Advances in Dynamical Systems and Applications (ADSA).
ISSN 0973-5321, Volume 16, Number 2, (2021) pp. 987-1005

https://www.ripublication.com/adsa.htm

Cardi Vascular Heart Disease Classification Using Machine Learning and Neuro-Fuzzy Models

[1]P.Bharath Kumar Chowdary, [2]Dr. R. Udaya Kumar

Department of Computer Science and Engineering,

[1]Reseach Scholar, Department of Computer science Engineering, BIST, Bharath Institute of Higher Education and Research (BIHER).

[2]Research Supervisor, Professor, Department of Information Technology, BIST, Bharath Institute of Higher Education and research (BIHER).

Abstract

Heart diseases are increasing day by day in the present world due to stress and work pressures. The early stages before severe heart diseases are obesity and diabetes. It is very challenging to predict heart diseases at the early stages. One of the common diseases that occur to heart patient is diabetes. Early detection and cure of diabetes can help to reduce the risk of heart failures. In this paper, we focus to work on early detection and prediction of heart failure with special attention to diabetes. The clinical data most of the times is incomplete and uncertain. In this paper we clean the data using a well formulated data cleaning procedures and apply prediction algorithms to evaluate the classification and moreover we designed a neuro fuzzy model also to predict the heart diseases. The proposed model using neuro fuzzy techniques with the mentioned data cleaning procedures in this work proved effective while evaluating the accuracy. The main motive of this research is to integrate the clinical data of medica domain with machine learning approaches to predict the heart diseases at early stages. The proposed model and various existing machine learning algorithms are tested, and the results are reported in the paper. It is observed from the results that cleaning of data improved the results of all the machine learning algorithms and the proposed model. The state-of-art models are also compared in this work and the results of the proposed model outperformed the existing models.

Keywords: Heart diseases, Diabetes, Feature extraction, Neuro-fuzzy model, Artificial Neural Network

1. INTRODUCTION

Research over the recent years has revealed that cardiovascular morbidity and mortality in patients with diabetes [1], [2] is a major contributor to heart failure. In some diabetic patients, the observation that in the absence of coronary artery disease, valvular disease, and related cardiovascular risk factors, myocardial dysfunction has led to the use of the misunderstood term "diabetes cardiomyopathy." It was first used in 1972 by Rubler et al. [9], where it explains myocardial dysfunction in the absence of coronary artery, hypertrophy or cardiac disorders in patients with diabetes.

Diabetes Mellitus Type 2 is the world's largest cause of disturbance of health of an individual or death of the individual, increasing the risk of deteriorating health problems such as coronary heart disease, stroke, chronic renal disease, and lower-end amputation. As a result of the high expenses of medical treatment and lost productivity, community health initiatives are now focusing on the prevention or early diagnosis of type 2 diabetes in individuals.

While the increased risk of these negative consequences associated with type 2 diabetes [3] is well-known, it is frequently assumed that the risk is the same regardless of age, gender, or race. However, research suggests that this isn't the case. Men and women with type 2 diabetes, for example, have distinct associations with coronary heart disease and stroke.

Recent study [4] has discovered that risk factors such as body weight, blood pressure, and cholesterol levels differ between those with and without type 2 diabetes based on demographic groups. It's possible that there's heterogeneity in the risk of these negative consequences based on age, gender, and ethnicity. However, there has yet to be a thorough examination of whether such heterogeneity occurs.

If population health managers had a systematic understanding of whether the additional hazards varied by age, gender, and ethnicity, they would be able to create interventions to avoid or identify these unfavourable events early among demographic groups at higher risk. This is especially relevant in Asia, where diabetes affects 60% of the population.

Type 2 diabetes recognition models [5], [6] require appropriate knowledge to efficiently detect the disease and thus help physicians to diagnose the disease early. Early diagnosis of the disease helps to overcome the impact that diabetes causes on various organs in the human body like kidneys, heart, eyes, etc.

The disease prediction models [7], [9] that handle the diabetes data often face issues like noisy, missing, irrelevant and inconsistent data. The performance of the model depends on the quality of the diabetes data presented to the model and hence the researcher must supplement accurate data to the classifier for effective disease prediction. In the machine learning domain, classification is an important task as it derives knowledge to handle real-world applications. Classification constructs a model to predict the target class of the data accurately.

In this paper the data involved with diabetic information and heart diseases are thoroughly examined and the data [8] is classified using machine learning algorithms

to predict the heart related diseases of the diabetic patients. Most classification models reduce the features at the pre-processing stage. This leads to loss of information and prediction capability will be affected if the features are reduced in the training process. Hence we examined all the features of the data for prediction.

Predicting data with machine learning algorithms though a better approach still there is a scope of improving the accuracies using neural techniques. Hence this paper also proposes a neuro fuzzy model to predict the heart failure involved with diabetic patients more effectively. The

The rest of the paper is organized as follows: the second section covers the existing work in this domain. Section 3 presents the proposed approach for heart failure for diabetic patients using classification approaches with a special attention to feature extraction. Section 4 presents the results of the proposed approach and the state-of-art classification approaches. Section 5 concludes the paper with a direction towards possible future aspects.

2. RELATED WORK

This section covers the recent work in the domain of heart diseases prediction using diabetes. The works discussed in this section gives an idea of the recent works going on in this domain.

Michael Lehrke, et al [10] provided the review of heart failure works. The authors discussed about various factors causing heart failures and also mentioned lifestyle changes that can help patients to recover or not to get affected from heart failures.

Ka Keat Lim, et al [11] provided a study for the inspecting the risks that persons with diabetes may get like heart strokes, kidney failures and other chronic diseases. This study is mainly focused on the various location-based groups of Asian population. This study examined all the persons related information to see the effects of diabetes on the chronic diseases. This work only studies the people using clinical data and gives only the location-based study.

The authors [12] proposed a model using Fuzzy Analysis to design a support system that takes clinical decisions to estimate a person that can develop heart disease. The model is developed to predict and help the clinical diagnosis experts to recommend the patients with few caring aspects that are likely to develop heart diseases. This study dealt with only 100 patients and to recommend clinical specialists that a patient may get heart failures.

The authors [13] collected a timeline data of patients suffering with heart failures along with type 2 diabetes and kidney diseases. The research made was with timeline data of the patients in UK. This work is also an observational study that examined various diseases and effects that are linked with heart failures.

The authors [14] presented a IoT based data mining approach to deal with the diabetes and cardiovascular diseases. The work tries to integrate the biosensors and granular computing aspects with chat boxes and built a hybrid framework. This work has not

provided any experimental results.

