Transcend Health: Healthcare Solution Using Machine Learning Approaches

Nilesh
School of Computer Science and Engineering
REVA University
Bengaluru, India
nilesh1722000@gmail.com

Prajwal U
School of Computer Science and Engineering
REVA University
Bengaluru, India
prajwal.u.34@gmail.com

Nikhil Gangneja
School of Computer Science and Engineering
REVA University
Bengaluru, India
nikhilgangneja@gmail.com

Neelakanth S
School of Computer Science and Engineering
REVA University
Bengaluru, India
neelakanth2000@gmail.com

Prof. Ranjitha U N
School of Computer Science and Engineering
REVA University
Bengaluru, India
ranjitha.un@reva.edu.in

***Abstract*—The health-care industry generates a lot of information. It's simple to tap into this data and predict diseases early, thanks to technological advancements. Chronic disorders such as coronary artery disease, chronic kidney disease, and diabetes have a high mortality rate and can be fatal if not diagnosed and treated early. In order for people to be able to heal themselves of these ailments, this machine learning prediction must be made publicly available to everyone. Our proposal is to create a one-stop health-care solution in this regard. People can utilize the web application to submit symptoms and other information related to the disease, and the machine learning model will predict the presence of a disease and display it to the user. Medical data is collected and these statistics are used as input for the models, and those abilities are then modeled for prediction of the diseases to provide high accuracy. This model is then used at the backend to forecast future medical conditions of people. This solves the issue of getting tested for disease without requiring considerable bodily action, traveling to the hospital, which in turn minimizes the number of tests to be conducted. This drastically reduces deceiving people to take many tests and lowering the costs.**

Keywords—Machine learning, chronic kidney disease, coronary artery disease, diabetes, web application, foolproof healthcare

I. INTRODUCTION

Healthcare is one of the most in-demand fields in this new post-covid era, where everything is geared toward the internet and virtual worlds. Much of the real-world problems may be handled with the aid of artificial intelligence and data mining tools. Data mining, which combines several disciplines, is employed to draw out valuable information from huge quantities of data. It's beneficial for extracting healthcare information for clinical decision-making and generating hypotheses from large quantities of medical data. The health-care business generates enormous amounts of data, and by analysing this data, many illness may be recognised and predicted early on, and even be treated early to prevent mortality. Simultaneously, rapid advancements are being made in clinical analytics, such as systems for evaluating and analyzing massive amounts of data and drawing new insights continuously from these analyses. This brings up incredible prospects to save health-care expenses while also making illness diagnosis considerably easier.

Chronic renal disease is a serious and major problem across the world, and it is a condition marked by a gradual decrease of kidney's performance over time. It affects around one-sixth of the world's population. The symptom of This disease is that it will not cleanse the blood as well as it should, leading to renal failure. The proliferation of Plaque is the most common cause of coronary artery disease. Coronary arteries constrict leading to a condition which reduces the blood flow to the heart. This disease can occur and may exhibit in a variety of ways, the symptoms can range from no symptoms to chest discomfort and ultimately

to a heart attack as well. Diabetes occurs when blood sugar levels remain high for an extended length of time. It has recently been identified as a risk factor for many diseases such as blindness and renal failure.

A web application is developed to allow anyone to predict if they are healthy or not using machine learning models and determine whether or not they have the condition. Because of the accessibility of access to the internet, web apps are frequently utilized. The major goal of establishing a one-stop health-care system is to provide a fool-proof health solution. People are frequently duped into undergoing many needless tests, resulting in higher hospital expenses. Doctors' workload is also increased as a result of this. The amount of time we spend interacting with the hospital is decreased when we use this app, as is the time we spend waiting for testing. In recent years, the hospital atmosphere has been viewed as unsafe. However, with the online web application, anybody from anywhere in the globe may utilize the services given to check for infections which saves time instead of waiting for the doctor’s appointment.

