Optimization of Logistic Regression Model by Feature Engineering using Fuzzy Logic

Rohit
Department of Computer Science and Engineering
ABES Engineering College
Ghaziabad, India
rohit.saklan@gmail.com

Vivek Kumar
Department of Computer Science and Engineering
ABES Engineering College
Ghaziabad, India
v.chaudhary1988@gmail.com

***Abstract*— In modern era, the Machine learning especially Supervised Machine Learning is iridescent field where many researchers are shaping their research work. Innovations and new ideas in field are growing fast. Still in various fields like medical, financial and other, we need to enhance the predictions of the Logistic Regression classification models. In this research paper, we focused on the optimization of the supervised Logistic Regression model which can predict more accurately when implemented on clinical dataset. Our purpose is to introduce and implement the fuzzy logic for Features engineering technique to optimize the logistic regression model before training of the Logistic regression, which can predict more accurately in the medical data sets. In this paper, we use Coronary Heart Disease (CHD) for implementing the features engineering using fuzzy logic. Final result showed that Optimized Logistic Regression Model prediction of Ten Years Heart risk is improved with accuracy of .863 when classical Logistic Regression classification model has been compared with it.**

Keywords*— *Machine learning, Fuzzy logic, Logistic regression, Coronary heart disease, Feature engineering.

I. INTRODUCTION

Today the Machine Learning can be defined as subfield of Artificial Intelligence. It is the process in which machine is able to learn by previous experience and perform intelligently like human being.

A. Machine Learning

The Machine learning has been become the milestone for the variety of applications like scientific research, academic activities, statistics, healthcare and other [1]. It can be viewed in three major categories like first is Supervised ML algorithms, Unsupervised ML algorithms and the third categories belongs to reinforcement learning [2].

In this research work we mainly focused on the supervised machine learning algorithms. Supervised Machine Learning or Supervised ML algorithms are the techniques that cause from outwardly completed instances to produce common hypotheses, which then make predictions about future instances [3].

The supervised machine learning further categorised in two different prediction processes like classification and regression. We have chosen classification supervised machine learning for implementation of our proposed methodology [4]. There are many classification algorithms are available in market. In which logistic regression has been chosen for prediction of ten years heart risk form CHD data set.

B. Logistic Regression

The Logistic regression or LR model prediction and analysis is a famous and appropriate method that gives binary response or categorical values as prediction result. This method is generally appropriate for classification models concerning disease state like healthy or diseased, decision making like yes or no. Generally logistic regression is widely used for studies, analysis and prediction in the health care fields [5].

The logistic regression function is defined in Eq. (1)

$$f(x) = \frac{1}{(1-e)^{-x}} \quad (1)$$

where, f(x) is outcome between the 0 and 1, x is input values given to the function f(x). The sigmoid function produces a curve also called the sigmoidal curve with input value x [6].

C. Data Pre-processing and Cleaning

It includes the process of Data cleaning and adding more relevant data to the dataset when some missing has been found. In this context the sample size of the dataset should be increased, likely to be tried to match data with new records. The authors presented several methods of missing handline techniques. Such as Removing the column having missing value, If a field or many fields are having many missing values then deletion of the column is one and maybe a good option but it cannot be the better or less effective choice [7-8]. The authors in [7] state that Mean/Median/Mode Imputation technique is quite easy which replace null values with the mean, median, or mode of that particular column.

D. Fuzzy Logic

In 1965, the fuzzy logic was introduced by author Zadeh, presented fuzzy logic for cellular robotics applications, and still today several investigators are incessantly contributing their research efforts in the field of Fuzzy logic [9]. Fuzzy logic can be stated as multi-valued logic of have truth values of variables in any real number ranges from 0 to 1.

