GUI BASED PREDICTION OF DIABETIC STAGES ACCURATELY USING MACHINE LEARNING APPROACH

S. Devika[#1], V. Gopikamani[#2], S. Mahima[#3], Mrs. k. Rejini[*4]

[#1, #2, #3] B.E, IV year, Department of Computer Science and Engineering
*[*4]Assistant Professor, Department of Computer Science and Engineering*
Anand Institute of Higher Technology, Chennai

ABSTRACT

In 2040, the world's diabetic patients will reach 642 million, which means that one`s of the ten adults in the future is suffering from diabetes. Diabetes Mellitus (DM) is defined a group of metabolic disorders exerting significant pressure on human health worldwide. DM is a chronic disease characterized by hyperglycaemia and it may cause many complications. To prevent this problem, to analyse the given hospital dataset by supervised machine learning technique (SMLT) while capturing several information's like variable identification uni-variate analysis, bi-variate analysis, multi-variate analysis, missing value treatments and analyse the data validation, data cleaning/preparing and data visualization. Our analysis provides a comprehensive guide to sensitivity analysis of model parameters with regard to performance in prediction of diabetic patients by given attributes of dataset with evaluation of GUI based user interface diabetes attribute prediction. Additionally, it observes to lead to an increase in the highest accuracy in diabetic prediction of attributes using a better classification report, identifying the confusion matrix and categorizing data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with the best accuracy with precision, Recall and F1 Score.

Keywords: *Dataset, Diabetes prediction, SMLT, Machine learning-Classification method, python, Prediction of Accuracy result.*

I. INTRODUCTION

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Process of training and prediction involves use of specialized algorithms. It feeds the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. Machine learning can be roughly separated in to three categories. There are supervised learning, unsupervised learning and reinforcement learning. Supervised learning program is both giving the input data and the corresponding labeling to learn data that has to be labeled by a human being beforehand. Unsupervised learning has no labels. It is provided to the learning algorithm. This algorithm has to figure out the clustering of the input data. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or negative feedback to improve its performance.

Data scientists use many different kinds of machine learning algorithms to discover patterns in python that lead to actionable insights. At a high level, these different algorithms can be classified into two groups based on the way they "learn" about data to make predictions: supervised and unsupervised learning. Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function from input variables(X) to discrete output variables(y). In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too. Some examples of classification problems are: speech recognition, handwriting recognition, bio metric identification, document classification etc. Supervised Machine Learning is the majority of practical machine learning. Supervised learning is where we have input variables (X) and an output variable (y) and use an algorithm to learn the mapping

function from the input to the output **is y = f(X).** The goal is to approximate the mapping function so well that when there is new input data (X) that can be predict the output variables (y) for that data. Techniques of Supervised Machine Learning algorithms include **logistic regression**, **multi-classes**, **DecisionTrees** and **support vector machines etc**. Supervised learning requires that the data used to train the algorithm is labeled with correct answers.

Preparing Dataset

The dataset is now supplied to machine learning model based on this data set the model is trained. In the first step of accumulating information, data from previously patients datasets from online sources are gathered together (like, www.kaggle.com / www.data.gov.in). These datasets are merged to form a common dataset, on which analysis will be done.

Table shows details of the datasets:

Variable	**Description**
Pregnancies	Number of times pregnant
Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test
Blood Pressure	Diastolic blood pressure (mmHg)
Skin Thickness	Triceps skin fold thickness (mm)
Insulin	2-Hour serum insulin (mu U/ml)
BMI	Body mass index (weight in kg/(height in m)^2)
Diabetes Pedigree Function	Diabetes pedigree function
Age	Age in years
Outcome	Class variable (0 or 1) 268 of 768 are 1, the others are 0

II. DESIGN ARCHITECTURE

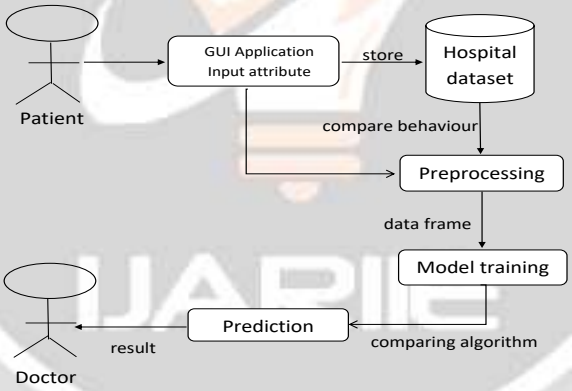

Fig 2.1 Overall Design

III. ALGORITHM AND TECHNIQUES

3.1 Software Description

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system "Conda".**Anaconda Navigator** is a desktop graphical user interface (GUI) included in Anaconda distribution that allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository, install them in an environment, run the packages and update them. It is available for Windows, macOS and Linux.

