Advances and Applications in Mathematical Sciences
Volume 21, Issue 6, April 2022, Pages 3233-3247

A SOFT COMPUTING MODEL WITH REDUCED FEATURES USING INTUITIONISTIC FUZZY LOGIC FOR DISEASE CLASSIFICATION

SHAHLA TARANNUM[1] and SURAIYA JABIN[2]

[1,2]Department of Computer Science
Faculty of Natural Sciences
Jamia Millia Islamia
New Delhi-25, India
E-mail: sjabin@jmi.ac.in

Abstract

During pregnancy a woman goes through lots of hormonal changes; the most important of these hormones is HPL (Human Placental Lactogen) causing gestational diabetes. It increases the maternal blood glucose level; and so the baby is unable to get proper nutrients. At the fourth month of pregnancy, another hormone HPG (Human Placental growth) also starts increasing which raises glucose level further and so baby is in danger to get right amount of nutrients. Medical practitioners are not sure about why some women develop it. Therefore sometimes, in some cases, they are hesitant to declare it confidently. In this paper, the Intuitionistic Fuzzy Decision Tree Algorithm is used over a dataset to develop a model that helps the medical expert in the process of disease diagnosis. But before applying intuitionistic fuzzy decision tree algorithm, prominent features of dataset were selected using intuitionistic fuzzy mathematical method.

1. Introduction

Diabetes has been one of the serious problems in India with having second largest number of people suffering with it. Unfortunately, India is at the top in case of gestational diabetes. The exact cause of GDM (Gestational Diabetes Mellitus) is still not known but it is a serious kind of diabetes which occurs during pregnancy. In India, the rate of GDM is estimated to be 10-14.3% which is much higher than the western countries [MHD, GOI (2018)].

2020 Mathematics Subject Classification: 03B52.
Keywords: Human Placental Lactogen, Human Placental growth, Gestational Diabetes, Intuitionistic Fuzzy; Decision Tree; Intuitionistic Fuzzy Decision Tree Algorithm.
[1]Corresponding author; E-mail: starannum@jmi.ac.in
Received November 28, 2021; Accepted January 31, 2022

It is becoming one of the major public health issues in India with reported rate between 4.6%-14% in urban regions and 1.7%-13.2% in rural regions [Anjana RM, et al. [1]]. It is found, globally, that around 10% pregnancies is associated with diabetes, and out which 90% are GDM. Undiagnosed or inadequately treated GDM can cause type-2 diabetes in later part of the life of mother and her child [Chudasama, Rajesh K., et al. [6]]. Many other complications may arise like increased frequency of macrosomia, hypoglycaemia, persistent hypoglycaemia, intra-uterine growth retardation, congenital heart disease, respiratory disease, renal and urinary defect, upper limb deficiencies etc. Therefore, the solution of these problems is the detection of GDM at early stage of pregnancy [Cho H.N [5]].

With the development of artificial intelligence and data analysis techniques, healthcare expert system, using machine learning, may be used to predict the risk of GDM by analyzing factors related to patients. The dataset has been used for modeling by using Intuitionistic Fuzzy Logic based feature selection and fuzzy decision tree for early detection of the disease [Wu Yan-Ting, et al., [28]], [Kim, et al. [13]], [Sacgdeva, et al. [24]].

The paper is organized as follows; Section 2 describes the review on literature survey. Section 3 is about methodology. Section 4 describes the proposed Gestational Diabetes Mellitus prediction Model. Section 5 explains the Evaluation and Result and section 6 is the conclusion.

2. Literature Review

Many researchers have developed different models for decision making process using various machine learning algorithm. In this section, we present various existing models. The research paper [Padmavathi, et al. [21]] highlights the steps in preprocessing, and then classifier performance is evaluated to observe the effectiveness of preprocessing. The proposed preprocessing method used Pima Indian Diabetes is classified using SVM, RF and KNN.

