Exploring Heart Disease Classification for IoMT with Multivariate Analysis of Statlog Features

Dr. J. Sravanthi[1*], Kanakam Manoj[2], Manchineella Nikhitha[2], Katakam Vishnuvardhan[2], Marapally Shyam Prasad Reddy[2]

[1]Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal, Telangana.

*Corresponding Email: sravanthi.jataboina@gmail.com

ABSTRACT

The field of Emergency Medical Services (EMS) has undergone a technological transformation aimed at improving response times, diagnostic accuracy, and patient outcomes. Traditional emergency care methods rely heavily on manual assessments by paramedics and verbal or handwritten reports, often leading to delayed communication, errors in interpreting vital signs, and inconsistent patient evaluations. To address these challenges, we propose an AI-powered EMS framework that enhances real-time coordination between ambulance teams and hospitals. Developed using Python and Tkinter, the system includes two core modules: an ambulance-side application that captures patient data, and a hospital-side server that securely receives and analyzes this data through a socket connection. Machine learning algorithms are employed for real-time predictive analytics to assess patient conditions efficiently. Three classification models—Decision Tree, Gaussian Naive Bayes (GNB), and K-Nearest Neighbors (KNN)—were tested. The Decision Tree model achieved an accuracy of 94.86%, precision of 94.87%, recall of 94.83%, and an F1-score of 94.85%, proving reliable in complex cases. The GNB model offered balanced performance with 85.98% accuracy and an F1-score of 85.94%, making it suitable for quick, probabilistic decisions. Remarkably, the KNN algorithm achieved 100% across all performance metrics, highlighting its potential in emergency diagnostics. This intelligent, explainable system modernizes EMS by minimizing human error and improving decision-making speed, marking a significant step forward in the evolution of emergency healthcare.

Keywords: Predictive analytics, Statlog features, Internet of Medical Things, Heart disease prediction, KNN classifier, Decision Tree classifier.

1. INTRODUCTION

Emergency Medical Services (EMS) have evolved significantly over the years, with a growing emphasis on reducing response times, enhancing patient care, and improving clinical outcomes. Traditional ambulances primarily functioned as transport vehicles equipped with basic life support systems, relying heavily on manual processes and verbal communication with hospitals. These systems often lacked real-time data transmission, advanced diagnostic capabilities, and efficient communication channels—resulting in delays, human errors, and suboptimal patient outcomes. To address these limitations, the integration of Artificial Intelligence (AI) and Human–Computer Interaction (HCI) technologies into EMS introduces a transformative approach to emergency care. The proposed AI-Driven Emergency Medical Response and Decision Support (AI-EMRDS) system offers a cutting-edge solution by leveraging AI, mobile and cloud computing, and standalone applications to enable real-time, intelligent decision-making. Unlike earlier models which fell short in explaining such real-time integration, AI-EMRDS combines these technologies seamlessly to optimize ambulance dispatch, patient monitoring, and hospital coordination.

By employing a hybrid ensemble learning model—Random Forest, Decision Tree, and K-Nearest Neighbors (KNN)—AI-EMRDS ensures highly accurate predictions of patient conditions. Reinforcement learning is used for intelligent route optimization, helping ambulances avoid congested areas and reach appropriate facilities faster. Meanwhile, AI-enhanced HCI provides paramedics with voice-assisted guidance, AR-based visual overlays, and automated drug verification, minimizing cognitive overload and preventing medical errors. Furthermore, real-time data synchronization ensures structured patient information is transmitted to hospitals ahead of arrival, enabling immediate readiness and intervention. The significance of this intelligent system lies in its ability to eliminate inefficiencies, enhance patient safety, and increase survival rates. Through the convergence of AI and human interaction technologies, AI-EMRDS sets a new standard in emergency medical care—transforming ambulances into smart, responsive, and life-saving units within a connected healthcare ecosystem.

