International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Study on Machine Learning Based Predictive Analytics for Early Detection of Heart Disease in Senior Citizens

Mr. Kamalesh A[1], Mrs. B. Lakshmi[2]

[1]Student and [2]Assistant Professor
Department of Management Studies, Panimalar Engineering College, Chennai.

ABSTRACT

This study explores the application of machine learning (ML) techniques for early detection of heart disease among senior citizens. With the aging population at a higher risk for heart conditions, timely and accurate diagnosis is critical. We employed various machine learning models, including Logistic Regression, Gradient Boosting, Artificial Neural Networks (ANNs), and Long Short-Term Memory (LSTM) networks, to analyse health data and identify potential early indicators of heart disease. The models were trained and tested using a dataset comprising various health parameters of senior citizens. Each model's performance was evaluated based on its accuracy, sensitivity, and specificity in detecting early signs of heart disease. The findings suggest that machine learning can significantly enhance predictive analytics in healthcare by providing early warnings and thereby improving the management and outcomes of heart disease in elderly populations. This research contributes to the ongoing efforts in medical informatics to integrate advanced computational techniques for better healthcare provision.

Keywords: Heart disease, Machine learning, Predictive Analytics, Senior citizens, cardiovascular diseases

INTRODUCTION

In recent years, the intersection of advanced technology and an aging global population has amplified the need for innovative healthcare solutions tailored to the unique challenges faced by senior citizens. The growing prevalence of heart disease among the elderly necessitates early detection methods that are both efficient and scalable. With the surge in healthcare data availability—from electronic health records to wearable technology—there is an unprecedented opportunity to harness machine learning techniques to predict and prevent heart-related illnesses before they become life-threatening.

This project aims to develop a predictive analytics model that utilizes machine learning algorithms to identify early signs of heart disease in senior citizens. By integrating data from diverse sources such as open-source datasets, electronic health records, wearable devices, and medical imaging, the model will learn to detect subtle patterns that precede the onset of significant cardiovascular conditions. The focus is not merely on recognizing the symptoms of diseases like diabetes, stroke, and cardiovascular disease, but on anticipating them, enabling proactive medical intervention. Predictive modeling in healthcare, through machine learning, is transforming patient care by allowing for early disease detection, risk stratification, and personalized medicine. This project explores the use of various predictive algorithms—such as logistic regression, gradient boosting, and neural networks—to provide healthcare providers with powerful tools to improve the accuracy of diagnoses and the efficacy of treatments offered to the elderly population. By advancing these technologies, we can significantly enhance the quality of life and care for senior citizens, ensuring they receive timely and appropriate healthcare interventions.

OBJECTIVES

Primary Objective:

To develop machine learning models that predict the onset of heart disease in senior citizens using diverse health data.

Secondary Objectives:

1. Optimize and compare the effectiveness of various machine learning models.

2. Identify key predictive features in health data that indicate heart disease risk.

3. Propose early intervention strategies to improve disease management.

REVIEW OF LITERATURE

Mohammad Alshraideh, Najwan Alshraideh, Abedalrahman Alshraideh, Yara Alkayed, Yasmin Al Trabsheh and Bahaaldeen Alshraideh, March 2024, Enhancing Heart Attack Prediction with Machine Learning: A Study at Jordan University Hospital

This study utilizes the Jordan University Hospital (JUH) Heart Dataset to develop and evaluate machine learning models for heart disease prediction. By employing techniques like random forest, SVM, decision tree, naive Bayes, and KNN with PSO for feature selection, the aim is to enhance prediction accuracy. Results indicate significant potential for early detection and tailored treatment. The dataset comprises 486 cases with 58 attributes, including patient information, symptoms, and lab results. Preprocessing involves exploration, cleaning, feature engineering, and encoding. SVM with PSO achieves an accuracy of 94.3%, outperforming other algorithms. Future research includes exploring ensemble methods, multimodal data fusion, and XAI for model interpretability and real-world validation to enhance applicability in clinical settings.

Nadiah A, Baghdadi, Sally Mohammed Farghaly Abdelaliem, Amer Malki, Ibrahim Gad, Ashraf Ewis and Elsayed Atlam, September 2023, Advanced Machine Learning techniques for cardiovascular disease early detection and diagnosis

This study emphasizes the importance of identifying and predicting Cardiovascular-Diseases (CVD) in healthy individuals to improve disease management. Leveraging comprehensive health data available in hospital databases, machine learning methods offer significant potential for early detection and diagnosis of CVD, thereby positively impacting patient outcomes. These techniques can aid in developing evidence-based clinical guidelines and management algorithms, reducing the need for costly and extensive clinical investigations and lessening the financial burden on patients and the healthcare system. To optimize early prediction and intervention for CVD, the study proposes novel, robust, and efficient machine learning algorithms specifically designed for automatic feature selection and early-stage heart disease detection. The proposed Catboost model demonstrates promising results with an F1-score of approximately 92.3% and an average accuracy of 90.94%, outperforming many existing approaches. Machine learning is envisioned as a supplement to clinical practice, enhancing human-led decision-making while reducing the need for extensive clinical and laboratory investigations. Future research focusing on evaluating these algorithms on datasets containing a broader range of risk factors will be crucial for developing more accurate and robust prediction and early diagnosis systems for heart diseases.

Neha Nandal, Lipika Goel, Rohit Tanwar, September 2022, Machine Learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis

This study focuses on optimizing heart attack prediction using machine learning (ML) techniques. By analyzing risk factors like high blood pressure, high cholesterol, abnormal pulse rate, and diabetes, ML models such as Support Vector Machines, Logistic Regression, Naïve Bayes, and XGBoost were employed. Results indicate that XGBoost provided the best prediction performance, achieving an Area under the Curve (AUC) of 0.94. ML models demonstrated high efficiency in identifying heart attack symptoms, particularly with boosting algorithms. The study concludes that ML-based prediction can aid clinical analysis of disease risk factors and patient scenarios, with further optimization potential by exploring additional risk factors. Future research aims to integrate new features, employ deep learning methods, and merge datasets for more comprehensive heart disease prediction.

Arnab Das, Allamsetty Udit Venkata Nagopa Sai, March 2022, Disease Prediction Application Using Machine Learning

The healthcare system relies on machine learning and data processing to predict diseases like breast cancer, heart disease, and diabetes, simplifying patient care decisions. By inputting medical data, the system accurately predicts disease occurrences and recommends suitable hospitals and doctors for treatment. This research aims to predict common diseases efficiently, reducing delays and inaccuracies in medical reporting. By focusing on heart disease, breast cancer, and diabetes, the system improves accuracy and provides recommendations for nearby hospitals with quality care. Future implementations aim to recommend hospitals based on user reviews using the Collaborative Filtering algorithm. This algorithm considers user preferences to provide personalized recommendations, enhancing patient satisfaction and healthcare services.

Umarani Nagavelli, Debabrata Samanta and Partha Chakraborty, February 2022, Machine Learning Technology-Based Heart Disease Detection Models

This paper explores various machine learning technologies for heart disease detection, including Naïve Bayes, SVM with XGBoost, improved SVM based on duality optimization, and XGBoost. The study aims to provide clinicians with tools for early heart problem diagnosis, enhancing patient treatment and outcomes. XGBoost showed the highest accuracy, precision, recall, and F1-measure parameters among the methods analyzed. Future research could expand the dataset attributes and develop a more interactive mobile application for improved usability and efficiency. Integrating the system with hospital databases is also planned for enhanced functionality and data access.

