

A Fusion Method Combining LSTM and MLP for Predicting Mortality Risk in Critically Ill Coronary Artery Disease Patients

Rongren Xu

School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
xurongren@mail.ustc.edu.cn

Haoran Zheng*

School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
*hrzheng@ustc.edu.cn

Abstract—Coronary artery disease (CAD) is one of the leading causes of death worldwide, and assessment of its prognosis remains challenging, especially in the intensive care unit (ICU). Existing tools for assessment often rely on static data, which cannot capture the dynamic physiological changes that are critical for timely risk stratification. To overcome this limitation, we extracted relevant patient data from the MIMIC-IV database and proposed a novel machine learning model to predict the 30-day mortality risk of patients. We constructed an end-to-end fusion model that uses Long Short-Term Memory (LSTM) networks to process time-series clinical data while using Multilayer Perceptron (MLP) to process static baseline features. The model's performance was primarily evaluated using the Area Under the Precision-Recall Curve (AU-PRC) and compared with traditional clinical scoring systems (SOFA, SAPS II, LODS, OASIS) and conventional machine learning models (logistic regression, random forest, LightGBM and XGBoost). The proposed fusion model achieved an AU-PRC of 0.5816 on the test set, significantly outperforming established clinical scoring systems such as SAPS II (AU-PRC = 0.3127) and the strongest traditional machine learning model, LightGBM (AU-PRC = 0.5614). Our findings suggest that by integrating patients' dynamic data, this fusion model provides a more powerful tool for predicting mortality risk in critically ill patients with CAD, potentially improving clinical decision-making quality by identifying high-risk individuals earlier.

Keywords—Coronary Artery Disease, MIMIC-IV, Mortality Prediction, LSTM.

I. INTRODUCTION

Coronary artery disease (CAD) is a leading cause of morbidity and mortality worldwide. In the intensive care unit (ICU), its acute presentation, acute coronary syndrome (ACS), has a particularly high mortality rate, with an in-hospital mortality rate exceeding 5% and reaching as high as 26.7% in certain subgroups [1]–[3]. A significant challenge in managing these patients is the accurate assessment of their prognosis. While numerous risk factors, from demographic data to biomarkers such as the triglyceride-glucose (TyG) index, hemoglobin glycation index (HGI), and neutrophil-to-lymphocyte ratio (NLR), have been identified [4]–[8], effectively integrating these scattered static indicators with patients' fluctuating physiological states remains a major obstacle in clinical practice.

For many years, clinical decision-making has relied on traditional scoring systems such as GRACE, TIMI, and SOFA [9]–[12]. While these tools are valuable, their limitations are increasingly apparent: they are often based on a limited set of

variables, assume linear relationships, and use only static patient characteristics, that cannot capture dynamic changes in a patient's condition. This static nature often results in suboptimal predictive performance in heterogeneous populations [13].

Machine learning has emerged as a powerful alternative, capable of analyzing high-dimensional data and modeling complex non-linear relationships. A growing body of research demonstrates that models like LightGBM and XGBoost can significantly outperform traditional scores in predicting adverse outcomes for CAD patients, achieving AUCs upwards of 0.83 [14]–[17]. However, a common limitation persists in these approaches: they typically process dynamic time-series data (e.g., continuous vital signs) by converting them into aggregated static features (e.g., mean, maximum). This simplification discards temporal information and trends that are crucial for reflecting real-time changes in a patient's state.

To overcome this challenge, deep learning models, like Recurrent Neural Networks (RNNs), have emerged as a powerful tool for modeling time series data. Researchers have applied architectures like LSTMs and TCNs to predict mortality in general ICU populations, achieving promising results [18], [19]. Despite this progress, a critical research gap remains: there is a lack of models specifically designed for patients with CAD that can effectively fuse dynamic time-series data with static baseline features in an end-to-end model to maximize predictive power. Currently, most deep learning applications do not specifically target the population with CAD but instead focus on the general ICU patient population or other specific conditions such as sepsis [20].

To bridge this gap, we propose a novel, end-to-end deep learning model to predict 30-day mortality of CAD patients in the ICU. We designed a fusion architecture that integrates a Long Short-Term Memory (LSTM) network with a Multi-Layer Perceptron (MLP). This architecture simultaneously processes dynamic data from the first 24 hours after ICU admission and static baseline features. By capturing temporal patterns through the LSTM and integrating baseline features through the MLP, our model learned a more comprehensive representation of patient risk, thereby improving the accuracy of mortality prediction in these critical patients.

