Journal of Informatics
Electrical and Electronics Engineering, 2024,
Vol. 05, Iss. 02, S. No. 110, pp. 1-13
ISSN (Online): 2582-7006

Multiple Disease Prediction System using Machine Learning

Ahad Rehman[1], Shikha Singh[2], Vineet Singh[3], Bramah Hazela[4]

[1,2,3,4]Department of Computer Science and Engineering, Amity School of Engineering and Technology, Lucknow, Amity University, Uttar Pradesh, India
[1]ahad.rehman@s.amity.edu, [2]ssingh8@lko.amity.edu, [3]vsingh@lko.amity.edu, [4]bhazela@lko.amity.edu

How to cite this paper: A. Rehman, S. Singh, V. Singh, B. Hazela, "Multiple Disease Prediction System using Machine Learning," *Journal of Informatics Electrical and Electronics Engineering* (*JIEEE*), Vol. 05, Iss. 02, S No. 110, pp. 1–13, 2024.

https://doi.org/10.54060/a2zjournals.jieee.110

Received: 05/01/2024
Accepted: 30/04/2024
Online First: 27/05/2024
Published: 31/12/2024

Copyright © 2024 The Author(s).
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Machine learning advancements have spurred a revolution in healthcare by making it possible to create prediction models for early disease diagnosis. This report introduces the Multiple Disease Prediction System (MDPS), a state-of-the-art approach that uses machine learning to forecast the likelihood of several diseases based on patient data, including medical history, lifestyle, and demographics. The MDPS addresses the growing difficulties in healthcare by focusing on the early detection of multiple diseases. Some of its crucial components include data preparation, feature selection, model training and disease Detection. Despite advantages like early detection and cost savings, dealing with data privacy, model interpretability, and continuous improvements is essential for MDPS's ethical and efficient usage in healthcare. As a result, MDPS has a significant potential to enhance public health and minimize the difficulties associated with chronic illnesses.

Keywords

Multiple Disease, machine learning, random forest classifier, KNN, logistic regression, liver disease, kidney disease, heart disease, diabetes.

1. Introduction

The integration of machine learning (ML) techniques into healthcare has revolutionized disease prediction and treatment, with ML algorithms driving the creation of numerous disease prediction systems. These systems utilize data-driven methodologies to assess targeted multiple medical factors simultaneously, enabling proactive healthcare through early identification, personalized risk assessment, and therapies. This introduction highlights the transformative potential of ML-based multiple illness prediction systems in improving patient outcomes and advancing medical practices.

1.1 Objective

The objective of this report is to develop and implement a smart multi disease prediction system leveraging machine learning algorithms. The focus is on creating a robust system that utilizes patients' data and predicts to provide an accurate disease prediction. By harnessing advanced technology, the aim is to empower medical field with precise prediction and detection of diseases guidance using multiple entries from the user, optimizing accuracy and enhancing overall system scalability and sustainability.

1.2. Motivation

The motivation behind choosing the Multiple Disease Prediction System (MDPS) stems from the urgent need to tackle healthcare challenges, especially in early disease detection and prevention. Integrating machine learning into disease prediction aims to empower individuals and healthcare providers with precise insights for early disease identification. By harnessing advanced machine learning techniques, MDPS offers personalized disease predictions based on diverse datasets, fostering proactive health management and optimizing healthcare resources. The focus is on enhancing decision making in healthcare, reducing risks of delayed or inaccurate diagnoses, and ultimately improving patient outcomes. Leveraging technology for tailored disease predictions is crucial in transforming healthcare practices and advancing precision medicine goals.

1.3. Problem Statement

The current healthcare landscape relies heavily on disease-specific machine learning models, resulting in fragmentation and inefficiencies for consumers and professionals. Managing multiple models for various illnesses poses challenges and risks inaccuracies, impacting patient outcomes. An integrated multi-disease prediction system is essential to streamline processes, provide comprehensive health insights, enhance diagnosis accuracy, and improve operational efficiency for both patients and physicians.

1.4. About Multiple Disease Prediction System

In the evolving landscape of healthcare, the integration of cutting-edge technology and data-driven methodologies has spurred a shift towards proactive disease management and prevention. This research endeavors to explore a Multiple Disease Prediction System (MDPS), recognizing the transformative potential of machine learning in healthcare practices. By leveraging diverse datasets and advanced machine learning techniques, the MDPS aims to transcend conventional diagnostic boundaries, offering personalized disease predictions. Emphasizing practical applicability, the project envisions an intuitive user interface for seamless integration into clinical decision-making processes, fostering collaboration between experts and healthcare professionals. Ethical considerations and data privacy are paramount, with robust security measures integrated into the system. The success of the MDPS hinges on its adaptability to evolving healthcare data, aiming to redefine healthcare boundaries through innovative, data-driven solutions that positively impact patient lives.

1.5. Multiple Diseases

The prevalence of diseases due to lifestyle and environmental factors necessitates early prediction of sickness. Relying solely on symptoms for accurate diagnosis poses limitations for doctors. Timely and precise examination of health issues is crucial yet challenging for prevention and treatment. An Automated Disease Prediction System is essential to address these challenges, especially in critical illness cases.