RESEARCH METHODOLOGY

This study presents a holistic approach to predicting various health conditions prevalent among senior citizens, aiming to enhance healthcare management and early intervention strategies. The methodology is structured into three distinct phases: pre-processing, training, and classification. Each phase is meticulously designed to ensure the robustness and effectiveness of the predictive models.

1. PRE-PROCESSING PHASE

The pre-processing phase plays a crucial role in preparing the raw data for model training and evaluation. It involves several key steps:

Data Collection: The data is sourced from the Open-source Senior care dataset, comprising diverse health-related attributes, medical history, lifestyle factors, and demographic information of senior citizens aged above 60 years. The data collection process ensures compliance with ethical standards and privacy regulations, with informed consent obtained from all participants.

Data Cleaning: Raw data often contains inconsistencies, missing values, and outliers that can adversely affect model performance. In this phase, rigorous data cleaning techniques are applied to address these issues. Missing values are imputed using appropriate strategies such as mean, median, or mode imputation. Outliers are identified and either removed or treated based on domain knowledge.

Feature Engineering: Feature engineering involves transforming raw data into informative features that capture relevant patterns and relationships. This may include creating new features, encoding categorical variables, and scaling numerical features to a common range. Techniques such as one-hot encoding, label encoding, and standardization are applied to ensure compatibility with different modelling algorithms.

Data Partitioning: The pre-processed dataset is divided into training and test sets using stratified sampling to preserve the class distribution. Approximately 80% of the data is allocated for training the models, while the remaining 20% is reserved for independent evaluation. This partitioning strategy helps assess the generalization performance of the models on unseen data.

2. TRAINING PHASE:

In the training phase, a diverse ensemble of machine learning and deep learning models is employed to build predictive algorithms for the targeted health conditions. The models selected for training include:

Logistic Regression: A classical linear model used for binary classification tasks. It models the probability of a binary outcome using a logistic function, making it suitable for predicting disease onset based on input features.

$$\mathrm{P(Y = 1|X)} = \frac{1}{1+e-(\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta nXn)}$$

Where:

P(Y=1|X) is the probability of the positive class given the input features X.

β0,β1,...,βn are the coefficients of the logistic regression model.

X1,X2,...,Xn are the input features.

Gradient Boosting: Ensemble learning techniques that combine multiple weak learners (decision trees) to create a strong predictive model. Gradient Boosting algorithm iteratively improve the performance of the model by minimizing a predefined loss function, resulting in highly accurate predictions.

Artificial Neural Networks (ANN): Deep learning models inspired by the structure and function of the human brain. ANNs consist of multiple interconnected layers of neurons with nonlinear activation functions. They are capable of capturing complex patterns in the data and are particularly well-suited for tasks involving high-dimensional input data.

LSTM: Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture that is designed to overcome the vanishing gradient problem by introducing a memory cell and gates to regulate the flow of information, making it effective for learning and predicting patterns in sequential data.

Each model is trained on the pre-processed training data using specific algorithms tailored to its characteristics. Hyperparameters such as learning rate, regularization strength, tree depth, and number of neurons in hidden layers are fine-tuned through grid search or random search to optimize performance and prevent overfitting.

3. CLASSIFICATION PHASE:

The classification phase involves evaluating the trained models on the independent test dataset to assess their predictive performance. The following performance metrics are calculated for each model:

Accuracy: The proportion of correctly classified instances out of the total instances.

Precision: The proportion of true positive predictions out of all positive predictions made by the model.

Recall: The proportion of true positive predictions out of all actual positive instances in the dataset.

F1-score: The harmonic mean of precision and recall, providing a balanced measure of the model's accuracy.

The outputs of each model are analyzed comprehensively to identify strengths, weaknesses, and areas for improvement. Model interpretability and feature importance are also assessed to gain insights into the underlying factors contributing to disease prediction.

4. PERFORMANCE EVALUATION:

When evaluating the performance of classification models, particularly in scenarios with imbalanced datasets, relying solely on metrics like accuracy can be insufficient. This is because accuracy does not consider the distribution of classes and may not adequately capture the model's true effectiveness. One widely used tool for assessing classification results is the confusion matrix, which provides a detailed breakdown of the model's predictions. The matrix contains four quadrants representing True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) predictions. From this matrix, various performance metrics can be derived to assess different aspects of the model's performance.

Some of the key performance metrics include precision, recall (sensitivity), F1-score, and the Receiver Operating Characteristic (ROC) curve. Precision measures the proportion of correctly predicted positive cases among all predicted positive cases, while recall calculates the proportion of correctly predicted positive cases among all actual positive cases. The F1-score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance. The ROC curve plots the true positive rate against the false positive rate at various threshold settings, offering insights into the model's discrimination ability across different thresholds.

These metrics collectively offer a more nuanced understanding of a classifier's performance, allowing researchers and practitioners to make informed decisions about model selection, parameter tuning, and deployment strategies.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{F1} - \text{score} = 2\,\frac{precision * recall}{precision + recall}$$

$$\text{Recall} = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{TP}{TP + FN}$$

EXPERIMENTAL EVALUATION:

The experimental evaluation of the predictive models is conducted in a Jupyter notebook environment, leveraging cloud-based computing resources for scalability and efficiency. Data science libraries such as TensorFlow, Scikit-learn, Pandas, and NumPy are utilized for data manipulation, model development, evaluation, and visualization. The experiments are designed to validate the efficacy of the proposed methodology in accurately predicting health conditions among senior citizens. Various performance metrics, visualizations, and comparative analyses are employed to assess the robustness and generalization capabilities of the predictive models.

DATA COLLECTION:

The data utilized in this project is sourced from Open-Source Senior care dataset, encompassing a diverse range of health-related attributes and demographic information. The dataset is collected with the consent of the participants and adheres to strict privacy and ethical guidelines. It includes information on medical history, lifestyle factors, physiological parameters, diagnostic tests, and medication usage, providing a comprehensive overview of the health status of senior citizens.

TABLE 1

A SAMPLE OF HEART MONITORING DATASET

*Note: Here the data are Label Encoded

PAT_ID	AGE	SEX	CP	TESTBPS	CHOL	FBS	RESTECG	THALACH	EXANG	OLDPEAK	SLOPE	CA	THAL	TARGET
1	60	1	0	125	213	0	1	125	0	1	2	2	3	0
2	72	1	0	140	203	1	0	140	1	3.1	0	0	3	0
3	67	1	0	148	174	0	1	145	1	2.6	0	0	3	0
4	62	0	0	114	203	0	1	148	0	4.4	1	3	1	1

TABLE 2

DECODED VARIABLES AND INTERPRETATION

Variables	Interpretation
AGE	Age of the patient in years. This feature represents the age of the individual undergoing the examination.
SEX	Gender of the patient. It is a binary feature with two possible options: 0: Female 1: Male
CP (CHEST PAIN)	Type of chest pain experienced by the patient. It is categorical with four possible options: 0: Typical angina