In [Nilashi, et al. [19]] researcher developed an intelligence system with clustering and classification approaches. For clustering, expectation maximization is used, for outliers Principal Component Analysis (PCA) were used and for classification SVM were used. The proposed method was executed for incremental situation by applying the incremental PCA and SVM. The results on Diabetes dataset shows that incremental approaches improve accuracy and reduces execution time as compared to non incremental approaches.

In [Barach, et al. [2]], research paper demonstrated a new approach based on induction of fuzzy decision tree in clinical data. This approach build decision models with different properties. Different types of fuzzy decision trees like stable, ordered and non-ordered are discussed in this paper. The prediction accuracy enhanced by using new approaches based on fuzzy decision trees.

In [Reddy, et al. [22]], the researcher worked on ensemble based machine learning model consisting of various Machine Learning (ML) Algorithms and experimented the algorithms with diabetic retinopathy dataset. Different ML algorithm like-K-NN, LR, RF, DT, were experimented. The dataset was normalized by min-max method then training of the ensemble model was done. The performance of the model was evaluated and comparative analysis was done and found that the ensemble machine learning model outperforms the individual machine learning algorithms.

In [Djatna, et al. [8]], to diagnosis the different types of stroke disease the researcher proposed Decision Tree by mapping observation data into Intuitionistic Fuzzy Set which have a membership function, non-membership function, and a hesitation degree for each record. Hamming Distance is calculated between the Intuitionistic Fuzzy sets as required. The study shows that the approach produces the one of the best diagnosis performance compared to the other.

In [El-Sappagh, et al. [9]], the system gives novel improvements to enhance the accuracy, make the system intelligent, smart, and interpretable Clinical Decision Support System. The system takes the ontology semantic similarity of diabetes complications and symptoms with the fuzzy rules. The system got tested with real dataset. The proposed system helps physicians and patients in accurately diagnosing diabetes mellitus.

In [Charbuty, et al. [4]], researcher compared the performance and most commonly used machine learning techniques. This paper present summary of different algorithm, datasets and the results achieved from different algorithms. In research paper [Lukmanto, et al. [15]], classification framework to classify the DM dataset using *f*-score feature selection and Fuzzy SVM has been proposed. By the Fuzzy SVM classifier data were effectively trained to generate fuzzy rules.

The outcome of the model was promising and in future the model can be enhanced for better accuracy. In paper [Rubini, et al. [23]], the researcher presents a comparative analysis of different machine learning techniques for the classification of cardiovascular disease. The accuracy of random forest, linear regression, SVM and Naive Bayes have been shown. The accuracy of random forest is highest with 84% while the accuracy of SVM (radial basis kernel function) is 58% and lowest naive bayes accuracy with 54% for predicting cardiovascular disease while in [Motarwar, et al. [18]] the author takes heart disease data to predict the disease with 5 different machine learning algorithms such as Random Forest, Naïve Bayes, Support Vector Machine, Hoeffding Decision Tree, and Logistic Model Tree. The random forest and naive bayes perform very well with 85% accuracy.

Hence, it can be concluded that various researchers applied fuzzy logic based machine learning algorithms to further improve disease classification, somehow IFS based machine learning algorithms are yet to be explored in the area and this paper is an attempt in this direction.

3. Methodology

In this section we describe all methods used in this work for fuzzification of dataset, feature extraction using IFS, IFS based decision tree for classification, and performance measures.

3.1. Intuitionistic Fuzzy Set

The Intuitionistic Fuzzy logic, introduced by Atanassov in 1986, is the generalization of Fuzzy Logic. IFS deals with incomplete and imprecise data of a given fuzzy element. In IFS the membership, non membership and the hesitation values are considered. Information is gathered about the Fuzzy Set element which helps in increasing the membership of the value. As the membership or non membership values increase the value of hesitation will decrease. And when the value of hesitation is zero the IFS set becomes a Fuzzy set. Therefore, it is also said that IFS is the generalization of Fuzzy Set which can be used for a wider range of applications in various fields [Atanassov K. T., (1999), (1989)].