II. MATERIALS AND METHODS

A. Data Source

In this study, data were extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database

(version 3.1). The database contains de-identified health data from patients admitted to Beth Israel Deaconess Medical Center (BIDMC) between 2008 and 2022 [21],[22]. Access to the database was granted to author Rongren Xu after completing the "Protection of Human Research Participants" examination (Record ID: 58289793). The establishment of the MIMIC-IV database was approved by the Institutional Review Boards of the Massachusetts Institute of Technology (MIT) and BIDMC. The requirement for individual patient consent was waived for this study because all data were anonymized.

B. Population and Cohort Selection

We identified adult patients (age ≥ 18 years) diagnosed with CAD using the International Classification of Diseases, Ninth and Tenth Revisions (ICD-9 and ICD-10) codes. The specific codes included were 410.xx-414.xx and 429.2 for ICD-9, and I20.x-I25.x for ICD-10. The inclusion criteria were: (1) an ICU stay of at least 24 hours and (2) age between 18 and 89 years. Exclusion criteria were: (1) age over 89, as the database had adjusts the ages of such patients; or (2) multiple ICU admissions for CAD, in which case only the final admission was retained to ensure the independence of each patient. a total of 18,486 patients were included in this study after the above screening process. The patient selection flowchart is shown in Figure 1. The primary endpoint of this study was all-cause mortality within 30 days following ICU admission.

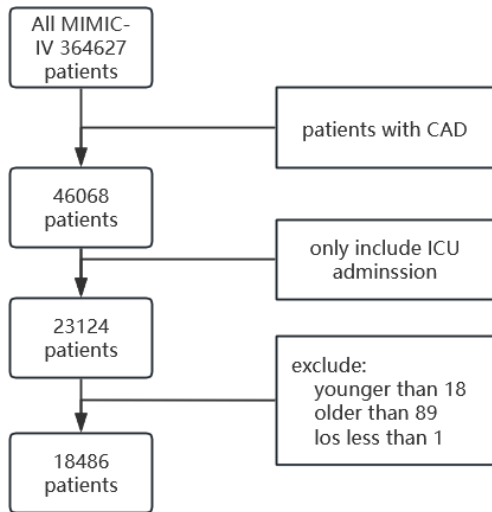


Figure 1 Flowchart of patient selection.

C. Feature Extraction and Preprocessing

Data were extracted from the first 24 hours of each patient's ICU stay using PostgreSQL 17. The features were categorized into static baseline features and time-series features.

Static Features included demographic data (age, gender, ethnicity, marital status), insurance status, and comorbidities. To ensure feature robustness, features with a missing rate over 10% were removed. Categorical variables were processed through meaningful grouping (e.g., merging admission types, see Table 1) and then converted into numerical format using one-hot encoding. This initial processing resulted in 44 static features.

Table 1 The merging rules for the type of initial ICU admission.

Original value	Merged value
Cardiac Vascular Intensive Care Unit (CVICU), Coronary Care Unit (CCU)	Cardiac
Medical Intensive Care Unit (MICU)	MICU
Medical/Surgical Intensive Care Unit (MICU/SICU)	MICU/SICU
Surgical Intensive Care Unit (SICU), Trauma SICU (TSICU)	SICU
Neuro Intermediate, Neuro Surgical Intensive Care Unit (Neuro SICU), Neuro Stepdown	Neuro

Time-Series Features included physiological measurements (such as heart rate and blood pressure) and laboratory results (such as hemoglobin and creatinine). We standardized the data to hourly intervals due to differences in measurement frequency. A feature was retained as a high-frequency dynamic feature if it had measurements in at least five separate hourly intervals in more than 70% of the patients. Features not meeting this criterion were considered low-frequency; their mean values over the 24-hour period were calculated and treated as static features to mitigate data sparsity while retaining their information. This process yielded seven high-frequency dynamic features and converted 21 low-frequency features into static representations.

Final Feature Set and Imputation: The 21 converted static features were combined with the original 44, resulting in a total of 65 static features. For the 7 retained dynamic features, missing values were first imputed using forward and backward filling. Any remaining missing values were imputed using the median value from the entire training set.

D. Model Architecture

To effectively integrate both data modalities, we propose an end-to-end deep learning model composed of an LSTM module for dynamic data and an MLP module for static data. The model architecture is shown in Figure 2.