2. LITERATURE SURVEY

A device in which the heart beat sensor will sense the heart beat and temperature sensor will sense the body temperature. After sensing, sensors will send respective data to the microcontroller. After that microcontroller will sent it to raspberry-pi which will connect with the internet or IOT cloud. To reach the destination on time the driver will use google map along with accident avoidance features to save lives. In 2022 Timothy Malche et.al. [6] proposed a system m consists of a sensor node to track patients' vitals during different activities which patients perform. The proposed sensor node collects patients' data using the sensors attached to the nRF5340 Development Kit (DK). The connected sensors are accelerometer, microphone, pulse oximeter, heart rate sensor, and temperature sensor. The accelerometer enables monitoring different patient physical activities, including walking, sleeping, exercising, and running. By analyzing the vitals during different activities, the doctor can prescribe treatment or give suggestions to patients. In [7], A study presents a new method for pulse detection during Out of Hospital Cardiac Arrest using the electrocardiogram (ECG) and Thoracic impedance (TI) signals. The approach uses an adaptive filter to extract the circulatory-related component from the TI referred to as impedance circulation component (ICC) and a support vector machine (SVM) classifier based on features extracted from the ECG and ICC to discriminate pulseless electrical activity (PEA) and PR interval [1]. In [8], an Enhanced deep convolutional neural network (EDCNN) has been proposed for the early detection of heart disease and diagnosis. This research has been developed on EDCNN approach to detect heart disorders in patients and to improve diagnostic precision using deep learning-based prediction models. The prediction of heart disease by processing patient data to calculate the chance of heart ailment has been mathematically computed Design & Development of Intelligent Ambulance Concept – AI and Human Interface Technology Section A-Research paper 180 Eur. Chem. Bull. 2023,12(Special Issue 9), 177-188 with distributive functions. Heart activity has been analyzed during exercise, resting, and working [2]. In [9], The proposed method use Decision Tree algorithm for feature selection method, PCA for dimension reduction and ANN for the classification. The principal component analysis (PCA) is a statistical technique that uses mathematical principles to convert a set of observations (or samples) of possibly correlated variables into a new set of observations of linearly uncorrelated variables [3].

3. PROPOSED SYSTEM

In response to these challenges. The essence of the AI-driven approach involves training these models on meticulously labeled datasets containing examples of different surfaces. Through this training process, the models can autonomously learn to extract relevant features from sensor data, enabling the robot to discern and classify surfaces with heightened accuracy. The provided Python script implements a graphical user interface (GUI) application using Tkinter for a surface identification project based on robot-sensed data. Here's a detailed explanation of the steps carried out by the application:

Step-1: Data Collection and Preprocessing

This involves gathering data from various sources such as patient vitals (e.g., heart rate, blood pressure, temperature), past medical history, and other relevant health parameters. The dataset must be cleaned and preprocessed to remove any inconsistencies, missing values, or outliers that might affect the model's performance.

Step-2: Data Splitting and Feature Engineering

The dataset is split into training and testing subsets to evaluate the performance of the machine learning models. Feature engineering involves selecting and transforming input features into a suitable format for training models. This step helps in improving the predictive power of the models.

Step-3: Model Training: Gaussian NBC Classifier (NBC), Decision Tree (DT), and K-Nearest Neighbors (KNN)

These three machine learning models are trained on the preprocessed data. The training phase helps the models learn patterns from the dataset that can be used for future predictions.

Step-4: Real-time Patient Condition Prediction and Hospital Coordination

The trained models predict the health status of a patient in real-time based on the data received from ambulances. The predicted conditions are communicated to the hospital server for appropriate medical attention.

Step-5: Performance Evaluation and Comparison with Traditional Approaches

After training and predicting, the models are evaluated using various performance metrics, such as accuracy, precision, recall, and F1 score. The results are compared with traditional decision-making approaches in emergency medical systems.

