BAOJ Biotechnology
An Open Access Journal

BAOJ Biotechnol
Short Commentary
Volume 6

On the applications of machine learning algorithms to predict the in-hospital mortality rate among heart failure patients

***Corresponding Author: Muhammad Adib Uz Zaman**
University of Florida, Gainesville, Florida, USA.
Email: a_u_z_ipe@yahoo.com

Abstract

With the enhancement of computational power and modern hardware, machine learning (ML) algorithms have been widely used in almost every sector these days. In modern healthcare systems, tons of data are generated each day and so the applications of machine learning in healthcare continue to grow rapidly. Modern ML algorithms require a very large amount of data to achieve somewhat reasonable performance. Electronic health records (EHR) data can be a good source of information in this case. In this study, the predictive performances of some popular machine learning algorithms on medical information mart for intensive care version 3 (MIMIC-III) EHR data to predict the mortality of heart failure patients, are discussed and presented.

Keywords: Machine learning; Gradient boost; Mortality prediction; ecision tree; MIMIC-III.

Received: Feb 04, 2022
Accepted: Mar 21, 2022
Published: Mar 25, 2022
Archived: www.bioaccent.org

Citation: Zaman MAU. On the applications of machine learning algorithms to predict the in-hospital mortality rate among heart failure patients. BAOJ Biotechnology. 2022; 6(1): 1003.

Introduction and related works

For the past few years, electronic health record (EHR) is undoubtedly one of the most important sources for data-driven patient care improvement approach. To harness the power of machine learning, EHR data can be used in many ways. However, EHR databases can be very large [1] and it might become very challenging to build mathematical models using conventional methods.

In the recent literature, many deep learning or artificial intelligence (AI) based models have been proposed to impute missing records [2] and predict the in-hospital mortality rate among critically ill patients [3]. However, obtaining a clinically relevant and implementable model is very difficult due to the complexity of human body and its different mechanisms. That is why many researchers have explored and compared multiple models [4] to assess the in-person mortality rate of hospitalized patients. It is also very important to preserve the privacy of those patients [5] while doing this type of research. So, there is a need for deidentified electronic health records [1].

While dealing with heart failure patients, it is found that a very common cause of death is acute myocardial infarction [7]. Although this study does not focus on the cause of death, it emphasizes the importance of the in-hospital mortality prediction for patients who have been diagnosed with heart failure. These patients might die because of myocardial infarction or other related conditions. In this study the patient cohort is selected based on ICD-9 code 4280 from MIMIC-III.

Some efforts have been found in the prediction of all-cause mortality [8]. But this would be even more challenging because all the systems in human body will have to be accounted for, which will be nearly impossible to build a model on.

In this era of modern technology, there is hardly a shortage of EHR data. But a study reports the methodology for predicting short-term mortality in acute heart-failure patients using minimal EHR data [9]. This kind of research is especially helpful if there is unavailability of good quality data from the EHR source.

A retrospective study attempts to reveal the underlying causes of mortality among heart failure patients using machine learning [10]. Like that study, there are many other studies [11] that report the mortality rate among heart failure patients using ML algorithms. A more recent article focuses on the mortality rate improvement for COVID-19 patients [12]. However, ML algorithms have been used not only for predicting heart failure but also managing it [13]. Recently, individual risk stratification strategy has been adopted for the prediction of worsening heart failure mortality of patients [14].

Overview of the machine learning algorithms

This section reviews some of the most popular machine learning methods that have been used in this study. All the methods are used for supervised learning purposes. For simplicity, the time variance [2] is not considered in this study, rather the last observed measurement for each patient is used as the imputed values.

Xgboost is the short form of extreme gradient boost, and it uses optimized distributed gradient boosting library to facilitate ML algorithms. Light GBM is the short form of light gradient boost machine that is usually faster and more accurate.

Catboost is another classifier that attempts to put more focus on categorical features than usual ML algorithms.

Ridge is a linear classifier that uses L2 regularization on the given data. Gaussian Naïve Bayes is a probabilistic method that relies on posterior distribution for prediction.

Figures 1 and 2 show the selected features and their imputed distributions, respectively.

