ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

HYPER PARAMETER OPTIMIZATION IN STACKED DEEP NEURAL NETWORK FOR MEDICAL DIAGNOSIS

Ambily Merlin Kuruvilla
Research Scholar
Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, India.
ambilykuruvilla@gmail.com

Dr. N V Balaji
Dean, Faculty of Arts, Science and Humanities, Karpagam Academy of Higher Education
Coimbatore, India.
balajinv@karpagam.com

Abstract— Diabetes is a metabolic disease where the blood sugar rate of an individual is consistently above normal. Due to the modern lifestyle and work culture, diabetics is widespread and affects the productivity and quality of life for an individual. A diabetics patient is at a very high risk of various health issues like organ failure and even it can even result in loss of life. An early prediction of this chronic disease can avoid health issues and save many lives. The aim of this article is to develop a better predictive model for diabetics using an automated hyper parameter optimization (HPO) approach in Multilayer Perceptron (MLP). This article provides an efficient way to increase the accuracy of Neural Network to a substantial level through the HPO process using Grid Search Optimization (GSO) through the stacking ensemble model. In order to run the ensemble model at an optimal level and to minimize errors, appropriate hyperparameters must be calculated. Three GSO methods are utilized to tune the hyper parameters. To build the stacking ensemble model, the PIMA data set was used.

Keywords: HPO, MLP, GSO: Hyper Parameter Optimization, Multilayer Perceptron, Grid Search Optimization.

INTRODUCTION

A study carried out by the WHO recently revealed that in 2016, diabetics was one of the leading causes of death worldwide. Diabetes has resulted in 1.6 million fatalities in 2016 and this statistic replaces HIV / AIDS with diabetes as one of the most frequent cause of death [4]. The burden of diabetes disease grew from 108 million in 1980 to 422 million [5] in 2014, and the percentage of diabetic patients amongst adults over 18 years of age rose from 4.7% in 1980 to 8.5% in 2014[5]. 642 million people i.e. (1 in 10 people) are expected to contract diabetes by 2040. 46.5% of people with diabetes have not been diagnosed officially [6]. This makes it necessary to develop techniques and procedures to assist in the early detection of diabetes in order to reduce the number of deaths related to diabetes, as late diagnosis is responsible for a majority of deaths linked to diabetes [7].

There is a need to implement sophisticated information processing to develop cutting-edge strategies for the early detection of diabetes. Data mining tools can also be effectively applied. The ability to remove and uncover previously unseen, secret, yet important patterns from a large database repository is given by data mining [7]. These tools can assist medical evaluation and decision making.

LITERATURE SURVEY

Roshan Birjais [1] conducted a research in many classification algorithms for diabetes prediction and his team found out that the Gradient Boosting algorithm outperform other classifiers and

obtained 86% accuracy. They have analysed various prime factors for the cause of diabetes disease.

Md. Maniruzzaman [2] wrote an article in which he used LR for identifying the prime features for the diagnosis of diabetes disease. Accuracy and Area Under curve are used as the performance measures for the classifiers. It revealed that LR along with the Random forest generates the maximum accuracy.

Atik Mahabub[3] used an ensemble voting classifier for forecasting diabetes disease. The Out of eleven classifiers, the best performing three will be used for the ensemble classifier. Accuracy, Precision, F-Measure and Recall [3] are used as the parameters for evaluating the classifiers. The outcome clearly shows that the ensemble method outperform the base classifiers.

BACKGROUND STUDY

Neural network is used for the prediction. Parameter tuning mechanism in neural network is applied for improving the accuracy.

ARTIFICIAL NEURAL NETWORK

Artificial Neural Network [17] is built by multiple nodes that reproduce the human brain's biochemical neurons. The neurons are linked and are communicating with each other. The nodes are capable of taking input data and executing the operations. The outcome of these operations is transferred to other neurons. The output is referred to as its activation or node value for a node. A technique known as Gradient Descent in the Artificial Neural Network, which takes places in the backpropagation period whereby it intends to regularly resample the gradient of the model parameter in the reverse direction based on the weight 'W', periodically updating till the global minimum of G(W) function is reached.

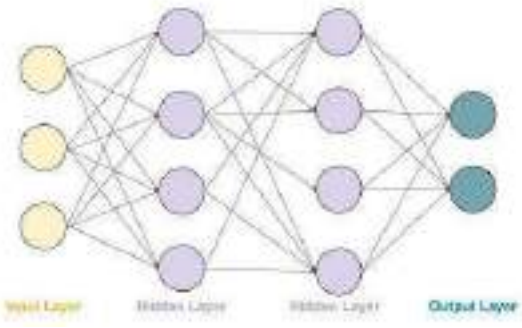

Figure 1: Hidden Layers in Artificial Neural Network

A loss represents the prediction error found in the Artificial Neural Network. In deep learning, this is estimated as a loss function. The Loss Function describes the model's operating efficiency. For the estimation of the loss function, stochastic gradient descent was used. For each iteration, weights are updated, and the model is trying to reach the global minimum point.

LOGISTIC REGRESSION

Logistic Regression is a well-known classification algorithm used to estimate the probability of a target variable. The design of the goal or dependent variable is binary, indicating that only two possible groups are available. Mathematically, f(Xi), a LR model predicts P(Yi=1).

b=wt0+wt1 x1+wt2x2+…+wtnxn.

A sigmoid function can be used to map the predicted values to the probabilities.

S(y)=1/(1+e(-y))

S(y) is the estimated output; y is the value inputted to the function and e is the base of natural log.

NAVIE BAYES

The Navie Bayes algorithm follows the Bayasian principle which belongs to a category of conditional probabilities (CP). The CP is the probability that an event B will happen, provided A has already occurred. Though Bayes Theorem provides a principled way for calculating conditional probability, in practice its computationally expensive and thus using some assumptions, bayes theorem is simplified by making some assumptions and turning it into an effective classification model referred to as Naive Bayes. Conditional probability can yield the probability of an event using prior information.

P(H/E) = (P(E/H) * P(H))/(P(E))

PROPOSED MODEL

Stacking Ensembles

To obtain better performance, ensemble methods allow for combining the results of many methods. More the models, better will be the performance of the ensemble strategies. An ensemble method where the no. of models is stacked in a way that their observations act as input to a new model is called Stacking.

Hyper Parameters in Neural Network

The variables that determines how the network is trained are called the Hyperparameters. Hyper Parameters also determine network structure. These variables are set before the training phase, i.e. before the weight and bias is optimized.

Network Weight Initialization

Based on different activation functions applied to each layer, it is preferred to use separate weight initialization schemes.

Activation function

It is used to apply nonlinearity to frameworks that will permit nonlinear prediction restrictions to be learned by deep learning models. The most popular among the activation functions is the rectifier activation function. While making predictions, for binary, Sigmoid is used while for multi-class predictions, softmax is deployed in the output layer.

Gradient Descent

The learning rate determines how fast its parameters are modified by a network. The learning process is slowed down by a low learning rate, but converges efficiently. The higher learning rate accelerates learning, but does not converge. A decreasing learning rate is usually prioritised.

Momentum

Momentum is used to eliminate oscillations and to know the course of the next step with knowledge of the previous phase. For momentum, usually, a value between 0.5 and 0.9 is used.

Number of epochs

The number of epochs measures the number of times the entire training data is given to the model during the training process. The number of epochs is raised until the accuracy begins to decline, while the accuracy of the training is improved due to overfitting.

Batch size

Refers to the no. of sub samples provided to network before the parameter is updated.

METHODOLOGY

Diabetes prediction using MLP Model is described in the following steps. In this model these steps are used to predict diabetes more accurately.

Data Collection

PIMA Indian Data set [17], from the UCI repository is used for the analysis. The data set consists of features including age, number of pregnancies, Body Mass index, Diabetes Pedigree Function, Glucose Level [17] etc…

HYPER PARAMETER TUNING USING GRID SEARCH.

It is one of the traditional hyper parameter tuning method [13]. Before the learning process starts, the value of the hyper parameter needs to be calculated. Grid Search is also known as a comprehensive search [13], with each mixture of hyper parameters explored by Grid Search. This means that each variation of the hyper parameter values listed would be tried. There can be several parameters for models, and it can be viewed as a search problem to find the right combination values to the parameters. The purpose of algorithm tuning is to find the best values of the parameter corresponding to the particular problem. Grid Search can be expanded to provide the highest results by using automatic approaches for finding optimum values to the parameters.

HYPER PARAMETER TUNING USING GRID SEARCH.

Also known as a comprehensive search, it's a traditional hyper parameter tuning method. The value of the hyper parameter needs to be measured before the start of the learning process. Here, each variation of the hyper parameter values is explored which means that each variation of the hyper parameter values that are listed is tested. On further expanding, grid search can provide the highest results using automatic approaches that would optimum values for the parameters.

