Review on Cardiovascular Disease Prediction System Using Machine Learning Algorithm

Dr.Taware Ganesh Gorakhnath[1], Miss.Dhanshri Dnyandeo Shinde[2]
[1]*Associate Professor, Dattakala group of institutions Faculty of engineering, swami chincholi, Daund, Pune, Maharashtra, India*
[2]*Dattakala group of institutions Faculty of engineering, swami chincholi, Daund, Pune, Maharashtra, India*

Abstract- Cardiovascular diseases (CVD) are a significant global health concern, necessitating effective tools for early diagnosis and risk prediction. This study introduces a web-based Cardiovascular Disease Prediction System that applies machine learning to analyze clinical and demographic factors, including age, weight, blood pressure, cholesterol, and glucose levels. The system is built using a dataset of 70,000 patient records obtained from Kaggle, which underwent preprocessing and feature selection using ANOVA to enhance data quality and model efficiency. Several algorithms were implemented, including Random Forest, Naive Bayes, K-Nearest Neighbors, Decision Tree, and Support Vector Machine. Among these, Random Forest achieved the highest accuracy of 74%, validated through performance metrics like precision, recall, and F1-score. The system features an intuitive interface supporting secure login, user registration, data input, and automated PDF report generation. By leveraging machine learning, the system serves as a reliable tool for early CVD detection, contributing to preventive healthcare and timely interventions.

Keywords: Cardiovascular disease prediction, machine learning algorithms, feature selection (ANOVA), hyperparameter tuning, clinical decision support system, health risk assessment, Web Application.

I. INTRODUCTION

Cardiovascular illnesses are one of the major executioners all inclusive and cause various passings around the world. Convenient and exact discovery of conditions related to heart-related issues is fundamental in avoiding passing and improving the survival and quality of life for patients. Ordinary conclusion includes ECG evaluation, angiography, and treadmill testing, which have gotten to be common. All of these strategies require physical translation by the doctor included, taking much time and gambling human mistake and are incomprehensible to perform in inaccessible or underdeveloped areas. With the headway of manufactured insights and machine learning, especially profound learning, restorative diagnostics have experienced critical change in exactness, speed, and robotization. Profound learning models have appeared promising results in analyzing expansive and complex therapeutic datasets for infection forecast and classification errands. In any case, profound learning models regularly require cautious tuning of parameters and can be computationally intensive. To improve the execution and productivity of profound learning frameworks, nature inspired optimization calculations have picked up ubiquity. HBO is competent of finetuning the hyperparameters of profound learning models to realize way better meeting and performance. In this venture, we propose a novel approach for automated cardiovascular disease determination employing a adjusted profound learning demonstrate optimized with the Honey Badger Optimization calculation. This framework has the potential to assist healthcare suppliers by giving a quick, precise, and versatile decision-support device that can be utilized in real time clinical utilize. The show is tried on a standard cardiovascular infection dataset, and its execution is compared to existing strategies to point out its adequacy.

II. LITERATURE SURVEY

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative technologies in the field of agriculture. The integration of these technologies enables

intelligent decision-making, enhances productivity, and helps achieve sustainable farming practices. Numerous research works have focused on developing predictive systems that analyze soil and environmental data to recommend appropriate crops, fertilizers, and irrigation methods.

Several machine learning algorithms such as Decision Trees, Random Forest, Naïve Bayes, Support Vector Machines (SVM), and XGBoost have been explored in agricultural prediction systems. These models are trained on datasets containing parameters like Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, humidity, and rainfall to determine the best crops for cultivation in specific conditions.

Researchers have demonstrated that AI-driven crop recommendation systems not only improve crop selection accuracy but also contribute to resource optimization by reducing fertilizer wastage, minimizing water usage, and improving soil health. The following section summarizes major works related to this study.

