ECG-Based Predictive Models for Assessment of Volume Overload: Leveraging MIMIC-IV and OMOP-CDM Standardization

Seung Wook Lee, MD
Department of Medicine, Metrowest Medical Center, Framingham, MA, USA

Background

Volume overload, characterized by excess total body sodium and water, has significant clinical consequences such as increased cardiovascular morbidity, pulmonary congestion, and edema in individuals with conditions like chronic kidney disease (CKD), heart failure (HF), and end-stage renal disease (ESRD).(1, 2) Accurate volume status assessment is crucial but challenging. Conventional approaches, including clinical exams and biomarker monitoring, require clinical visits and evaluations. Techniques such as chest X-ray, echocardiography, and Point-of-Care Ultrasound (POCUS) for visualizing the inferior vena cava or internal jugular vein require involvement of imaging methodologies.(3-5).

Recent advancements in machine learning and big data analytics present potential improvements for volume status assessment. Utilizing electrocardiogram (ECG) data and demographic information from extensive healthcare databases can yield a more detailed understanding of volume overload and improve predictive accuracy. This study aims to develop an objective measure for assessing volume overload using the MIMIC-IV (Medical Information Mart for Intensive Care IV) database, standardized according to OMOP-CDM (Observational Medical Outcomes Partnership Common Data Model) guidelines. Throughout this study, we aim to demonstrate the application of standardized health data for prediction of relevant clinical condition.

Methods

We utilized the MIMIC-IV and MIMIC-IV-ECG datasets from PhysioNet, which include comprehensive records of 299,712 patients admitted to an intensive care unit (ICU) or the emergency department (ED) of Beth Israel Deaconess Medical Center (BIDMC) from 2008 to 2019.(6-9) Information including demographics, comorbidities, and ECG data upon ED admission were collected. Data were standardized to the OMOP-CDM using an ETL (Extract, Transform, Load) process defined by a previous research group to ensure consistent data representation and analysis.(10) Custom vocabularies were utilized to describe ECG features not covered by existing standards.

Patients included in the study were those who had an ECG performed within six hours of their ED admission. Among these patients, those who received diuretics within six hours were labeled as positive cases. Patients whose ECGs were taken after the administration of diuretics were excluded from the study, and for patients who had multiple admissions, only the first admission was considered.

An XGBoost model was developed using demographic data, ECG-derived features (e.g., QT interval, p-onset, T axis), and patient comorbidities (e.g., history of CHF, atrial fibrillation). Additionally, a 1D Convolutional Neural Network (CNN) was incorporated to analyze waveforms from the 12-lead ECGs. Each ECG waveform, consisting of 10-second recordings sampled at 500 samples per second, was processed. The two distinct approaches aimed to leverage both structured tabular data and raw ECG waveform data to predict the need for diuretic administration.