The authors [15] analyzed various aspects of diabetes related complications. This work has given a deeper analysis of more than 50,000 patients.

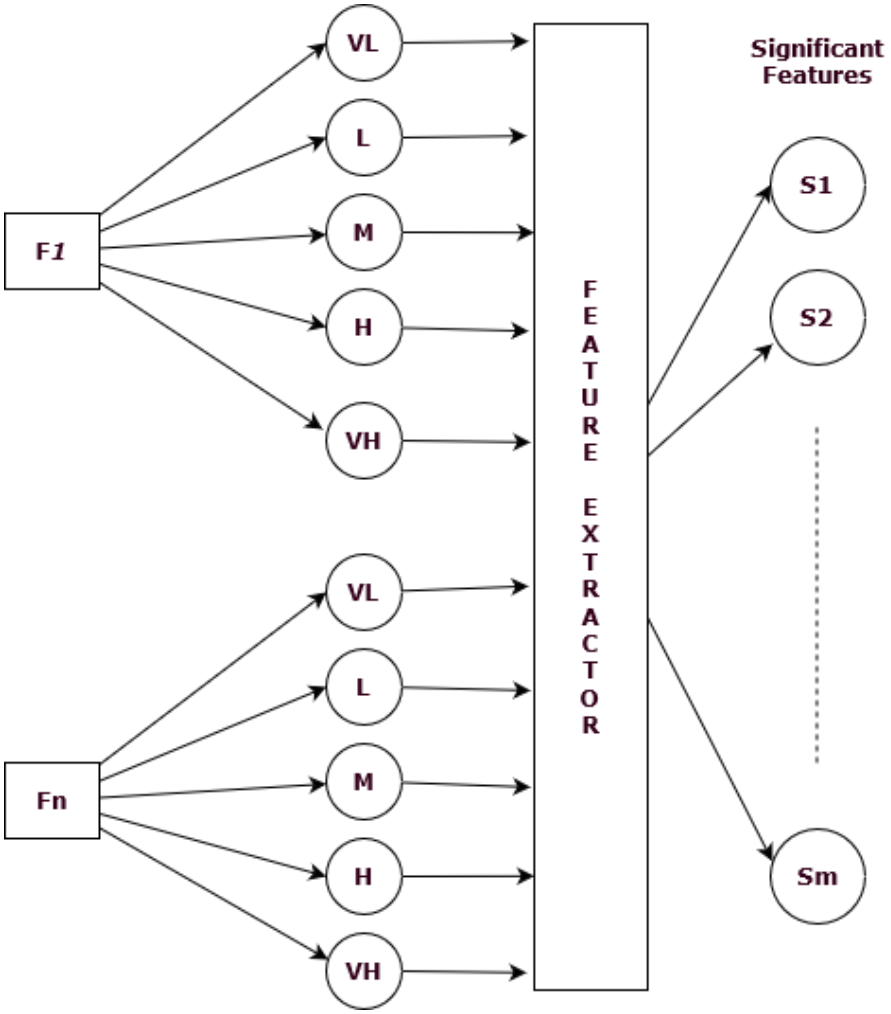

Figure 1: Feature extraction of the proposed model

The works presented in this section proves that there is a lot of research happening in the domain of clinical data. Moreover, the findings that the researchers presented in this domain of diabetic related diseases needs an extensive research. The existing works lacks with a) classification and prediction analysis of diabetic and heart failures, b) the works are either limited to a location or a very few records and c) machine learning and neural networks are not properly explored on diabetic and heart related data.

The presented work in the next section discusses about the application of machine learning and neuro fuzzy model applications to the diabetic and heart diseases.

3. PROPOSED SYSTEM TO PREDCIT HEART DISEASES WITH SPECIAL ATTENTION TOWARDS DIABETES

In this section the proposed model is defined to predict the heart failures of the clinical data using fuzzy membership values.

A. Assigning fuzzy membership values to the features in the dataset using fuzzification and performing classification of data

The initial phase is to perform the fuzzification process on the data. Each entry in the dataset consists of many features. In the first phase, we convert the feature into a linguistic term like VL, L, M, H and VH.

If dataset is having 'P' records with each 'n' features per record,

$$P_i = [f_{i1}, f_{i2}, \ldots\ldots f_{in}] \qquad (1)$$

Where 'i' represents the i^{th} record in the dataset and f_{ij} represents the j^{th} feature of the record.

For each feature in the record membership values are assigned using the π-type function. The π-type assigns fuzzified values based on the five linguistic terms. Hence, the feature vector contains 5*n fuzzified features, if there are 'n' features per record.

$$f_{ij} = [\mu_{VL}(f_{ij}), \mu_{L}(f_{ij}), \mu_{M}(f_{ij}), \mu_{H}(f_{ij}), \mu_{VH}(f_{ij})] \qquad (2)$$

The data is transformed into fuzzy membership values initially. Once each feature is assigned with five membership values, the size of the feature vector grows significantly which increases complexity. To overcome this problem we used principal component analysis to store all the significant features and discard the unnecessary features.

Moreover, the data in the dataset has many missing values. We have used imputation techniques to fill the missing data and after the fuzzification process, all the significant features are restored. The procedure of the proposed feature extraction methodology is shown in Figure 1.

There are two phases in the proposed approach. Figure 1 shows the working principle of the feature extraction procedure followed in this paper. The first phase performs two tasks, a) fuzzification of data and b) feature extraction. The second phase handles the neural network aspects of the model.

After the feature extraction model, a suitable neural network is built to perform classification of data. The process of the proposed approach is shown in Figure 2. The input to the ANN are membership values of the feature vector. The initial weights of the network are in the range (0, 1). The input layer has as many nodes equivalent to the number of features after feature reduction. The output layer has nodes equivalent to the number of the classes in the dataset

The output of the feature extraction layer is given to the hidden layer and the result of the hidden layer is given to the output layer.

The overall dataflow of the proposed model is shown in Figure 2.