Sr.No	Title / Research Focus	Author & Year	Key Findings
1	Machine Learning-Based Cardiovascular Disease Detection Using Optimal Feature Selection	TAHSEEN ULLAH (2024)	This study developed a framework for CVD detection using machine learning, with FCBF achieving 78% accuracy alongside Extra Tree and Random Forest models. It highlights FCBF's effectiveness in improving performance on large, categorical datasets.
2	An efficient iot-based patient monitoring and heart disease pre diction system using deep learning modified neural network	S. S. Sarmah (2020	The Deep Learning Modified Neural Network (DLMNN) achieved 96.8% accuracy for heart disease prediction using real-time IoT data.
3	A gradient boosted decision tree with binary spotted hyena optimizer for cardiovascular disease detection and classification	G. R. Reddy (2023)	GBDT-BSHO hybrid model achieved 97.89% accuracy on the UCI dataset, outperforming 12 ML classifiers including SVM, KNN, and Logistic Regression.
4	An integrated machine learning frame work for effective prediction of cardio vascular diseases	A. Rahim (2021	The MaLCaDD framework using Logistic Regression achieved 99.1% accuracy on the Framingham dataset..
5	Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV	G. N. Ahmad (2022)	The paper explores machine learning models like XGBoost, SVM, and GBC for heart disease prediction, using datasets from UCI Kaggle and other sources. With GridSearchCV and 5-fold cross-validation, XGBoost achieves the highest accuracy of 100% and 99.03%, showcasing the impact of hyperparameter tuning on diagnostic accuracy.
6	An automatic diagnostic model for the detection and classification of cardiovascular diseases based on swarm intelligence technique	C. Venkatesh (2024)	CNN with PSO achieved 98.58% accuracy for cardiovascular disease detection using ECG signal preprocessing and multimodal data fusion.

III. PROPOSED SYSTEM

The system architecture follows a modular design comprising data preprocessing, feature selection, model training, prediction, and user interaction through a user-friendly graphical interface. During preprocessing, the dataset is cleaned, normalized, and encoded to ensure uniformity and compatibility with machine learning algorithms. Feature selection techniques like ANOVA are employed to identify significant predictors, improving model efficiency and accuracy.

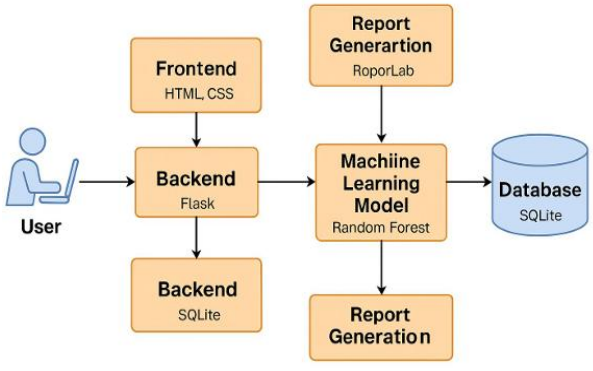

Fig: System Architecture

IV. METHODOLOGY

The proposed system is a Cardiovascular Disease Prediction application that leverages machine learning techniques to deliver accurate and timely risk assessments. It is designed to assist in the early diagnosis of cardiovascular conditions by analyzing clinical and behavioral parameters such as age, blood pressure, cholesterol levels, glucose levels, smoking habits, and physical activity.

This study aims to predict the likelihood of cardiovascular disease using an automated prediction system, offering decision support to medical practitioners and patients. To achieve this goal, a comprehensive machine learning pipeline was implemented, including data preprocessing, feature selection, model training, and performance evaluation. The cardiovascular disease prediction model was developed using a dataset of 70,000 patient records obtained from Kaggle kaggle˙dataset.

Category	Feature	Description / Values
Demographic Characteristics	Age	Age in years
	Gender	0: Male, 1: Female
	Height	Height in centimeters (cm)
	Weight	Weight in kilograms (kg)
Clinical Measurements	Systolic Blood Pressure	Measured in mmHg
	Diastolic Blood Pressure	Measured in mmHg
	Cholesterol Levels	1: Normal, 2: Elevated, 3: High
Lifestyle Factors	Smoking Status	0: Non-smoker, 1: Smoker
	Alcohol Consumption	0: Non-drinker, 1: Drinker
	Physical Activity	0: Inactive, 1: Active
Target Variable	Cardio	1: Diagnosed with cardiovascular disease 0: No cardiovascular disease

