Early Sepsis Prediction

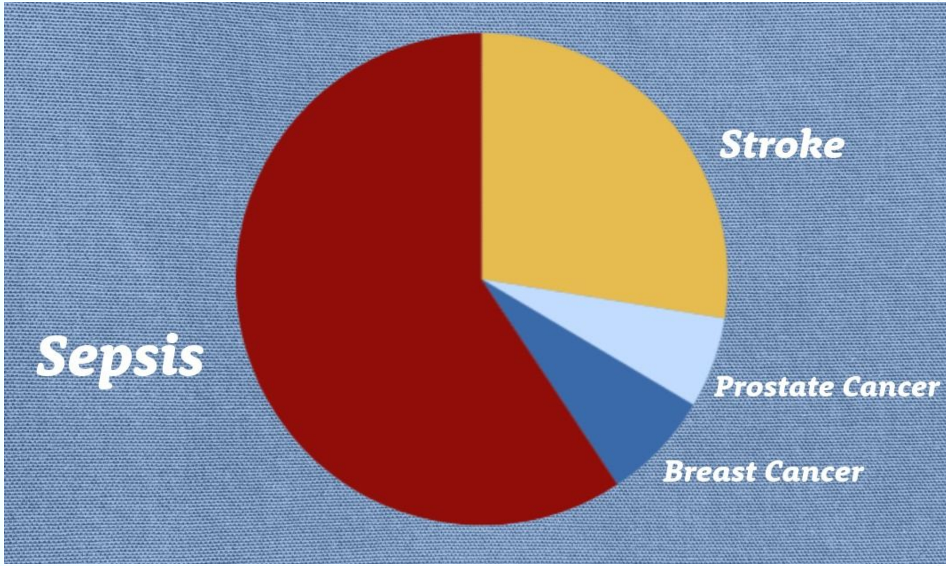
The Erdős Institute Deep Learning Bootcamp, Fall 2025

Yang Mo, Sayantan Sarkar, Cristopher Thompson, Alexandria Wheeler

Table of contents

Sepsis: A Critical Healthcare Crisis

- Leading cause of death in US hospitals
- 1.7M cases, ~350K deaths annually in U.S.
- $24B annual healthcare cost
- **Every hour of delay: 4–9% mortality increase**



Our Goal: Deep learning model for early ICU sepsis detection

Exploratory Data Analysis

Dataset:

- 40,000+ ICU patients from 2 hospitals
- Hourly time-series: 40 clinical variables + demographics
  - e.g.: HR, temp, potassium levels, age, gender

Key Challenges:

- Severe class imbalance:

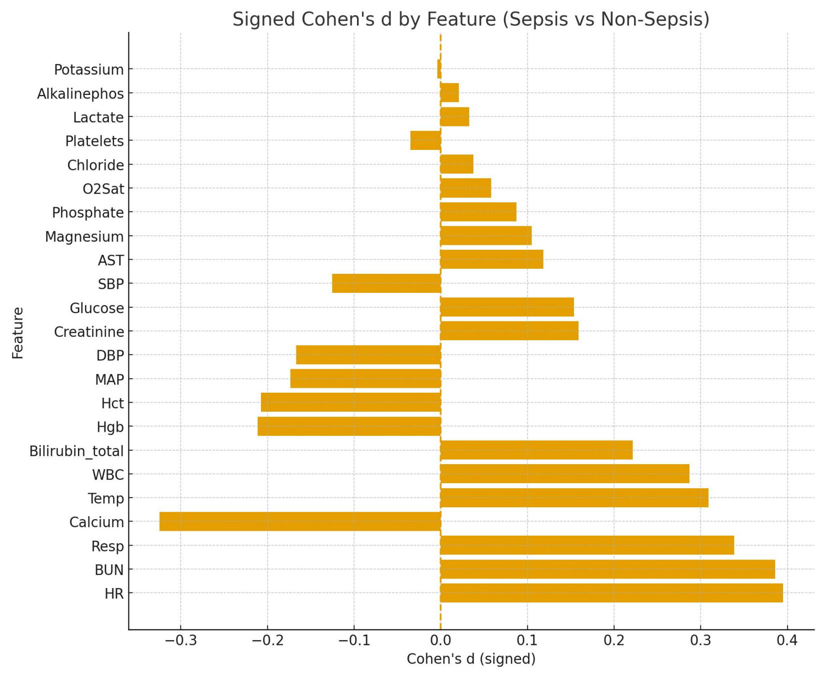
  7% sepsis rate
- Massive scale:

  1.4M+ hourly records
- Lots of missing values:

  ~20% data density

Exploratory Data Analysis

- Cohen's *d* measures **how far apart two group means are.**
- |d| < 0.2: negligible; 0.2–0.5: small; 0.5–0.8: moderate; ≥ 0.8: large.
- Separation is small but consistent across several vitals/labs
- Top separators (by |d|) highlight HR, BUN, Resp, Calcium, Temp
- Positive d ⇒ higher in septic; negative d ⇒ higher in non–septic



Data Preprocessing

- **For** very less sparse features (like temp) we used linear interpolation.
- **Missingness handled** via **KNN imputation Iterative/Model–based imputation (RandomForest)** where applicable.
- **Outliers capped** (winsorization) to enforce physiologic plausibility.
- **Standardization/scaling** applied; explicit **missingness masks** created for analysis/modeling.
- The data were **padded and masked** to align variable–length time series for deep learning input.