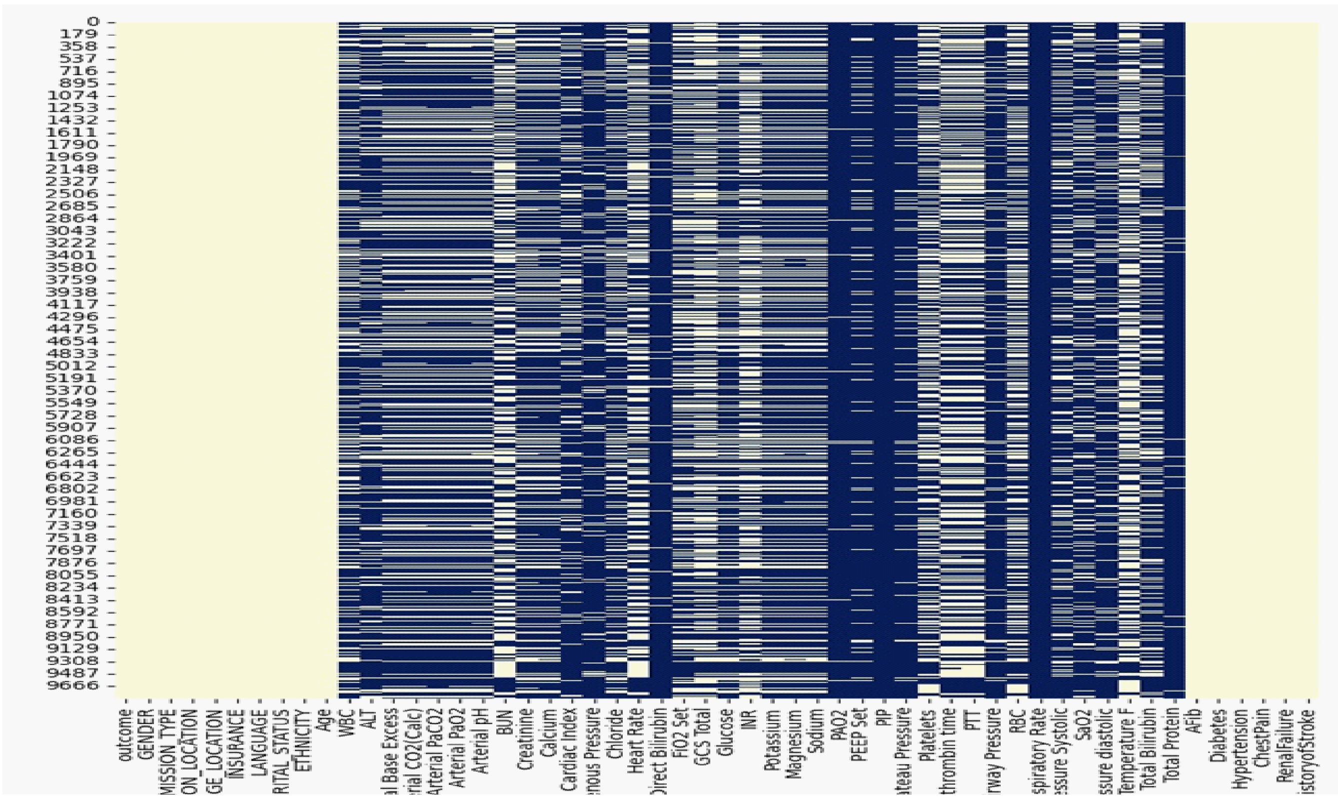

Figure 1: Selected features and their missing percentages.

