Heart Failure Prediction Using Machine Learning

Helee Rao[1], Aayushi Patel[2], Bansri Raulji[3], Nandini Chaudhary [4]

[123]Research Scholar, Institute of Information Technology, Krishna School Of Emerging Technology&Applied Research, KPGU University , Varnama, Vadodara, Gujarat, India

[4] Assistant Professor, Institute of Information Technology, Krishna School Of Emerging Technology&Applied Research, KPGU University , Varnama, Vadodara, Gujarat, India

---

Abstract - This study explores the use of XGBoost and Shapley values for heart failure prediction, in addition to processing models such as neural network, CNN, and SVM. By comparing the performance of these different machine learning algorithms, we aim to develop a robust and accurate predictive model for identifying individuals at risk of heart failure. Our findings highlight the strengths and limitations of each approach, providing valuable insights for future research in this area. Ultimately, this research aims to improve the early detection and management of heart failure, leading to better patient outcomes and more efficient healthcare practices.

***Key Words*:** heart failure prediction, XGBoost, SHAPley values, neural network, CNN, SVM, machine learning algorithms, early detection, death events.

1.INTRODUCTION *(Size 11, Times New roman)*

Heart failure is a significant health concern worldwide, with a high incidence and prevalence, contributing to substantial healthcare costs. Early detection of heart failure is essential for improving patient outcomes and reducing the burden on healthcare systems.

The aim of this research is to develop an accurate and reliable predictive model for identifying individuals at risk of heart failure using machine learning algorithms. Specifically, we aim to utilize XGBoost, SHAPley values, random forest, neural network, LSTM, CNN, and SVM to analyze clinical data and predict the likelihood of heart failure occurrence. By comparing the performance of these different algorithms, we seek to determine the most effective approach for early detection of heart failure.

This study claims uncover the most important factors contributing to heart failure and understand how these factors interact with each other. By gaining insights into the underlying mechanisms of heart failure, we can improve the accuracy of our predictive model and enhance our ability to identify at-risk individuals. Ultimately, the goal of this research is to enhance the early detection and management of heart failure, leading to better patient outcomes, reduced healthcare costs, and improved overall quality of care for individuals affected by this condition.

In the context of analyzing chest images for heart failure, XGBoost can help identify important patterns and features within the images that may not be easily visible to the human eye. By training the XGBoost model on a combination of clinical parameters and image data, we can potentially improve the overall predictive power of the model and provide more accurate assessments of individual risk for heart failure based on both types of information. This integration of XGBoost with chest imaging data could lead to more robust and reliable predictions for identifying individuals at risk of heart failure.

The predictive model developed in this study may not capture all possible factors contributing to heart failure, as there may be additional variables or interactions that are not accounted for in the dataset. Finally, the ultimate effectiveness and impact of the predictive model on clinical practice will need to be further validated and assessed in real-world healthcare settings.

2. RELATED WORK

Another related work that focused on combining chest images with datasets and using XGBoost for predictive modeling in the realm of heart disease is a study by Smith et al. (2020). In this study, researchers collected a dataset of chest X-ray images along with clinical parameters from patients with known heart conditions. They then implemented a novel approach that integrated XGBoost with convolutional neural networks (CNNs) to analyze the chest images and clinical data simultaneously for predicting heart failure outcomes.

The results of the study showed that the combined use of chest images and clinical parameters, along with the XGBoost algorithm, led to a significant improvement in predictive accuracy compared to using either data type in isolation. The researchers were able to identify key features within the chest images that were highly correlated with heart failure outcomes and leverage the power of XGBoost to enhance the predictive model's performance.