It is a notion of partial truth. A Machine Learning methodology integrated with fuzzy logic has been technologically advanced to extract association or relationships, modelled as rules [10]. Fuzzy Logic is a procedure of logical reasoning that seems like way of thinking and analysis of human being. Like human being, how humans make their decisions on any reasoning, fuzzy logic performs similar to this approach. The fuzzy approach distinguishes range of possibilities between Yes or No [11].

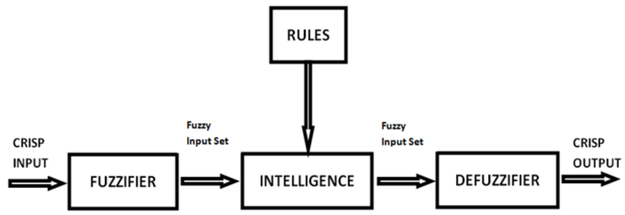

Fig. 1. Boolean logic and Fuzzy logic

In real life the human faces different situations that arise in real life and it cannot be decide whether any statement or any idea is actually in truth condition or not. The fuzzy logic concept came with the idea of offering very treasured elasticity for that kind of reasoning [12]. In Boolean algebraic system when true, absolutely true has value 1 and if false then it will absolutely false with value 0. Where the fuzzy logic performs on concept of that it covers middle value too called partially true or partially false [13].

E. Architecture of Fuzzy Logic

The architecture of the fuzzy logic has four phases like – Rules, Fuzzification, Intelligence and defuzzification (see in Figure 1).

1) Rule base

It is a set of rules having some conditional statements such as IF-THEN. This rule base created and delivered by the user to manage or design new decision making system [14].

2) Fuzzification

It is the process of converting crisp/ hard inputs, exact input taken from the sources into non crispy fuzzy sets [13].

Inference engine or intelligent system- It governs with identifying similar degree of the current fuzzy sets input with respect to each rule of the rule base. Hence it resolves which rules can be fit to accumulate desire control system [15].

3) Defuzzification

It is reconversion of fuzzy sets that are governed by the intelligent system into a crisp input values [16].

Our proposed methodology i.e. fuzzy logic feature engineering has been oriented on a membership function represented as a graph. It shows the input feature relationship or mapping to the values of membership i.e. 0 and 1 [17]. A membership function for a fuzzy set A on the universe of discourse S can be defined using Eq. (2):

$$Set\ A: s = \{0,1\} \quad (2)$$

It quantifies the degree of membership of the element in S to the fuzzy set A. In fuzzy logic we can use no of membership functions that are available to apply [18]. The membership functions for LN (i.e. Large Negative), MN (say Medium Negative), S (say Small), MP (i.e. Medium Positive), LP (say Large Positive) has been calculated for triangular membership function which are not common when compared to other membership function shapes (see Table V.).

This research paper has been organized in different sections. Section one concerned with general introduction of Machine learning (ML) and its types and Fuzzy Logic. Section II represents related work reviews of different researchers in the field of Fuzzy logic and features engineering as state-of-the-art. Section III presented Research methodology and its practical implementation. Section IV depicts the result discussions and conclusion, and in section V represents the future scope.

II. RELATED WORK

Since 1990 many contributors and researchers are continuously surveying and studying, and introducing new algorithms, methodologies to make machine learning more efficiently. From the beginning, the Researchers are providing different techniques and methodologies continuously in ML and Artificial Intelligence [28].

Nowadays, feature selection and engineering is the most challenging task for research contributors. This section depicts the insights of detailed reviews of the most swaying techniques used to optimize the ML model by feature selection and engineering.

The researchers, in [19], compared categorization and classification ML model for classification of 'Ten Years Heart Risk' prediction in coronary heart disease (CHD0 dataset. They used as logistic regression classifier to predict the heart disease also. When they compared the Logistic Regression with other classification algorithms, the logistic regression model performs better. Authors, in [20], did improvements in existing categorization methods to eliminate basic drawbacks. They also did the survey and tests for the possibility heart illness prediction using novel algorithms. They targeted the sensitivity and specificity measures of the confusion matrices of the classifier. They used Logistic regression classifiers that were used to filter out the ultrasonic readings and then developed the classification procedure. They found that their proposed model avoids over-fitting and produced more truthful results.

