Prediction of Length-of-stay at Intensive Care Unit (ICU) Using Machine Learning based on MIMIC-III Database

Md Nahid Hasan *, Sammi Hamdan†, Samir Poudel †, Jorge Vargas ‡, Khem Poudel * †
Email: mh2ay, sah9j, sp2ai@mtmail.mtsu.edu, jorge.vargas@mtsu.edu, khem.poudel@mtsu.edu,
*Department of Computational and Data Science,
†Department of Computer Science,
‡Department of Engineering Technology,
Middle Tennessee State University, Murfreesboro, TN 37132, USA

2023 IEEE Conference on Artificial Intelligence (CAI) | 979-8-3503-3984-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/CAI54212.2023.00142

***Abstract*—The length-of-stay (LOS) is critical for patient care and accommodation in the intensive care unit (ICU). In this work, we developed a framework to predict the LOS using the Medical Information Mart for Intensive Care (MIMIC-III) database. We extracted six features from individual patients and submitted them to the regressors model and examined how well these features could predict LOS. We considered four prediction regimes; extreme gradient boosting (XGBoost), support vector regressor, random forest, and voting regressor. Our analysis reveals that XGBoost yields the best result among other regressors with R^2 0.86 and root mean square error (RMSE) 1.2. Remarkably, our results show that ICD9 (9^{th} International classification of diseases code), saline intake per hour, and drug rates are the top three critical features for predicting the LOS.**

***Index Terms*—Regression, extreme gradient boosting, voting regressor, MIMIC, machine learning, feature selection.**

I. Introduction

Modern electronic healthcare records (EHRs) contain an increasingly large amount of data to improve healthcare services such as better treatment, hospital operation, and explore scientific questions [1]. Additionally, some of the information has a great influence for providing better care and identification of morality rate. The application of machine learning (ML) techniques provides a better understanding and interpretation of health care. Moreover, it improves the facilities of EHR systems without any intervention from humans [2], [3]. Presently, ML techniques are used in EHR for diagnosis, and to understand how the multiple factors are associated with diseases [4].ML is a subset of artificial intelligence that *"learns a model"* from the past data to predict the future data [5]. Deep learning (DL) and classical ML are widely used in healthcare systems and many other applications [5]. DL needs a huge amount of data and higher computation power, however, classical ML works well with a small set of data and less computing power. Extreme gradient boosting (XGBoost), support vector regressor (SVR), random forest (RF), and voting regressor (VR) are widely used as robust frameworks for prediction (regression and classification) in various fields, including the medical sector and industry [6].

Our present study demonstrates a new analysis on the Medical Information Mart for Intensive Care (MIMIC-III) database for the prediction of length-of-stay (LOS) using machine learning regressors with a small subset of feature spaces. We hypothesized that some features have more influence on the LOS, however, we do not know what features are more important for predicting the LOS. Using ML techniques, we explore what features have the most influence on the prediction of LOS. To the best of our knowledge, this is the first classical ML approach for the prediction of LOS and how the demographics, microbiology, drug rate, ICD9 (9^{th} International classification of diseases code) [7], and current service features (e.g., medication or surgery) are associated with LOS on MIMIC-III database. We demonstrate a comprehensive data-driven approach for the prediction of LOS on the MIMIC-III database.

The main contributions of our work are as follows:

- We conducted four regimes of regression analysis (i.e., SVR, RF, XGBoost, and VR) on MIMIC-III database to predict the LOS.
- We found that XGBoost regression shows the best performance among all other regressors.
- We demonstrated that ICD9 code, saline intake per hour, and drug rate are strongly associated with LOS.

The rest of the paper is organized as follows. We discuss the previous related research work on MIMIC-III database in Section II and brief description of the data and methods in Section III. Subsequently, we discuss the results in Section IV. Finally, We highlight our key findings and limitations in Section V.