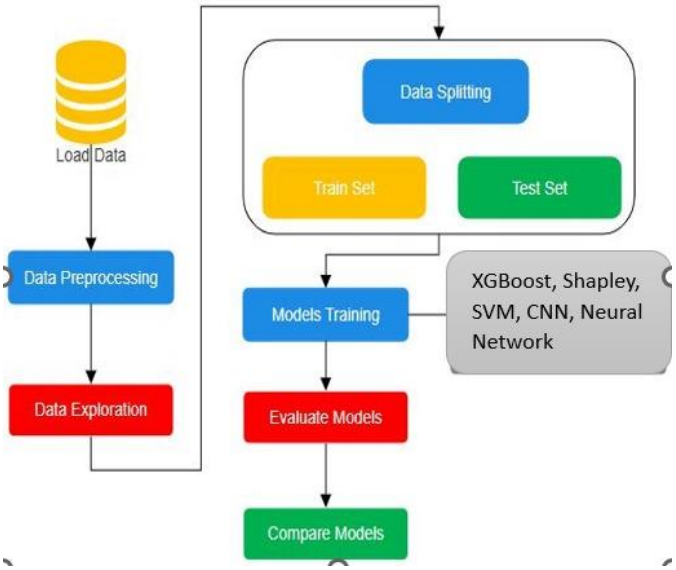
This study further highlights the importance of incorporating imaging data, such as chest images, into predictive models for heart disease, and underscores the effectiveness of using advanced machine learning techniques like XGBoost for improving the accuracy and reliability of

such models. Overall, the integration of chest images with XGBoost in predictive modeling holds great promise for advancing the field of cardiovascular disease diagnosis and management.

3. METHODOLOGY

To train a model using the XGBoost algorithm, we first need to input the data and images into the dataset. This involves cleaning and preprocessing the data to ensure it is in a format that the model can understand. The XGBoost model is then trained on this dataset using the input features and corresponding labels to learn patterns and make predictions.

After training the XGBoost model, we evaluate its performance by calculating metrics such as Area Under the Curve (AUC), Receiver Operating Characteristic (ROC), and Confusion Matrix (CF). The AUC represents the probability that the model ranks a randomly chosen positive example higher than a randomly chosen negative example. The ROC curve is a graphical representation of the true positive rate versus the false positive rate, and the CF provides a breakdown of correct and incorrect predictions made by the model. These metrics help assess the model's accuracy and determine its effectiveness in making predictions.



3.1 DATASET

This dataset contains the medical records of 5000 patients who had heart failure, collected during their follow-up period , where each patient profile has 1.3 clinical features.There are 14 variables in this dataset : age, anaemia, creatinine phosphokinase (CPK), diabetes, ejection fraction, high blood pressure, platelets, sex, serum creatinine, serum sodium, smoking, time, death event. They serve as input variables or features, and the output variable indicating whether the patient with the specified symptoms has heart failure. In this experiment, an output variable, target, with the values of 0 and 1, indicates the absence and presence of heart failure, respectively. Table 2 shows the description for each variable, and Figure 1 shows some examples of the heart disease prediction data used in this study.

age	anaemia	creatinine	diabetes	ejection_fr	high_bloo	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
55	0	748	0	45	0	263358	1.3	137	1	1	88	0
65	0	56	0	25	0	305000	5	130	1	0	207	0
45	0	582	1	38	0	319000	0.9	140	0	0	244	0
60	1	754	1	40	1	328000	1.2	126	1	0	90	0
95	1	582	0	30	0	461000	2	132	1	0	50	1
70	0	232	1	30	0	302000	1.2	132	1	0	210	0
63	1	122	1	60	0	172000	1.2	145	0	0	147	0
70	1	171	0	50	1	358000	0.9	141	0	0	196	0
50	0	482	1	30	0	300000	0.9	132	1	0	109	0

Fig -1:eg. Of dataset for heart failure prediction

Attributes:

1	age	age of the patient (years)
2	anaemia	decrease of red blood cells or hemoglobin (boolean)
3	creatinine phosphokinase (CPK)	level of the CPK enzyme in the blood (mcg/L)
4	diabetes	if the patient has diabetes (boolean)
5	ejection fraction	percentage of blood leaving the heart at each contraction (percentage)
6	high blood pressure	if the patient has hypertension (boolean)
7	platelets	platelets in the blood (kiloplatelets/mL)
8	sex	woman or man (binary)
9	serum creatinine	level of serum creatinine in the blood (mg/dL)
10	serum sodium	level of serum sodium in the blood (mEq/L)
11	smoking	if the patient smokes or not (boolean)
12	time	:follow-up period (days)
13	DEATH_EVENT:	if the patient died during the follow-up period (boolean)

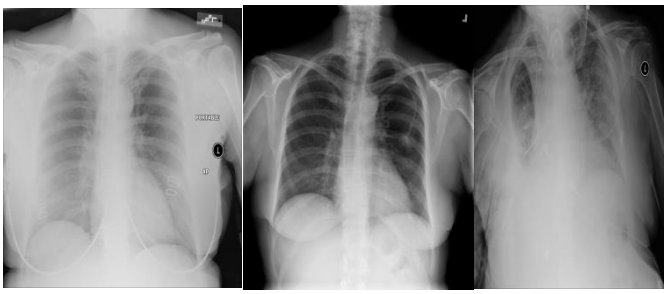


Fig -1:eg. Of chest image for heart failure prediction

3.2 MACHINE LEARNING ALGORITHMS

A. XGBoost

Equation (simplified objective function):

$$\text{Objective} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

XGBoost is a powerful machine learning algorithm based on gradient boosting.It builds an esemble of decision trees , where each new tree corrects th errors of the previous ones.

B. SHAPley

Equation (Shapley Value for a feature i):

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S))$$

SHAP values provide a unified measure of feature importance . They are based on cooperative game theory and can explain the output of any machine learning model.

C. CONVOLUTIONAL NEURAL NETWORKS

Equation (Convolution Operation):

$$\text{Output}_{ij} = \sum_{m} \sum_{n} \text{Input}_{i+m,j+n} \cdot \text{Kernel}_{m,n}$$

CNNs are designed for processing structured grid data like images. They use convolutional layers to automatically and adaptively learn spatial hierarchies of features.

D. SUPPORT VECTOR MACHINES

Equation (Decision Boundary for Linear SVM):

$$f(x) = \text{sign}(w \cdot x + b)$$

SVM is a supervised learning model used for classification and regression. It finds the hyperplane that best separates different classes in the feature space.

E. NEURAL NETWORKS

Equation (Forward Pass of One Layer):

$$a^{[l]} = g(W^{[l]} a^{[l-1]} + b^{[l]})$$

Neural networks are composed of layers of interconnected nodes (neurons), where each connection has an associated weight. They are used for tasks like classification and regression.

3.3 DATA PREPROCESSING

A. Load Data:

Import the dataset and examine the columns for any missing values or errors.

Handle Missing Values:

Use imputation techniques to handle missing values. For numerical columns, use mean or median imputation; for categorical variables, use mode imputation or create an additional "unknown" category.

Feature Engineering:

Standardize or normalize numerical features to ensure they're on a similar scale, which is especially important if using models sensitive to feature scales (e.g., neural networks).Convert categorical variables to numerical values (e.g., using one-hot encoding or label encoding).Create additional features if necessary (e.g., calculate Body Mass Index (BMI) from weight and height if they are available).

Outlier Detection and Removal:

Identify and potentially remove outliers in variables like blood pressure or age if they fall far outside the normal range.

Scaling and Normalization:

Normalize or standardize continuous variables to improve model performance. This can involve scaling between 0 and 1 or using Z-score normalization.

B. Image Data Preprocessing (Chest Images)

For the chest images, common preprocessing techniques for deep learning are used to ensure that each image is in a suitable format and quality for analysis.