Implementation

Input values are converted to vectorised format using Logistic regression. The output of Logistic Regression is saved as a collection in Python and this will be given as the input for the MLP. Grid Search is used for doing the parameter optimization. The model optimizes the following hyper parameters

- Epoch
- Batch Optimization.
- Gradient descent

By using Keras Classifier Grid parameters such as Batch size and Epochs are optimized. In this model stochastic Gradient is introduced for training the Artificial Neural Network. The internal parameters such as epoch and Batch values are optimized and the model is automized so the best values for Epoch and Batch optimization is find out and it is assigned. This leads to an increase in accuracy and reduced the loss and MSR (Mean Square Error). The Vanishing Gradient Problem is overcome by using ReLU (Rectified Linear unit) activation function. The formulae for ReLU is

R= {0, z}

{1 for z > 0}

{0 for z < 0}

The input of each neuron is passed to the activation function and there it is processed. ReLU overcoming the problem of vanishing gradient by keeping the derivative value as positive. So always there is a difference between Wold and Wnew in ReLU. Feedback connections were introducing LSTM and the best score for grid search is found out. By using standard Grid Search Value of Epoch and Batch size is automated and it is found out that for this dataset the best result was obtained for Batch size =20 and Epoch=20. Forward propagation is improved by adding the features of LSTM [12]. So a new memory part was introduced to store [12] the activation details of the hidden layers. Various steps involved in this model are described below:

- Choose the data set
- Convert all the input values to Vectorised format for standardization Find out the training values
- Initialise the model using Keras Classifier
- Three sub models are introduced, and, in each model, Optimal parameters are finding out using hyper parameter tuning in Grid Search Model.
- In model1 the best values for Epoch and Batch size was find out using Grid search in Keras Classifier. The model is tested with three batch sizes (10,20 and 30) and three epochs (10,20 and 30). All combinations were tested and the best accuracy will be find out for the combinations of batch size 20 and epoch 20.
- In Model2 Forward propagation is improved by introducing LSTM and Adam Optimizer. By automating the model best values for epoch and batch size was predicted. Learning rate and dropout rate is find out and best classifying accuracy is calculated.
- In Model3 ReLU activation function is introduced along with Softmax function. Best predicted result was observed.
- Stacked the three Models using tensor flow method and the accuracy matrix was printed. Logistic Regression is used for stacking the Models.

FINDING OPTIMAL PARAMETERS USING HYPER PARAMETER TUNING IN GRID SEARCH MODEL.

- In model1 the best values for Epoch and Batch size was find out using Grid search in Keras Classifier. The model is tested with three batch sizes (10,20 and 30) and three epochs (10,20 and 30). All combinations were tested and the best accuracy will be findout for the combinations of batch size 20 and epoch 20.
- In Model2 Forward propagation is improved by introducing LSTM and Adam Optimizer. By automating the model best values for epoch and batch size was predicted. Learning rate and dropout rate is find out and best classifying accuracy is calculated.
- In Model3 ReLU activation function is introduced along with Softmax function. Best predicted result was observed.
- Stacked the three Models using tensor flow method and the accuracy matrix was printed. Logistic Regression is used for stacking the Models.

RESULT AND DISCUSSION

From the observations it is found out that the accuracy is incremented in the new model comparing with the base classifiers. Here three base classifiers are used for the accuracy comparison. Naive Bayes, Logistic Regression, Multilayer Perceptron. From the observations it is found out that the automated hyperparameter optimization with stacked ensembled model increases the accuracy level.

Algorithm	Accuracy	Precision	Recall
Navie Bayes	76.3	75.9	76.3
Logistic	77.2	76.7	77.2

Regression			
MLP	78.3	75.0	74.0
Stacked Model	90.7	94.0	82.0

Table 1: Performance Comparison of Classifiers

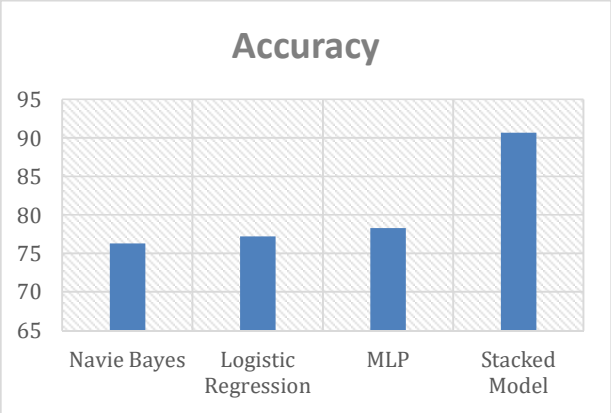

Figure 2: Classifier Performance based on Accuracy

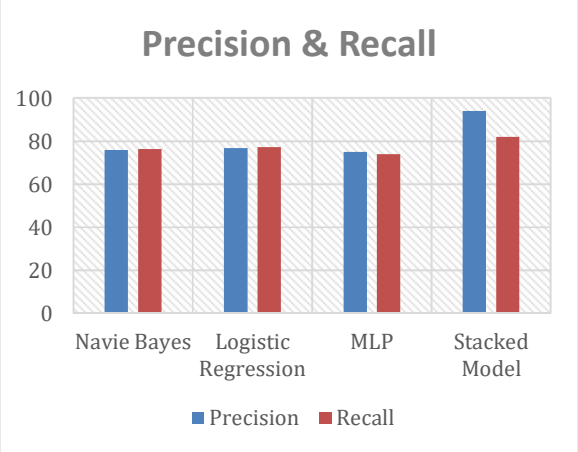

Figure 3: Classifier Performance based on Precision and Recall.

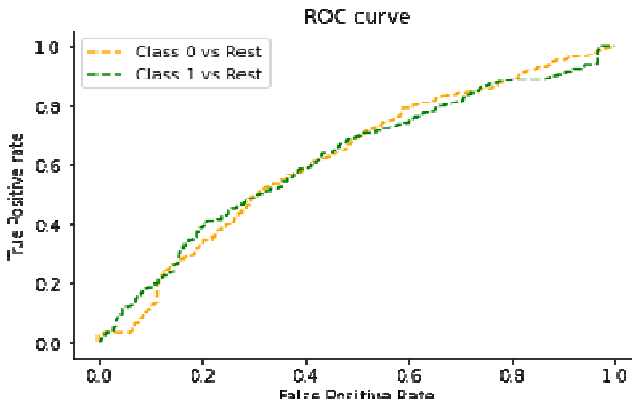

Figure 4: ROC Curve

ROC curve is a graphical plot that shows the system's diagnostic potential as its discriminating threshold is varied. The ROC curve is generated by plotting the true positive rate at different threshold settings against the false positive rate.

CONCLUSION

In this article we are introducing an ensembled hyper parameter tuning mechanism to tackle the deficiencies and improving the accuracy. For this purpose, we have used an automated stacked ensemble method which combines various hyper parameters. Grid Search Optimisation method is used, and three different models were created as base learners using Neural Network by combining various activation functions. The optimum value for each parameter is calculated and stored into an external file by each model. Three output files are created, and these output files are inputted to a logistic regression model which is a learning model. We have used LR Model as the learning model. It is found out that this model improves the accuracy to good extend.

REFERENCES

1.Roshan Birjais, Ashish Kumar Mourya,Ritu Chauhan, Harleen Kaur " Prediction and diagnosis of future diabetes risk : a machine learning approach". Springer.

2.Md.Maniruzzaman,JahanurRahman, BenojirAhammed, Md.Menhazul bedin "Classification and prediction of diabetes disease using machine learning paradigm", Springer

3. Atik Mahabub "A robust voting approach for diabetes prediction using traditional machine learning techniques", Springer.

4.www.geeksforgeeks.org/ml-chi-square-test-for-feature-selection

5.www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes Retrieved.

6.www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death, Accessed 27th Jul 2018.

7.www.who.int/news-room/fact-sheets/detail/diabetes retrieved 27/07/2018.

8.www.diabetesdaily.com/learn-about-diabetes/what-is-diabetes/how-many-people-have-diabetes.

