is individual classes consisting of different values and conditionally independent. Furthermore, Naïve Bayes uses predictions to identify the probabilities of specified data points which belong to a certain class. The class with the highest probability is the accepted class. Multi class Naïve Bayes classifiers are adapted from Naïve Bayes that utilize a multi class distribution for each individual class, with possible outcomes of two or more classes. The Bayes theorem classification implemented in this research are BayesNet and the updateable version of NaïveBayes.

2.2.4 AdaBoostM1

AdaBoostM1, used in this research is an ensemble iterative approach that learns from the previous misclassification of vectors with the initiation of increasing misclassification weight. AdaBoost is also known as meta-learning. Adaboost performs by originally initiating the data points weights, then training the model using a decision tree moreover, to calculate the weighted error rate, which is the number of incorrect predictions designated from the weight of the vector. In other words, AdaBoost learn from the mistakes of weak classifiers and transform them into strong ones.

2.2.5 Bagging

Bagging or bootstrap aggregation is an ensemble method consisting of parallel techniques, where a set of independent models are trained with random subsets supplied by the dataset. The subsets are gathered to enable training of the base learners using the bootstrap sampling technique. For an aggregation of the base learner values, bagging utilizes the voting method for classification.

2.2.6 Stacking

Stacking is a parallel ensemble method that uses a combination of multiple regression and multiple classification models to define a prediction output regarding to various predictions of weaker models. The stacking model contains numerous machine learning algorithms to train homogeneous weak learners using meta-models, based on the utilization of the complete dataset.

2.3 Models for Decision-Making

Models based on machine learning have been designed to support the decision-making process. A machine learning approach combined with IoT technology has been proposed for classification, early-stage identification, and prediction of diabetes which used the PIDD as a benchmark for experimental evaluation [31]. Authors compared machine learning, data mining and Neural Network (NN) techniques and tested using the PIDD [32]. A machine learning model named as twice-growth deep neural network (2GDNN) for diabetes prediction and diagnosis using the PIDD and the laboratory of the Medical City Hospital (LMCH) diabetes dataset has been proposed [33]. Machine learning is utilized to distinguish and predict three types of diabetes: pre-diabetes, Type 1 Diabetes and Type 2 Diabetes based on a Saudi Arabian hospital dataset [18]. ShanghaiT1DM and ShanghaiT2DM datasets have been developed to promote and facilitate the research in diabetes management especially for data-driven machine learning methods [28]. Machine learning algorithms are experimented using secondary dataset from the medical dataset record review in Afar, Northeastern Ethiopia to classify and predict Type 2 diabetes [7]. An intelligent decision support healthcare model based on multi-agent approach was designed and simulated diabetes treatments datasets are tested using machine learning algorithms [2].

Moreover, fuzzy is also a technique to create models for decision-making. In the medical area, fuzzy inference is adopted because it is widely accepted for capturing expert knowledge, suitable to be applied in medical diagnosis, is more intuitive and human-like manner. A fuzzy rule-based classification system which provides an interpretable knowledge base to explain the decision-making process utilizes machine learning algorithms and experimented using the Cleveland, Hungarian and Va long beach heart disease datasets to predict heart disease [19]. A fuzzy rule-based system combined with the cosine amplitude method and fuzzy classifier has been invented for the classification of diabetes and tested using PIDD [20]. An expert fuzzy logic model for classification of surgical risks was developed to assist physicians in the prediction of postoperative complications of prostatic hyperplasia before surgery [21]. A fuzzy-based insulin advisory system that adopts a non-linear delay mechanism has been developed to assist an artificial pancreas for diabetes Type 1 patients [34]. A study to utilizing optimal decision tree algorithm, Modified Adaptive Neuro Fuzzy Inference System (M-ANFIS) and K-Nearest Neighbor (K-NN) was implemented to diagnose diabetes and validated using the PIDD [35]. An Enhanced Genetic Algorithm (EGA) based Fuzzy Weight updating Support Vector Machine (FWSVM) algorithm to diagnose early heart disease is proposed and tested using the Cleveland dataset [36]. Other than the medical area, a decision-making model is also created for phrase classification based on a fuzzy framework [22]. Table 2 in Appendix summarizes the models for decision-making explained in this section.

Based on our related work in Table 2, half of the research used publicly available datasets [19]-[20], [31]-[33], [35]-[36], whereas the remaining half research work in Table 2 used collected dataset from implemented research [2], [7], [18], [21], [28], [33]-[34]. This indicates the importance of developing and validating datasets to improve decision-making by providing acceptance and usable models. The dataset used in our research work is based on the simulated diabetes treatment dataset validated by medical expert, that provides predictor and target attributes for diabetes diagnosis and level of care.

In Table 2, Artificial Intelligence (AI), ML and data mining techniques have been proposed for decision-making models. Majority of the research focus on a single technique [7], [18]-[19], [21]-[22], [31]-[33], [35] and the best accuracy was produced from a fuzzy model with accuracy of 97.7% [22]. Several techniques have also been combined to improve the model [2], [20], [36] and the best accuracy produced is 99% [2].