(a) Diabetes

Diabetes, a prevalent metabolic disorder, remains a global health priority due to its impact on individuals and healthcare systems. Recent advancements in diabetes research emphasize personalized medicine, innovative treatments, and preventive strategies. Breakthroughs in understanding genetic and environmental factors drive innovation in diabetes management, including technologies like continuous glucose monitoring and artificial pancreas devices. Public health initiatives focus on promoting healthy lifestyles and early detection to mitigate diabetes risk factors. However, challenges persist in ensuring equitable access to diabetes care, necessitating efforts to address socioeconomic disparities and improve healthcare infrastructure globally.

(b) Heart Disease

Heart disease, a leading cause of mortality, poses a significant threat due to its impact on vital organs. With hectic lifestyles and neglect of health, individuals are increasingly susceptible to cardiac ailments. Globally, heart-related diseases contribute to over 31% of all deaths, highlighting the urgency of addressing this issue. However, prediction and diagnosis remain challenging, especially in resource limited settings with limited diagnostic tools and medical expertise. Patients often face inconvenience and delays in diagnosis, hindering timely treatment initiation. This research aims to enhance heart disease diagnosis to mitigate its grave consequences, recognizing the critical importance of the heart in human health.

(c) Kidney Disease

Like Kidney disease, marked by impaired kidney function and damage to filtering units, is a major global health concern exacerbated by factors diabetes and hypertension. Recent research breakthroughs have led to targeted therapies and personalized treatment for conditions such as chronic kidney disease and glomerulonephritis. Innovations in kidney transplantation techniques, along with advancements in regenerative medicine, offer hope for improved outcomes, particularly for patients in advanced stages. Efforts to raise awareness and promote preventive measures through public health campaigns aim to reduce the burden of kidney disease worldwide. Despite progress, challenges like limited access to care and organ shortage persist, underscoring the need for a comprehensive approach involving healthcare providers, policymakers, and communities. A multi-disease detection system can aid early identification and treatment initiation, serving as a crucial tool in addressing various illnesses.

(d) Liver Disease

Liver disease encompasses various conditions affecting the liver, from mild inflammation to severe damage and cirrhosis, posing a global health burden linked to factors like alcohol consumption, viral hepatitis, and obesity. Recent advancements in liver disease research highlight breakthroughs in understanding molecular mechanisms and offer hope with targeted therapies, including antiviral drugs and pharmacological agents. Lifestyle modifications and public health campaigns play crucial roles in prevention, alongside innovations in transplantation and regenerative medicine for end-stage liver disease. Despite progress, challenges like limited healthcare access and rising NAFLD prevalence persist, necessitating a comprehensive approach involving stakeholders to promote liver health and reduce disease burden.

2. Literature Survey

The study focused on Parkinson's disease detection using reduced facial expressions and hand-drawn images, highlighting their correlation with early PD symptoms. Utilizing ensemble techniques with two base models, the study achieved high metrics: accuracy (90.18%), precision (88.21%), recall (92.29%), and F1-measure (90.29%). The proposed method offers a straightforward approach for integrating PD detection into daily routines.[1]

The heart disease prediction project accomplished its objectives by creating a user-friendly web app for accurate heart disease prediction, validated through extensive evaluation. It addresses critical healthcare challenges, including accurate diagnosis in resource limited settings and providing a cost-effective alternative to expensive cardiac tests, thus proving to be a valuable asset in healthcare.[2]

The multiple disease prediction model allows users to input health records from prescribed tests, bridging the gap between testing and medical consultation. Enhancing efficiency, it predicts three diseases from a single input set, potentially reducing consultations and lowering mortality rates. Advancing existing systems, it offers advanced disease prediction, flexibility, and focuses on specific diseases.[3]

Emphasizing the heart's significance, accurate heart disease (HD) prediction is crucial for averting life-threatening outcomes. Introducing an enhanced HDNN system, incorporating diverse ML methods and data imputation techniques, outperforming conventional approaches. The findings demonstrate superior predictive performance, achieving high metrics across various datasets, including accuracy, precision, sensitivity, MCC, specificity, f measure, and AUC.[4]

Recognizing the significance of early illness detection and the drawbacks of traditional methods, an Automated Disease Prediction System using a Random Forest-based algorithm was developed to enhance accuracy. Through extensive experimentation with symptom datasets, the model showcased substantial improvement, achieving a 95% accuracy rate by correctly classifying 39 out of 41 test records, outperforming existing systems.[5]

The study focuses on legal infectious diseases in Shandong Province, noting their yearly trend's impact on model accuracy. Recommendations include using seasonal ARIMA for diseases with seasonal peaks, the grey model for grassroots personnel, and the BP neural network for accurate epidemic trend prediction by scientific institutions. It concludes that selecting the right model based on characteristics and demand is crucial for optimal prediction outcomes for various epidemics.[6]