Data Preprocessing Prior to model training, the dataset underwent several preprocessing steps to ensure quality and consistency. The preprocessing pipeline included the following steps: • Irrelevant Feature Removal: The id column being a non-informative unique 37 Chapter 7.4.3 Feature Engineering identifier, was removed from the dataset as it provided no predictive value. • Age Conversion: The age attribute, originally recorded in days, was converted into years to enhance intepretability using the transformation:

$$Age\_year = Age\_days / 365.25$$

Feature Engineering: The Body Mass Index (BMI) was computed as an additional health indicator using the standard formula:

$$BMI = weight(kg) / height(m2)$$

V. RESULT

The Cardiovascular Disease Prediction System is a web-based application designed to predict cardiovascular disease using machine learning. The Random Forest algorithm, achieving 74% accuracy, emerged as the most effective model after comparing it with Logistic Regression, Naive Bayes, KNN, Decision Tree, and SVM. Preprocessing techniques such as normalization and encoding, combined with feature selection using ANOVA, identified significant predictors like age, weight, systolic and diastolic blood pressure, and cholesterol levels, optimizing the model's performance. The system includes a user friendly interface with a Home Page, Registration Page, Login Page, and a Prediction Form, ensuring accessibility for all users. Database integration provides secure storage for user credentials and prediction data, while a report generation feature allows users to download detailed PDF summaries of their prediction results. This project effectively demonstrates the application of machine learning in healthcare, offering a reliable and user-centric tool for early cardiovascular disease prediction.

VI. CONCLUSION

This project successfully developed a machine learning-based system for predicting cardiovascular disease using key health indicators such as age, blood pressure, cholesterol, glucose levels, and lifestyle

factors. The Random Forest algorithm achieved the highest accuracy of 74%, demonstrating its effectiveness and reliability compared to other models. Feature selection using ANOVA improved model efficiency by identifying the most significant attributes, reducing complexity without compromising performance. With its user-friendly interface, seamless data input, and automated PDF report generation, the system provides a practical solution for early detection and preventive healthcare, emphasizing the potential of machine learning in enhancing medical diagnostic.

REFERENCE

[1] T. Ullah, S. I. Ullah, K. Ullah, et al., "Machine learning-based cardiovascular disease detection using optimal feature selection," IEEE Access, vol. 12, pp. 1643116446, 2024. doi: 10.1109/ACCESS.2024. 3359910.

[2] M. S. Al Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani, and A. Shaikh, "A robust heart disease prediction system using hybrid deep neural networks," IEEE Access, vol. 11, pp. 121574–121591, 2023. doi: 10.1109/ACCESS.2023.3321234.

[3] M. Obayya, J. M. Alsamri, M. A. AlHagery, A. Mohammed, and M.A.Hamza, "Automated cardiovascular disease diagnosis using honey badger optimization with modified deep learning model," IEEE Access, vol. 11, pp. 64274–64281, 2023. doi: 10.1109/ACCESS.2023.3286661.

[4] T. Ullah, S. I. Ullah, K. Ullah, et al., "Machine learning-based cardiovascular disease detection using optimal feature selection," IEEE Access, vol. 12, pp. 1643116446, 2024. doi: 10.1109/ACCESS.2024. 3359910.

[5] M. Obayya, J. M. Alsamri, M. A. AlHagery, A. Mohammed, and M.A.Hamza, "Automated cardiovascular disease diagnosis using honey badger optimization with modified deep learning model," IEEE Access, vol. 11, pp. 64274–64281, 2023. doi: 10.1109/ACCESS.2023.3286661.

[6] S. Kiran, G. R. Reddy, S. P. Girija, S. Venkatramulu, K. Dorthi, and C. S. V. Rao, "A gradient boosted decision tree with binary spotted hyena optimizer for cardiovascular disease detection and classification," Healthcare Analytics, vol. 3, p. 100173, 2023. doi: 10.1016/j.health. 2023.100173.

[7] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim, and A. W.Muzaffar, "An integrated machine learning framework for effective prediction of cardiovascular diseases," IEEE Access, vol. 9, pp. 106575–106588, 2021. doi: 10.1109/ ACCESS.2021.3098688.

