Heart Failure Mortality Prediction

Machine Learning Analysis for Clinical Decision Support

299
Patient Records

81.67%
Accuracy

86.14%
AUC Score

August 2025

Project Overview

Primary Goal

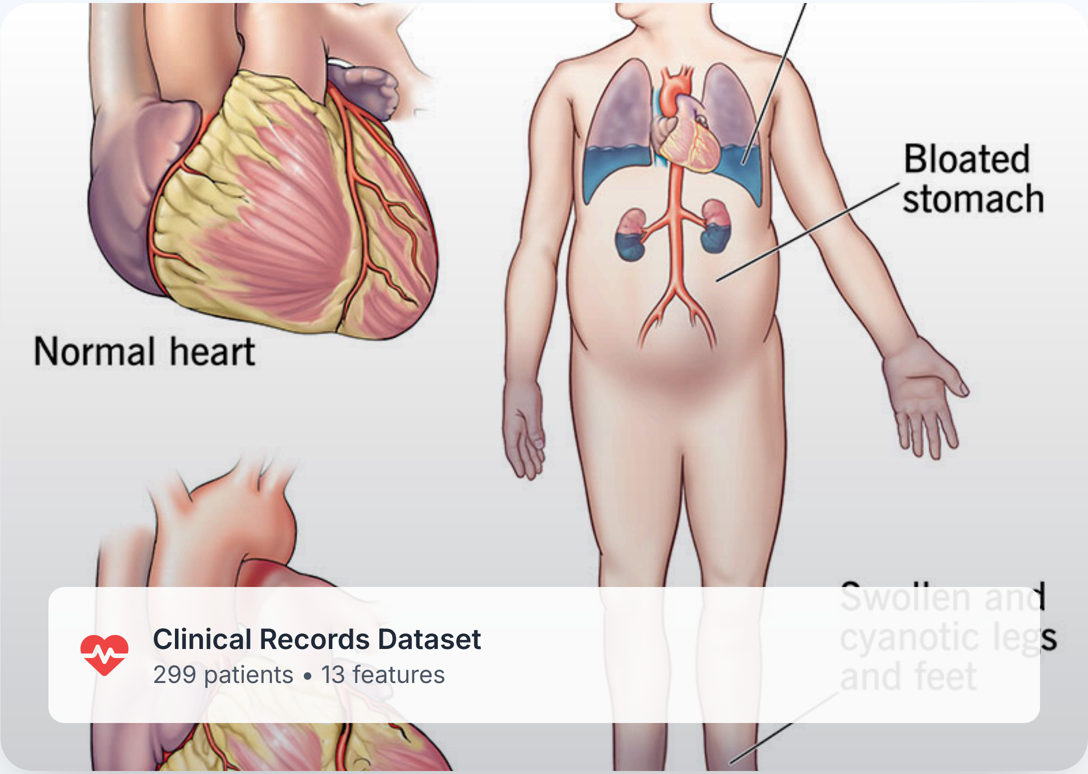
Develop a predictive model to assess heart failure mortality risk using clinical patient data

Clinical Data Analysis

Utilizing comprehensive patient records including demographics, lab results, and medical history

Machine Learning Approach

Implementing logistic regression to identify key predictive factors for mortality outcomes



