Serial 12-Lead ECG–Based Deep-Learning Model for Hospital Admission Prediction in Emergency Department Cardiac Presentations: Retrospective Cohort Study

Arda Altintepe, Kutsev Bengisu Ozyoruk

Submitted to: JMIR Cardio
on: August 25, 2025

Table of Contents

Serial 12-Lead ECG–Based Deep-Learning Model for Hospital Admission Prediction in Emergency Department Cardiac Presentations: Retrospective Cohort Study

Arda Altintepe[1]; Kutsev Bengisu Ozyoruk[2] PhD

[1] Horace Mann School New York US
[2] Emory University Atlanta US

Corresponding Author:
Kutsev Bengisu Ozyoruk PhD

Emory University
Health Sciences Research Building II (HSRB II)
1760 Haygood Dr NE Suite N647
Atlanta
US

Abstract

Background: Emergency Department (ED) crowding is often attributed to a slow hospitalization process, leading to reduced quality of care. Predicting early disposition with cardiac-presenting patients is challenging: most are ultimately discharged, yet those with a cardiac etiology frequently require hospital admission. Existing scores rely on single-time-point data and often underperform when patient risk evolves during the visit.

Objective: To develop and validate a real-time deep-learning model that fuses serial 12-lead electrocardiogram (ECG) waveforms with sequential vitals and routinely available clinical data to predict hospital admission early in ED encounters.

Methods: We conducted a retrospective cohort study using the MIMIC-IV, MIMIC-IV-ED, and MIMIC-IV-ECG databases. Adults presenting with chest pain, dyspnea, syncope, or presyncope and at least one ECG within their ED stay were included. Two evaluation cohorts were defined: all stays with ?1 ECG (N=30,421) and a subset with ?2 ECGs during the encounter (N=11,273). To predict hospital admission, we first established two baseline models: a tabular model (random forest) trained on structured clinical variables including demographics, triage acuity, past medical history, medications, and laboratory results, and an ECG-only model that learned directly from raw 12-lead waveforms. We then developed a multimodal deep-learning model that combined ECGs with sequential vital signs as well as the same static tabular features. All models were restricted to data available during the stay up to the time of the last ECG. Performance was assessed with stratified 5-fold cross-validation using identical splits across models.

Results: The multimodal model achieved an Area Under Receiver Operating Characteristic (AUROC) of 0.911 when trained on all eligible stays. The model predicted disposition after the final ECG was taken, which was a median of 0.3 hours after triage and 4.6 hours before ED departure. Baseline models performed worse: the ECG-only model had an AUROC of 0.852, and the tabular random forest had an AUROC of 0.886. In the subset requiring at least two ECGs within the stay, ECG-only reached an AUROC of 0.859, and random forest, with the longer interval to chart tabular data, reached a higher AUROC of 0.911. The multimodal model had AUROC 0.924, and outperformed baselines in each cohort (paired DeLong P<.001).

Conclusions: Serial ECGs, when integrated with evolving vitals and routine clinical features, enable accurate, early prediction of ED disposition in cardiac-presenting patients. This open-source, reproducible framework highlights the potential of multimodal deep learning to streamline ED flow, prioritize higher-risk cases, and detect evolving, time-critical pathology.

(JMIR Preprints 25/08/2025:80569)
DOI: https://doi.org/10.2196/preprints.80569

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Serial 12-Lead ECG–Based Deep-Learning Model for Hospital Admission Prediction in Emergency Department Cardiac Presentations: Retrospective Cohort Study

Abstract

Background: Emergency Department (ED) crowding is often attributed to a slow hospitalization process, leading to reduced quality of care. Predicting early disposition with cardiac-presenting patients is challenging: most are ultimately discharged, yet those with a cardiac etiology frequently require hospital admission. Existing scores rely on single-time-point data and often underperform when patient risk evolves during the visit.

Objective: To develop and validate a real-time deep-learning model that fuses serial 12-lead electrocardiogram (ECG) waveforms with sequential vitals and routinely available clinical data to predict hospital admission early in ED encounters.

