IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Multiple Disease Prediction System Using Machine Learning

[1]Sandeep Singh
[1]Assistant Professor
[1]Maharaja Surajmal Institute of Technology, New Delhi, India

Abstract: Machine learning and Artificial Intelligence is playing a huge role in today's world. From self-driving cars to medical fields, we can find them everywhere. The medical industry generates a huge amount of patient data which can be processed in a lot of ways. So, with the help of machine learning, we have created a Prediction System that can detect more than one disease. Many of the existing systems can predict disease at a time with lower accuracy. Lower accuracy can seriously put a patient's health in danger. We have considered three diseases for now that are Heart, Parkinson's, and Diabetes and in the future, many more diseases can be added. The user has to enter various parameters of the disease and the system would display the output whether he/she has the disease or not. This project can help a lot of people as one can monitor the persons' condition and take the necessary precautions thus increasing the life expectancy

Index Terms - Diabetes, Heart, Parkinson, SVM, KNN, Logistic Regression, Disease Prediction

I. INTRODUCTION

In the dynamic landscape of contemporary healthcare, the integration of advanced technologies plays a pivotal role in revolutionizing diagnostic and predictive methodologies. One such groundbreaking advancement is the development of Multiple Disease Prediction Systems (MDPS), a cutting-edge approach designed to enhance early detection and intervention strategies. This research paper delves into the intricacies of MDPS, exploring its significance, challenges, and potential impact on public health. The ever-growing complexity of diseases necessitates a shift from traditional reactive healthcare models to proactive and preventive paradigms. Multiple Disease Prediction Systems represent a paradigm shift in this direction, leveraging data analytics, machine learning algorithms, and artificial intelligence to analyze vast datasets and discern patterns that may indicate the onset of various medical conditions. By harnessing the power of predictive analytics, MDPS offers the potential to identify health risks at an early stage, enabling timely and targeted interventions. This research endeavors to provide a comprehensive understanding of the technological underpinnings of MDPS, examining the integration of diverse data sources, including electronic health records, genetic information, and lifestyle data. As we navigate the era of precision medicine, the promise of Multiple Disease Prediction Systems in reshaping the healthcare landscape is profound. This research paper seeks to shed light on the transformative potential of MDPS, offering insights into its current state, challenges, and future directions. By doing so, it aspires to contribute to the ongoing dialogue surrounding innovative healthcare solutions and the realization of a proactive and personalized approach to disease prevention and management.

II. LITERATURE REVIEW

The main aim of the paper is, as the heart plays an important role in living organisms. So, the diagnosis and prediction of heart related disease should be perfect and correct because it is very crucial which can cause death cases related to heart. So, Machine learning and Artificial Intelligence supports in predicting any kind of natural events. So, in this paper they calculate accuracy of machine learning for predicting heart disease using decision tree, linear regression and SVM. They also compared the algorithm and their accuracy SVM 83 %, Decision tree 79%, Linear regression 78% [1]. In the dynamic landscape of diabetes prediction through machine learning, diverse algorithms have been scrutinized for their efficacy in early disease detection. Decision Tree algorithms have an accuracy of approximately 80%, demonstrating their proficiency in discerning patterns within multifaceted datasets encompassing clinical records, genomic information, and lifestyle factors. Similarly, Logistic Regression has exhibited noteworthy accuracy, with an impressive rate of approximately 81%, underscoring its practicality in identifying key risk factors for diabetes. Additionally, Support Vector Machines (SVM) have played a pivotal role, yielding an accuracy of approximately 77.3% and showcasing their ability to capture intricate patterns indicative of diabetes. These collective insights illuminate the pivotal role of machine learning algorithms in reshaping diabetes prediction, providing a proactive and personalized framework for healthcare interventions. [2].