9.Changsheng Zhu, Christian Uwa Idemudia "Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques, ScienceDirect.
10.https://towardsdatascience.com/ways-to-detect-and-remove-the-outliers-404d16608dba
11.https://mc.ai/deep-learning-in-the-real-world-how-to-deal-with-non-differentiable-loss-functions/
12.Merijn Beeskma, Suzan Verberne, Antal van den Bosch Predicting life expectancy with a long short-term memory recurrent neural network using electronic medical records.
13.Sampath Kumar Palaniswamy, Hyperparameters tuning of ensemble model for software effort estimation, Springer.
14.Aramesh Rezaeian, Marzieh Rezaeian "Prediction of mortality of premature neonates using neural network and logistic regression", Journal of Ambient Intelligence and Humanized Computing, Springer.
15. AH. Habbi and M. Zelmat "Fuzzy Logic Based Gradient Descent Method with Application to a P1-type Fuzzy Controller Tuning: New Results", IEEE 2017.
16. Huaping Zhou, Raushan Myrzashova and Rui Zheng "Diabetes prediction model based on an enhanced deep neural network" EURASIP Journal on Wireless Communications and Networking, Springer.
17. Kalaiselvi, G. M. Nasira, "A New Approach for Diagnosis of Diabetes and Prediction of Cancer Using ANFIS", IEEE
18. Pahulpreet Singh Kohli, Shriya Arora "Application of Machine Learning in Disease Prediction", IEEE.
19. Abdulhakim Salum Hassan, I. Malaserene, A. Anny Leema , "Diabetes Mellitus Prediction using Classification Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE)
20. Sai Poojitha Nimmagadda, Sagar Yeruva, Rakesh Siempu, "Improved Diabetes Prediction Model for Predicting Type-II Diabetes", International Journal of Innovative Technology and Exploring Engineering (IJITEE).

Karpagam Academy of Higher Education

(Deemed to be University)
(Established Under Section 3 of UGC Act,1956)
Coimbatore – 641 021

Department of Computer Science, Applications & Information Technology

CERTIFICATE

This is to certify that Dr. / Mr. / Ms. Ambily Merlin Kuruvilla of Ph-D Scholar KAHE has participated / presented a paper entitled Predicting Diabetes Mellitus Using feature selection methods in machine Learning Algorithms. in the 7th National Conference on "Recent Trends and Advances in the Field of Information Technology" held at Department of Computer Science, Applications & Information Technology, KAHE, Coimbatore on 27th September, 2019.

Convenor | Organizing Secretary | HoD 27/9/19 | Dean, FASH

Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach.

To cite this article: Ambily Merlin Kuruvilla and N.V Balaji 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1085** 012028

Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach.

Ambily Merlin Kuruvilla[1] **and Dr.N.VBalaji**[2]

[1] Research Scholar, Department of CS,CA and IT, Karpagam Academy of Higher Education
[2]Dean, Faculty of Arts, Science and Humanitics, Karpagam Academy of Higher Education

E-mail: E-mail:ambilykuruvilla@gmail.com

Abstract. Cardiac disease prediction helps physicians to make accurate recommendations on the treatment of the patients. The use of machine learning (ML) is one of the solution for recognising heart disease-related symptoms. The goal of this study is to suggest a methodology for identifying the most relevant features of cardiac disease characteristics by applying a feature selection technique. The data set used in this study was Framingham heart disease dataset (FHS). It was collected from KAGGLE Machine Learning repository. There are 16 attributes and a mark in the dataset that has been validated by four ML classifiers. There are two feature selection methods, Correlation Based Feature selection (CBFS) and Principle Component Analysis (PCA) was used for the comparison in the study. By using CBFS Method five highly correlated features are selected for the study, and by using PCA thirteen features are selected. The experimental result shows that Correlation Based Feature Selection with Multilayer perceptron (CBFS with MLP) obtained the highest accuracy for this dataset.

1. Introduction

The research concentrates on the two feature selection methods for data reduction before building the predictive models by classification algorithms. These reduced features are then passed into the classification algorithms to design the models for the heart disease prediction. These models are used for the comparison of accuracy of the classifier. Principle Component Analysis and Correlation Based feature selection methods are used for finding out the reduced features. The selected features are inputted to four different classifiers such as Navie Bayes, ADABOOST, MLP and SMO. The accuracy of each model is compared with the other.

2. Background Study

Devansh Shah studied various attributes related to heart disease[1]. The study was conducted with Naïve Bayes, decision tree, K-nearest neighbor, and random forest algorithms[1]. The experimental result proves that K-nearest neighbor algorithm exhibits the highest accuracy.

Hamidreza Ashrafi Esfahani[2] formulated a model to predict cardiovascular disease. The model includes decision trees, Neural Networks, Rough set, Naïve Bayes and SVM for implementation. On comparing the results achieved, it was revealed that the hybrid model of Rough Set, Naïve Bayes and Neural Network obtained the highest accuracy. An ensemble strategy was implemented that allowed for the output to be combined that would result in

better accuracy. The performance of the classifiers was compared with the parameters such as its precision, sensitivity, accuracy and F-Measure.

3. Proposed Methodology

Framingham Heart Study (FHS) from Kaggle Machine Learning repository is used for the study. Two feature selection methods along with four classification algorithms are used for the study. CBFS and PCA are the methods used for the dimensionality reduction. MLP, Navie Bayes, Sequential Minimum Optimiser (SMO) and ADABOOST algorithms are used for classification. The reduced feature set from both the feature selection methods are inputted to different classifiers. Eight different Machine Learning Models were created for Heart Disease Prediction. Accuracy of these Models are compared with each other.

4. Description of the Data Set

The Framingham Heart Study (FHS) dataset was collected from Kaggle. The dataset consists of 4241 records. It contain sixteen features including AGE, PREVALENT HYP, SYSBP, DIABP, GLUCOSE, SEX, EDUCATION, CURRENT SMOKER, CIGSPERDAY, BPMEDS, PREVALENT STROKE,BMI, HEART RATE, DIABETES, TOTCHOL and PREDICTOR VARIABLE.

5. Classification Algorithms

In Machine Learning various forms of classification techniques are available. Classification techniques used for this study was described below.

6. Multilayer Perceptron (MLP)

MLP is a subset of Artificial Neural Network. MLP comprises one or more than one hidden layers aside from one input and one output plate. The Perceptron is made of an input layer and a totally linked output layer. MLPs have the same levels of input and output, but could have several levels concealed within them.

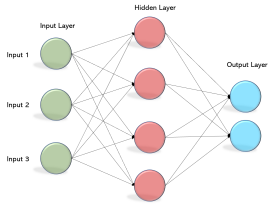

Figure 1. Different layers in MLP.

7. Adaboost

In machine learning, AdaBoost(Adaptive Boosting) is a supervised learning algorithm. It is used for combining several weak classifiers together to generate a strong classifier.

8. Naive Bayes

Naive Bayes is a Machine Learning algorithm based on Probability theory in statistics. The term naive suggests that the elements that go through the software are autonomous of each other,That is, the value of one characteristic, does not explicitly influence or alter the value of any of the other characteristics used in the algorithm. The Bayes theorem tells us how we can compute the conditional probability. The equation for conditional probability is,

P(Ai/Bi) =(P(Bi/Ai) * P(Ai)) /P(Bi)

P(Ai/Bi) defines the probability of an event Ai occurs corresponding to the event Bi has occurred.

P (Bi / Ai) is the conditional probability and it defines the probability of occurrence of event Bi corresponding to the occurrence of the event Ai.

P(Ai) and P(Bi) defines the probability of the events Ai and Bi occurs.

9. Sequential Minimal Optimization (SMO)

The sequential minimal optimization is more effective to solve the SVM problem compared to traditional Quadratic Programming algorithms such as the interior-point method. The SMO algorithm can be viewed as a method of decomposition by which a problem of optimization of multiple variables is decomposed into a set of sub problems, each optimizing an objective feature of a limited number of variables, usually only one, whereas all other variables are treated as constants which remain unchanged in the sub problem.

10. Feature Selection

During feature selection the most relevant features are extracted from the data set. Redundancy can be avoided using this method. Since irrelevant features are excluded from the input data, feature selection can increase the accuracy of prediction. In this study Correlation Based Feature Selection (CBFS) and Principle Component Analysis(PCA) is used for feature selection. After feature selection the reduced data set is applied to four different classification Algorithm.

11. Correlation Based Feature Selection(CBFS)

Correlation values are calculated by CBFS. The five highly correlated features are selected for the analysis. These features are given as the inputs for the classifiers.

Table 1. Features selected for the analysis by CBFS along with the correlation values.

S/N	Selected Features	Correlated Values
1	AGE	0.2254
2	PREVALENTHYP	0.2164
3	AGE	0.2254
4	PREVALENTHYP	0.2164
5	PREVALENTHYP	0.2164

12. Principal Component Analysis (PCA)

Thirteen features were selected by PCA during feature selection. The features selected by the PCA algorithm are AGE, PREVALENTHYP, SYSBP, DIABP, DIABETES, SEX, BPMEDS, TOTCHOL, PREVALENTSTROKE, CIGSPERDAY, EDUCATION, BMI, CURRENT SMOKER.

13. Result and Discussion

In the study two feature selection methods are used for comparison - Principle Component Analysis (PCA) and Correlation Based Feature Selection (CBFS). After dimensionality reduction the reduced dataset is applied to four different classification Algorithm such as

Multilayer Perceptron (MLP), AdaBoost, Navie Bayes and SMO. Five most correlated features were selected and applied to Classifiers in CBFS Method. Thirteen features were selected by Principle Component Analysis (PCA). The result is shown in Table4. From the results it is found out that Correlation Based Feature Selection (CBFS) along with MLP algorithm shows maximum accuracy.