Methods: We conducted a retrospective cohort study using the MIMIC-IV, MIMIC-IV-ED, and MIMIC-IV-ECG databases. Adults presenting with chest pain, dyspnea, syncope, or presyncope and at least one ECG within their ED stay were included. Two evaluation cohorts were defined: all stays with ≥1 ECG (N=30,421) and a subset with ≥2 ECGs during the encounter (N=11,273). To predict hospital admission, we first established two baseline models: a tabular model (random forest) trained on structured clinical variables including demographics, triage acuity, past medical history, medications, and laboratory results, and an ECG-only model that learned directly from raw 12-lead waveforms. We then developed a multimodal deep-learning model that combined ECGs with sequential vital signs as well as the same static tabular features. All models were restricted to data available during the stay up to the time of the last ECG. Performance was assessed with stratified 5-fold cross-validation using identical splits across models.

Results: The multimodal model achieved an Area Under Receiver Operating Characteristic (AUROC) of 0.911 when trained on all eligible stays. The model predicted disposition after the final ECG was taken, which was a median of 0.3 hours after triage and 4.6 hours before ED departure. Baseline models performed worse: the ECG-only model had an AUROC of 0.852, and the tabular random forest had an AUROC of 0.886. In the subset requiring at least two ECGs within the stay, ECG-only reached an AUROC of 0.859, and random forest, with the longer interval to chart tabular data, reached a higher AUROC of 0.911. The multimodal model had AUROC 0.924, and outperformed baselines in each cohort (paired DeLong P<.001).

Conclusions: Serial ECGs, when integrated with evolving vitals and routine clinical features, enable accurate, early prediction of ED disposition in cardiac-presenting patients. This open-source, reproducible framework highlights the potential of multimodal deep learning to streamline ED flow, prioritize higher-risk cases, and detect evolving, time-critical pathology.

Keywords: Emergency Service, Hospital; Electrocardiography; Deep Learning; Hospitalization; Machine Learning

Introduction

Chest pain, dyspnea, and syncope are among the most common Emergency Department (ED) chief complaints that eventually result in a cardiac diagnosis [1]. They make up around 16 million encounters yearly in the United States, with chest pain accounting for almost 11 million visits a year [2-4]. Despite their cardiac connotation, the majority prove non-cardiac: observational series show that 58.7% of chest-pain cases are discharged with a non-cardiac diagnosis, adjudication of an international dyspnea cohort found cardiac etiology in 47% and non-cardiac causes in the remaining 53% of patients, and only 7–10% of syncope presentations to ED are ultimately attributed to a cardiac mechanism [5-6]. When a cardiac condition is confirmed, however, hospitalization becomes far more likely: more than 80% of acute-heart-failure presentations and up to 86% of high-risk syncope cases are admitted, while admission is far less common for patients whose symptoms are ultimately non-cardiac [7-8].

ED providers typically rely on an initial assessment that includes history and physical examination, vital signs, cardiac biomarker tests, and clinical risk scores like the Emergency Severity Index (ESI) triage level or the History, Electrocardiogram (ECG), Age, Risk Factors, and Troponin (HEART) score [2,14]. However, these traditional risk stratification tools have important limitations. Scores such as ESI and HEART are calculated at a single time point and may not fully reflect evolving patient risk. In practice, initial risk stratification for possible cardiac-presenting patients can be insufficient, contributing to ED crowding, which is associated with delayed care, mortality, and generally poorer patient outcomes [25]. This motivates the exploration of advanced machine learning (ML) methods that can integrate multiple data sources and time points to improve predictive performance.

Recent studies have shown that ML and deep learning models can outperform traditional triage and risk scores in predicting ED patient outcomes. These models often utilize triage data in combination with vitals, lab results, free-text notes, and past medical history (PMH) to successfully predict general hospitalization or specific critical care outcomes like acute coronary syndrome [9-23]. Fewer recent studies include features within the ED stay, including medications administered, lab tests, and early diagnoses [21-23].

One promising avenue is leveraging deep learning to fuse heterogeneous data sources, including sequential time-series data like waveforms, for outcome prediction. Many studies have incorporated ECGs into a disposition model; however, they use implied ECG findings indicated within physician notes or a simple flag indicating whether an ECG was abnormal or conducted [13-23]. The implementation of waveforms or more advanced ECG features remains unexplored. Chest pain, dyspnea, and (pre)syncope patients often undergo ECGs and vital sign measurements over the course of their ED evaluation. Important prognostic information may lie in the trends and changes in these data. Prior work suggests that sequential data modeling can improve the prediction of patient outcomes. For instance, Bouzid et al. demonstrated that analyzing serial ECGs can enhance the detection of acute coronary syndromes: in patients with suspected non-ST-elevation myocardial infarction, combining the pre-hospital ECG with the initial ED ECG and applying an ML classifier improved diagnostic accuracy and an AI-augmented model further boosted performance to AUROC 0.83 [24].