In [21-22], the contributors surveyed various introduced fuzzy logic applied logistic regression (LR) model where they found the crisp dependent variable and fuzzy set's independent variables. They concluded that their crisp variable outcome was based on odds prediction. They used dataset of Pima India Diabetes which constituted the eight independent variables and one binary dependent variable (fuzzy). They found that their model obtained the parameters of estimation perform better than ordinary logistic regression (LR) model when they compared the sensitivity.

The study of the authors in [23] presented a classification methodology with integration of logistic regression ML algorithm and fuzzy logic for identification sample of feral fish and Nile Tilapia. In their proposed method they statistically tested different ways related to target. Their proposed integrated worked satisfactorily and give better statistical analysis

The researchers, in [24], broadly observed that when using feature selection techniques applied on various Classification ML algorithm, Decision Tree and SVM ML algorithms are most effective for prediction of heart disease risk when using CVD dataset. They showed in their result that SVM predicts with 91.0 %. The authors described fuzzy logic intelligent system that used the swarm optimization tactic in [25]. They took UCI cardiac data sets for their experiments. The authors used decision tree ML algorithm to find the key characteristics useful in prediction accurate diagnosis and health treatment. They found that their proposed fuzzy expert system gave precision of 93.27 %. The authors, in [26], examined the statistical approach of Logistic Regression in deep to prove the usability of the logistic regression classifier for analysis. They declared that Logistic regression is a statistical analysis technique for proliferated data, which worked on binary dependent variables. The authors, in [27], suggested an optimization algorithms oriented on the classification ML algorithms. They surveyed, studied and conclude that if they used the roughest fuzzy learning techniques for treatment and prognosis of heart disease, the outcomes of two methodologies like SVM and ANN performs better.

In paper [28], the contributors surveyed and did deep study about the novel approaches to ML methodology previously applied used that proven for better prediction. As per their bottomless study they conclude and suggested that Logistic Regression (LR) classification methodology and SVM techniques would be the game changer for optimization binary classification. The author targeted, in paper [29], that they studied and introduced an incorporated fuzzified logistic regression (FLR) method for features engineering by analysis of crisp independent and fuzzified dependent output. They presented the numerical implementation using data set and found the results. They compared the results with classic logistic regression classified, performance was improved.

In [30], the researchers estimated the permeability with low porosity. They studied and concluded that Fuzzy logic approach is extensively used in approximation of permeability. The authors applied fuzzy logic feature engineering using estimation of correlation method. The correlation coefficients has been investigated which reached 76% to 70%, improved the accuracy of the model. In [8], the target of researchers is survey as state of the art for the methodologies applied to the ML algorithm optimization that performed better than old model. Their experimental result and compared the efficiency of various ML algorithms and lastly they found that SVM worked with accuracy of 86.8%. In [31], the authors applied feature selection and classification algorithms. They suggested a selection algorithm that was implemented on Cleveland heart disease dataset. Their result shows that K-Nearest Neighbor classification model gave accuracy of 0.90 after using feature selection and engineering.

Researchers, in this paper [32], proposed a clinical support system to predict and diagnose heart diseases. They implemented Supervised ML algorithms to predict heart disease. They did several experiments; their findings showed that Naïve Bayes performed better and predict disease with accuracy of 84.28 from 82.17. The author, in [33], compared best algorithms in a collection of ML algorithms. They presented a brief conclusion on soft computing-based techniques and machine learning algorithms which helped in prediction and categorization of cardiac disease. Their findings impacted on the medical field that will increase the classifier's accuracy of the data mining techniques such as expectation maximization and other rule-based inference.