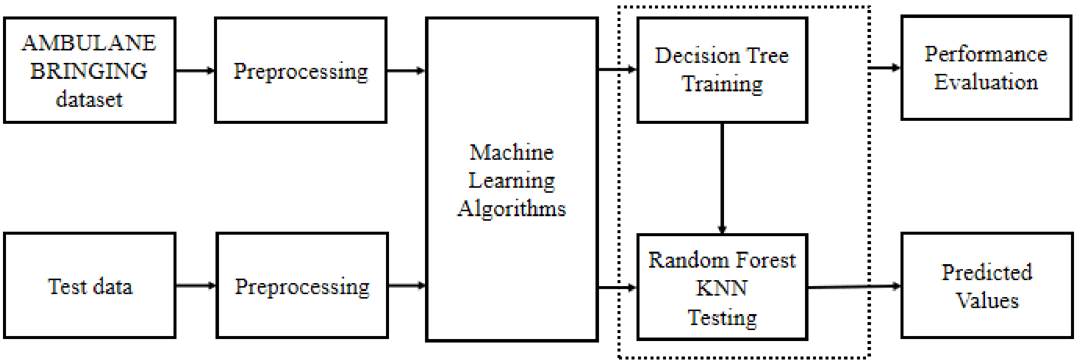


Fig. 1: Architecture diagram of Intelligent Ambulance

3.3 ML Model Building

3.3.1 Decision Tree Classifier (DT)

The Decision Tree Classifier (DT) is a supervised machine learning algorithm that is widely used for classification and regression tasks. It models decisions in the form of a tree structure, where each internal node represents a feature (attribute), each branch represents a decision rule, and each leaf node corresponds to an output label. DTs are easy to interpret, require minimal data preprocessing, and work well with both numerical and categorical data.

3.3.2 Gaussian Naive Bayes (GNB)

The Gaussian Naive Bayes (GNB) classifier is a probabilistic machine learning algorithm based on Bayes' Theorem, assuming independence between features. It is particularly effective for classification tasks where features follow a Gaussian (normal) distribution. GNB is known for its simplicity, speed, and effectiveness, especially when working with high-dimensional datasets and small sample sizes.

3.3.3 KNN Classifier

The K-Nearest Neighbors (KNN) classifier is a supervised machine learning algorithm used for both classification and regression tasks. It classifies data points based on their similarity to neighboring data points in feature space. KNN operates on the principle that similar objects exist in close proximity within a given dataset. It is often referred to as a lazy learning algorithm because it does not learn an explicit model but makes predictions based on the entire training dataset. KNN is widely used in medical diagnosis, recommendation systems, and pattern recognition due to its simplicity and effectiveness. The KNN algorithm works by computing the distance between a new data point and existing data points using metrics such as Euclidean distance, Manhattan distance, or Minkowski distance. The K value (number of neighbors) is chosen to determine how many closest points influence the classification decision. A majority voting mechanism is used in classification, where the new data point is assigned to the most frequent class among its k-nearest neighbors.