The interrelated decision-making model in healthcare proposed allows information to be shared between key decision-makers and provides iterative data flows that, mimics the real world [2]. However, vagueness and uncertain data exists and gives room to be improved. Due to the fact, that fuzzy technique has been widely accepted for capturing expert knowledge in a more intuitive and human-like manner, therefore fuzzy is proposed to be embedded in the existing interrelated decision-making model [2], for a better comprehensive decision-making. Figure 1 shows the interrelated decision-making model for diabetes diagnosis [37], consisting of five level of care: Primary, Secondary, Tertiary, Quaternary and Palliative Care. However, only Primary and Secondary level of care are focused on this research work because studies have been proved that the interrelated decision-making model was working as expected with only Primary and Secondary care [2], [37]. Therefore, the proposed fuzzy model which provides sufficient information based on the interrelated decision-making model handles the lack of information in the diabetes dataset, as well as the vagueness and uncertainty issues. The proposed model is vital in order to make accurate decision-making to diagnose diabetes.

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

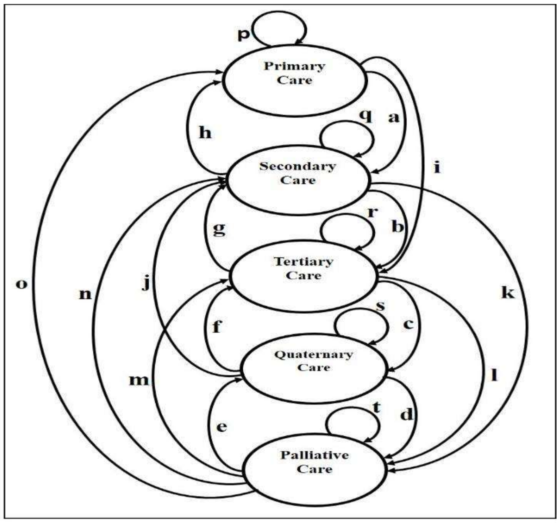

Figure 1: Interrelated decision-making model for diabetes diagnosis [37]

3. METHODOLOGY

A fuzzy model for diabetes diagnosis and level of care is proposed, as shown in Figure 2. The first step is preprocessing of the datasets. The second step is the design of the fuzzy model by defining fuzzy sets that represents the linguistic modifiers. The third step is construction of the fuzzy rules. The fourth step is modeling of the fuzzy diabetes diagnosis and level of care using ML algorithms. The final step is the evaluation of the fuzzy model using accuracy, precision, recall, F1-Score and confusion matrix performance evaluation methods for classification and prediction.

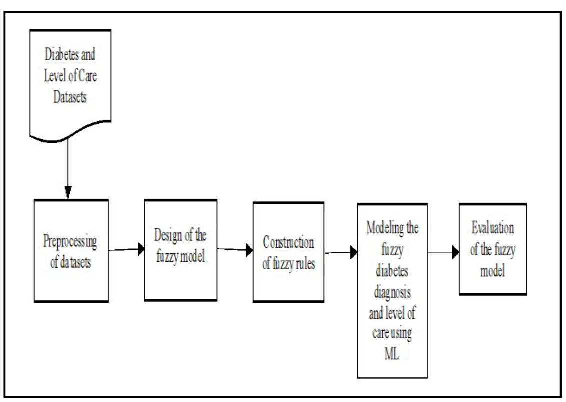

Figure 2: Proposed Diabetes Diagnosis and Level of Care Fuzzy Model

3.1 Description and Preprocessing of Datasets

As mentioned in section 2.1, this research work is based on the simulated diabetes datasets that were validated by medical experts [2]. The dataset consists of diabetes diagnosis dataset and the level of care dataset. The diabetes diagnosis dataset consists of 5000 records and 37 attributes. The level of care dataset consists of 5000 records and 37 attributes. There are 36 predictor attributes and 1 target attribute for each dataset. Example of the predictor attributes are shown in Table 1. The target attribute for the diabetes diagnosis datasets consists of "HealthyPrediabetes", "IFGIGTPrediabetes" and "T2DMDiabetes". The target attribute for the level of care dataset consists of "PrimaryCare" and "SecondaryCare". These target attributes will be further explained in section 4, which is associated to the confusion matrix.

Due to the large number of missing values in the diabetes diagnosis and the level of care datasets, the datasets were cleaned by removing the missing values. The cleaned diabetes diagnosis dataset consists of 2616 records, while the level of care dataset consists of 2422 records, with both datasets containing 37 attributes.

3.2 Fuzzy Model Design and Fuzzy Rules Construction

The fuzzy model design and fuzzy rules construction is improved and extended from a previous model produced [38]. The fuzzy model is designed starting with the initialization of input and output variables. The crisp input variables or predictor attributes and output variable or target attribute are initialized. The crisp input and output variables are then transformed to fuzzy linguistic variables and the membership functions are constructed for each fuzzy variable. Table 3 shows the fuzzy representation, where the number of membership functions is associated to each of the linguistic labels for some of the predictor attributes and target attribute for diabetes diagnosis and level of care datasets.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: Fuzzy Representation