In [34], the authors proposed an optimization function integrated with SVM. They select the relevant important characters to predict heart disease with the help of objective function. They used genetic algorithm which offer an efficient feature engineering process. The prediction of the proposed model has increased the accuracy to 88.34%. In [35], the author’s main aim is to predict the cardiac disease. In this paper the authors suggested a three-layered binary classifier worked on neural networks techniques. This process has been used for to create feature space. They used Cleveland UCI heart dataset. Their surveys made use of many data features engineering approaches to enhance the accuracy which has been increased to a .92 from .88. In [36], this research, authors proposed a deep learning methodology for heart disease diagnose by using Multiple Kernel Learning. They used enhanced Neuro-Fuzzy Inference System. Their proposed model yields 80% of prediction accuracy when applied on UCI heart disease dataset.

In [37], the authors compared seven different feature selection methods using different classifier models. They also fixed their domain as knowledge-based feature selection method. After comparison they proposed probabilistic ensemble Feature Selection method. Then performance metrics have been compared, evaluated and resulted that the accuracy has been improved as 91.77%, 85.54%, and 85.45% for Z-Alizadeh Sani, Statlog, and Cleveland data sets. In [38], authors proposed and implemented a prediction classifier for CVSS susceptibility classification of ICS. Authors suggested many methods for analysis of dataset like statistical regression, cluster. In this paper they focused on employment of fuzzified logistic regression model and declared as one potential alternative technique. In [39], paper, the authors proposed a new Intuitionist Fuzzy LR (Logistic Regression) classifier which dealt with the inaccurate parameters having ambiguity and indecision degree. [40]. They applied this model on the birth-weight dataset, and observe in their findings that the intuitionist fuzzy logistic regression values indicate the better fit and may measure for the best fitting. Their proposed model can perform better even when the dataset is vague. Their proposed model can be employed on other similar predictions space. In [41-42], contributor proposed a novel fuzzified Logistic Regression Model by considering the correlation attribute, namely oriented fuzzy rough sets. They studied and analyzed attribute focused fuzzy rough and achieved optimistic and pessimistic estimation fuzzfied sets and logistic regression.

III. Proposed Methodology

In this paper, we have optimized the logistic regression model using proposed Fuzzy Logic Feature Engineering Techniques for prediction of Ten Years Heart Risk in CHD data set. The Python code has been used for practical implementation of our proposed methodology, which has been executed on the Anaconda Python Environment using Jupyter Notebook and Framingham Coronary Heart Disease Data set.

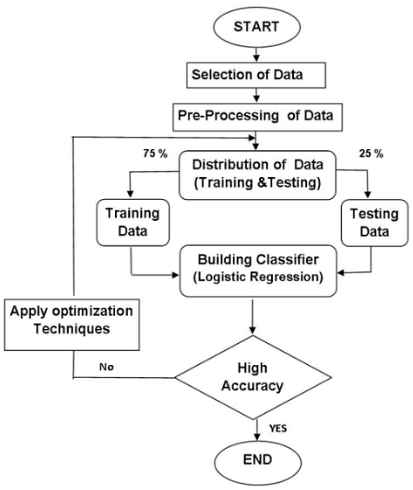

Fig. 2. Flow chart for proposed methodology

We proposed and implement our research methodology as given in flowchart in figure 2 which generally includes three processes such as

- Pre-processing of the data
- Fuzzy logic feature engineering
- Building and evaluation of Logistic regression model classifier.

A. *Pre-processing of Data*

For pre-processing of data, the CHD dataset has been selected in which there are 4238 records and 16 features. The data set contains 15 features {'male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds', 'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP', 'diaBP', 'BMI', 'heartRate', 'glucose'} and one dependent or target feature {'TenYearCHD'}.

B. *Handling Missing Data*

Some features have missing values in the CHD data set given in Table I.

