International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 10 ǁ October 2023 ǁ PP. 360-368

Development of Disease Diagnostic System Using Machine Learning and Deep Learning Techniques

Dr. Yelepi Usha Rani, Bolledla Deekshitha , Meda Chandana Reddy , Mudarapu Ramya, Palla Raghavi Reddy.

[1]*Associate Professor, VNR Vignana Jyothi Institute of Engineering and Technology, UGC Autonomous, Hyderabad, 500090, India.*

[2345]*UG VNR Vignana Jyothi Institute of Engineering and Technology, UGC Autonomous, Hyderabad, 500090, India.*

Abstract: *Main purpose is to build an application which can predict more than one disease and which can provide more accurate results in health care sector. Building a programmer that can forecast several diseases with greater accuracy will be key goal in the health sector. The issue whichever is relatedto health should be analyzed correctly to prevent it or cure it. The tough and challenging task is to make a correct diagnosis and prediction of a particular disease. For this purpose, various classification algorithms play a crucial role for prediction of sickness. In this project, for the purpose of disease prediction, we are attempting to implement various classification algorithms. To choose the most accurate algorithm for prediction, each algorithm's accuracy is verified and contrasted with one another. To reach the highest level of anticipated results accuracy, we're going to merge numerous datasets.For end user to make things easy , we as a team came forward to build a web application where the user can easily predict the disease, which they want, just by entering the respective attribute(input) values of that specific disease. Our goal is building a web application which is able to find out many diseases using machine learning algorithms.*

Index Terms: *Machine Learning, Deep Learning, Support Vector Machine (SVM), Random Forest, Logistic Regression, CNN, KNN.*

Date of Submission: 18-10-2023 Date of acceptance: 02-11-2023

I. Introduction:

These days, People often deal with many ailments, it may be because of their way of living, food, their choices and surrounding environmental conditions. The growing volume of data growth in the medical and healthcare industries has benefited early patient care through efficient medical data analysis. The data of many diseases can be used for making predictions using data mining classification algorithms.[1] As a result, it is crucial to predict diseases early on. Currently, many models just concentrate on one particular disease for disease analysis. To give an example, it may be analyzed only for skin illnesses. For this, there isn't a mechanism in place that can analyze multiple diseases at once.

For predicting various diseases there isn’t any application which makes the user move from one application to another application to predict different diseases.[2] This application which can predict only one disease is required to fill lengthy questionnaires by the user. The accuracy is also not so high and reliable. Low accuracy prediction is very dangerous for the person's health .Health is the most important aspect of human life which needs to be taken care of by them.

Therefore, we made the decision to create a system that could forecast several diseases using a single user interface. If left untreated, diseases including diabetes, cancer, and heart disease pose a risk to humanity.

In this paper, we'll predict various types of diseases or illnesses. For the purpose of disease prediction, various classification algorithms were implemented and tested . To choose the most accurate algorithm for prediction, each algorithm's accuracy is verified and contrasted with one another. To reach the highest level of anticipated results accuracy, we're going to merge numerous datasets. To create a webapplication, the algorithm which has high accuracy for each disease is selected and included. By giving the appropriate values for particular sickness, the user can quickly forecast the needed disease.

II. Literature Review:

This section will include the review of previous projects and research on the existing system. Numerous analysis works are conducted for disease detection and diagnosis. This study helps in knowing that

any system involves the detection of multiple diseases like diabetes, heart disease, kidney and cancer diseases. The following papers helped in knowing the drawbacks of the existing systems.

Related Works:

"Keniya, Rinkal and Khakharia, Amanand Shah, Vruddhi and Gada, Vrushabh and Manjalkar, Ruchi and Thaker, Tirth and Warang, Mahesh and Mehendale, Ninad and Mehendale, Ninad" et al.[3] proposed a paper named **"**Disease Prediction from Various Symptoms using Machine Learning" in 2020. They have created an excel-sheet which consists all signs which are related to many diseases. The diseases were around 230 and all the unique symptoms for them were written in the excel sheet. The signs, age, and gender of a person are taken as input from a user. In this system, ML algorithms such as "Naïve Bayes Algorithm", "KNNAlgorithm", "Decision tree" and "SVM" classifier are used. After predicting the output, the application itself will suggest to the user if he or she needs a doctor to be consulted. Drawbacks for this was Some models were dependent on the parameters, they could not predict the disease and accuracy percentage was less.