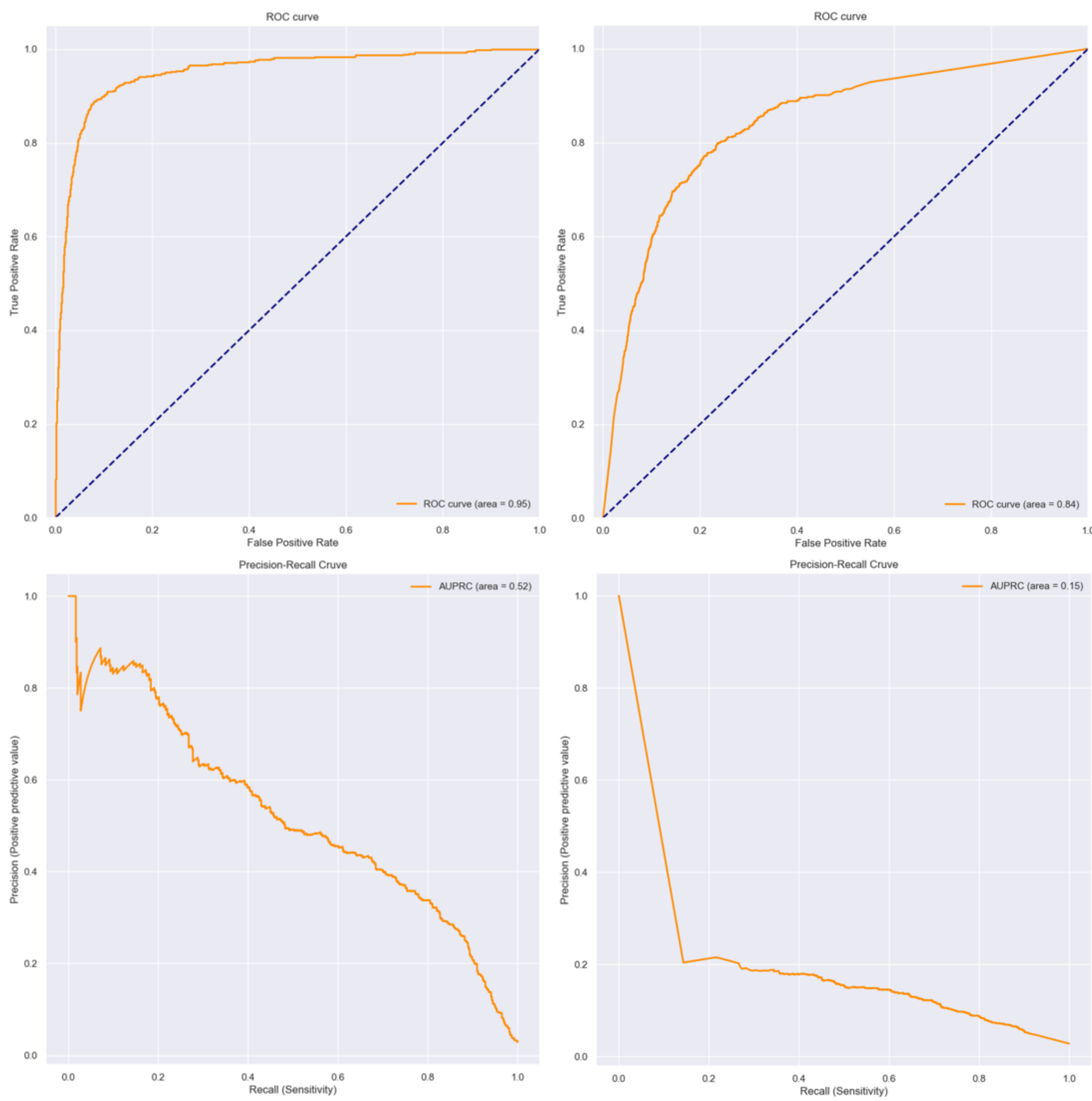

Figure 1: ROC (top) and PRC (bottom) for XGBoost model (Left) and 1D CNN model (Right)

Results

In total, 96,422 subjects were included in the study. The dataset was highly imbalanced, with only 3% of subjects requiring diuretics. Our analysis identified key ECG features and demographic factors that correlate with the use of diuretics within 6 hours of ED admission. The XGBoost model achieved an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.95 and an Area Under the Precision-Recall Curve (AUPRC) of 0.52 (Figure 1). The most significant predictors included CHF, T axis, RR interval, history of atrial fibrillation, and QRS onset. The 1D CNN model yielded an AUROC of 0.84 and an AUPRC of 0.15, reflecting challenges in precision and recall due to the imbalanced dataset.

Conclusion

This study introduces an innovative, automated approach to assessing volume status and predicting diuretic use by utilizing ECG data and demographic information from the MIMIC-IV database, standardized according to OMOP-CDM guidelines. Our model provides a reliable tool for monitoring and determining volume overload, demonstrating the potential of advanced machine learning to enhance clinical decision-making and improve outcomes for patients at risk of volume overload, such as those with HF or CKD. By relying on non-invasive ECG data and demographic information, this method reduces the need for clinician intervention in assessing volume overload.

Future directions for the study include concentrating on populations with CHF or CKD to gain a deeper understanding of the predictors of volume overload in these high-risk groups. Additionally, we plan to adapt this approach for use with single-lead ECG signals, which can be monitored with wearable devices at home, thus enabling the identification of volume overload in home settings. This research highlights the importance of standardized health data and diverse data integration in tackling complex clinical challenges. By focusing on specific patient populations and incorporating wearable technology, we aim to further enhance the clinical applicability and accuracy of our predictive models.

REFERENCES

1. Hung SC, Kuo KL, Peng CH, Wu CH, Lien YC, Wang YC, et al. Volume overload correlates with cardiovascular risk factors in patients with chronic kidney disease. Kidney Int. 2014;85(3):703-9.
2. Hung SC, Lai YS, Kuo KL, Tarng DC. Volume overload and adverse outcomes in chronic kidney disease: clinical observational and animal studies. J Am Heart Assoc. 2015;4(5).
3. Bonow RO, Carabello BA, Chatterjee K, de Leon AC, Jr., Faxon DP, Freed MD, et al. 2008 Focused update incorporated into the ACC/AHA 2006 guidelines for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 1998 Guidelines for the Management of Patients With Valvular Heart Disease): endorsed by the Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. Circulation. 2008;118(15):e523-661.
4. Alexandrou ME, Balafa O, Sarafidis P. Assessment of Hydration Status in Peritoneal Dialysis Patients: Validity, Prognostic Value, Strengths, and Limitations of Available Techniques. Am J Nephrol. 2020;51(8):589-612.
5. Vaidya GN, Kolodziej A, Stoner B, Galaviz JV, Cao X, Heier K, et al. Bedside ultrasound of the internal jugular vein to assess fluid status and right ventricular function: The POCUS-JVD study. Am J Emerg Med. 2023;70:151-6.
6. Johnson A, Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. MIMIC-IV (version 2.2). PhysioNet. 2023.
7. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1.
8. Gow B, Pollard, T., Nathanson, L. A., Johnson, A., Moody, B., Fernandes, C., Greenbaum, N., Waks, J. W., Eslami, P., Carbonati, T., Chaudhari, A., Herbst, E., Moukheiber, D., Berkowitz, S., Mark, R., & Horng, S. MIMIC-IV-ECG: Diagnostic Electrocardiogram Matched Subset (version 1.0). In: PhysioNet, editor. 2023.
9. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):E215-20.
10. Kallfelz M, Tsvetkova, A., Pollard, T., Kwong, M., Lipori, G., Huser, V., Osborn, J., Hao, S., & Williams, A. MIMIC-IV demo data in the OMOP Common Data Model (version 0.9). In: PhysioNet, editor. 2021.