TABLE I. Missing values in following features in CHD dataset

Features	Missing Values
Education	105
cigsPerDay	29
BPMeds	53
totChol	50
BMI	19
Glucose	388

When we used the random imputation logistic regression model prediction accuracy has been improved from 84.7% to 84.8%. So finally we decided to use Random Imputation Technique to handle missing values. We use random imputation techniques to handle the missing values for the CHD data which was proven in our research in paper [7] as table II.

TABLE II. Comparison of Performance measures using missing handling Techniques

Performance Parameters	Mean Imputation	Median Imputation	Random Imputation
Accuracy	0.846	0.845	0.847
F-Score	0.085	0.0751	0.094
Recall score	0.045	0.04	0.05
Precision	0.642	0.571	0.667

C. *Fuzzy Logic Implementation*

We entrenched the fuzzy logic feature engineering algorithms for optimization of the logistic regression model by using the following process.

Step - 1: After pre-processing of the data, the continuous features ('cigsPerDay' ,'BPMeds','totChol', 'sysBP' ,'diaBP' , 'BMI' , 'glucose') have been chosen for implementation of Fuzzy logic algorithm, and the code for the algorithm has been executed.

Step - 2: Apply the fuzzy logic feature engineering algorithm (see Figure 3) for fuzzification of the continuous features to find the min, max and mid value, membership grade and the degree presented in output by executing the python code.

```
Step-1 class Fuzzify(object):
    def __init__(self, data_series, Min, Max, Mid):
    Create a triangular membership function and its max, min and median
Step-2 Set  self.data = data_series
    self.low = Min,  self.high = Max and self.mid = Mid
    self.universe = np.arange(np.floor(self.low), np.ceil(self.high)+0.2, 0.1)
    self.trimf_low  = fuzz.trimf(self.universe, [self.low, self.low, self.mid])
    self.trimf_mid = fuzz.trimf(self.universe, [self.low, self.mid, self.high+0.1])
    self.trimf_hi =  fuzz.trimf(self.universe, [self.mid, self.high+0.1, self.high+0.1]
Step-3 def get_universe(self):
    return self.universe
    def get_membership(self):
    Assign fuzzy membership to each observation and return a dataframe
    Set new_df = pd.DataFrame(self.data)
    new_df['low'] = fuzz.interp_membership(self.universe, self.trimf_low, self.data)
    new_df['mid'] = fuzz.interp_membership(self.universe, self.trimf_mid, self.data)
    new_df['high'] = fuzz.interp_membership(self.universe, self.trimf_hi, self.data)
    new_df['membership'] = new_df.loc[:, ['low', 'mid', 'high']].idxmax(axis = 1)
    new_df['degree'] = new_df.loc[:, ['low', 'mid', 'high']].max(axis = 1)
    return new_df
Step-4 end
```

Fig. 3. Algorithm for Formation of Class for Fuzzy Logic Feature Engineering

Step - 3: We defined the class 'fuzzify', which takes the data values from the vector, and assigns the estimated grade of relation to fuzzy sets. In definition of the class the fuzzy.trim() used to make the triangular membership function with data series and minimum, maximum and the median. The fuzzy.interp_membership() function create the membership grade to fuzzy sets.

Step - 4: In continuation, now when membership and the degree of the features individually has been calculated for every values of the calculated , we replace the features actual values by the degree of that particular instance as algorithm is given in Figure 4. The output after the execution of the code has been given in Figure 5, we can see that how the membership degree is replaced in particular feature which is near to binary classification.

