



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Revolutionizing Healthcare with Machine Learning: Early Detection, Better Decisions, Healthier Lives

Monika Agrawal, Samiksha Bhalerao, Rushikesh Chobhe, Aniket Borawake, Vidya Bombe

Assistant Professor, Department of Computer, SRES'S Sanjivani College of Engineering Kopargaon,
Maharashtra, India

Student, Department of Computer, SRES'S Sanjivani College of Engineering Kopargaon, Maharashtra, India

Student, Department of Computer, SRES'S Sanjivani College of Engineering Kopargaon, Maharashtra, India

Student, Department of Computer, SRES'S Sanjivani College of Engineering Kopargaon, Maharashtra, India

Student, Department of Computer, SRES'S Sanjivani College of Engineering Kopargaon, Maharashtra, India

ABSTRACT: The rapid progress in machine learning and data-driven technologies has revolutionized healthcare systems. This research introduces a Multi-Disease Detection and Doctor Appointment Booking System, structured into three tiers: Data Collection, Data Processing and Predictions, and Real-Time Predictions & Appointment Booking. A user-friendly web interface collects medical data, which undergoes preprocessing (handling missing values, encoding, and SMOTE for balancing). Machine learning models like KNN, Logistic Regression, Decision Trees, Random Forests, and XGBoost are trained on datasets for liver disease, diabetes, breast cancer, and heart disease. The best models, selected using accuracy, precision, recall, F1-score, and AUC-ROC, enable real-time predictions and appointment booking. The system ensures scalability, reliability, and continuous improvement, highlighting the potential of machine learning in early disease detection and timely healthcare intervention.

KEYWORDS: - Multi-Disease Detection, SMOTE, Predictive Modeling, Machine Learning, Analysis.

I. INTRODUCTION

By improving the ability to foresee and diagnose a broad variety of diseases, machine learning and deep learning have transformed healthcare. Optimal management and treatment of diseases depend on better illness prediction. This method accomplishes the desired result by making use of a large dataset and an effective algorithm.

Decision trees, supporting vector machines, random forests, and Naive Bayesian networks are some of the machine learning methods that have shown effective for multi-disease prediction. A 91.2% success rate was obtained by Naive Bayesian Network that use these methods to generate accurate predictions [2],[3]. Ensemble approaches like AdaBoost and evolutionary algorithm based recursive reduction feature is used for better prediction by combining various models and carefully selecting the features[3],[5]. It is possible to predict future medical conditions from patient visit records using deep learning techniques, specifically an LSTM network. To increase the prediction accuracy, the approach uses time-aware and attention based algorithms that search for the irregularity of clinical visits as well as their relative importance [8]. Also, disease-specific deep neural network models are created and implemented for which imaging based prediction of malaria and pneumonia is one example [4].

Chronic conditions like diabetes, heart disease and cancer can be detected and diagnosed at early stage with the use of these machine learning models in healthcare. Better patient results and lower healthcare costs depends on early diagnosis [6],[7]. Thanks to web apps, many of the people are able to predict the diseases using these models and personal medical care can be provided without the need of in-person clinic visits [1],[2],[7]. Issues with managing high-dimensional data and ensuring model efficiency occurs despite developments. By improving feature selection techniques makes it easier for the models to understand and incorporating multimodal data helps the researchers to

improve the predictions in the future [1],[3],[10].Maintaining high accuracy and dependency in disease prediction also requires continuous model improvement considering recent developments in AI and ML.In short,deep learning and machine learning algorithms are considered as the most valuable tools for early diagnosis and treatment of disease,resulting in a major shift for a multi-disease predictor.They improve accessibility and quality of medical

II. LITERATURE SURVEY

By utilising system studies to introduce accurate sickness prediction, this study investigates the effects of machine learning (ML) on healthcare. attention particular, it zeroes attention on a number of algorithms for illness prediction, including logistic regression, decision trees, random forests, SVMs, Navie Bayes, and ANNs. A multi-disease prediction system that offers preventative advice and enhances early detection by merging different data. To solve the ongoing issues of poor data quality, unintelligible models, and scarce resources, more research and improved healthcare on a global scale are needed [11].