"Naved, Mohd & Shinde, Priyanka & Leiva- Chauca,Orlando &Huaman-Osorio, Antonio & Gonzales-Yanac, Tatiana" et al.[4] proposed a paper named "Multiple Disease Prediction Using Different Machine LearningAlgorithms Comparatively" in 2020 . In this application ML techniques like "Naïve Bayes Algorithm", "KNN Algorithm","Decision Tree Algorithm", "SVM classifiers" are used. After predicting output the application itself will suggest to the user if he or she needs a doctor to be consulted. Drawbacks for this system was the dataset used for this project is not large enough to predict more accurate results for any other testing dataset.

"Alanazi, Rayan" et al.[5] proposed a paper named "Identification and Prediction of Chronic Diseases using Machine Learning Approach" in February of 2022 . The data is taken from different sources. Then after collection of data, the data has to be preprocessed in order to avoid errors and wrong prediction of output. Then, the data is trained with CNN andKNN classifiers. Then, after many cycles, as the output is obtained, that particular model can be tested. Drawbacks of this system is that, this system uses only two ML algorithmsfor prediction.

There were many other ML algorithms which can provide good and accurate results other than KNN and CNN.

"K., Arumugam & Naved, Mohd & Shinde, Priyanka & Leiva-Chauca, Orlando & Huaman- Osorio, Antonio & Gonzales-Yanac, Tatiana" et al.[6] proposed a paper named "Multiple Disease prediction using Machine Learning algorithms" in August 2021 . In this, the data set which is used is Cleveland data. Then, preprocessing is done for this and all the null values are removed.Finally, data is clear. Now the data will be fed into ML algorithms.These will do the job of classification and will predict the disease.. The Drawback of this system is that it is mainly tuned for best performances in the prediction of likelihood of having heart diseases in diabetes individuals only.

"Mitushi Soni , Dr. Sunita Varma" et al.[7] proposed a paper named "Diabetes prediction using Machine Learning Techniques" in September 2020 . In this project, the dataset contains 2000 records. The main objective is to forecast disease based on various measures to know if that individual has higher sugar levels in blood or not. Classification algorithms like "KNN", "Logistic Regression", Decision Tree", "Random Forest", "SVM" has been used. Goal is about building the model which can give more accuracy in detectingif diabetes is present or not. The whole data is collected from a uci repository that is under the name of pima Indian diabetes dataset.The data is collected from hospitals in china. ML classification algorithms like decision tree, RF , neuralnetworks has been used for detecting diabetes. Firstly, two separate datasets are pre-processed. At last, the model that gave more accuracy is chosen and then the project is finally deployed in a web-application that uses flask.The above paper shows that KNN and logistic regression were best for the detectionof heart disease.

III. Proposed Approach:

This paper can forecast many diseases using various classification algorithms like "Naive-Bayes, KNN, Random Forest, Logistic Regression and SVM-Classifier".

Validation of each and every algorithm has been performed and to find out the best accuracy. We will integrate multiple datasets to get best results with good accuracy.

We as a team came forward to build a web application that can be used easily by anyone to predict the disease just by filling some of the attributes for the related disease.