Step-1 *Copy feature values to new attribute*

$X_{new} = df['X_i']$

Step-2 *Set* $X_{new}\_min = X_{new}.min()$

$X_{new}\_max = X_{new}.max()$

$X_{new}\_mid = np.median(np.arange(X_{new}\_min, X_{new}\_max, 0.1))$

Step-3 *Finds the object of class Fuzzify*

$X_{new}\_object = Fuzzify\ (X_{new}, X_{new}\_min, X_{new}\_max, X_{new}\_mid)$

Step-4 *Copy the membership to new variable*

$fuzzified\_X_{new} = X_{new}\_object.get\_membership()$

Step-5 *Replace the fuzzified degree to actual features*

$df['X_i'] = fuzzified\_X_{new}['degree']$

Step-6 *End*

Fig. 4. Finding the degree and replacing the attribute value by degree

	cigsPerDay	low	mid	high	membership	degree
0	0.0	1.000000	0.000000	0.000000	low	1.000000
1	0.0	1.000000	0.000000	0.000000	low	1.000000
2	20.0	0.427754	0.572246	0.000000	mid	0.572246
3	30.0	0.141631	0.858369	0.000000	mid	0.858369
4	23.0	0.341917	0.658083	0.000000	mid	0.658083
...	...	...	...	...	...	...
4233	1.0	0.971388	0.028612	0.000000	low	0.971388
4234	43.0	0.000000	0.770982	0.229018	mid	0.770982
4235	20.0	0.427754	0.572246	0.000000	mid	0.572246
4236	15.0	0.570815	0.429185	0.000000	low	0.570815
4237	0.0	1.000000	0.000000	0.000000	low	1.000000

4238 rows × 6 columns

Fig. 5. Output of the fuzzification of the feature 'CigsPerDay'

D. Building and Evaluation of Logistic Regression Model Classifier

After the application and implementation of fuzzy logic feature engineering we split and distribute the feature into X (independent variables) and y (dependent vcariable) with 70% trained model and 30% test model for logistic regression classification.

We fitted the trained test model in both classical Logistic Regression and the Gradient Descent Theta parameter optimized Logistic Regression Model for comparison of the prediction accuracy. The outcome (prediction of Ten years heart risk of both classical and optimized logistic regression model) has been find out as given in table III after execution of the code.

Finally after optimization of Logistic Regression classical model we got better accuracy of prediction (from 84.7 % to 86.3%) using Fuzzy Logic Feature Engineering.

TABLE III. RESULTS OF CLASSICAL AND OPTIMIZED LOGISTIC REGRESSION

Performance Measures	Classical LR	LR using FLFE
Confusion matrix	[1068 5] [189 10]	[1085 10] [164 13]
Precision	0.667	0.570
Accuracy	0.847	0.863
F-score	0.094	0.130
Recall-Score	0.050	0.073

IV. CONCLUSION

As per result table (Table III), it can be observed that before applying Fuzzy Logic Features Engineering, the logistic regression model is predicting with the accuracy of .847 and after applying Fuzzy Logic Features Engineering techniques, the Logistic Regression model has been optimized and improved the accuracy to 0.863.

We focused on the optimization of the supervised Logistic Regression model which can predict more accurately when we apply our proposed methodology. The aim was to introduce and implement the fuzzy logic for Features Engineering of the medical data sets to optimize the Logistic Regression Model. Hence our research methodology is proven for optimization for better prediction accuracy then classical LR Model.

Considering the global situation, the data is being generated in exponential form but it is not being utilized properly, having challenging to handle and prediction knowledge from the data set. This work will be the landmark for researchers to enhance the prediction accuracy in the field of supervised machine learning and the fuzzy logic.

REFERENCES

[1] Nemati, M., Ansary, J., and Nemati, N.: Machine-Learning Approaches in COVID-19 Survival Analysis and Discharge-Time Likelihood Prediction Using Clinical Data. Patterns 1, 1-8, (2020).

[2] Rohit, M and Chaurasia, A.: Selection of Classification and Regression Algorithms for Knowledge Discovery –A Review. 2022-Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 144-149.Solan, Himachal Pradesh, India, (2022).

[3] A.V. Sharmila: Web Based Technologies to Support High Level Process Maturity, BIJIT-BVICAM's International Journal of Information Technology, July-December, 5(2), (2013).

