Developing machine learning models to predict probability of ICU mortality within 24 hours post admission

CHL5230H Datathon #4 High-fidelity Report

By: Jacqueline Jia, Paijani Sheth, Xiao Yan

Introduction

Acute physiology and chronic health evaluation (APACHE) is a prognostic system that describes intensive care unit (ICU) patients and evaluates their care (1). Delayed admission into the ICU has been associated with increased mortality, and physicians have to make important decisions when newly admitted patients are in critical conditions, especially if there are not enough ICU beds (2). Hence, the prediction of mortality in the first 24 hours after admission into the ICU is very important, and machine learning algorithms can inform health care workers in their decision making, such as deciding on what patients to prioritize, and how to allocate limited resources.

The APACHE IVa probabilistic prediction of in ICU mortality utilizes the highest APACHE III score in the APACHE scoring system, because it is calculated using measurements from physiological variables made at the patient's most critical state within the first 24 hours of ICU admission. The physiological variables used in determining the APACHE III score includes demographic variables (like age, sex, and weight), vital measurements (like arterial partial pressure of carbon dioxide, body temperature, and heart rate) and other covariates (2).

The objectives of this study are to develop a model that predicts APACHE IVa probability of ICU mortality, identify what variables measured in the first 24 hours after admission of ICU best predict this outcome, and determine what model has the best performance in predicting this outcome.

Data Engineering Process

This study uses data provided in collaboration with MIT's GOSSIS community initiative and includes over 90,000 ICU patient observations over 1 year. The dataset consisted of 91713 observations and 155 variables. Since the outcome variable was calculated using APACHE III, all APACHE covariates (measurements made at the highest APACHE III score) were removed in order to reduce bias. Instead, patient measurements and demographic variables during the first 24 hours post admission were considered in the model building. Model building involved a combination of selecting variables based on a conceptual model (literature search), and computational analysis (low variance, lasso regression, and multicollinearity). Previous literature showed evidence that sex, oxygen saturation, heart rate, age, and hypertension are important predictors of ICU mortality (3-6). Variables in the dataset were removed if they had a variance threshold below 1% and then lasso regression was conducted. The lasso regression determined that in addition to some of the variables already mentioned, the length of stay of the patient between hospital admission and unit admission, ethnicity, presence of metastatic tumor, plasma potassium, sodium and hemoglobin concentrations, are important predictors in ICU mortality probability. The final selected covariates in the model were 'ethnicity', 'pre_icu_los_days', 'd1_spo2_max', 'd1_spo2_min', 'd1_mbp_min', 'd1_mbp_max', 'pre_icu_los_days', 'hepatic_failure', 'solid_tumor_with_metastasis', 'icu_type', 'gender', 'icu_stay_type', 'age', 'd1_sodium_max', 'd1_hemaglobin_max', 'd1_potassium_min', 'death_prob_level'

Furthermore, missing values were median imputed for numerical variables and removed for categorical variables. Removal for the latter was justified because missing data took up less than 5% of the dataset. After data cleaning and feature selection, there were 87414 observations and 17 variables in the dataset.

Analysis

This study used three deep learning models (AdaBoost, Random Forest, Neural Network) to predict the ICU death probability. For the AdaBoost and Random Forest model, the continuous outcome variable (APACHE icu death probability) was categorized into three different levels ('low', 'moderate', 'high') based on their distribution and set the outcome variable as death probability level. Since the icu death probability's distribution was left skewed, it was categorized as any icu death probability less than 3% as low, 3% to 20% as moderate, 20% to 100% as high. The outcome variable was renamed 'death_prob _level'.

The dataset was split into training and testing sets, with 80% of the data allocated for training and 20% reserved for testing. This split ensures that the model is tested on unseen data to evaluate its predictive performance. We implemented a ColumnTransformer with OneHotEncoder to transform categorical variables into a machine-learning-friendly format. We integrated the random forest classifier/Adaboost classifier into a pipeline alongside the preprocessing step. This pipeline approach ensured a streamlined process where data is automatically pre-processed before being fed into the classifier. The model was then trained on the training subset of the data. Finally, accuracy was calculated separately for both the training and testing datasets. This helped to understand the model's ability to generalize, as well as in detecting any overfitting or underfitting issues. After initial evaluation of performance, the Adaboost model was hyperparameter tuned using GridSearch with 10-fold cross validation, and the following parameters were specified: the total number of trees (50, 100, 200) and learning rate (0.1, 0.2) and checked again for performance.

In addition, a neural network model was employed to better capture the patterns lying in the probability of ICU mortality. Neural network parameters were initialized using the PyTorch library. It was aimed to have two hidden layers of neural network. By setting hidden units in both layer 1 and layer 2 to 5, 5 neurons were in each layer, which were connected to the input layer with 36 neurons. The tensors were then set representing the weights and biases for the second and output layer similarly. Afterwards, the forward pass of the neural network was defined, with hyperbolic tangent activation function for the first and second hidden layer, and sigmoid activation for the output layer because the outcome variable was a probability ranging from 0 to 1. Then, training hyperparameters were tuned, specifically, epoch was set to 200, learning rate to 0.01 and batch size to 32. The learning rate was set to reduce by 20% within every 50 epochs. Finally, the training loss over epochs was plotted and model performance on both training and validation data were evaluated.

Findings

The initial training and testing accuracy of the random forest model was 0.6399 and 0.6447, respectively. The proximity of both scores suggest that the model is generalizing well to unseen data. The initial training and testing accuracy of the AdaBoost model was 0.6388 and 0.6462, respectively. Similarly, the model generalized well to test data. After hyperparameter tuning for the AdaBoost model, it was found that the optimal parameters were learning rate = 0.1 and total number of trees = 50. The model’s performance after setting optimal parameters did not improve, as it resulted in an accuracy of 0.6459.