Table 2. Comparison of Predictive accuracy of Models.

S/N	Alogrithm	CBFS	PCA	Before FS
1	MLP	84.9057	83.9151	84.1509
2	ADABOOST	84.8113	84.8821	84.8821
3	Navie Bayes	81.1792	80.0472	80.0472
4	SMO	84.8113	84.8113	84.8113

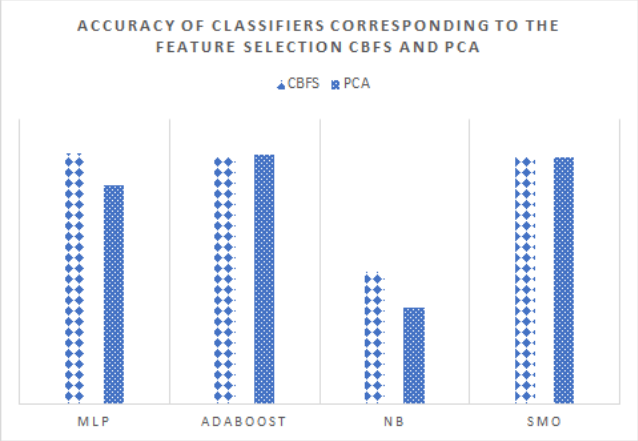

Figure 2. Comparison of Accuracy of classifiers corresponding to CBFS and PCA Feature Selection Methods.

Eight Models are generated by combining two feature selection algorithms and four classifiers. They are CBFS-MLP, CBFS-ADABOOST, CBFS-NB, CBFS-SMO,PCA-MLP, PCA-ADABOOST, PCA-NB and PCA-SMO.The accuracy of various Models are shown in Table.1 The results from table proves that Correlation Based Feature Selection along with Multilayer Perceptron (CBFS-MLP) Model perform better than the other Models with the accuracy of 84.9057 Percentage.

The result from Table 2 proves that CBFS-MLP combination shows better performance than MLP classifier. Also when we are comparing CBFS-MLP and PCA-MLP models accuracy measures proves that CBFS-MLP combination shows more performance.

Table 3. Comparison of Predictive Accuracy of MLP.

MLP	Precision	Recall	F Measure	ROC
Before FS	0.790	0.842	0.798	0.671
After CBFS	0.805	0.849	0.791	0.649
After PCA	0.784	0.795	0.668	3.62

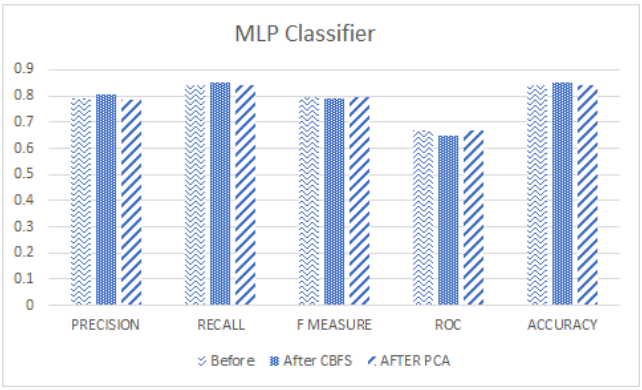

Figure 3. Comparison of Accuracy of MLP Classifier.

14. Conclusion and Findings

During the study the performance of two different feature selection methods CBFS and PCA are evaluated. Eight different classifier models are developed by combining the feature selection and classification algorithms.The performance of each model was evaluated.Performance measures such as Accuracy, Precision, Recall, F Measure and ROC are evaluated for finding out the best classifier. From the result it is proven that the model CBFS with MLP Classifier shows the maximum performance for FHS dataset.

References

[1] Devansh Shah, Samir Patel and Santosh Kumar Bharti 2020 Heart Disease Prediction using Machine Learning Techniques Springer.
[2] Hamidreza and Ashrafi Esfahani 2017 Cardiovascular disease detection using a new ensemble classifier IEEE International Conference on Knowledge-Based Engineering and Innovation (KBEI).
[3] V.V.Ramalingam 2018 Prediction of Heart Diseases Using Machine Learning International Journal of Engineering and Technology.
[4] Amin Ul Haq 2019 A Hybrid Intelligence System Frame Work for Prediction of Heart Disease Using ML, Mobile Information Systems.
[5] Shadman Nashif Heart Disease Detection by Using Machine Learning Algorithm and a Real time Cardiovascular Health Monitoring System World journal of Engineering and Technology,vol.06 no.04,2018,article id 88650.
[6] Poornima Singh 2018 Effective heart disease prediction system using data mining techniques International journal of Nano medicine.
[7] M. Gunay and T. Ensarı 2019 Predictive churn analysis with machine learning methods IEEE 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1- 4.
[8] R. Suguna, M. Shyamala Devi and Rincy Merlin Mathew 2019 Customer Churn Predictive Analysis by Component Minimization using Machine Learning International Journal of Innovative Technology and Exploring Engineering.
[9] Shyamala Devi, Munisamy, Suguna Ramadass and Aparna Joshi 2020 Cultivar Prediction of Target Consumer

Class using Feature Selection with Machine Learning Classification Springer's book series "Learning and Analytics in Intelligent Systems Springer, LAIS vol. 3, pp. 604-612, 2019.
[10] Suguna Ramadass and Shyamala Devi 2019 Prediction of Customer Attrition using Feature Extraction Techniques and its Performance Assessment through dissimilar Classifiers Springer's book series Learning and Analytics in Intelligent Systems, Springer.
[11] R.Suguna, M. Shyamala Devi, Rupali Amit Bagate and Aparna Shashikant Joshi 2019 Assessment of Feature Selection for Student Academic Performance through Machine Learning Classification Journal of Statistics and Management Systems, Taylor Francis.
[12] M. Shyamala Devi, Rincy Merlin Mathew and R. Suguna 2019 Feature Snatching and Performance Analysis for Connoting the Admittance Likelihood of student using Principal Component Analysis International Journal of Recent Technology and Engineering.
[13] Debjani Panda, Ratula Ray, Azian Azamimi Abdullah and Satya Ranjan Dash 2019 Role of Feature Selection in Prediction of Heart Disease International Conference on Biomedical Engineering (ICoBE) IOP Publishing
[14] Bandari Sai Santosh, Dharma Sahith Reddy, M Sai Vardhan and Dr. Shaik Subhani 2019 Heart Disease Prediction with PCA and SRP International Journal of Engineering and Advanced Technology .
[15] Mothe Sunil Goud 2019 Heart Disease Prediction and Performance Assessment through Attribute Element Diminution using Machine Learning International Journal of Innovative Technology and Exploring Engineering.

IMPROVED ARTIFICIAL NEURAL NETWORK THROUGH METAHEURISTIC METHODS AND ROUGH SET THEORY FOR MODERN MEDICAL DIAGNOSIS

Ambily Merlin Kuruvilla

Research Scholar, Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore.
E mail: ambilykuruvilla@gmail.com

Dr. N.V. Balaji

Dean, Faculty of Arts, Science and Humanities,
Karpagam Academy of Higher Education, Coimbatore.
E mail: balajinv@karpagam.com

Abstract- A novel meta-heuristic soft computing model with feature selection implemented using Rough set (RS) theory for the diagnosis of coronary artery disease in diabetes patients is proposed in this study. The binary classification method in multiclass classification problems is applied by the One Versus Rest approach (OVR) is incorporated. To avoid the redundancy problem, a mathematical approach known as rough-set theory (RS) is applied to identify the most significant features from the dataset. The Artificial Neural Network with improved hidden layers is used as the classifier which is optimized through a metaheuristic population-based method, known as the Grasshopper Optimization Algorithm (GOA) with a single objective optimization approach integrated for improving the accuracy of the model. Mean Square Error (MSE) is taken as the objective function and the result shows that the accuracy of the model has been improved significantly from 89.1% to 95.25% after optimization.

***Keywords*: One Versus Rest (OVR); Artificial Neural Network (ANN); Rough Set (RS); Grasshopper Optimization Algorithm (GHO); Mean Square Error (MSE).**

1.Introduction

Diabetes is a chronic condition that has the potential to lead to many side effects and heart disease is one among them. According to the report of the World Health Organization, diabetics patients have a higher probability of heart diseases. The statistics show that the death rate due to heart diseases are on a rise. Diabetes can cause damage to the blood vessels as well as the nerves which control the functions of the heart. This will lead to cardiac disorders and coronary artery problems. Heart disease being a major health hazard is an issue that immediately requires more academic research. The risk of heart disease can be avoided or reduced by controlling the diet, blood glucose level, and regular physical activity. Therefore, early prediction of the possibility of heart disease in diabetes patients will help to improve their health conditions. The main objective of the study is to find out the possibility of heart disease especially among diabetes patients. Artificial Neural network is used as the classifier for the proposed study. It is widely accepted that backpropagation in the Neural Network is a very strong tool for forecasting and estimation. The main power of the Neural Network is that we can do many complex problems with large datasets that can be estimated without any complex mathematical expressions. In Neural Networks, weight optimization plays an important role in the accuracy obtained. By using the activation function and error value obtained in each iteration, the weights of the Neural Network are adjusted until it reaches the optimum value. But during these steps, the local optimization problem can arise. To overcome this issue, metaheuristic population-based algorithms such as grasshopper optimization algorithms are used to optimize the weight of the Artificial Neural Network and the optimal weight value is used for the prediction.

