

# Reproducibility Project of Density-Aware Personalized Training for Risk Prediction in Imbalanced Medical Data

Kaley Nguyen

{kaleynn2}@illinois.edu

Group ID: 103

Paper ID: 81 (Huo et al., 2022)

Presentation link: [https://www.youtube.com/working\\_on\\_it](https://www.youtube.com/working_on_it)

Code link: <http://github.com/kaleynguyen/>

## 1 Introduction

Since public health data follow a long tail distribution (MW et al., 2014), many medical events are rare events. In the past, researchers have attempted to solve the issue by either undersampling the majority class, oversampling the minority class (Kotsiantis et al., 2005) or use synthetic examples to account for imbalance (Chawla et al., 2002). The authors of the article suggest the following framework for training the data with the imbalance problem: 1. Separate the feature extraction and classification processes; 2. Train the network using a loss function that takes into account density and a learnable cost matrix for misclassifications. Using the real-world medical datasets (MIMIC-III), the model's performance metrics are improved, as evidenced by the AUC-ROC, AUC-PRC, and Brier Skill Score.

## 2 Scope of reproducibility

The paper introduces a new approach to handle imbalanced medical data set by exploiting the imbalanced densities and training the network with a density-aware hinge loss function to improve the performance metric by 4.7% (max percent change AUC-ROC) and 14.2% percent (max percent change AUC-PRC) compared to the baselines of mortality prediction on MIMIC-III dataset.

### 2.1 Predictive Power and Performance Metrics

The authors propose a framework for addressing class imbalance density (7-10%) and utilizing the imbalance to provide density-aware training for enhanced risk prediction performance.

The framework proposes some ideas:

- The risk prediction performance can be improved by decoupling the training of repre-

sentation learning and classification to extract class-specific features and predictions.

- Differences in density are some features that should be learned, rather than attempted to be eliminated.
- AUC-ROC is too confident to use as the only performance metric to address the rare medical event predictive power.

### 2.2 Discriminative Power

The authors propose that the patients who survived would be more similar to each other compared to the patients who did not survive.

## 3 Methodology

I aim to implement the approach of the author as below:

1. Sampling 2 batches of data using 2 different sampling strategies, regular random sampling and stratified random sampling. Decoupling training for imbalance classes to separate the feature extraction and classifier output.
2. Density-aware outlier detection loss replaces the log loss or cross entropy loss to address the imbalance issue.
3. Trainable cost matrix is used to reduce the bias between the big class and small class.

### 3.1 Model descriptions

- Split the data using 80/10/10 splits
- Train the model on flexEHR, which is a GRU based method architecture.
- Train the model with 50 epochs, batch size is 128 and the Adam optimizer is used.

- AUC-ROC, AUC-PRC and BSS are used to measure the difficulty of discriminating power under imbalance.
- softmax is used in the last layer of the flex-EHR to obtain the predicted probability for the calibration study.

### 3.2 Data descriptions

For predicting a patient's mortality and phenotype, the authors only look at the first 48 hours after they are admitted to the ICU. The idea is that the precaution treatment can be done for early risk prognosis and phenotyping since the average stay in the ICU is 100 to 200 hours. The paper uses the same data pre-processing steps as a set of standard models in MIMIC-III (Harutyunyan et al., 2019) to build 34 time-series features using the same 17 clinical measurements.

### 3.3 Hyperparameters

- Density-aware component  $\Delta_c = \frac{K}{|N_c|^{1/4}}$  for  $c \in \{1, \dots, |C|\}$  where  $K$  is a hyper-parameter and  $|N_c|$  is number of examples in class  $c$ .
- Relax the constrained problem as an unconstrained one and rewrite  $C_{FN} = \theta C_{FP} + D$  where  $D$  is a regularization term,  $\theta$  is a hyper-parameter,  $C_{FP}$  is a type I error, and  $C_{FN}$  is a type II error in the confusion matrix. Notice that the trainable confusion matrix is only applicable for a binary classification problem.

### 3.4 Ablation Study

In the paper, two new components are introduced: decoupling the training and a loss function that takes into account the density of the data. The paper suggests using the same architecture - flexEHR or GRU. Specifically, one GRU as a baseline, one GRU with the trainable cost matrix, one GRU with the decoupling method, and one GRU with both the decoupling method and the trainable cost matrix. In the ablation study, the AUC-ROC does not change significantly, but the AUC-PRC and BSS improve significantly. However, GRU-decoupling has poor calibrated range compared to the combined framework.