[8] T. Ullah, S. I. Ullah, K. Ullah, et al., "Machine learning-based cardiovascular disease detection using optimal feature selection," IEEE Access, vol. 12, pp. 1643116446, 2024. doi: 10.1109/ACCESS.2024. 3359910.

[9] C. Venkatesh, B. V. V. S. Prasad, M. Khan, J. C. Babu, and M. V. Dasu, "An automatic diagnostic model for the detection and classification of cardiovascular diseases based on swarm intelligence technique," Heliyon, vol. 10, no. 3, e25574, 2024. doi: 53

[10] I. D. Apostolopoulos, D. I. Apostolopoulos, T. I. Spyridonidis, N. D. Papathanasiou, and S. Panayiotakis, "Multi-input deep learning approach for cardiovascular disease diagnosis using myocardial perfusion imaging and clinical data," Physica Medica, vol. 84, pp. 168–177, 2021. doi: 10.1016/j.ejmp.2021.04.011.

[11] A. R. Vijayaraj and S. Pasupathi, "Nature inspired optimization in context-awarebased coronary artery disease prediction: A novel hybrid harris hawks approach," IEEE Access, vol. 12, pp. 92635–92651, 2024. doi: 10 . 1109 / ACCESS. 2024. 3414662.

[12] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 8154281554, 2019. doi: 10.1109/ACCESS.2019. 2923707.

[13] A. Mahajan, B. Kaushik, M. K. I. Rahmani, and A. S. Banga, "A hybrid feature selection and ensemble stacked learning models on multi-variant cvd datasets for effective classification," IEEE Access, vol. 12, pp. 87023–87038, 2024. doi: 10. 1109/ACCESS.2024.3412077.

[14] M.-H. Tayarani-Najaran, "A novel ensemble machine learning and an evolutionary algorithm in modeling the COVID19 epidemic and optimizing government policies," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 10, pp. 6362–6372, 2022. doi: 10. 1109/TSMC.2022.3143955.

[15] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O.

Chow, J. H. Chuah, and H. M. Gheni, "An effective heart disease detection and severity level classification model using machine learning and hyperparameter optimization methods," IEEE Access, vol. 10, pp. 79974– 79984, 2022. doi: 10. 1109/ACCESS.2022.3194736.

[16] S. S. Sarmah, "An efficient iot-based patient monitoring and heart disease prediction system using deep learning modified neural network," IEEE Access, vol. 8, pp. 135784– 135797, 2020. doi: 10.1109/ ACCESS.2020.3007561.

[17] D. Y. Omkari and K. Shaik, "An integrated two-layered voting (tlv) framework for coronary artery disease prediction using machine learning classifiers," IEEE Access, vol. 12, pp. 56275–56290, 2024. doi: 10.1109/ACCESS.2024.3389707.

[18] M. Z. Khan, R. Mangayarkarasi, C. Vannathi, and M. Angulakshmi, "Bio-inspired pso for improving neural based diabetes prediction system," Journal of ICT Standardization, vol. 10, no. 2, pp. 179–200, 2022. doi: 10.13052/jicts2245-800X. 1025. 54

[19] G. N. Ahmad, H. Fatima, Shafiullah, A. S. Saidi, and Imdadullah, "Efficient medical diagnosis of human heart diseases using machine learning techniques with and without gridsearchcv," IEEE Access, 1109/ACCESS.2022.3165792. vol. 10, pp. 80151–80173, 2022. doi: 10

[20] A. M.Qadri, A. Raza, K. Munir, and M. S. Almutairi, "Effective feature engineering technique for heart disease prediction with machine learning," IEEE Access, vol. 11, pp. 56214–56224, 2023. doi: 10.1109/ ACCESS.2023.3281484.

[21] G. Muhammad and et al., "Enhancing prognosis accuracy for ischemic cardiovascular disease using K nearest neighbor algorithm: A robust approach," IEEE Access, vol. 11, pp. 97879–97895, 2023. doi: 10.1109/ACCESS.2023.3312046.