Attributes		Number of Linguistic Labels	Lists of Linguistic Labels
Predictor Attributes	Age	3	[Young, Adult, Old]
	Sex	2	[Male, Female]
	Acanthiosis Nigricans	2	[No, Yes]
	A1c	3	[Normal, Prediabetes, Diabetes]
	FPG	3	[Normal, IFG, DM]
	RPG	3	[Normal, OGTT, Second RPG]
	OGTT	6	[FPGNormal, IPGIFG, FPFDM, PPGNormal, PPGIGT, PPGDM]
	HDL	3	[Low, Borderline, High]
	TG	3	[Optimal, Elevated, High]
	CVD	2	[No, Yes].
Target Attribute	Diabetes Diagnosis	3	[HealthyPrediabetes, IFGIGTPrediabetes, T2DMDiabetes]
	Level of Care	2	[PrimaryCare, SecondaryCare]

Figure 3 shows the variables Age, Sex and OGTT which are chosen to indicate the fuzzy partitions with 3, 2 and 6 linguistic labels respectively.

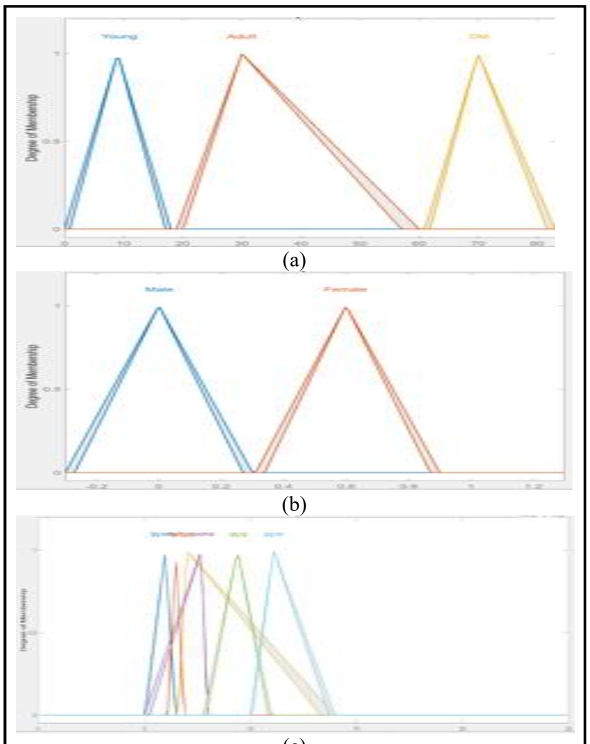

Figure 3: The Fuzzy Partitions of Linguistic Variables (a) Age (b) Sex (c) OGTT

The next process is to construct fuzzy rules. Fuzzy rules are constructed to infer an output based on the input variables. The fuzzy rule is based on the implication [39]-[41]:

IF x is A THEN y is B

Where, the premise *x* is *A,* and the consequent *y* is *B* can be true to a degree, instead of entirely true or entirely false. The linguistic variables A and B are represented using fuzzy sets. The lists of linguistic labels illustrated in Table 3 are the fuzzy sets.

The design of the fuzzy model and fuzzy rules are done by referring to academic research work and Ministry of Health Malaysia sources [2], [37]-[38]. The fuzzy rule evaluation is based on the operator OR. Table 4 shows some of the proposed fuzzy rules. For example, the fuzzy rule for row 2 in Table 4 is:

IF (age is Adult) OR (sex is Male) OR (acanthosis nigricans is Yes) OR (A1c is Normal) OR (FPG is Normal) OR (RPG is OGTT) OR (OGTT is Nil) OR (HDL is High) OR (TG is Optimal) OR (CVD is No) THEN (Diabetes Diagnosis is Prediabetes)

Table 4: Fuzzy Rules

Age	Sex	Acanthosis Nigicans	A1c	FPG	RPG	OGTT	HDL	TG	CVD	Diabetes Diagnosis	Level of Care
Adult	Male	Yes	Nomal	Nomal	OGTT	Nil	High	Optimal	No	Healthy Prediabetes	Primary Care
Adult	Female	Yes	Nomal	IFG	Second RPG	FPG Nomal	High	High	No	IFGIGT Prediabetes	Primary Care
Adult	Female	Yes	Diabetes	DM	Second RPG	Nil	Low	High	No	T2DM Diabetes	Primary Care
Old	Male	No	Diabetes	DM	Second RPG	Nil	Low	High	Yes	T2DM Diabetes	Secondary Care
Old	Male	No	Diabetes	DM	Second RPG	Nil	Borderline	High	Yes	T2DM Diabetes	Secondary Care
Old	Female	No	Prediabetes	IFG	OGTT	PPGIGT	High	High	No	IFGIGT Prediabetes	Primary Care

3.3 Fuzzy Modeling Utilizing Machine Learning

The constructed fuzzy model is tested using ten machine learning techniques. Six machine learning technique selected are J48, Logistic, NaiveBayes Updateable, RandomTree, BayesNet and AdaBoostM1 which is based on our main references [2], [37]. Additional four machine learning technique chosen are Random Forest, Multilayer Perceptron, Bagging and Stacking which is based on techniques utilized in others previous research works [19], [22], [31]-[32].

