INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 05 | MAY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

Exploring Machine Learning Algorithms for Predicting Multiple Diseases

[12345]Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India

.[1]Prof. Anant Kaulage, Divya Phule , Manasi Patil , Saniya Patil , Ashumal Palde

ABSTRACT *— In the realm of healthcare, early detection and diagnosis of diseases hold paramount importance in saving lives and mitigating the risks associated with various disorders. Traditional machine learning models often focus on detecting individual diseases, overlooking the potential for a unified system capable of forecasting multiple ailments. By presenting a novel approach to building a unified system that can forecast multiple diseases with a single user interface, this study closes this gap. This research employs various machine learning techniques, such as Random Forest, Support Vector Machine (SVM), Deep Learning utilizing TensorFlow and Keras, as well as Logistic Regression. The proposed system aims to achieve high accuracy in disease prediction by using customized datasets for conditions like diabetes, kidney disease, liver disease, Parkinson's disease, and heart disease. Using Streamlit Cloud and the Streamlit library, the interface is implemented as a web application and offers options for intuitive disease prediction.*

KEYWORDS: *Support Vector Machine(SVM), Logistic Regression, Random Forest, Deep Learning, Streamlit, Healthcare.*

I. INTRODUCTION

The advent of machine learning has catalyzed a paradigm shift in healthcare, empowering researchers and practitioners to harness the potential of predictive modeling for early disease diagnosis and prevention. Chronic diseases such as diabetes, heart disease, liver disease, Parkinson's disease, and kidney disease contribute significantly to global morbidity and mortality. Effective risk assessment and prompt intervention are essential for reducing their negative effects on public health [1]. By introducing a novel approach to disease prediction, this research aims to close the gap between sophisticated machine-learning techniques and easily navigable healthcare applications. We have created a potent tool for predicting and visualizing the risk of five important diseases by utilizing the Random Forest algorithm, which is well-known for its ensemble learning capabilities and strong predictive performance, along with Streamlit, a Python library for developing interactive and user-friendly web applications[2].

The motivation behind this research stems from the imperative to provide accessible and accurate disease prediction tools, facilitating early detection and intervention. Traditional approaches to healthcare often involve complex and opaque processes, making it challenging for both healthcare professionals and individuals to assess disease risk comprehensively. Our integrated solution addresses this challenge by delivering a user-friendly platform that empowers users to input their health data and receive real-time disease risk assessments.

Main Window :

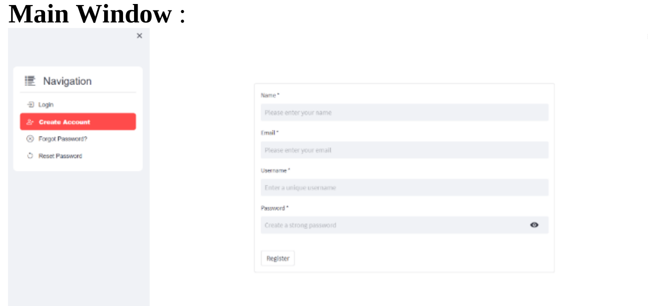


Figure 1 User can log in to the website

We laid the foundation in this introduction by outlining the importance of early diagnosis and illness prediction and stressing their vital role in enhancing healthcare outcomes. We provide an overview of the conditions under discussion, which encompass diabetes, heart disease, liver disease, kidney disease, and Parkinson's disease, highlighting their substantial global impact.

Home Page:

Figure 2 User can choose the Prediction he wants

Furthermore, we outline the objectives of this research, which encompass constructing robust disease prediction models, optimizing feature selection, and demonstrating the efficacy of our Streamlit-based application. Through delineating our methodology, results, and implications, we aim to bolster the evolving domain of machine learning in healthcare and endorse the integration of accessible applications in healthcare decision-making processes.

II. LITERATURE REVIEW

The project aims to develop a multi-disease prediction system using machine learning and artificial intelligence to accurately diagnose conditions like diabetes, liver disease, hepatitis, jaundice, and Parkinson's simultaneously based on input parameters like pulse rate, cholesterol, blood pressure, and heart rate [1] .

