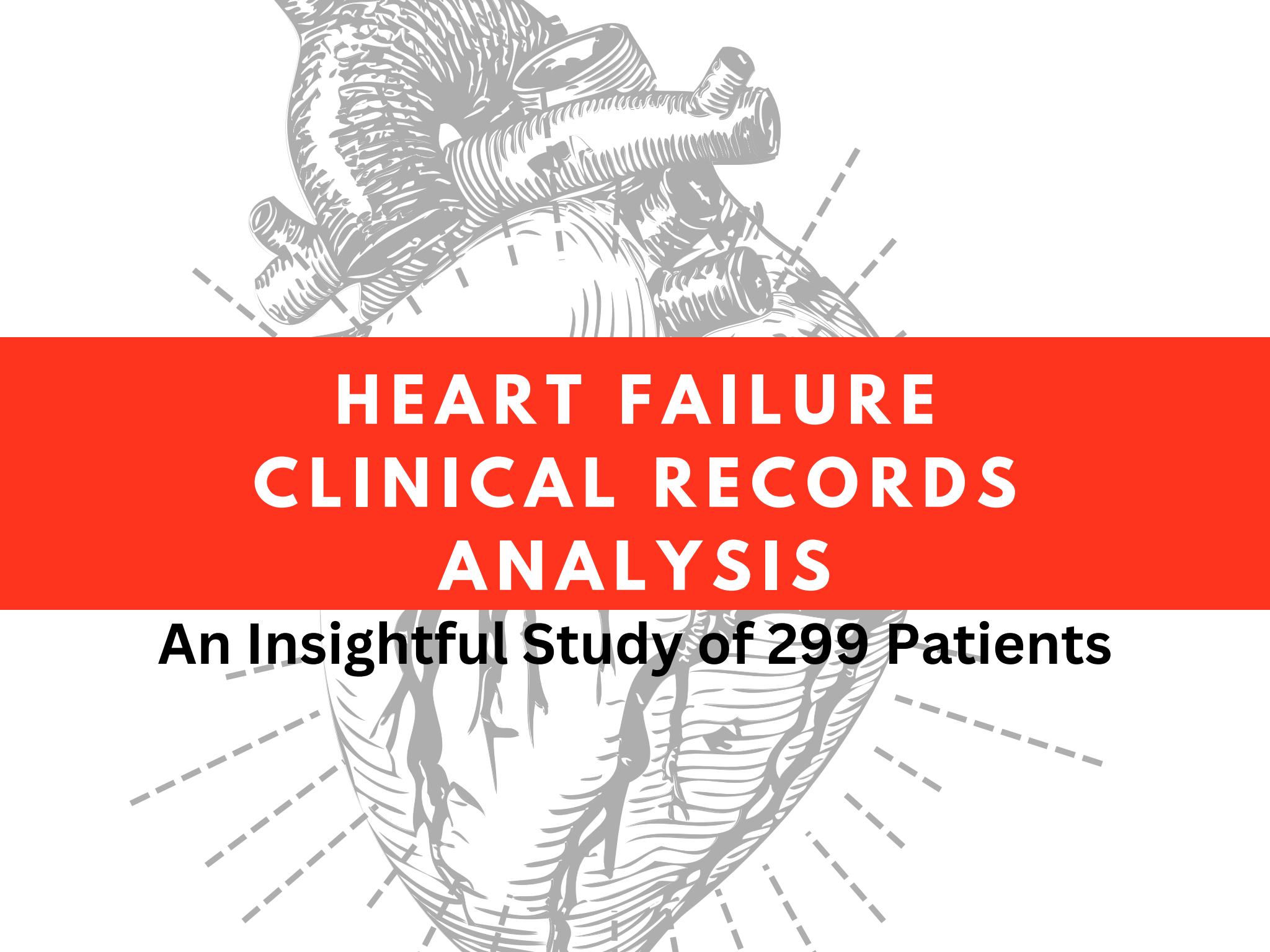


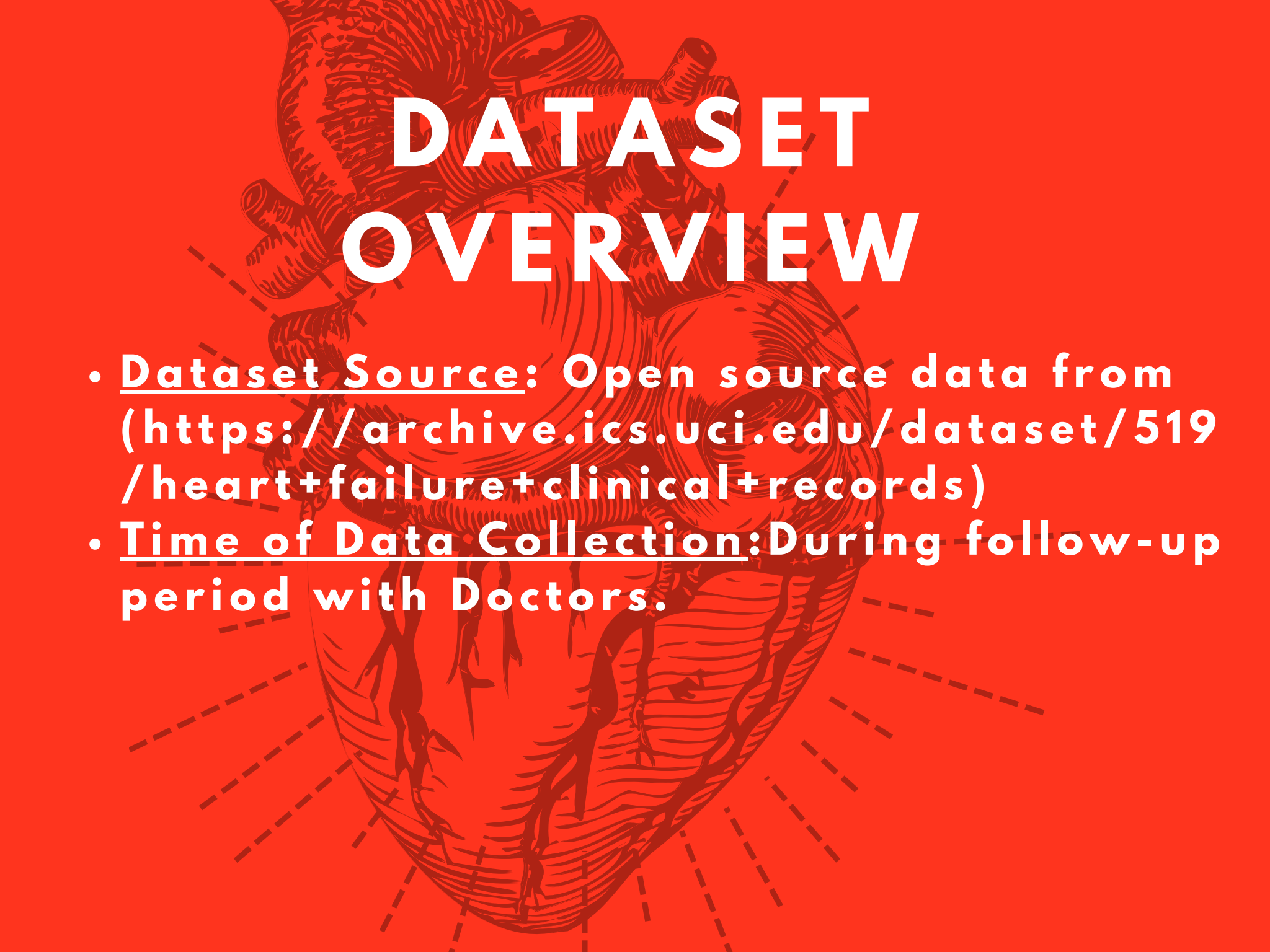
HEART FAILURE CLINICAL RECORDS ANALYSIS

An Insightful Study of 299 Patients

HEART DISEASE

Heart disease refers to a range of conditions that affect the heart's structure and function. These conditions can lead to various cardiovascular issues, including heart attacks, heart failure, and arrhythmias.

The term "heart disease" is often used interchangeably with "cardiovascular disease," which encompasses disorders of both the heart and blood vessels.
as a heart attack)



DATASET OVERVIEW

- Dataset Source: Open source data from (https://archive.ics.uci.edu/dataset/519/heart+failure+clinical+records)
- Time of Data Collection:During follow-up period with Doctors.

CLINICAL FEATURES

Fourteen (14) clinical features:

- Age
- Anaemia
- Creatinine Phosphokinase (CPK)
- Diabetes
- Ejection Fraction
- High Blood Pressure
- Platelets
- Sex
- Serum Creatinine
- Serum Sodium
- Smoking
- Time (follow-up period)
- Death Event (target)

CLINICAL FEATURES

Fourteen (14) clinical features:

- AGE: AGE OF THE PATIENT (YEARS)
- ANAEMIA: DECREASE OF RED BLOOD CELLS OR HEMOGLOBIN (BOOLEAN)
- CREATININE PHOSPHOKINASE (CPK): LEVEL OF THE CPK ENZYME IN THE BLOOD (MCG/L)
- DIABETES: IF THE PATIENT HAS DIABETES (BOOLEAN)
- EJECTION FRACTION: PERCENTAGE OF BLOOD LEAVING THE HEART AT EACH CONTRACTION (PERCENTAGE)
- HIGH BLOOD PRESSURE: IF THE PATIENT HAS HYPERTENSION (BOOLEAN)
- PLATELETS: PLATELETS IN THE BLOOD (KILOPLATELETS/ML)
- SEX: WOMAN OR MAN (BINARY)
- SERUM CREATININE: LEVEL OF SERUM CREATININE IN THE BLOOD (MG/DL)
- SERUM SODIUM: LEVEL OF SERUM SODIUM IN THE BLOOD (MEQ/L)
- SMOKING: IF THE PATIENT SMOKES OR NOT (BOOLEAN)
- TIME: FOLLOW-UP PERIOD (DAYS)
- [TARGET] DEATH EVENT: IF THE PATIENT DIED DURING THE FOLLOW-UP PERIOD (BOOLEAN)

DATA IMPORT AND PREPARATION

- imported the dataset using CSV files
- Performed initial data cleaning and preprocessing

```