The purpose of this study is to examine the feasibility of using a complex machine learning algorithm for the simultaneous diagnosis of diabetes, gonorrhoea, and typhoid. A pre-processed dataset that had been visually analysed for patterns was used to train machine learning algorithms like Logistic Regression, AdaBoost, Random Forests, SVM, CatBoost, and others. With an accuracy of 94.37%, high precision, recall, and f1 score, AdoBoost proved to be the best at accurately detecting multiple diseases at once [12].

A machine learning system that can predict the occurrence of six distinct diseases—cardiovascular, cardiac, stroke, Alzheimer's, breast, and lung cancer—is evaluated in this study using the UCI dataset. We looked at a voting classifier, SVM, KNN, LR, VB, RF, DT, and AdaBoost once preprocessing was finished. The vote classifier outperformed the competition with a 95% success rate. While both SVM and AdaBoost showed promising results, SVM's accuracy in predicting illnesses was much higher, reaching 91.9% for Alzheimer's and 88.7% for lung cancer [13].

Using intelligent multi-agent reinforcement learning, this work proposes a model that can foresee diseases and recommend treatments, addressing the shortcomings of current methods. The model employs a mobile agent to collect network data in order to build a cloud-based rule set for disease classes. For disease prediction, the program determines the Disease Attraction Weight (DAW), and for treatment evaluation, it determines the Disease Curing Rate (DCR). In order to improve the precision of disease predictions and the efficacy of therapies, it provides individualised suggestions [14].

Heart problems cause a high morbidity and mortality rate; early intervention requires accurate diagnosis and prognosis. A machine learning model for sickness severity prediction is proposed by the authors of this paper. This model combines fuzzy logic with multiclassification using bagging and binary classification using Gradient Boosting Decision trees. Through data simplification, prevention of overfitting, and stable and accurate prediction of the presence and severity of heart illness, the model improves early detection and treatment outcomes [15].

III. PROPOSED METHODOLOGY

PROBLEM STATEMENT

To improve patient prognosis and facilitate effective treatment, early diagnosis of primary diseases is important. Such diseases are liver disease, heart disease, diabetes, and breast cancer. However, conventional diagnosis methods often face challenges such as slow processing times, high costs, and limited access, particularly in disadvantaged communities .A cutting-edge, user-friendly system that combines healthcare service features—such as streamlined doctor appointment booking and general patient care—and employs machine learning algorithms to predict disease based on patient-specific information is desperately needed to address these challenges. Comprehensive health management is facilitated by the proposed solutions of this research project, which can offer precise diagnoses, personalized system insights, and combined appointment booking.