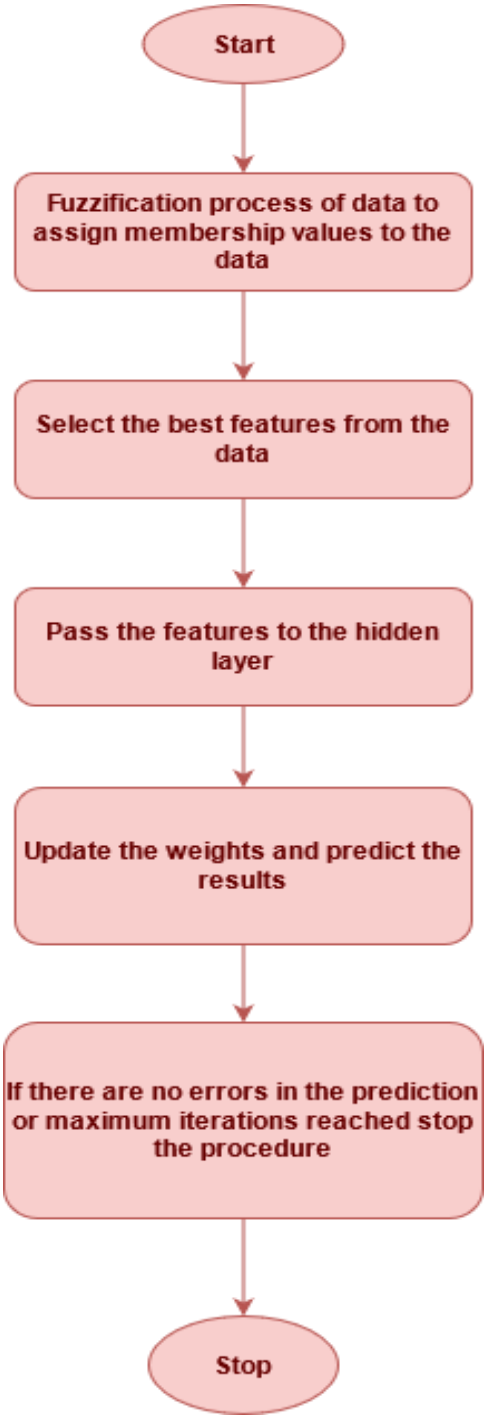

Figure 2: Data flow of the proposed fuzzy neuro model.

The proposed model architecture after feature extraction is depicted in Figure 3.

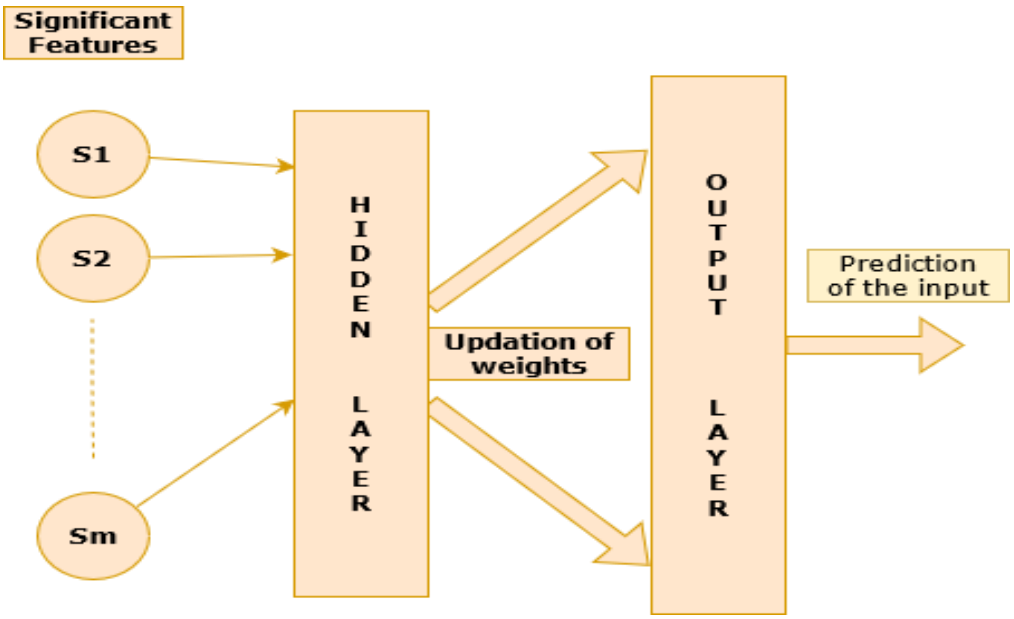

Figure 3: Prediction of class using hidden and output layers of the significant features

Figure 3 gives the second phase in neuro fuzzification process. In this phase the proposed model identifies the significant features from the input. The significant features are supplied to the hidden layer and the hidden layer passes the results to the output layer and if there are any errors in the classification the weights are updated till the maximum iterations are reached or the classification results are correct.

B. Applying Machine Learning approaches on the datasets

In this work we also considered to apply the machine learning approaches to be applied on the data. In order to apply the machine learning algorithms, we have considered the best features of the dataset and then supplied to the machine learning algorithms like SVM, naïve Bayes, Decision trees, Random Forest Classifier, K-nearest neighbourhood algorithms.

The procedure of using these algorithms is mentioned in Figure 4. In Figure 4, initially we will select and load the dataset later the dataset will be pre-processed. We have selected optimum data pre-processing mechanisms to do the data filling of missing values.

The most correlated features with the target and the dataset are split into test and train sets. The next step is to select the appropriate model and evaluate the model using various parameters such as precision, recall and support etc.