Let X be a universal set, an intuitionistic fuzzy set A in X is defined below as a set of ordered triplets $A = \{x, \mu_A(x), \upsilon_A(x) : x \in X\}$ where $\mu_A(x) : X \to [0, 1]$ and $\upsilon_A(x) : X \to [0, 1]$ are functions called "membership function" and "non-membership function" respectively such that $0 \leq \mu_A(x) + \upsilon_A(x) \leq 1 \ \forall \ x \in X$.

For each $x \in X$, the values $\mu_A(x)$ and $\upsilon_A(x)$ represent the "degree of membership" (or membership value) and "degree of non-membership" (or nonmembership value) of the element x to the intuitionistic fuzzy set A respectively. And $\pi_A(x) = 1 - \mu_A(x) - \upsilon_A(x)$ is called hesitation value. Thus the IFS set is a superset of the fuzzy set. In other words, a fuzzy set is a particular case of the intuitionistic fuzzy set if $\pi_A(x) = 0$, $x \in X$ [De, Kumar S., et al. [7]].

3.2. Feature Selection using Intuitionistic Fuzzy Logic. During modeling, feature selection plays an important role for exhibiting better performance of the model. The root of the decision tree contains entire data sets because no rule has been imposed on it yet. Each node is repetitively splitted by clustering of data sets with remaining attributes.

For the proposed work, the fuzzified attributes are described by the membership degree, non membership degree and hesitation, and we call them Intuitionistic Fuzzy attributes. In an intuitionistic fuzzy set there are three components: membership, non membership and hesitation, which may be used to identify the most relevant attributes. The values of each component of an attribute is different for every object (patient).

$$(\overline{\mu_{Ac}}) = \frac{1}{n} \sum_{i=1}^{n} \mu_{Ac}(x_i) \tag{1}$$

$$(\overline{\upsilon_{Ac}}) = \frac{1}{n} \sum_{i=1}^{n} \upsilon_{Ac}(x_i) \tag{2}$$

$$(\overline{\pi_{Ac}}) = \frac{1}{n}\sum_{i=1}^{n}\pi_{Ac}(x_i) \tag{3}$$

here $(\overline{\mu_{Ac}})$ represents the average membership value of attribute Ac, where n is the number of patients.

For any intuitionistic fuzzy attribute, the IFS value expressed in equation-3 should be low while the difference between membership and non-membership value should be maximum. The simplest function $f(A_c)$ for expressing such a condition is [Szmidt, et al. [26]].

$$f(A_c) = [(1 - \overline{\pi_{Ac}})(\overline{\mu_{Ac}} - \overline{\nu_{Ac}})] \tag{4}$$

On the basis of the value of $f(A_c)$, feature is selected which is described in Table-2.

3.3. Implementation of Intuitionistic Fuzzy Decision Tree. ID3 uses Entropy and Information Gain to construct a decision tree. A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). A completely homogeneous sample has entropy of 0. An equally divided sample has entropy of 1. Information gain is the amount of information gained by knowing the value of the attribute [Meng, et al. [17]], [Srinivasan V., [25]].

Decision tree is not robust. If you change the dataset, the structure of the tree also changes completely. This creates instability in the algorithm. To solve these issues, a Fuzzy decision tree is introduced. In this mechanism, the given dataset is converted into a fuzzy dataset; and with this fuzzy dataset, splitting of node and tree-structure is determined. A fuzzy decision tree reduces the classification ambiguity also. It is more robust in tolerating imprecise, uncertain or missing information [Tangirala S., [27]], [Hamsagayathri and Sampath, [10]], [Ludwig, et al. [14]]. It can also handle symbolic values like tall person, very tall person, short person etc.

In the decision tree, the cut-point test does not work properly on the dataset whose attributes values are closer. Two values closer to each other may happen to be split or divided on different branches. Therefore, they are placed at faraway distance. To overcome this problem, the decision tree is used over a fuzzified dataset which utilizes splitting criteria based on fuzzy restriction. In this paper we have applied a decision tree over the Intuitionistic fuzzy dataset [Tangirala S., [27]] [Janikow and Cezary, [11]].