# Create a SparkSession
spark = SparkSession.builder\
    .appName("SparkSQL")\
    .config("spark.sql.debug.maxToStringFields", 2000)\
    .config("spark.driver.memory", "2g")\
    .getOrCreate()


heart_df = spark.read.csv("heart_failure_clinical_records_dataset.csv", sep=",", header=True)
heart_df.show()
```

```
+---+-------+-----------------------+--------+----------------+------------------+---------+----------------+------------+---+-------+----+-----------
+
|age|anaemia|creatinine_phosphokinase|diabetes|ejection_fraction|high_blood_pressure|platelets|serum_creatinine|serum_sodium|sex|smoking|time|DEATH_EVENT
|
+---+-------+-----------------------+--------+----------------+------------------+---------+----------------+------------+---+-------+----+-----------
+
| 75|      0|                    582|       0|              20|                 1|   265000|             1.9|         130|  1|      0|   4|          1
|
| 55|      0|                   7861|       0|              38|                 0|263358.03|             1.1|         136|  1|      0|   6|          1
|
| 65|      0|                    146|       0|              20|                 0|   162000|             1.3|         129|  1|      1|   7|          1
|
| 50|      1|                    111|       0|              20|                 0|   210000|             1.9|         137|  1|      0|   7|          1
|
| 65|      1|                    160|       1|              20|                 0|   327000|             2.7|         116|  0|      0|   8|          1
|
| 90|      1|                     47|       0|              40|                 1|   204000|             2.1|         132|  1|      1|   8|          1
|
| 75|      1|                    246|       0|              15|                 0|   127000|             1.2|         137|  1|      0|  10|          1
|
| 60|      1|                    315|       1|              60|                 0|   454000|             1.1|         131|  1|      1|  10|          1
```