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.4 Performance Evaluation Method

The measurements used to evaluate the validity of the fuzzy model are accuracy, precision, recall, F1-Score, and confusion matrix. A confusion matrix is a table that defines the predicted class and the actual class, showing the number of predictions which are correct and incorrect per class. Confusion matrix consists of binary classification with only two classes to classify and multiclass classification with more than two classes to classify. This research produced binary classification confusion matrix for the level of treatment dataset and three class classification confusion matrixes for the diabetes diagnosis dataset. Figure 4 shows a binary classification confusion matrix for positive and negative classes, explanation as follows:

True Positive (TP): Refers to the number of predictions where the classifier correctly predicts the positive class as positive.
False Negative (FN): Refers to the number of predictions where the classifier incorrectly predicts the positive class as negative.
False Positive (FP): Refers to the number of predictions where the classifier incorrectly predicts the negative class as positive.
True Negative (TN): Refers to the number of predictions where the classifier correctly predicts the negative class as negative.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 4: Binary Classification Confusion Matrix

Figure 5 shows an example of a three-class classification confusion matrix for classes A, B and C, calculated as follows:

TP = 1, FN = 2+3 = 5, FP = 4+7 = 11 and TN = 5+6+8+ 9 = 28

		Predicted Class		
		A	B	C
Actual Class	A	1	2	3
	B	4	5	6
	C	7	8	9

Figure 5: Three-class Classification Confusion Matrix

Accuracy measures the percentage of data points that are correctly identified, providing the overall accuracy of the model. Accuracy is calculated as follows:

Accuracy=(TN+TP)/(TP+FN+FP+TN) (1)

Precision indicates what fraction of predictions as a positive class were positive. Precision is calculated as follows:

Precision=TP/(TP+FP) (2)

Recall implies what fraction of all positive samples were correctly predicted as positive by the classifier. Recall is calculated as follows:

Recall=TP/(TP+FN) (3)

F1-Score combines the precision and recall into a single measure. F1-Score is calculated as follows:

F1-Score=2TP/(2TP+FP+FN) (4)

3.5 Experiment Setting

The fuzzy diabetes diagnosis and level of care model was designed using the MATLAB R2022b Fuzzy Logic Designer application and the experiments for the fuzzy datasets were implemented using WEKA version 3.8.6. The processor used is the 11th Gen Intel Core i7 with the speed of 2.80 GHz and RAM memory of 16 GB.

The experiment applied the ten machine learning algorithms which are J48, Logistic, NaiveBayes Updateable, RandomTree, BayesNet, AdaBoostM1, Random Forest, Multilayer Perceptron, Bagging and Stacking uses the 10-fold cross validation method.

4. RESULTS AND DISCUSSION

The diabetes diagnosis dataset consists of 2616 records or instances. The accuracy result for the eight algorithms: J48, Logistic, NaiveBayes Updateable, RandomTree, BayesNet, Random Forest, Multilayer Perceptron and Bagging are 100%. This proved the effectiveness of the fuzzy model executed using the machine learning algorithms. The accuracy for AdaBoostM1 is 79.82% and the accuracy for Stacking is 67.89%. The confusion matrix for AdaBoostM1 is shown in Figure 6 and the confusion matrix for Stacking is shown in Figure 7. The actual and predicted classes in Figure 6 and Figure 7 consist of the categories of diabetes, which are “HealthyPrediabetes” (HP),

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

"IFGIGTPrediabetes" (IFGIGTP) and "T2DMDiabetes" (T2DMD).

		Predicted Class		
		IFGIGTP	T2DMD	HP
Actual Class	IFGIGTP	0	0	528
	T2DMD	0	312	0
	HP	0	0	1776

Figure 6: AdaBoostM1 Algorithm Confusion Matrix

Figure 6 shows that the classifier correctly predicts 312 of the T2DMDiabetes class and 1776 of the HealthyPrediabetes class based on the actual class. However, all 528 of the IFGIGTPrediabetes classes are incorrectly classified by the classifier. As a result, this reduces the percentage of the accuracy.

		Predicted Class		
		IFGIGTP	T2DMD	HP
Actual Class	IFGIGTP	0	0	528
	T2DMD	0	0	312
	HP	0	0	1776

Figure 7: Stacking Algorithm Confusion Matrix

Figure 7 shows that the classifier only predicts HealthyPrediabetes class as the correct class with a number of 1776, based on the actual class. However, all 528 of the IFGIGTPrediabetes and 312 of the T2DMDiabetes classes are incorrectly classified by the classifier. As a result, this reduces the percentage of accuracy more for the Stacking algorithm compared to AdaBoostM1.

The level of care dataset consists of 2422 records or instances. The accuracy, precision, recall and F1-Score result for the all the ten-machine learning algorithm tested is shown in Table 5. The level of care datasets produced the highest accuracy of 97.15% for MLP and Bagging algorithms and the lowest accuracy of 91.66% for stacking algorithm. For the Stacking algorithm, no value is given to precision and F1-Score, and this is associated to the confusion matrix which will be explained shortly.