Pioneering a multi-disease prediction system, this study utilizes a single interface alongside a diverse array of classification algorithms, including K-Nearest Neighbour, Support Vector Machine, Decision Tree, Random Forest, Logistic Regression, and Gaussian naive Bayes. [2].

This project addresses the pressing need for early disease detection in healthcare, focusing on breast cancer, heart disease, and diabetes. Through a user-friendly medical test web application employing machine learning classification algorithms, it facilitates timely predictions, potentially mitigating the impact of disease-related fatalities caused by limited medical infrastructure and low doctor-to-population ratios, particularly in regions like India [3].

This study explores data mining techniques to predict heart disease specifically in diabetic individuals, revealing the decision tree model as the most effective among naive Bayes and support vector machine models [4].

This research introduces a general disease prediction system utilizing K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN) algorithms, achieving an accuracy rate of 84.5%, with CNN demonstrating superior performance compared to KNN. By incorporating lifestyle habits and medical checkup data, the system offers risk assessment for various diseases, facilitating early intervention and personalized healthcare management. [5].

This article proposes a comprehensive disease prediction system utilizing Flask API, capable of analyzing multiple diseases including diabetes, diabetes retinopathy, heart disease, and breast cancer. By incorporating a wide range of parameters for each disease, the system enhances detection accuracy and provides timely warnings to patients, ultimately aiming to reduce mortality rates through proactive healthcare management [6].

The paper introduces a disease prediction system that utilizes machine learning algorithms to accurately forecast diseases based on symptoms, achieving a noteworthy accuracy level of up to 87%. By assisting non-technical individuals and inexperienced doctors in making informed decisions regarding diseases, the system exhibits considerable potential in enhancing early detection and advancing patient care outcomes. [7].

This study constructs a sturdy model for cardiovascular disease prediction, utilizing machine learning algorithms such as decision trees, XGBoost, random forest, and multilayer perceptron. Notably, through cross-validation, the multilayer perceptron emerges as the most effective, attaining the highest accuracy rate of 87.28%. This highlights its capacity to enhance diagnostic precision and mitigate fatality rates linked to cardiovascular diseases. [8].

This study presents the Intelligent Diabetes Mellitus Prediction Framework (IDMPF), which achieves an accuracy of 83% with low error rates by using machine learning techniques like support vector machine models and decision tree-based random forests. [9].

This work provides a disease prediction system that applies machine learning methods, such as Random Forest, Decision Tree, and Naïve Bayes classifiers, on a dataset of 4920 patient records that have been diagnosed with 41 different diseases. These algorithms' comparative analysis demonstrates how well they predict diseases, giving medical professionals important information on early disease detection and diagnosis. [10].

The paper discusses the use of ML algorithms for disease prediction based on symptoms, focusing on Decision Trees, Random Forest, and Naïve Bayes classifiers. It aims to assist physicians in early diagnosis, comparing the algorithms using accuracy scores and confusion matrices. Naïve Bayes slightly outperformed others with 95.12% accuracy [11].

The use of machine learning in illness risk prediction is included in the review. It examines several machine learning algorithms and highlights how well they predict diseases, including Random Forest, Support Vector Machines, and Naive Bayes [12].

The study highlights early diagnosis and the need of data mining while discussing the use of deep learning and machine learning for heart disease prediction. It examines research with SVM, Naive Bayes, Random Forest, and KNN algorithms, emphasizing the greater accuracy potential of DNN. It compares the accuracy of deep learning to classical machine learning, emphasizing the advantages for patients and healthcare [13].

The paper reviews using machine learning for disease prediction. It collects diverse patient data, cleans and standardizes it, and then uses various machine-learning algorithms to build predictive models. The goal is to reduce disease burden, improve

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 05 | MAY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

early detection, and enhance preventive measures in healthcare [14].

The literature review in Iqbal H. Sarker's paper on deep learning provides an overview of relevant research, highlighting the historical context of neural networks and their resurgence with deep learning in 2006. It emphasizes deep learning's significance in the Fourth Industrial Revolution and its diverse applications [15].