Our study builds on these advancements by introducing a multimodal deep learning approach for early prediction of ED disposition in adult patients presenting with cardiac-related complaints. We developed a multimodal deep-learning model that fuses serial ECG waveforms, sequential vital signs, and key clinical features to predict, in real time, whether a chest-pain patient will require

hospital admission. The model will predict disposition after the final ECG available has been taken, requiring anywhere from 1 to 6 ECGs. In contrast to prior works that often focus on diagnostic endpoints or use data available only at presentation, we target the practical outcome of patient disposition and leverage data collected during the ED stay—focusing on ECG waveforms, a novel datastream when predicting ED disposition.

The vast majority of previous studies utilized private hospital data that was not available for public use [10-14, 16-23]. The Medical Information Mart for Intensive Care IV is a large deidentified dataset of patients admitted to the emergency department or an intensive care unit at the Beth Israel Deaconess Medical Center in Boston, MA [26-28]. All data used in this project can be found in the MIMIC-IV, MIMIC-IV-ED, and MIMIC-IV-ECG modules on PhysioNet [26-30].

We hypothesized that this fusion of multimodal sequential data would yield more accurate disposition predictions than conventional methods. If successful, our approach could improve early identification of patients with chest pain, dyspnea, and syncope who require admission (or conversely, those who are safe for early discharge), ultimately enhancing ED decision-making, resource utilization, and patient outcomes.

Methods

Study Cohort

The MIMIC-IV-ED module was filtered to obtain a cohort of 82,907 unique patients with at least one mapped electrocardiogram from the MIMIC-IV-ECG module within the duration of an ED stay [29, 30]. ED stays without ECGs were excluded, and one stay per patient was retained.

The cohort was then filtered only to include patients whose chief complaints contained keywords relating to (pre)syncope, dyspnea, and chest pain. After filtering patients with a disposition other than discharge or admission, 30,421 unique ED patients with possible cardiac-related symptoms and at least one ECG were left in the study. The number of ECGs per stay and the length of stay before the final ECG from which the model predicts disposition were additionally noted as static features.