In the realm of Parkinson's disease prediction using machine learning, recent studies have employed diverse algorithms to achieve accurate and early detection. Notably, Decision Tree algorithms have demonstrated clinical utility with an accuracy of approximately 80%, shedding light on early indicators of Parkinson's disease. Logistic Regression, with an accuracy of approximately 78%, stands out for its practicality in identifying key risk factors associated with Parkinson's disease. The transparency and interpretability of Logistic Regression make it a valuable tool in clinical contexts. Additionally, Support Vector Machines (SVM), recognized for its ability to capture complex patterns, has proven effective with an accuracy of approximately 81%. These findings underscore the significance of machine learning in advancing predictive models for Parkinson's disease, providing clinicians with valuable tools for timely intervention and improved patient outcomes. [3-5].

III. SYSTEM ANALYSIS

Functional Requirement-

- The system allows the patient to predict the disease
- The user adds the input for the particular disease and based on the trained model of the user input the output will be displayed.

Non-Functional Requirement-

- The website will provide a range of the values during the prediction of the disease.
- The website should be reliable and consistent.

IV. DESIGN

I. *ARCHITECTURE DESIGN*

Initially the dataset is collected and prepared the data by per-processing that includes removing or replacing null values, removing outliers, standardizing the data etc. Now the data is split into training and testing data in the ratio 80:20 or 70:30. Now the model is trained with training data. If under-fitting or over-fitting occurs then the model is changed and try with another model likewise, we do the training. Now we choose the models for testing purposes that give us the best training accuracy. If the results are accurate then our model is ready for deployment. Figure 1. shows the complete flow chart of the machine learning multiple disease process (MLMDP).