Experimentation with the CNN-UDRP algorithm combines structured data for disease risk prediction, particularly in heart disease. Naïve Bayes achieves an 82% accuracy, surpassing KNN, while structured data yields a 65% accuracy, emphasizing efficient risk assessment. The authors aim to expand this approach to encompass more diseases in future studies.[7]

An innovative health prediction system utilizes diverse ML algorithms for accurate multi-disease forecasting, including diabetes, heart disease, and Parkinson's. Future goals involve expanding disease coverage, refining accuracy, and enhancing user accessibility for proactive healthcare decisions, aiming to reduce mortality rates. The project prioritizes precise disease forecasting and user-friendly features, like a chatbot, to positively impact public health.[8]

The study proposes a new method for assessing MS severity and predicting disease progression using machine learning on a diverse dataset. The Random Forest classifier achieved 94.87% accuracy in categorizing MS patients based on EDSS scores and 83.33% accuracy in forecasting disease advancement, without relying on MRI scans. This approach offers accurate MS severity assessment and prognosis over a 6-month period, beneficial for clinics with limited MRI access and aiding in early diagnosis for patients facing extended radiology waiting times.[9]

The study investigates machine learning's capability in detecting multiple diseases in healthcare, utilizing various algorithms trained on diverse datasets for high accuracy. AdaBoost demonstrated exceptional performance in precision, accuracy, recall, and F1 score across all dataset classes. Challenges acknowledged include limited disease coverage and suboptimal performance of some algorithms.[10]

The article addresses the limitation of single-disease-focused healthcare analysis and proposes a Flask API-based system for predicting multiple diseases, including diabetes, retinopathy, heart disease, and breast cancer, with scalability to encompass other conditions. By leveraging machine learning algorithms, TensorFlow, and Python pickling, the system aims to enable early detection and monitoring, ultimately improving patient care and reducing mortality rates.[11]

Soft computing techniques improve chronic disease prediction, enhancing healthcare system reliability and streamlining diagnostics for optimal treatment decisions. An ANFIS-based model achieves 94% accuracy in early-stage CKD prediction, integrating fuzzy logic and neural estimation.[12]

Network techniques for robust mathematical Incorporating data mining techniques into patient medical records has improved healthcare standards and decision support systems. A novel hybrid model, combining Neural Networks and Genetic Algorithms, optimizes ANN connection weights for enhanced heart disease prediction accuracy, achieving a training accuracy of 96.2% and a validation accuracy of 89% across 50 patient datasets.[13]

The survey highlights the significant factors affecting heart disease determination and examines multiple research endeavors in predicting the condition. It notes that not all researchers incorporate every attribute, with some opting to exclude certain ones to improve accuracy. A thorough discussion explores the unresolved key challenges present in various research initiatives aimed at predicting heart disease. [14]

3. Methodology

The methodology involves gathering and processing medical data, selecting pertinent features, training ML algorithms, and evaluating model performance. Following validation, integrated models enable practitioners to input data for prompt disease predictions, supporting early detection and proactive management. This method ensures efficacy and integrity through meticulous data processing, rigorous evaluation, and intuitive system integration.

3.1. Technology Used

Leveraging a variety of technologies and techniques, an organized and effective strategy was used in the development of the MDPS project. Google-Collab, a cloud-based Python coding environment, was used primarily to implement the project. The adaptable nature of Python was utilized, and key libraries including Matplotlib, Seaborn, Pandas, and NumPy were used for procedures like data cleansing and exploratory data analysis (EDA).

To enhance the predictive capabilities of the Multiple Disease Prediction System (MDPS), diverse datasets were obtained from Kaggle.com. The focus during development was on improving predictive accuracy through rigorous data preprocessing and exploratory data analysis (EDA) using essential tools like Matplotlib, Seaborn, Pandas, and NumPy. These analytical steps enabled the identification of patterns and trends within patient health records, forming the basis for further machine learning model development and predictive analytics in the MDPS.