### 3.5 Implementation

I plan to use the code from Homework 2,3 and 4 as my code for the project. I also plan to read the original paper that proposes the main backbone architecture - flexEHR.

### 3.6 Computational requirements

I plan to use paperspace to run the code on a P5000 30GB 8CPU 16 GiB GPU. It costs 8\$ per month to borrow several machines with similar computing power.

## 4 Results

Start with a high-level overview of your results. Does your work support the claims you listed in section 2.1? Keep this section as factual and precise as possible, reserve your judgement and discussion points for the next "Discussion" section.

Go into each individual result you have, say how it relates to one of the claims, and explain what your result is. Logically group related results into sections. Clearly state if you have gone beyond the original paper to run additional experiments and how they relate to the original claims.

Tips 1: Be specific and use precise language, e.g. "we reproduced the accuracy to within 1% of reported value, that upholds the paper's conclusion that it performs much better than baselines." Getting exactly the same number is in most cases infeasible, so you'll need to use your judgement call to decide if your results support the original claim of the paper.

Tips 2: You may want to use tables and figures to demonstrate your results.

#### 4.1 Result 1

#### 4.2 Result 2

#### 4.3 Additional results not present in the original paper

Describe any additional experiments beyond the original paper. This could include experimenting with additional datasets, exploring different methods, running more ablations, or tuning the hyperparameters. For each additional experiment, clearly describe which experiment you conducted, its result, and discussions (e.g. what is the indication of the result).

## 5 Discussion

Describe larger implications of the experimental results, whether the original paper was reproducible, and if it wasn't, what factors made it irreproducible.

Give your judgement on if you feel the evidence you got from running the code supports the claims of the paper. Discuss the strengths and weaknesses of your approach – perhaps you didn't have time

to run all the experiments, or perhaps you did additional experiments that further strengthened the claims in the paper.

### 5.1 What was easy

Describe which parts of your reproduction study were easy. E.g. was it easy to run the author's code, or easy to re-implement their method based on the description in the paper. The goal of this section is to summarize to the reader which parts of the original paper they could easily apply to their problem.

Tips: Be careful not to give sweeping generalizations. Something that is easy for you might be difficult to others. Put what was easy in context and explain why it was easy (e.g. code had extensive API documentation and a lot of examples that matched experiments in papers).

### 5.2 What was difficult

Describe which parts of your reproduction study were difficult or took much more time than you expected. Perhaps the data was not available and you couldn't verify some experiments, or the author's code was broken and had to be debugged first. Or, perhaps some experiments just take too much time/resources to run and you couldn't verify them. The purpose of this section is to indicate to the reader which parts of the original paper are either difficult to re-use, or require a significant amount of work and resources to verify.

Tips: Be careful to put your discussion in context. For example, don't say "the math was difficult to follow," say "the math requires advanced knowledge of calculus to follow."

### 5.3 Recommendations for reproducibility

Describe a set of recommendations to the original authors or others who work in this area for improving reproducibility.

## 6 Communication with original authors

Document the extent of (or lack of) communication with the original authors. To make sure the reproducibility report is a fair assessment of the original research we recommend getting in touch with the original authors. You can ask authors specific questions, or if you don't have any questions you can send them the full report to get their feedback.

## References

- Nitesh Chawla, Kevin Bowyer, Lawrence Hall, and W. Kegelmeyer. 2002. [Smote: Synthetic minority over-sampling technique](#). *J. Artif. Intell. Res. (JAIR)*, 16:321–357.
- Hrayr Harutyunyan, Hrant Khachatrian, David Kale, and Aram Galstyan. 2019. [Multitask learning and benchmarking with clinical time series data](#). *Scientific Data*, 6.
- Zepeng Huo, Xiaoning Qian, Shuai Huang, Zhangyang Wang, and Bobak J. Mortazavi. 2022. [Density-aware personalized training for risk prediction in imbalanced medical data](#).
- Sotiris Kotsiantis, D. Kanellopoulos, and P. Pintelas. 2005. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30:25–36.
- Kreuter MW, Hovmand P, Pfeiffer DJ, Fairchild M, Rath S, Golla B, and Casey C. 2014. [The “long tail” and public health: new thinking for addressing health disparities](#). *Am J Public Health*, 104(12):2271–2278.