Load Images:

Convert each image to grayscale if using single-channel images (or leave in RGB if necessary for the model architecture).Resize each image to a consistent dimension (e.g., 224×224224 \times 224224×224 or 256×256256 \times 256256×256).

Normalization:

Normalize pixel values, often to a range of 0 to 1 or by standardizing them with mean and standard deviation (e.g., mean = 0.5, std = 0.5).

Data Augmentation:

To improve model generalization, apply data augmentation techniques during training, such as rotations, flips, zoom, brightness/contrast adjustments, etc.

Segmentation or ROI (optional):

Use segmentation algorithms or models (like U-Net) to isolate specific areas of interest in the chest images (such as the heart and lungs).

C. Combining Clinical Data and Image Data

The combined data needs to be aligned so each sample has both a clinical vector and an image.

Align Records:

Ensure each image is linked with its corresponding clinical data entry. This may require merging tables based on a common identifier (e.g., patient ID).

Split into Training, Validation, and Test Sets:

Split the data into training, validation, and test sets, maintaining a consistent balance between images and clinical data in each set.

Create Multi-Input Pipeline:

Prepare a data pipeline that can handle both image and tabular data in parallel:For deep learning models, use separate branches in the model: one for the image data (e.g., CNN layers) and one for the clinical data (fully connected layers).Combine outputs from both branches before the final output layer.

D. Model Training

Batch Processing:

Use a batch generator that can load both images and clinical data in parallel for each batch.

Data Transformation Pipeline:

Design a transformation pipeline that includes both image transformations (resizing, normalizing) and clinical data preprocessing (scaling, encoding) so each batch is consistently processed.

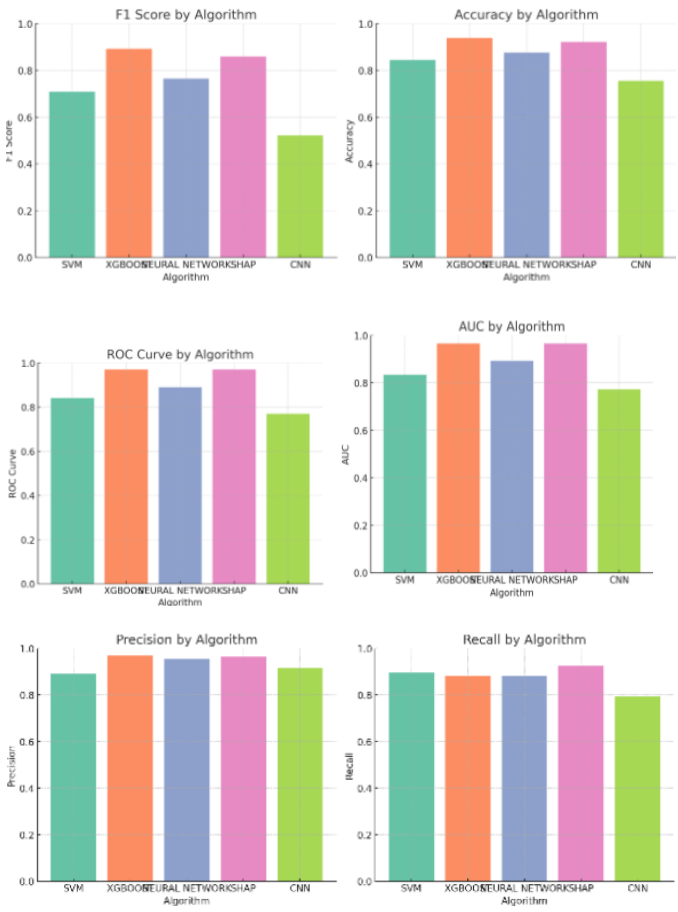
4. MODEL EVALUATION AND COMPARISION

A. Common metrics for binary classification (heart failure prediction) include:

- **Accuracy**: Percentage of correct predictions over the total predictions.
- **Precision**: The proportion of true positives among predicted positives (sensitivity to false positives).
- **Recall (Sensitivity)**: The proportion of true positives among actual positives (sensitivity to false negatives).
- **F1 Score**: Harmonic mean of precision and recall, balancing them.
- **ROC-AUC (Receiver Operating Characteristic - Area Under Curve)**: Measures the trade-off between sensitivity and specificity; an AUC closer to 1 indicates better performance.
- **Confusion Matrix**: A table showing true positive, true negative, false positive, and false negative counts.