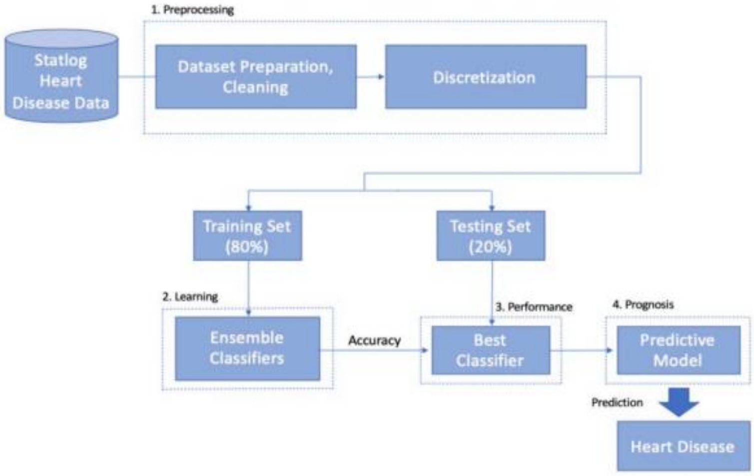

Fig 1: Architecture diagram for MLMDP system

II. USER INTERFACE DESIGN

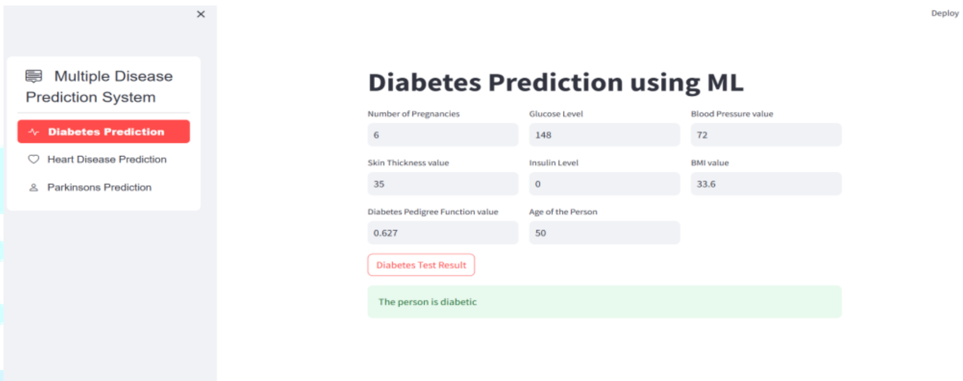

Fig 2: User Interface Design

V. RESULTS AND DISCUSSIONS

Evaluation Methods For the experiment's performance assessment. First, we identify True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). True positive refers to the number of cases successfully forecasted as necessary, false positive refers to the number of instances mistakenly predicted as required, and so on. The following are the four measurements that are obtained: accuracy, precision, recall, and F1-measure.
Accuracy = (TP+TN)/(TP+TN+FP+FN)
Precision = TP/(TP+FP)
Recall = TP/(TP+FN)
F1 Score = TN/(TN+FP)

The results of all the employed machine algorithms is presented in Table 1. Overall, the results indicate that Support Vector Machine (SVM) consistently delivers high performance across different disease datasets, achieving the highest accuracy for both. Heart Disease (83.12%) and Parkinson's Disease (81.25%). This suggests that SVM is particularly effective in handling complex, high-dimensional data often seen in medical diagnoses. Logistic Regression performs best for Diabetes with an accuracy of 81.50%, showing that even simpler, linear models can be highly effective when the data has clear linear patterns. In contrast, K-Nearest Neighbors (KNN) consistently shows the lowest accuracy across all three diseases, which may be due to its sensitivity to noisy data and the effects of high-dimensionality, making it less suitable for these types of medical datasets. Decision Tree algorithms perform moderately well but do not achieve the highest accuracy in any category, indicating that while they are interpretable and useful, they may benefit more from being part of ensemble methods like Random Forests. Overall, SVM stands out as the most reliable algorithm across the board, followed by Logistic Regression, with KNN being the least effective in this context. The accuracy is also demonstrated in terms of bar graph in Fig. 3.

Table 1: Disease based results of MLMDP system

Disease Name	Algorithm	Accuracy (%)	Precision (%)	F1 Score (%)
Heart Disease	SVM	83.12	84.50	83.70
	Decision Tree	79.40	78.10	78.50
	KNN	74.31	73.50	73.80
	Logistic Regression	78.00	77.00	77.40
Diabetes	SVM	77.30	76.80	77.00
	Decision Tree	80.70	80.20	80.40
	KNN	76.56	75.90	76.10
	Logistic Regression	81.50	81.00	81.20
Parkinson’s	SVM	81.25	82.00	81.50
	Decision Tree	80.62	80.00	80.30
	KNN	76.12	75.50	75.80
	Logistic Regression	78.25	77.60	77.90