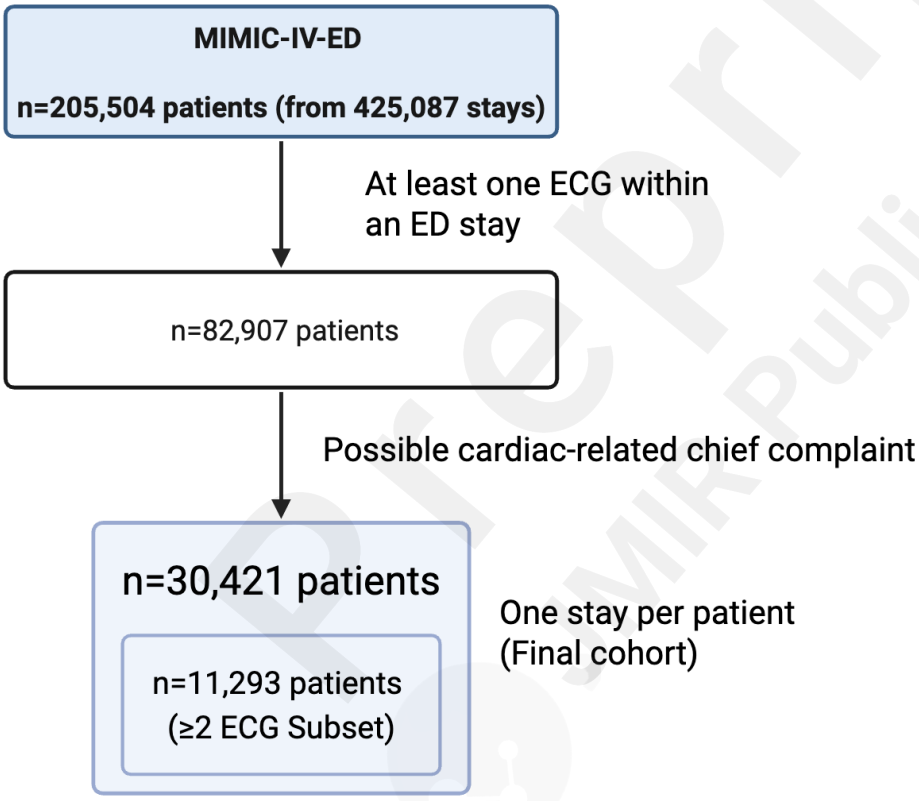


Figure 1. Flowchart of study population.

Feature Extraction

ED Lab Results and Medications (static)

From MIMIC-IV hospital data, we extracted ED labs and medications recorded after ED arrival and before the final ECG. The lab results included were Troponin T, creatinine, lactate, C-reactive protein, B-type natriuretic peptide, hemoglobin, potassium, magnesium, and white blood cell count.

For each lab, we derived four features: first value, peak value, an abnormal flag, and a “missing” flag. ED medications were identified through Pyxis GSN codes, which were mapped to Enhanced Therapeutic Classification (ETC) codes and grouped by their first six digits. Each group contributed a binary feature (38 features). All lab tests and medications were incorporated as static features and were not treated sequentially or iteratively imputed.

Past Medical History (static)

We included the disposition of the prior visit and the number of previous stays visible in the MIMIC-IV database. Prior ICD diagnoses from a MIMIC hospital visit were binned to CCS-ICD9 and CCSR-ICD10 categories (253 variables). Outpatient prescriptions in the past year were binned to 7 broad ETC groups.

Other static features

Age, sex, and acuity were included along with arrival mode and chief complaint Booleans. We added total counts for ED medications, prior medications, and prior diagnoses.

Vitals (Sequential)

We incorporated temperature, heart rate, respiratory rate, oxygen saturation, systolic and diastolic blood pressure, and pain recorded before the final ECG. All vitals taken at triage were included, as well as vitals drawn throughout the ED stay, with chart time before the last ECG for that stay. All vitals except for pain were non-missing at triage, but contained a large number of missing values when later charted during the stay. Missing vitals, excluding pain, were imputed using a Bayesian Regression formula fit on the training set per fold. Missing flags were created for each vital to note whether the vital was initially missing before imputation. Two more columns were added, stating whether the patient was sleeping or unable to answer based on the string value of pain.

Electrocardiogram Feature Extraction

In order to best implement ECGs for the downstream task of predicting hospitalization, an AI model was used to encode a lower-dimensional ECG feature vector. Specifically, a one-dimensional residual neural network model (ResNet-18) was chosen due to its ability to learn discriminative temporal and spatial features from high-dimensional ECG signals while avoiding vanishing gradients through skip connections. The network takes in an ECG with 5000 samples (500Hz for 10 seconds) for each of the 12 leads, and outputs one singular 512-dimensional feature vector. In the first approach, the ResNet-18 was initialized for hospitalization prediction with random weights. In the second, it was first trained on a larger ECG dataset with labels unrelated to ED disposition, to capture general waveform structure and patterns. Training beforehand allows the network to initialize with physiologically relevant weights, enhancing feature extraction of QRS–T morphology and inter-lead patterns, while significantly reducing overfitting through faster convergence and freedom to freeze layers. Reduced overfitting may be important given the added complexity of the multimodal architecture.