The Jupyter Notebook :The Jupyter Notebook is an open-source web application that allows to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Used packages are

sklearn:

- In python, sklearn is a machine learning package which include a lot of ML algorithms.
- Here, we are using some of its modules like train_test_split, DecisionTreeClassifier or Logistic Regression and accuracy_score.

NumPy:

- It is a numeric python module which provides fast maths functions for calculations.
- It is used to read data in numpy arrays and for manipulation purpose.

Pandas:

- Used to read and write different files.
- Data manipulation can be done easily with data frames.

Matplotlib:

- Data visualization is a useful way to help with identify the patterns from given dataset.
- Data manipulation can be done easily with data frames.

IV. IMPLEMENTATION

Modules:

1. Data validation process and preprocessing
2. Exploration data analysis of visualization process
3. Performance of logistic regression and decision tree algorithms
4. Performance of random forest and support vector machine
5. Performance of k-nearest neighbor and Naive Bayes
6. GUI based prediction results of Diabetes in the form of best accuracy

4.1 Data validation process and preprocessing

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, There is no need of the validation techniques. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers uses this data to fine-tune the model hyper parameters. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time-consuming to-do list. During the process of data identification, it helps to understand data and its properties; this knowledge will help to choose which algorithm to use to build model.

4.1.1 Data Validation/ Cleaning/Preparing Process:

Importing the library packages with loading given dataset. To analyzing the variable identification by data shape, data type and evaluating the missing values, duplicate values. A validation dataset is a sample of data held back from training model that is used to give an estimate of model skill while tuning model's and procedures that can be used to make the best use of validation and test datasets when evaluating models. Data cleaning / preparing by rename the given dataset and drop the column etc. to analyze the uni-variate, bi-variate and multi-variate process. The steps and techniques for data cleaning will vary from dataset to dataset. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making.

4.2 Data analysis of visualization process

Data visualization is an important skill in applied statistics and machine learning. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Many machine learning algorithms are sensitive to the range and distribution of attribute values in the input data. Outliers in input data can skew and mislead the training process of machine learning algorithms resulting in longer training times, less accurate models and ultimately poorer results.Even before predictive models are prepared on training data, outliers can result in misleading representations and in turn misleading interpretations of collected data. Outliers can skew the summary distribution of attribute values in descriptive statistics like mean and standard deviation and in plots such as histograms and scatterplots, compressing the body of the data. Finally, outliers can represent examples of data instances that are relevant to the problem such as anomalies in the case of fraud detection and computer security. It couldn't fit the model on the training data and can't say that the model will work accurately for the real data. For this, we must assure that our model got the correct patterns from the data, and

it is not getting up too much noise. Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.

Data Visualization types:

- Introduction to Matplotlib
- Line Plot
- Bar Chart
- Histogram Plot
- Box and Whisker Plot
- Scatter Plot

Bar Chart: A bar chart is generally used to present relative quantities for multiple categories. The x-axis represents the categories and are spaced evenly. The y-axis represents the quantity for each category and is drawn as a bar from the baseline to the appropriate level on the y-axis.