TRAIN-TEST SPLIT

- **Utilized scikit-learn to split data into training and testing sets**
- **Ensured balanced distribution of target variable (death event)**

```
LogisticRegression(max_iter=200, random_state=1)
```

Out[8]:

```
LogisticRegression(max_iter=200, random_state=1)
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [9]:

```
print(f"Training Data Score: {classifier.score(X_train, y_train)}")
print(f"Testing Data Score: {classifier.score(X_test, y_test)}")
```

```
Training Data Score: 0.8348214285714286
Testing Data Score: 0.7733333333333333
```

In [10]:

```
predictions = classifier.predict(X_test)
results = pd.DataFrame({"Prediction": predictions, "Actual": y_test}).reset_i
results.head(10)
```

Out[10]:

RANDOM FOREST CLASSIFIER

- **Accuracy Score: .80**

```
Accuracy Score : 0.8
Classification Report
              precision    recall  f1-score   support

           0       0.85      0.86      0.85        51
           1       0.70      0.67      0.68        24

    accuracy                           0.80        75
   macro avg       0.77      0.76      0.77        75
weighted avg       0.80      0.80      0.80        75


In [18]:
```

LOGISTIC REGRESSION

- **Accuracy Score: 77.3%**

```
from sklearn.metrics import accuracy_score
# Display the accuracy score for the test dataset.
accuracy_score(y_test, predictions)
```

Out[11]:

```
0.7733333333333333
```

In [12]:

```
 from sklearn.metrics import classification_report
target_names = ["Death Event", "Comorbidity"]
print(classification_report(y_test, predictions, target_names=target_names))
```

```
              precision    recall  f1-score   support

 Death Event       0.83      0.84      0.83        51
 Comorbidity       0.65      0.62      0.64        24