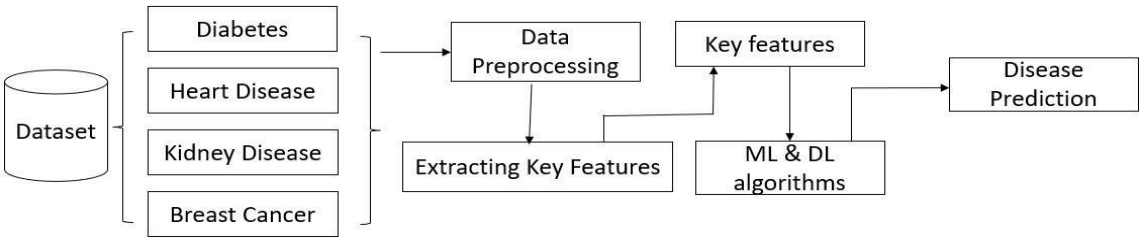


Fig 1 : Proposed Methodology

For each disease, the available classification algorithms are applied, and the algorithm which has high accuracy, that model for that particular disease is considered, unlike the existing models, where a single model is applied for all the diseases detection. All these models are finally integrated during the web application. The users need to give input, the information which is required in order to detect a disease. Then the model processes the input and then finally predicts whether a user is affected with aparticular disease or not. The same methodology is followed for other diseases as well. By doing this, the accuracy of each disease detection will be increased and even predictions will be accurate.

IV. Objectives of Proposed System :

The primary goal in the healthcare sector is to develop a program that can predict a variety of diseases more accurately. Prevention ispreferable to treatment. For preventing health-related problems, it should be accurately assessed in order to limit further harm. The most difficult task is correctly diagnosing a condition. We applied ML and deep learning algorithms,which are crucial for forecasting or predicting a disease, to solve this issue.

V. Work Plan of Project:

After formulating a concise problem definition and choosing the essential algorithm to be employed as part of the model, we proceed to the plan of carrying out the construction of our model. The following stages were developed as part of the work plan:

Stage 1: Dataset Collection

The primary requirement for our project is the collection of disease-related datasets. Once these are obtained, we might move on to the following phase.

Stage 2: Pre-processing

Every time the required data is collected from various other sources, this is done so that unprocessed data is prevented from analysis. After that, the raw data is translated into a formatthat can be read by computers and used by staff members across an organization (graphics, papers, etc.). We can move forward with trainingthe model after pre- processing.

Stage 3: "Data is split into Train and Test"

Here, the dataset is splitted into the above two data in proportion 8:2, or 80% for training model and 20% for testing model after it is built. Then use the train dataset to apply each method.

Stage 4: Apply the algorithm

Select the model with the highest level of accuracy following the application of all methods.

Stage 5: Selecting the disease

After deciding on the appropriate model, choose the disease on the interface, and then enter the required information. Based on the information given, the system predicts the ailment.

VI. MODEL WORKFLOW:

One of the most concerning industries is healthcare. The demand to accurately detect a variety of diseases is currently quite great. Due to the complexity of the various disease processes and underlying symptoms of the patient population, the creation of a tool for early diagnosis and an effective therapy faces significant obstacles. Thanks to machine learning, some of these issues can be handled by academics, medical professionals, and patients.

The majority of currently employed algorithms can only accurately predict one ailment at a time, and even then, only one at a time. The patient's health could be seriously threatened by lower precision. The process of looking through several websites for disease analysis take time. In this application, we used methods like SVM, KNN, Random Forest, etc.

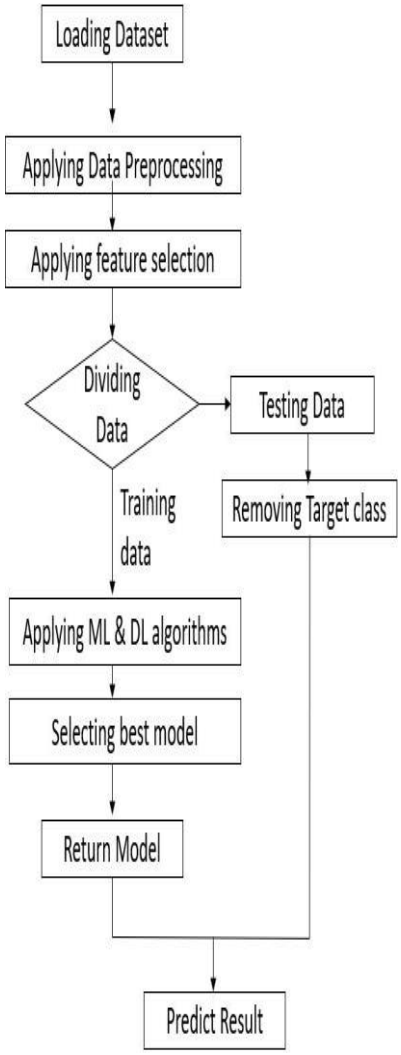


Fig 2 : Work Flow

Availability of data and material:

Diabetes :
The dataset contains 2000 observations With 9 aattributes.