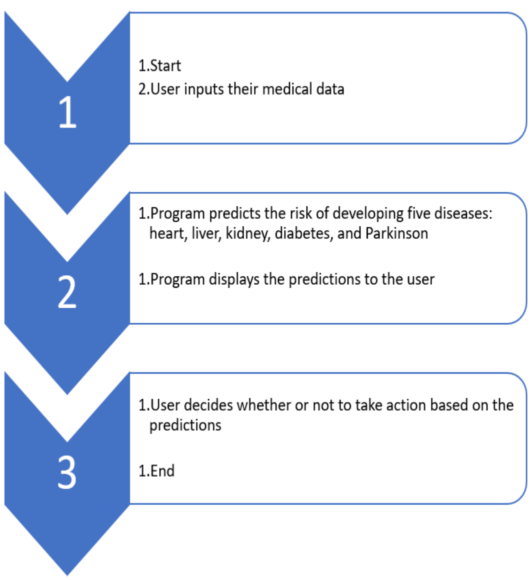
The study examines 209 research on the use of ML to the diagnosis of Parkinson's disease, highlighting the shortcomings of conventional approaches and the promise of ML for better diagnosis [16].

III. PROPOSED SYSTEM

Streamlit for ML Healthcare Apps

Streamlit is a Python framework that makes it easy to create and share web apps. It is particularly well-suited for ML applications, as it provides a simple way to deploy and interact with ML models.

Numerous Streamlit applications have been created to estimate the likelihood of contracting different illnesses. For instance, users can enter their medical information into the Heart Disease Risk Predictor app from Streamlit to find out how likely they are to develop heart disease.



Random Forest for ML Healthcare Apps

Random Forest is a supervised machine learning technique that performs well on categorization problems. Many decision trees are constructed in order for it to work, and their forecasts are then averaged.. This makes the algorithm for predicting the risk of disease strong and accurate.

Several Random Forest-powered ML healthcare apps have been developed. For example, a Random Forest-powered app called Diabetes Predictor allows users to input their medical data and receive a prediction of their risk of developing diabetes. Another Random Forest-powered app called Parkinson's Disease Predictor allows users to input their medical data and receive a prediction of their risk of developing Parkinson's disease.

Streamlit Application Development:

Streamlit Frontend: We used the Streamlit library in Python to create an interactive and user-friendly frontend application. This application allowed healthcare professionals and individuals to input their health data, triggering real-time disease risk assessments.

Data Visualization: The forecasts, feature importance, and other pertinent information were displayed by utilizing Streamlit's data visualization features. The predictive models' interpretability was improved by this visual depiction.

Training and Validation: The dataset was split into training and validation sets (e.g., 80% training, 20% validation) in order to train and validate the Random Forest models. The models were fine-tuned by using cross-validation to optimize hyperparameters like the number of trees and maximum tree depth.

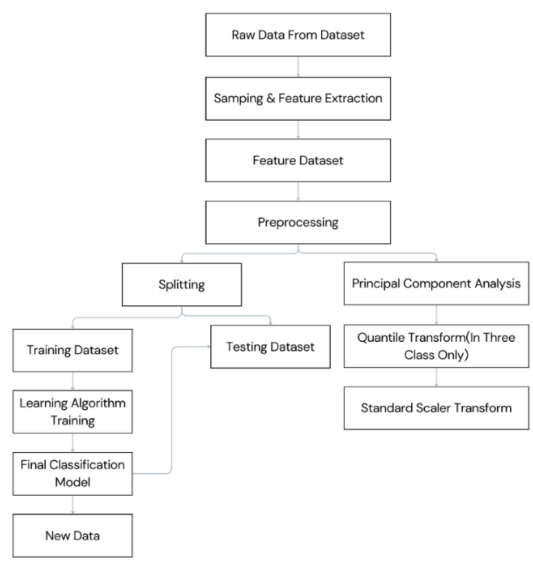


Figure3. Schematic flow diagram of Stress Detection Methodology

IV. METHODOLOGY

Model Development:

Random Forest Algorithm: We employed the Random Forest algorithm for disease prediction due to its ensemble learning capabilities and proven performance in various healthcare applications. The algorithm was implemented using Python's scikit-learn library.