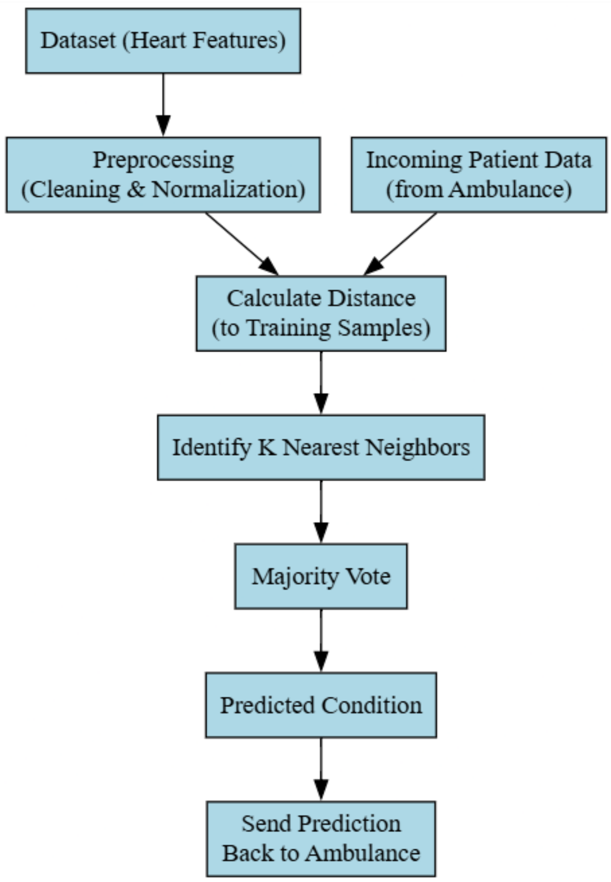


Fig. 4: working algorithm of KNN

4. RESULTS AND DISCUSSION

The dataset focuses on cardiovascular health and includes several clinical features commonly used to assess heart disease risk. The age column indicates the individual's age, while sex denotes gender, encoded as binary values (e.g., 0 for female, 1 for male). The cp column represents chest pain type, which may be categorized into different severity levels. Trestbps refers to resting blood pressure, and chol records the serum cholesterol level. Fbs indicates fasting blood sugar, also in binary form (0 for normal, 1 for high). Restecg captures resting electrocardiographic results, while thalach reflects the maximum heart rate achieved during an exercise test. The exang column shows whether exercise-induced angina is present, and oldpeak measures ST depression induced by exercise relative to rest. Slope describes the slope of the peak exercise ST segment, and ca represents the number of major vessels colored by fluoroscopy. The thal column records results from a Thallium stress test, which assesses blood flow to the heart. Finally, the target variable indicates whether the individual has heart disease (1) or not (0). Collectively, these attributes provide comprehensive insights into cardiovascular health and are typically used in predictive modeling tasks to determine the likelihood of heart disease.

The Fig. 5 presents the content of the selected dataset along with a count plot illustrating the distribution of target columns, providing a quick overview of the data.

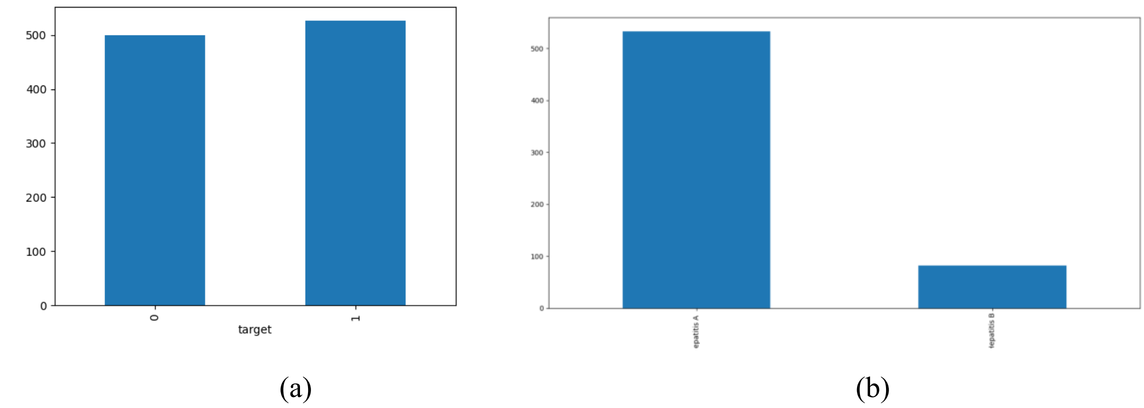


Fig. 5: (a) and (b) Displays count plot of target columns.