Heart Disease:
This dataset contains 918 observations with 12 attributes.
Chronic Kidney Disease :
This dataset has 400 observations with
26 attributes .
Breast Cancer :
The dataset contains 500 observations and 5 Features.

VII. Accuracy of Prediction:

In this paper 6 algorithms were implement for each disease which include KNN, SVM, Random Forest, Logistic Regression, Naïve-Bayes, Decision Tree. The accuracies of above algorithms are compared with each other and the best one has been chosen. For Diabetes Disease among the all-algorithms Random Forest gives highest accuracy. For Heart Disease among the all-algorithms SVM gives highest accuracy, for Chronic Kidney among all-algorithms Random Forest gives highest accuracy,for cancer disease the accuracy was higher by using Neural networks i.e., 95.65 percentage .

Disease/ Classifier	KNN	SVM	Random Forest	Decision Tree	Naive Bayes	Logistic Regression
Diabetes Disease	82.25	78.5	**99.5**	98	78.75	78.75
Heart Disease	85.21	**90**	86.08	77.82	89.13	89.56
Chronic Kidney Disease	78.33	97.5	**99.16**	97.5	93.33	97.05

Table 1: Accracy Table

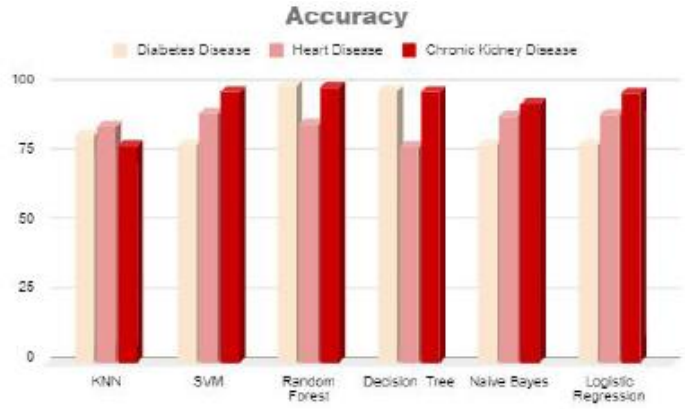


Fig 3: Graphical representation of accuracy

Disease/Classifier	KNN	SVM	Random Forest	Decision Tree	Naive Bayes	Logistic Regression
Diabetes Disease	0.1775	0.215	**0.005**	0.02	0.2125	0.2125
Heart Disease	0.1478	**0.1**	0.1391	0.2217	0.1086	0.1043
Chronic Kidney Disease	0.425	0.0416	**0.0083**	0.0333	0.125	0.0416

Table 2: Mean Absolute Error Rate Table

Fig 4: Mean Absolute Error Rate

MODEL ACCURACY AND MODEL LOSS FOR BREAST CANCER:

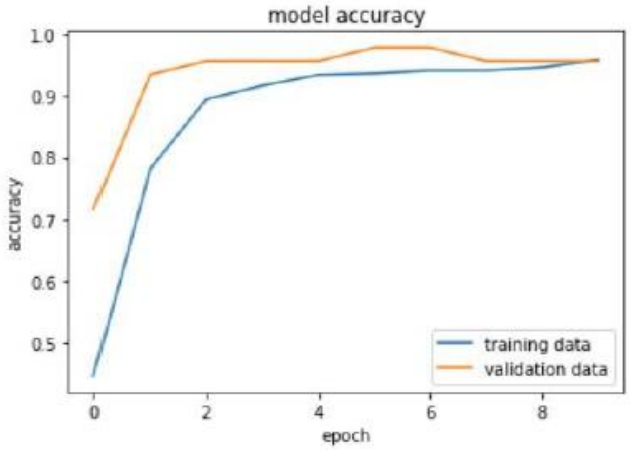


Fig 5: Model Accuracy

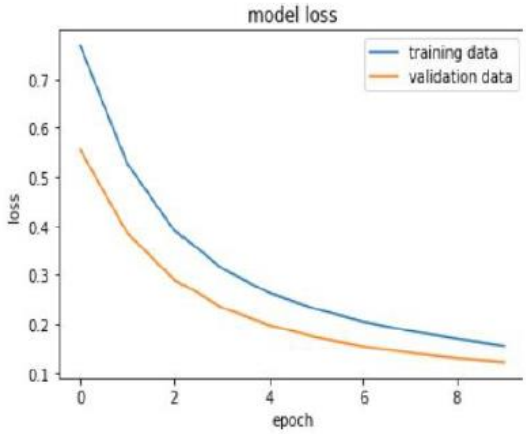


Fig 6: Model Loss

VIII. Results:

This paper explains a web based application which can predict more than one disease with most accuracy rate and it has been built by using streamlit. The application is having two frames. The leftmost frame consists of the various types of diseases . When a particular disease is selected the respective attributes of the disease will be displayed in the right side frame.

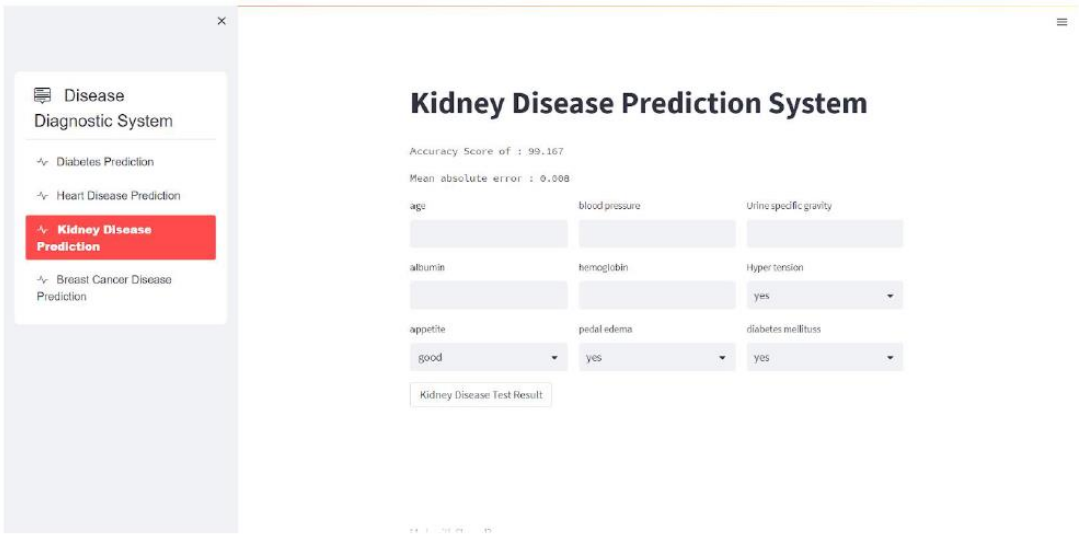
When the user clicks on Diabetes Disease, it shows a list of features that need to be filled in order to predict the disease.After filling , on clicking "Diabetes Test Result" , button displays the output on the screen according to the symptoms (yes or no). Yes indicates that the user is affected by that disease. NO indicates the user isn't affected by the diabetes disease.If the user is affected by that particular disease then it automatically suggests a list of few doctors that the user can contact if the user wishes to.
When the user clicks on Heart Disease, it shows a list of attributes that need to be filled in order to predict the disease. On clicking "Heart Test Result".button displays the output on the screen according to the symptoms (yes or no). Yes indicates that the person is affected by that disease. NO indicates the person isn't affected by the heart disease.