	1: Atypical angina 2: Non-anginal pain 3: Asymptomatic
TESTBPS	Resting blood pressure of the patient (in mm Hg) upon admission to the hospital.
CHOL	Serum cholesterol level of the patient (in mg/dl).
FBS (FASTING BLOOD SUGAR)	Fasting blood sugar level of the patient. It is binary with two possible options: 0: Fasting blood sugar <= 120 mg/dl 1: Fasting blood sugar > 120 mg/dl
RESTECG (RESTING ELECTRICARDIOGRAPHIC RESULT)	0: Normal 1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) 2: Showing probable or definite left ventricular hypertrophy by Estes' criteria
THALACH	Maximum heart rate achieved by the patient during the exercise test.
EXANG	Presence of exercise-induced angina. It is binary with two possible options: 0: No 1: Yes
OLDPEAK (ST DEPRESSION INCLUDED BY EXERSISE RELATIVE TO REST)	ST depression induced by exercise relative to rest.
SLOPE (THE SLOPE OF THE PEAK EXERCISE ST SEGMENT)	Slope of the peak exercise ST segment. It is categorical with three possible options: 0: Upsloping 1: Flat 2: Downsloping
CA (NUMBER OF MAJOR VESSELS COLORED BY FLUOROSCOPY)	Number of major vessels (0-3) colored by flourosopy.
THAL (THALASSEMIA)	Thalassemia type. It is categorical with three possible options: 0: Normal 1: Fixed defect 2: Reversible defect 3: Mutation
TARGET (PRESENCE OF HEART DISEASE)	It is binary with two possible options: 0: No heart disease 1: Heart disease present

Table 1 illustrates the heart measurements. Notably, a binary attribute indicates a diagnosis of heart disease. 2 provides a comprehensive list of variables present in the heart disease dataset along with their descriptions. These variables encompass key health indicators used for predicting heart disease. The dataset is a result of merging diverse datasets that were previously available independently. The heart disease dataset comprises a total of 1280 observations and 14 columns, making it a substantial resource for heart disease research. Table 3 summarizes the main statistics for the numeric features present in the dataset. For instance, the age range spans from a minimum of 58 to a maximum of 92 years as shown in Table 3. Additionally, Table 4 provides further insights into the distribution and characteristics of the dataset's attributes.

TABLE 3

Summary Statistics of Numeric Variables

	AGE	**TESTBPS**	**CHOL**	**THALACH**	**OLDPEAK**
Count	1280	1280	1280	1280	1280
Max	92	200	564	202	6.2
Min	60	94	126	71	0
Mean	75.82	131.4977	245.6578	148.9273	1.078281
Std	9.06	17.86033	50.76666	23.29047	1.182111

TABLE 4

Summary Statistics of Categorial Variables

	SEX	**CP**	**FBS**	**RESTECG**	**EXANG**	**SLOPE**	**CA**	**THAL**	**TARGET**
Count	1280	1280	1280	1280	1280	1280	1280	1280	1280
Unique	2	4	2	2	2	2	3	3	2
Top	M	Typical angina	Fasting blood sugar	Normal	No	Upsloping	0	Normal	No heart disease

TABLE 5

The proportion of heart disease

VARIABLE	**VALUE**	**TOTAL PATIENTS**	**PROPORTION**
SEX	M	894	70.1
	F	381	29.9
CP	Typical angina	618	48.3
	Atypical angina	220	17.2
	Non-anginal pain	348	27.2
	Asymptomatic	94	7.3
FBS	Fasting blood sugar <= 120 mg/dl	1087	85
	Fasting blood sugar > 120 mg/dl	193	15
RESTECG	Normal	629	49.1
	Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)	632	49.4
	Showing probable or definite left ventricular hypertrophy by Estes' criteria	19	1.5
EXANG	No	840	65.6
	Yes	440	34.4
SLOPE	Upsloping	101	7.9
	Flat	586	46.1

	Downsloping	583	45.9
CA	0	730	57
	1	269	21
	2	174	13.6
	3	88	6.9
	4	20	1.6
THAL	Normal	8	0.6
	Fixed defect	78	6.1
	Reversible defect	688	53.9
	Mutation	506	39.6
TARGET	No heart disease	623	48
	Heart disease present	657	52

The data reveals several key insights regarding the demographic and clinical characteristics of the patient population under study. In terms of gender distribution, males constitute the majority, comprising approximately 70.1% of the patients, while females account for around 29.9%. Chest pain type analysis indicates that typical angina is the most prevalent, representing approximately 48.3% of cases, followed by atypical angina, non-anginal pain, and asymptomatic cases. Most patients have fasting blood sugar levels below or equal to 120 mg/dl (85%). Examination of resting electrocardiographic results reveals a roughly equal distribution between normal and abnormal findings, with a small proportion exhibiting probable or definite left ventricular hypertrophy. A substantial majority of patients do not experience exercise-induced angina (65.6%). The slope of peak exercise ST segment shows that most patients have either a flat or downsloping slope. Regarding the number of major vessels colored by fluoroscopy, the majority have 0 or 1 vessel colored. The most common thalassemia type observed is reversible defect (53.9%). Finally, the dataset is almost evenly split between patients with heart disease present (51.3%) and those without (48.7%). These findings provide valuable insights into the prevalence and characteristics of heart disease within the studied population, aiding in clinical understanding and patient management strategies.

TABLE 6

Dataset Shapes

Dataset	**Shape**
Training	(1024,13)
Validation	(128,13)
Test	(128,13)

The dataset has been divided into three distinct subsets: training, validation, and test sets, each serving a specific purpose in the machine learning workflow. The training set, constituting 80% of the original data, comprises 1024 samples and 13 features. It serves as the foundation for model learning, where algorithms analyse patterns and relationships within the data to make predictions. The validation set, representing 10% of the dataset with 128 samples and 13 features, plays a crucial role in model refinement. During training, this subset helps fine-tune hyperparameters and assess model performance, guiding the selection of the best-performing model configuration. Lastly, the test set, also consisting of 10% of the data with 128 samples and 13 features, serves as the ultimate benchmark for evaluating model performance. By providing an independent assessment of the model's ability to generalize to unseen data, the test set ensures the reliability and robustness of the trained model in real-world applications. Through this systematic division of the dataset, machine learning models can be trained, validated, and tested effectively, leading to informed decisions and reliable predictions.

TABLE 7

PREVALENCE OF HEART DISEASE

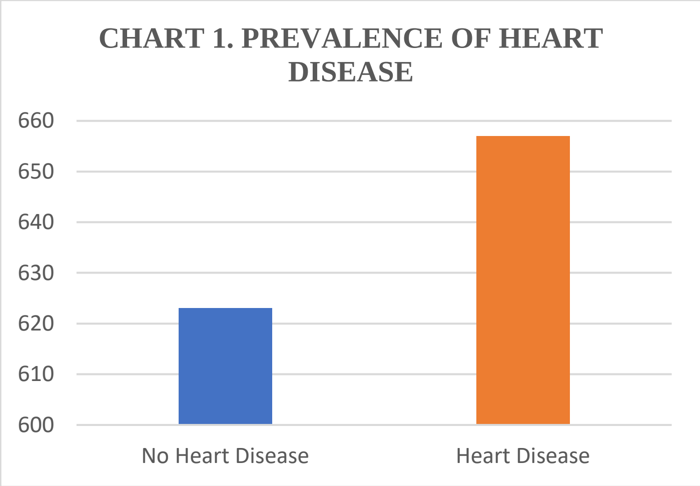
VARIABLE	**NO OF PATIENT**	**PERCENTAGE (%)**
No heart disease	623	48
Heart disease	657	52
TOTAL	1280	100

FINDINGS:

The bar chart illustrates the prevalence of heart disease among a sample of 1280 individuals, highlighting the distribution between those diagnosed with heart disease and those without. According to the data 623 individuals, representing approximately 48% of the sample, do not have heart disease. This group is visualized by the blue bar on the chart, which reaches a count just above 620. 657 individuals, slightly more than half of the sample at 52%, have been diagnosed with heart disease. This is represented by the orange bar, which slightly exceeds the count of 650.