II. Related Work

There is a tremendous development in ML over the past several decades in medical diagnostics and automatic detection of casual relationships [8]. Research works show that multiple factors are associated with the prediction of LOS and mortality rate in ICU. *Gentimis et al.* [2] demonstrated the prediction of LOS in a specific time span using neural networks. Researchers examined that risk factors of the first admission day are strongly related for predicting the mortality in ventilated patients [9]. Scherpt et al. predicted the sepsis of MIMIC-III database using recurrent neural network [10]. The trajectory of patients' interactions with health care was demonstrated via deep learning model [11]. Some research encoded the ICD9 code using deep learning [7]. However, there is no combined study for LOS prediction and feature selection using classical ML.

979-8-3503-3984-0/23/$31.00 ©2023 IEEE
DOI 10.1109/CAI54212.2023.00142

III. Materials and Methods

In this work, we used MIMIC III database [12]. This work reflects a new analysis of the MIMIC-III database using the classical machine learning approach. We extracted six general features. Presumably, these features link with LOS. We used these features as input to XGBoost, SVR, RF, and VR. We trained the regressors and examined the model performance using unseen test data. The analysis flowchart of this work is illustrated in Fig. 1.

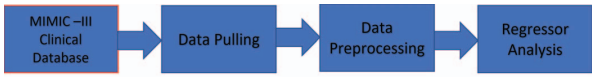

Fig. 1: Analysis flowchart.

A. Participants and Data Description

In this work, we used a demo subset of the MIMIC-III database that consists of 100 participants (male: 45, Female: 55) which is freely available. [12], [13]. This data was originally collected over a decade of intensive care unit (ICU) patient stays at Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA. The participants are of different ethnicity and currently taking various medical services. More detail about this database is available in [12].

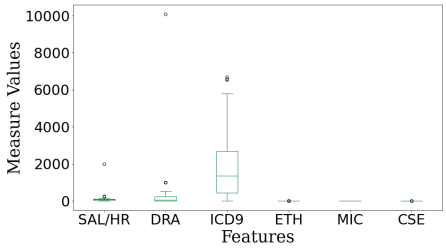

Fig. 2: Representation of features. SAL/HR: Saline intake per hour, DRA: Drug rate, ICD9: International classification of diseases [7] code, ETH: Ethnicity, MIC: Microbiology events (e.g., Stool, sputum, and urine), CSE: Current services (e.g., Medication or surgery).

B. Data Preprocessing

We preprocessed the freely available demo 100 participants of the MIMIC-III database. After preprocessing, we found that only 48 patients have six features. We used these extracted six features (i.e., demographics, microbiology, drug rate, ICD9, saline intake per hour, and current service) as input of the different regressor models to examine the prediction of LOS. Drug rate, ICD9, and saline intake per hour of the features are numerical values. Demographics, microbiology, and current service are categorical. First, we transformed the categorical variable into numerical values using the scikit-learn [14] function's OrdinalEncoder. Then we used these numerical feature values for further analysis using different regressors to predict the LOS. The box plot representation of these six features is represented in Fig. 2.

C. Extreme Gradient Boosting (XGBoost) Regressor

In healthcare, it is important to know how the different factors are associated with the target (e.g., LOS). The classical ML can find the relation and the pattern with the target variables. ML models learn a discriminated function from the previous data (i.e., training data) and predict the future (i.e., test data). XGBoost regressor is faster and widely used for classification and regressor analysis in health care and industry [6]. . It is a highly efficient and flexible algorithm. It could be used as a classifier and regressor. In this work, we used an XGBoost regressor for predicting the LOS of the MIMIC-III database. We spilt our data randomly into training and test sets of 80%, and 20%, respectively [15]. The XGBoost model performance significantly varies based on hyper-parameters [16]. During the pilot training phase, we fine-tuned the XGBoost regressor's hyper-parameters (e.g., C, γ, n_estimators, max_depth, etc.,) with a grid search approach. Once the model has been trained then we selected the best parameters of the XGBoost regressor. In this study, the best parameter are: objective ='reg: tweedie', colsample_bytree = 0.3, subsample = 0.8, learning_rate = 0.1, max_depth = 5, booster ='dart', alpha, n_estimators = 100). Then we predicted the LOS by feeding the feature vectors only. After that, we computed the model performance metrics (e.g., R^2 and root mean square error (RMSE)) by using the predicted LOS with the actual LOS. An excellent model has higher R^2 close to 1 and lower RMSE; this means that it is a good fit model with generalized. On the other hand, a bad model has R^2 near to 0 and higher RMSE; it means that the model does not generalize. The model predicted LOS and R^2 performance are presented in Fig. 3 and Fig. 4, respectively. We also examined the importance of the features from our best model (e.g., XGBoost regressor). The feature importance is delineated in Fig. 5.