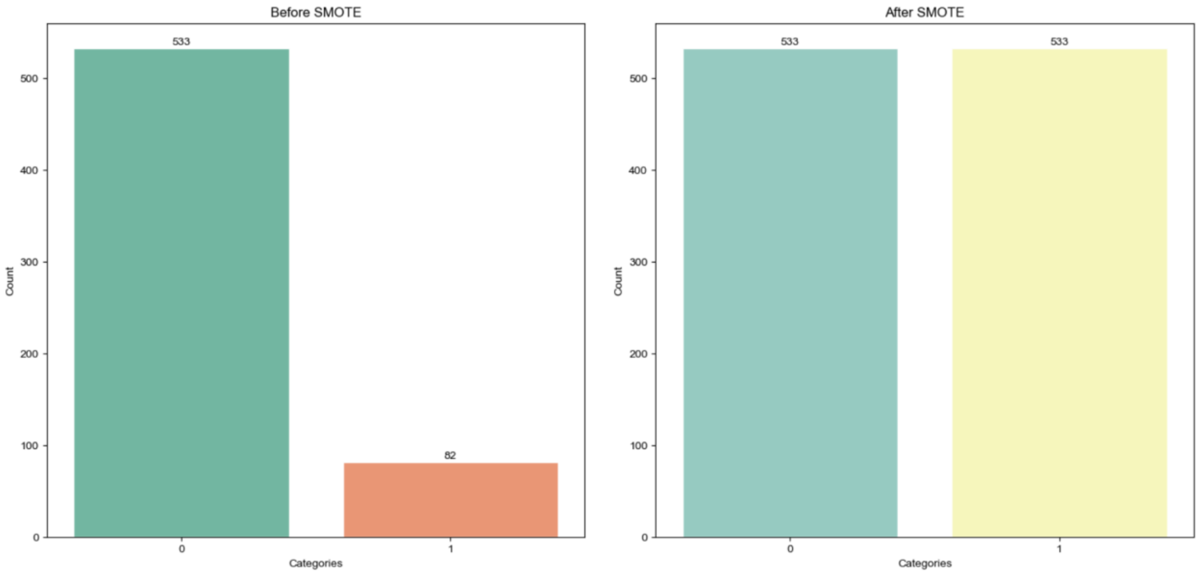
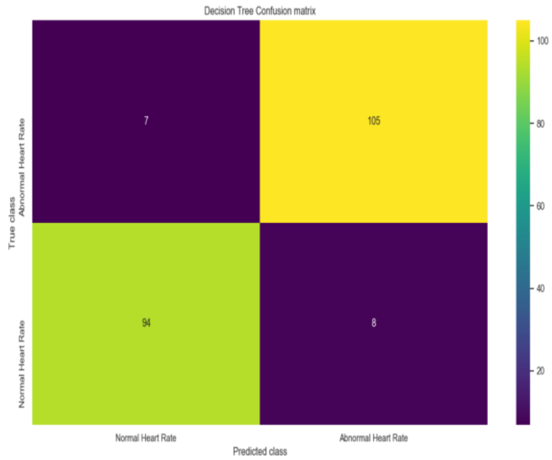
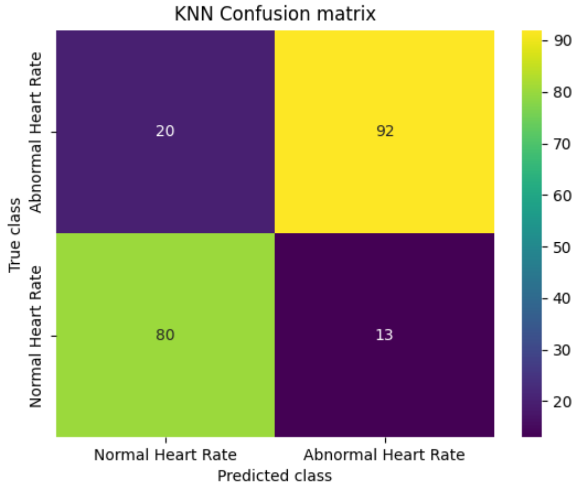


Fig. 6: Data balancing after applying SMOTE algorithm.