[4] Vijay, N. Kalbande, D. Handa, C.C.: Developing A Model To Predict Employability Of Engineering Students In Campus Placement For IT Sector. IJAREST 2(6) (2015).

[5] Rohit, Chaurasia, A.: A Literature Review of Ensemble of Supervised Machine Learning Methods and Algorithms. Journal of Optoelectronics Laser. 41(6), 81–87(2022).

[6] Park, Hyeoun-Ae: An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. J Korean Academy of Nursing 43(2), 154-64 (2013).

[7] Rohit and Chaurasia, A.: An adaptive knowledge discovery framework using optimal techniques in Logistic Regression machine learning. NeuroQuantology 20(6), 4034-4044 (2022).

[8] Rohit, Gupta, B. and Gola, K. K.: Analysis of Machine Learning for Processing Big Data in High Performance Computing: A Review. EAI Endorsed Transactions on Cloud Systems 6 (1) , 1-12 (2020).

[9] Hani, M., Haresh, R., Daniel, L. and Arpita, C.: On Improving the Performance of Logistic Regression Analysis Via Extreme Ranking. Journal Of Applied Statistics 4(44) 753–766 (2017).

[10] Dhilsath, F., Mohammed, M., Justin, S., Savari, R., Raja, S.: Regression Imputation and Optimized Gaussian Naïve Bayes Algorithm for an Enhanced Diabetes Mellitus Prediction Model. Brazilian Archives of Biology and Technology 64 (2021).

[11] Liu, Z., Zheng, Z. & Li, Y.: Enhancing fault-tolerant ability of a nine-phase induction motor drive system using fuzzy logic current controllers. IEEE Trans. Energy Convers 32, 759–769 (2017).

[12] Çeven, S., Albayrak, A., and Bayır, R.: Real-time range estimation in electric vehicles using fuzzy logic classifier. Computer Electric Engineer 83, 106577 (2020).

[13] AP Gopi, RNS Jyothi, VL Narayana: KS Sandeep Classification of tweets data based on polarity using improved RBF kernel of SVM, International Journal of Information Technology, 2023.

[14] Tanaka, H., Uejima, S., Asai, K.: Linear Regression Analysis with Fuzzy Model, IEEE Transactions on Systems, Man, and Cybernetics, 12, 903-907 (1982).

[15] Saw, Montuet et al.: Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning. Conference: 2020 International Conference on Computer Communication and Informatics (ICCCI), 1-6(2020).

[16] Zadeh, L.A., "In Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh" World Scientific: Singapore, pp. 394–432, 1996.

[17] Stephan, T., Sharma, K.A., Shankar, S., Punitha, V., Varadarajan, and Liu, P., "Fuzzy-logic-inspired zone-based clustering algorithm for wire-less sensor networks," International Journal of Fuzzy Systems, 23, pp. 1–12, 2020.

[18] Saeedeh P., et al.: Fuzzy logistic regression based on the least squares approach with application in clinical studies,Computers & Mathematics with Applications. 62(9), 3353-3365(2011).

[19] Kurt, and Imran, U. M.: Comparing performances of logistic regression, classification and regression trees and artificial neural networks for predicting albuminuria in type-2 diabetes mellitus. Journal of Expert System with Applications 31(10), 173–187 (2008).

[20] Pencina, M. J., and Derpan,:Heart disease diagnosis using machine learning approach. Journal of NIH public access 119(24), 3078–3084 (2010).

[21] Pourahmad, S. Ayatollahi, and S.M. Taheri,: Fuzzy Logistic Regression: A New Possibilistic Model and Its Application In Clinical Vague Status. Iranian Journal of Fuzzy Systems, 8, 1-17 (2011a)

[22] S. Pourahmad and S.M.T. Ayatollahi and S.M. Taheri and Z.H. Agahi,: Fuzzy logistic regression based on the least squares approach with application in clinical studies. Computers & Mathematics with Applications, 62, 3353-3365 (2011b).