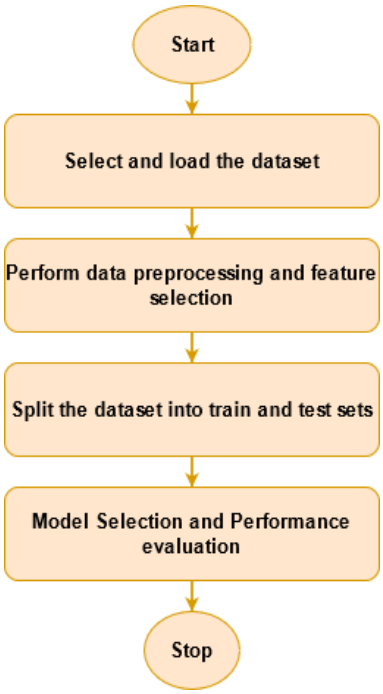

Figure 4: Prediction of class using machine learning algorithms

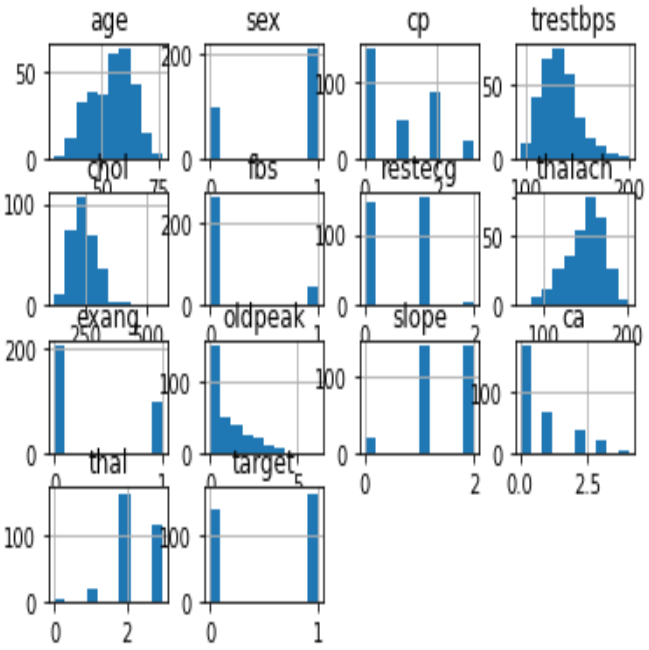

Figure 5: Attributes of Heart disease dataset

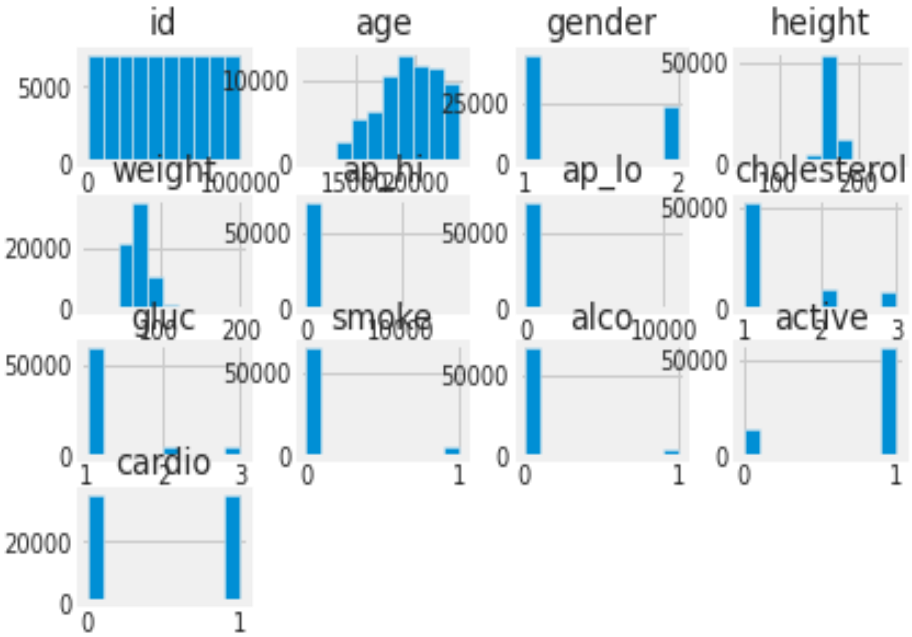

Figure 6: Attributes of cardiovascular disease dataset

In this section, we have proposed a neuro fuzzy model to predict the heart diseases. We have used few machine learning algorithms to compare the results with the results of the proposed model. The next section covers the dataset information and the results obtained on the dataset using the proposed model and the machine learning algorithms.

4. EXPERIMENTATION AND RESULTS

The following datasets are used in this paper for evaluation.

a. Heart Disease Data Set: This dataset consists of 303 records with 75 attributes, but out of all the 75 attributes, 14 attributes are recommended by the researchers. The attributes of this dataset are clearly shown in Figure 5.

b. Cardiovascular-disease-dataset: This dataset has 70,000 records and has 13 attributes. The different attributes of this dataset are shown in Figure 6.

We have performed an extensive experimentation on all the datasets and reported the results in this section.

At the initial stage we have performed analysis on Heart disease dataset. The proposed model was tested with all the machine learning algorithms. We have tested the Heart disease dataset on the following machine learning algorithms

A. Linear Regression (LR)

B. Linear discriminant analysis (LDA)

C. K-nearest neighborbood algorithm (KNN)

D. CART

E. Naïve Bayes algorithm (NB)

F. Enhanced Naïve Bayes (ENB) classification Algorithm [16]

G. Support Vector Machine (SVM)

The pictorial representation of the results is shown in Figure 7.

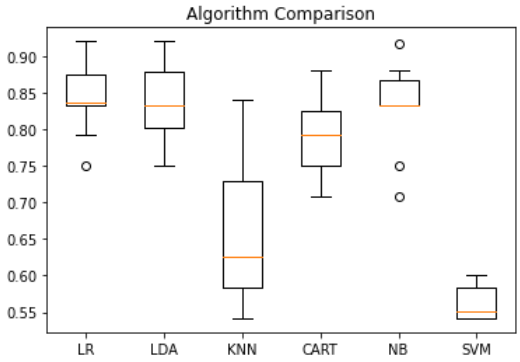

Figure 7: Performance analysis of various machine learning algorithms on the heart disease dataset

The results of the accuracy of the proposed model and various machine learning algorithms are reported in Table 1.

TABLE 1: RESULTS OF ACCURACY ON VARIOUS MACHINE LEARNING ALGORITHMS ON HEART DISEASE DATASET WHEN THE TRAIN AND TEST SET ARE IN THE RATIO 80:20

Model	Accuracy
Linear Regression	0.84
Linear Discriminant Analysis	0.84
K nearest neighborhood	0.65
CART	0.78
Naïve Bayes	0.83
Enhanced Naïve Bayes	0.85
Support Vector Machine	0.56
Neural Networks	0.88
Proposed model	0.89

The proposed model and the other machine learning algorithm results of the Heart disease dataset are presented in the Table 1. In the dataset there is a feature called diabetic and based on the diabetes we have framed the model.

The results of all the machine learning are in the range of 0.85 mostly. Enhanced naïve bayes classifier proposed in the work [16] is considered to perform the experimentation.

The results shows that the proposed model outperformed all the machine learning algorithms taken for comparison.

The results shown in Table 1 are obtained when the test and train tests are split in the ratio of 0.8 and 0.2 respectively.

TABLE 2: RESULTS OF ACCURACY ON VARIOUS MACHINE LEARNING ALGORITHMS ON HEART DISEASE DATASET WHEN THE TRAIN AND TEST SET ARE IN THE RATIO 90:10

Model	Accuracy
Linear Regression	0.83
Linear Discriminant Analysis	0.82
K nearest neighborhood	0.63
CART	0.77
Naïve Bayes	0.82
Enhanced Naïve Bayes	0.83
Support Vector Machine	0.55
Neural Networks	0.87
Proposed model	0.88

The results shown in Table 2 are obtained when the test and train tests are split in the ratio of 0.9 and 0.1 respectively.