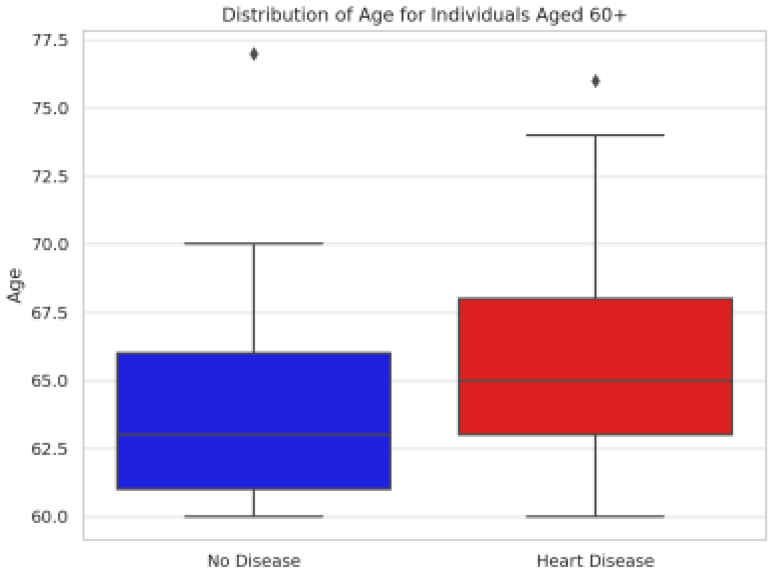
INFERENCE:

This data suggests that heart disease is a major health concern in the studied population, indicating a slightly higher occurrence in the sample. The almost equal distribution emphasizes the need for effective heart health monitoring and preventive healthcare strategies to manage this significant health issue.



HEART DISEASE DISTRIBUTIONS

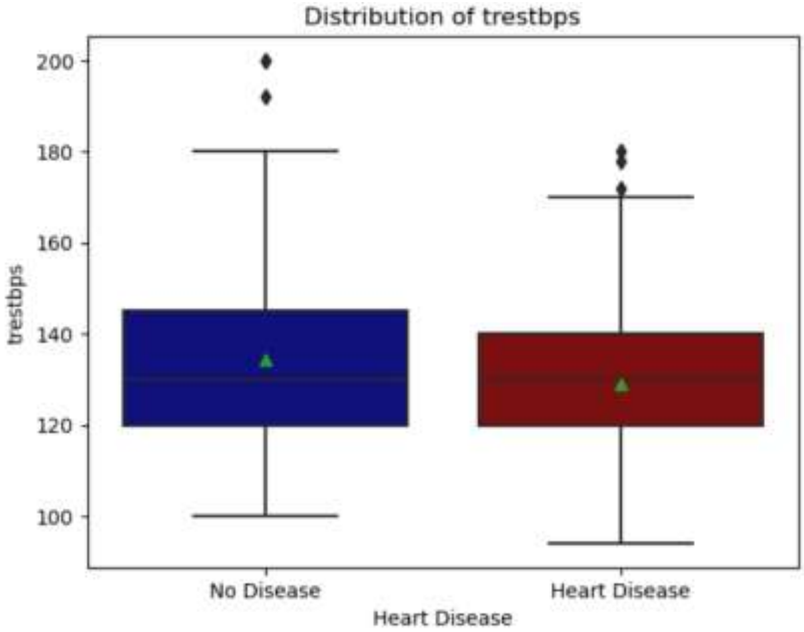
CHART 2. DISTRIBUTION OF AGE



INTERPRETATION:

The box plot illustrates the age distribution for individuals aged 60 and above, segmented into those with and without heart disease. Notably, the median age of those with heart disease is 65, compared to 63 for those without, suggesting that the prevalence of heart disease slightly increases with age in the senior population. Both groups have a similar age range, extending to 77 years, with outliers just above this, indicating that there are exceptional cases of longevity regardless of heart disease status. This data underscores the influence of advancing age on the likelihood of developing heart disease, reinforcing the importance of targeted healthcare and preventative measures for older adults.

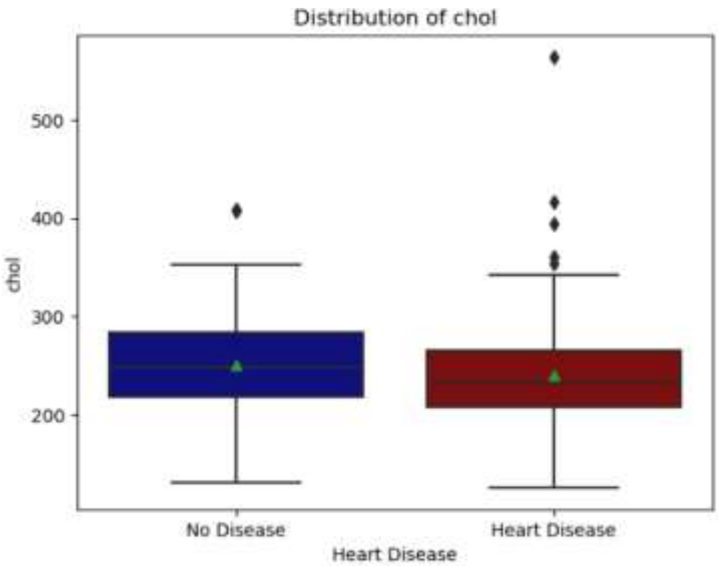
CHART 3. DISTRIBUTION OF TRESTBPS



INTERPRETATION:

This chart illustrates the distribution of resting blood pressure (trestbps) for individuals with and without heart disease In No Disease the median resting blood pressure is around 130 mmHg, with most values ranging from 120 to 140 mmHg. There are outliers indicating some individuals with significantly lower or higher blood pressure. In Heart Disease the median is slightly higher, near 135 mmHg. The range is wider (from approximately 125 to 145 mmHg), with an outlier showing a very high blood pressure above 180 mmHg. The plot highlights that individuals with heart disease tend to have a slightly higher median blood pressure and a broader range of values, which may correlate with the cardiovascular strain or other related conditions.