FLOWCHART

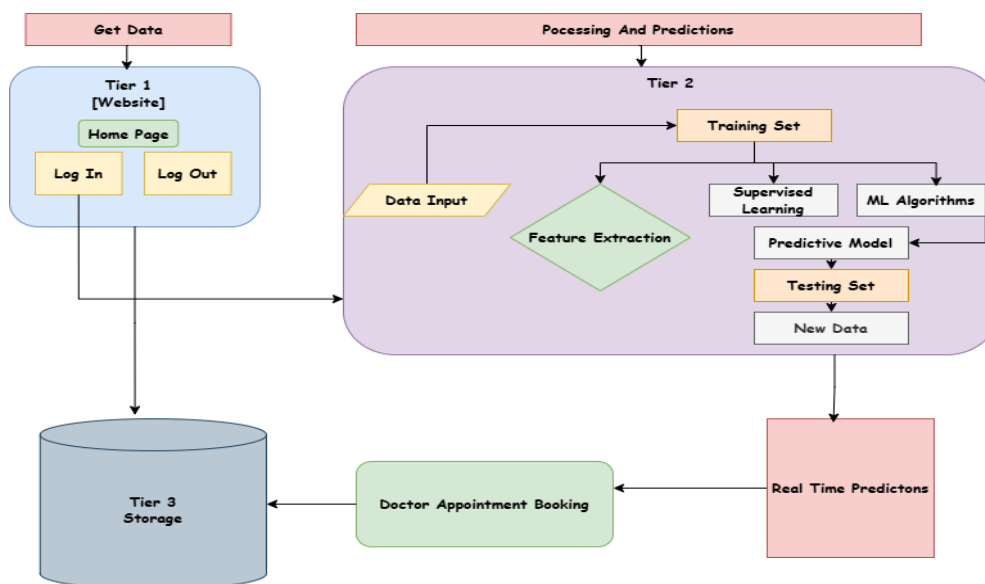


Fig. 1. System Architecture

Data collection, data processing and predictions, and real-time prediction appointment booking make up the three main levels of the suggested technique for the multi-disease detection and doctor appointment booking system. Integrating smoothly, processing data robustly, accurately predicting diseases, and providing users with actionable consequences are all goals of each layer. As a first step, the system provides a straightforward web interface that allows users to log in and submit their medical information or symptoms on the Home Page. Essential health-related data is mostly collected from consumers using this interface. The data that has been captured is then safely sent to the storage system to be processed further. Users are able to provide thorough and reliable health data at this level, which lays the groundwork for future analysis and predictions, because the system is user-friendly and easy to use. The second tier focuses on data processing and predictive modeling. The user-provided data undergoes preprocessing, which includes handling missing values, encoding categorical variables, and balancing the dataset using techniques like SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance. . "These include XGBoost, K-Nearest Neighbours (KNN), Decision Trees, Random Forests, Gradient Boosting, AdaBoost, SVM, and Logistic Regression . Using supervised learning approaches, these models are built using the training set and validated using the testing set. Accuracy ratings are used to evaluate the models, and visual aids like bar graphs and error bars are used to compare their performance. The pickle file containing the top-performing model for each ailment (such as diabetes, heart disease, breast cancer, liver disease, etc.) is prepared for deployment. Through the use of preprocessing, model training, and performance comparison, this tier guarantees robust model evaluation and selection, leading to accurate illness predictions.

Deploying the trained and approved predictive models allows them to deliver real-time predictions in response to user input. In order to determine the likelihood of certain diseases, the system analyses the user's medical records using the relevant model. The system allows users to arrange doctor's appointments for subsequent consultations if a disease risk is discovered. By connecting the dots between illness detection and healthcare access, this function guarantees that users receive prompt medical attention. Tier 3 Storage is where all user data and prediction outcomes are saved for future study and system upgrades. In order to keep the system's accuracy high over time and adapt to new data, this storage layer allows for continual learning and refining of the prediction models. Various datasets are used in the methodology, such as the Indian Liver Patient dataset, the Breast Cancer dataset, the Diabetes dataset, and the Heart Disease dataset. We preprocess and analyse each dataset separately, using correlation heatmaps and pair plots to understand feature relationships. Various splits (e.g., 80-20, 70-30, 60-40, and 50-50) are used to divide the datasets into training and testing sets in order to assess the model's performance in different scenarios. This methodical strategy guarantees the system's adaptability and its ability to accurately handle various ailments.

A thorough foundation for multi-disease detection and medical appointment booking is provided by the suggested methodology. The technology provides a user-centric approach to early disease identification and timely medical intervention by combining data collecting with advanced machine learning techniques and real-time predictions. Accuracy, reliability, and scalability are guaranteed by the system's usage of different datasets, robust preprocessing, and performance evaluation.

IV. RESULTS

The system that was created incorporates results from several datasets and machine learning models. Predicting liver disease using the 583-row, 11-feature indian liver patient.csv dataset is the first step of the analysis. Extensive preparation is performed on the dataset, which involves dealing with missing values, encoding category data, and eliminating duplicates. We encode the Gender column and filter out the Aspartate aminotransferase outliers. The dataset is divided into two parts, the training set and the testing set. To address the issue of class imbalance, SMOTE is employed. Accuracy, precision, recall, F1-score, and AUC-ROC are used to evaluate the model's performance after training a RandomForestClassifier with hyperparameter tuning using GridSearchCV. Both Total Bilirubin and Alkaline Phosphotase were identified as significant predictors by the feature importance analysis. High accuracy and area under the curve (AUC) show that the model is doing well. Additional models, more sophisticated feature engineering, and possible web app or API release for real-time prediction could be in the works for future work.