Table 5: Result for Level of Care Dataset

ML Algorithm	Accuracy (%)	Precision	Recall	F1-Score
J48	97.11	0.975	0.971	0.972
Logistic	96.33	0.967	0.963	0.965
NaiveBayes Updateable	93.10	0.937	0.931	0.934
RandomTree	97.07	0.975	0.971	0.972
BayesNet	93.10	0.937	0.931	0.934
AdaBoostM1	95.38	0.956	0.954	0.955
Random Forest	97.11	0.972	0.971	0.972
Multilayer Perceptron	97.15	0.974	0.972	0.972
Bagging	97.15	0.975	0.972	0.973
Stacking	91.66	-	0.917	-

Figure 8 shows the bar chart comparison in terms of accuracy for the ten ML algorithms. All the algorithms produced a good accuracy of 90% above. Figure 9 shows the bar chart comparison in terms of precision, recall and F1-Score for the nine ML algorithms except the Stacking algorithm because of the non-existence value for the precision and F1-Score as shown in Table 5.

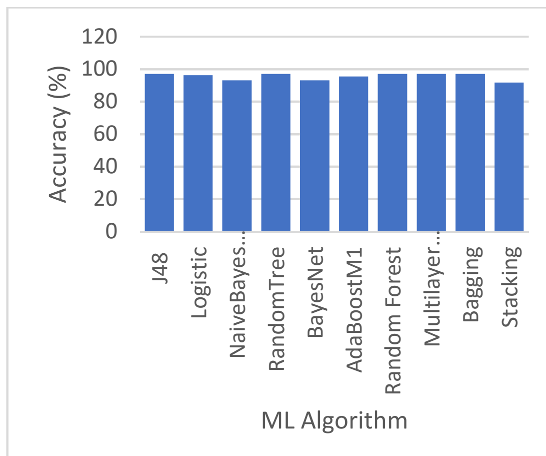

Figure 8: Accuracy for Level of Care Dataset

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

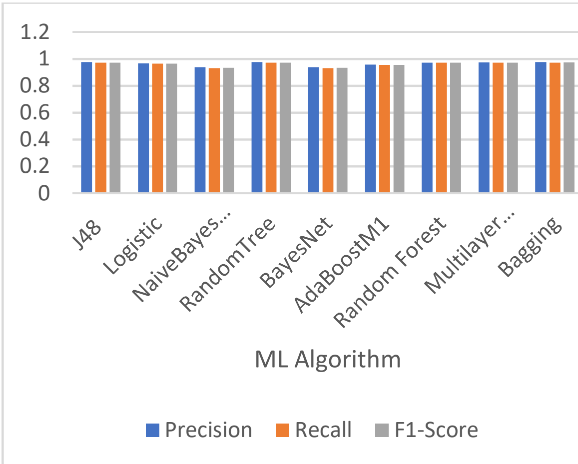

Figure 9: Precision, Recall and F1-Score for Level of Care Dataset

Figure 10 and Figure 11 shows the confusion matrix for the MLP and Bagging algorithms with the highest accuracy of 97.15% and Figure 12 shows the confusion matrix for Stacking algorithm with the lowest accuracy of 91.66%. From Figures 10 and 11, a number of 69 items are incorrectly classified as “SecondaryCare” and “PrimaryCare” (21+48 from MLP algorithm and 16+53 from Bagging algorithm) which is highlighted. For the Stacking algorithm, a larger number of 202 items are incorrectly classified highlighted in Figure 12. As mentioned, the non-existence values for precision and F1-Score is associated to the Stacking confusion matrix, meaning that the Stacking algorithm is not a good algorithm for the proposed fuzzy model.

		Predicted Class	
		SecondaryCare	PrimaryCare
Actual Class	SecondaryCare	181	21
	PrimaryCare	48	2172

Figure 10: MLP Algorithm Confusion Matrix for Level of Care Dataset

		Predicted Class	
		SecondaryCare	PrimaryCare
Actual Class	SecondaryCare	186	16
	PrimaryCare	53	2167

Figure 11: Bagging Algorithm Confusion Matrix for Level of Care Dataset

		Predicted Class	
		SecondaryCare	PrimaryCare
Actual Class	SecondaryCare	0	202
	PrimaryCare	0	2220

Figure 12: Stacking Algorithm Confusion Matrix for Level of Care Dataset

From the results, it shows that the diabetes diagnosis dataset with three classes produced better results compared to the level of care dataset with two classes in terms of accuracy, precision, recall and F1-Score. Further investigation for performance improvement needs to be done to analyze the AdaBoostM1 and Stacking algorithms for the diabetes diagnosis dataset with three classes and the Stacking algorithm for the level of care dataset with two classes. Generally, the proposed fuzzy model produced an accuracy of 67.89% to 100%. Furthermore, the proposed fuzzy model which is embedded into the interrelated decision-making model for diabetes diagnosis proposed by Normadiah [2], [37] is working as expected.