Clinical Records Dataset
299 patients • 13 features

Dataset Overview

Clinical Data Analysis

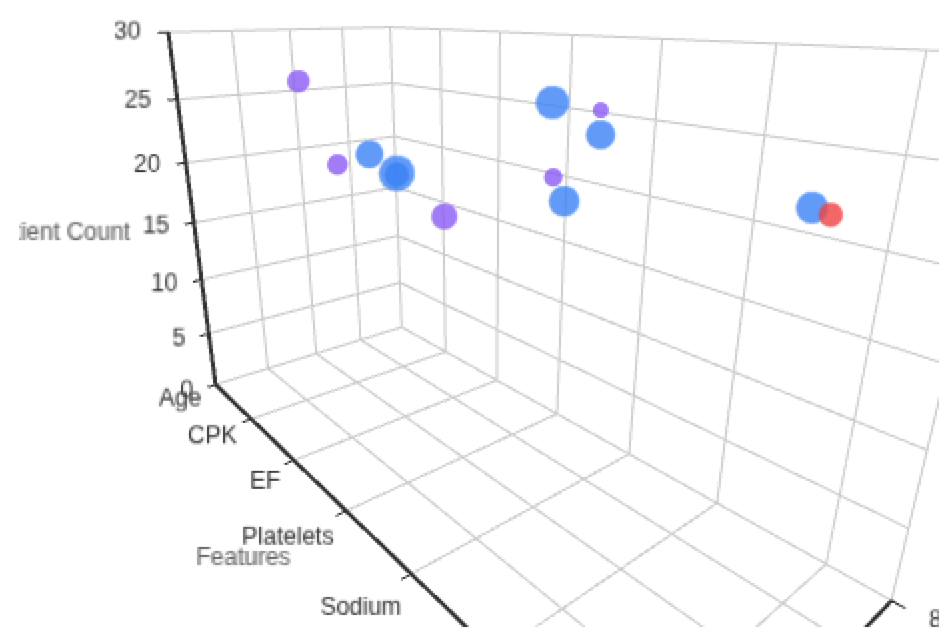
Key Statistics

Total Patients	299
Total Features	13
Data Quality	100% Complete

Feature Categories

- Age & Demographics
- Clinical Measurements
- Binary Health Conditions
- Target Variable

3D Feature Distribution



7
Numerical Features

5
Binary Features

Data Preprocessing

Preparing data for machine learning analysis

Feature Scaling

StandardScaler normalization applied to numerical features:

```
numerical_features = ['age',
'creatinine_phosphokinase',
'ejection_fraction', 'platelets', 'serum_creatinine',
'serum_sodium', 'time']
```

Train-Test Split

Data split with stratification:

80%
Training

20%
Testing

Data Processing Pipeline

Raw Dataset (299 patients, 13 features)

↓

StandardScaler Normalization

↓

Train-Test Split (239/60 samples)

Dataset Shapes

`X_train: (239, 12)` `X_test: (60, 12)`

Machine Learning
Model Implementation

Logistic Regression

- Binary Classification Algorithm
- Predicts Mortality Probability (0-1)
- Sigmoid Function for Probability Mapping

```
model = LogisticRegression()
model.fit(X_train, y_train)
```

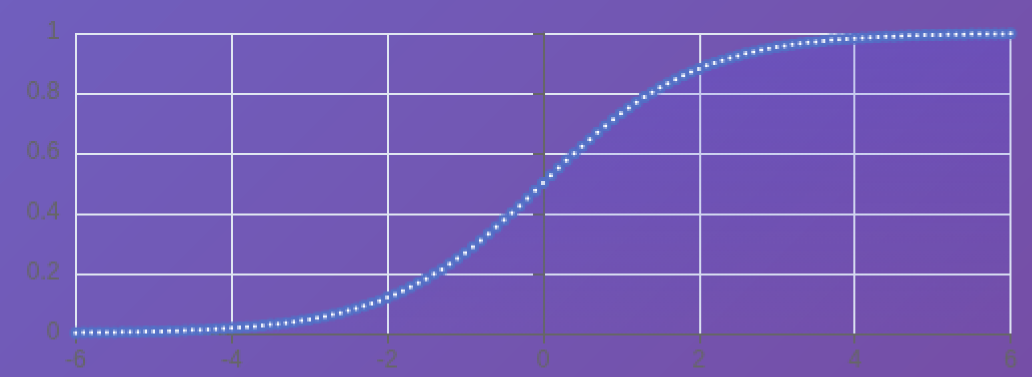
Algorithm Workflow

Input Features
12 Clinical Variables

Linear Combination
$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{12} x_{12}$