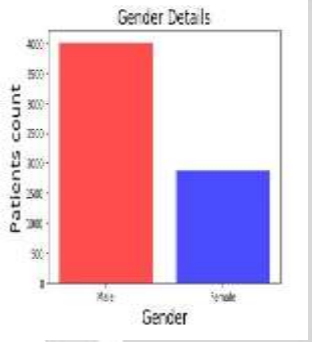

Fig 4.2.1 Gender info of dataset

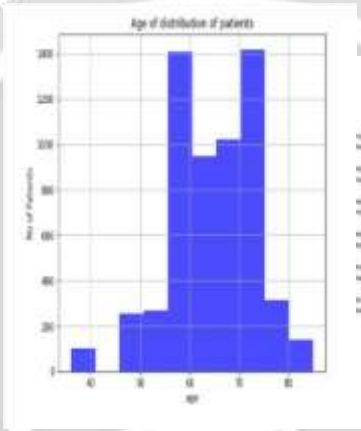

Fig 4.2.2 Age distribution of patients

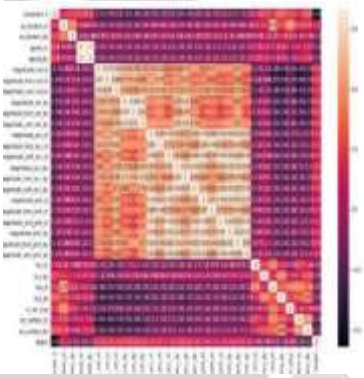

Fig 4.2.3 Heat map diagram

Box and Plot: A box and whisker plot, or boxplot for short, is generally used to summarize the distribution of a data sample. The x-axis is used to represent the data sample, where multiple boxplots can be drawn side by side on the x-axis if desired. The boxplot is a graphical technique that displays the distribution of variables. It helps us see the location, skewness, spread, tile length and outlying points. The boxplot is a graphical representation of the Five Number Summary.

Heat map: A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. It is a bit like looking a data table from above. It is really useful to display a general view of numerical data, not to extract specific data point. It is quite straight forward to make a heat map, as shown on the examples below.

4.3 Performance of logistic regression and decision tree algorithms

Logistic Regression: It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).

Decision Tree: It is one of the most powerful and popular algorithm. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables. Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and

smaller subsets while at the same time an associated decision tree is incrementally developed. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. Decision tree builds classification or regression models in the form of a tree structure. It utilizes an if-then rule set which is mutually exclusive and exhaustive for classification. The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covered by the rules are removed.This process is continued on the training set until meeting a termination condition. It is constructed in a top-down recursive divide-and-conquer manner. All the attributes should be categorical. Otherwise, they should be discretized in advance. Attributes in the top of the tree have more impact towards in the classification and they are identified using the information gain concept.

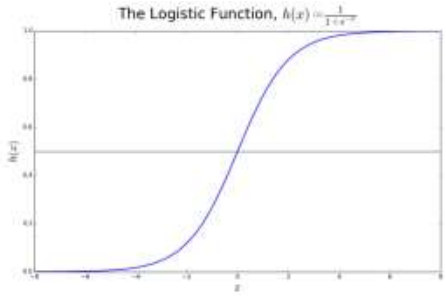

Fig 4.3.1 Logistic Regression function graph

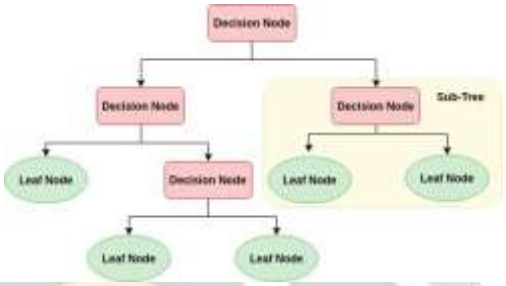

Fig 4.3.2 Decision tree

4.4 Performance of random forest and support vector machine

Random Forest: Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where different types of algorithms or same algorithm multiple times are joined to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

Support Vector Machines: A classifier that categorizes the data set by setting an optimal hyper plane between data. I chose this classifier as it is incredibly versatile in the number of different kernelling functions that can be applied and this model can yield a high predictability rate. Support Vector Machines are perhaps one of the most popular and talked about machine learning algorithms. They were extremely popular around the time they were developed in the 1990s and continue to be the go-to method for a high-performing algorithm with little tuning**.**

Fig 4.4.1 Random forest

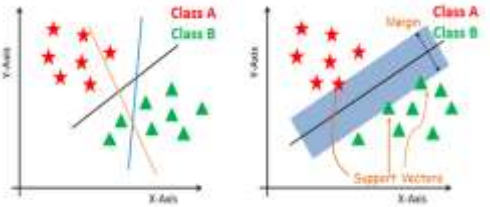

Fig 4.4.2 Support vector machine

4.5 Performance of k-nearest neighbor and Naive Bayes

K-Nearest Neighbor (KNN): K-Nearest Neighbor is a supervised machine learning algorithm which stores all instances correspond to training data points in n-dimensional space. When an unknown discrete data is received, it analyzes the closest k number of instances saved (nearest neighbors) and returns the most common class as the prediction

and for real-valued data it returns the mean of k nearest neighbors. In the distance-weighted nearest neighbor algorithm, it weights the contribution of each of the k neighbors according to their distance using the following query giving greater weight to the closest neighbors. Usually KNN is robust to noisy data since it is averaging the k-nearest neighbors. The k-nearest-neighbors algorithm is a classification algorithm, and it is supervised: it takes a bunch of labeled points and uses them to learn how to label other points. To label a new point, it looks at the labeled points closest to that new point. Makes predictions about the validation set using the entire training set. KNN makes a prediction about a new instance by searching through the entire set to find the k "closest" instances. "Closeness" is determined using a proximity measurement (Euclidean) across all features.