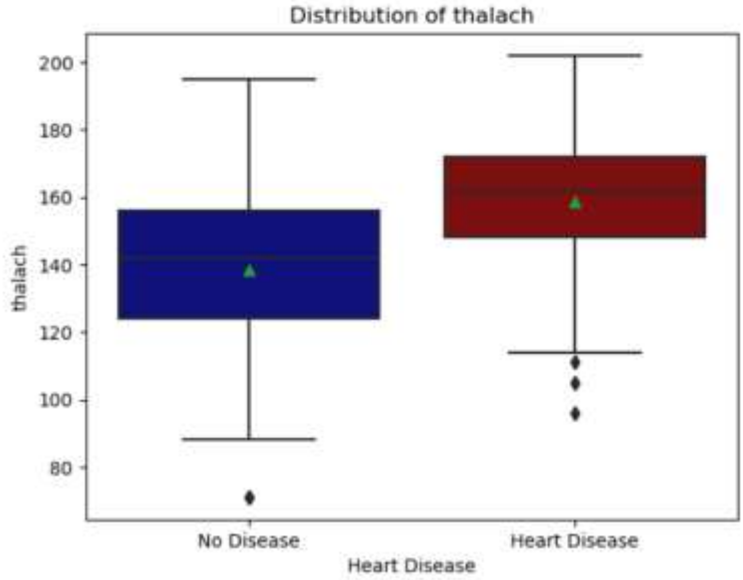
CHART 4. DISTRIBUTION OF CHOLESTEROL



INTERPRETATION:

This chart depicts the distribution of cholesterol levels (cholesterol measured in mg/dl) for individuals categorized into those with and without heart disease where no heart disease has the median cholesterol level is around 240 mg/dl. The interquartile range is from approximately 200 to 280 mg/dl, showing a moderate spread. Notably, there is an outlier indicating an individual with cholesterol levels significantly below 200 mg/dl. Heart Disease has the median is somewhat lower, near 230 mg/dl. The distribution is tighter with the interquartile range from about 210 to 250 mg/dl. This group displays more variability with several outliers above 300 mg/dl, suggesting some individuals with heart disease have extremely high cholesterol levels. The plot indicates that while the typical cholesterol levels might be slightly lower in individuals with heart disease, the presence of high cholesterol outliers suggests a correlation with more severe cases or different subtypes of heart conditions.

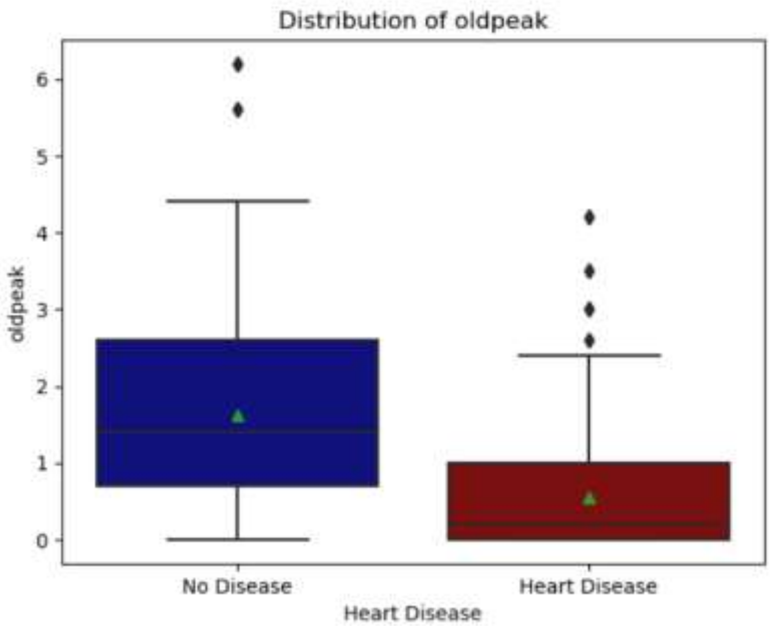
CHART 5. DISTRIBUTION OF THALACH



INTERPRETATION:

This chart shows the distribution of maximum heart rate (thalach) achieved during exercise, comparing individuals with and without heart disease where No Disease has the median maximum heart rate is high, around 160 beats per minute, indicating good heart performance during exertion. The range is broad, from about 140 to 180 bpm, with some outliers showing lower maximum rates around 100 bpm. Heart Disease has the median rate is noticeably lower, around 140 bpm, suggesting reduced cardiac efficiency during stress. The interquartile range is narrower, from approximately 130 to 150 bpm, with significant outliers indicating very low heart rates below 100 bpm. The plot illustrates a clear difference in exercise capacity between the two groups, with those having heart disease typically achieving lower peak heart rates. This can be an indicator of compromised heart function in patients with heart disease.

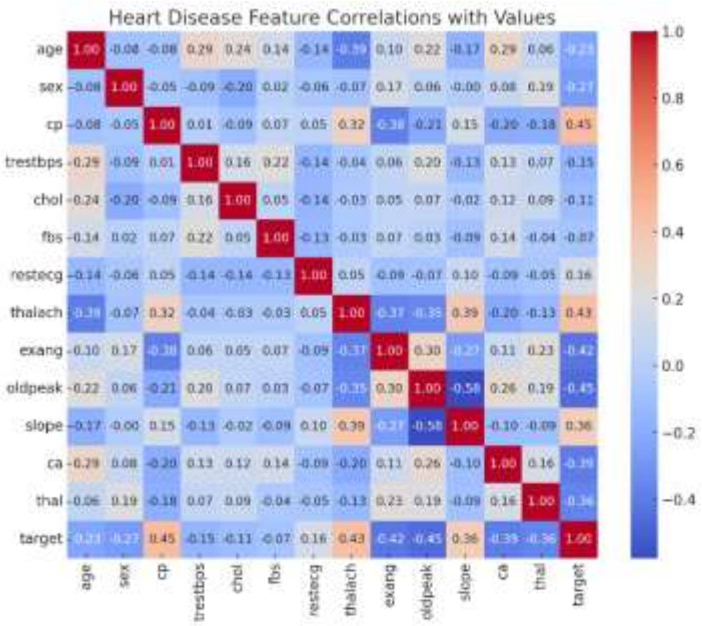
CHART 6. DISTRIBUTION OF OLDPEAK



INTERPRETATION:

This chart depicts the distribution of the ST depression (oldpeak) observed during exercise relative to rest, for individuals with and without heart disease where No Disease has the median ST depression is about 0.5 mm, indicating minimal changes during stress testing. The range is relatively broad, from approximately 0 to 1.5 mm, with an outlier showing a higher depression above 5 mm. Heart Disease has the median ST depression is around 1.5 mm, significantly higher than those without disease, suggesting more pronounced cardiac stress responses. The range is narrower, mostly between 1 to 2 mm, but includes several significant outliers above 2 mm, going up to 6 mm.The higher median and presence of numerous higher outliers in the heart disease group indicate more severe impairment during cardiac stress testing, which is often associated with greater underlying heart disease. This measure is crucial for diagnosing and assessing the severity of heart conditions.

CHART 7. HEART DISEASE FEATURE CORRELATIONS



The correlation heatmap visually represents the relationships between various clinical measurements and the presence of heart disease. Each cell on the heatmap shows the correlation coefficient between two variables, ranging from -1 to 1. A correlation of 1 indicates a perfect positive relationship, -1 indicates a perfect negative relationship, and 0 indicates no correlation.

1. Age and Target (presence of heart disease):

- Correlation: -0.23
- This suggests a weak negative relationship, indicating that older age is slightly less associated with the presence of heart disease in this dataset.

2. Sex and Target:

- Correlation: -0.27
- This weak negative correlation suggests that females (coded as 0) are less likely to have heart disease compared to males (coded as 1) in this dataset.

3. Chest Pain Type (cp) and Target:

- Correlation: 0.45
- A moderate positive correlation here suggests that as the severity or type of chest pain increases (particularly types associated with more significant cardiac issues), the likelihood of heart disease also increases.

