IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CHRONIC DISEASES DETECTION SYSTEM

[1]Shreenidhi B S, [2]Jigyasa Mangal, [3]Kashish Mehta, [4]Keerthana R, [5]Parnika Aravind

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student

[1]Department of Computer Science Engineering

[1]Dayananda Sagar Academy of Technology and Management, Bengaluru, India

Abstract: The primary objective of our project is to develop a comprehensive health monitoring system capable of detecting the presence of critical medical conditions such as Heart Disease, Diabetes, and Parkinson's Disease. Users can input relevant health parameters, and the system employs machine learning algorithms to analyze the data, providing real-time feedback on potential health risks. This project utilizes a user-friendly interface that allows individuals to input their health information easily. The system then processes this data, leveraging machine learning and statistical models to assess the likelihood of heart disease, diabetes, and Parkinson's disease. The results are promptly communicated to the user, enabling them to take proactive measures for their well-being. The machine learning algorithms taken into consideration for detecting Heart Disease is Logistic Regression, Diabetes is Random Forest and Parkinson's Disease is Support Vector Machine. Once the data is taken from the users for the selected disease detector, the values will be used by the classifiers to detect if the individual has the particular disease or not. This system is implemented using Python programming and Streamlit library for integrating the models to front end.

Index Terms - Chronic Diseases, Diabetes, Health Monitoring System, Heart Disease, Logistic Regression, Machine Learning Algorithm, Parkinson's Disease, Python programming, Random Forest, Streamlit, Support Vector Machine.

I. INTRODUCTION

In today's world, Healthcare is evolving at a rapid pace. It is becoming incredibly crucial to detect and manage critical medical conditions proactively. The prevalence of these medical conditions emphasizes the demand for personalized health monitoring solutions. In response to this need, our project is dedicated to developing a robust health monitoring system designed to identify the presence of Heart Disease, Diabetes, and Parkinson's Disease. Heart disease, a serious health concern, refers to a range of conditions affecting the heart and blood vessels. Diabetes, a prevalent health condition, involves the body's difficulty in regulating blood sugar levels. Parkinson's disease, a complex neurological condition, affects movement control and can have a profound impact on one's daily life. All of these chronic diseases are the leading cause of health issues globally. Our approach takes into consideration the active role of the individuals in their wellbeing. Users can easily input relevant health parameters into our system, which, powered by advanced algorithms, meticulously analyzes these values. The goal is to pinpoint potential risks linked to Heart Disease, Diabetes, and Parkinson's Disease, promoting health awareness and enabling early detection and intervention. The significance of early detection cannot be overstated, and our project stands as a testament to the transformative potential of preventive healthcare. By deciphering potential risks embedded in user-inputted data, our system not only aids in early intervention The ultimate goal is to contribute to the progression of preventive healthcare on a societal scale. By leveraging technology to enhance health awareness, promote early detection, and facilitate informed decision-making, our project aspires to shape a future where individuals are not just recipients of healthcare but active stewards of their own well-being.

II. PRIOR AND RELATED WORK

Nuria Caballé-Cervigón, José L. Castillo-Sequera, Juan A. Gómez-Pulido, José M. Gómez-Pulido and María L. Polo-Luque in **[1]** makes a comparative study on all the commonly occurring chronic diseases. The study involves searching for different techniques to detect the presence of different types of diseases. A comparison has been made of all the models which prove effective in detecting a particular disease. A systematic review is provided of the intelligent data analysis tools existing in the medical field. They have analysed the possible trends in the focused application areas and provided a few algorithms used in those areas of the medical field. Moreover, advantages and disadvantages of each technique is laid out to provide information about which technique is most suitable. Additionally, relationships between all technique is provided.

Shrinivas D. Desai, Shantala Giraddi, Prashant Narayankar, Neha R. Pudakalakatti and Shreya Sulegaon in **[2]** has provided a detailed study of Heart Disease. The objectives are to identify the best classification model among parametric and nonparametric for effective heart disease prediction and optimizing the number input attributes for classification model which were achieved by various systematic and engineering approaches. Experiments were conducted to identify the best classifier for predicting the existence of heart disease. The dataset taken into consideration was the Cleveland heart disease data set. In this, 13 fields and 270 records are present which include the most influencing attributes.

Mohammed Jawwad Ali Junaid and Dr. Rajeev Kumar in **[3]** has attempted to predict the possibility of heart disease at an early stage. A hybrid algorithm-based prediction of heart disease is proposed. Additionally, suitable suggestion to the patient is proposed. The model has increased attributes namely BMI, waist circumference, family history, and physical activities in routine life. This Hybrid model has shown a sensitivity and higher specificity by having 82.11% and 91.47 % respectively. This research showed a new path to the research area. Revolution of the cure and heart disease diagnosis can be achieved by these concepts.