The performance of the proposed model is evaluated using standard metrics such as Accuracy, Precision, Recall, and F1 measure. Accuracy of a model is related to True Positives and True negatives. Precision is related to the predicted positive vs. Actual positive. Recall actually calculates correctly the predicted positives out of total positive classes. F1-Value is a more informative metric than that of accuracy. It is the weighted average of Recall value and Precision value [Novakovic, et al. [20]].

4. Proposed Gestational Diabetes Mellitus Prediction Model

The workflow of proposed model is as follows:

Step-1. Convert a given dataset into a fuzzy dataset by selecting suitable linguistics and membership functions.

Step-2. Selection of the most prominent features using simplest function $f(A_c)$.

Step-3. Determine entropy and information-gain to find appropriate attribute for the root node of decision tree.

Step-4. Split root node into small subsets containing the suitable possible values for the best attributes.

Step-5. Reproduce the node containing the best attribute to construct another sub trees down the tree path.

Step-6. Recursive reproduction of new branch of decision trees and keep repeating step-4 to 5 until no further splitting is possible.

4.1. Dataset Description. The aim is to predict whether or not a patient has diabetes based on certain diagnostic features available in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients considered here are females who are at least 21 years old.

The datasets consist of several medical predictor variables (features) and one response variable. Predictor variables include the number of pregnancies the patients had, their BMI, insulin level, glucose, and so on which is mentioned below.

- Pregnancies: Number of times conceived
- Glucose: Checking the Plasma glucose concentration after 2 hours
- Blood Pressure: blood pressure measurement
- Skin Thickness: Triceps skin thickness
- Insulin: 2-Hour serum insulin
- BMI: Body mass index (weight in kg/(height in m)^2)
- Diabetes Pedigree Function: Diabetes pedigree function
- Age: Age (years)
- Outcome: class variable (0 or 1). Out of 768 data, 268 are in class 1, the others are in class 0.

The primary objective is to develop an efficient soft computing based model with the help of an Intuitionistic fuzzy dataset to predict the case of GDM. We build a soft computing model to accurately predict whether or not the patients in the dataset have diabetes or not

4.2. Data Pre-Processing and Feature Extraction using IFS. We pre-processed the dataset by eliminating tuples with missing values for more than 40% of the attributes, otherwise we substituted average value as the missing value for an attribute.

After eliminating the missing values, the dataset has been converted into a Fuzzy dataset. Fuzzy profiling of Input/ Output variables is as follows. For each attribute the possible values are low, medium, and high for their full range of values. The membership function of the intuitionistic fuzzy set has been taken as approximation type (triangular function) membership as we have defined the range in 3 categories and accordingly we have constructed the Intuitionistic fuzzy set for all the attributes. For the attribute BMI, the triangular membership is shown in figure 1. Similarly, we do for all attributes.

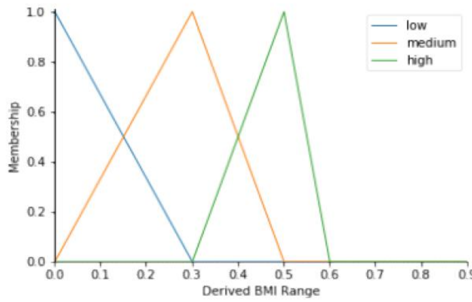

Figure 1. Membership function for BMI attribute.

The range of each feature is categorized in Table 1

Table 1. Range of Attributes.

Feature name	low	medium	high
Insulin	<150	150-300	300 and above
BMI	<18	18-24	24 and above
Glucose	<140	140-200	200 and above
Pregnancy	<1	1-3	3 and above
BP	<70	70-120	120 and above
ST	<25	25-60	60 and above
DPF	<0.5	0.5-0.7	0.7 and above
Age	<35	35-50	50 and above

Once the samples are fuzzified, we convert them into the Intuitionistic Fuzzy Set. The hesitation value is set as 0.1 for low and high category. If the membership values are 0.3 and 0.6 respectively for low and high category, their hesitation value becomes 0.1. Whereas for medium category if both the membership and non membership values are 0.5, the hesitation value becomes 0. Using the same way, we can convert the value of each fuzzified attribute to IFS for all samples.