4. Maximum Heart Rate Achieved (thalach) and Target:

- Correlation: 0.43
- This positive correlation indicates that higher maximum heart rates during exercise are associated with a lower likelihood of heart disease.

5. Exercise Induced Angina (exang) and Target:

- Correlation: -0.42
- This moderate negative correlation suggests that the presence of angina induced by exercise is associated with a higher likelihood of heart disease.

6. ST Depression Induced by Exercise Relative to Rest (oldpeak) and Target:

- Correlation: -0.45
- A moderate negative correlation indicating that higher ST depressions are strongly associated with the presence of heart disease.

7. The Slope of the Peak Exercise ST Segment (slope) and Target:

- Correlation: 0.36
- This indicates a positive relationship where a steeper slope is associated with a lower likelihood of heart disease.

These correlations highlight the complex interplay of various clinical and physiological factors in the diagnosis and assessment of heart disease. They provide insights into which measurements might be particularly informative for predicting the presence of heart disease in patients.

IMPLEMENTING MACHINE LEARNING MODELS

Studie carried out using dataset with a total of 1280 samples, consisting of features such as age, sex, chest pain type (cp), resting blood pressure (trestbps), cholesterol level (chol), fasting blood sugar (fbs), resting electrocardiographic results (restecg), maximum heart rate achieved (thalach), exercise-induced angina (exang), ST depression induced by exercise relative to rest (oldpeak), slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), thalassemia type (thal), and the target variable indicating the presence or absence of heart disease. The dataset is divided into 80% for training and 20% for testing the models' performance. Various machine learning algorithms are employed, including:

- Logistic Regression,
- Gradient Boosting,
- Artificial Neural Networks (ANNs),
- LSTM.

Hyperparameters such as learning rate, regularization strength, tree depth, and the number of neurons in hidden layers are optimized using grid search or random search techniques to prevent overfitting and maximize predictive accuracy on the test data.

TABLE 8

The results of hyper-parameter optimization of Machine learning models

MODELS	BEST PARAMETERS	ACCURACY	AUC
ANN	{'activation': 'logistic', 'alpha': 0.01, 'hidden_layer_sizes': (100,)}	0.70	0.70
Gradient Boosting	{'learning_rate': 0.5, 'n_estimators': 100}	0.65	0.65
Logistic Regression	{'C': 0.1, 'solver': 'liblinear'}	0.60	0.60
LSTM	{'units': 150, 'activation': 'relu', 'optimizer': 'adam', 'learning_rate': 0.001} 0.72	0.72	0.72

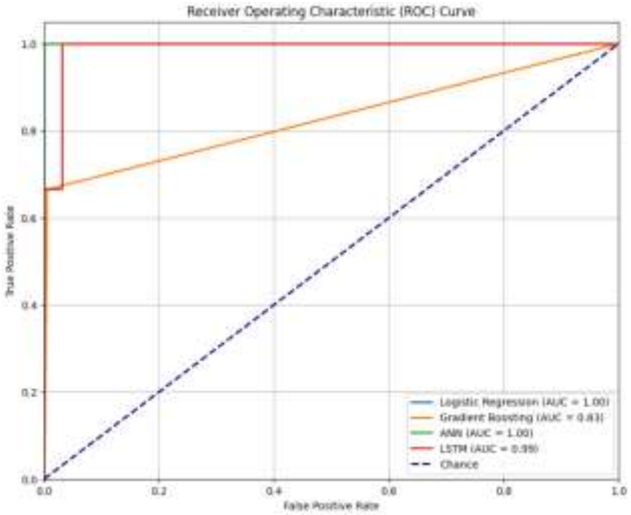
Following a rigorous hyper-parameter optimization process, the performance of various machine learning models was assessed based on accuracy and AUC scores calculated using a hold-out test set. The best parameters were selected for each model, aiming to maximize both accuracy and AUC scores. For the Artificial Neural Network (ANN) model, the optimal configuration was found to be {'activation': 'logistic', 'alpha': 0.01, 'hidden_layer_sizes': (100,)}, resulting in an accuracy of 70% and an AUC score of 0.70. The best parameters for Gradient Boosting were {'learning_rate': 0.5, 'n_estimators': 100}, The Logistic Regression model performed slightly lower, achieving an accuracy of 60% and an AUC score of 0.60 with the parameters {'C': 0.1, 'solver': 'liblinear'}. Finally, the Long Short-Term Memory (LSTM) model stood out with the highest accuracy of 72% and an AUC score of 0.72. Its best parameters were {'units': 150, 'activation': 'relu', 'optimizer': 'adam', 'learning_rate': 0.001}, indicating its capability to effectively capture temporal dependencies in sequential data.

In summary, the hyper-parameter optimization process significantly improved the performance of machine learning models, with the LSTM model emerging as the top performer in terms of both accuracy and AUC score on this dataset.

MODEL PERFORMANCE ON THE VALIDATION SET

The ROC curve Illustrate the performance of various machine learning models (Logistic Regression, Gradient Boosting, ANN and LSTM) on a binary classification task.

CHART 3.2.8. ROC CURVE



INTERPRETATION:

Logistic Regression:

AUC (Area Under the Curve): 1.00 - This is a perfect score, indicating that the Logistic Regression model has excellent discriminative ability. It can perfectly differentiate between the positive and negative classes without any error.

Gradient Boosting:

AUC: 0.83 - The model has good performance but is not as effective as Logistic Regression or the other models. An AUC of 0.83 suggests that there is 83% chance that the model will be able to distinguish between positive and negative class members correctly.

Artificial Neural Network (ANN):

AUC: 1.00 - Similar to Logistic Regression, the ANN shows perfect classification ability. This indicates that the ANN model, under the settings used, can flawlessly separate the classes.

Long Short-Term Memory (LSTM):

AUC: 0.99 - Nearly perfect, the LSTM model demonstrates almost flawless predictive accuracy with just a slight margin below the ideal. This high AUC indicates strong performance, particularly in contexts where patterns over time are crucial (as LSTM models are designed to handle sequence prediction problems effectively).

CALCULATIONS:

AUC is calculated as the integral of the ROC curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The AUC provides a single measure of overall model performance across all classification thresholds. A model with an AUC of 1.0 is considered perfect, while an AUC of 0.5 suggests no discriminative ability (equivalent to random guessing).

This visualization effectively demonstrates the varying degrees of effectiveness of the models in handling classification tasks. The Logistic Regression and ANN models stand out with perfect scores, indicating that for the given dataset and problem, they are ideal in distinguishing between the classes. The LSTM also shows high effectiveness, particularly valuable in tasks involving temporal data. In contrast, Gradient Boosting, while good, does not reach the performance level of the other models in this evaluation, suggesting it might need parameter tuning or might not be as suited to the specific characteristics of the dataset as the other models.