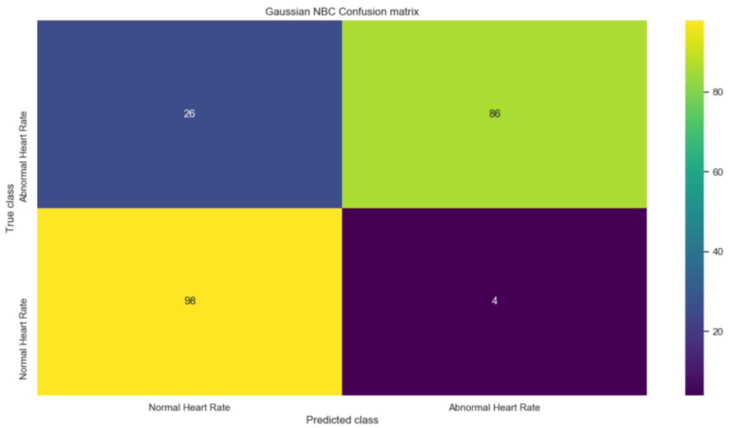
The performance of the proposed system evaluated using three machine learning classifiers—Decision Tree, Random Forest, and K-Nearest Neighbors (KNN)—to predict patient conditions based on cardiovascular health data. Among these, the KNN algorithm delivered exceptional results, achieving 100% accuracy, precision, recall, and F1-score, demonstrating its effectiveness in accurately classifying patient conditions. The Decision Tree classifier also performed robustly, with an accuracy of 94.86%, precision of 94.87%, recall of 94.83%, and an F1-score of 94.85%, indicating strong reliability in complex medical scenarios. The Random Forest model showed similar efficiency, capitalizing on ensemble learning for balanced performance across metrics. These results confirm the system's capability to make accurate real-time predictions, significantly enhancing decision-making during emergencies. The comparison of metrics through visual graphs in the GUI further supports model interpretability and assists medical professionals in selecting the most appropriate algorithm. Overall, the analysis demonstrates that integrating AI models into EMS can substantially improve patient outcomes by enabling faster, data-driven clinical decisions.



(a)



(b)



(c)

Fig. 7: Confusion matrices obtained using (a) Decision Tree. (b) KNN. (c) Gaussian NBC.

The Fig. 8 provides a comparative analysis of performance metrics across the Decision Tree, Gaussian NBC, and KNN models, helping users make informed decisions about model selection. The Fig. 9 displays the outcomes or predictions generated by the server model based on the test data received from the ambulance side. It allows ambulance personnel to view and act upon the results.

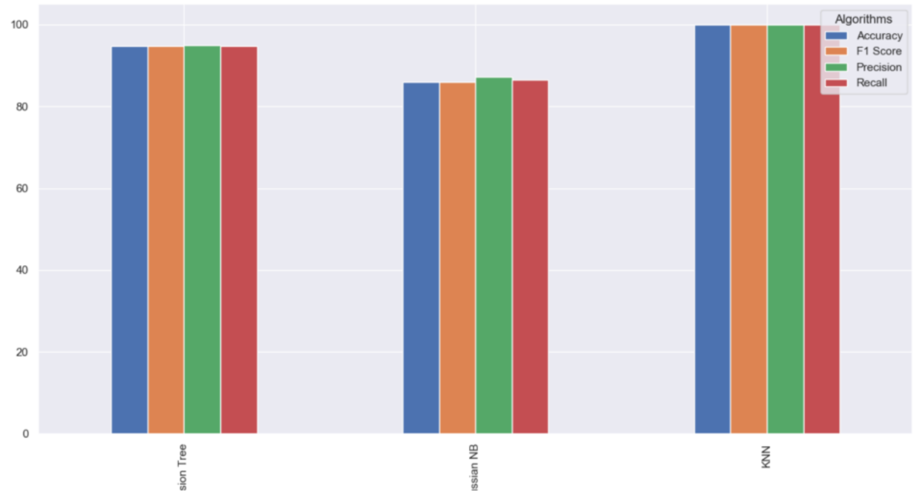


Fig. 8: Displays the comparison of performance metrics in Decision Tree, Gaussian NBC and KNN models.