3.2. Layout Design

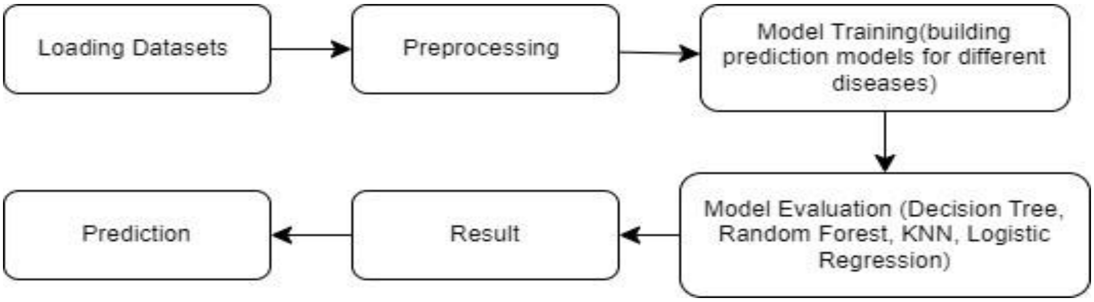

Figure 1. Multi-Disease Prediction workflow diagram

The workflow flow of our machine learning-based multiple disease prediction system is illustrated in the schematic presented in Figure 1 Step one is loading the dataset, which collects the necessary data for further analysis. Next, the preprocessing phase takes precedence, covering activities like handling missing data and utilizing Exploratory Data Analysis (EDA) to learn more about the features of the dataset. Afterwards, the process of training models entails creating several models for various diseases using distinct algorithms on selected datasets. The next step in the workflow is model evaluation, which involves evaluating the accuracy of the various algorithms used in the various models for the MDPS. It also involves comparing the accuracy, precision, and F1-score of each algorithm across the models. Following the patient's data entry into the MDPSystem, the workflow diagram's Prediction step determines if the patient has the specific ailment or not.

Once the multiple datasets are meticulously pre-processed, they undergo a pivotal transformation, where they are partitioned into an 80% training set and a 20% test set, setting the stage for model training endeavors. Four meticulously selected algorithms - namely Decision Tree, Random Forest, Logistic Regression and K- Nearest Neighbors (KNN) - are then harnessed for the purpose of model training, each offering unique methodologies for pattern recognition and predictive analysis.

After the model is rigorously trained, it is subjected to a thorough evaluation system wherein performance measures including accuracy, precision, recall, and F1 score are carefully examined. These metrics provide an informed comparison to determine which algorithm is the best fit for predictive tasks by acting as standards for evaluating each algorithm's robustness and effectiveness.

4. Implementation

The Multiple Disease Prediction System (MDPS) involves integrating trained machine learning models into a unified software architecture, with programming algorithms and designing user interfaces for seamless interaction. Robust data pipelines are established for efficient data flow and preprocessing, accompanied by rigorous testing and validation to ensure system accuracy and reliability. Considerations for scalability, security, and regulatory compliance are meticulously addressed, ensuring suitability for healthcare deployment. Once implemented, MDPS provides valuable insights and support to healthcare professionals, enhancing disease management and patient care strategies.

Step 1: Loading Datasets

The project initiates by importing essential libraries for data manipulation, visualization, and analysis, as depicted in the workflow block diagram. Libraries like NumPy, Pandas, Seaborn, and Matplotlib facilitate robust data management and exploration for multiple diseases, including Diabetes, Heart Disease, Kidney Disease, and Liver Disease datasets. Loading diverse datasets into different models of the prediction system enables conversion of numerical data into categorical formats to achieve balance and normalization. The conversion process is executed separately for diabetes, heart disease, liver, and kidney disease datasets using the [dataset.head()] function.

The initial phase of project implementation involves utilizing the Pandas library for data processing and filtering. This includes loading the dataset from a CSV file, separating input features and target variables, and performing necessary preprocessing tasks such as handling missing values and encoding categorical variables. Data preprocessing aims to clean, organize, and transform raw data into a suitable format for analysis or machine learning, enhancing data quality and preparing it for further analysis or modelling. The specific preprocessing steps depend on the nature of the data and the goals of the analysis or modelling task. Proper data preprocessing helps improve the accuracy and reliability of the results obtained from the analysis or machine learning models, making the data more suitable for meaningful insights and decision-making.

Step 2: Applying ML algorithms

(a) Logistic Regression

Logistic regression is a pivotal tool in machine learning for binary classification tasks like disease prediction within the Multiple Disease Prediction System. It models the relationship between input features and the probability of disease occurrence, optimizing coefficients through parameter estimation for accurate predictions. Its interpretability and computational efficiency make it well-suited for healthcare applications, aiding clinicians in understanding disease factors from large-scale datasets.

Step 1: Initialize an empty logistic regression model to serve as the predictive tool for outcome prediction based on input data preparation.

Step 2: Train the model using the training dataset to enable pattern recognition, utilizing features (X_train) such as age and gender to predict the target variable (y_train), representing disease presence.

Step 3: Utilize the trained model to make predictions on new, unseen data (X_test), assessing its performance in predicting outcomes it hasn't encountered previously.

Step 4: Evaluate the model's accuracy by comparing predictions made on the training data (y_train) with the actual outcomes (y_train), providing insights into its learning efficacy from the training data.

(b) KNN

The K-Nearest Neighbors (KNN) algorithm, suitable for the Multiple Disease Prediction System, operates on similarity principles to classify diseases based on feature similarities. It accommodates diverse datasets, including numerical and categorical features, and handles imbalanced data through techniques like weighted voting. However, KNN's performance relies heavily on distance metric selection and appropriate k value, with computational complexity increasing with dataset size.

Step 1: Initialize an empty K-Nearest Neighbors (KNN) classifier, a supervised learning algorithm designed for classification tasks.

Step 2: Train the KNN model using the training data (`X_train` and `y_train`), enabling it to learn patterns for prediction based on similar patterns in new data.

Step 3: Utilize the trained model to predict outcomes for the test data (`X_test`) using the `predict()` method, generating predicted values based on provided features.

