ICU Mortality Prediction Checkpoint #2

Introduction:

ICU datasets can look very different from one hospital to another, especially in how often measurements are recorded and how variables are defined. Because of this, a model that performs well in one ICU setting may not work as well in another. We want to explore this problem using modern deep learning methods, so our project focuses on Transformer-based time-series models and how well they transfer across datasets.

We train our main model on the HiRID dataset, which contains high-resolution ICU data that tends to work well with deep learning architectures. After building a strong baseline on HiRID, we test how well the same model performs on MIMIC-IV, a widely used ICU dataset with different recording patterns and clinical characteristics. We also try several domain-adaptation strategies to see whether we can narrow the performance gap between training on one dataset and applying the model to another.

Overall, this is a supervised learning problem and a binary classification task. Using the first 24 hours of ICU data, we predict whether a patient will die during their ICU stay.

Challenges: What has been the hardest part of the project you've encountered so far?

Our biggest challenge so far has been being familiar with an entirely new technical and clinical domain. Although all of us have ML experience, working with critical-care time-series data, ICU workflows, and real-world EHR conventions has required learning a completely different set of assumptions about how the data is generated and where models tend to fail. Understanding how HiRID and MIMIC-IV define and record variables, how the absence of certain measurements carries clinical meaning, and how to structure patient stays into meaningful sequences turned out to be more complex than the typical ML datasets we are used to working with. At the same time, we had to familiarize ourselves with Transformer-based architectures for time-series data and with domain adaptation techniques such as adversarial training. So far, this steep learning curve has been the biggest challenge, but it has also clarified the conceptual motivations behind our project and helped us design more realistic and thoughtful experiments.

Insights: Are there any concrete results you can show at this point?

- How is your model performing compared with expectations?

While we don't yet have concrete benchmarks on the performance of our model, we have spent time reviewing the data and ensuring that our model architecture aligns with established baselines in prior work. We have also formulated a concrete plan and division of labor to efficiently carry out training, evaluation, and domain adaptation experiments so that we can generate meaningful results in a timely manner.

Plan: Are you on track with your project?

- What do you need to dedicate more time to?

We need to allocate more time to implementation and experimental execution rather than conceptual exploration, so we can begin producing concrete results.

- What are you thinking of changing, if anything?

We're happy with our current approach and are confident that we can move ahead as planned.