Sigmoid Function
$P = 1 / (1 + e^{-z})$

Mortality Prediction
Death Event: 0 or 1



Model Performance Results

Logistic Regression Classification Metrics

Accuracy
81.67%

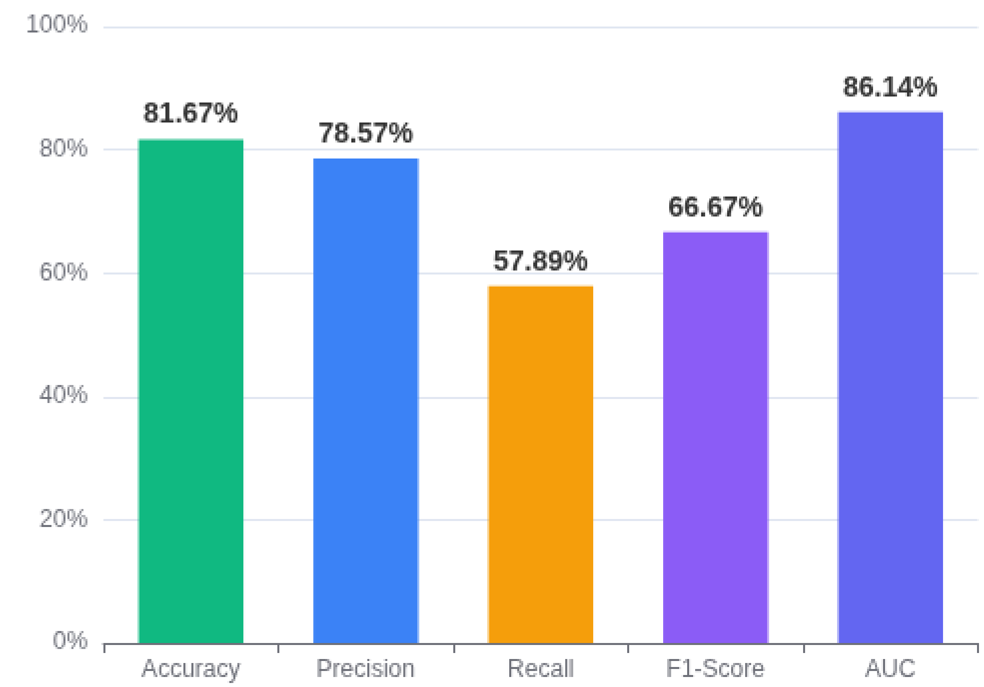
Precision
78.57%

Recall
57.89%

F1-Score
66.67%

AUC Score
86.14%

Performance Metrics Overview



Confusion Matrix

True Negative **38**	False Positive **3**
False Negative **8**	True Positive **11**

Confusion Matrix Analysis

- **True Negatives:** 38 patients correctly predicted as survivors
- **False Positives:** 3 patients incorrectly predicted as deceased
- **False Negatives:** 8 patients incorrectly predicted as survivors
- **True Positives:** 11 patients correctly predicted as deceased

Model Insight: High specificity (92.7%) indicates excellent ability to identify survivors, while sensitivity (57.9%) suggests room for improvement in detecting mortality risk.

Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	38 True Negative	3 False Positive
Actual Positive	8 False Negative	11 True Positive

Sensitivity (Recall)
57.9%

Specificity
92.7%

Key Conclusions

Strong Predictive Performance

AUC score of **86.14%** demonstrates excellent model discrimination capability

Balanced Accuracy

Overall accuracy of **81.67%** with good precision-recall balance

Clinical Relevance

Model successfully identifies high-risk patients for early intervention

Future Directions

Advanced Algorithms

Explore ensemble methods, neural networks, and gradient boosting for improved performance

Feature Engineering

Incorporate additional clinical markers and patient history data

Clinical Validation

Validate model with larger, multi-center datasets for broader applicability

Clinical Integration

Develop user-friendly interface for real-time clinical decision support

Successful ML Model for Heart Failure Mortality Prediction