From this study, it was found that the proposed fuzzy model that incorporate the features of interrelated decision-making caters the problem of vagueness and uncertainty in the simulated diabetes datasets. The evaluation analysis criteria based on the accuracy, precision, recall, F1-Score, and confusion matrix results show a detailed analysis of two and three class classification in a way that is satisfactory comprehend with the application of fuzzy technique, compared to the previous recent models [7], [8], [9], [22], [28], [34], [35], [36], [37]. Four popular supervised machine learning algorithms utilized in recent researches, namely the Random Forest, Multilayer Perceptron, Bagging and Stacking totaling ten algorithms were added from the research work used as our benchmark which only used six supervised machine learning algorihms [2]. The reason is to strengthen and proof the effectiveness of the proposed fuzzy rule-based model using the evaluation analysis criteria mentioned. Moreover, the evaluation analysis criteria implemented in this research work signifies an improved human understanding model for decision-making in the real world. To date, there have not exist any studies which invented a model that associates types of diabetes diagnosis with the level of care in a human view approach to assist decision-making in the medical real life.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. CONCLUSION AND FUTURE WORKS

The critical need for a human understanding mechanism in medical datasets especially diabetes datasets and the relation between these datasets, are important for an effective decision-making that represents the real world. Researches have proved that fuzzy logic is a successful method that imitates the human practice in decision-making. The proposed fuzzy rule-based model which measures the accuracy, precision, recall, F1-Score, and confusion matrix validates the outstanding results that were produced, which is tested with most of the machine learning algorithms.

As a conclusion, the proposed fuzzy model tested with most of the machine learning techniques provides a good performance. The vagueness and uncertainty of the proposed fuzzy model is handled with the utilization of linguistic labels for a comprehensive and reliable decision-making, which is beneficial to the healthcare sector.

For future works, the AdaBoostM1 and Stacking algorithms need to be improved with the application of the proposed fuzzy model. Additional studies are required on the detailed algorithms to analyze the reasons of the low accuracy produced. Furthermore, the speed of the machine learning algorithms can also be further investigated to produce a more efficient model.

ACKNOWLEDGEMENT

Thank you to my supervisor, Prof. Dr. Azuraliza Abu Bakar and co-authors, Dr. Normadiah Mahiddin and Dr. Maslina Zolkepli for their expertise in assisting me to complete this paper. Thank you also to Universiti Putra Malaysia for letting me to conduct my sabbatical research at Universiti Kebangsaan Malaysia. My appreciation, to the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, for providing financial support under project code 6236400.

REFERENCES:

[1] M. S. L. M. Roobinia, "Prediction of Diabetes Mellitus using Statistical Methods", *International Journal of Pure Applied Mathematics,* Vol. 118, 2018, pp. 385-390.

[2] Normadiah Mahiddin, Zulaiha Ali Othman, Azuraliza Abu Bakar, and Nur Arzuar Abdul Rahim, "An Interrelated Decision-Making Model for an Intelligent Decision Support System in Healthcare", *IEEE Access*, Vol. 10, 2022, pp. 31660-31676.

[3] S. Jaiswal, and T. Jaiswal, "Machine Learning-Based Classification Models for Diagnosis of Diabetes", *Recent Advances in Computer Science and Communications*, Vol. 15, No. 6, 2022, pp. 813-821.

[4] J.B. Awotunde, S. Misra, and Q.T. Pham, "An Enhanced Diabetes Mellitus Prediction Using Feature Selection-Based Type-2 Fuzzy Model", *Communications in Computer and Information Science*, 1688 CCIS, 2022, pp. 625-639.

[5] J.A. Nettleton, A.E. Burton, and R.C. Povey, "No One Realizes What We Go Through as Type 1s: A Qualitative Photo-elicitation Study on Coping with Diabetes", *Diabetes Research and Clinical Practice*, Vol. 187, 2022, pp. 1-7.

[6] A. Ehtasham, L. Soo, L. Roberta, R.W. David, and J.D. Melanie J, "Type 2 Diabetes", *Lancet*, Vol. 400, 2022, pp. 1803-1820.

[7] A.E Oumer, and D. Getachew, "Application of Supervised Machine Learning Algorithms for Classification and Prediction of Type-2 Diabetes Disease Status in Afar Regional State, Northeastern Ethiopia", Scientific Reports, Vol. 13, 2023, pp. 1-12.

[8] K. Kangra, and J. Singh, "Comparative Analysis of Predictive Machine Learning Algorithms for Diabetes Mellitus", *Bulletin of Electrical Engineering and Informatics*, Vol. 12, No. 3, 2023, pp. 1728-1737.

[9] P.N. Thotad, G.R. Bharamagoudar, and B.S. Anami, "Diabetes Disease Detection and Classification on Indian Demographic and Health Survey Data using Machine Learning Methods", Diabetes & Metabolic Syndrome: Clinical Research & Reviews, Vol. 17, No. 102690, 2023, pp. 1-7.

[10] K.C. Howlader, M.S. Satu, M.A. Awal, M.R. Islam, S.M.S. Islam, J.M.W Quinn, and M.A. Moni, "Machine Learning Models for Classification and Identification of Significant Attributes to Detect Type 2 Diabetes", *Health Information Science and Systems*, Vol. 10, No. 1, 2022, pp. 1-13.

[11] C.C. Olisah, L. Smith, and M. Smith, "Diabetes Mellitus Prediction and Diagnosis from a Data Preprocessing and Machine Learning Perspective", *Computer Methods and Programs in Biomedical*, Vol. 220, No. 106773, 2022, pp. 1-12.

[12] J.J. Khanam, and S.Y. Foo, "A Comparison of Machine Learning Algorithms for Diabetes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Prediction", *ICT Express*, Vol. 7, 2021, pp. 432-439.