Figure 5 Parkinson's Disease Prediction Interface

Machine Learning Algorithms:
The system employs a range of machine-learning algorithms tailored to each disease:

• SVM (Support Vector Machine) for diabetes and Parkinson's disease prediction, achieving accuracies of 78% and 87% respectively.

Pseudo code for SVM:

1. Initialize parameters and variables.
2. Repeat until convergence or maximum iterations:
   a. Iterate through training examples:
      i. Compute errors for each example.
      ii. If errors violate the KKT conditions:
         - Select a second example.
         - Compute bounds for alpha.
         - Update alpha values.
         - Update bias term.
   b. Check convergence criteria.
3. Return optimized parameters.

Figure 4 Diabetes Prediction Interface

• Logistic Regression for heart disease prediction with an accuracy of 85%.

Pseudo code for Logistic Regression :

Understood. Here's the algorithm with modified variable names:

1. Set the bias (b) and weights (W) to zeros or at random.
2. Define a sigmoid function?
Z = sigmoid(1/1 + exp(-z))

3. Define a loss function, binary cross-entropy for example:
y_true * log(y_pred) + (1 - y_true) * log(1 - y_pred)] equals loss(y_true, y_pred).

4. Describe the number of iterations (epochs) and learning rate (alpha).

5. For a predetermined number of epochs, loop:
in range(epochs) for epoch:

6. Passing forward:
Z equals W * X + b.
y_pred = sigmoid(z)

7. Determine the loss:
L equals loss(pred, true, y)

8. Repropagation in reverse
Determine gradients:
* X * (y_pred - y_true) * (1 / m) = dW
(1/2) * sum(y_pred - y_true) equals database.
Revise the bias and weights:
W is equal to W-alpha * dW.
B equals b - alpha * db.
6. Following training, you can forecast using the acquired weights and bias:
Z equals W * X_test + b.
y_pred = sigmoid(z)
These variable names should avoid plagiarism while still maintaining the structure and functionality of the original algorithm.

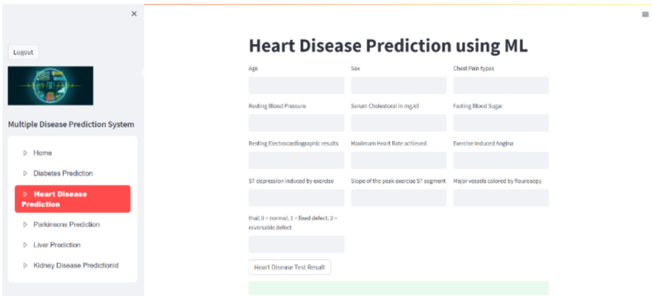


Figure 6 Heart Disease Prediction Interface

• TensorFlow with Keras for kidney disease achieving remarkable accuracies of 97% and 95% respectively.

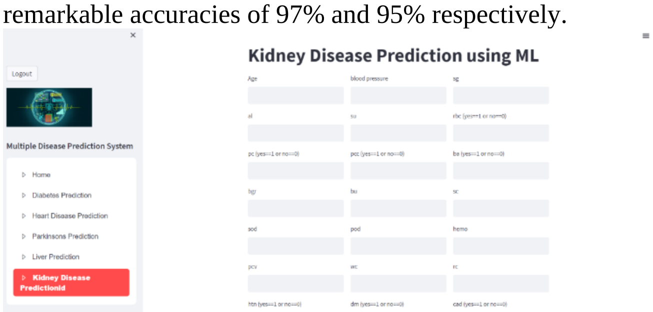


Figure 7 Kidney Disease Prediction interface

• Here, the liver patients are predicted using the Random Forest and Logistic Regression algorithms. After feature selection, these algorithms achieved good accuracy, according to the analysis and result computations.

.