Naive Bayes algorithm: The Naive Bayes algorithm is an intuitive method that uses the probabilities of each attribute belonging to each class to make a prediction. It is the supervised learning approach would come up with yhe want to model a predictive modeling problem probabilistically. Naive bayes simplifies the calculation of probabilities by assuming that the probability of each attribute belonging to a given class value is independent of all other attributes. This is a strong assumption but results in a fast and effective method. The probability of a class value given a value of an attribute is called the conditional probability. By multiplying the conditional probabilities together for each attribute for a given class value, we have a probability of a data instance belonging to that class. To make a prediction we can calculate probabilities of the instance belonging to each class and select the class value with the highest probability. Naive Bayes is a statistical classification technique based on Bayes Theorem. It is one of the simplest supervised learning algorithms. Naive Bayes classifier is the fast, accurate and reliable algorithm. Naive Bayes classifiers have high accuracy and speed on large datasets. Naive Bayes classifier assumes that the effect of a particular feature in a class is independent of other features.

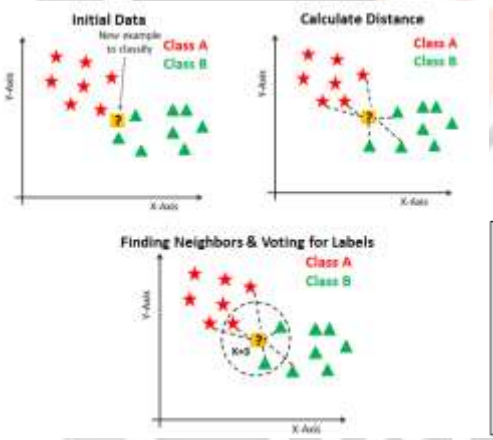

Fig 4.5.1 K - Nearest Neighbours

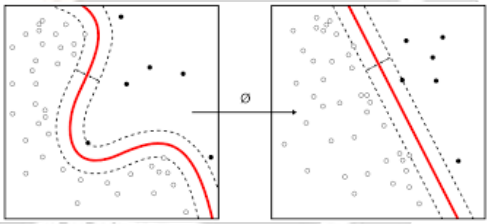

Fig 4.5.2 Naive Bayes

4.6 GUI based prediction results of Diabetes in the form of best accuracy

Tkinter is a python library for developing GUI (Graphical User Interfaces). We use the tkinter library for creating an application of UI (User Interface), to create windows and all other graphical user interface and Tkinter will come with Python as a standard package, it can be used for security purpose of each users.

False Positives (FP): When actual class is no and predicted class is yes.

False Negatives (FN): When actual class is yes but predicted class in no.

True Positives (TP): These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes.

True Negatives (TN): These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no.

The positive data points predicted as positive will be considered true positive, the negative data points predicted as negative will be considered true negative, and the respective counterparts will be considered false positive and false negative.

Accuracy - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

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

Precision - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

Precision =(TP) / (TP+FP)

Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class .

Recall = (TP) / (TP+FN)

F1 score - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

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

Sensitivity: Sensitivity is a measure of the proportion of actual positive cases that got predicted as positive (or true positive). Sensitivity is also termed as Recall. This implies that there will be another proportion of actual positive cases, which would get predicted incorrectly as negative (and, thus, could also be termed as the false negative). This can also be represented in the form of a false negative rate. The sum of sensitivity and false negative rate would be 1.Mathematically, sensitivity can be calculated as the following:

Sensitivity = (True Positive) / (True Positive + False Negative)

The following is the details in relation to True Positive and False Negative used in the above equation.

- True Positive = Persons predicted as suffering from the disease (or unhealthy) are actually suffering from the disease (unhealthy).
- False Negative = Persons who are actually suffering from the disease (or unhealthy) are actually predicted to be not suffering from the disease (healthy).
- The higher value of sensitivity would mean higher value of true positive and lower value of false negative. The lower value of sensitivity would mean lower value of true positive and higher value of false negative.