When the user clicks on Kidney Disease, it shows a list of attributes that need to be filled in order to predict the disease.After filling , on clicking "Kidney Disease Test Results" , button displays the outputon the screen according to the symptoms (yes or no). Yes indicates that the user is affected by kidney disease. NO indicates the user isn't affected by kidney disease.

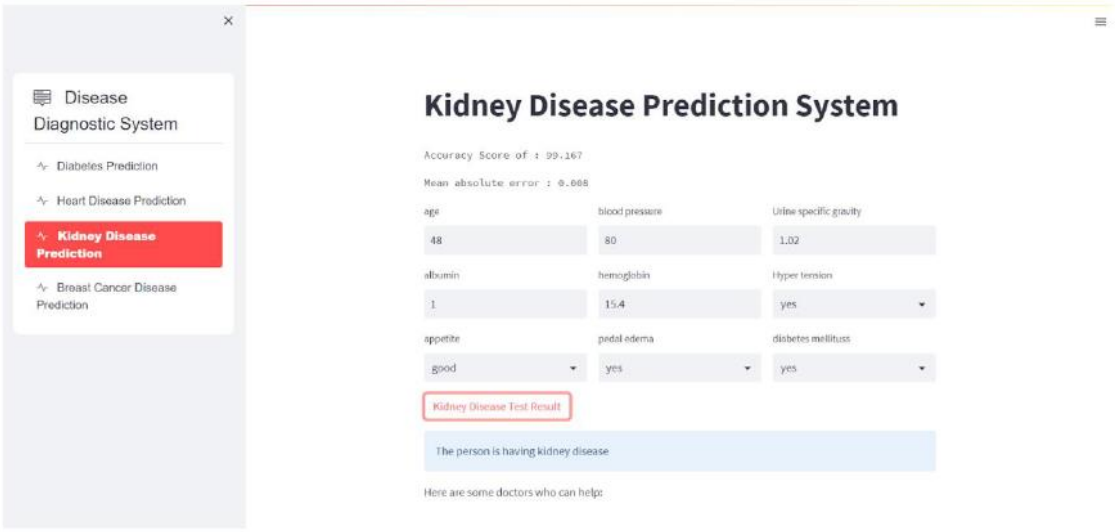
When the user clicks on Breast Cancer Prediction , it shows a list of attributes that need to be filled in order to predict the disease.On clicking " Breast Cancer Disease Test Results", button displays the output on the screen according to the symptoms(Yes or No). Yes indicates that the user is affected by Breast Cancer . NO indicates the user isn't affected by Breast Cancer disease.

As the below pictures shows how the output window screen looks when the user accesses the application .
Each and every disease will show their respective attributes to be filled by the user in order to detect the disease exists or not.
If the application predicts that there is no disease then it displays the person isn't effected by the disease .If the person is affected by the disease then the application itself gives some of the best doctor suggestion to the user so that it won't be a hard task for the user to find the specialist in specific domain. The following pictures shows the steps of the application and doctor suggestion of respective specialist in the specific domain or area.

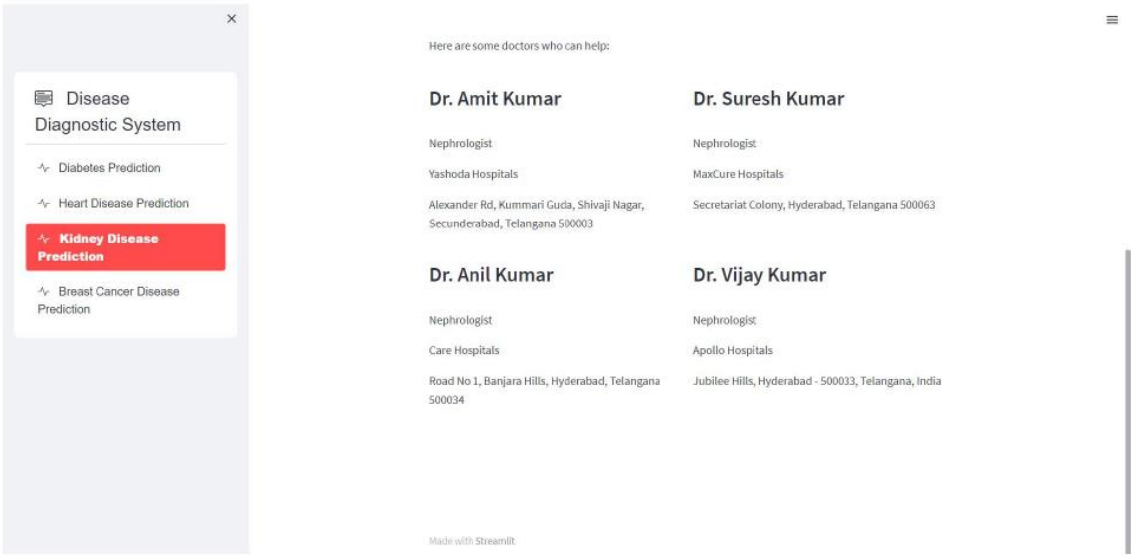
STEP 1:



STEP 2:



STEP 3 :



IX. Conclusion:

This paper discusses about prediction of various diseases, application predicts results for Heart diseases, chronic kidney disease, cancer and Diabetes by taking input values from user. By using these users can easily find all possible results at one place. Early diagnosis by decoding medical patterns can be done. We have applied each and every algorithm on the model to predict the best accuracy for the disease. So by trying each and every algorithm on the model based on the accuray we have decided on using the final algorithms for the application to get the best results. Based on the above study "Random forest algorithm" gave the best accuracy prediction for Diabetes and Chronic kidney Disease. And "SVM" gives best accuracy in the prediction of "Heart Disease" .For Breast cancer disease the accuracy was higher by using Neural networks i.e., 95.65 percentage .

Prediction of Diabetes :

- Random Forest Algorithm
- Accuracy 97.75

Prediction of Heart Disease:

- SVM Algorithm

- Accuracy 89.1

Predicting Chronic Kidney Disease:

- "Random Forest Algorithm"
- Accuracy 99.16

Prediction of Breast Cancer:

- Neural Networks
- Accuracy 95.65

References:

[1]. Latif, Jahanzaib & Xiao, Chuangbai & Tu, Shanshan & Rehman, Sadaqat Ur & Imran, Azhar & Bilal, Anas. (2020). Implementation and Use of Disease Diagnosis Systems for Electronic Medical Records Based on Machine Learning: A Complete Review. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3016782.

[2]. Ahirrao, Aditya & Bhagwat, Aditya & Desai, Pranali & Kaneri, Sourabh & Shaikh, & Mohammad, Sameer. (2020). Multi Disease Detection and Predictions Based On Machine Learning. SSRN Electronic Journal. 7. 950-953.

[3]. Keniya, Rinkal and Khakharia, Aman and Shah, Vruddhi and Gada, Vrushabh and Manjalkar, Ruchi and Thaker, Tirth and Warang, Mahesh and Mehendale, Ninad and Mehendale, Ninad, Disease Prediction From Various Symptoms Using Machine Learning (July 27, 2020). Available at SSRN: https://ssrn.com/abstract=3661426or http://dx.doi.org/10.2139/ssrn.3661426

[4]. Naved, Mohd & Shinde, Priyanka & Leiva-Chauca, Orlando & Huaman-Osorio, Antonio & Gonzales- Yanac, Tatiana. (2021). Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings. 10.1016/j.matpr.2021.07.361.

[5]. Alanazi, Rayan. (2022). Identification and Prediction of Chronic Diseases Using Machine Learning Approach. Journal of Healthcare Engineering. 2022. 1-9. 10.1155/2022/2826127.

[6]. K., Arumugam & Naved, Mohd & Shinde, Priyanka & Leiva-Chauca, Orlando & Huaman- Osorio, Antonio & Gonzales-Yanac, Tatiana. (2021). Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings. 10.1016/j.matpr.2021.07.361.