Step 4: Evaluate the accuracy of the model by comparing predicted outcomes (`y_pred`) with actual outcomes from the test data (`y_test`) using the `accuracy_score()` function, determining the proportion of correctly predicted outcomes.

Step 5: Print the accuracy score to assess the KNN model's performance on the test data, providing insights into its effectiveness in predicting unseen data.

(c) SVC

Support Vector Machine (SVM) is a powerful classification algorithm, pivotal in predicting diseases like diabetes, heart, liver, and kidney ailments in the "Multiple Disease Prediction System using ML" project. SVM segregates data into distinct classes using a hyperplane, leveraging various features to categorize individuals accurately. Through rigorous training on labelled datasets, SVM learns to discern patterns and make precise predictions on unseen data, excelling in high-dimensional medical datasets. Its interpretability empowers healthcare professionals to make informed decisions, enhancing patient outcomes in disease prediction and intervention.

Step 1. Model Training: The `svc.fit(X_train, y_train)` method trains a Support Vector Classifier (SVC) model using the training data.

- `fit()`: Trains the SVC model.
- `X_train`: Contains the features of the training dataset.
- `y_train`: Contains the corresponding target labels for the training dataset.

Step 2. Making Predictions: `svc.predict(X_test)` generates predictions for the target variable based on the features in the test dataset(Utilizes the trained SVC model to make predictions on new data).

Step 3. Calculating Accuracy: The `accuracy_score()` function compares actual target values (`y_test`) with predicted values from the SVC model(Calculates the proportion of correct predictions over all observations in the test dataset).

Step 4. Printing Accuracy: The calculated accuracy score, representing the proportion of correct predictions, is printed to the console using the `print()` function(Provides insight into the model's performance on the test data).

(d) Decision Tree

The Decision Tree classifier is integral to the Multiple Disease Prediction System, utilizing machine learning to forecast diseases like diabetes, heart disease, liver disorders, and kidney ailments. By iteratively partitioning feature space into homogenous subsets, Decision Trees offer insights into predictive factors and facilitate informed decision-making in patient care. Ensemble methods like Random Forest or Gradient Boosting enhance prediction accuracy and mitigate risks, while the interpretability of Decision Trees aids in understanding disease mechanisms. Their versatility extends to feature selection, outlier detection, and risk stratification, contributing to advanced diagnostic capabilities and personalized medicine initiatives.

Step 1. Model Creation: Initialize an empty Decision Tree classifier, a type of algorithm that learns patterns and makes decisions based on them.

Step 2. Model Training: Train the Decision Tree classifier using the training data (`X_train` and `y_train`). This involves exposing the model to examples with known outcomes to learn from.

Step 3. Prediction Generation: Utilize the trained model to predict outcomes for the test data (`X_test`). This process involves applying the learned patterns to new data to make educated guesses.

Step 4. Accuracy Evaluation: Compare the model's predictions (`y_pred`) with the actual outcomes from the test data (`y_test`) to assess prediction accuracy.

Step 5. Accuracy Printing: Display the accuracy score to quantify the model's performance on the test data, indicating the proportion of correct predictions made.

(e) RFC

In our Multiple Disease Prediction System, the Random Forest classifier is pivotal for predicting diabetes, heart disease, liver disease, and kidney disease, harnessing ensemble learning to mitigate overfitting and enhance generalization. Its robustness to noisy data, feature importance insights, and meticulous preprocessing ensure accurate predictions and empower proactive disease management. By integrating Random Forest, our system offers personalized risk scores for early disease detection and enables healthcare providers to prioritize patient care effectively, revolutionizing medical diagnostics and interventions.

Step 1. Instantiate the Random Forest Classifier: Create an instance of the RandomForestClassifier class, specifying parameters such as criterion, max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators.

Step 2. Set Parameters: Define the parameters for the Random Forest Classifier:

- criterion='entropy': Measure impurity using entropy for node splitting.
- max_depth=15: Set the maximum depth of each decision tree to 15.
- max_features=0.75: Consider 75% of features when searching for the best split.
- min_samples_leaf=2: Specify the minimum number of samples required at a leaf node as
- min_samples_split=3: Set the minimum number of samples required to split an internal node to 3.
- n_estimators=130: Determine the number of decision trees in the forest as 130.

Step 3. Fit the Model: Train the Random Forest classifier using the fit() method with training data (X_train and y_train). X_train contains features, and y_train contains corresponding labels. The model learns to predict labels based on features in the training data.

5. Model Evaluation

The F1 Score, a harmonious blend of recall and precision, offers a holistic assessment of model performance, especially in imbalanced datasets. Precision gauges the accuracy of positive predictions, recall highlights the model's ability to identify positive instances, while accuracy provides an overall view of model efficacy, albeit sensitive to class imbalances. Researchers tailor evaluations to align with project objectives and dataset nuances, as each metric serves a distinct role in classification model assessment.

The precision, recall, accuracy, and F1-score of the algorithms employed for the various disease models are listed in the tables below.