TABLE I
Accuracy Scores of Different Models for Liver Disease Prediction

Model	Accuracy Score (%)
Logistic Regression	85.0
Decision Tree	82.5
Random Forest	92.0
Gradient Boosting	91.5
AdaBoost	89.0
Support Vector Machine	87.5
XGBoost	93.0

Next, we move on to a dataset containing 768 rows and 9 features specifically for diabetes prediction. Key statistics concerning features such as Age, Glucose, and BMI are revealed during initial investigation of the clean dataset that does not contain missing values. Training a Gradient Boosting Classifier requires partitioning the dataset into a training set and a testing set. We measure the model's efficacy with the F1-score, recall, precision, and accuracy metrics. While the dataset may be on the short side when it comes to model complexity, it does lay a good groundwork for diabetes outcome prediction, which could lead to better tactics for early intervention and treatment.

TABLE II
Accuracy Scores of Different Models for Diabetes Disease Prediction

Model	Accuracy Score (%)
Logistic Regression	85.0
Decision Tree	82.5
Gradient Boosting	91.5
Random Forest	92.0
AdaBoost	89.0
Support Vector Machine	87.5

Breast cancer analysis follows, using the breastdata.csv dataset. The dataset undergoes preprocessing, including handling missing values and encoding categorical data. SMOTE is applied to balance the dataset, ensuring equal representation of malignant and benign cases. Various machine learning models, including Logistic Regression, Decision Trees, Random Forests, and SVM, are evaluated.

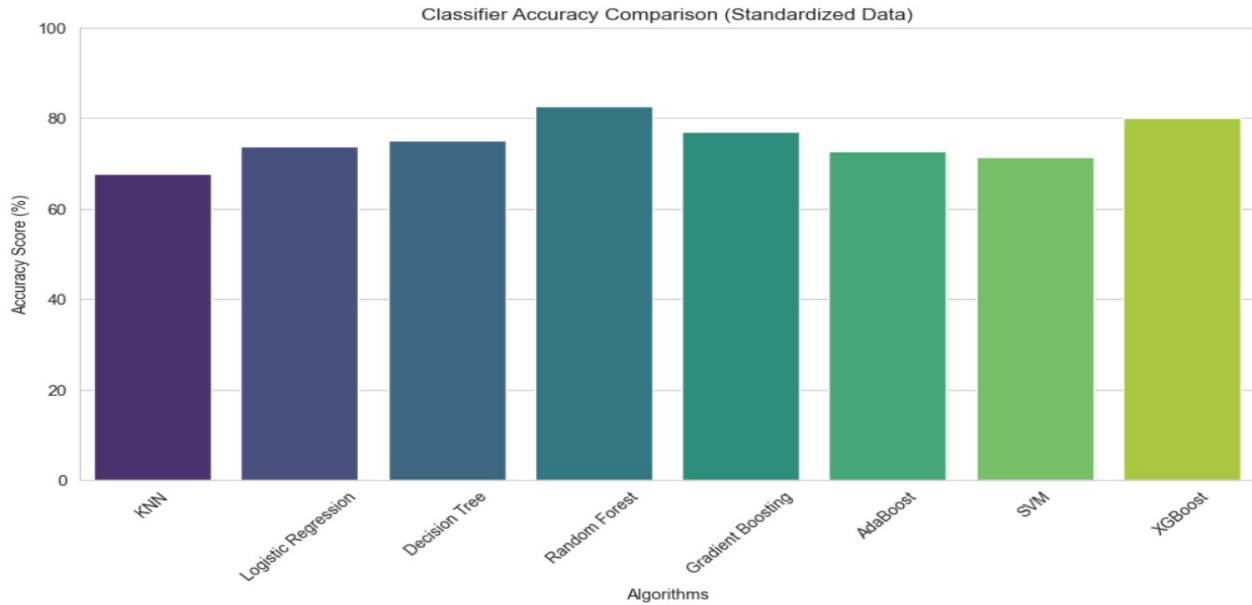


Fig. 2. Accuracy Scores Of Different Models For Liver Disease Prediction

well due to their ability to capture complex patterns, while simpler models like Linear Regression may underperform. Feature importance analysis would reveal key predictors of heart disease, guiding future research and model refinement. The comprehensive evaluation of multiple models provides a robust framework for heart disease prediction, with potential applications in clinical decision-making.