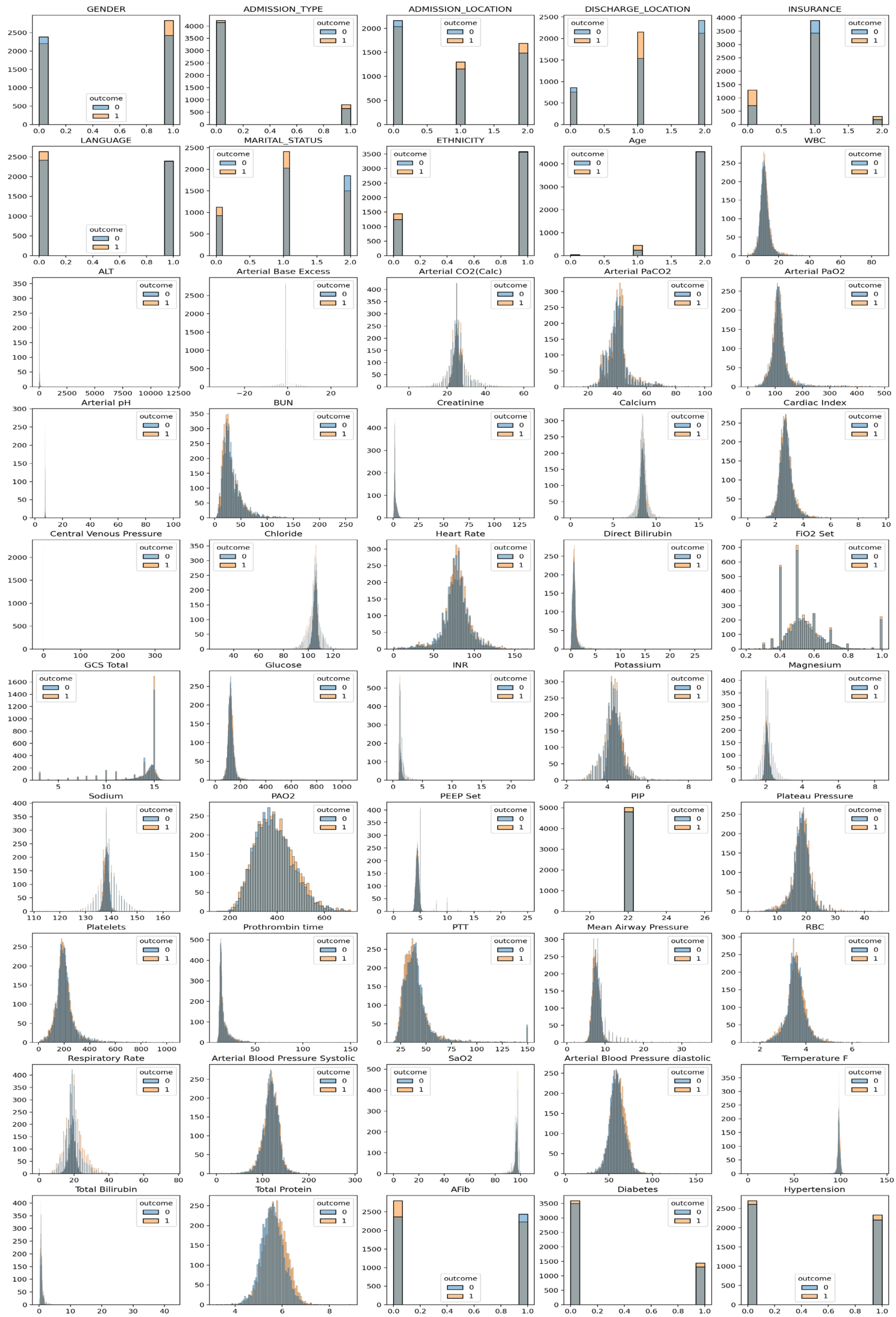

Figure 2: Distribution of features after imputation.

Results

This study reports the traditional F1 score is the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{t_p}{t_p + \frac{1}{2}(f_p + f_n)}$$

Here, tp = True positive (real: mortal, predict: mortal)

fp = False positive (real: non-mortal, predict: mortal)

The codes used in this study have been modified from this link.

Table 1 shows the comparison of performance for different ML algorithms that have been used in this study. From the comparison, Xgbost, light GBM and Catboost did well. However, these resultsmight not be clinically relevant yet.

Table 1 : Summary of performance for different ML algorithms.

Algorithm name	F1 score
Xgboost	0.63 ± 0.02
Light GBM	0.63 ± 0.03
CatBoost	0.62 ± 0.02
Gaussian NB	0.58 ± 0.02
Ridge	0.48 ± 0.01
SGD	0.43 ± 0.12

Conclusion

In this study, a comparison between different ML algorithms has been presented which shows that there is still scope for improvement in this area of research. An AI-powered tool is very much possible in near future with the help of these advanced algorithms. If the in-hospital mortality can be predicted with acceptable accuracy, it will greatly benefit the physicians and the hospital authority to improve healthcare and patient experience. Considering these factors and impacts, AI in healthcare is undoubtedly the new direction to follow.