Table 1. Performance for diabetes detection Model

Diabetes disease	Precision	F1-score	Recall	Accuracy
LR	0.94	0.92	0.90	89.47%
KNN	0.92	0.91	0.90	88.16%
SVM	0.94	0.93	0.92	90.79%
DT	0.91	0.90	0.88	82.24%
RFC	0.93	0.92	0.92	90.13%

Table 2. Performance for kidney disease detection Model

Kidney disease	Precision	F1-score	Recall	Accuracy
LR	0.92	0.92	0.92	90%
KNN	0.86	0.78	0.71	73.75%
SVM	0.83	0.83	0.83	77.50%
DT	1.00	0.98	0.96	97.50%
RFC	0.98	0.99	1.00	98.75%

Table 3. Performance for heart disease detection model

Heart disease	Precision	F1-score	Recall	Accuracy
LR	0.84	0.82	0.87	80.51%
KNN	0.81	0.78	0.85	75.82%
SVM	0.52	0.68	1.00	51.62%
DT	0.74	0.75	0.77	73.62%
RFC	0.89	0.85	0.91	83.53%

Table 4. Performance for liver disease detection model

Liver disease	Precision	F1-score	Recall	Accuracy
LR	0.78	0.85	0.94	76.99%
KNN	0.79	0.46	0.50	66.37%
SVM	0.72	0.84	1.00	71.68%
DT	0.83	0.75	0.68	73.45%
RFC	0.48	0.80	0.84	69.91%

6. Result and Discussion

The machine learning experiment yielded promising results in forecasting multiple diseases, utilizing diverse techniques like Support Vector Machine, Random Forest Classifier, Decision Tree Classifier, and Logistic Regression. Each algorithm showcased its effectiveness, collectively enhancing the comprehensive analysis of illness risk factors and prediction outcomes. The collaborative efforts of these algorithms bolstered the system's capability to detect various diseases simultaneously, underscoring the transformative potential of early disease identification in healthcare.

1. Diabetes Machine Learning Techniques

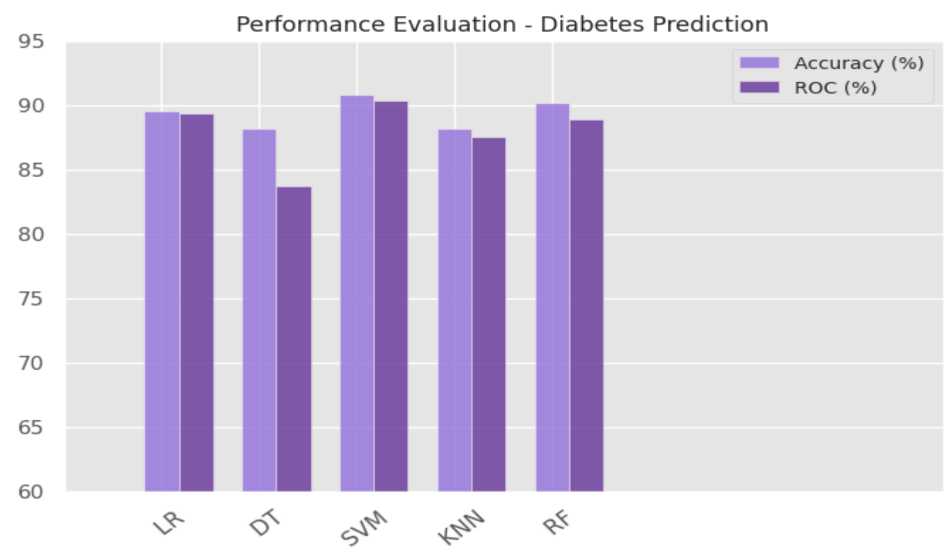

Figure 2. Bar Graph of Algorithm vs Accuracy

Fig. 2 Insightful results were obtained from models trained with Random Forest, Decision Tree, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN) algorithms. SVM emerged as the most accurate model, showcasing its efficacy in predicting diabetes mellitus from input parameters. These findings highlight SVM's potential for improving the accuracy of multiple disease prediction systems, suggesting avenues for further exploration into machine learning algorithms.