[13] N. Gupta, H. Singh, and J. Singla, "Fuzzy Logic-based Systems for Medical Diagnosis", *Proceedings of the Third International Conference on Electronics and Sustainable Communication Systems (ICESC 2022)*, 2022, pp. 1058-1062.

[14] R.H. Abiyev, and H. Altiparmak, "Type-2 Fuzzy Neural System for Diagnosis of Diabetes", Mathematical Problems in Engineering, No. 5854966, 2021, pp. 1-9.

[15] G.M. Bressan, B.C.F. de Azevedo, and R.M. de Souza, "A Fuzzy Approach for Diabetes Mellitus Type-2 Classification, *Brazilian Archives of Biology and Technology*, Vol. 63, No. e20180742, 2020, pp. 1-11.

[16] S. Khokhar, Q. Peng, R. Touqir, and M.S. Khan, "Medical Condition Monitoring System using Fuzzy Logic", *2020 IEEE International Conference on Artificial Intelligence and Infoermation Systems (ICAIIS)*, 2020, pp. 211-216.

[17] K. Uma, and U.R. Devi, "A Study on Fuzzy Expert System for Diagnosis of Diabetes Mellitus", *International Journal of Applied Engineering Research*, Vol. 14, No. 4, 2019, pp. 129-139.

[18] G. Mohammed, A. Aisha, A. Heba, A. Meelaf, S. Rasha, A. Reem, A. Mohammed, A. Maiadah, A. Dania, A.A. Reem, and A. Waleed, "A Novel Stacking Ensemble for Detecting Three Types of Diabetes Mellitus using Saudi Arabian Dataset: Pre-diabetes, T1DM, and T2DM, *Computers in Biology and Medicine*, Vol. 147, 2022, pp. 1-12.

[19] B. Khalid, M. Mohammed, and R. Mohammed, "An Accurate Fuzzy Rule-based Classification Systems for Heart Disease Diagnosis", *Scientific African*, Vol. 14, 2021, pp. 1-14.

[20] K.A. Aamir, L. Sarfraz, M. Ramzan, M. Bilal, J. Shafi, and M. Attique, "A Fuzzy Rule-Based System for Classification of Diabetes", Sensors, Vol. 21, No. 8095, 2021, pp. 1-17.

[21] F. Sergey, T.A. Riad, S. Olga, K. Nikolay, S, Ashraf, P. Zeinab, I. Maksim, and L. Mikhail, "Biotechnical System Based on Fuzzy Logic Prediction for Surgical Risk Classification using Analysis of Current-voltage Characteristics of Acupuncture Points", *Journal of Integrative Medicine*, Vol. 20, 2022, pp. 252-264.

[22] D.V. Michael, and K.I Dimitris, "Fuzzy Similarity Phrases for Interpretable Data Classification", *Information Sciences*, 2023, Vol. 624, pp. 881-907.

[23] T. Harshil, S. Vaishnavi, Y. Hiteshri, and S. Manan, "Comparative Anatomization of Data Mining and Fuzzy Logic Techniques used in Diabetes Prognosis", *Clinical eHealth*, Vol. 4, 2021, pp. 12-23.

[24] CPG Secretariat, Health Technology Assessment Section, Medical Development Division, Ministry of Health Malaysia, "Clinical Practice Guidelines Management of Type 2 Diabetes Mellitus", 5th Edition, 2015, pp. 1-129.

[25] A.M. Egan, and S.F. Dinneen, "What is Diabetes? Key Points", *Medicine*, Vol. 1, No. 47, 2018, pp. 1-4.

[26] https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[27] https://archive.ics.uci.edu/

[28] Z. Qinpei, Z. Jinhao, S. Xuan, L. Chuwen, Z. Yinjia, L. Yuxiang, C. Baige, L. Jiangfeng, L. Xiang, R. Weixiong, and W. Congrong, "Chinese Diabetes Datasets for Data-driven Machine Learning", *Scientific Data*, 2023, pp. 1-8.

[29] C. Jyotismita, S.T. Ganesh, S.K. Cidham, and S.A. Theertan, "Machine Learning and Artificial Intelligence Based Diabetes Mellitus Detection and Self-management: A Systematic Review", *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, 2022, pp. 3204-3225.

[30] M. Anthony, P. John, and L. Lu, "A Review of Regression and Classification Techniques for Data Analysis of Common and Rare Variants and Gene-environmental Factors", *Neurocomputing*, Vol. 489, 2022, pp. 466-485.

[31] M.B. Umair, L. Sukumar, A. Mubashir, H.H. Fadratul, B. Anees, and H.R.S. Hafiz, "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications", *Journal of Healthcare Engineering*, Vol. 2021, 2021, pp. 1-17.

[32] J.K. Jobeda, and Y.F. Simon, "A Comparison of Machine Learning Algorithms for Diabetes Prediction", *ICT Express*, Vol. 7, 2021, pp. 432-439.

[33] C.O. Chollette, S. Lyndon, and S. Melvyn, "Diabetes Mellitus Prediction and Diagnosis from a Data Preprocessing and Machine Learning Perspective", *Computer Methods and Programs in Biomedicine*, Vol. 220, 2022, pp. 1-12.