Grasshopper Optimization is one among the stochastic methods of search optimization [Janmenjoy Nayak, (2020)] which will help to overcome the issue of local optima problem in Neural Network. It is developed based on the life cycle and food search pattern of insects known as Grasshoppers. During their life cycle, grasshoppers follow

two stages: The nymph and adult stage. During the Nymph stage, it follows slow movement, and during the adult stage, it follows the abrupt movement. These two movements will be adjusted in the algorithm with the help of mathematical methods. This feature of GOA helps to find out the optimum results by overcoming the stochastic gradient descent problems in the Neural Network.

Another main concept used in the study is dimensionality reduction. A mathematical foundation tool called Rough Set is used for feature selection by finding the most relevant subsets. A reduced subset retains all the features of the parent feature set. In the study, the Framingham dataset is used for classification purposes.

2. Literature Review

Ashir Javeed and Sanam Shahla Rizvi proposed a floating window with adaptive size [Ashir Javeed, (2020)] method for dimensionality reduction purposes. They have proposed this method to create a subset feature and is given as the input to the Neural Network architecture. ANN and Deep Neural Networks are used as the classifiers. The cleveland dataset was used for the analysis. The result shows that the proposed model achieves an accuracy rate of 91.11% for ANN and 93.33% by the Deep Neural Network.

A hybrid Model of Emotional Neural Networks (EmNNs) and Particle Swarm Optimization (PSO) [Afzal Hussain Shahid, (2020)] was introduced for the detection of Coronary Artery Disease (CAD) by Afzal Hussain Shahid. Feature selection was carried out by four different methods. To improve the learning capability of the Neural Network, emotional parameters are also taken into consideration along with the conventional weights and bias.

Nithyavishnupriya in her article used DNN for heart disease prediction. They used the statistical analysis Chi-Square analysis [Nithyavishnupriya and Ramprakash (2020)] along with DNN architecture. They utilized the Dataset collected from the UCI Repository which consists of 303 instances with 14 columns.

In the model, proposed by Archana Singh, they utilized Python programming in Anaconda (Jupiter) notebook [Archana Singh (2020)] as a simulation environment. They have tested various supervised machine learning algorithms and stated that the KNN algorithm is the most suited method compared with the other classifiers they have tested.

A Machine learning-based medical intelligence system was made by Amin Ul Haq [Amin Ul Haq, (2018)]. In his study, Relief, minimal-redundancy-maximal-relevance (mRMR), Shrinkage and Selection Operator (LASSO) were used to find out the most relevant and highly correlated features. Seven supervised classification algorithms are used for the prediction. Reduced feature sets obtained from each feature selection process were given as input to the classifiers. Time for prediction is also calculated.

Awais Mehmood proposed a new method of study known as cardio to help their study. CNN algorithm [Awais Mehmood, (2021)] was used for the construction of the model. The proposed model is used to study the probability of the presence of cardiovascular disease in a patient and the model achieved an accuracy of 97%.

Zubar in his studies proposes a hybridized method using KNN and the Spiral optimization method [A. H. Zubar and R. Balamurugan, (2020)] for the classification of cardiovascular diseases. The method enhances the clustering quality and also PCA is used for dimensionality reduction.

The author of the article [Irfan Javid, (2020)] used deep learning methods along with conventional machine learning techniques to predict heart disease. A voting-based method is adopted to strengthen the weak classifiers. LSTM is added along with RNN and is used as a deep neural network. 85.71% accuracy is achieved by the proposed model.

Kalaselvi and G N Nasira analysed the complex relationship between cancer and diabetes in their study. They propose a novel method known as ANIFS [Kalaiselvi and G.N. Nasira, (2014)] in their study. They have achieved 80% accuracy in the PIMA dataset using ANIFIS with adaptive KNN approach.

3.Proposed Methodology

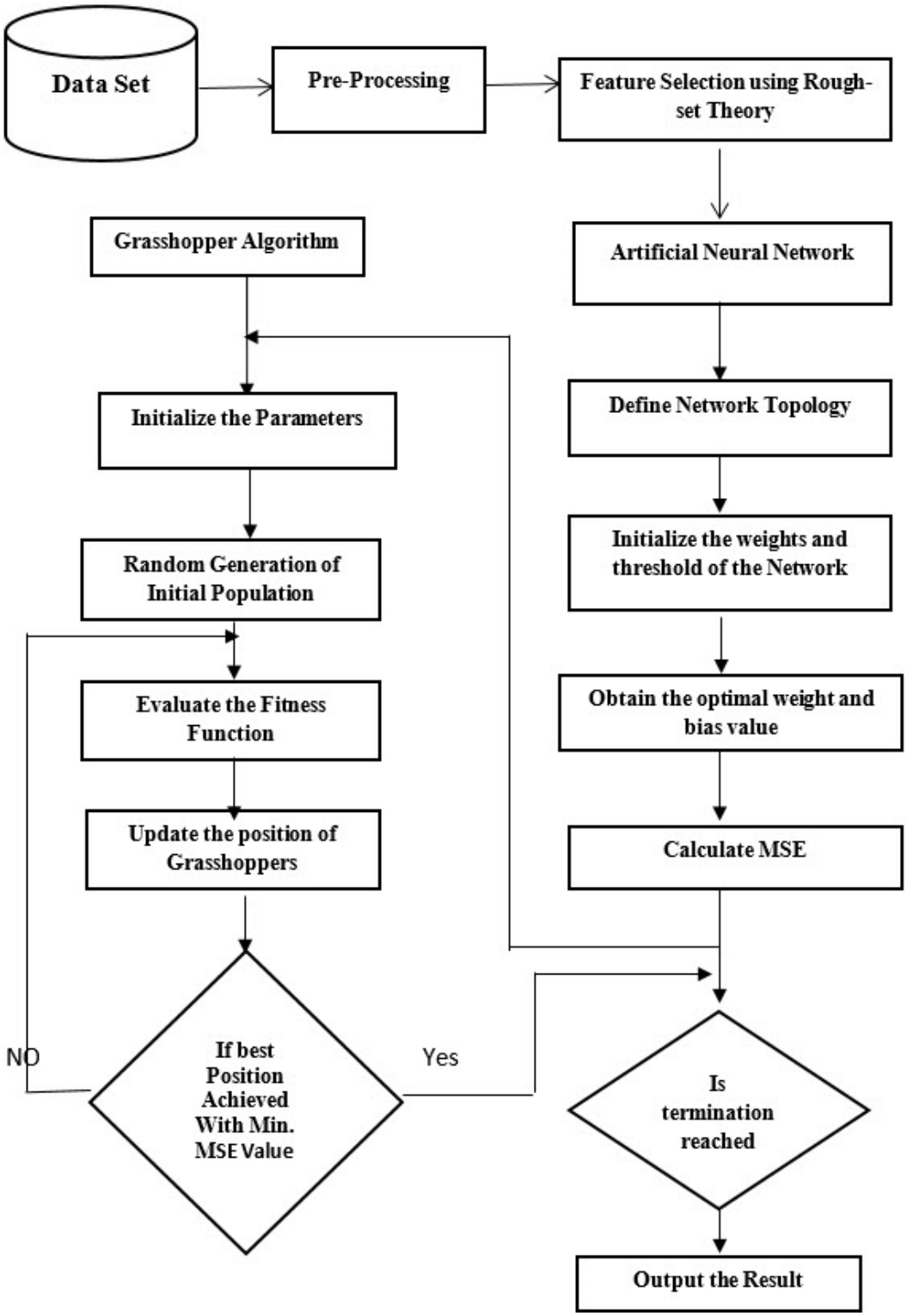

Figure 1: Work Flow of GOA based Artificial Neural Network for Coronary artery disease.

The proposed work consists of three phases. In the first phase, the Rough Set-based feature selection approach finds the most significant subset attributes from the dataset. In the second phase, Artificial Neural Network with improved hidden layers is used for the classification. LSTM is added along with the hidden layers of the Neural Network. In the third phase where a metaheuristic algorithm known as the Grasshopper Optimization algorithm is used for optimizing the model.