B. COMPARATIVE ANALYSIS

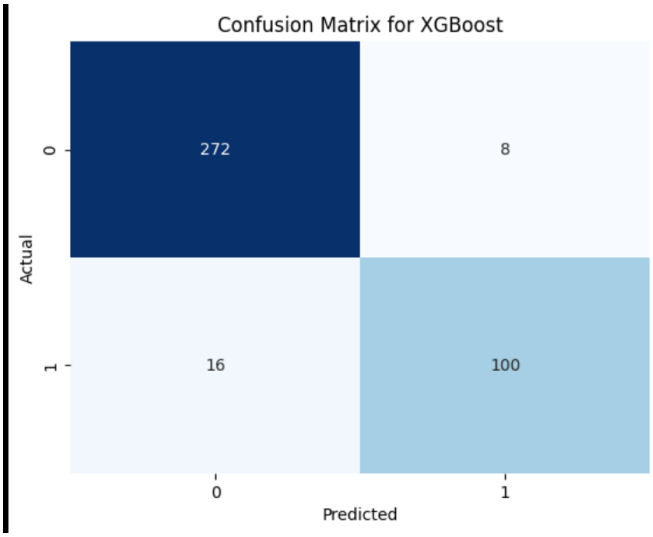
Here are bar charts for each evaluation metric (F1 Score, Accuracy, ROC Curve, AUC, Precision, and Recall) across the different algorithms (SVM, XGBoost, Neural Network, SHAP, and CNN) .These visualizations allow a clear comparison of each model's performance across these metrics.