For each disease, we applied various machine learning models in our study. Classification algorithms are used in the models which is a common technique that is utilised in a variety of applications, including locating unknown samples. Diseases are diagnosed using a combination of clinical indicators and laboratory results. To mention a few, the models include decision tree classifier, extra boost classifier, support vector machine, logistic regression, random forest classifier, and k nearest-neighbour. All of the models were evaluated, and the best algorithm was chosen and incorporated with the online application to accurately forecast the disease's outcome. The data was first processed to make it usable for the models, while also cleaning and refining it to improve accuracy.

II. Related Work

A. Chronic Kidney Disease

Ebrahime Mohammed Senan[1] and teams suggested that in the data pre-processing stage the mean statistical is used to replace missing numerical values whereas mode statistical analysis method weas used to replace the missing categorical values. And then the feature selection was done, to do this the features were removed recursively using RFE algorithm to end up with the most important features in the set for the modelling purpose. Four popular machine learning algorithms were used namely SVM, KNN, decision tree, and random forest for modelling to extract the best diagnostic results. The random forest classifier gave the best accuracy of one hundred percent, and classified all the positive and negative samples accurately.

Imesh Udara Ekanayake[2], in his research concentrated on data cleaning. Where the with attributes with more than one fifth of data had missing values were deleted as it did not contribute much for prediction. As a result, the features such as blood cells count etc. are left out of the analysis. In the next step the missing values in the other features were handled. To ensure satisfactory accuracy, missing values must be handled in the pre-processing step depending on their distributions. Little's MCAR test was used in this study to confirm the randomization of missing values. The missing values were substituted by taking average value using the KNN model instead of just substituting it with the mean or mode statistics. When this was compared to data which had been substituted with the mean and mode values, the results showed that the KNN infused dataset gave better accuracy for the models in prediction of the disease. This increases the scope to investigate and invest in KNN imputer based techniques to handle missing values in the sets of data which in turn will increase the accuracy of prediction. Additionally, by taking into consideration factors such as water intake patterns, and dietary kinds into the research, more information about CKD can be learned.

Chronic kidney disease was studied and evaluated by S.Revathy[3]. Chronic Kidney Disease(CKD) is a disorder wherein the kidneys are unable to fulfil their normal blood filtration and other functions. Kidneys' main function is to filter-out excess water and waste from the blood. processing. Pre-processing data acquired from numerous sources makes it easier to use data mining methods. On the CKD datasets, this paper depicts how to develop and use a data mining technique for extracting useful information. A large number of CKD datasets are gathered. The usual data mining approaches of data preparation and pre-processing are used. To predict the early onset of CKD using data mining, three machine learning methods were used namely support vector machine, Decision tree classifier and Random Forest classifiers. Each algorithm's effectiveness is evaluated using various parameters, and the model which had the highest effectiveness was selected.

B. Coronary Artery Disease

Rahul Gupta[4] and his team worked on diabetes detection. Humans are at risk from diseases such as coronary artery disease, cancer, and tumour illness. The recommended study in this paper was conducted in this regard to forecast coronary disease early on. AI techniques are frequently used to detect heart disease. Medical data is collected and these statistics are used as input for the prediction of illness. These techniques can then be used to predict medical condition in the future. In the suggested approach, they apply three alternative data mining strategies. The model's accuracy while employing each of the methods is intended. The algorithm with the highest accuracy is then chosen for predicting the diseases. The proposed approach uses machine learning to detect and predict the likelihood of developing coronary artery disease. For predicting the CAD various machine learning techniques were used such as SVM, random forest and Naïve bayes classifier. The data for this analysis was gathered from the uci data repository, and it also contains data collected from other sources. The training dataset comprises the prediction of heart disease incidence for various age groups.

Santosh kumar[5] and team utilised a publicly available dataset, machine learning systems can detect the existence of heart disease with excellent accuracy. This paper showed that when it comes to detecting CAD, the neural network model beats other machine learning methods. Many machine learning techniques were developed and applied to the data after the dataset was pre-processed. Evaluation metric such as F1 score and accuracy used to evaluate the models results. The six machine learning models performed admirably, with accuracies of better than 80 percent. Tools such as R was used in this research and investigation, this has also been. Using a publicly available dataset, machine learning methods may be used to detect the presence of heart disease with high F1 score and accuracy. Here they also showed that the random forest classifier was prone to overfitting compared to the other models. Hence the neural networks models was used to predict the disease accurately.