    accuracy                           0.77        75
   macro avg       0.74      0.73      0.74        75
weighted avg       0.77      0.77      0.77        75
```

In [13]:

KEY FINDINGS

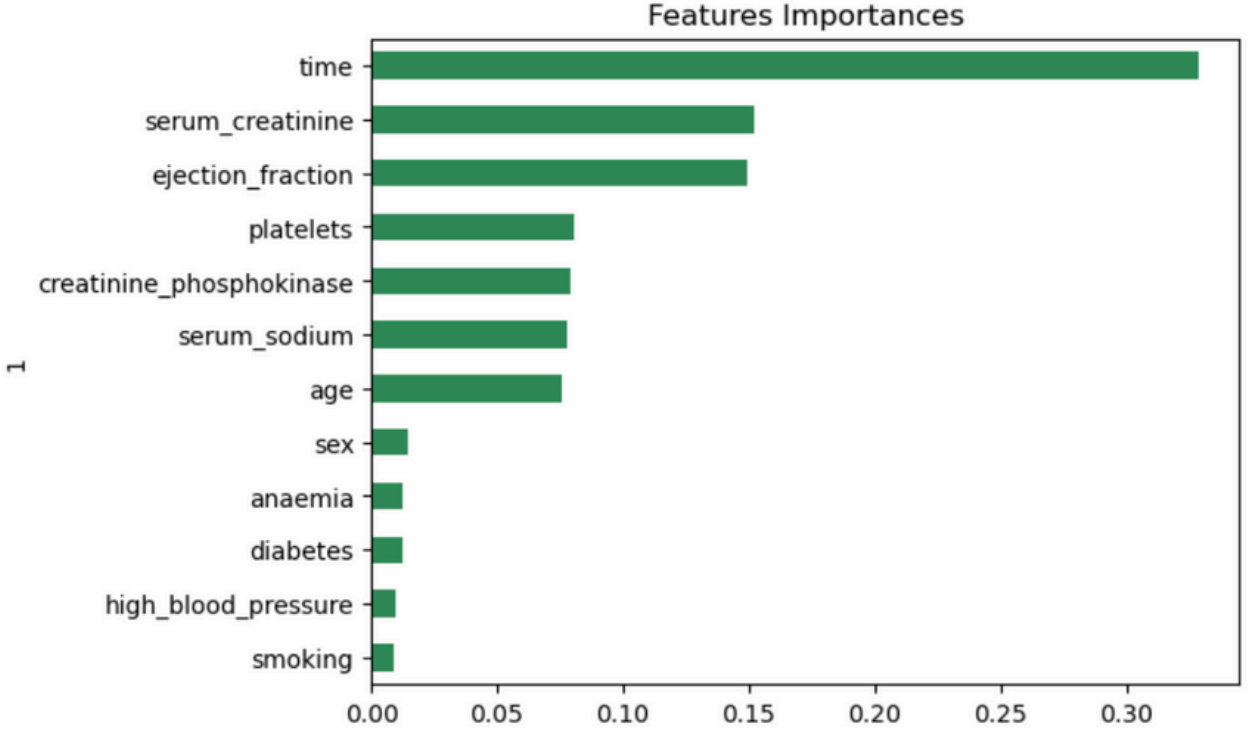
Model Performance:

- **Achieved an accuracy score of 77.3% with the logistic regression model and 80% with Random Forest Classifier**

- **High precision and recall for the "Death Event" class indicate the model's effectiveness in identifying high-risk patients.**

FEATURES BY IMPORTANCE

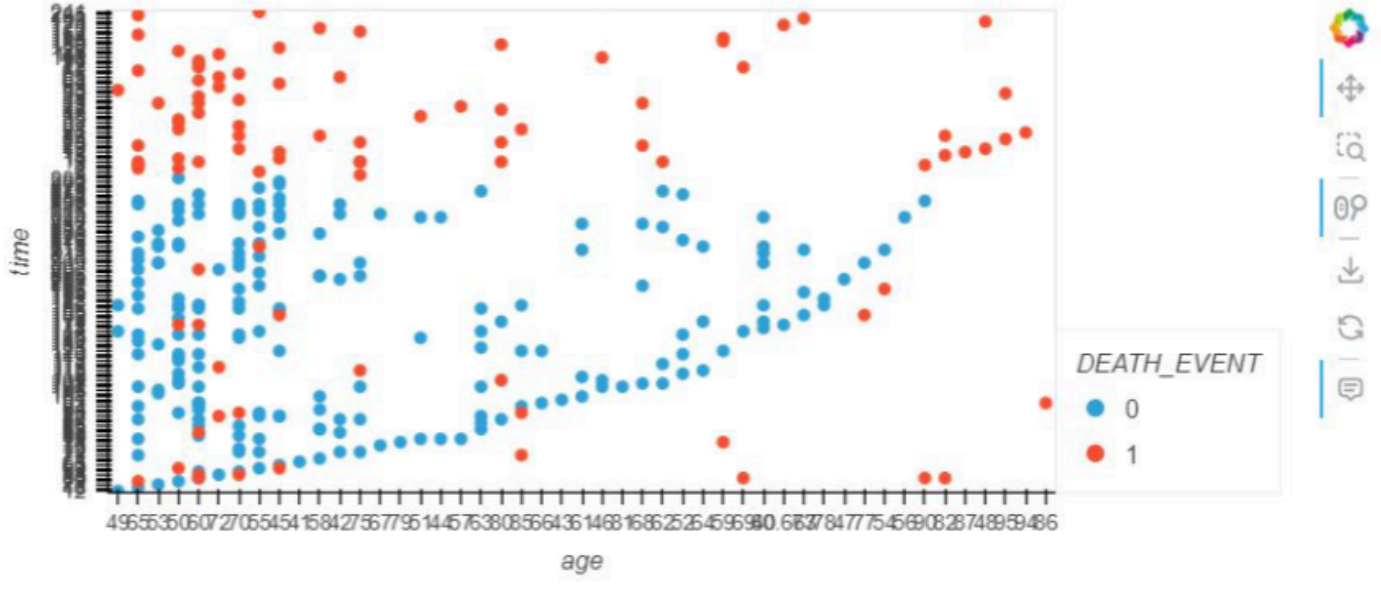
- **Analyzed the relationship between various clinical features and the death event**
- **Key features: Platelets, Serum Creatinine, CPK, Serum Sodium, Ejection Fraction, Time, Age**



FEATURES BY IMPORTANCE: TIME (0.328)

Significance: The follow-up period (time) is the most significant feature in predicting the death event. This indicates that the duration of monitoring has a strong impact on patient outcomes.