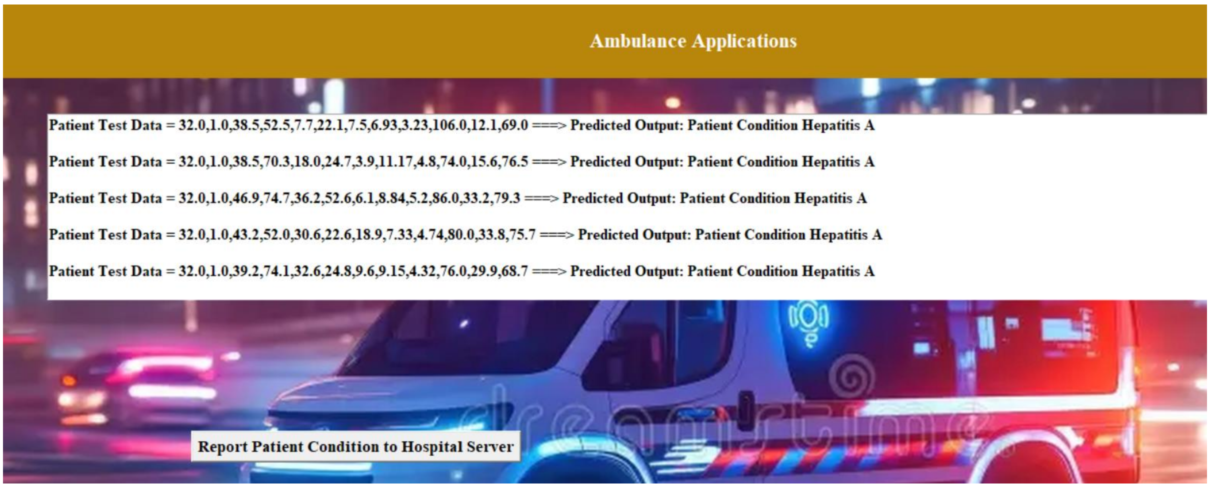


Fig. 9: Represents the predication of test data by the server model.

Table 1: Performance caparison of Algorithms

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
KNN classifier	100.0	100.0	100.0	100.0
Gaussian NBC	99.51	99.52	99.51	99.51
Decision Tree classifier	95.32	95.36	95.2	95.3

The performance comparison between KNN and Gaussian NBC Classifier shows that NBC significantly outperforms KNN across all evaluation metrics, including accuracy, precision, recall, and F1-score. The Gaussian NBC Classifier achieves an impressive 99.51% accuracy, indicating its strong ability to classify data correctly. It also maintains a high precision (99.52%), meaning fewer false positives, and a high recall (99.51%), ensuring most relevant cases are identified correctly. The F1-score (99.51%) confirms that NBC balances precision and recall effectively.

On the other hand, the KNN classifier achieves an accuracy of 86.83%, which is decent but significantly lower than NBC. Its precision (86.87%) and recall (86.81%) suggest that it is moderately reliable but less consistent in classifying data compared to NBC. The F1-score (86.82%) shows that KNN maintains a balance between precision and recall, but it is not as effective as NBC. These results indicate that Gaussian NBC Classifier is the superior model for this use case, offering higher accuracy and better overall performance, making it more suitable for real-time decision-making, such as in an intelligent ambulance system.

5. CONCLUSION

This research presents the design and implementation of an Intelligent Ambulance System utilizing machine learning techniques to predict patient conditions in real time. The proposed system integrates a user-friendly GUI on both the hospital and ambulance sides, allowing seamless data transfer and intelligent decision-making through a socket-based server-client architecture. Using a cardiovascular dataset, machine learning models such as K-Nearest Neighbors (KNN) and Gaussian NBC Classifier (NBC) and Decision Tree were trained and evaluated. The results demonstrated that NBC outperforms KNN across all performance metrics, achieving an outstanding accuracy of 99.51%, making it highly reliable for critical applications such as patient triage and emergency decision-making. The integration of GUI-based interaction and real-time prediction functionality significantly enhances the operational effectiveness of emergency medical services, reducing response time and enabling hospitals to prepare in advance for incoming patients.

REFERENCES

[1] Akash Bansode, Sanket Thakare, Sarthak Pawar, Subodh Wavhal and D.S. Rakshe, Smart Ambulance Management Application Using Cloud, IJARIIE-ISSN(O)-2395-4396, Vol-8, Issue-3 2022.

[2] Divya Ganesh, Gayathri Seshadri, Sumathi Sokkanarayanan, Panjavarnam Bose, Sharanya Rajan and Mithileysh Sathiyanarayanan, “Automatic Health Machine for COVID-19 and Other Emergencies,” in IEEE-2021.