2. Kidney Disease Machine Learning Techniques

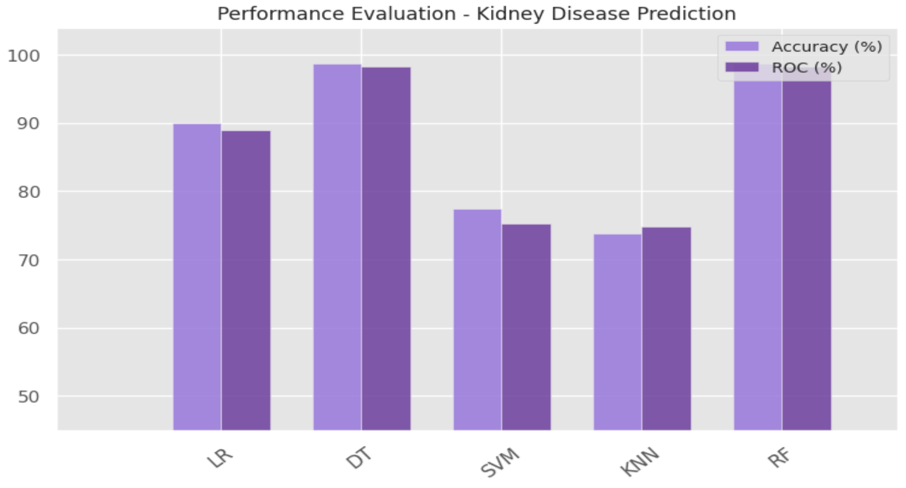

Figure 3. Bar Graph of Algorithm vs Accuracy

Fig. 3 Models trained with Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, and K-Nearest Neighbors (KNN) algorithms provided insightful outcomes. Among them, Random Forest stood out as the most accurate, showcasing its ability to predict diabetes mellitus based on input parameters. Its superior performance suggests Random Forest's effectiveness in capturing intricate data patterns, thereby enhancing multiple disease prediction accuracy.

3. Heart Disease Machine Learning Technique.

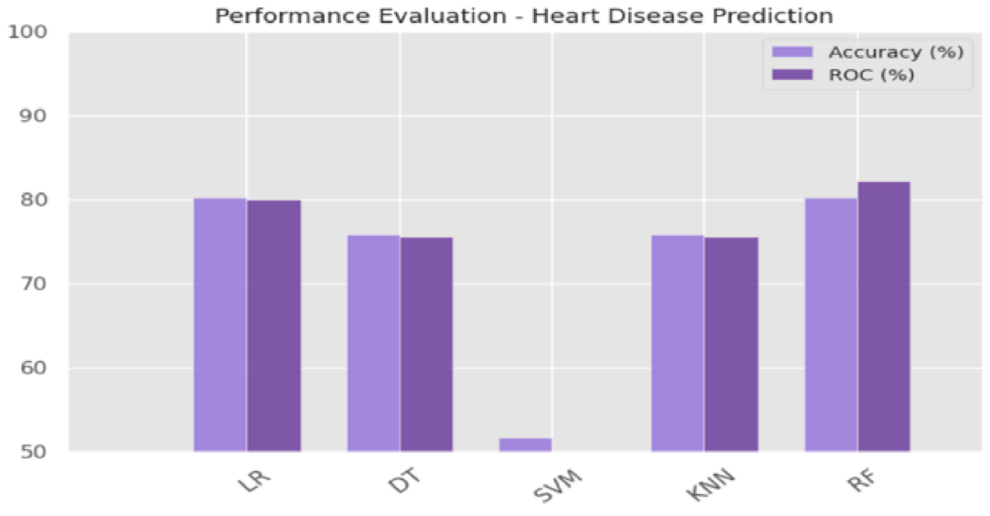

Figure 4. Bar Graph of Algorithm vs Accuracy

Fig. 4 The trained models, employing sophisticated algorithms like Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, and K-Nearest Neighbors (KNN), produced insightful outcomes. Among these, Random Forest stood out as the most accurate, demonstrating its remarkable predictive capability in identifying diabetes mellitus based on diverse input parameters. These findings highlight Random Forest's potential to enhance disease prediction accuracy, urging further exploration for improved diagnostic and prognostic capabilities.

4. Liver Disease Machine Learning Technique.

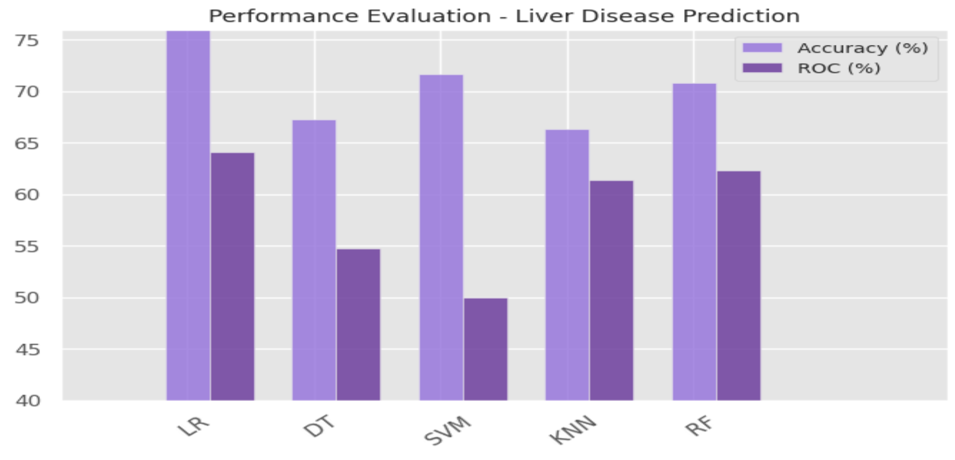

Figure 5. Bar Graph of Algorithm vs Accuracy

Fig. 5 Various algorithms including Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, and K-Nearest Neighbors (KNN) were employed to train the models. Logistic Regression emerged as the most accurate method, with Random Forest displaying commendable performance, underscoring the significance of algorithm selection in disease prediction accuracy. Overall, these results highlight the potential of machine learning algorithms to enhance diagnostic and prognostic capabilities in medical settings.