The goal of Rashmi[6] and her team's research is to use machine learning to diagnose cardiac disease using a minimum number of attributes. They also used Hadoop and big data analytics concepts to run the models and provide faster outputs. This random forest classifiers were running on different nodes on Apache Spark to provide a solution in a more accurate and faster manner. This method provides health care analysts a huge potential to use these techniques to tackle the hurdle of analysing huge amounts of data. The focus of this project is on the data analysis portion. The healthcare data collected every now and then in the Medicare industry may be handled quickly using big data analytics for illness prediction with minimal overhead. They suggested and evaluated a scalable approach for predicting heart disease characteristics. They used the Spark framework to develop the random forest method for predicting heart

illness, and they demonstrated that it can operate with as few as six hundred data records and were able to achieve ninety eight percent accuracy.

C. Diabetes

Aishwarya Mujumdara[11] suggested that in order to provide more accurate and exact findings, the model deals with inconsistencies in data. There were missing values in the dataset. Because these attributes cannot have 0 values, these values were substituted with other values and was handled appropriately for a few selected attributes. The dataset was then scaled to normalize all values to achieve better results. Then multiple machine learning classification techniques were applied to the preprocessed data, and this classification results were evaluated. Different models which were used are Logistic Regression, AdaBoost classifier, Extra trees classifier etc. Among these algorithms the model with the best accuracy was chosen which turned out to be the AdaBoost classifier with an accuracy of 98.8%. They had seen a comparison of the accuracies of two distinct machine learning algorithms using two separate datasets. They used numerous evaluation criteria, such as classification accuracy, confusion matrix, and f1-score, to assess the prediction outcomes.

Mitushi Soni[12] suggested several measures that were made in this project. Different categorization and ensemble approaches are used in the suggested strategy. In this study, it was discovered that random forest classifiers outperformed the remaining classifiers in terms of accuracy. All in all, they applied the few of the most advanced data mining techniques for detection and prediction of the diseases with excellent efficiency and accuracy. In this case, characteristics played a significant impact in prediction. The main intention behind this research was to develop a diabetes prediction model using data mining techniques. This research also evaluates and analyze the performance of all the developed models and ultimately select the best model with the best accuracy. The different models used ranges from random forest classifiers to logistic regressions to SVM. Gradient Boosting classifiers were also used in this research paper, as well as other classification and ensemble learning approaches. Furthermore, a classification accuracy of 77% was achieved using Random Forest. The outcomes of the experiment can help people to diagnose diabetes at an early stage in order to cure and save lives.

Ayush Anand[13] used data acquired manually through questionnaires in this study. The questionnaire was created solely through discussions with doctors, two of whom were Diabetologists. For a highly categorized dataset, the CART(Classification And Regression Trees) prediction model was used which resulted in 75% accuracy. The k-fold cross-validation approach was used to validate the results, which was followed by the CART technique. The complete data set was divided into five portions, with each fold serving as a test set once and a training set for the remaining iterations. As a result, they were able to improve the accuracy of the created model while still providing completely unbiased data. Many key attributes were discovered in this research. These key attributes are daily lifestyle activities which can influence and cause diabetes to occur. These attributes are eating at roadside eateries, fluctuation of blood pressure, irregular sleep cycle, high intake of rice and also inheritance of diabetes from parents. As a result, while one should appreciate every part of one's life, a little prudence in one's everyday routine is not harmful.

III. METHODOLOGY

A number of different data mining techniques were experimented to analyse the datasets. For each disease a dataset was collected and used for the analysis. We utilized machine learning techniques such as KNN, random forest, decision tree and SVM. In pre-processing, handling of missing values is done.

Various strategies were used to pick important features. Disease diagnosis was based on a few selected characteristics. Three diseases were identified in this study: Chronic kidney disease, Coronary heart disease, and Diabetes. KNN, random forest, decision tree and SVM and boosting approaches were utilised as algorithms. All of the classifiers produced promising and exceptional results for predicting and diagnosing the given variable into diseased or not.