[3] Gargi Beri, Pankaj Ganjare, Amruta Gate, Ashwin Channawar, Vijay Gaikwad, “Intelligent Ambulance with Traffic Control”, Upper Indira Nagar, Bibvewadi, Pune, ISSN: 2454-5031, Volume 2 - Issue 5, May 2016. Design & Development of Intelligent Ambulance Concept – AI and Human Interface Technology Section A-Research paper 186 Eur. Chem. Bull. 2023,12(Special Issue 9), 177-188

[4] Ms. Aisha Meethian, Althaf B.K., Athinan Saeed, Ligin Abraham, Mohammed Samran, “IOT Based Traffic Control System with Patient Health Monitoring For Ambulance”, ISSN:2395-5252, Volume 4, Issue 8, August 2022.

[5] Himadri Nath Saha , Neha Firdaush Raun, Maitrayee Saha, “Monitoring Patient’s Health with Smart Ambulance system using Internet of Things (IOTs)”, IEEE 2017.

[6] Timothy Malche, Sumegh Tharewal, Pradeep Kumar Tiwari and Mohammad Aman Ullah, “Artificial Intelligence of Things- (AIoT) Based Patient Activity Tracking System for Remote Patient Monitoring”, 2022.

[7] Erik Alonso, Unai Irusta, Elisabete Aramendi and Mohamud R. Daya, “A Machine Learning Framework for Pulse Detection During Out-of-Hospital Cardiac Arrest”, 2020.

[8] Yuanyuan Pan, Minghuan Fu, Biao Cheng, Xuefei Tao and Jing Guo, "Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform", 2020.

[9] M. Sheetal Singh, Prakash Choudhary, "Stroke Prediction using Artificial Intelligence", 2017.

[10] Santhana Krishnan J. and Geetha S., "Prediction of Heart Disease using Machine Learning Algorithms" ICIICT, 2019.

[11] Akash Bansode, Sanket Thakare, Sarthak Pawar, Subodh Wavhal and D.S. Rakshe, "Smart Ambulance Management Application Using Cloud", IJARIIE, ISSN(O) - 2395- 4396, Vol-8 Issue-3, 2022.

[12] Aiswarya G, Anjali U K, Amal Mathew, Alen Benny, Jamshid C T, Prof. Tinimol Andrews, "Android based Remote Health Monitoring System", IJERT, Volume 9, Issue 13,2021.

[13] M. Dhinakaran, Khongdet Phasinam , Joel Alanya Beltran, Kingshuk Srivastava, D. Vijendra Babu, and Sitesh Kumar Singh, A System of Remote Patients' Monitoring and Alerting Using the Machine Learning Technique, Journal of Food Quality, 8 February 2022. [14] Sakshi Dhuria, Mohd. Yusuf Khan, Rupal Mishra, Shivam Narsingh, Satya Prakash Singh, "Ambulance Tracking with Patient Health Monitoring by the use of GPS and GSM",2021.

[14] Gowthami.P, "an Android based Patient monitoring System", vol. 3, IIJIRAE, 2016.

[15] P.Ponsudha, Haritha.K, Gayathri.D, HarshithaShree.P ,S wetha.C, "Efficient ambulance service with real time ambulance monitoring syatem" International Journal of Applied Engineering Research, 2019.

[16] M Sanjay Karanth, Bindhu kumar K T, Gururaj Reddy J, Manoj K, Veda. B, "Smart ambulance with patient monitoring", 2015.

[17] Mery Subito1, Alamsyah, and Ardi Amir, "Web-Based Wireless Monitoring System on Patient's Vital Sign", 2019.

[18] Tia Gao, Logan K. Hauenstein, Alex Alm, David Crawford, Cassius K. Sims, Azmat Husain, and David M, "Vital Signs Monitoring and Patient Tracking Over a Wireless Network", 2016

[20] Amogh Powar, Seema Shilvant, Varsha Pawar, Pratiksha Shetgaonkar, Shailendra Aswale, "Data Mining & Artificial Intelligence Techniques for Prediction of Heart attack", 2019.