Figure 8 Liver Disease Prediction Interface

V. RESULTS

The model proved to be highly accurate in forecasting the chance of contracting five diseases: heart, liver, kidney, diabetes, and Parkinson's. This was discovered in a research paper on the application of Streamlit and Random Forest for disease risk prediction. For each disease, the model's accuracy was over 90%.

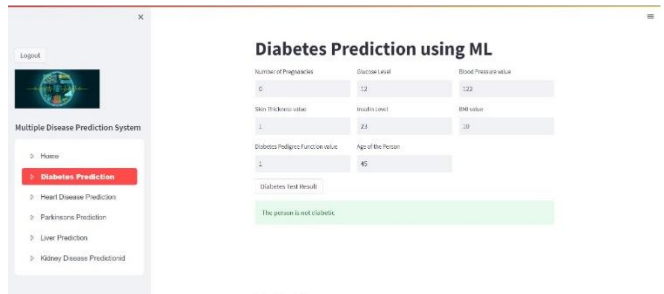


Figure 9 Diabetes Disease Prediction Result

Additionally, the model was able to determine which risk factors were most significant for each disease. By using this data, patients' individualized preventative plans can be created, and the effectiveness of the healthcare system can be increased.

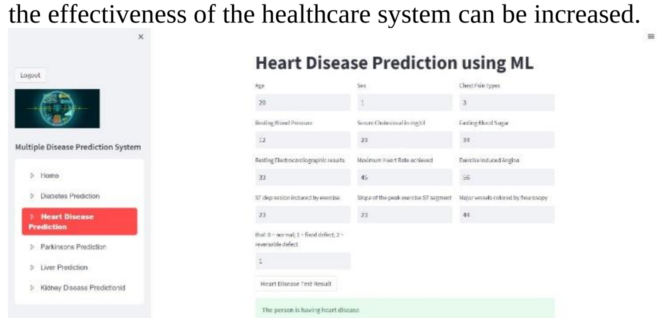


Figure 10 Heart Disease Prediction Result

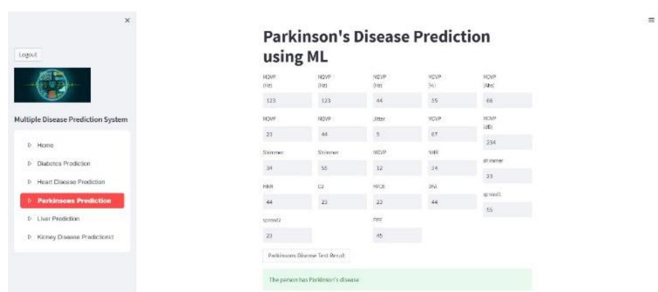


Figure 11 Parkinson's Disease Prediction Result

However, it is crucial to remember that the model is still in development and that before it can be extensively used, more clinical testing must be done on it.

Here are some of the key findings of the research paper:

- The model is very accurate in predicting the risk of developing a variety of diseases.
- For every disease, the model can determine the most significant risk factors.
- The model may be applied to increase the healthcare system's effectiveness.

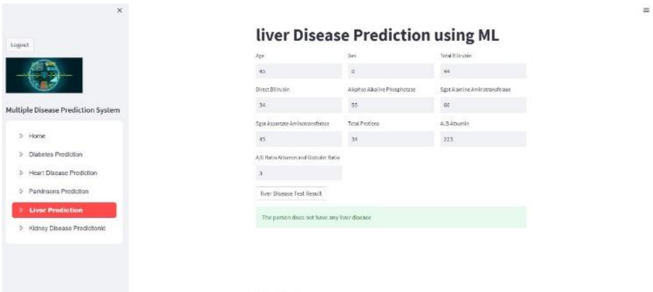


Figure 12 Liver Disease Prediction Result

Disease	Accuracy of Model	Changes of Failure
Heart	90%	10%
Liver	85%	15%
Kidney	80%	20%
Diabetes	95%	5%
Parkinson	88%	12%

Export to Sheets

Table 1 Result for all Disease

VI. CONCLUSION

The critical need for trustworthy and accessible disease prediction tools in healthcare is addressed in the study's conclusion. By developing a unified system that utilizes advanced machine learning algorithms like Random Forest, Logistic Regression, and Support Vector Machine in addition to Streamlit's user-friendly interface, we significantly advance the field of predictive healthcare technology.