A. Dataset

There were 400 entries and 25 characteristics in the kidney disease dataset. Attributes such as Age, haemoglobin, appetite, blood pressure, diabetes, sugar, red blood cells, and other characteristics are included. CKD and not CKD are the two values in the diagnostic class. Except for the diagnostic feature, all features had missing entries.

There are 768 entries in the Diabetes dataset, and it has 9 characteristics. Age, blood pressure, pregnancy, hyperglycaemia, and other factors are all considered. There are two values in the diagnostic class: 0 and 1. There were some missing values in some features.

There were 303 entries and 14 characteristics in the Heart disease dataset. Age, kind of chest discomfort, resting blood pressure, sex and other characteristics are included. There are two values in the diagnostic class: 0 and 1. There were no missing values in any of the features.

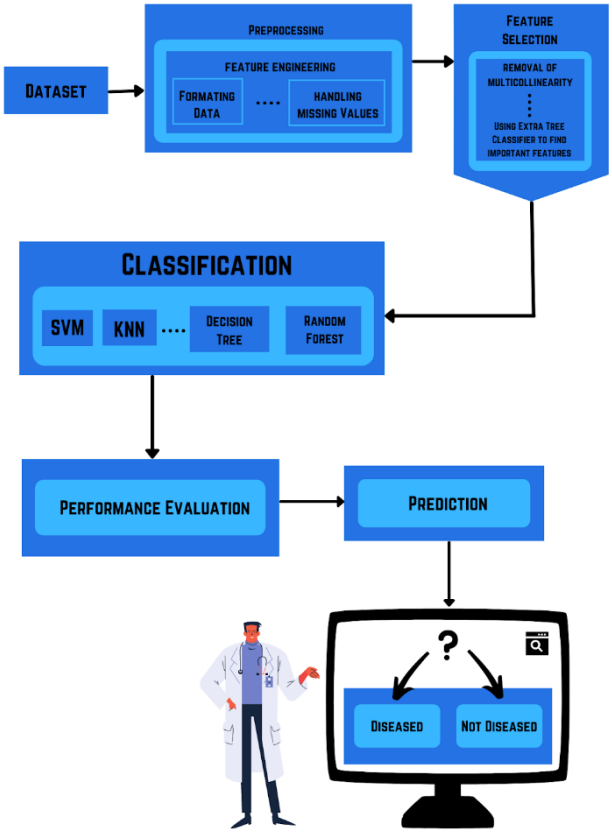

Fig. 1. Methodology.

B. Pre-Processing

Outliers and noise were present in the dataset, therefore it needed to be cleaned up in this stage. Estimating missing values and removing noise, outliers were all part of the preprocessing stage as was standardized scaling of variables, testing for unbalanced data, and variable encoding. The most straightforward way to deal with missing entries is to remove those entry or ignore the records, although this is ineffective on a tiny dataset. Instead of eliminating records, we have applied methods to compute missing values. Missing entries for numerical attributes can be calculated using statistical measures like mean, mode and median, whereas missing entries for categorical characteristics was calculated using the mode statistical approach, wherein the missing values are replaced with the most frequently occurred value in each of the attributes.

C. Features Selection

Here, key features with a high and positive association for disease diagnosis are identified, while features that are irrelevant meaning the features that do not contribute significantly for the prediction are eliminated from the dataset. These useless features will hinder the robust performance of the model as it increases the complexity of the model. Here, columns with high collinearity with each other were removed, and Extra Trees Classifier was used to find top features influencing the target, and those features were only used for prediction because they contributed most to the model's accuracy, and the achieved accuracy had minimal changes.

D. Classification

To create categorization templates, data mining and machine learning techniques were used to develop new, understandable and insightful patterns. The creation of machine learning models established on preceding analysis is required for both supervised and unsupervised learning approaches. These techniques are utilized for diagnostics in the medical industry, these models fall into the category of regression and classification. SVM, KNN, random forest, decision tree, logistic regression and boosting are popular ML algorithms that produced the best diagnostic findings. ML techniques are partitioned into two stages, the models work around these stages to build patterns which can then be used for classification. Two stages are: The training stage, in this stage the model is developed from the training dataset with the expected results. The second stage is the validation stage in which the constructed model is validated for its quality using a test dataset. All of the techniques are supervised methods for solving classification and regression issues.