Upon constructing the neural network model, the training loss over epochs was plotted, and it was found that the loss dropped sharply initially, and did not have a notable decrease after 25 epochs. Therefore epoch was set to 200 with a learning rate 0.01. By evaluating the performance of the neural network model on validation and training data, the model’s loss on the validation set was 0.235, and the loss on the training set was 0.232, indicating that there is no overfitting and the model performed well in terms of its generalization to unseen data.

Conclusion

In conclusion, AdaBoost performed slightly better than Random Forest in predicting ICU death probability, however, both accuracies were similar. A possible reason why both models performed so similarly could be because of the complexity and high dimensionality of the model. A model with relatively high dimensionality can limit the number of weak classifiers, and as a result, the AdaBoost would behave similar to a Random Forest model. The neural network model had a high performance on the validation set as its validation loss was similar to its training loss. All three models had a high generalizability with no sign of overfitting.

Limitations in this study included the imbalanced outcome variable. Since the class distributions were not balanced, accuracy may not have been the most suitable metric to use. Additionally, other performance metrics like recall should be assessed in the future when there is more balanced data. Recall may be more relevant as false negatives should be minimized (ie. the number of patients who are at higher probability of mortality don’t get treated are minimized). Furthermore, since deep learning models were used for predictive analysis, the interpretability of describing how the model outputs ICU death probability was sacrificed due to the black-box nature of these models. Future work should look into developing other models.

Individual Contributions

Jacqueline Jia: Conducted EDA. Performed random forest modeling and data cleaning. Compiled the analysis, findings and conclusion part for random forest in the report.
Paijani Sheth: Conducted EDA. Performed AdaBoost modeling and data cleaning. Wrote the introduction, data engineering, conclusion, and adaboost related parts in the report.
Xiao Yan: Built the neural network model for analysis. Compiled the analysis, findings and conclusion part on neural networks in the report.

Code and Presentation

GitHub: [JacquelineeJia/datathon4-Team14-CHL5230-F23 (github.com)]
Slides: Datathon4_Presentation

Reference

1. Leong, J., Madhok, J., & Lighthall, G. K. (2020). Mortality of Patients Requiring Escalation to Intensive Care within 24 Hours of Admission in a Mixed Medical-Surgical Population. *Clinical Medicine & Research*, *18*(2–3), 68–74. https://doi.org/10.3121/cmr.2019.1497
2. Wagner, D. P., & Draper, E. A. (1984). Acute physiology and chronic health evaluation (APACHE II) and Medicare reimbursement. *Health Care Financing Review*, *1984*(Suppl), 91–105.
3. Awad, A., Bader-El-Den, M., McNicholas, J., Briggs, J., & El-Sonbaty, Y. (2020). Predicting hospital mortality for intensive care unit patients: Time-series analysis. *Health Informatics Journal*, *26*(2), 1043–1059. https://doi.org/10.1177/1460458219850323
4. Khajehali, N., Khajehali, Z., & Tarokh, M. J. (2023). The prediction of mortality influential variables in an intensive care unit: A case study. *Personal and Ubiquitous Computing*, *27*(2), 203–219. https://doi.org/10.1007/s00779-021-01540-5
5. Li, L., Bai, M., Zhang, Q., & Sun, S. (2022). Characteristics and predictors of ICU-mortality in critically ill patients with hyperlactatemia requiring CRRT: A retrospective cohort study. *The International Journal of Artificial Organs*, *45*(12), 973–980. https://doi.org/10.1177/03913988221126728
6. Li, X., Ge, P., Zhu, J., Li, H., Graham, J., Singer, A., Richman, P. S., & Duong, T. Q. (2020). Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. *PeerJ*, *8*, e10337. https://doi.org/10.7717/peerj.10337

Appendix

Figure 1. Distribution of Classes in *ICU death probability level*

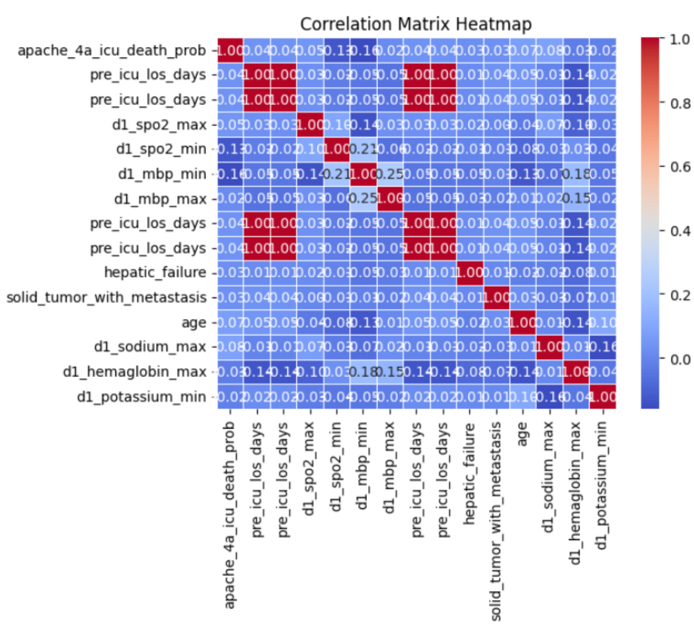

Figure 2. Correlation Matrix Heatmap of variables