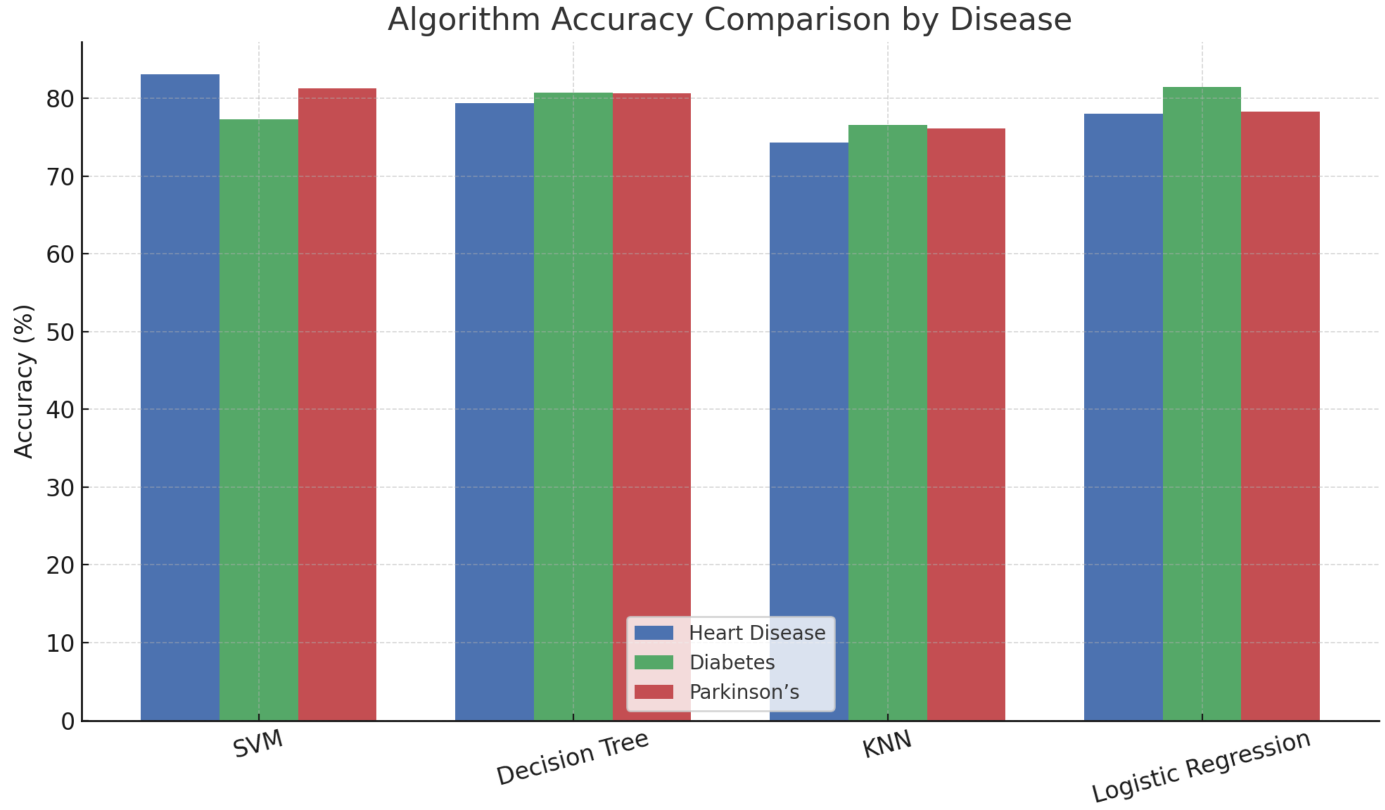

Fig. 3: Accuracy results of algorithms.

VI. Conclusion

In conclusion, this research paper has navigated the intricate landscape of developing a Multiple Disease Prediction System (MDPS) leveraging advanced machine learning techniques. The exploration of algorithms, including Support Vector Machines (SVM), Decision Trees, KNN, and Logistic Regression, has provided a comprehensive understanding of their roles in predicting diverse health conditions. The significance of early disease detection cannot be overstated, and the results presented herein demonstrate the potential for these models to revolutionize healthcare practices. The comparison of these algorithms in the contexts of diabetes, heart and Parkinson's disease highlights their varying strengths and accuracies, enabling a nuanced selection based on the specific requirements of each medical condition.

REFERENCES

[1] Sonar, P., & Jaya Malini, K. (2019, March). Diabetes prediction using different machine learning approaches. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 367-371). IEEE

[2] Singh, A., & Kumar, R. (2020, February). Heart disease prediction using machine learning algorithms. In 2020 international conference on electrical and electronics engineering (ICE3) (pp. 452-457). IEEE

[3] Pratihar, R., & Sankar, R. (2024). Advancements in Parkinson’s Disease Diagnosis: A Comprehensive Survey on Biomarker Integration and Machine Learning. Computers, 13(11), 293.

[4] Ibrahim, A. M., & Mohammed, M. A. (2024). A comprehensive review on advancements in artificial intelligence approaches and future perspectives for early diagnosis of Parkinson's disease. International Journal of Mathematics, Statistics, and Computer Science, 2, 173-182.

[5] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Computers in Biology and Medicine, 136, 104672.