[7]. Mitushi soni , Dr. Sunita varma, 2020, Diaabetes prediction using machine learning techniques INTERNATIONAL JOURNALOF ENGINEERING RESEARCH & TECHNOLOGY(ijert) 2020.

[8]. Prediction Of Diabetics Using Machine Learning Classifiers: A Review November 2021 DOI:10.1109/I-SMAC52330.2021.9640806 Conference: 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)

[9]. Zou Q, Qu K, Luo Y, Yin D, Ju Y and Tang H (2018) Predicting Diabetes Mellitus With Machine Learning Techniques. Front. Genet. 9:515. doi: 10.3389/fgene.2018.00515

[10]. Baby, Steffy & Karunakaran, V.. (2021). Prediction Of Diabetics Using Machine Learning Classifiers: A Review. 530-537. 10.1109/I- SMAC52330.2021.9640806.

[11]. Ionita, Irina & Ioniță, Liviu. (2016). Applying Data Mining Techniques in Healthcare. Studies in Informatics and Control. 25. 385-394. 10.24846/v25i3y201612.

[12]. Frank, Eibe & Hall, Mark & Holmes, Geoffrey & Kirkby, Richard & Pfahringer, Bernhard & Witten, Ian & Trigg, Len. (2010). Weka-A Machine Learning Workbench for Data Mining. 10.1007/978-0-387 09823-4_66.

[13]. Pang-Ning Tan; Michael Steinbach; Anuj Karpatne; Vipin Kuma Introduction to Data Mining 2nd ed, Publisher: Pearson, 2019, Print ISBN: 9780133128901, 0133128903 eText ISBN: 9780134080284, 013408028.

[14]. Pandey, Dr. Subhash. (2016). Data Mining Techniques for Medical Data: A Review. 10.1109/SCOPES.2016.7955586.

[15]. Sisodia, Deepti & Sisodia, Dilip. (2018). Prediction of Diabetes using Classification Algorithms. Procedia Computer Science. 132.1578-1585. 10.1016/j.procs.2018.05.122.

[16]. Tharak Roopesh, Asadi Srinivasulu and K.S.Kannan EasyChair, Prediction of Diabetes Disease Using Data Mining and Deep Learning Techniques, Easy hair Preprint, № 1608, October 9, 2019.

[17]. K.Priyadarshini, I.Lakshmi, Predictive Analysis of Diabetes Using Bayesian Network and Naive Bayes Techniques, International Conference on Advancements in Computing Technologies - ICACT 2018, Volume: 4 Issue: 2, ISSN: 2454-4248.

[18]. Giveki, D., Salimi, H., Bahmanyar, G., & Khademian, Y. (2012). Automatic Detection of Diabetes Diagnosis using Feature Weighted Support Vector Machines based on Mutual Information and Modified Cuckoo Search. ArXiv, abs/1201.2173.

[19]. Jegan, Chitra. (2013). Classification Of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications. 3. 1797 - 1801.

[20]. Sahana Shetty, Kaveri B. Kari and Jayantkumar. A. Rathod,Detection of Diabetic Retinopathy Using Support Vector Machine (SVM) , International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 23 Issue 6 –OCTOBER 2016 (SPECIAL ISSUE)

[21]. Bashir, Saba & Qamar, Usman & Khan, Farhan & Javed, Muhammad. (2014). An Efficient Rule-based Classification of Diabetes Using ID3, C4.5 & CART Ensembles. Proceedings - 12th International Conference on Frontiers of Information Technology, FIT 2014. 10.1109/FIT.2014.50.

[22]. Faruque, Md &, Asaduzzaman & Sarker, Iqbal. (2019). Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus, 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.

[23]. Talha Mahboob Alam, Muhammad Atif Iqbal, Yasir Ali, Abdul Wahab, Safdar Ijaz, Talha Imtiaz Baig, Ayaz Hussain, Muhammad Awais Malik, Muhammad Mehdi Raza, Salman Ibrar, Zunish Abbas, A model for early prediction of diabetes, Informatics in Medicine Unlocked, Volume 16, 2019,100204, ISSN 2352-9148.