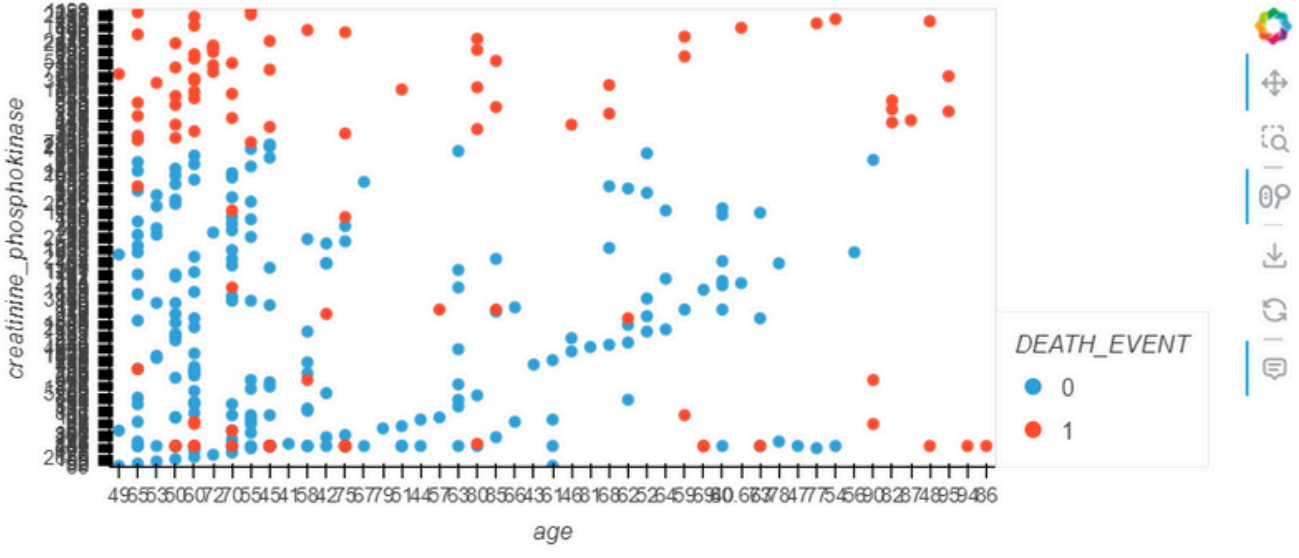
```
heart_pd.hvplot.scatter(x="age", y="time", by='DEATH_EVENT')
```



FEATURES BY IMPORTANCE: CREATININE PHOSPHOKINASE (0.079):

Significance: The level of the CPK enzyme, which can indicate muscle damage (including heart muscle), is an important predictor. Elevated CPK levels are often seen after heart attacks or other muscle injuries.

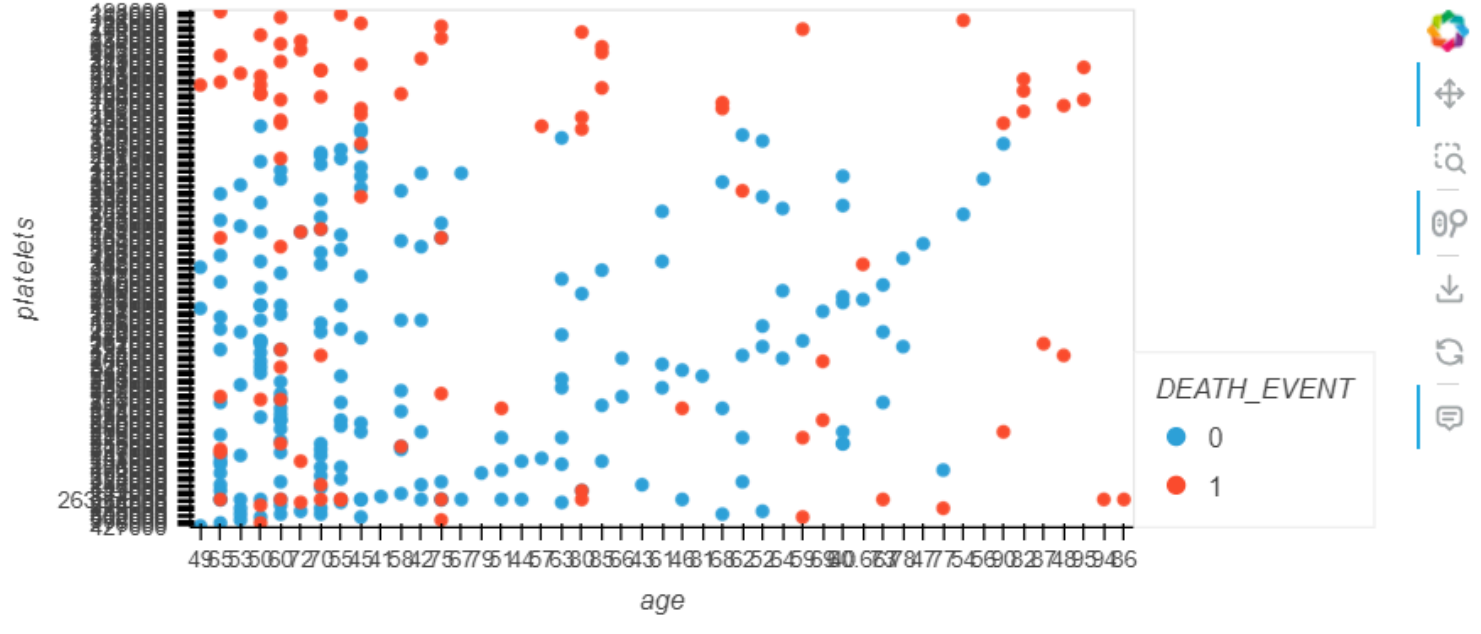
```
#Scatter plot for age vs creatinine_phosphokinase(help diagnose and monitor various conditions related to muscle damage, heart conditions, and o
heart_pd.hvplot.scatter(x="age", y="creatinine_phosphokinase", by='DEATH_EVENT')
```



FEATURES BY IMPORTANCE: PLATELETS (0.080):

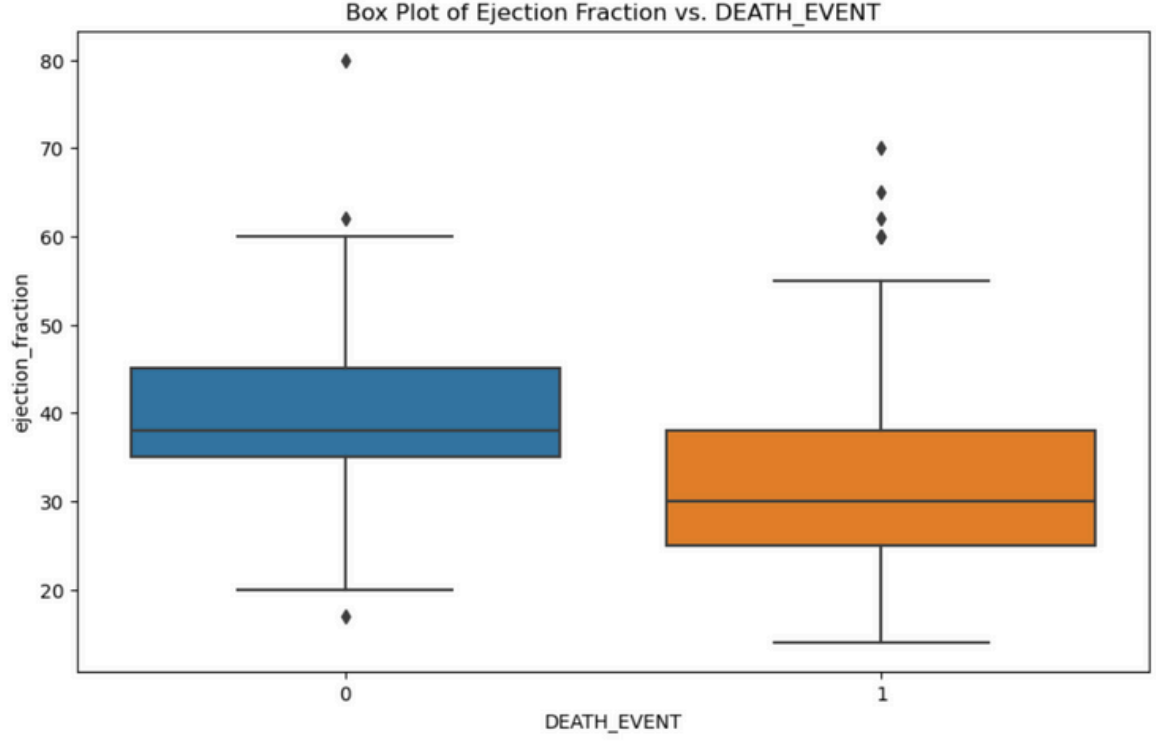
- **Significance: Platelet count in the blood, which affects clotting, also plays an important role. Abnormal platelet levels can indicate various health issues affecting heart failure outcomes.**