TABLE 3: RESULTS OF ACCURACY ON VARIOUS MACHINE LEARNING ALGORITHMS ON HEART DISEASE DATASET WHEN THE TRAIN AND TEST SET ARE IN THE RATIO 70:30

Model	Accuracy
Linear Regression	0.83
Linear Discriminant Analysis	0.82
K nearest neighborhood	0.64
CART	0.81
Naïve Bayes	0.82
Enhanced Naïve Bayes	0.83
Support Vector Machine	0.54
Neural Networks	0.87
Proposed model	0.88

The results shown in Table 3 are obtained when the test and train tests are split in the ratio of 0.7 and 0.3 respectively.

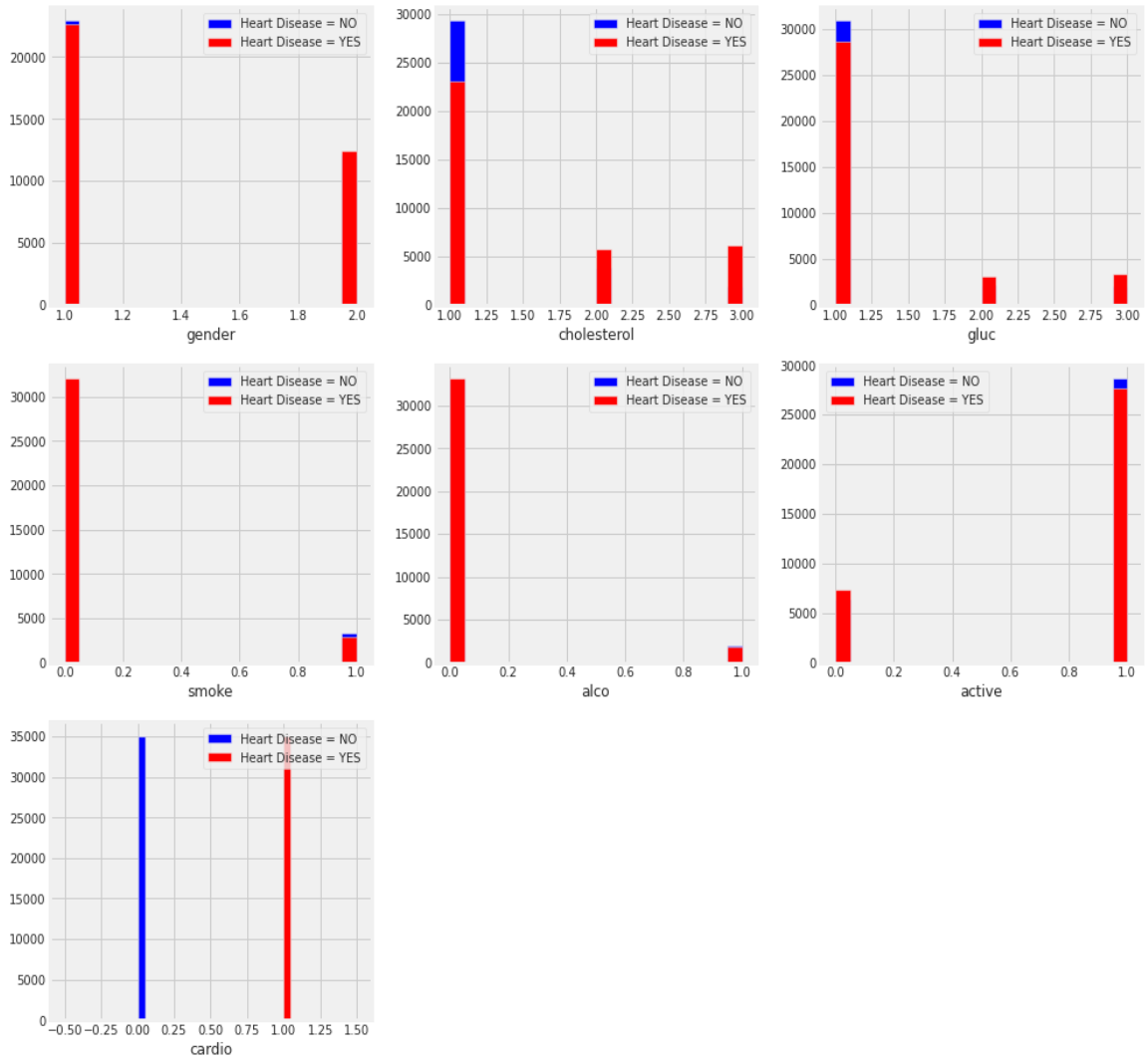

Figure 8: Correlation of the attributes of cardiovascular dataset with the target 'cardio' attribute

From Tables 1,2,3 it can be observed that the results are consistent from the proposed model and the proposed model outperformed all the machine learning models.

The results of all the machine learning algorithms like LR, LDA, CART, NB, ENB are also consistent, but the results of these algorithms are less when compared with the proposed and the neural networks.

The ENB algorithm is an enhanced version of naïve bayes algorithm, and it is implemented in the work [16]. The results of ENB are also consistent. We have implemented this algorithm to compare the results with the proposed model and the machine learning algorithms.

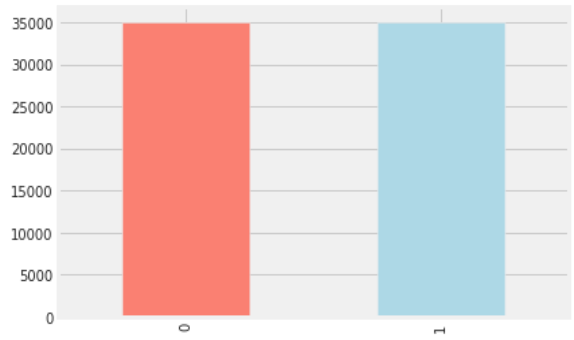

Figure 9: Number of records of cardiovascular dataset. 'O' indicates no disease and '1' indicates record with disease