Baseline Models

Logistic Regression Baseline:

- Fast training/prediction
- Test AUROC: 66%
- Precision: 23% | Recall: 1%
- Insufficient for clinical use

KNN with DTW:

- Could not scale to large dataset
- Poor performance due to class imbalance

Deep Learning Models Explored

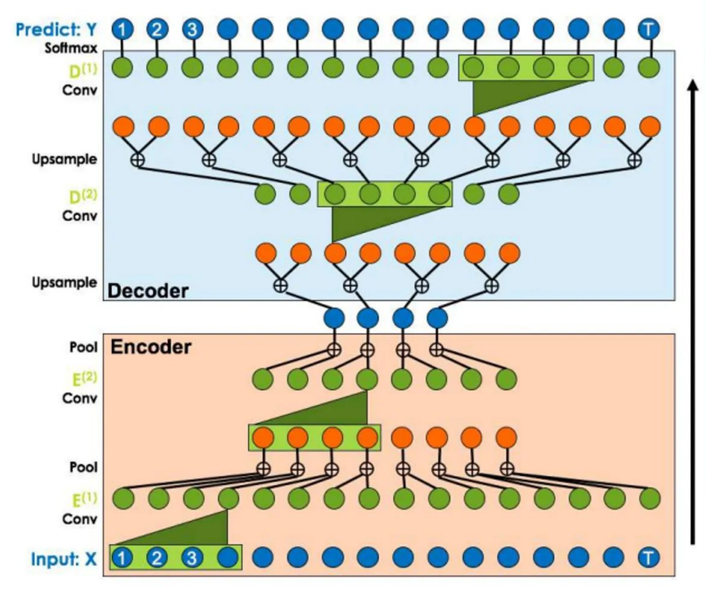
Temporal Convolution Network (TCN)

TCN= 1D CNN + causal convolutions

Performance (on validation data):

```
train_acc=0.9792 |
| Val loss=0.3226 |
AUROC=0.703 AUPRC=0.151
```

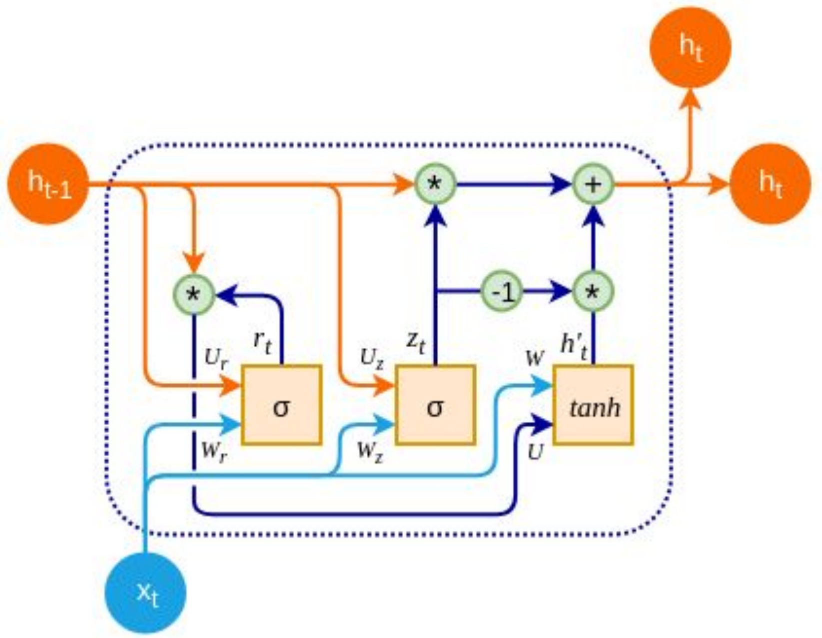
Not very good because of low AUPRC.



Lea et al. (2016)

GRU– D

- **Gated Recurrent Unit with Decay:** a deep learning model for irregular and missing time series data.
- Designed specifically for medical time series like ICU data, where sampling is irregular and values are often missing.
- Learns temporal changes and missingness patterns that reveal rare events early, rather than relying on frequency of labels– thus is very useful when there is a class imbalance.



courtesy

- **Patient−level**
    - **AUROC = 0.82**
    - **AUPRC = 0.51**
- **@ threshold = 0.24**
    - **Precision = 0.53**
    - **Recall = 0.41**
- **@ recall ≥ 0.90**, threshold = 0.03 → **Recall = 0.88**

Temporal Residual Transformer (TRT)

Model Design – Temporal Residual Transformer (TRT): A deep learning model that uses two Transformer Encoder blocks with 4-head self-attention, 256-dimensional feed-forward layers, and residual connections.

Epoch [20/20] | Train Loss: 0.3541 | Val Loss: 0.4449, Acc: 0.8400, Prec: 0.2734, Rec: 0.7248, AUC: 0.8793

Final Model/Results

Model selected: Temporal Residual Transformer (TRT)

--- Test Set Evaluation ---

Test Loss: 0.4401 Test Accuracy: 0.8351 Test Precision: 0.2717 Test Recall (Sensitivity): 0.7554 Test F1 Score: 0.3997 **Test AUROC: 0.8802 Test Avg Precision (AUPRC): 0.5543**

Future Goals

Improve the model to where it can predict the hour in which the patient gets sepsis.

Compare hospitals A and B to see if there is a difference.

Continue to further improve the model’s accuracy.