```
#Scatter plot for age vs platelets(small blood cells that help stop bleeding by forming clots)
heart_pd.hvplot.scatter(x="age", y="platelets", by='DEATH_EVENT')
```



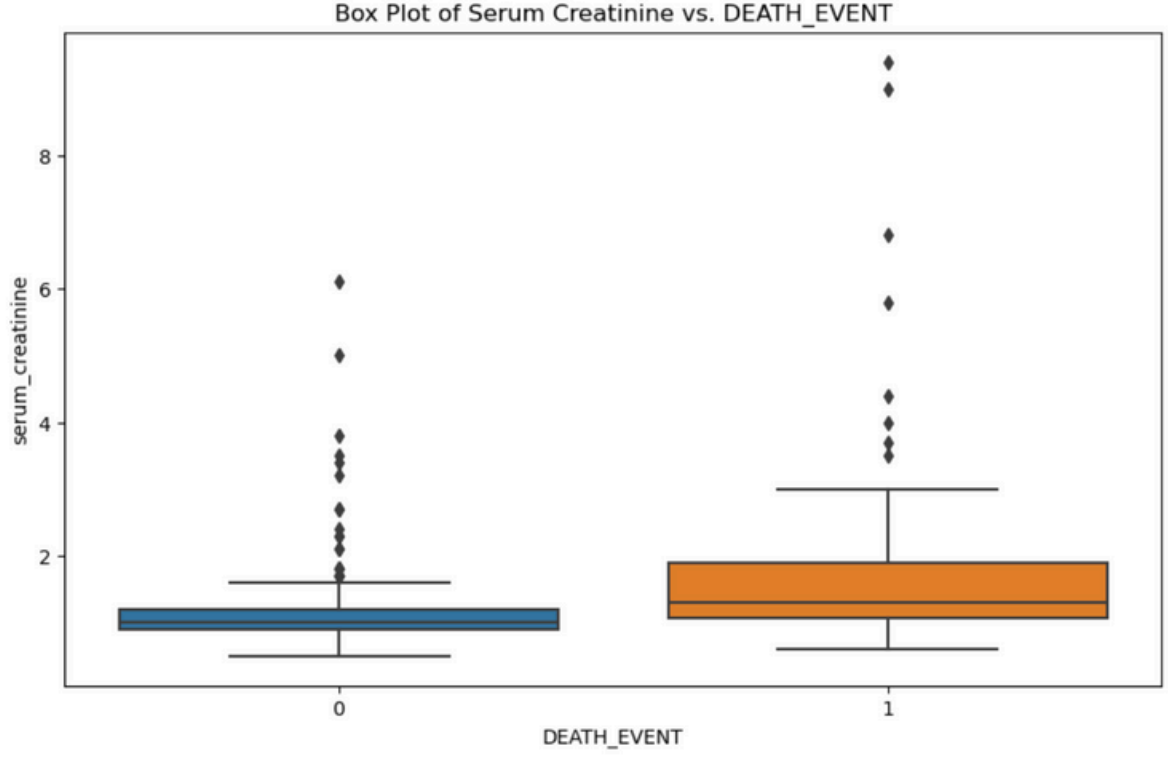
FEATURES BY IMPORTANCE: EJECTION FRACTION (0.149):

Significance: The percentage of blood leaving the heart at each contraction is crucial. Lower ejection fraction values indicate worse heart function, making it a key predictor of mortality.



FEATURES BY IMPORTANCE: SERUM CREATININE (0.152):

Significance: High levels of serum creatinine, which indicate kidney function, are highly predictive of heart failure outcomes. Kidney function is a critical factor in the health of heart failure patients.



KEY FINDINGS

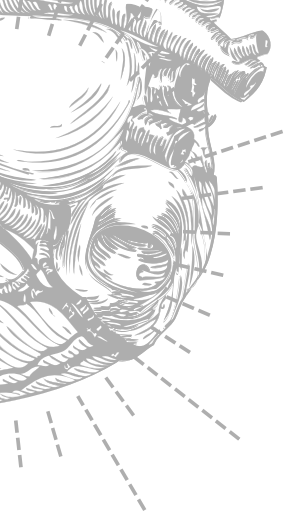
Important Predictive Features:

- **Time: The follow-up period is the most significant predictor, emphasizing the importance of monitoring duration.**
- **Serum Creatinine and Ejection Fraction: Indicators of kidney function and heart efficiency are critical in predicting mortality.**
- **Platelets, Creatinine Phosphokinase, and Serum Sodium: Blood-related metrics and enzyme levels also play significant roles.**

KEY FINDINGS