Specificity: Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative). This implies that there will be another proportion of actual negative, which got predicted as positive and could be termed as false positives. This proportion could also be called a false positive rate. The sum of specificity and false positive rate would always be 1.Mathematically, specificity can be calculated as the following:

Specificity = (True Negative) / (True Negative + False Positive)

The following is the details in relation to True Negative and False Positive used in the above equation.

- True Negative = Persons predicted as not suffering from the disease (or healthy) are actually found to be not suffering from the disease (healthy).
- False Positive = Persons predicted as suffering from the disease (or unhealthy) are actually found to be not suffering from the disease (healthy).
- The higher value of specificity would mean higher value of true negative and lower false positive rate. The lower value of specificity would mean lower value of true negative and higher value of false positive.

V. RESULT

According to the results shown in Table5.1 and 5.2 , we can figure out that the proposed model has better accuracy than other classification models for Type 2 diabetes in the related studies. Comparing with the above results, it is clear to see the proposed model obtains quite promising results in classifying the possible Type 2 diabetes patients. With the rapidly growing demand for medical data analysis, the proposed model can be fairly useful to the researchers and doctors for their decision-making on the patients as by using such an efficient model they can make more accurate decisions.
The result obtained from the above Supervised machine learning algorithm accuracy comparison is with the best performance accuracy score of 96.15% using the K-Nearest Neighbor along with the dataset splitted into 70% of training set and 30% set of test set. The Table 5.3 gives the sensitivity and specificty score of Logistic Regression, Decision tree,

Random forest, Support vector machine, Naive bayes and K-Nearest neighbor algorithms. The Fig 5.4 is the Bar chart analysis of the comparison chart of the above mentioned six Supervised Machine Learning techniques and their Diabetes prediction accuracies.

Table comparison of affected case & not affected case :

Algorithm	Precision	Recall	F1-Score	Accuracy (100%)
LR	0.79	0.85	0.82	92.0
DT	0.72	0.75	0.74	80.77
RF	0.78	0.89	0.83	88.0
SVC	0.65	1.00	0.79	66.66
NB	0.81	0.81	0.81	92.31
KNN	0.76	0.83	0.79	96.15

Table.5.2 Table for Unaffected case

Algorithm	Precision	Recall	F1-Score	Accuracy (100%)
LR	0.68	0.58	0.63	92.0
DT	0.51	0.47	0.49	80.77
RF	0.72	0.54	0.62	88.0
SVC	0.00	0.00	0.00	66.66
NB	0.65	0.64	0.65	92.31
KNN	0.61	0.51	0.55	96.15

Table 5.1 Table for Affected case

Sensitivity & Specificity Analysis:

Algorithm	Sensitivity	Specificity
LR	0.85	0.58
DT	0.75	0.46
SVC	1.00	0.00
RF	0.88	0.53
NB	0.81	0.64
KNN	0.82	0.50

Table 5.3 Sensitivity and Specificity chart

Graph Analysis:

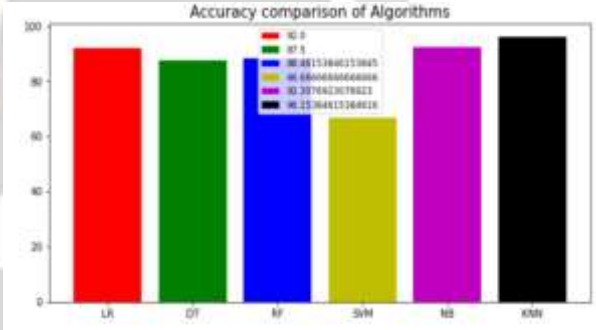

Fig 5.4 Graph Analysis of Accuracy rate

VI. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. Finding the patient stages and grade with parameter like accuracy, classification report and confusion matrix on public test set of given attributes by Supervised machine learning algorithm method. There are Various data mining method and its application were studied or reviewed application of machine learning algorithm were applied in different medical data sets including machine Diabetes dataset. Machine learning methods have different power in different data set. We obtained 768record diabetes data set from UCI the comparison of individual algorithm and the proposed method is done on this study.In this study the proposed method provide high accuracy with accuracy value of 96.15% with the help of K- Nearest Neighbor algorithm. This proposal exceeds the previous work of finding the hospital admission for avoidable causes of Diabetes Mellitus obtained by LDA balanced and SVM balanced both with 92% accuracy and the RNN with a balanced dataset for the prediction of incidence of atleast one disease with an accuracy of 94.6%, 7.4% above LDA balanced and SVM weighted.