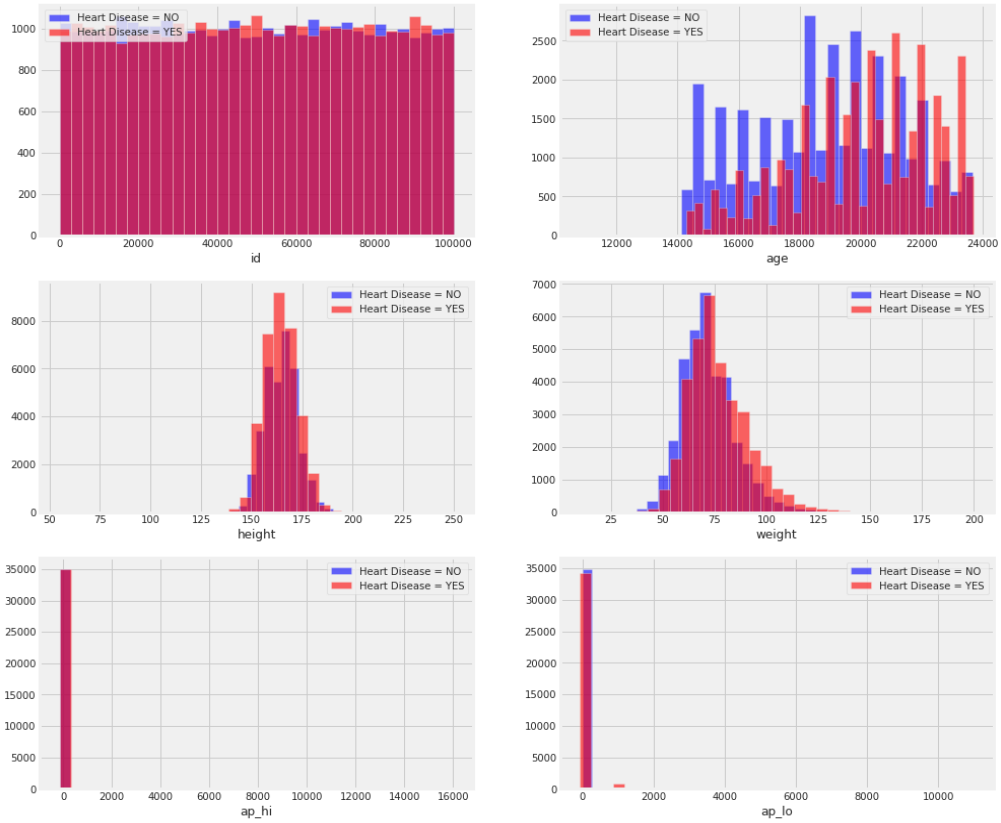

Figure 10: Correlation of the attributes of cardiovascular dataset with the target 'cardio'

The neural networks model is implemented by us taking an input layer and two hidden layers and an output layer. The results of this model are also consistent, but the proposed model outperformed all the models. Even though the train and test splits are varying there is no much difference in the accuracy.

The next part of the results is for the cardiovascular heart disease dataset. There are around 70,000 records in the cardiovascular dataset. The dataset has almost equal

distribution of records with no disease and with disease. Figure 9 shows the statistics of the dataset.

Figures 8, 10 show the correlation of all the attributes with the target attribute ‘cardio’ of the dataset. In the cardiovascular dataset the outcome is placed in the attribute ‘cardio.’

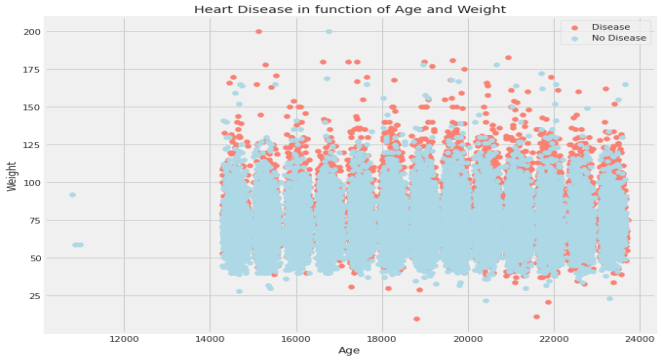

Figure 11: Cardio as a function of age and weight attributes

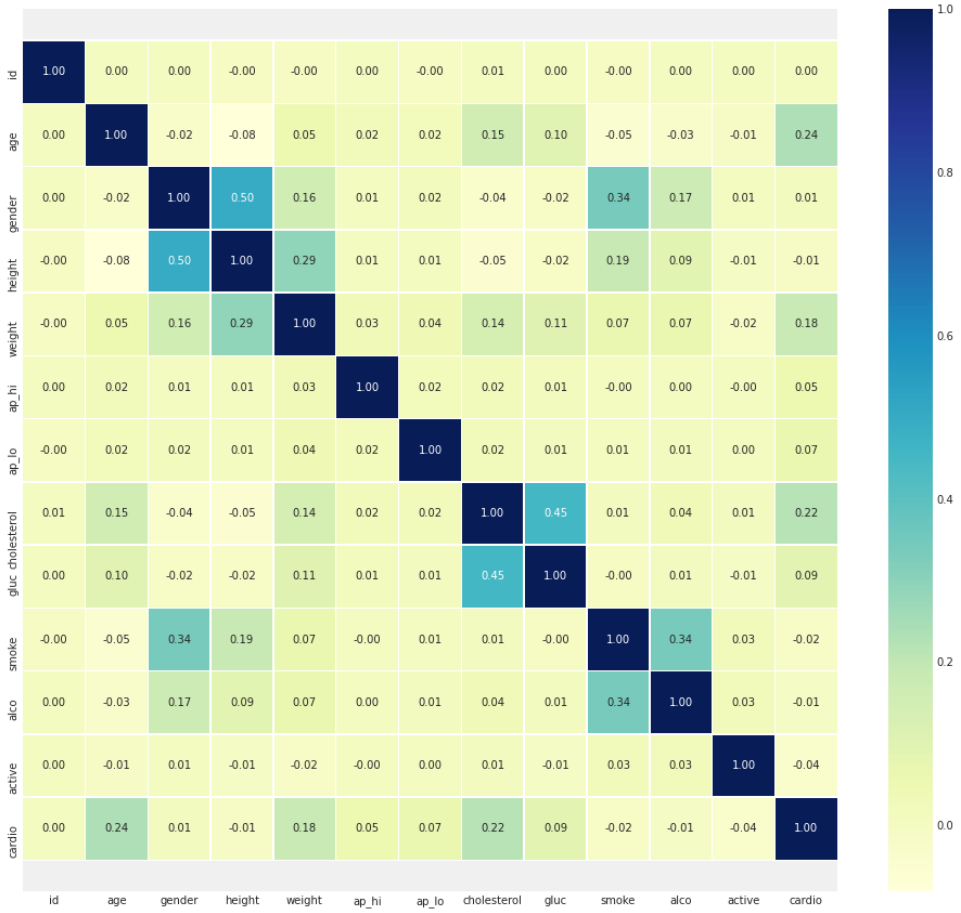

Figure 12: Correlation matrix of cardiovascular dataset attributes

Figure 11 gives the correlation of the target attribute 'cardio' with age and weight. The dataset does not contain any missing values and the results are presented here in this section. The dataset does not contain any missing values and the results of this dataset are as follows.