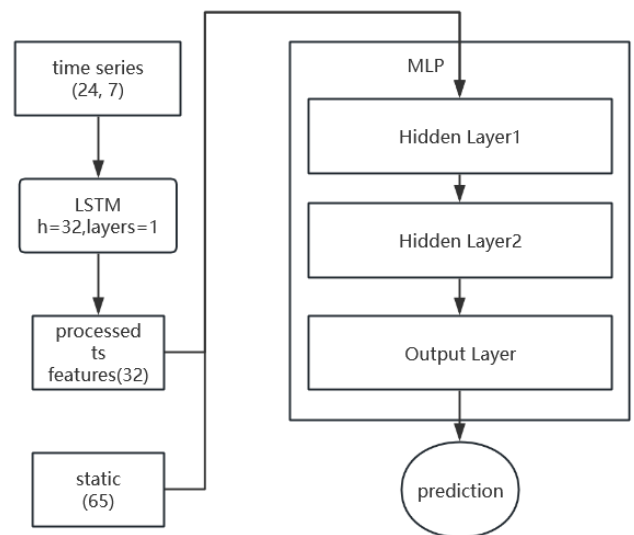


Figure 2 Model architecture.

The model takes two parts of inputs:

Time series input (x_{ts}): A tensor of shape $(N, 24, 7)$, representing 24 hourly measurements for the 7 high-frequency dynamic features.

Static input (x_{static}): A vector of 65 baseline features.

The time-series input is fed into a single-layer LSTM network with 32 hidden units. The final hidden state of the LSTM is a 32-dimensional vector representing the temporal dynamics. This vector is then concatenated with the 65 static features to form a comprehensive fused feature vector. This vector is passed to an MLP module. This module consists of two hidden layers (128 neurons for each, with ReLU activation) and a dropout rate of 0.45 after each layer to prevent overfitting. The final output layer is a single neuron producing a prediction score for mortality risk.

E. Experimental Setup and Evaluation

We benchmarked our proposed LSTM-MLP model against two categories of baseline models: (1) clinical scoring systems, including SOFA, SAPS II, OASIS and LODS; and (2) traditional machine learning models, including Logistic Regression (LR), Random Forest (RF), XGBoost and LightGBM.

Considering the imbalanced distribution of the dataset (the number of survivors is approximately six times that of non-survivors), we chose Area Under the Precision-Recall Curve (AU-PRC) as the primary evaluation metric because AU-PRC provides a more informative assessment on imbalanced classification tasks than Area Under the Receiver Operating Characteristic Curve (AU-ROC). We additionally reported AU-ROC, accuracy, precision, recall and F1-Score for a comprehensive comparison.

III. RESULTS

A. Baseline characteristics

We compared some baseline characteristics of patients with different outcomes, and the results are shown in Table 2.

The final study cohort comprised 18,486 patients with CAD, who were stratified into a 30-day mortality group ($N = 2412$, 13.05%) and a survival group ($N = 16074$, 86.95%). The detailed baseline characteristics of both groups are presented in Table 2.

Table 2 Baseline characteristics of patients according to survival or dead.

Feature	Death(N=2412)	Survival(N=16074)	All(N=18486)
Age	75.97±10.73	70.37±11.45	71.05±11.51
Weight	79.51±21.47	86.19±21.27	85.38±21.41
Hadmicc	8.00±2.76	5.62±2.66	5.91±2.78
Sapsii	49.55±14.61	37.34±11.71	38.82±12.74
Oasis	37.77±8.90	31.45±7.83	32.21±8.23
Lods	6.96±3.22	4.47±2.58	4.77±2.79
Ventstatus	2.37±1.95	2.12±1.98	2.15±1.98
Gender			
Male	1505 (62.4%)	11063 (68.8%)	12568 (68.0%)
Female	907 (37.6%)	5011 (31.2%)	5918 (32.0%)

Several differences were observed between two groups. Patients in the death group were, on average, older and had a lower body weight compared to the survival group. Demographically, the percentage of female patients was higher in the death group than that in survival group. Clinically, patients in the death group underwent mechanical ventilation more frequently.

B. Model Performance

The predictive performance of the proposed LSTM-MLP fusion model and all baseline models was evaluated on the independent test set, with the results summarized in Table 3.

The fusion model we proposed demonstrated the highest performance for predicting 30-day mortality risk. This model achieved a primary evaluation metric, the AU-PRC, of 0.5816. This performance significantly surpassed that of traditional clinical scoring systems, with SAPS II yielding an AU-PRC of only 0.3127. Moreover, our model also outperformed all traditional machine learning baseline models, with the best performing model, LightGBM, achieving an AU-PRC of 0.5614.