The following Table 2 shows the intuitionistic fuzzified values of all features.

Table 2. Gestational Diabetes Mellitus Data in terms of IFS

Insulin			BMI			Glucose			Age			BP			ST			DPF			Pregnancy		
low			high			medium			low			medium			medium			medium			high		
Values of attributes in IFS Form																							
Insulin			BMI			Glucose			Age			BP			ST			DPF			Pregnancy		
μ	υ	π	μ	υ	π	μ	υ	π	μ	υ	π	μ	υ	π	μ	υ	π	μ	υ	π	μ	υ	π
0.3	0.6	0.1	0.6	0.3	0.1	0.5	0.5	0	0.3	0.6	0.1	0.5	0.5	0	0.5	0.5	0	0.5	0.5	0	0.6	0.3	0.1

Next, making use of the IFS (Table 2) we compute an average values of μ, υ and π for each attribute as per equations 1, 2, and 3. These average values are given in Table-3.

Table 3. Characteristics of Gestational Diabetes Mellitus IFS attributes.

Features	$\bar{\mu}$	$\bar{\upsilon}$	$\bar{\pi}$	$\bar{\mu}-\bar{\upsilon}$	$1-\bar{\pi}$	$f(A_c)=(1-\bar{\pi})(\bar{\mu}-\bar{\upsilon})$
BMI	0.59	0.32	0.09	0.27	0.91	0.25
Insulin	0.35	0.57	0.08	0.22	0.92	0.20
Glucose	0.35	0.57	0.07	0.22	0.93	0.20
Pregnancy	0.38	0.54	0.08	0.16	0.92	0.15
ST	0.40	0.55	0.05	0.15	0.95	0.14
BP	0.42	0.54	0.04	0.12	0.96	0.12
DPF	0.39	0.53	0.21	0.14	0.79	0.11
Age	0.52	0.42	0.06	0.1	0.94	0.09

The last column of Table-3 gives the value of $f(A_c)$ which is obtained by equation-4 for each attribute. The first four features with highest value of $f(A_c)$ are BMI, Insulin, Glucose and Pregnancy which are shown in Figure 4.

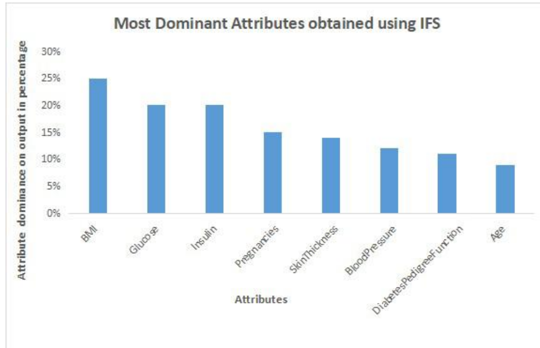

Figure 4. Dominating Features by IFS.

This way the most prominent features i.e., BMI, Insulin, Gucose, and Pregnancies, will be used for the prediction purpose. And it is expected that accuracy of the model with these 4 features should be higher than that of with given 8 features.

5. Experimental results and Evaluation of the Performance of Proposed Model

The performance of the proposed model is evaluated using standard metrics such as Accuracy, Precision, Recall, and F1 measure. Accuracy of a model is related to True Positives and True negatives. Precision is related to the predicted positive vs. Actual positive. Recall actually calculates correctly the predicted positives out of total positive classes. F1-Value is a more informative metric than that of accuracy. It is the weighted average of Recall value and Precision value [Brownlee J., [3]].

We developed model with all 8 features and with 4 prominent features. Once model is built, its performance is evaluated over the Test dataset.

The experiment with 8 features and with 4 most prominent features were also done with different machine learning techniques. The below table 4 summarizes the accuracy of different models:

Table 4. Performance comparison of different models.