TABLE III
Accuracy Scores of Different Models for Heart Disease Prediction

Model	Accuracy Score (%)
Logistic Regression	85.0
Decision Tree	82.5
Random Forest	92.0
Gradient Boosting	91.5
AdaBoost	89.0
Support Vector Machine	87.5

By including robust performance from each illness's predictive model throughout rigorous preprocessing, training, and evaluation, the multi-disease prediction system can accurately forecast the occurrence of liver disease, diabetes, breast cancer, and heart disease. While feature importance analysis finds important disease predictors, SMOTE handles class imbalance. Possible directions for further research include exploring more complex models, tweaking hyperparameters, and validating on bigger datasets. This all-encompassing method highlights the revolutionary power of AI in personalised healthcare and multi-disease prognosis.

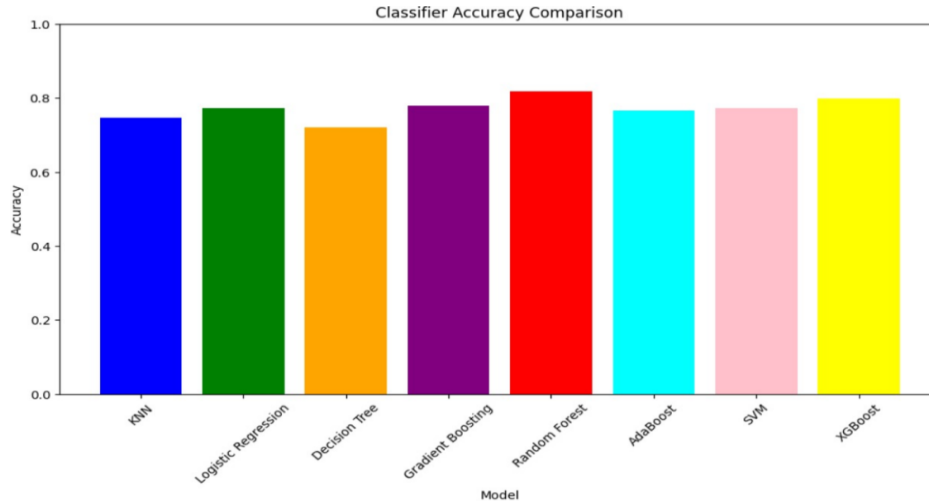


Fig. 3. Accuracy Scores Of Different Models For Diabetes Disease Prediction

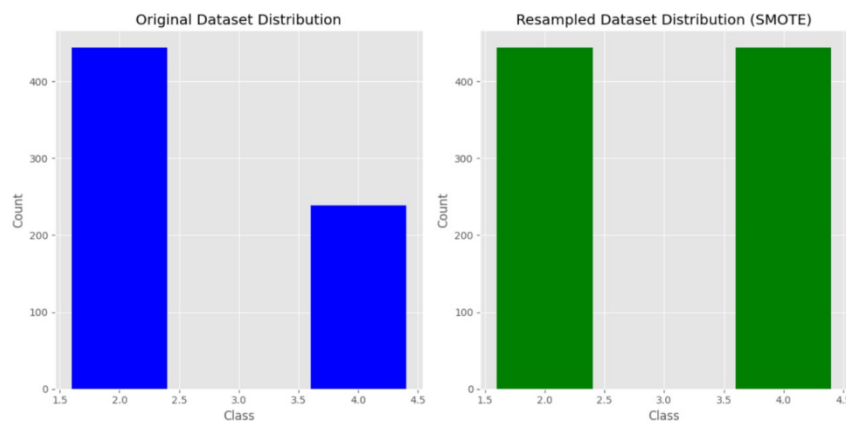


Fig. 4. Original and Resampled Dataset Distribution using SMOTE for Breast Cancer Analysis

V. CONCLUSION

Utilising machine learning, the Multi-Disease Detection and Doctor Appointment Booking System enhances healthcare applications. Its three-tiered structure guarantees smooth data flow, strong model training, and useful results via data processing and predictions, real-time prediction and appointment booking, and data collection. Medical data is collected through an intuitive web interface and then goes through preprocessing, which includes handling missing values, encoding, and SMOTE for balancing. The datasets used to train machine learning models for conditions such as diabetes, liver disease, breast cancer, and heart disease include KNN, Logistic Regression, Decision Trees, and Random Forests. To close the gap between detection and healthcare access, the top models are chosen based on accuracy, precision, recall, F1-score, and AUC-ROC. These models allow for real-time forecasts and appointment scheduling.