By using Rough-set theory the most relevant features were extracted [Youness, (2018)] from the dataset for the classification. The reduced dataset is given as input to the Neural Network. Weights and bias of the hidden layers are initialized by the random initialization methods. Mean Square Error (MSE) is calculated. A single-objective optimization method is followed in the study. Grasshopper Optimization algorithm, optimizes the model, by setting the objective function as MSE.

In the beginning, the initial population is constructed using the random initialization method, and the fitness value for the population is calculated. Positions of grasshoppers are adjusted and the new fitness value is calculated. The process repeated until it received the best fit value. The best fit value is returned to the Neural Network. GOA computes the optimized weights by adjusting the initial parameters and the fitness value for each solution is evaluated. The best fit solution is returned to the Neural Network.

3.1. Rough set-based feature selection

Rough Set Theory (RS) is a mathematical tool [Javad Rahimipour Anaraki, (2013)] that successfully determines the data dependencies among attributes of a dataset and reduces the dataset through structural approach. The features selected by the rough set produce very useful and informative results while other attributes can be removed with less loss of information, as they have a high degree of dependency with the other attributes in the dataset. This will help to reduce redundancy. In this work, the dataset with discretized values of attributes are used to discover the subset of all attributes using the concept of RS, which selects the attributes that provide more information helping to predict the most relevant class attributes accurately.

3.1.1. Fundamental notations for attribute reduction in rough set theory.

In this study, all information is stored in the table format. The dataset contains all the information corresponding to the Information System [Richard Jensen, (2007)]. Assume that the Information System IS = (DS, A) where DS is the set of instances in the dataset A and Att is the non-empty set of finite attributes and att: $DS \rightarrow SV_{att}$ for each att $\in$ Att, where SV_{att} is the set of values that an attribute att may have. With any IP→Att, it has its associated equivalence relation, ie (Independency) which is denoted as IND(IP) [Kanchan Shailendra Tiwari (2013)] as shown in the equation below [Javad Rahimipour Anaraki, (2013)]

$$IND(IP) = \{(u, v) \in DS^2 | \forall att \in IP, att(u) = att(v)\} \quad (1)$$

IND(IP) partitions the instances corresponding to the dataset and it is represented as DS / IND (IP) (or *DS / IP*). If (u, v)) $\in$ IND(IP) then u and v are indiscernible by attribute of IP. In rough set-based feature selection dependency between the attributes is calculated, and if there is interdependency between the attributes exists, these features are removed. So that it can avoid redundancy in the dataset.

The attribute in an information system is classified into different categories using roughest. Decision attributes [Omnia S. Elazab and Hany M. Hasanien, (2020)] decides in which class the attributes belong at the same time. Conditional attributes, do not decide the class of an object, but it will help to decide the class which belongs.

Let Y $\subseteq DS$. Then y can be approximated by calculating the information contained in IP by constructing IP Upper and IP lower approximations [Richard Jensen, (2007)]. IP-lower. and IP-upper approximations [Richard Jensen, (2007)] of Y as formulated as

$$\underline{IP}Y = \{y | [Y]_{IP} \subseteq Y\} \quad (2)$$

$$\overline{IP}Y = \{y | [Y]_{IP} \cap Y \neq \emptyset\} \quad (3)$$

$\underline{IP}Y$ represents the lower approximation and $\overline{IP}Y$ represents the upper approximations. The attribute values belonging to these regions can also be considered in the subset.

Assume that the IP and IQ are equivalence relations over DS, then their positive area can be represented as

$$PSV_{IP}(IQ) = \bigcup_{Y \epsilon DS/IQ} \underline{IP}Y \quad (4)$$

The positive region is comprised of all instances of DS which can be classified to classes of DS/IQ with the information in attributes IP. With this RST defines the degree of dependency of a set of attributes IQ on a set of IP attributes is formulated as For IP, IQ $\subset$ Att, it is signified as IQ depends on IP in a degree l ($0 \leq l \leq 1$) represented as IP→IQ if

$$l = \gamma_{IP}(IQ) = \frac{|PSV_{IP}(IQ)|}{|DS|} \quad (5)$$

The dimensionality of attributes is accomplished by associating equivalence relations produced by sets of attributes. Once the redundant attributes are removed as the reduced set of attributes provides the same predictive ability of decision feature as the original dataset.

3.1.2 Algorithm.

For attribute reduction, an RS-based Quick Reduct [Javad Rahimipour Anaraki, (2013)] Algorithm is implemented. Att is a set of all features, CF is the conditional variable and DF is the decision features [Richard Jensen (2007)] RST denotes the reduced dataset corresponding to the conditional attributes CF.

RST ← {}

Do

T ← RST

$\forall y \in (CF - RST)$

If $(\gamma_{RST \cup \{y\}}(DF) > \gamma_T(DF))$ then

T← RST ∪ {y}

RST ← T

until $\gamma_{RST}(DF) = \gamma_{CF}(DF)$

Return RST

The reduced set of features were computed by the algorithm without exhaustively producing all probable subsets. It begins with an empty set and starts adding attribute one in the incremental order [Javad Rahimipour Anaraki, (2013)], attributes are chosen, which consequence in the greatest increase in rough set dependency, until it generates maximum possible value for the dataset.

3.2 Artificial Neural Network

A kind of computing approach and artificial intelligence which simulates the functioning and analyzing the behavior of the human brain is known as an artificial neural network (ANN). It has the ability of self-learning which enables it to generate better results when the dataset is voluminous. ANN is composed of three different layers. Each layer in turn comprised of processing unit sets. ANN has an input layer, output layer, and hidden layer. Figure2 shows the overall structure of a standard Artificial Neural Network. Each layer nodes are connected to the forwarded layer nodes through the links. The link among neural nodes or perceptrons indicates the stream of information passing from the previous layer node to the forwarding layer node. The neural network function receives the predictor values at the left side and the computations are done on the hidden layer and the output layer shows the predicted outcome.

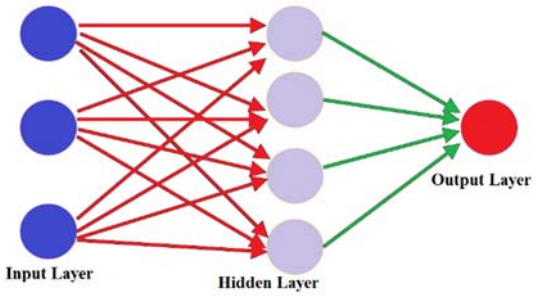

Figure2: General Structure of an Artificial Neural Network

The Figure 3 depicts the weight and bias assignment for each Neuron in an Artificial Neural Network. The hidden nodes receive the value from the previous layer and which is multiplied by weights, whose values are predetermined. The weighted input values received from all the nodes of the previous layer are summed together to generate a single value as shown in Figure 3. Next, an activation function known as ReLU is applied to the summed values which control the output of the nodes. Same as the hidden layer, the output layer's active node also does the relevant computations and produces the output. If the output value is binary then the sign of positive represented the presence and negative represented the absence which is highly dependent on the input data. The input nodes are passive, its function is to receive the input from the outside world and pass it to the next layer,

the hidden layer performs the multiplication of the received value with its concern weight assigned on each.

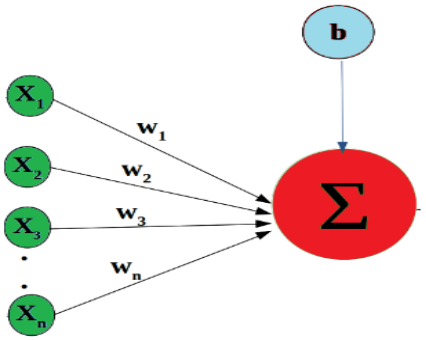

Figure 3: Assignment of weights and bias [20] of a Neuron in an Artificial Neural Network.

The weight for an input corresponding to an individual node is calculated as in the Equation (6)

$$netwrk_j = \sum_{i=1}^{n} w\, gt_{ij}.\, O_i \qquad (6)$$

where wgt_{ij} signifies weights between two nodes i and j, and output from the unit 'i' is denoted as O_i. To generate the final output 'O_t' the sum is passed to the nonlinear filter φ called activation function which generates the output given in Equation (7).

$$O_t = \varphi(netwrk_j) \qquad (7)$$

where ϕ refers to activation function and the weighted sum of input values prior to the signal processes to the subsequent layer. Finally, the generated output is compared with the expected output and their difference is assigned as an error. The standard neural network uses the backpropagation method to update the weights on trial and error basis repeatedly until the accepted error rate is obtained. This result determining the appropriate weights to produce more accurate results. The error rate is defined in equation (8)

$$\text{Error} = \frac{1}{2}\sum_k (desireout_k - resultout_k) \qquad (8)$$

3.3 Grass Hopper Optimization Algorithm.

The Grasshopper Algorithm is a bio inspired metaheuristic method which mimics the life cycle of grasshopper insects. The life cycle of grasshopper consists of two different stages: Nymph and Adult. Even though the grasshoppers are seen individually, they will foam large swarms during their Nymph and Adult stage. The movement of this swam during the Nymph stage is very slow compared to that of the adult stage. At the adult stage, they will show long-range abrupt movements [Ali Asghar Heldari, (2018)].