References

1. Johnson, A.E.; Pollard, T.J.; Shen, L.; Li-Wei, H.L.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L.A.; Mark, R.G. MIMIC-III, a freely accessible critical care database. Scientific data 2016, 3, 1–9.
2. Zaman, M.A.U.; Du, D. A Stochastic Multivariate Irregularly Sampled Time Series Imputation Method for Electronic Health Records. BioMedInformatics 2021, 1, 166–181.
3. Kim, S.Y.; Kim, S.; Cho, J.; Kim, Y.S.; Sol, I.S.; Sung, Y.; Cho, I.; Park, M.; Jang, H.; Kim, Y.H.; et al. A deep learning model for real-time mortality prediction in critically ill children. Critical care 2019, 23, 1–10. 97
4. Lagu, T.; Pekow, P.S.; Shieh, M.S.; Stefan, M.; Pack, Q.R.; Kashef, M.A.; Atreya, A.R.; Valania, G.; Slawsky, M.T.; Lindenauer, P.K. Validation and comparison of seven mortality prediction models for hospitalized patients with acute decompensated heart failure. Circulation: Heart Failure 2016, 9, e002912. 100
5. Barish, M.; Bolourani, S.; Lau, L.F.; Shah, S.; Zanos, T.P. External validation demonstrates limited clinical utility of the interpretable mortality prediction model for patients with COVID-19. Nature Machine Intelligence 2021, 3, 25–27. 102
6. Kerkouche, R.; Acs, G.; Castelluccia, C.; Genevès, P. Privacy-preserving and bandwidth-efficient federated learning: An application to in-hospital mortality prediction. Proceedings of the Conference on Health, Inference, and Learning, 2021, pp. 25–35
7. Payrovnaziri, S.N.; Xing, A.; Salman, S.; Liu, X.; Bian, J.; He, Z. Assessing the Impact of Imputation on the Interpretations of Prediction Models: A Case Study on Mortality Prediction for Patients with Acute Myocardial Infarction. MedRxiv 2020. 107
8. Cerna, A.E.U.; Jing, L.; Good, C.W.; Raghunath, S.; Suever, J.D.; Nevius, C.D.; Wehner, G.J.; Hartzel, D.N.; Leader, J.B.; Alsaid, A.; et al. Deep-learning-assisted analysis of echocardiographic videos improves predictions of all-cause mortality. Nature Biomedical Engineering 2021, 5, 546–554. 110
9. Radhachandran, A.; Garikipati, A.; Zelin, N.S.; Pellegrini, E.; Ghandian, S.; Calvert, J.; Hoffman, J.; Mao, Q.; Das, R. Prediction of short-term mortality in acute heart failure patients using minimal electronic health record data. BioData mining 2021, 14, 1–15. 112
10. Lv, H.; Yang, X.; Wang, B.; Wang, S.; Du, X.; Tan, Q.; Hao, Z.; Liu, Y.; Yan, J.; Xia, Y. Machine Learning–Driven Models to Predict Prognostic Outcomes in Patients Hospitalized with Heart Failure Using Electronic Health Records: Retrospective Study. J Med Internet Res 2021, 23, e24996. doi:10.2196/24996. 115
11. Mpanya, D.; Celik, T.; Klug, E.; Ntsinjana, H. Predicting mortality and hospitalization in heart failure using machine learning: A systematic literature review. IJC Heart & Vasculature 2021, 34, 100773. 117
12. Vaid, A.; Jaladanki, S.K.; Xu, J.; Teng, S.; Kumar, A.; Lee, S.; Somani, S.; Paranjpe, I.; De Freitas, J.K.; Wanyan, T.; et al. Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach. JMIR Med Inform 2021, 9, e24207. doi:10.2196/24207. 120
13. Jing, L.; Ulloa Cerna, A.E.; Good, C.W.; Sauers, N.M.; Schneider, G.; Hartzel, D.N.; Leader, J.B.; Kirchner, H.L.; Hu, Y.; Riviello, D.M.; et al. A machine learning approach to management of heart failure populations. Heart Failure 2020, 8, 578–587. 122
14. Zile, M.R.; Koehler, J.; Sarkar, S.; Butler, J. Prediction of worsening heart failure events and all-cause mortality using an individualized risk stratification strategy. ESC Heart Failure 2020, 7, 4277–4289. 124