Nour Abdulhadi Amjed Al-Mousa in **[4]** has studied the various methods that can be incorporated to detect Diabetes. The main objective is to detect the possible presence of diabetes -specifically in females- at an early stage with the help of different machine learning techniques. The main aim is to help assist doctors in detecting diabetes at an early stage to save a patient's life. This is achieved by studying the various models that will help in accurate detection and build a model using supervised learning methods. The technique which is used is the Random Forest Classifier which achieved a highest accuracy of 82%.

Anirudh Hebbar P, Manoj Kumar M V and Sanjay H A in **[5]** have used a system called DRAP in detecting the Diabetes disease. DRAP used recorded data of the existing patients and classified the unknown instances into one of the two classes - yes or no (whether diabetic or not). The algorithms considered for developing the system is a hybrid of random forest and decision tree. The classification and prediction are based on the feature set mainly consisting of BMI, age, blood pressure, insulin level, and glucose level. Modified random forest and decision tree algorithm for learning, classification and prediction has been used. The accuracy observed is 75% and 72% for random forest and decision tree respectively.

K.VijiyaKumar, B.Lavanya, I.Nirmala, S.Sofia Caroline in **[6]** have put systematic efforts in creating a model which predicts illness due to genetic defect. The algorithm used throughout the work is the Random Forest Algorithm. The units were tested on varied measures. The objective of the project was to build a model using machine learning techniques for early prediction of diabetes for a patient with a higher accuracy. It provides a high accuracy rate and thus provides advanced support.

Wu Wang, Junho Lee, Fouzi Harrou and Ying Sun in **[7]** have proposed a deep learning model to automatically discriminate normal individuals and patients affected by Parkinson's Disease based on premotor features. Good detection capacity of the system deep learning model was observed with an accuracy of 96.45%. The desirable characteristics of deep learning model helped in achieving a high accuracy. The boosting methods also provide comparable performances. The work can be viewed as a promising first step towards the application of cutting-edge research for early disease detection.

Surendrabikram Thapa, Surabhi Adhikari, Awishkar Ghimire and Anshuman Aditya in **[8]** have utilized Twin-Support Vector Machine (TVSM) for the initial stage detection of Parkinson's Diseases. TSVM has proved to be a better classifier for binary classification problems like delineation of Parkinson's disease and healthy subjects. Using TSVM for feature selection has shown promising results in classification of Parkinson's disease patients from healthy individuals. Experiments were conducted with various learning algorithms to get an overview of the performance of TSVM as compared to traditional ML algorithms. For each of the learning algorithms, 10-fold cross-validation has been done for obtaining more consistent results. Of all, TVSM provided better results with a higher accuracy of 0.932 without feature selection and 0.939 with feature selection.

III. DETECTION METHOD

3.1 Mechanism for Detection

a) Heart Disease: For the detection of Heart Disease, Logistic Regression is used. Using Logistic Regression, Heart Disease detection involves linking health indicators like age and cholesterol to the likelihood of the condition. Once data preprocessing and model training is completed, coefficients unveil each feature's impact. This method provides a straightforward approach, which emphasizes the simplicity and interpretability in medical predictions.

b) Diabetes: The detection of Diabetes is achieved through the usage of Random Forest Algorithm. The Random Forest algorithm will help in detection by training decision tree ensembles on user-input health data to predict the likelihood of diabetes. With its ability to handle non-linear relationships, Random Forest provides accurate and interpretable results.

c) Parkinson's Disease: For the detection of the presence of Parkinson's Disease, Support Vector Machine is utilized. It involves training the model on the dataset containing relevant features such as tremor amplitude and speech characteristics. This model detects the presence of Parkinson's disease based on the trained model and the data which the user has provided as input. With its ability to find optimal hyperplanes for classification, Support Vector Machine provides accurate results, making it a valuable tool for efficient and interpretable Parkinson's disease detection.

3.2 Mechanism for taking user input

Once, the specific models are trained and tested with respect to the values of the chronic disease, the models used for detection are first extracted. The extracted models are then integrated with User Interface using Streamlit and the necessary libraries. Separate sections are maintained for each of the disease. Labels are created to help user identify the key values to be entered. This way, a user-friendly interface is developed where the users can input the data which is required for the identification of presence of the disease and receive the appropriate result.

IV. EXISTING SYSTEM

Many systems exist with respect to the detection of chronic diseases like Heart Disease, Diabetes and Parkinson's Disease. An existing system for Heart Disease is the Framingham Heart Study Risk Calculator which is based on the inventor's findings. It estimates a person's risk of developing cardiovascular disease within a certain time frame. Another existing system for Diabetes is the Indian Diabetes Risk Score (IDRS) which is a less complex used tool in India for prediction of the risk of development of type 2 diabetes. Finally, in the field of Parkinson's Disease, the Parkinson's Progression Markers Initiative (PPMI) is a research initiative which collects extensive data, including clinical, imaging and biological data, to identify and improve the knowledge on Parkinson's disease progression. It is not a diagnostic tool but helps in providing valuable insights for researchers. As per the observation made, few existing systems detects or predicts the chronic diseases which is for a particular region and for a particular community. Moreover, the detection or prediction of diseases is done only for particular diseases. Systems in place use different platforms for different disease detection.