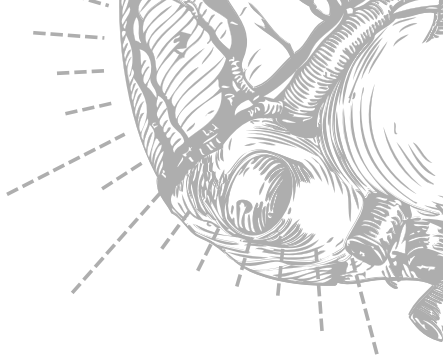
Comorbidity Analysis:

- **Certain comorbidities, such as anaemia and diabetes, although less impactful individually, still contribute to the overall risk profile of heart failure patients.**



82%

of people who die of coronary heart disease are 65 and older. The risk of stroke double every decade after age 55.



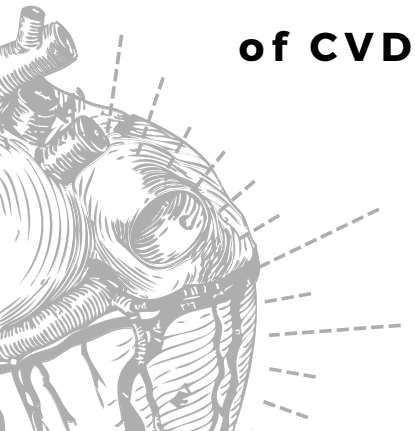
CARDIOVASCULAR DISEASES ARE THE LEADING CAUSE OF DEATH GLOBALLY

80%

of CVD deaths in MALES

75%

of CVD deaths in FEMALES



IMPLICATIONS FOR CLINICAL PRACTICE:

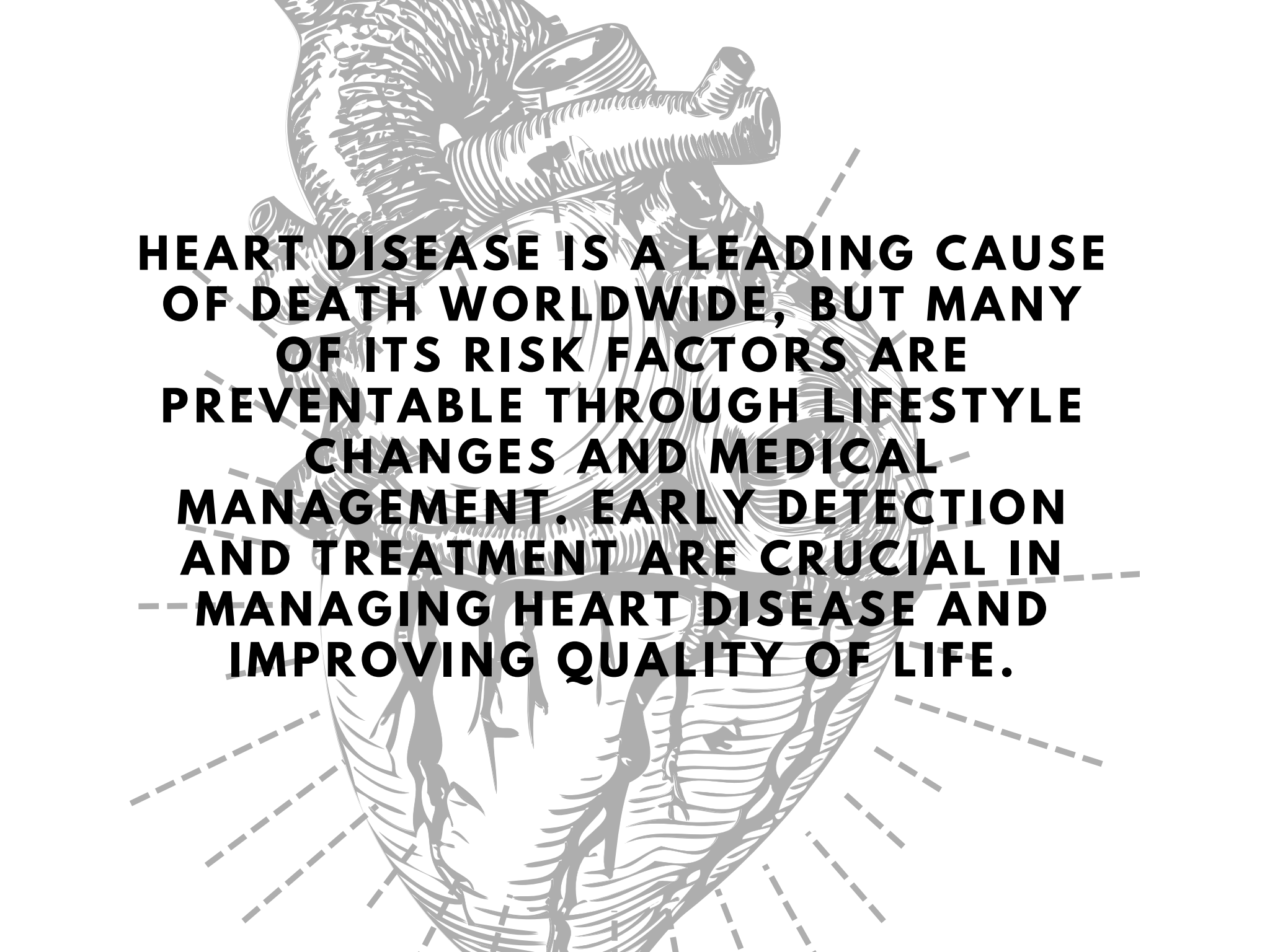
1. **Enhanced Monitoring:**
   - **Prioritize longer follow-up periods for heart failure patients to better manage and predict adverse outcomes.**
   - **Regular monitoring of serum creatinine and ejection fraction levels to assess and manage patient risk effectively.**
2. **Targeted Interventions:**
   - **Develop personalized treatment plans based on the most impactful features, such as kidney function and heart efficiency.**
   - **Implement interventions for patients with significant blood-related anomalies (e.g., platelet count, CPK levels).**
3. **Holistic Patient Management:**