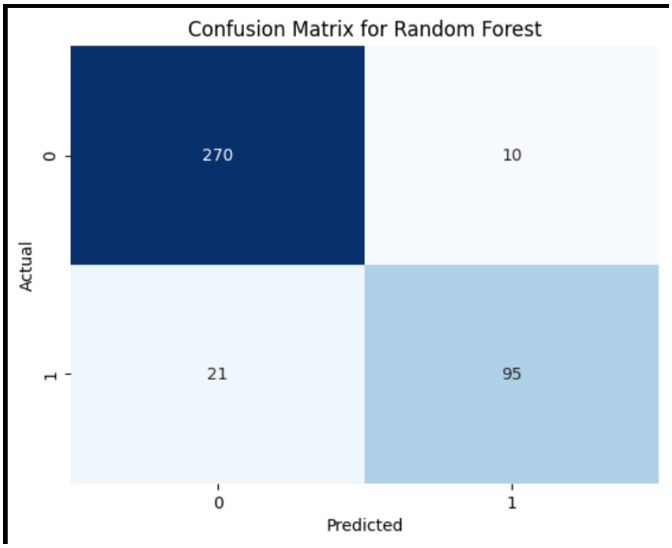


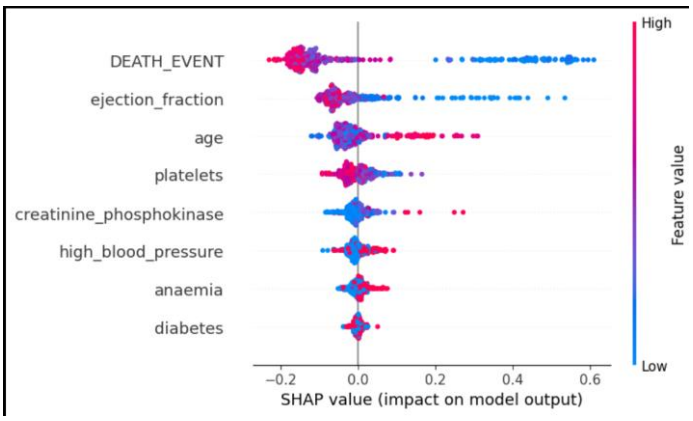
NAME OF ALGORITHM	F1 SCORE	ACCURACY	ROC CURVE	AUC	precision	recall
SVM	0.709	0.844	0.84	0.835	0.892	0.896
XGBOOST	0.892	0.939	0.97	0.966	0.971	0.881
NEURAL NETWORK	0.765	0.876	0.89	0.893	0.954	0.881
SHAP	0.859	0.921	0.97	0.967	0.964	0.928
CNN	0.521	0.755	0.77	0.772	0.918	0.793

XGBoost performed best in F1 Score, Accuracy, Precision, and tied for ROC Curve.**SHAP** performed best in AUC and tied for ROC Curve.**SVM** achieved the highest Recall.