[23] Hauser-Davis, and et al,: Logistic regression and fuzzy logic as a classification method for feral fish sampling sites. Environ Ecol Stat 19, 473–483 (2012).

[24] Ahmed, A. and et al.: Data mining techniques to find out heart diseases: an overview. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 1(4), pp.18-23 (2012).

[25] Muthukaruppan, J.:A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. Journal of Expert Systems with Applications 9(4), 11657–11665(2012).

[26] Hssina, Merbouha, A. and Ezzikouri.: A comparative study of decision tree ID3 and C4.5'. International Journal of Advanced Computer Science and Applications 4(2), 13–19(2014).

[27] Long, Nguyen Cong, M. H.: A highly accurate firefly based algorithm for heart disease prediction. Journal of Expert Systems with Applications 42(21), 8221–8231(2015).

[28] Rohit, B. Gupta, R. Kumar and A. Kumar.:Towards Information Discovery on Large Scale Data: state-of-the-art. International Conference on Soft-computing and Network Security (ICSNS),pp. 1-9, Coimbatore, India, (2018).

[29] Yapıcı Pehlivan, Nimet, Yonar, Aynur.: An Integrated approach for fuzzy logistic regression(2018).

[30] Wang, X., and. et al. Improved permeability prediction based on the feature engineering of petrophysics and fuzzy logic analysis in low porosity–permeability reservoir. J Petrol Explor Prod Technol 9, 869–887 (2019).

[31] M. A. Nabaouia Louridi, Bouabid El Ouahidi.: Identification of Cardiovascular Diseases Using Machine Learning," presented at the 2019 7th Mediterranean Congress of Telecommunications (CMT) (2019).

[32] J. L. Amin ul Haq, and et al.: Heart Disease Prediction System Using Model of Machine Learning and Sequential Backward Selection Algorithm for Features Selection. 2019 IEEE 5th International Conference for Convergence in Technology (I2CT)(2019).

[33] S. B. Halima El Hamdaoui, Nour El Houda Chaoui, Mustapha Maaroufi.: A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)(2020).

[34] Sajja, Guna Sekhar.: A Comprehensive Review of Various Machine Learning Techniques for Heart Disease Prediction. International Journal of Computer Applications 183(37) 53-56 (2021).

[35] Dzik-Walczak A. and Heba M., An implementation of ensemble methods, logistic regression, and neural network for default prediction in Peer-to-Peer lending[J], ZbornikradovaEkonomskogfakulteta u Rijeci: casopis za ekonomskuteorijuipraksu 39(1) (2021), 163–197.

[36] C. B. Gokulnath, S. P. Shantharajah.: An optimized feature selection based on genetic approach and support vector machine for heart disease," Cluster Computing, vol. 22, no. S6, pp. 14777-14787 (2018).

[37] Burak Kolukisa, Burcu Bakir-Gungor, Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis, Computer Standards & Interfaces,Volume 84, 2023, 103706,

[38] Ahmet Murat Dere, Mehmet Kabak, An application of fuzzy logistic regression for predicting CVSS severity category of industrial control systems", Volume 56(6) pp. 4083 – 4111 (2022).

[39] Abdullah Ali H. Ahmadini.: A novel technique for parameter estimation in intuitionistic fuzzy logistic regression model, Ain Shams Engineering Journal, 13(1) 101-518 (2022).

[40] Saru D. and Swati S.: Crime and Criminal Tracking Networks & Systems Using Agile Methodology, BIJIT - BVICAM's International Journal of Information Technology Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA), 8 (1), pp 973 –5658 (2016).

[41] Yu, Bin et al.: A Twin Logistic Regression Method Based on Attribute-oriented Fuzzy Rough Set. 9581 – 9597 (2023).

[42] Sijin, P., Champa, H. N.:Fuzzy conceptualization of the search queries, International Journal of Information Technology, BIJIT, pp. 2511-2112, (2023)