The fusion model also achieved the highest AU-ROC of 0.8995, consistent with the AU-PRC results, compared to 0.8934 for LightGBM and 0.7505 for SAPS II. Table 3 provides a comprehensive comparison across all evaluated metrics, including. Our proposed model achieved the highest values in half of the six metrics, with the remaining three metrics only slightly lower than one or two others. These findings collectively highlight the superior performance of the proposed fusion model in leveraging both dynamic and static clinical data.

Table 3 Model performance.

Model	AU-PRC	AU-ROC	Accuracy	Precision	Recall	F1-Score
Clinic scoring systems						
SOFA	0.2443	0.6485	0.7257	0.2132	0.4680	0.2929
SAPSII	0.3127	0.7505	0.7933	0.2969	0.5135	0.3762
OASIS	0.2757	0.7004	0.8129	0.2905	0.3749	0.3274
LODS	0.2971	0.7291	0.7723	0.2766	0.5415	0.3662
Static only						
LR	0.4470	0.8513	0.8304	0.3900	0.6645	0.4915
RF	0.4740	0.8637	0.8469	0.4196	0.6294	0.5035
Light-GBM	0.4606	0.8579	0.8426	0.4092	0.6228	0.4939
XG-Boost	0.4542	0.8545	0.8388	0.4030	0.6382	0.4941
high-freq time series only						
LSTM	0.3782	0.7963	0.7793	0.3056	0.6206	0.4096
static and time series as static						
LR	0.5237	0.8843	0.8634	0.4621	0.6557	0.5422
RF	0.5509	0.8922	0.8661	0.4683	0.6316	0.5378
Light-GBM	0.5562	0.8934	0.8856	0.5342	0.5658	0.5495
XG-Boost	0.5294	0.8924	0.8653	0.4668	0.6469	0.5423
fused static and time series(end-to-end)						
LSTM-MLP	0.5816	0.8995	0.8818	0.5168	0.6404	0.5720

IV. DISCUSSION

In this study, we have proposed an end-to-end model integrating LSTM and MLP to predict 30-day mortality risk in critically ill patients with CAD. Our results show that combining dynamic time-series data with static baseline features allows the model to achieve a marked improvement in prediction, outperforming traditional approaches, particularly on the AU-PRC metric.

The performance metrics of the model highlight its potential clinical applications.. The AU-PRC of 0.5816 demonstrates its robustness in an imbalanced scenario, and the AU-ROC of 0.8995 illustrates its excellent overall ability to discriminate between high-risk and low-risk patients. Notably, the model achieved a recall of 0.6404. This is clinically significant as it signifies the model's capacity to identify nearly two-thirds of patients who will actually experience a fatal outcome, thus providing a critical window for clinical interventions. While a precision of 0.5168 suggests room for improvement, the model's utility as a screening tool remains high and can be optimized by adjusting decision thresholds based on specific clinical needs.

The superiority of our fusion model is clear when contrasted with the baseline methods. Traditional scoring systems like SAPS II as well as advanced machine learning models like LightGBM were significantly outperformed. This gap highlights a fundamental limitation of the baseline methods: by aggregating time-series features into static representations, they discard the temporal dependencies and dynamic trends in the process of a patient's physiological changes. In contrast, our LSTM-based model is specifically designed to capture and utilize these complex patterns to better effect. This suggests that capturing changes in a patient's condition, rather than just a snapshot of his or her state at a given moment or an aggregated state over a period of time, is critical for accurate mortality prediction.

A. Strengths

A primary strength of this study is its novel end-to-end fusion architecture. This architecture outperforms other methods that use only a single data modality or employ simple feature aggregation methods. Secondly, our use of Optuna for hyperparameter optimization to ensure that the reported performance is robust and approximately optimal under the selected architecture. Lastly, we use AU-PRC as the primary evaluation metric, which provides a more reliable performance assessment for this imbalanced classification task.

B. Limitations

Although the study results are encouraging, this study still has several limitations. First, as a single-center study based on the MIMIC-IV database, the generalizability of this model needs to be validated using external, especially multi-center, datasets. Second, this is a retrospective study, meaning it may be subject to potential biases arising from data quality and recording practices. Third, the inherent "black-box" nature of deep learning models limits clinical interpretability, which can be a barrier to clinical application. Finally, our use of a 24-hour data window may omit critical information present before or after this time period.