Based on these movements of grasshoppers, the search process in the GOA algorithm usually follows two methods: exploration and exploitation. During exploration, an abrupt movement is shown by the search agents in exploration [Peng Chen and Xiaoqiang Zhao, (2020)] but a local movement is followed during exploitation. The swarming [Ali Asghar Heldari, (2018)] behavior of Grasshoppers can be mathematically expressed in the formulae given below:

$$Y_j = S_j + G_j + W_j \qquad (9)$$

Where Y_j defines the position [Ali Asghar Heldari, (2018)] of the individual grasshopper" j". S_j defines the social interaction of the j^{th} individual, G_j defines the gravitational force of the j^{th} individual and W_j defines the advection. The social interaction between the grasshopper is one of the main parameters for the search mechanism. It can be computed by the mathematical formulae,

$$S_i = \sum_{j=1\ j\#i}^{N} S(A_{ij})\hat{A}_{ij} \qquad (10)$$

A_{ij} is the distance between [Ahmed A. Ewees and Zhang Jianhua (2020)] the grasshopper i and j which can be mathematically expressed as $A_{ij} = | Y_i - Y_j |$. $\hat{A}_{ij}$ is a singular vector between grasshopper i and j and is mathematically denoted as

$$\hat{A}_{ij} = (Y_j - Y_i)/ A_{ij} \qquad (11)$$

The social force [Omnia S. Elazab and Hany M. Hasanien, (2020)] between grasshoppers are mathematically calculated as

$$S(r) = f e^{(-r/l)} - e^{-r} \quad (12)$$

The intensity of gravitation is represented by f and l denotes the length of the social interaction. The impact of the interaction between the grasshoppers is denoted as S. A parameter C represents the mathematical formulation that can be used for balancing the movements [Omnia S. Elazab and Hany M. Hasanien, (2020)] of search agents.

$$C = (\text{Maximum Iteration}) - (\text{Current Iteration}) \frac{\text{Maximum Iteration} - \text{Mininmum Iteration}}{Total\ No\ of\ Iteration} \quad (13)$$

3.3.1Fitness Evaluation.

The model follows a single objective optimization method. For finding the best fit solution, the fitness value of each solution is calculated. MSE is taken as the parameter for fitness evaluation. The experiment is repeated for 1000 iterations and the best fit value shows minimum MSE. MSE can be calculated using the Equation (14)

$$MSE = \frac{1}{n} (r_j - \hat{r})^2 \quad (14)$$

Where '**r**' represents the actual result and '**r̂**' represents the predicted result, the total number of instances is denoted as 'n'. The experiment is repeated for 1000 iterations and the best fit value is taken as the one with minimum MSE value. The fitness value corresponding to different iterations is plotted in Figure 4.

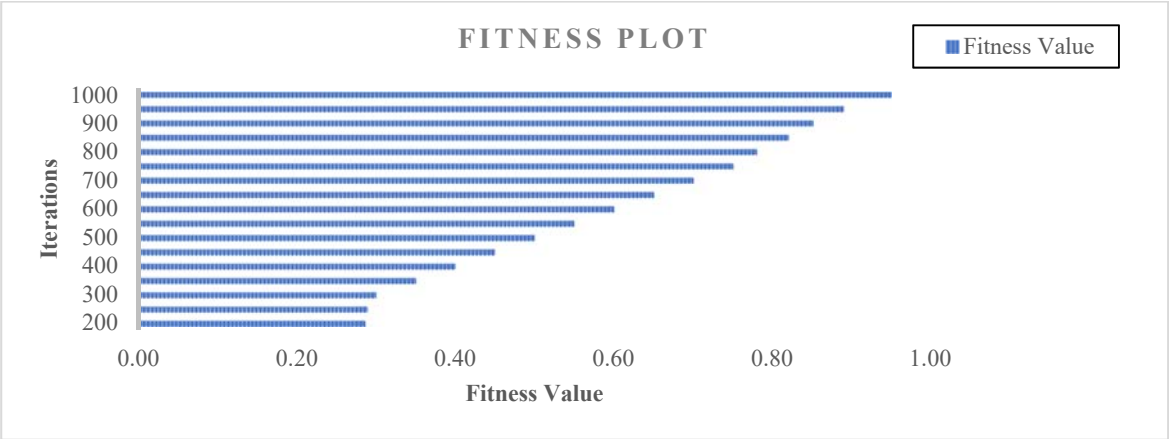

Figure 4: Fitness Plot generated by the Grasshopper Algorithm

4. Experimental Results and Discussions

4.1 Dataset description

Framingham Dataset from Kaggle Machine Learning Repository is used for the study. The dataset is comprised of 4240 instances with 16 attributes. They are gender information, age of the person, education of the person, current Smoker or not, cigarettes Per Day, Blood Pressure, prevalent Stroke, Prevalent Hyper Tension, diabetes, total Cholesterol, Systolic blood pressure, Diastolic blood pressure, Body Mass Index, heartbeat Rate, Blood glucose level with the class attribute Chronic Heart Disease present or not.

To predict diabetic heart disease, a new class variable is generated to label the patients in four different risk levels such as without diabetic without heart disease as 0, presence of diabetic and absence of heart disease as 1, absence of diabetic, and presence of heart disease as 2 and presence of both diabetes and heart disease as 3. Inter Quartile Range is used to preprocess the dataset in this study. Ambiguous and erroneous data is replaced by the mathematical and statistical approach used in IQR.

4.2 Results & discussions

This work is implemented with the help of two different class variables in the dataset, they are responsible for generating the new class variable which determining the presence or absence of diabetes and the consequent heart failure possibilities among the patients. A new class variable is generated based on the values of these two class variables.

The proposed model RS-GANN is designed using Python and its performance is examined and compared with other prediction models like Naive Bayes and Logistic Regression. The parameters used for evaluating the performance of the model are Precision, Recall, F1-Score, Specificity and, Accuracy. The results are discussed in Table 1. The comparison of results is represented in a graphical format using Figure5. The proposed model is compared with the model before optimization (RS-ANN). Comparison results are shown in Table2 and Figure 6. The mathematical formulae for calculating the performance measurement parameters are given below.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive+False\ Positive}. \quad (15)$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive+False\ Negative} \quad (16)$$

$$\text{Specificity} = \frac{True\ Negative}{True\ Negative+False\ Positive} \quad (17)$$

F1 Score = 2*((Precision* Recall) / (Precision +Recall)). (18)

$$\text{Accuracy} = \frac{True\ Positive+True\ Negative}{True\ Positive+True\ Negative+FAlse\ Positive+False\ Negative} \quad (19)$$

Algorithm	Precision	Recall	Specificity	F1-Score	Accuracy
Navie Bayes	97.8	79.9	84.5	88.0	80.41
Logistic Regression	98.3	86.1	87.4	91.8	86.28
RS-GANN	99.1	95.6	92.2	97.3	95.25

Table1: showing the Performance of Comparison of Classifiers

The rate of positive predicted value can be explained by the parameter Precision. The recall is the result of the analysis which explains the percentage of relevant values predicted by the model. Specificity measures the ability of a classifier to identify non-disease individuals. The weighted average of Precision and Recall can be represented as the F1-Score of the Classifier. The accuracy of the model is computed by the mathematical formulae given in equation (19).