7. Conclusion and Future Scope

The insights from this project offer actionable guidance for healthcare stakeholders, advocating for evidence-based approaches in disease detection and management. Integrating machine learning into healthcare systems presents an opportunity to enhance diagnostic accuracy and treatment effectiveness. Continued collaboration between healthcare and technology sectors is vital for maximizing the potential of intelligent disease detection initiatives. This project marks a milestone toward data-driven healthcare, ensuring improved patient outcomes and resilient healthcare systems. The MDPS project heralds an era of precision health management, with implications for healthcare stakeholders and public health advancement. In the future, expanding the system's scope, enhancing data analytics, and fostering collaboration will further its impact on public health, early diagnosis, and treatment efficacy. Incorporating wearable technology and real-time data collection will enable continuous monitoring and early disease identification. Collaboration with legislators and healthcare providers is crucial for system integration and widespread acceptance. Overall, ongoing innovation and adaptation will drive improved healthcare outcomes globally.

References

[1]. S. Karunarathne, S. Vasanthapriyan and K. Chathumini, "Utilizing Ensemble Learning in Detecting Parkinson's Disease with Reduced Facial Expressions and Hand-Written Drawings," *7th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), Kuliyapitiya, Sri Lanka*, pp. 1-6, 2023. doi: 10.1109/SLAAI-ICAI59257.2023.10365024.

[2]. N. Sabri et al., "HeartInspect: Heart Disease Prediction of an Individual Using Naïve Bayes Algorithm," *IEEE 11th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia*, pp. 350-354, 2023. doi: 10.1109/ICSPC59664.2023.10420149.

[3]. R. Shanthakumari, C. Nalini, S. Vinothkumar, E. M. Roopadevi and B. Govindaraj, "Multi Disease Prediction System using Random Forest Algorithm in Healthcare System," *International Mobile and Embedded Technology Conference (MECON), Noida, India*, pp. 242-247, 2022. doi: 10.1109/MECON53876.2022.9752432.

[4]. M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," *in IEEE Access*, vol. 11, pp. 121574-121591, 2023, doi: 10.1109/ACCESS.2023.3328909.

[5]. A. N. V. K. Swarupa, V. H. Sree, S. Nookambika, Y. K. S. Kishore and U. R. Teja, "Disease Prediction: Smart Disease Prediction System using Random Forest Algorithm," *IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), Visakhapatnam, India*, pp. 48-51, 2021. doi: 10.1109/ICISSGT52025.2021.00021.

[6]. Y. Zhao and Y. Su, "Comparison of Three Prediction Models for the Incidence of Epidemic Diseases," *International Conference on Communications, Information System and Computer Engineering (CISCE), Kuala Lumpur, Malaysia*, pp. 131-136, 2020. doi: 10.1109/CISCE50729.2020.00033.[6]

[7]. S. Ambekar and R. Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network," *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India*, pp. 1-5, 2018. doi: 10.1109/ICCUBEA.2018.8697423.

[8]. N. R. Rusyana, F. Renaldi and D. Destiani, "Prediction Analysis Of Four Disease Risk Using Decision Tree C4.5," *International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), Jakarta, Indonesia*, pp. 90-94, 2023. doi: 10.1109/ICCoSITE57641.2023.10127710.[8]

[9]. M. K, S. S and T. E, "Streamlit-Powered Comprehensive Health Analysis and Disease Prediction System," *International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India*, pp. 1-7, 2023. doi: 10.1109/ICERCS57948.2023.10434221.

[10]. P. Bhardwaj, Y. Kumar and S. Mishra, "Machine Learning-Based Approaches for the Prognosis and Prediction of Multiple Diseases," *International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bangalore, India*, pp. 1-6, 2024. doi: 10.1109/IITCEE59897.2024.10467881.[10]

[11]. A. Yaganteeswarudu, "Multi Disease Prediction Model by using Machine Learning and Flask API," *5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India*, pp. 1242-1246, 2020. doi: 10.1109/ICCES48766.2020.9137896.

[12]. K. Damodara and A. Thakur, "Adaptive Neuro Fuzzy Inference System based Prediction of Chronic Kidney Disease," *7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India,* pp. 973-976, 2021. doi: 10.1109/ICACCS51430.2021.9441989.

[13]. S. U. Amin, K. Agarwal and R. Beg, "Genetic neural network-based data mining in prediction of heart disease using risk factors," *IEEE Conference on Information & Communication Technologies, Thuckalay, India*, pp. 1227-1231, 2013. doi: 10.1109/CICT.2013.6558288.

[14]. G. Shanmugasundaram, V. M. Selvam, R. Saravanan and S. Balaji, "An Investigation of Heart Disease Prediction Techniques," *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)*, Pondicherry, India, pp. 1-6, 2018. doi: 10.1109/ICSCAN.2018.8541165.