C. Perspectives

Our future work will focus on addressing these limitations. First, we will conduct rigorous external validation across a diverse patient population. Second, we will explore more advanced architectures, such as introducing attention mechanisms to enhance feature and temporal importance weights. Third, we will deploy interpretability techniques (such as SHAP or LIME) to explain our prediction results, thereby improving model transparency. Finally, we will attempt to develop personalized models for specific patient subgroups.

V. CONCLUSION

This study proposed a deep learning model that fuses LSTM and MLP networks to predict 30-day mortality risk of patients admitted in ICU with CAD. By integrating dynamic time-series features and static baseline features, our end-to-end model achieved AU-PRC of 0.5816 and AU-ROC of 0.8995 on the independent test set, significantly outperforming both traditional clinical scoring systems and machine learning models.

These results suggest that accurately capturing the changing of a patient's conditions is critical for precise risk prediction. Despite the limitations of single-center data and the need for improved interpretability, the superior discriminatory ability of the model proposed in this study over the baseline makes it a promising clinical decision support tool that can help clinicians identify high-risk individuals early on, thereby guiding timely intervention.

ACKNOWLEDGMENT

This work was supported by the National Key Technologies R&D Program of China [2017YFA0505502] and the Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38000000].

REFERENCES

- [1] X. Zhang, X. Wang, L. Xu, J. Liu, P. Ren, and H. Wu, "The predictive value of machine learning for mortality risk in patients with acute coronary syndromes: A systematic review and meta-analysis," *European Journal of Medical Research*, vol. 28, no. 1, p. 451, Oct. 20, 2023, issn: 2047-783X. doi: 10.1186/s40001-023-01027-4.
- [2] M. Y. Chan, X. Du, D. Eccleston, et al., "Acute coronary syndrome in the asia-pacific region," *International Journal of Cardiology*, vol. 202, pp. 861–869, Jan. 1, 2016, issn: 0167-5273. doi: 10.1016/j.ijcard.2015.04.073.
- [3] A. Ariza-Solé, O. Alegre, F. J. Elola, et al., "Management of myocardial infarction in the elderly. insights from spanish minimum basic data set," *European Heart Journal: Acute Cardiovascular Care*, vol. 8, no. 3, pp. 242–251, Apr. 2019, Publisher: Oxford University Press (OUP), issn: 2048-8726, 2048-8734. doi: 10.1177/2048872617719651.
- [4] R. Nicoll, U. Wiklund, Y. Zhao, et al., "Gender and age effects on risk factor-based prediction of coronary artery calcium in symptomatic patients: A euro-CCAD study," *Atherosclerosis*, vol. 252, pp. 32–39, Sep. 2016, issn: 00219150. doi: 10.1016/j.atherosclerosis.2016.07.906.
- [5] X. Wei, X. Chen, Z. Zhang, et al., "Risk analysis of the association between different hemoglobin glycation index and poor prognosis in critical patients with coronary heart disease-a study based on the MIMIC-IV database," *Cardiovascular Diabetology*, vol. 23, no. 1, p. 113, Mar. 30, 2024, issn: 1475-2840. doi: 10.1186/s12933-024-02206-1.
- [6] D. Chen, B. Hu, X.-h. Chen, X. Wei, J. Feng, and Z.-p. Hu, "Association between different hemoglobin glycation index and prognosis in patients with a first diagnosis of acute myocardial infarction: A retrospective study based on the MIMIC-IV database," *Frontiers in Cardiovascular Medicine*, vol. 12, p. 1 447 420, May 26, 2025, issn: 2297-055X. doi: 10.3389/fcvm.2025.1447420.