E. Splitting Dataset

Dataset was divided into two parts namely the training dataset and testing dataset in the specific ratio to get a better trained model and then validation is done on the testing data.

F. Building and deploying the GUI for the built models

We planned to develop refined structure and user interface to the prediction model. To do this, the best model was uploaded to the backend of our live website, different conditions and parameters were added, and a basic simple and easy to use web application using HTML and CSS was developed, in which the user enters the values for the parameters as described. The app would then use these inputs provided for the prediction of the diseases. The app will then inform you of its essential prediction. The user would

receive the desired answers based on the analysis and be informed as to whether he or she is healthy or should seek medical advice.

Various machine learning models were developed using the UCI and Kaggle datasets with multiple features. In Python, the models were saved in a pickle file. Then with the help of Flask the web application was integrated with the model at the backend. Before being redirected to the form on that page, the user will be asked to select the condition he or she wants to check for. In the website's form, the user can fill out the essential information. The input is passed as a request to the models on the backend using flask, and the models revert the prediction, which is displayed in the result area of our website. The Flask web application was later released on the Heroku platform, which is a cloud-based platform for developing and running web applications. The outcome of the application is dependent on probability. Models will assess whether the person is healthy or in need of medical help. It will also be displayed in the results section of the website.

IV. Result and Analysis

A. *Evaluation Metrics*

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All Samples}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Fig. 2. Evaluation metrics.

The metrics for evaluation which are available in python were used to validate the performance of the classifiers. The confusion matrix is one of these metrics, from which the true negatives, true positives, false negative and false positives are displayed in a tabular form. Using this matrix accuracy, F1 score, recall and precision calculated by computing the properly categorised samples True Positives and True Negatives and the wrongly classified samples False Positives and False Negatives, as illustrated in the formulae above.

B. *Chronic Kidney Disease Classification*

Multiple models was used to find the best suitable algorithm to predict CKD. The model with the best result is the Random forest classifier. As the name implies, a random forest classifier comprised of massive number of individual decision trees that work together as an ensemble model. The class prediction of each tree in the random forest is aggregated and the class which has the most votes will be considered as the prediction of the model. It was chosen for our final CKD illness prediction because it performed better than

other models and had a 96 percent accuracy rate. The different models that were also considered were SVM, Logistic regression and Gradient boost which gave an accuracy of 90, 87.5 and 95 percent respectively.

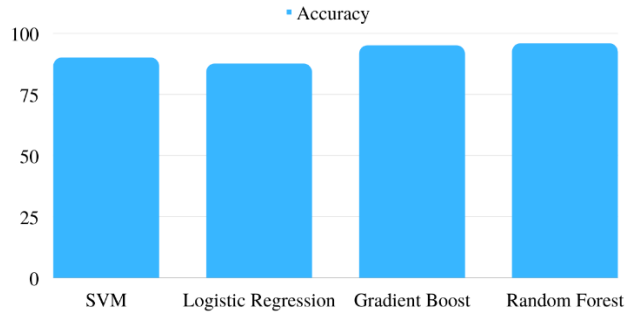

Fig. 3. Chronic Kidney Disease Model Accuracy.

C. *Heart Disease Classification*

SVM is used to predict cardiac disease. Support Vector Machine is a supervised data mining technique which is used for classification and regression techniques. It locates a hyper-plane that establishes a distinction between different data types. This hyper-plane is nothing more than a line in 2 dimensional space. When the dataset has two classes and is separable, linear SVM is utilised. Different kernels such as linear, RBF, poly, and sigmoid are used to forecast heart disease, and accuracy is calculated from them. SVM is used for final predictions since it performed better than other algorithms. Poly kernel obtained best accuracy for training data, but RBF kernel performed better than poly for testing data, hence SVM is used with RBF kernel for final output prediction. The different models that were also considered were KNN, Logistic regression and random forest which gave an accuracy of 79, 86 and 81 percent respectively.