4.1 Comparison

TITLE	OBJECTIVE	TECHNOLOGIES USED	ADVANTAGES	LIMITATIONS OR FUTURE ENHANCEMENTS
[1] Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review (2020)	Focusing on modern techniques related to the development of Machine Learning applied to the diagnosis of human diseases in order to discover useful parameters in decision making	Different technologies for the different types of human diseases.	This helps in identifying the algorithms and classifiers which help detect the presence of diseases accurately.	One specific model recommendation for a particular type of human disease.
[2] Back-Propagation Neural Network Versus Logistic Regression in Heart Disease Classification (2019)	To assess the accuracy of classification model for the prediction of heart disease for Cleveland dataset.	Binarized Neural Network (nonparametric) and Logistic Regression (parametric) models.	Accuracy of the model which uses Logistic Regression is very high (about 92.58%).	1. Cannot be a replacement to the clinical experts. 2. The futuristic scope of this project is to collect the dataset from local regional healthcare unit, along with additional input attributes such as smoking/alcoholic habits, food habit, and genetic conditions.
[3] Data Science And Its Application In Heart Disease Prediction (2020)	To make a hybrid-algorithm based model for the heart disease prediction at very early stage	Naive Bayes NB, Support Vector Machine SVM and other useful algorithms	This Hybrid model also shows higher specificity and sensitivity by having 82.11% and 91.47 % respectively.	1. System is very complex. 2. Training time is more. 3. Computational resources required is more.
[4] Diabetes Detection Using Machine Learning Classification Methods (2021)	To predict the possible presence of diabetes - specifically in females- at an early stage using different machine learning techniques.	Random Forest Classifier	1. Highest accuracy of 82%. 2. Handles missing values while training.	1. Computational Complexity 2. Memory usage is more

[5] DRAP: Decision Tree and Random Forest Based Classification Model to Predict Diabetes (2019)	Usage of DRAP for early detection and classification of Diabetes.	DRAP- Decision Tree and Random Forest	1. Feature Importance Analysis. 2. Combination of two powerful algorithms.	1. Computationally intensive. 2. Probability of redundancy while training or testing.
[6] Random Forest Algorithm for the Prediction of Diabetes (2019)	To develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by using machine learning technique which provides advance support for predicting the accuracy rate of diabetes.	Random Forest	1. High accuracy. 2. Tolerant to missing values while training.	Memory usage.
[7] Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning (2020)	To propose a deep learning model to automatically discriminate normal individuals and patients affected by PD based on premotor features.	Deep Learning Techniques	Accuracy of 96.45% is achieved.	1. Overfitting 2. Complexity is more.
[8] Feature Selection Based Twin-Support Vector Machine for the Diagnosis of Parkinson's Disease (2020)	To use Twin-Support Vector Machine (TVSM) for the early and accurate detection of Diabetes.	Twin-Support Vector Machine (TVSM)	A better classifier for binary classification problem like delineation of Parkinson's disease and healthy subjects.	1. Binary Classification only. 2. Dependency on data quality.

V. CONCLUSION

The main aim of the project is to revolutionize chronic diseases detection through a thoughtful integration of machine learning techniques. The plan is to employ Logistic Regression for Heart Disease, Random Forest for Diabetes, and Support Vector Machine for Parkinson's Disease. The choice of algorithm is based on the observation derived from the existing models and papers with respect to accuracy, ease of the usage of algorithms and several other factors which helps in accurate detection of the diseases. This project strives to redefine early detection by not only embracing cutting-edge technology but also embodying a commitment to personalized and proactive healthcare.

REFERENCES

[1] Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review, Nuria Caballé-Cervigón, José L. Castillo-Sequera, Juan A. Gómez-Pulido, José M. Gómez-Pulido and María L. Polo-Luque, 2020**.**

[2] Back-Propagation Neural Network Versus Logistic Regression in Heart Disease Classification, Shrinivas D. Desai, Shantala Giraddi, Prashant Narayankar, Neha R. Pudakalakatti and Shreya Sulegaon, 2019.

[3] Data Science And Its Application In Heart Disease Prediction, Mohammed Jawwad Ali Junaid and Dr. Rajeev Kumar, 2020.

[4] Diabetes Detection Using Machine Learning Classification Methods, Nour Abdulhadi and Amjed Al-Mousa, 2021.

[5] DRAP: Decision Tree and Random Forest Based Classification Model to Predict Diabetes, Anirudh Hebbar P, Manoj Kumar M V and Sanjay H A, 2019.

[6] Random Forest Algorithm for the Prediction of Diabetes, K.VijiyaKumar, B.Lavanya, I.Nirmala, S.Sofia Caroline, 2019.

[7] Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning, Wu Wang, Junho Lee, Fouzi Harrou and Ying Sun, 2020.

[8] Feature Selection Based Twin-Support Vector Machine for the Diagnosis of Parkinson's Disease, Surendrabikram Thapa, Surabhi Adhikari, Awishkar Ghimire and Anshuman Aditya, 2020.