- [7] R. Zhang, S. Shi, W. Chen, et al., "Independent effects of the triglyceride-glucose index on all-cause mortality in critically ill patients with coronary heart disease: Analysis of the MIMIC-III database," *Cardiovascular Diabetology*, vol. 22, no. 1, p. 10, Jan. 13, 2023, issn: 1475-2840. doi: 10.1186/s12933-023-01737-3.
- [8] F. Cao, J.-j. Jiang, G. Zhang, et al., "Prognostic value of inflammatory markers for all- cause mortality in patients with acute myocardial infarction in the coronary care unit: A retrospective study based on MIMIC-IV database," *Frontiers in Cardiovascular Medicine*, vol. 12, p. 1 439 650, Jan. 22, 2025, issn: 2297-055X. doi: 10.3389/fcvm.2025.1439650.
- [9] J. -L. Vincent, R. Moreno, J. Takala, et al., "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," *Intensive Care Medicine*, vol. 22, no. 7, pp. 707–710, Jul. 1996, doi: 10.1007/BF01709751.
- [10] E. M. Antman, M. Cohen, P. J. L. M. Bernink, et al., "The TIMI risk score for unstable angina/non–ST elevation MI: A method for prognostication and therapeutic decision making," *JAMA*, vol. 284, no. 7, p. 835, Aug. 16, 2000, issn: 0098-7484. doi: 10.1001/jama.284.7.835.
- [11] F. A. Wenzl, S. Kraler, G. Ambler, et al., "Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and switzerland: A multinational analysis with external cohort validation," *The Lancet*, vol. 400, no. 10354, pp. 744–756, Sep. 2022, issn: 01406736. doi: 10.1016/S0140- 6736(22)01483-0.
- [12] L. Wang, Z. Zhang, and T. Hu, "Effectiveness of LODS, OASIS, and SAPS II to predict in-hospital mortality for intensive care patients with ST elevation myocardial infarction," *Scientific Reports*, vol. 11, no. 1, p. 23 887, Dec. 13, 2021, issn: 2045-2322. doi: 10.1038/s41598-021-03397-3.
- [13] F. Muşat, D. N. Păduraru, A. Bolocan, et al., "Machine learning models in sepsis outcome prediction for ICU patients: Integrating routine laboratory tests—a systematic review," *Biomedicines*, vol. 12, no. 12, p. 2892, Dec. 19, 2024, issn: 2227-9059. doi: 10 . 3390 / biomedicines12122892.
- [14] H. Yang, Z. Chen, H. Yang, and M. Tian, "Predicting coronary heart disease using an improved LightGBM model: Performance analysis and comparison," *IEEE Access*, vol. 11, pp. 23 366–23 380, 2023, issn: 2169-3536. doi: 10.1109/ACCESS.2023.3253885.
- [15] X. Lin, X. Pan, Y. Yang, et al., "Machine learning models to predict 30-day mortality for critical patients with myocardial infarction: A retrospective analysis from MIMIC-IV database," *Frontiers in Cardiovascular Medicine*, vol. 11, p. 1 368 022, Sep. 20, 2024, issn: 2297-055X. doi: 10.3389/fcvm.2024.1368022.
- [16] S. Peng, Q. Chen, W. Ke, and Y. Wu, "Predicting in-hospital mortality in myocardial in- farction: A nomogram-based retrospective analysis of the MIMIC-IV database," *Vascular Health and Risk Management*, vol. Volume 21, pp. 461–476, Jun. 2025, issn: 1178-2048. doi: 10.2147/VHRM.S511277.
- [17] J. Li, Y. Sun, J. Ren, Y. Wu, and Z. He, "Machine learning for in-hospital mortality prediction in critically ill patients with acute heart failure: A retrospective analysis based on the MIMIC-IV database," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 39, no. 3, pp. 666–674, Mar. 2025, issn: 10530770. doi: 10.1053/j.jvca.2024.12.016.
- [18] Y.-w. Chen, Y.-j. Li, P. Deng, et al., "Learning to predict in-hospital mortality risk in the intensive care unit with attention-based temporal convolution network," *BMC Anes- thesiology*, vol. 22, no. 1, p. 119, Dec. 2022, issn: 1471-2253. doi: 10.1186/s12871-022- 01625-5.
- [19] K. Niu, K. Zhang, X. Peng, Y. Pan, and N. Xiao, "Deep multi-modal intermediate fusion of clinical record and time series data in mortality prediction," *Frontiers in Molecular Biosciences*, vol. 10, p. 1 136 071, Mar. 8, 2023, issn: 2296-889X. doi: 10.3389/fmolb. 2023.1136071.
- [20] S. Dalal, A. K. Ardabili, and A. S. Bonavia, "Time-series deep learning and conformal pre- diction for improved sepsis diagnosis in primarily non-ICU hospitalized patients," *Com- puters in Biology and Medicine*, vol. 193, p. 110 497, Jul. 2025, issn: 00104825. doi: 10.1016/j.combiomed.2025.110497.
- [21] A. E. Johnson, L. Bulgarelli, L. Shen, et al., "Mimic-iv, a freely accessible electronic health record dataset," *Scientific data*, vol. 10, no. 1, p. 1, 2023.
- [22] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. A. Celi, and R. Mark, "Mimic-iv," *PhysioNet*, pp. 49–55, 2024.