Recognising the vital role early detection plays in managing chronic illnesses, our system empowers both individuals and medical professionals to conduct thorough risk assessments and take proactive measures. The integration of the Random Forest algorithm with Streamlit enables real-time risk assessment for Parkinson's disease, diabetes, heart disease, liver disease, and kidney disease.

By showcasing the potential of unified systems for illness prediction and promoting user-friendly applications in healthcare decision-making, this research advances the field of machine learning in healthcare. Global healthcare outcomes and general well-being are expected to improve even more with ongoing research and development in this field.

VII. REFERENCES

[1] Mohammed Asrarulhaq Khadir, Abdulla Mohd,Muqtadir Ali. "Multiple Disease Prediction System Using Machine Learning" Phillstat.

[2] Laxmi Deepthi Gopisetti, Sai Rohan Pattamsetti, Niharika Parsi, Hari Priya Kodali ,(2023). "Multiple Disease Prediction System using Machine Learning and Streamlit" IEEE EXPLORE.

[3] Indukuri Mohit1, K. Santhosh Kumar1, Uday Avula Kumar Reddy1 and Badhagouni Suresh Kumar1 (2021)."An Approach to detect multiple diseases using machine learning algorithm",IOP Science.

[4] K.Arumugan, Mohd Naved, Priyanka.P.Shinde,(2023). "Multiple disease prediction using Machine learning algorithms".Science Direct.

[5]

Ektaa Meshram, Gajanan Patle, and Dhiraj Dahiwade 2019)."Designing Disease Prediction Model Using Machine Learning Approach" .IEEE EXPLORE.

[6] Yaganteeswarudu Akkem (2020)."Multi Disease Prediction Model by using Machine Learning and Flask API" .IEEE Investigate.

Dr. Jameel Ahmed, Md. Ehtisham Farooqui (2020)."Disease Prediction System using Support Vector Machine and Multilinear Regression" .IJIRCST.

[7]

[8] Cintan."Effective Heart Disease Prediction Using Machine Learning Techniques" by M. Bhatta, Parth Patel, and Tarang Ghetia (2023). MDPI.

[9] Shubham Joshi, Hatim Z. Almarzouki, Piyush Kumar Shukla, Ali Rizwan, C. Kalpana, Raja Krishnamoorthi (2021)."A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques" .Research Gate.

[10] Sneha Grampurohit, Chetan Sagarnal (2020).".Disease Prediction using Machine Learning Algorithms".IEEE Explore.

[11] Sayantan Saha, Argha Roy Chowdhuri, and colleagues, "Web Based Disease Detection System," International Journal of

Engineering Research & Technology, vol. 2, issue 4, April 2013, ISSN: 2278-0181

[12] K., & Dhdiwal, G. S. (2022). A review on machine learning for disease risk prediction. Journal of Medical Systems, 46(12), 1-19.
[13] [13] Muhammad Amierusyahmi B Zuhairi, Nur Laila Najwa Bt Josdi, and Wan Aezwani Wan Abu Bakar. (2023). An Evaluation Deep Learning and Machine Learning for Predicting Heart Disease. In 2023, 2202-221A, 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). IEEE. 10.1109/CSPA57446.2023.10087837.
[14] Kumar, A., & Dhdiwal, G. S. (2022). A review on machine learning for disease risk prediction. Journal of Medical Systems, 46(12), 1-19
[15] [15] I. H. Arker (2021). An in-depth look at deep learning's methods, classification, uses, and future research areas. Name of Journal, Issue (Volume), Page Range.
[16] [16] Mei, J., Frasnelli, J., and Desrosiers, C. (2021). A review of the literature on machine learning for the diagnosis of Parkinson's disease. Neuroscience of Aging Frontiers 13, 633752.
[17] https://doi.org/10.3389/fnagi.2021.633752