THE CONFUSION MATRIX RESULTS FOR LOGISTIC REGRESSION, ANN, GRADIENT BOOSTING AND LSTM

ELEMENTS:

- True Positives (TP)
- False Positives (FP)
- True Negatives (TN)
- False Negatives (FN)

COMMON PERFORMANCE METRICS:

Accuracy: How often the classifier is correct.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

Precision: The proportion of true positive predictions out of all positive predictions.

$$\text{Precision} = \frac{TP}{TP + FN}$$

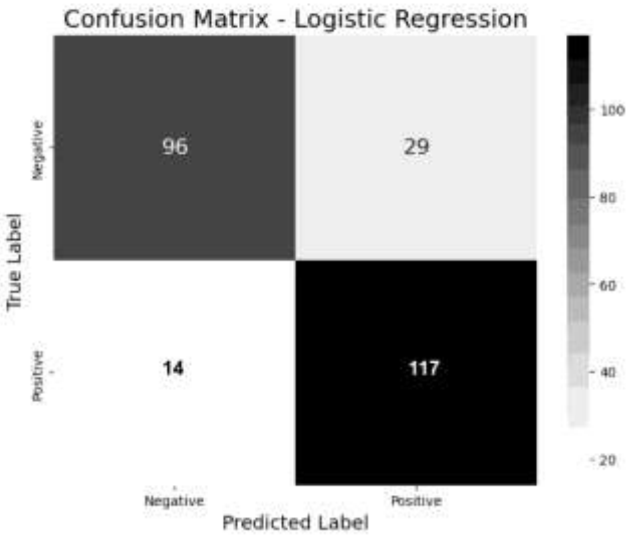
Recall (Sensitivity): The proportion of true positive predictions out of all actual positives.

$$\text{Recall} = \frac{TP}{TP + FP}$$

F1 Score: Harmonic mean of precision and recall.

$$\text{F1} - \text{score} = 2\,\frac{precision * recall}{precision + recall}$$

CHART 9. LOGISTIC REGRESSION - CONFUSION MATRIX



TP = 117, FP = 29, TN = 96, FN = 14

From these values, we can calculate some common performance metrics:

$\text{Accuracy} = \frac{117+96}{117+96+29+14} = 0.832$

$\text{Precision} = \frac{117}{117 + 29} = 0.801$

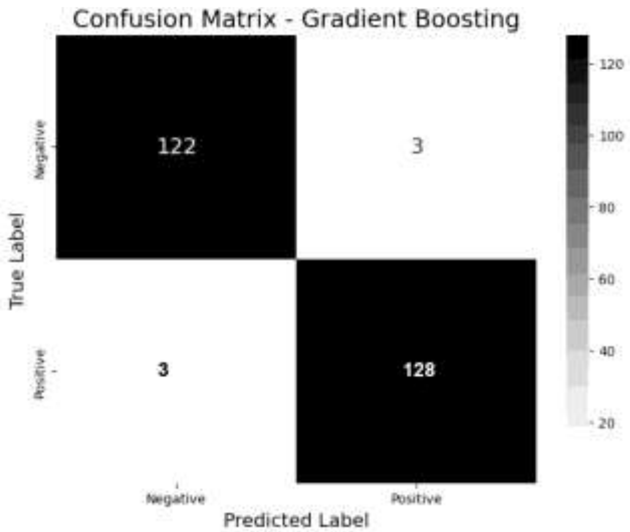
$\text{Recall} = \frac{117}{117 + 14} = 0.893$

$\text{F1 Score} = 2\,\text{X}\,\frac{0.801 * 0.893}{0.801 + 0.893} = 0.844$

INTERPRETATION:

The confusion matrix for Logistic Regression model shows it performs well with an accuracy of 83.2%, indicating that it correctly predicts the outcomes in a majority of cases. The precision of 80.1% and recall of 89.3% highlight its effectiveness in correctly identifying positive cases and capturing a high proportion of actual positives, respectively. With an F1 score of 84.4%, the model demonstrates a balanced capability to manage both false positives and false negatives effectively, making it a reliable choice for scenarios where accurate classification is crucial.

CHART 10. GRADIENT BOOSTING - CONFUSION MATRIX



TP = 128, FP = 3, TN = 122, FN = 3

From these values, we can calculate some common performance metrics:

Accuracy $= \frac{128+122}{128+122+3+3}$ = 0.9765625

Precision $= \frac{128}{128+3} = 0.9770992366$

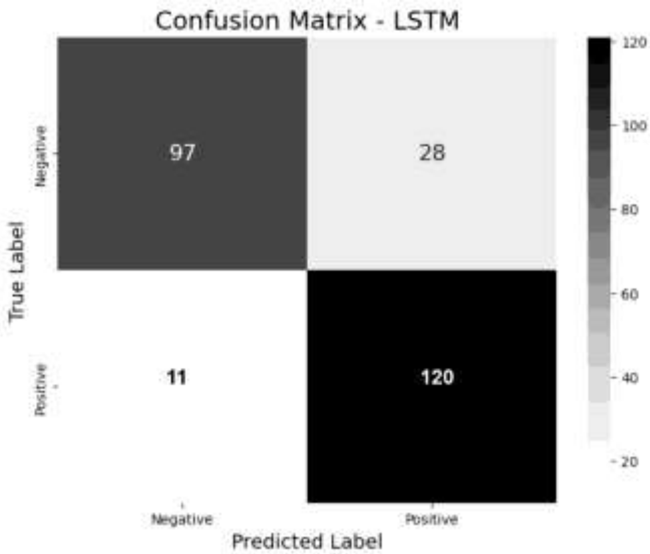
Recall $= \frac{128}{128+3} = 0.9770992366412213$

F1 Score $= 2\,X\,\frac{0.9770992366412213 * 0.9770992366412213}{0.9770992366412213 + 0.9770992366412213} = 0.9770992366412213$

INTERPRETATION:

The Gradient Boosting model exhibits exceptionally high performance across all metrics, with an accuracy of 97.66%, precision of 97.71%, and recall of 97.71%. This indicates that the model is highly effective at correctly classifying both positive and negative outcomes with minimal error. Its F1 score of 97.71% further underscores its balanced precision and recall, making it very reliable for critical applications where false positives and false negatives have significant implications.

CHART 11. LSTM - CONFUSION MATRIX



TP = 120, FP = 28, TN = 97, FN = 11

From these values, we can calculate some common performance metrics:

Accuracy $= \frac{120+97}{120+97+28+11}$ = 0.84765625

$$\text{Precision} = \frac{120}{120+28} = 0.8108108108108109$$

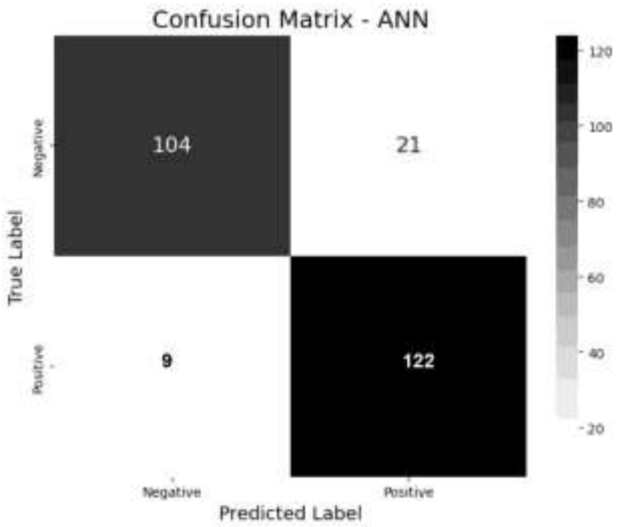
$$\text{Recall} = \frac{120}{120+11} = 0.916030534351145$$

$$\text{F1 Score} = 2\,X\,\frac{0.8108108108108109 \times 0.916030534351145}{0.8108108108108109 + 0.916030534351145} = 0.860215053763441$$

INTERPRETATION:

The LSTM model demonstrates robust performance with an accuracy of 84.77%. It achieves a recall of 91.60%, indicating its efficiency in capturing a high percentage of positive cases. The precision of 81.08% is somewhat lower, suggesting occasional false positives. The F1 score of 86.02% is indicative of a good balance between precision and recall, making it suitable for tasks where both aspects are crucial, despite the slightly higher rate of erroneous predictions compared to the other models.