FUTURE ENHANCEMENT

- Indian hospitals wants to automate the detection of patients whether affected with diabetes or not with the help of given attributes from the eligibility process (real time) & also to automate this process by showing the prediction result in web application or desktop application in future.
- To optimize the work to implement in Artificial Intelligence environment.

REFERENCE

[1]. Ali Al-Matouq,Taous-meriem Laleg-Kirati,CarloNovara,Ivana Rabbone , Tyrone Vincent, "Sparse Reconstruction of Glucose Fluxes using Glucose Monitors ",IEEE/ACM Transaction on computational Biology and Bioinformatics.

[2]. Yunlei Sun, Dalin Zhang, "Diagnosis and Analysis of Diabetic Retinopathy based on Electronic Health Records" .College of Computer & Communication Engineering, China University of Petroleum(East China) Qingdao, 266580, China National Research Center of Railway Safety Assessment,BeijingJiaotongUniversity,Beijing,100044, China. DOI 10.1109/ACCESS.2019.2918625,IEEE Access

[3]. Rammah Abohtyra, YossiChait,M.J. Germain, C.V. Hollot, Fellow, IEEE, and Joseph Horowitz."Individualization of Ultrafiltration in Hemodialysis",DOI 10.1109/TBME 2018.2884931 IEEE Transactions on biomedical Engineering.

[4]. Debopriya Ghosh, Javier Cabrera,Tarek N. Adam, Petros Levounis and Nabil R. AdamRutgers University, NYP/Weill Cornell Medical Center."Comorbidity Patterns and its Impact on Health Outcomes: Two-way Clustering Analysis", IEEE Transactions on Big Data.

[5]. S.Benbelkacem and B. Atmani, "Random Forests for Diabetes Diagnosis," 2019 International Conference onComputer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-4.

[6]. W. Chen, S. Chen, H. Zhang and T. Wu, "A hybrid prediction model for type 2 diabetes using K-means and decision tree," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 386-390.

[7]. H. Kaur and S. Batra, "HPCC: An ensembled framework for the prediction of the onset of diabetes," 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, 2017, pp. 216-222.

[8]. R. Vaishali, R. Sasikala, S. Ramasubbareddy, S. Remya and S. Nalluri, "Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, 2017, pp. 1-5.

[9]. Manu Goyal, Student Member, IEEE, Neil D. Reeves, Satyan Rajbhandari, Moi Hoon Yap, Member,IEEE "Robust Methods for Real-time Diabetic Foot Ulcer Detection and Localization on Mobile Devices"IEEE journal of Biomedical and Health Informatics

[10]. K.S. Thulasi, E. S. Ninu and K. K. M. Shiva, "Classification of diabetic patients records using NaïveBayesclassifier," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2017, pp. 1194-1198.

[11]. S.Fong, J. Fiaidhi, S. Mohammed and L. Moutinho, "Real-time Decision Rules for Diabetes Therapy Management by Data Stream Mining," in IT Professional.

[12]. M.Di Ferdinando, P. Pepe, P. Palumbo, S. Panunzi and A. De Gaetano, "Semiglobal Sampled-Data Dynamic Output Feedback Controller for the Glucose–Insulin System," in IEEE Transactions on Control Systems Technology, vol. 28, no. 1, pp. 16-32, Jan.2020

[3]. A. Beneyto, A. Bertachi, J. Bondia and J. Vehi, "A New Blood Glucose Control Scheme for Unannounced Exercise in Type 1 Diabetic Subjects," in IEEE Transactions on Control Systems Technology, vol. 28, no. 2, pp.593-600, March 2020.

[14]. Yang Guo, Guohua Bai and Yan Hu, "Using Bayes Network for Prediction of Type-2 diabetes," 2012 International Conference for Internet Technology and Secured Transactions, London, 2012, pp. 471-472.

[15]. Pradhan, M., Sahu, R.K.: "Predict the onset of diabetes disease using Artificial Neural Network". Intl J Comp Sci& Emerging Tech. 2:303-11 (2011)

[16]. Qingchen Zhang, Changchuan Bai, Laurence T. Yang, Zhikui Chen, Peng Li, and Hang YuA Unified Smart Chinese ,"Medicine Framework for Healthcare and Medical Services "Zhang Q, et al. IEEE/ACM Trans Comput Biol Bioinform. 2019.

[17]. K.Li, J. Daniels, C. Liu, P. Herrero and P. Georgiou, "Convolutional Recurrent Neural Networks for Glucose Prediction," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 603-613, Feb. 2020.


IJARIIE