Tutorial for Heart Failure Mortality Prediction Using MIMIC-IV

JESUS MINJARES

UT AUSTIN: AI IN HEALTHCARE

Assignment Goals:

- Apply ML/DL to EHR data for risk management.
- Extract features, fit models, derive insights.

Objective: Predict in-hospital mortality in heart failure patients.

Dataset: MIMIC-IV (v3.1), 42,990 admissions (34,392 train, 8,598 test).

Models: XGBoost, Advanced FNN, Ensemble (XGBoost + FNN).

Task: Binary classification (mortality vs. non-mortality).

Mortality Rate: 5.40% (class imbalance).

Introduction

Data Preprocessing and Feature Extraction

Data Loading:

- Loaded MIMIC-IV: admissions, patients, diagnoses_icd, icustays.
- Filtered for heart failure (ICD-10: I50.x).

Feature Extraction:

- Demographics: Age (capped at 90), gender (M=0, F=1).
- Clinical: ICU stay (binary), admission type (categorical).
- Comorbidities: Top ICD-10 codes (one-hot encoded).
- Target: mortality (1 if deathtime not null, 0 otherwise).

Preprocessing: 80/20 train-test split, scaled features.

ML Model - XGBoost

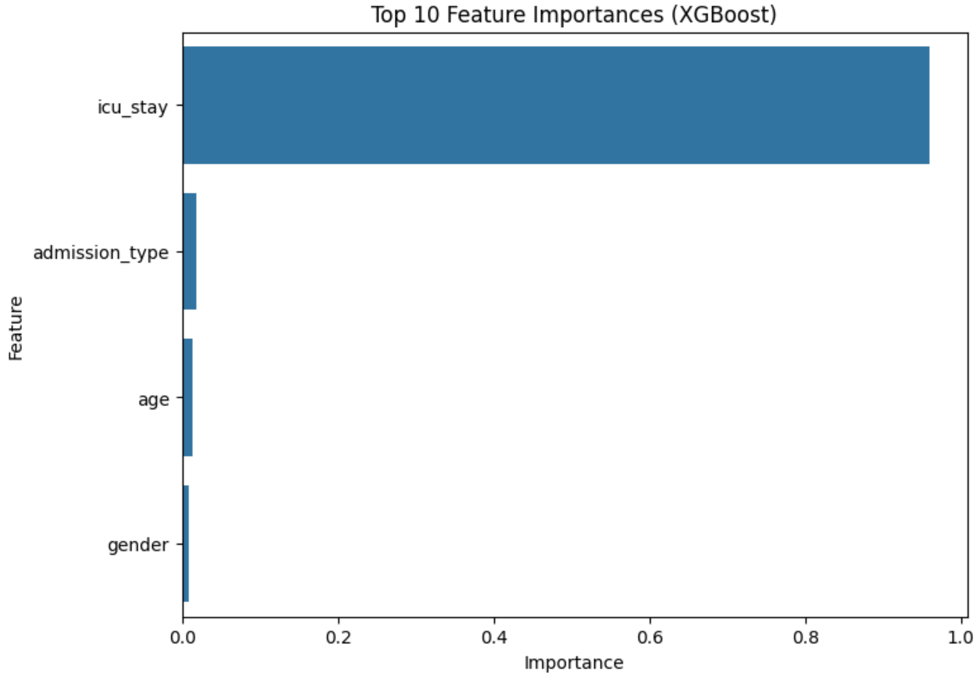
Model: XGBoost (gradient boosting classifier).

Setup: eval_metric='logloss', trained on 34,392 samples.

Feature Fitting: Used scaled features (age, gender, ICU stay, etc.).

Evaluation: Predicted on 8,598 test samples, computed metrics.

Feature Importance: Top features (e.g., ICU stay, age).




```
XGBoost Metrics (Full Dataset)
Accuracy: 0.9457
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
AUC-ROC: 0.8392
```

DL Model - Advanced Feedforward Neural Network (FNN)

```
Advanced Feedforward Neural Network Metrics (with Threshold Tuning)
Accuracy: 0.7534
Precision: 0.1601
Recall: 0.8405
F1 Score: 0.2690
AUC-ROC: 0.8408

FNN training and evaluation completed in 20.66 seconds.

Ensemble Metrics
Accuracy: 0.9458
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
AUC-ROC: 0.8430
```

Model: Advanced FNN with residual connections.

Architecture: input_dim→256→128→64→32→1, focal loss.