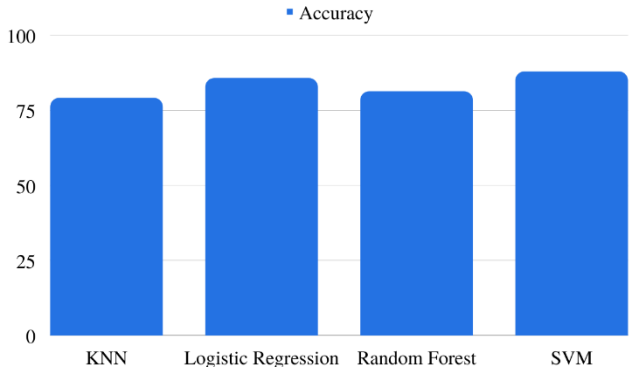

Fig. 4. Coronary Artery Disease Model Accuracy.

D. *Diabetes Classification*

Diabetes is detected using KNN. K Nearest Neighbours is a regression and classification algorithm and a machine learning technique. In order to predict which class a data point belongs to K-Nearest Neighbours analyses the labels of a specific number of observations neighbouring it. The Euclidean distance is used to determine the neighbours. Because it performed better than other models, with an accuracy of roughly 81 percent, it was picked for our final diabetes illness prediction. The different models that were also considered were SVM and random forest which gave an accuracy of 76 and 77 percent respectively.

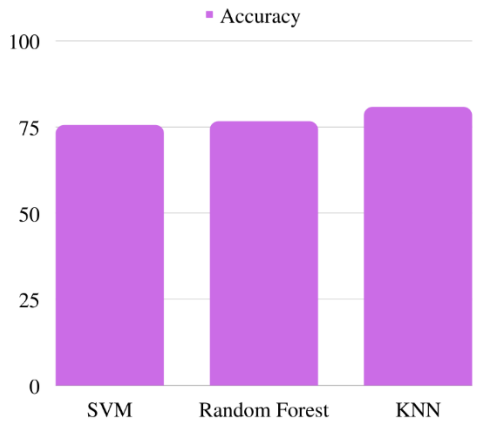

Fig. 5. Diabetes Disease Model Accuracy.

V. Limitation and Future Work

In the online web application, just three distinct diseases have been included. More illnesses can be anticipated and incorporated in the future utilizing data supplied by the medical sector. More real-time data which could be extracted from different sources could be used to keep updating the models to increase accuracy. Using data science approaches, a whole virtual healthcare system can be created. Doctor recommendation can be added for people to consult for additional medical assistance.

VI. Conclusion

In today's world, when everything is virtual, we devised a strategy to build a virtual healthcare system based on machine learning that can identify disease early on and avert life-threatening scenarios. The healthcare data collected every now and then in the Medicare industry may be handled quickly using big data analytics for illness prediction with minimal overhead. Predicting illnesses with greater overall accuracy allows more individuals to become aware of life-threatening diseases early on, allowing them to be treated at minimized costs and with less risk-factors. Effective feature engineering reduces the amount of attributes needed for the machine learning algorithms, this in turn reduces the number of medical tests that have to be taken up. Machine learning techniques have been used to provide a one-stop solution for forecasting certain chronic illnesses. This reduces reliance on hospitals by allowing patients to be tested for illness in the privacy of their own homes at minimal cost and with high accuracy.

References

[1] Ebrahime Mohammed Senan ,Mosleh Hmoud Al-Adhaileh ,Fawaz Waselallah Alsaade, “Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques”, *Hindawi, Journal of Healthcare Engineering, Volume 2021, Article ID 1004767*

[2] Imesh Udara Ekanayake, Damayanthi Herath, “Chronic Kidney Disease Prediction Using Machine Learning Methods”, *Moratuwa Engineering Research Conference (MERCon) 2020.*

[3] S.Revathy, B.Bharathi, P.Jeyanthi, M.Ramesh, “Chronic Kidney Disease Prediction using Machine Learning Models”, *International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019.*

[4] Rahul Gupta, Nidhi Koli, Niharika Mahor, N Tejashri, “Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease”, *2020 International Conference for Emerging Technology (INCET)*.