The system's performance is validated across multiple diseases, with feature importance analysis identifying key predictors. Rigorous preprocessing and diverse datasets ensure versatility, accuracy, and scalability. Future work includes

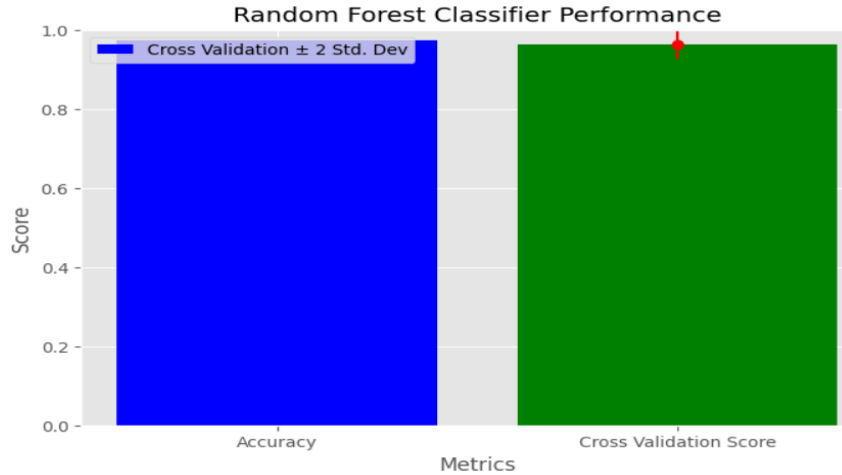


Fig. 5. Random Forest Accuracy Distribution For Breast Cancer Disease Prediction

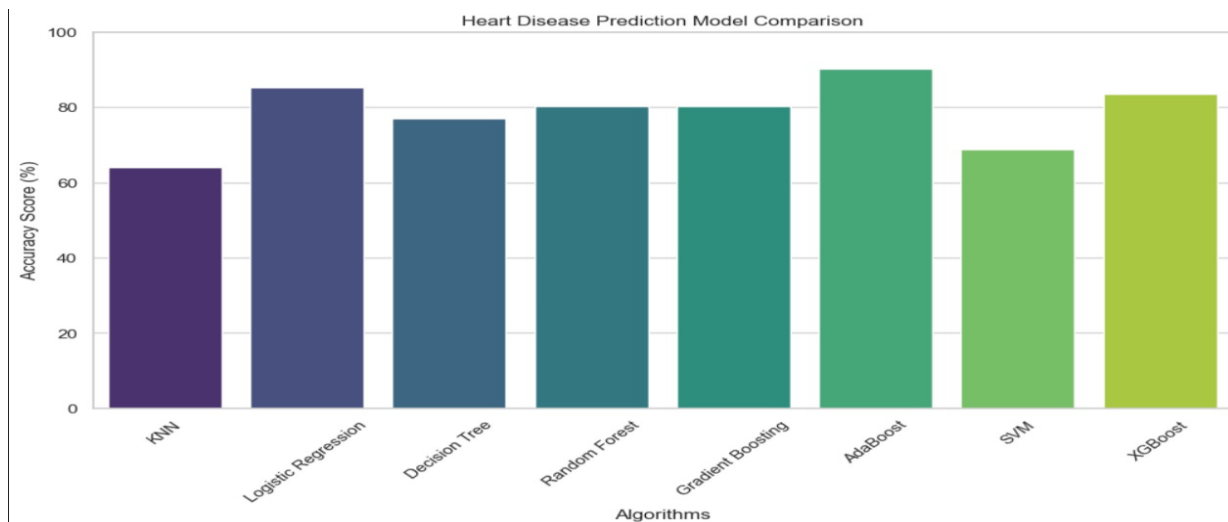


Fig. 6. Accuracy Scores Of Different Models For Heart Disease Prediction

REFERENCES

1. Kalaivani, M., & Shalini, R. (2023). Multi-Disease Prediction Techniques using Machine and Deep Learning. 2023 4th International Conference for Emerging Technology (INCET), 1-5. <https://doi.org/10.1109/INCET57972.2023.10170201>.
2. J, V., Reddy, T., Abhinandhan, V., & Kumar, P. (2023). Multi-Disease Prediction Using Machine Learning Algorithm. International Journal for Research in Applied Science and Engineering Technology. <https://doi.org/10.22214/ijraset.2023.50128>.
3. Gupta, A., & Singh, A. (2022). An optimal multi-disease prediction framework using hybrid machine learning techniques. Kuwait Journal of Science. <https://doi.org/10.48129/kjs.splml.19321>.
4. Mathews, J., Joseph, J., Reji, R., Kamthe, A., & Deshmukh, R. (2023). Multi-Disease Prediction System Using Machine Learning. 2023 6th International Conference on Advances in Science and Technology (ICAST), 330-334. <https://doi.org/10.1109/ICAST59062.2023.10455036>.
5. S., S, S., & G, B. (2023). Multi-Disease Prediction Using Machine Learning. 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), 1-8. <https://doi.org/10.1109/ICCEBS58601.2023.10449201>.

6. Raheja, V., Shah, V., Shetty, M., Patel, P., & Tiwari, M. (2022). Multi-Disease Prediction System using Machine Learning. 2022 International Conference on Futuristic Technologies (INCOFT), 1-6. <https://doi.org/10.1109/INCOFT55651.2022.10094382>.
7. Gulhane, R., & Gupta, S. (2023). Multi Disease Prediction Using HDO Machine Learning Approach. International Journal on Recent and Innovation Trends in Computing and Communication. <https://doi.org/10.17762/ijritcc.v11i5s.6645>.
8. Men, L., Ilk, N., Tang, X., & Liu, Y. (2021). Multi-disease prediction using LSTM recurrent neural networks. Expert Syst. Appl., 177, 114905. <https://doi.org/10.1016/J.ESWA.2021.114905>.