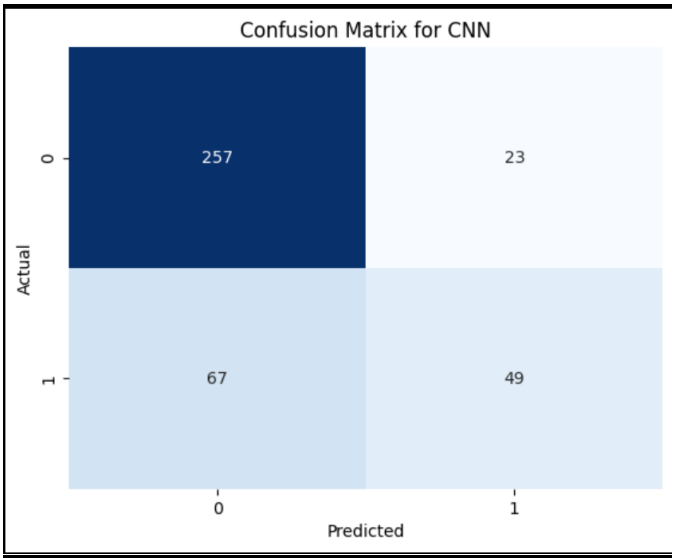
C. CONFUSION MATRICES

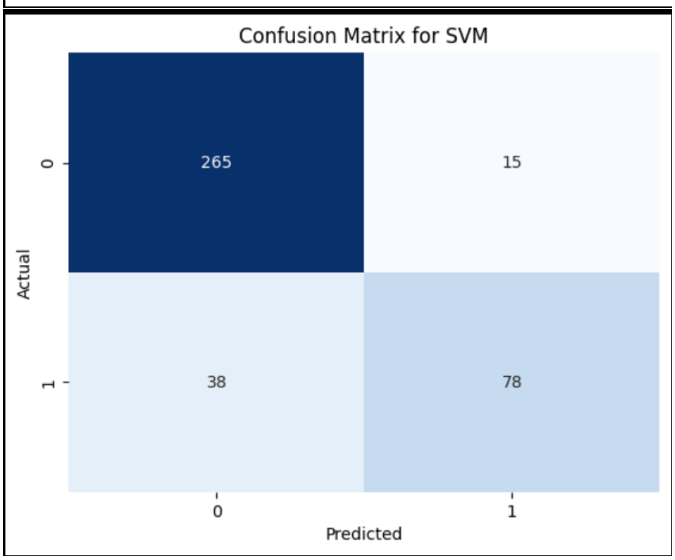


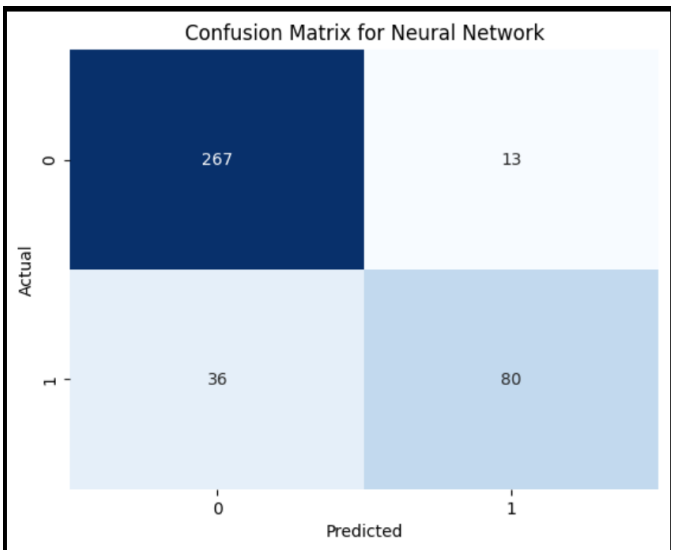




We used a Random Forest Classifier in conjunction with SHAP (Shapley Additive Explanations) to not only achieve strong predictive performance but also to interpret model outputs effectively. SHAP values were applied to the Random Forest model to quantify the contribution of each feature to the prediction outcomes, providing insights into the factors most associated with heart failure risk. This approach allows us to leverage the robustness of the Random Forest algorithm while enhancing interpretability, making it possible to understand and explain the model's decisions in a transparent manner.







The results indicate that **XGBoost** and **SHAP** are the top-performing models for heart failure prediction, with XGBoost excelling in F1 Score, Accuracy, Precision, and achieving a high ROC Curve, while SHAP had the highest AUC and shared the top ROC Curve score. These findings suggest that XGBoost, a robust gradient boosting algorithm, is highly effective for this structured clinical data, offering reliable predictions with high precision and balanced performance across metrics. SHAP's strong performance, particularly in interpretability (via Shapley values), not only provides excellent prediction accuracy but also valuable insights into feature importance, making it a good choice for models where interpretability is crucial, such as in medical research. Meanwhile, **SVM** demonstrated the highest Recall, indicating it is particularly effective at identifying true positives (patients at risk of heart failure), which is vital in clinical applications where missing a diagnosis could have serious consequences. Overall, XGBoost is highly competitive for predictive accuracy, while SHAP adds interpretability, and SVM offers value in recall-focused applications.

5. CONCLUSIONS

In this study, we conclude that combining clinical and imaging data with advanced machine learning models offers a powerful approach for predicting heart failure risk. Among the models tested, **XGBoost** and **SHAP-enhanced models** demonstrated the highest predictive performance, with XGBoost excelling in accuracy, precision, and F1 Score, while SHAP provided valuable interpretability by highlighting the most influential features contributing to heart failure risk. **SVM** achieved the highest recall, suggesting it is particularly suited for applications where identifying true positives is crucial, even at the cost of some false positives.The combination of predictive accuracy and interpretability is especially important in healthcare, where understanding why a model makes certain predictions is as valuable as the predictions themselves. SHAP's interpretability could aid healthcare professionals in identifying key clinical indicators, allowing for more transparent, data-driven decisions in patient care.Overall, this study suggests that **XGBoost** and **SHAP-enhanced Random Forest models** are highly effective for heart failure prediction, balancing performance and interpretability. The findings encourage further research into integrating machine learning with clinical insights, aiming for predictive tools that are both accurate and actionable in real-world healthcare settings.

REFERENCES

1.https://pubmed.ncbi.nlm.nih.gov/30810291/

2.https://pubmed.ncbi.nlm.nih.gov/38969836/

3.https://www.kaggle.com/datasets/nih-chest-xrays/data

4.https://www.kaggle.com/datasets/aadarshvelu/heart-failure-prediction-clinical-records

5.https://www.sciencedirect.com/science/article/abs/pii/S0002870320302155