D. Voting Regressor (VR)

A Voting regressor is an ensemble regressor that consists of several meta-regressors. In our analysis, we considered XGBoost, RF, and SVR as a base regressor of the voting regressor. The voting regressor prediction is the average prediction of all base regressors. We separately conducted our analysis using XGBoost, RF, SVR, and VR. Since we already know the best parameters of the XGBoost regressor, we kept it the same but fine-tuned the hyperparameters of the other two models (e.g., RF, SVR). For SVR, we fine-tuned 10 different values of (C, γ) in the following range for the C = [2^{-1} to 2^4], and γ = [2^{-2} to 2^4], $eta = [0.1, 0.2, 0.3]$ and kernel= rbf. For RF, parameters was n_estimators= 100, max_depth = 5, criterion= mse. Our analysis results for each regressor model are reported in the results section and illustrated in Fig. 3 and Fig. 4.

IV. Results

The prediction of LOS using XGBoost, SVR, RF, and VR are presented in Fig. 3. In this figure, it is clear that the prediction of XGBoost is closer to the actual LOS. However, SVR and RF predicted LOS to have a larger difference than the true LOS. So our analysis demonstrated that XGBoost alone showed the best-predicted ability and VR showed a little less performance as compared to XGBoost. However, SVR and RF yielded less performance as compared to XGBoost and VR. The R^2 of XGBoost, VR, SVR, and RF are 86%, 76%, 60%, and 52%, respectively. Moreover, we also observed another

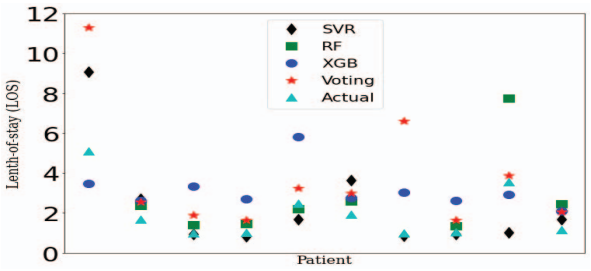

Fig. 3: Prediction of LOS using different regressors and actual LOS. SVR: support vector regressor, RF: random forest; XGBoost: extreme gradient boosting, voting: voting regressor that combined of SVR, RF, and XGBoost.

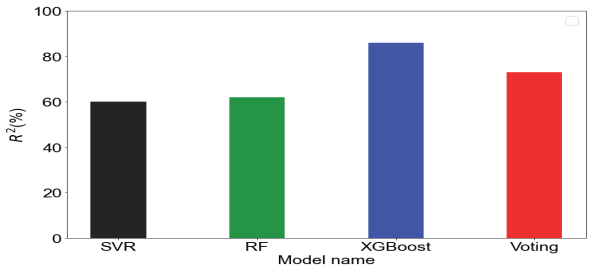

Fig. 4: Model performance in R^2.

performance metric (e.g., RMSE) of these regressors are 1.2 %, 1.9%, 17.8%, and 28.8%, respectively.

The XGBoost regressor outperformed all of the other regressors. Hence, we selected the XGBoost as the best framework for identifying feature importance and LOS prediction. Then we identified the feature importance of the XGBoost from the training model. This is illustrated in Fig. 4. Here the feature procedure events (ICD9, F-score: 29.0) is the top one-ranked. Input events (SAL/HR, F-score: 26.0) is the top-second ranked feature to predict LOS. The drug rate is the top-third ranked feature (DRA, F-score: 24.0). The lowest two F-scores shown are in microbiology (MIC, F-score: 10.0) and ethnicity (ETH, F-score: 9.0). Ethnicity and microbiology have less importance as compared to others.