[5] S. K. K. L, N. K. G and M. J. A, "Coronary Artery Disease Prediction using Data Mining Techniques," *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 693-697, doi: 10.1109/ICISS49785.2020.9316014.*

[6] R. G. Saboji, "A scalable solution for heart disease prediction using classification mining technique," *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*,2017, pp. 1780-1785, doi: 10.1109/ICECDS.2017.8389755.

[7] Aravind Akella, Sudheer Akella, “Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution”, *Future Science, 29 March 2021*

[8] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *2021 6th International Conference on Inventive Computation Technologies (ICICT),2021,pp.1329-1333,doi: 10.1109/ICICT50816.2021.9358597.*

[9] Repaka, Anjan & Ravikanti, Sai & Franklin, Ramya. (2019). “Design And Implementing Heart Disease Prediction Using Naives Bayesian”. *292-297. 10.1109/ICOEI.2019.8862604.*

[10] Rahul Katarya, Polipireddy Srinivas , “Predicting Heart Disease at Early Stages using Machine Learning: A Survey”, *International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108*

[11] Aishwarya Mujumdara, Dr. Vaidehi V , “ Diabetes Prediction using Machine Learning Algorithms ”, *INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 201*

[12] Mitushi Soni, Dr. Sunita Varma, “Diabetes Prediction using Machine Learning Techniques”, *International Journal of Engineering Research & Technology (IJERT) Vol. 9 Issue 09, September-2020*

[13] Ayush Anand, Diva Shakti, “Prediction of Diabetes Based on Personal Lifestle Indicators”, *International Conference on Next Generation Computing Technologies (NGCT-2015) Dehradun, India, 4-5 September 2015*

[14] K.VijiyaKumar, B.Lavanya, I.Nirmala, S.Sofia Caroline, “Random Forest Algorithm for the Prediction of Diabetes”, *International Conference on Systems Computation Automation and Networking 2019*

[15] Gazi Mohammed Ifraz, Muhammad Hasnath Rashid, Tahia Tazin,Sami Bourouis, Mohammad Monirujjaman Khan, “Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods”, *Hindawi Computational and Mathematical Methods in Medicine Volume 2021, Article ID 6141470*

[16] Zvi Segal , Dan Kalifa , Kira Radinsky , Bar Ehrenberg , Guy Elad , Gal Maor, “Machine learning algorithm for early detection of end-stage renal disease”, *Segal et al. BMC Nephrology (2020)*

[17] Hamida Ilyas, Sajid Ali, Mahvish Ponum , Osman Hasan , Muhammad Tahir Mahmood, Mehwish Iftikhar, “Chronic kidney disease diagnosis using decision tree algorithms”, Ilyas et al. BMC Nephrology, s12882-021-02474-z

[18] Yazan Jian, Michel Pasquier, Assim Sagahyroon Fadi Aloul, “A Machine Learning Approach to Predicting Diabetes Complications”, *Healthcare 2021, 9, 1712. https://doi.org/ 10.3390/healthcare9121712*

[19] Basha, S. M., Poluru, R. K., & Ahmed, S. T. (2022, April). A Comprehensive Study on Learning Strategies of Optimization Algorithms and its Applications. In 2022 8th International Conference on Smart Structures and Systems (ICSSS) (pp. 1-4). IEEE.

[20] Quan Zou, Kaiyang Qu, Yamei Luo, Dehui Yin, Ying Ju, Hua Tang, "Predicting diabete mellitus with machine learning technologies", *National library of medicine, nationale center for biotechnology information*

[21] Nagashree N, Premjyoti Patil, Shantakumar Patil, Mallikarjun Kokatanur, "Performance Metrics for Segmentation Algorithms in Brain MRI for Early Detection of Autism", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2S, December 2019.

[22] N. Nagashree, Premjyoti Patil, Shantakumar Patil, and Mallikarjun Kokatanur, "InvCos Curvature Patch Image Registration Technique for Accurate Segmentation of Autistic Brain Images", Springer Nature Singapore Pte Ltd. 2022 V. S. Reddy et al. (eds.), Soft Computing and Signal Processing, Advances in Intelligent Systems and Computing 1340, https://doi.org/10.1007/978-981-16-1249-7_62, pp 659-666.