Figure 12 shows the correlation matrix of all the attributes of the cardiovascular dataset and the dataset has attributes age, gender etc.

The attributes height and gender are more correlated compared with other attributes.

The attributes glucose and gender are also more related with each other. The dataset has the target attribute 'cardio.'

The attribute 'cardio' is more related with age, weight and cholesterol. The attributes relations are shown in the form of graph in Figure 13.

Figure 13 shows the values of the correlation of all the attributes with the target attribute 'cardio.'

Various machine learning algorithms are considered to evaluate the dataset and the results of the dataset are presented in Table 4. The results presented in the table are when the train and test sets are 80:20.

From Table 4 it can be observed that the proposed model outperformed all the other algorithms results.

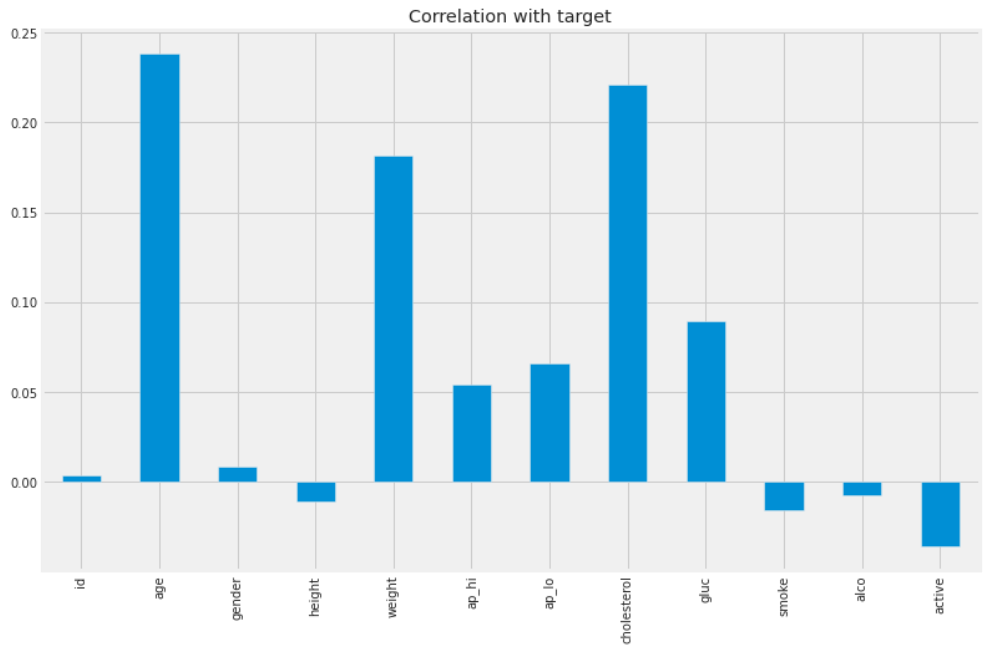

Figure 13: Correlation matrix of cardiovascular dataset attributes with the attribute 'cardio'

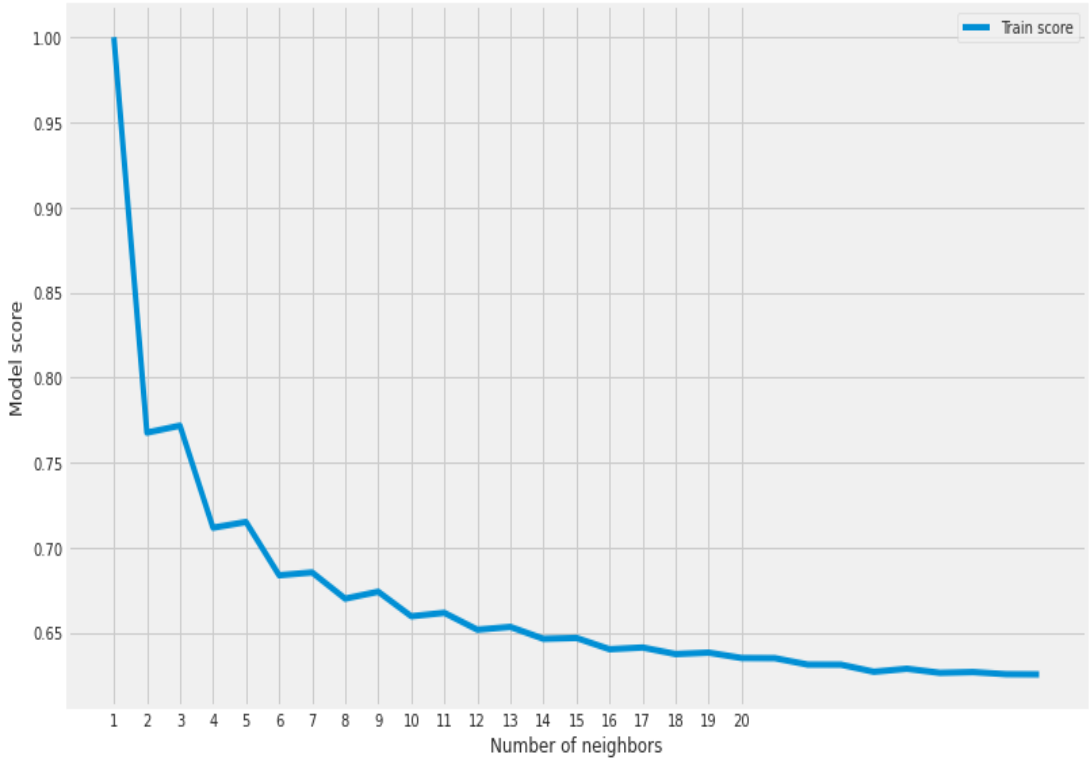

Figure 14: Accuracy of the K nearest neighborhood algorithm for different 'k'(1 to 20) values on the cardiovascular dataset

Figure 14 shows the accuracy values for KNN algorithm for different 'k' values and the from the graph it can be observed that the KNN has shown better performance n cardiovascular dataset when k=2.

Table 4 presents the results of the proposed model and various machine learning algorithms for the train and test ratio of 80:20. From the results it can be observed that the propose model outperformed all the other algorithms.