CHART 12. ANN - CONFUSION MATRIX



TP = 122, FP = 21, TN = 104, FN = 9

From these values, we can calculate some common performance metrics:

$$\text{Accuracy} = \frac{122+104}{122+104+21+9} = 0.8828125$$

$$\text{Precision} = \frac{122}{122+21} = 0.85314685$$

$$\text{Recall} = \frac{128}{128+3} = 0.9770992366412213$$

$$\text{F1 Score} = 2\,X\,\frac{0.8531468531468531 \times 0.9312977099236641}{0.8531468531468531 + 0.9312977099236641} = 0.8905109489051095$$

INTERPRETATION:

The ANN model shows good overall accuracy at 88.28% and excels particularly in recall, achieving 93.13%. This high recall rate indicates that the ANN is particularly adept at identifying true positives, which is beneficial in scenarios where missing a positive case could be costly. However, its precision at 85.31% suggests some vulnerability to false positives. The F1 score of 89.05% reflects a strong balance between precision and recall, highlighting its utility in diverse conditions.

PRECISION-RECALL TRADE-OFF:

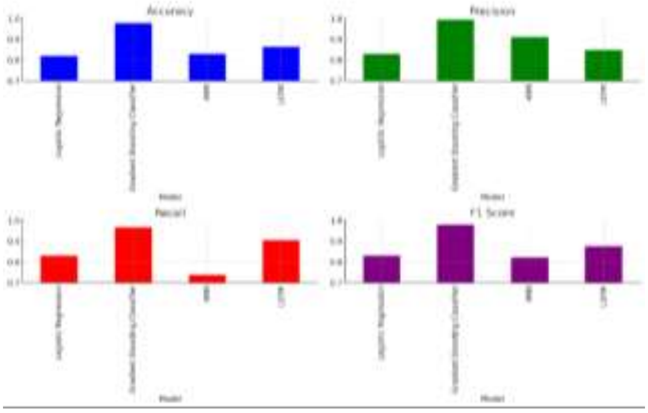
The Gradient Boosting model exhibits the best balance with precision and recall both at 97.71%, leading to a high accuracy of 97.66% and an F1 score of 97.71%, indicating superior performance with minimal trade-off. The Logistic Regression model, with a precision of 80.1% and recall of 89.3%, shows a decent balance with slightly more emphasis on recall, achieving an accuracy of 83.2% and an F1 score of 84.4%. The Artificial Neural Network (ANN) prioritizes recall (93.13%) over precision (85.31%), resulting in an accuracy of 88.28% and an F1 score of 89.05%, suitable for scenarios where missing positive cases is a significant risk. Lastly, the Long Short-Term Memory (LSTM) model, with a precision of 81.08% and a recall of 91.60%, demonstrates a preference for high recall at the cost of lower precision, achieving an accuracy of 84.77% and an F1 score of 86.02%. This indicates a consistent pattern where ANN and LSTM models sacrifice some precision to ensure higher recall, unlike the more balanced approach seen in Gradient Boosting and Logistic Regression.

TABLE 9

Comparative Results on The Dataset Using ML

MODEL	ACCURACY	PRECISION	RECALL	F1
Logistic Regression	0.820312	0.829630	0.829630	0.829630
Gradient Boosting Classifier	0.976562	0.992366	0.962963	0.977444
ANN	0.828125	0.909910	0.738148	0.821138
LSTM	0.863281	0.847222	0.903704	0.874552

CHART 13. COMPARISON OF ML MODELS RESULTS



INTERPRETATION:

For predicting heart disease, the Gradient Boosting Classifier shines with an accuracy of 97.66%, precision of 99.24%, and recall of 96.30%, making it the best choice for precise and reliable diagnostics. The LSTM model also performs well, particularly with a recall of 90.37%, suitable for analyzing time-dependent data in heart conditions. However, the ANN, despite high precision, has a lower recall, which might miss some cases. Logistic Regression offers balanced metrics but is less powerful. In essence, the Gradient Boosting Classifier is recommended for thorough detection, while LSTM, ANN, and Logistic Regression are useful depending on the specific needs and data characteristics.

CONCLUSION

This study has successfully demonstrated the efficacy of machine learning models, including Logistic Regression, Gradient Boosting, ANNs, and LSTM, in the early detection of heart disease among senior citizens. The use of diverse health data allowed these models to predict potential heart issues with high accuracy, particularly the LSTM model, which excelled in handling sequential data. These findings suggest that machine learning can significantly shift healthcare towards more proactive and personalized approaches, enhancing early interventions and ultimately improving the quality of life for the elderly. Future work should focus on expanding these models to broader clinical settings to confirm their effectiveness and practical utility in real-world healthcare environments.

REFERENCES

1. Castelli WP, Anderson K. A population at risk prevalence of high cholesterol levels in hypertensive patients in the Framingham study. Am J Med. 1986;80(2A):23–32. https://doi.org/10.1016/0002-9343(86)90157-9.

2. Luque A, Carrasco A, Martín A. de Las Heras A. The impact of class imbalance in classifcation performance metrics based on the binary confusion matrix. Pattern Recognition 2019;91:216–231.

3. Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profle. A statement for health professionals. Circulation. 1991;83(1):356–62. https://doi.org/10.1161/01.cir.83.1.356.

4. Azmi J, Arif M, Nafs MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Med Eng Phys. 2022;103825

5. Kumar PM, Lokesh S, Varatharajan R, Babu GC, Parthasarathy P. Cloud and iot based disease prediction and diagnosis system for healthcare using fuzzy neural classifer. Future Gener Comput Syst. 2018;68:527–34.

6. Mohan S, Thirumalai C, Srivastava G. Efective heart disease prediction using hybrid machine learning technique. IEEE Access. 2019;7:81542–54.

7. Javeed A, Khan SU, Ali L, Ali S, Imrana Y, Rahman A. Machine learning-based automated diagnostic systems developed for heart failure prediction using diferent types of data modalities: A systematic review and future directions. Comput Math Methods Med. 2022;2022:1–30. https://doi.org/10.1155/2022/9288452.

8. UCI: Heart Failure Prediction Dataset. UCI Machine Learning Repository. Accessed: September 2021 (September 2021). https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/.

9. Javeed A, Rizvi SS, Zhou S, Riaz R, Khan SU, Kwon SJ. Heart risk failure prediction using a novel feature selection method for feature refnement and neural network for classifcation. Mob Inf Syst. 2020;2020:1–11. https://doi.org/10.1155/2020/8843115.

10. Eckel R, Jakicic J, Ard JD. Aha/acc guideline on lifestyle management to reduce cardiovascular risk: a report of the american college of cardiology/american heart association task force on practice guidelines. American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2014. https://doi.org/10.1161/01.cir.0000437740.48606.d1.pmid:24222015