   - **Address comorbid conditions like anaemia and diabetes as part of comprehensive heart failure management to improve overall patient outcomes.**

FUTURE RESEARCH DIRECTIONS:

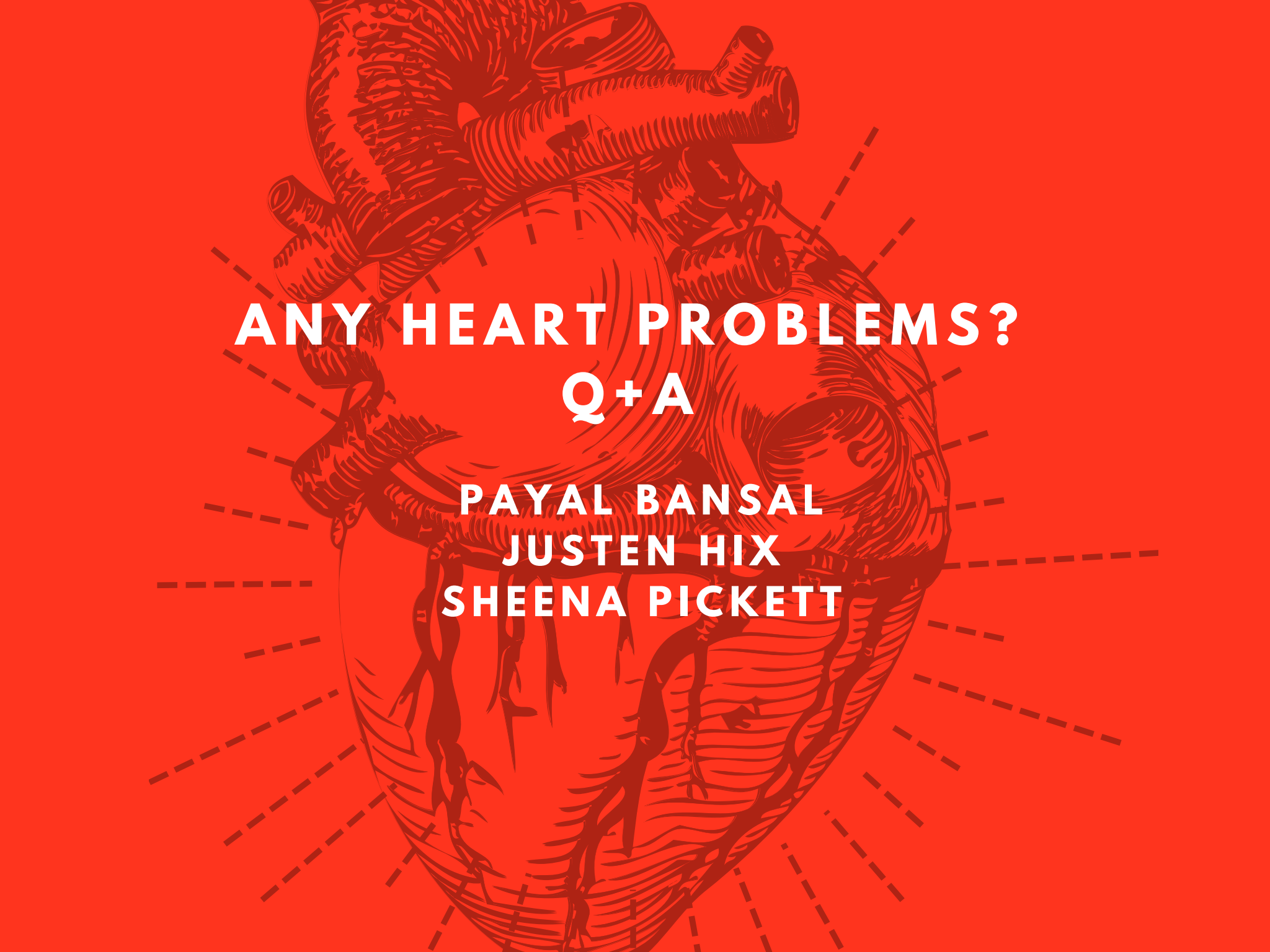
1. **Model Improvement:**
- **Explore advanced machine learning models to improve prediction accuracy and incorporate more complex interactions between features.**

2. **Feature Exploration:**
- **Investigate additional clinical features and biomarkers that could enhance predictive power.**
- **Study the impact of lifestyle factors and medication adherence on heart failure outcomes.**

3. **Longitudinal Studies:**
- **Conduct long-term studies to validate the findings and refine predictive models.**
- **Analyze the effectiveness of targeted interventions based on model predictions in real-world clinical settings.**



HEART DISEASE IS A LEADING CAUSE OF DEATH WORLDWIDE, BUT MANY OF ITS RISK FACTORS ARE PREVENTABLE THROUGH LIFESTYLE CHANGES AND MEDICAL MANAGEMENT. EARLY DETECTION AND TREATMENT ARE CRUCIAL IN MANAGING HEART DISEASE AND IMPROVING QUALITY OF LIFE.



ANY HEART PROBLEMS?
Q+A

PAYAL BANSAL
JUSTEN HIX
SHEENA PICKETT


CONGRATULATIONS
ON YOUR
ACHIEVEMENT!