Model	Accuracy with 8 features using ID3 (non-fuzzy)	Accuracy with 4 most prominent features using ID3 (non-fuzzy)	Accuracy with 4 most prominent features using Intuitionistic fuzzy decision tree
Naive Bayes	78.65	80.01	80.08
Linear Regression	78.68	79.22	80.02
SVM	63.20	63.60	70.99
Decision Tree (ID3)	78.05	79.22	81.95

Thus, finally, it is found that the accuracy of the proposed model for 4 most prominent features using an intuitionistic fuzzy decision tree shows the highest accuracy in comparison to other models as discussed in table 4.

The summary of the performance matrix of the proposed model is discussed below:

Table 5. Performance Metrics of the Proposed Model (IFS based Decision tree).

Metrics	With 8 features using ID3 (non- fuzzy)	With 4 most prominent features using ID3 (non- fuzzy)	With 4 most prominent features using Intuitionistic Fuzzy decision tree
Accuracy of model	78.05%	79.22%	81.95%
Precision of model	71%	79%	80%
Recall value of model	71%	79%	81%
F1 value of model	71%	79%	80%

The Accuracy of our developed model with all given 8 features is found to be 78.05%, while with 4 prominent features accuracy is enhanced to 79.22%.

The accuracy of the model is further enhanced to 81.95 % by using IFL with 4 prominent features.

The Precision of the proposed model with all given 8 features is found 71% while with 4 prominent features, precision is enhanced to 79%.The precision of the model is further enhanced to 80 % by using IFL with 4 prominent features.

The recall value of the proposed model with all given 8 features is found 71% while with 4 prominent features, it is 79%.The recall value of the model is further enhanced to 81 % by using IFL with 4 prominent features.

The F1 value of the proposed model with all given 8 features is found 71% while with 4 dominating/prominent features, it is enhanced to 79%. The accuracy of the model is further enhanced to 80 % by using IFL with 4 prominent features.

6. Conclusion

In this paper, we developed an efficient model for early detection of GDM disease. We used a intuitionistic fuzzy decision tree algorithm and Pima Indians Diabetes database taken from Kaggle to develop the model. To enhance the accuracy and other performance metrics of the model we used only those features which are prominent over other features by using an efficient IFS based feature extraction method. These prominent features play a vital role in detecting the disease more accurately. With the reduced features, the medical practitioners get less confused to reach the final clinical conclusion, and machine learning based medical expert systems can be designed with more accuracy. We first converted the chosen diabetes dataset into intuitionistic fuzzy dataset by incorporating suitable hesitation to reduce the range of data which helps in increasing the efficiency of the model. In the future, we plan to extend proposed model for designing an IFS based medical expert system on a large sized dataset by applying IFS based deep learning algorithm.

References

[1] R. M. Anjana, R. Pradeepa, M. Deepa, M. Datta, V. Sudha, R. Unnikrishnan, et al. Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose

tolerance) in urban and rural India: Phase I results of the Indian council of medical research-India Diabetes (ICMR-INDIA) study, Diabetologia 54 (2011), 3022-3027.

[2] P. Barach, V. Levashenko and E. Zaitseva, Fuzzy decision trees in medical decision making support systems, Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 8(1) (2019).

[3] J. Brownlee, What is a confusion matrix in machine learning, Code Algorithm from Scratch, (2020).

[4] Charbuty, Bahzad and Adnan Abdulazeez, Classification based on decision tree algorithm for machine learning, Journal of Applied Science and Technology Trends 2(01) (2021), 20-28.

[5] Cho Han Nam, ET health world, Gestational diabetes is a more serious problem in India than in other parts of the world, (2017).

[6] R. K. Chudasama et al., Magnitude of gestational diabetes mellitus, its influencing factors and diagnostic accuracy of capillary blood testing for its detection at a tertiary care centre, Rajkot, Gujarat, Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive and Social Medicine 44(2) (2019), 142-146.

[7] De, Supriya Kumar, Ranjit Biswas and Akhil Ranjan Roy, An application of intuitionistic fuzzy sets in medical diagnosis, Fuzzy Sets and Systems (2001), 209-213.