V. Conclusion

We developed a robust computationally efficient framework for predicting LOS on the MIMIC-III database. Our analysis demonstrates that a XGBoost regressor can predict the LOS greater than 85% R^2 from six features. Notably, three features exhibited a more substantial influence on the prediction of LOS. Due to limitations of our work, we only considered six features and a few patients. In future work, we will incorporate more features and more data to explore the mortality rate in ICU.

References

[1] M. Ghassemi, T. Naumann, F. Doshi-Velez, N. Brimmer, R. Joshi, A. Rumshisky, and P. Szolovits, "Unfolding physiological state: Mortality modelling in intensive care units," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 75–84.

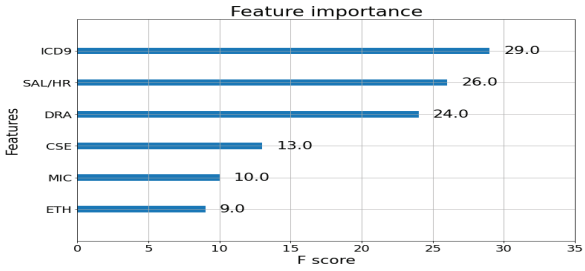

Fig. 5: Feature importance with ranked. Higher F score indicates most important; lower score indicates less importance.

[2] T. Gentimis, A. Ala'J, A. Durante, K. Cook, and R. Steele, "Predicting hospital length of stay using neural networks on mimic iii data," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE, 2017, pp. 1194–1201.

[3] A. Vaid, S. K. Jaladanki, J. Xu, S. Teng, A. Kumar, S. Lee, S. Somani, I. Paranjpe, J. K. De Freitas, T. Wanyan *et al.*, "Federated learning of electronic health records to improve mortality prediction in hospitalized patients with covid-19: Machine learning approach," *JMIR medical informatics*, vol. 9, no. 1, p. e24207, 2021.

[4] F. S. Ahmad, L. Ali, H. A. Khattak, T. Hameed, I. Wajahat, S. Kadry, S. A. C. Bukhari *et al.*, "A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (ehrs)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3283–3293, 2021.

[5] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future healthcare journal*, vol. 6, no. 2, p. 94, 2019.

[6] J. Taninaga, Y. Nishiyama, K. Fujibayashi, T. Gunji, N. Sasabe, K. Iijima, and T. Naito, "Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[7] M. Li, Z. Fei, M. Zeng, F.-X. Wu, Y. Li, Y. Pan, and J. Wang, "Automated icd-9 coding via a deep learning approach," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 16, no. 4, pp. 1193–1202, 2018.

[8] N. Louridi, S. Douzi, and B. El Ouahidi, "Machine learning-based identification of patients with a cardiovascular defect," *Journal of Big Data*, vol. 8, no. 1, pp. 1–15, 2021.

[9] Y. Zhu, J. Zhang, G. Wang, R. Yao, C. Ren, G. Chen, X. Jin, J. Guo, S. Liu, H. Zheng *et al.*, "Machine learning prediction models for mechanically ventilated patients: Analyses of the mimic-iii database," *Frontiers in Medicine*, vol. 8, p. 955, 2021.

[10] M. Scherpf, F. Gräßer, H. Malberg, and S. Zaunseder, "Predicting sepsis with a recurrent neural network using the mimic iii database," *Computers in biology and medicine*, vol. 113, p. 103395, 2019.

[11] B. K. Beaulieu-Jones, P. Orzechowski, and J. H. Moore, "Mapping patient trajectories using longitudinal extraction and deep learning in the mimic-iii critical care database," in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium*. World Scientific, 2018, pp. 123–132.

[12] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[13] S. Wang, M. B. McDermott, G. Chauhan, M. Ghassemi, M. C. Hughes, and T. Naumann, "Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii," in *Proceedings of the ACM Conference on Health, Inference, and Learning*, 2020, pp. 222–235.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[15] A. Rácz, D. Bajusz, and K. Héberger, "Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification," *Molecules*, vol. 26, no. 4, p. 1111, 2021.

[16] H. Nguyen, T. Vu, T. P. Vo, and H.-T. Thai, "Efficient machine learning models for prediction of concrete strengths," *Construction and Building Materials*, vol. 266, p. 120950, 2021.