ML/DL using MIMIC Data

Naveen Kilaru

Introduction

Title: Predicting Early ICU Transfer and Care Unit Assignment from MIMIC Data

Objective: Using Machine learning and Deep learning on EHR data

Speaker Notes:

Welcome to our project tutorial. We explore how to apply both traditional machine learning and deep learning techniques on real-world EHR data to predict ICU transfer risk within 24 hours of hospital admission and identify likely care unit destination.

Motivation & Objective

- ICU resources are limited
- Early ICU transfer prediction helps triage patients
- Objective: Predict ICU transfer within 24 hours of admission
- Bonus: Predict likely care unit destination (multitask learning)

Speaker Notes:

Accurate early risk stratification can improve care outcomes. We aim to build models that can anticipate ICU transfer to support proactive clinical decisions. With multitask learning, we also predict care unit to assist with real-time resource allocation.

Mimic Data Used

- PATIENTS: Age, gender
- ADMISSIONS: Admission time
- ICUSTAYS: ICU entry time
- PRESCRIPTIONS: Early medication use
- TRANSFERS: Initial care unit

Speaker Notes:

We join multiple MIMIC-III tables to construct meaningful features. These include both static demographics and early clinical actions.

Feature Engineering

- Age at admission (ADMITTIME - DOB)
- Gender (binary)
- Drug count within first 12h
- First transfer care unit (one-hot)
- Label 1: ICU entry within 24 h of ADMITTIME
- Label 2: Care unit classification (multiclass)

Speaker Notes:

We created a binary target variable and normalized inputs. Features were selected for clinical relevance and early availability. Multitask learning enables predicting both outcomes simultaneously

Machine Learning Pipeline

- Preprocessing: StandardScaler
- Train/Test split (80/20)
- Model: Random Forest Classifier
- Metrics: Accuracy, AUC, Precision/Recall

Speaker Notes:

Random Forest was chosen for robustness and interpretability. We evaluate using common classification metrics to understand performance.

Deep Learning Pipeline

- Model: 3-layer feedforward neural net (64-32-2)
- Framework: PyTorch
- Inputs: Scaled features
- Outputs:
  a. Binary (ICU transfer, Sigmoid)
  b. Multiclass (Care unit, Softmax)

Speaker Notes:

We implemented a multitask deep neural network with dual outputs. One predicts ICU transfer as binary, and the other predicts care unit using softmax classification.

Evaluation Results (Model Performance Summary)

Random Forest:

- ICU Transfer Accuracy: 74%
- AUC-ROC: 0.44
- F1 (Class 1): 0.85
- Poor at identifying non-ICU transfers

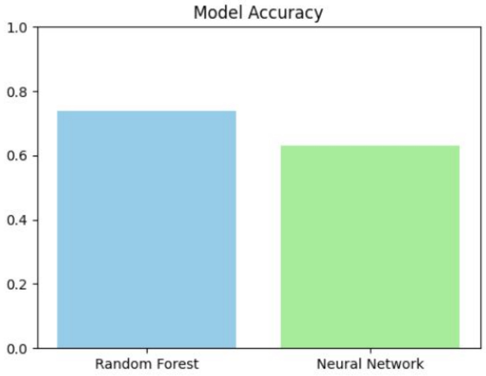
Deep Learning:

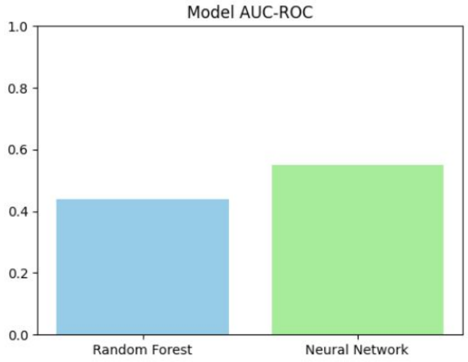
- ICU Transfer Accuracy: 85%
- AUC-ROC: 0.55
- F1 (Class 1): 0.92
- Better recall for ICU transfers

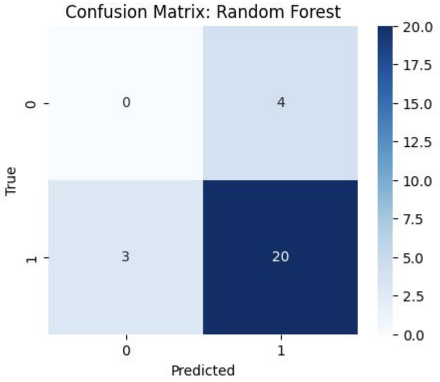
Speaker notes:

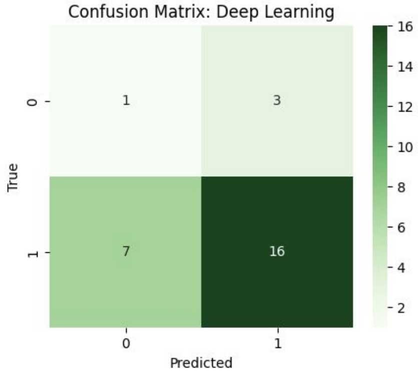
Deep learning showed higher accuracy and recall for ICU prediction, and was capable of predicting care unit destination with decent performance. Future tuning may further improve care unit classification.

Data Visualization









Conclusion & Next Steps

- ICU transfer is predictable from early features
- RF is strong with tabular data
- DNNs support multitask learning
- Future:
  - Add time series (e.g. vitals, labs)
  - Add NLP features (clinical notes)
  - Multitask refinement (joint loss optimization)

Speaker notes:

We demonstrated how structured EHR data can be transformed into predictive insights. Future work includes integrating more modalities and refining our multitask loss balancing.

Resources & Submission

- GitHub Url: NaveenKilaru_MIMIC_Data_MLDL_HW4.ipynb

Speaker notes:

Thanks for reviewing our tutorial. Please visit the repo and notebook for full code, multitask model architecture, and reproducibility.