[34] S.A. Nilam, and H.P. Singh, "Computer-Controlled Diabetes Disease Diagnosis Technique Based on Fuzzy Inference Structure

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for Insulin-dependent Patients", *Applied Intelligence*, Vol. 53, 2023, pp. 1945-1958.

[35] S. Lakshmi, V. Reshma, and D.N. Mysore, "Early Prediction of Diabetes Diagnosis using Hybrid Classification Techniques", *IAES International Journal of Artificial Intelligence*, Vol. 12, 2023, pp. 1139-1148.

[36] G. Sugendran, and S. Sujatha, "Earlier Identification of Heart Disease using Enhanced Genetic Algorithm and Fuzzy Weight Based Support Vector Machine Algorithm", *Measurement: Sensors*, Vol. 28, 2023, pp. 1-9.

[37] N.Mahiddin, "An Intelligent Clinical Decision Support System Based on Multiagent System for Healthcare", PhD Thesis, 2020.

[38] Teh Noranis Mohd Aris, Azuraliza Abu Bakar, Normadiah Mahiddin, and Maslina Zolkepli, "A Fuzzy Inference Model for Diagnosis of Diabetes and Level of Care", *Journal of Theoretical and Applied Information Technology*, Vol. 101, No. 15, 2023, pp. 5962-5975.

[39] N. Michael, Artificial Intelligence: A Guide to Intelligent Systems, 3rd ed., United Kingdom: Addison Wesley, 2011.

[40] H.T. Nguyen, and E.A. Walker, "A First Course in Fuzzy Logic", CRC Press, Inc., Florida, United States of America, 1999.

[41] https://en.wikipedia.org/wiki/Fuzzy_rule

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

APPENDIX

Table 2: Summary of Models for Decision-Making

Ref.*	Year	Dataset	Techniques	Comments
[19]	2021	Cleveland, CombinedHunVa	ML: NN, Support Vector Machine (SVM), K-NN, Naïve Bayes (NB), Random Forest (RF)	Cleveland dataset: Naïve Bayes achieved highest accuracy of 84.51% and Random Forest achieved the lowest accuracy of 82.17% CombinedHunVa dataset: SVM achieved the highest accuracy of 82.67% and Random Forest achieved the lowest accuracy of 75.34%
[20]	2021	PIDD	Fuzzy + cosine amplitude method	96.47%
[31]	2021	PIDD	ML: J48, K-NN, Feed Forward Neural Network, RB-Bayes, NB, NN, Proposed Long Short-Term Memory (LSTM)	Proposed LSTM achieved the highest accuracy of 87.26% J48 achieved the lowest accuracy of 67.9%
[32]	2021	PIDD	ML: Decision Tree (DT), KNN, RF, NB, AdaBoost, LR, SVM	KNN (Splitting) and AdaBoost (Splitting) achieved the highest accuracy of 79.42% DT (Splitting) achieved the lowest accuracy of 73.14%
[2]	2022	Simulated diabetes treatments	Multi-agent + ML: J48, Logistic, Naivebayes Updateable, RandomTree, BayesNet, AdaBoost	J48 achieved the highest accuracy of 99% AdaboostM1 achieved the lowest accuracy of 46%
[18]	2022	Saudi Arabian hospital	ML: SVM, RF, K-NN, DT, Bagging and Stacking	Stacking achieved the highest accuracy of 94.48% and DT achieved the lowest accuracy of 84.50%
[21]	2022	Homogeneous patients underwent prostate surgery	Expert fuzzy model	97%
[33]	2022	PIDD and Laboratory of the Medical City Hospital (LMCH) diabetes	Proposed twice-growth deep neural network (2GDNN)	PIDD: 97.25% Lowest accuracy: 97.33%
[7]	2023	Secondary dataset from the medical dataset record review in Afar, Northeastern Ethiopia	ML: DT, J48, NN, K-NN, SVM, Binary Logistic Regression, RF, NB	RF achieved the highest accuracy of 93.8% SVM achieved the lowest accuracy of 85.5%
[22]	2023	Publicly available datasets: Appendicitis, Australian, Banana, Bands, Diabetes, Haberman, Ionosphere, Liver, Ring, Letter Recognition	Fuzzy Similarity Phrases (FSPs)	Highest accuracy achieved for Ring dataset: 97.7% Lowest accuracy achieved for Liver dataset: 62.1%
[28]	2023	ShanghaiT1DM and ShanghaiT2DM	Preparation of datasets for future research in data-driven machine learning classification techniques	Research work describes the ShanghaiT1DM and ShanghaiT2DM datasets on Type 1 and Type 2 diabetes patients
[34]	2023	Insulin-dependent patients	Fuzzy non-linear delay controller	Four experiments tested over the proposed controller in terms of (i) no meal consumption, (ii) multiple meal in take in a day, (iii) atypical meal input, (iv) uncertainties in model's parameter. The results support that the proposed controller supports an artificial pancreas of clinical patients.
[35]	2023	PIDD	Optimal Decision Tree, M-ANFIS, K-NN	M-ANFIS achieved the highest accuracy of 97.5% K-NN achieved the lowest accuracy of 77.1%
[36]	2023	Cleveland	Proposed EGA + FWSVM	93.26%

* - References