[8] Djatna, Taufik, Medria Kusuma Dewi Hardhienata and Anis Fitri Nur Masruriyah, An intuitionistic fuzzy diagnosis analytics for stroke disease, Journal of Big Data 5(1) (2018), 1-14.

[9] Shaker EL-Sappagh, et al., An ontology-based interpretable fuzzy decision support system for diabetes diagnosis, IEEE Access 6 (2018), 37371-37394.

[10] P. Hamsagayathri and P. Sampath, Performance analysis of breast cancer classification using decision tree classifiers, Int. J. Curr. Pharm. Res 9(2) (2017), 19-25.

[11] C. Z. Janikow, Fuzzy decision trees: issues and methods, IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) 28(1) (1998), 1-14.

[12] K. T. Atanassov, Intuitionistic fuzzy sets, Physica-Verlag, Heidelberg, New York, (1999) and K. T. Atanassov, More on intuitionistic fuzzy sets, Fuzzy sets and systems 33(1) (1989), 37-45.

[13] Kim, Kyoungok and Jung-Sik Hong, A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis, Pattern Recognition Letters 98 (2017), 39-45.

[14] Ludwig, A. Simone, Stjepan Picek and Domagoj Jakobovic, Classification of cancer data: analyzing gene expression data using a fuzzy decision tree algorithm, Operations research applications in health care management, Springer, Cham (2018), 327-347.

[15] Lukmanto, Rian Budi, Ariadi Nugroho and Habibullah Akbar, Early detection of diabetes mellitus using feature selection and fuzzy support vector machine, Procedia Computer Science 157 (2019), 46-54.

[16] Maternal Health Division, MHFW, GOI Diagnosis and Management of Gestational Diabetes Mellitus Technical Guidelines, GOI (2008).

[17] Meng, Xiangfei, et al. Construction of decision tree based on C4. 5 algorithm for online voltage stability assessment, International Journal of Electrical Power and Energy Systems 118 (2020), 105793.

[18] Motarwar, Pranav, et al. Cognitive Approach for Heart Disease Prediction using Machine Learning, International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) IEEE, 2020.

[19] M. Nilashi, O. B. Ibrahim, A. Mardani, A. Ahani, A. Jusoh, A soft computing approach for diabetes disease classification, Health Informatics Journal 24(4) (2018), 379-93. PMid: 30376769. https://doi. org/10.1177/1460458216675500

[20] Novaković, Jasmina Dj, et al. Evaluation of classification models in machine learning, Theory and Applications of Mathematics and Computer Science 7(1) (2017), 39-46.

[21] M. S. Padmavathi and C. P. Sumathi, A new method of data preparation for classifying diabetes dataset, Indian Journal of Science and Technology 12 (2019), 22.

[22] G. T. Reddy, et al., An ensemble based machine learning model for diabetic retinopathy classification, International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) IEEE, 2020.

[23] P. E. Rubini, et al., A Cardiovascular disease prediction using machine learning algorithms, Annals of the Romanian Society for Cell Biology (2021), 904-912.

[24] Sachdeva, Kavita, Madasu Hanmandlu and Amioy Kumar, Real life applications of fuzzy decision tree, International Journal of Computer Applications 42(10) (2012), 24-28.

[25] V. Srinivasan, Selection for share market dataset using entropy and information gain, (2019).

[26] Szmidt, Eulalia and Janusz Kacprzyk, Selection of the attributes in intuitionistic fuzzy models, Notes on Intuitionistic Fuzzy Sets 24(4) (2018), 63-71.

[27] Tangirala and Suryakanthi, Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm, Int. J. Adv. Comput. Sci. Appl. 11 (2020), 612-619.

[28] Wu Yan-Ting, et al., The Journal of Clinical Endocrinology and Metabolism Early Prediction of Gestational Diabetes Mellitus in the Chinese Population via Advanced Machine Learning 106(3) (2021), e1191-e1205.