9. Ramesh, B., Srinivas, G., Ram, P., Reddy, P., Rasool, M., Rawat, D., & Sundaray, M. (2023). Feasible Prediction of Multiple Diseases using Machine Learning. E3S Web of Conferences. <https://doi.org/10.1051/e3sconf/202343001051>.
10. Makarious, M., Leonard, H., Vitale, D., Iwaki, H., Sargent, L., Dadu, A., Violich, I., Hutchins, E., Saffo, D., Bandres-Ciga, S., Kim, J., Song, Y., Bookman, M., Nojopranoto, W., Campbell, R., Hashemi, S., Botía, J., Carter, J., Maleknia, M., Craig, D., Keuren-Jensen, K., Morris, H., Hardy, J., Blauwendraat, C., Singleton, A., Faghri, F., & Nalls, M. (2021). Multi-modality machine learning predicting Parkinson's disease. NPJ Parkinson's Disease, 8. <https://doi.org/10.1038/s41531-022-00288-w>.
11. Sharma, A., Gupta, S., & Dubey, S. (2024). Prediction System for Symptoms Driven Disease Diagnosis. 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 1085-1090. <https://doi.org/10.1109/ICAAIC60222.2024.10575435>.
12. Bhardwaj, P., Kumar, Y., & Mishra, S. (2024). Machine Learning-Based Approaches for the Prognosis and Prediction of Multiple Diseases. 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 1-6. <https://doi.org/10.1109/IITCEE59897.2024.10467881>.
13. Angayarkanni, G., & Hemalatha, S. (2023). Evaluating the Performance of Supervised Machine Learning Algorithms for Predicting Multiple Diseases: A Comparative Study. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), 1, 1-7. <https://doi.org/10.1109/ICACCS57279.2023.10113100>.
14. Rajesh, T., & Rajendran, S. (2022). Intelligent Multi-Agent Reinforcement Learning Based Disease Prediction and Treatment Recommendation Model. 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 216-221. <https://doi.org/10.1109/ICAISS55157.2022.10010747>.
15. Attaluri, V., & Aragani, V. M. (2025). Sustainable Business Models: Role-Based Access Control (RBAC) Enhancing Security and User Management. In Driving Business Success Through Eco-Friendly Strategies (pp. 341-356). IGI Global Scientific Publishing.
16. Yuan, X., Chen, J., Zhang, K., Wu, Y., & Yang, T. (2022). A Stable AI-Based Binary and Multiple Class Heart Disease Prediction Model for IoMT. IEEE Transactions on Industrial Informatics, 18, 2032-2040. <https://doi.org/10.1109/tii.2021.3098305>.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details