Algorithm	Precision	Recall	Specificity	F1 -Score	Accuracy
RS-ANN (Before Optimization)	98.8	88.9	90.9	93.6	89.1
Proposed Model	99.1	95.6	92.2	97.3	95.25

Table 2: Comparison of Classifier Performance

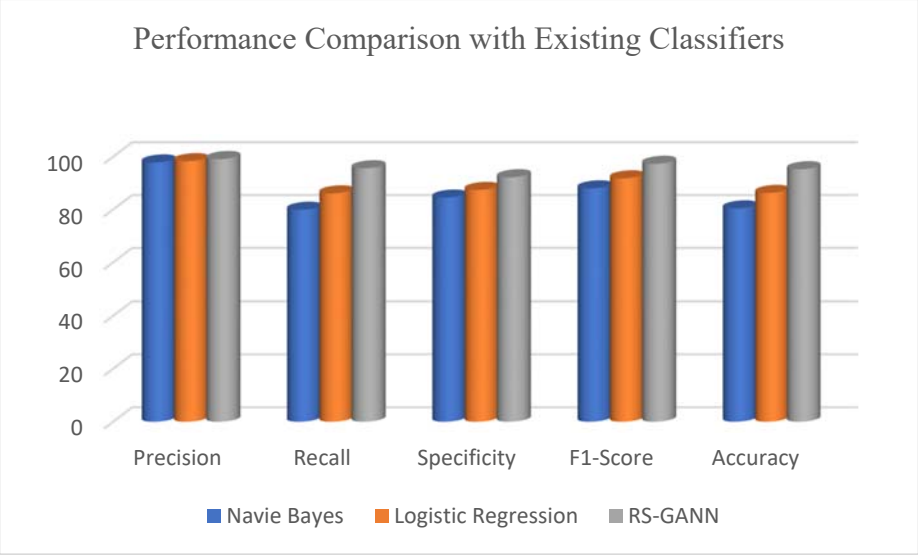

Figure 5: Performance Comparison with Existing Classifiers

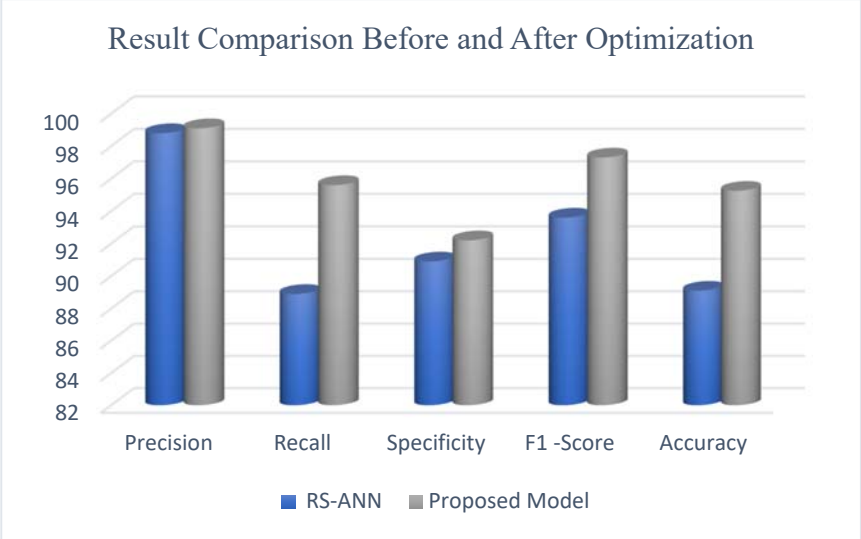

Figure 6: Comparison of Result Before and After Grasshopper Optimization

5. Conclusion.

In this study, a novel soft computing model is proposed for heart disease prediction among diabetes patients based on Rough set-based feature selection with an Artificial Neural Network optimized using a metaheuristic stochastic algorithm known as grasshopper optimization. Evaluation of results is plotted in tabular as well as in the graphical format. The results revels that the proposed system shows better performance than the existing algorithms. After executing the proposed methodology, it is found that the classifier shows an improvement of 6.15% accuracy, 0.3% in precision, 6.7% recall, 3.7% in F1-Score, and 1.3% in Specificity while compared with the performance before optimization and that is a significant improvement.

References

[1] Javad Rahimipour Anaraki, Mahdi Eftekhari(2013): Rough set-based feature selection: A Review IEEE, The 5th Conference on Information and Knowledge Technology.
[2] Ashir Javeed, Sanam Shahla Rizvi, Shijie Zhou, Rabia Riaz, Shafqat Ullah Khan, and Se Jin Kwon (2020): Heart Risk Failure Prediction Using a Novel Feature Selection Method for Feature Refinement and Neural Network for Classification. Personal Communication Technologies for Smart Spaces, Hindawi.
[3] Afzal Hussain Shahid, M.P. Singh (2020): A Novel Approach for Coronary Artery Disease Diagnosis using Hybrid Particle Swarm Optimization based Emotional Neural Network. IEEE.
[4] Janmenjoy Nayak, Bighnaraj Naik, Paidi Dinesh, Kanithi Vakula, Pandit Byomakesha Dash2(2020): FireflyX Algorithm in Biomedical and Health Care: Advances, Issues and Challenges. springer, SN Computer Science.
[5] Youness Khourdifi, Mohamed Bahaj (2018): Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization. International Journal of Intelligent Engineering & Systems.
[6] S. Nithyavishnupriya, P. Ramprakash,, Sarumathi (2020): "Disease Prediction Using Deep Neural Network ", IEEE.
[7] Archana Singh, Rakesh Kumar (2020). Heart Disease Prediction Using Machine Learning Algorithms. International Conference on Electrical and Electronics Engineering (ICE3-2020), ieeexplore.ieee.org / document/9122958.
[8] Rahul Katarya, Sunit Kumar Meena (2020): Machine Learning Techniques for Heart Disease Prediction A Comparative Study and Analysis. Health and Technology, Springer Nature.
[9] Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon Shah Nazir and Ruinan Sun (2018): A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms. Wearable Technology and Mobile Applications for Healthcare, Hindawi.
[10] Richard Jensen, Qiang Shen (2007): Rough Set-Based Feature Selection: A Review. ResearchGate.
[11] Kanchan Shailendra Tiwari, Ashwin Kothari (2013). Design and Implementation of Rough Set Algorithms on FPGA: A Survey. International Journal of Advanced Research in Artificial Intelligence, 2013.
[12] Awais Mehmood, Munwar Iqbal, Zahid Mehmood, Aun Irtaza, Marriam Nawaz, Tahira Nazir & Momina Masood (2021): Prediction of Heart Disease Using Deep Convolutional Neural Networks. Springer.
[13] A. H. Zubar, R. Balamurugan(2020): Green Computing Process and its Optimization Using Machine Learning Algorithm in Healthcare Sector. Mobile Networks and Applications, Springer.
[14] Irfan Javid, Ahmed Khalaf Zager Alsaedi, Rozaida Ghazali (2020): Enhanced Accuracy of Heart Disease Prediction using Machine Learning and Recurrent Neural Networks Ensemble Majority Voting Method. (IJACSA) International Journal of Advanced Computer Science and Applications.
[15] C. Kalaiselvi, Dr. G.M. Nasira(2014): A New Approach for Diagnosis of Diabetes and Prediction of Cancer using ANFIS. 2014 World Congress on Computing and Communication Technologies.
[16] Peng Chen, Xiaoqiang Zhao, Qixian Zhu (2020): A novel classification method based on ICGOAKELM for fault diagnosis of rolling bearing. Applied Intelligence.
[17] Omnia S. Elazab, Hany M. Hasanien, Ibrahim Alsaidan, Almoataz Y. Abdelaziz, S. M. Muyeen (2020): Parameter Estimation of Three Diode Photovoltaic Model Using Grasshopper Optimization Algorithm. Energies.
[18] Ali Asghar Heldari, Hossam Faris (2018): An Efficient Hybrid Multilayer Perceptron Neural Network with Grasshopper Optimization. Soft Computing, Springer.

[19] Ahmed A. Ewees, Zhang Jianhua (2020): Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price volatility, Resources Policy, Elsevier.
[20] https://towardsdatascience.com/whats-the-role-of-weights-and-bias-in-a-neural-network-4cf7e9888a0f

Authors Profile

Ambily Merlin Kuruvilla is pursuing her research in Computer Science in the area of "Data Mining and Machine Learning" from, Karpagam Academy of Higher Education, Coimbatore. She holds a Master Degree in Computer Applications from Sri Ramakrishna College of Arts and Science for Women affiliated to Bharathiar University, Coimbatore, and M.Phil from Vinayaka Missions University, Salem. She has more than 19 years of graduate teaching experience and has published research papers in Journals and Proceedings of International repute.

Dr. N.V. Balaji is a highly acclaimed academician with dedication and commitment towards education. As a veteran educationalist for more than a decade he renders his service as an academic contriver with zeal for 21 years. He enriched training and placements with immense enthusiasm scaling the academy to great heights. Dr. Balaji a pertinacious personality brought laurels upon the institutions by representing it at Cambridge University for Business English Certifications. He is a honorable recipient of the award of Ambassador for Computer based Learning and Assessment category in the year 2015.

Adding to his reputation he has accumulated 15 years of research experience in the field of computer science. He exhibits deep interest towards various genres such as Neural Networks, Fuzzy Logic, Image Processing, Classification and Data Mining. Embellishing the annals of education at Karpagam institutions, he has been an influential face between prime IT firms namely Infosys, Wipro, Cognizant and Zoho. Through his constant coordination with competent companies, he incorporated industry collaborated electives in the curriculum of computer science.

He is meticulously marching towards success through significant strategies of prior planning and precise amalgamation of academics with practical experience. Emphasizing industry-based syllabus, he evolved a new theory of learning in order to elevate education as a wholesome experience. As a part of his educational expedition, he has visited United Kingdom and Israel for academic assignments. He is committed to the continuous growth of the institution by presently serving as a Dean of Arts, Science and Humanities at KAHE.