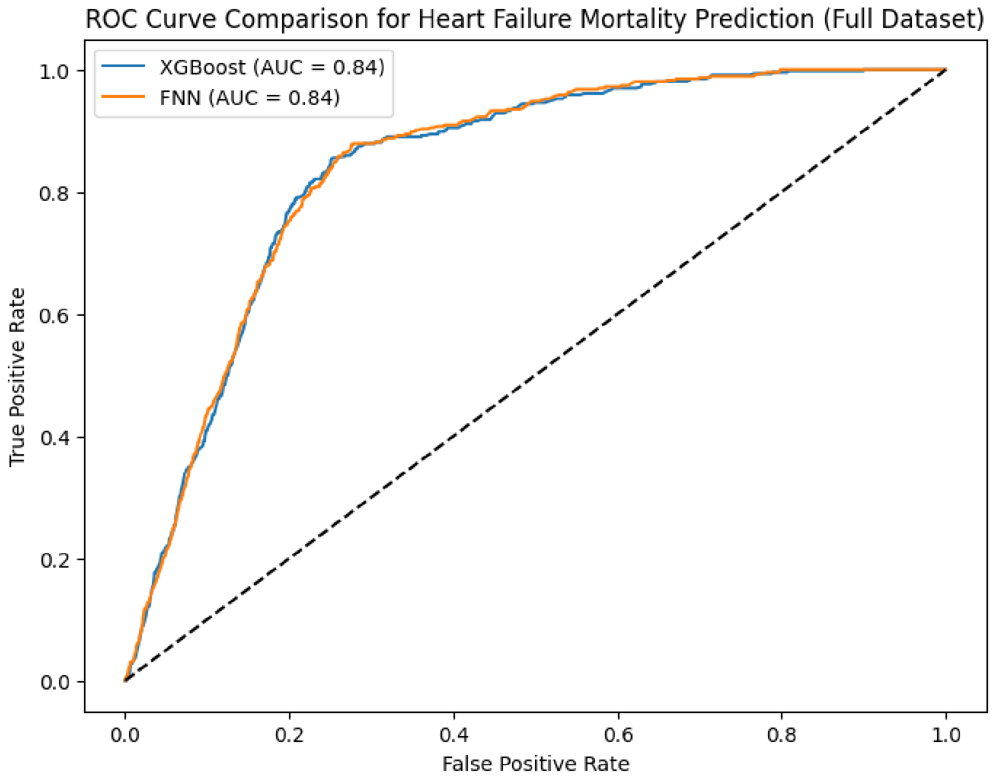
Feature Fitting: Used scaled features, learned complex patterns.

Training: 80/20 train-validation split (27,513 train, 6,879 val).

Data augmentation (noise), early stopping (epoch 26).

Evaluation: Threshold tuning (best = 0.30, min recall 0.7).

Evaluation of Results



ROC Curve:

- XGBoost AUC = 0.844, FNN AUC = 0.844.
- Nearly identical curves, excellent discriminative power.

Key Metrics:

- **XGBoost**: Accuracy: 0.9457, Precision: 0.0, Recall: 0.0, F1: 0.0.
- **FNN**: Accuracy: 0.7534, Precision: 0.1601, Recall: 0.8405, F1: 0.2690.
- **Ensemble**: Accuracy: 0.9458, Precision: 0.0, Recall: 0.0, F1: 0.0.

Analysis:

- FNN's high recall (0.8405) identifies most high-risk patients.
- Low precision (0.1601) due to class imbalance (5.40% mortality).

Plot/Output: ROC curve plot (XGBoost AUC = 0.844, FNN AUC = 0.844).

Discussion and Insights

Strengths:

- AUC of 0.844 for both models, suitable for risk stratification.
- FNN's high recall (0.8405) identifies most high-risk patients.
- XGBoost's feature importance (e.g., ICU stay, age) aids decisions.

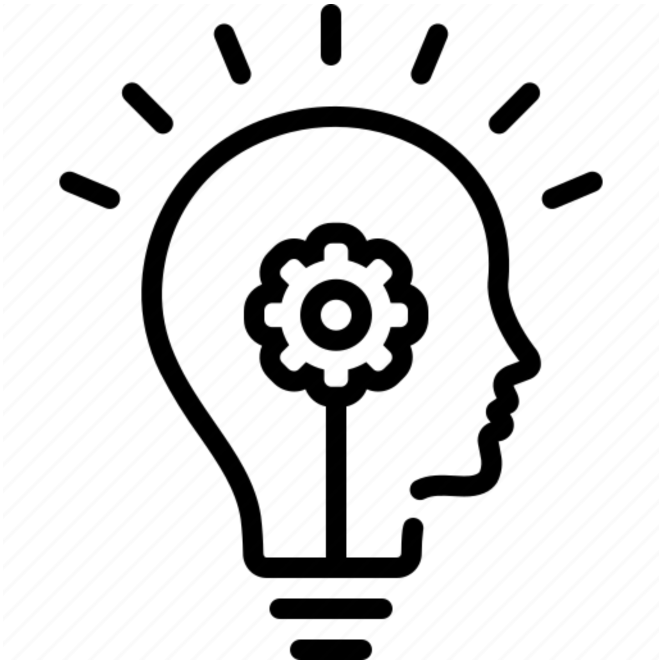
Insights:

- ICU stay and age are key predictors of mortality.
- FNN supports timely interventions for at-risk patients.

Challenges:

- Class imbalance (5.40%) leads to low precision for FNN.
- Ensemble threshold (0.5) needs tuning for better recall.

Conclusion



Summary:

- Developed ML/DL models for mortality prediction using MIMIC-IV.
- Achieved AUC of 0.844, FNN recall of 0.8405.
- Ensemble AUC of 0.8437 shows potential.

Future Work:

- Optimize ensemble threshold for better recall and F1.
- Explore advanced DL (e.g., transformers) with better hardware.

Clinical Impact:

- FNN identifies high-risk patients for timely interventions.
- XGBoost provides evidence-based insights for decisions.