The following reasons made the proposed model more robust,

i) assigning fuzzy membership values to the attributes

ii) designing a framework with neural networks

iii) calculating the error and updating weights.

TABLE 4: RESULTS OF ACCURACY ON VARIOUS MACHINE LEARNING ALGORITHMS AND THE PROPOSED MODEL ON CARDIOVASCULAR DISEASE DATASET WHEN THE TRAIN AND TEST SET ARE IN THE RATIO 80:20

Model	Training accuracy	Testing Accuracy
Linear Regression	70.96	71.50
K nearest neighborhood	71.50	55.66
Decision tree	100.00	63.65
Random Forest	100.00	72.54
XGBoost Classifier	73.76	74.00
Naïve Bayes	71.50	55.66
Neural Networks	82.50	78.66
Proposed model	84.50	80.66

This section covered the experimentation and results of the proposed and various machine learning algorithms on the heart disease related datasets.

5. CONCLUSION

As the heart diseases are increasing very rapidly now a days due to stress and work pressures, predicting it at the early stages will help to cure without serious damage. In this paper, we mainly focused on early detection and prediction of heart failure with special attention to diabetes. We have designed a neuro fuzzy model also to predict the cardiovascular diseases. In this paper various machine learning algorithms are examined intensively and their results are reported. The state-of-art models are also compared in this work and the results of the proposed model outperformed the existing models. This work focused mainly on neural networks and in future we try to focus on long short term memory based neural networks to predict the heart diseases.

REFRENCES:

[1] Rutledge, G. E., Lane, K., Merlo, C., & Elmi, J. (2018). Coordinated approaches to strengthen state and local public health actions to prevent obesity, diabetes, and heart disease and stroke. *Preventing chronic disease*, *15*.

[2] Park, B. Z., Cantrell, L., Hunt, H., Farris, R. P., Schumacher, P., & Bauer, U. E. (2017). Peer Reviewed: State Public Health Actions to Prevent and Control Diabetes, Heart Disease, Obesity and Associated Risk Factors, and Promote School Health. *Preventing Chronic Disease*, *14*.

[3] Schwingshackl, L., Knüppel, S., Michels, N., Schwedhelm, C., Hoffmann, G., Iqbal, K., ... & Devleesschauwer, B. (2019). Intake of 12 food groups and disability-adjusted life years from coronary heart disease, stroke, type 2 diabetes, and colorectal cancer in 16 European countries. *European journal of epidemiology*, *34*(8), 765-775.

[4] Dale, C. E., Fatemifar, G., Palmer, T. M., White, J., Prieto-Merino, D., Zabaneh, D., ... & Casas, J. P. (2017). Causal associations of adiposity and body fat distribution with coronary heart disease, stroke subtypes, and type 2 diabetes mellitus: a Mendelian randomization analysis. *Circulation*, *135*(24), 2373-2388.

[5] Khan, Hassan, Stefan D. Anker, James L. Januzzi Jr, Darren K. McGuire, Naveed Sattar, Hans Juergen Woerle, and Javed Butler. "Heart failure epidemiology in patients with diabetes mellitus without coronary heart disease." *Journal of cardiac failure* 25, no. 2 (2019): 78-86.

[6] Mancini, GB John, David J. Maron, Pamela M. Hartigan, John A. Spertus, William J. Kostuk, Daniel S. Berman, Koon K. Teo, William S. Weintraub, William E. Boden, and COURAGE Trial Research Group. "Lifestyle, glycosylated hemoglobin A1c, and survival among patients with stable ischemic heart disease and diabetes." *Journal of the American College of Cardiology* 73, no. 16 (2019): 2049-2058.

[7] Chen, Hua- Fen, Ching- An Ho, and Chung- Yi Li. "Risk of heart failure in a population with type 2 diabetes versus a population without diabetes with and without coronary heart disease." *Diabetes, Obesity and Metabolism* 21, no. 1 (2019): 112-119.

[8] Rubler S, Dlugash J, Yuceoglu YZ, Kumral T, Branwood AW, Grishman A. New type of cardiomyopathy associated with diabetic glomerulosclerosis. Am J Cardiol. 1972;30:595–602.

[9] Paolillo, Stefania, Fabio Marsico, Maria Prastaro, Francesco Renga, Luca Esposito, Fabiana De Martino, Pierfrancesco Di Napoli et al. "Diabetic cardiomyopathy: definition, diagnosis, and therapeutic implications." *Heart failure clinics* 15, no. 3 (2019): 341-347.

[10] Lehrke, Michael, and Nikolaus Marx. "Diabetes mellitus and heart failure." *The American journal of cardiology* 120, no. 1 (2017): S37-S47.

[11] Lim, Ka Keat, Vivian Shu Yi Lee, Chuen Seng Tan, Yu Heng Kwan, Zoey Hui Xian Lim, Hwee Lin Wee, Truls Østbye, and Lian Leng Low. "Examining the heterogeneity in excess risks of coronary heart disease, stroke, dialysis, and lower extremity amputation associated with type 2 diabetes mellitus

across demographic subgroups in an Asian population: A population-based matched cohort study." *Diabetes Research and Clinical Practice* 171 (2021): 108551.

[12] Nazari, Somayeh, Mohammad Fallah, Hamed Kazemipoor, and Amir Salehipour. "A fuzzy inference-fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases." *Expert Systems with Applications* 95 (2018): 261-271.

[13] Birkeland, Kåre I., Johan Bodegard, Jan W. Eriksson, Anna Norhammar, Hermann Haller, Gerard CM Linssen, Amitava Banerjee et al. "Heart failure and chronic kidney disease manifestation and mortality risk associations in type 2 diabetes: a large multinational cohort study." *Diabetes, obesity and metabolism* 22, no. 9 (2020): 1607-1618.

[14] Sharma, Manik, Gurvinder Singh, and Rajinder Singh. "An advanced conceptual diagnostic healthcare framework for diabetes and cardiovascular disorders." *arXiv preprint arXiv:1901.10530* (2019).

[15] P. Chen and C. Pan, "Evaluation of the relationship between diabetes and large blood vessel disease," 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), 2017, pp. 200-207, doi: 10.2316/P.2017.852-008.

[16] Susilawati, Desi Susilawati, and Dwiza Riana. "Optimization the Naive Bayes Classifier Method to diagnose diabetes Mellitus." *IAIC Transactions on Sustainable Digital Innovation (ITSDI)* 1, no. 1 (2019): 78-86.