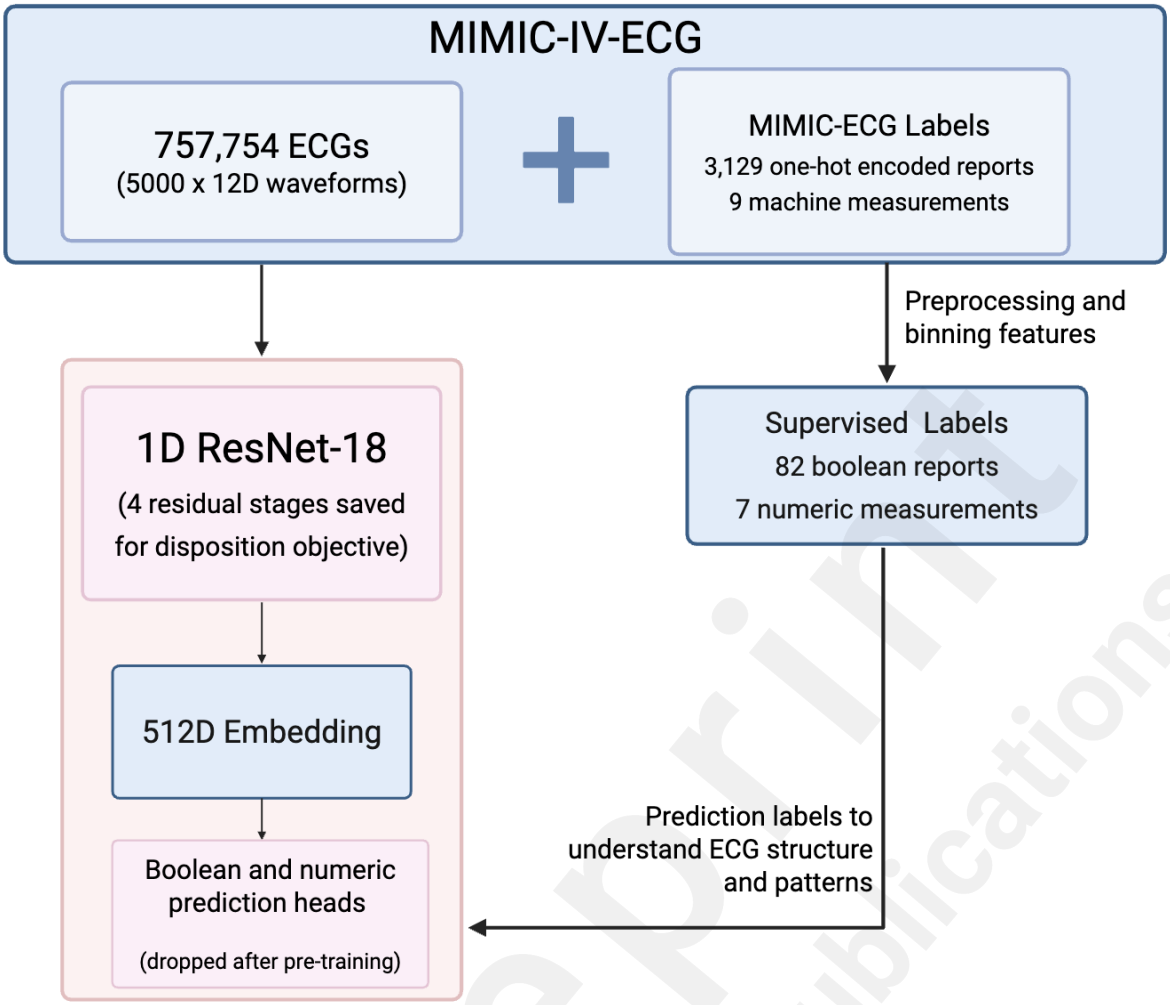


Figure 2. Flow chart of supervised pretraining workflow for the transfer-learning approach.

The full MIMIC-IV-ECG dataset was used as a means of supervised pretraining, in which each ECG waveform was used as an input to predict MIMIC machine measurements and machine-generated reports. These measurements and reports are provided on the MIMIC-IV-ECG module for every available ECG. Machine measurements reflect quantitative ECG characteristics across all leads, including the average RR interval, QRS axis, T-axis, and several more. Machine-generated reports are stored as strings in columns labeled report 1 through report 17, with example strings being “Atrial Fibrillation”, “ST-elevation”, or “Normal ECG.” After one-hot-encoding, 3,129 columns were initially created. After removing labels that had too few positive instances, along with combining truth values of columns with slight syntax variance or similar clinical significance, such as “ventricular pacing” and “ventricular-paced rhythm”, 82 unique Boolean features remained.

Before training, all ECGs were cleaned using the Neurokit2 signal processing library [31]. A 1D ResNet-18 Model was trained to predict both numeric and Boolean features corresponding to each of the total 757,754 ECGs. Mean Squared Error (MSE) was used to evaluate the loss of numeric features (machine measurements), and Binary Cross-Entropy (BCE) was used to evaluate the loss of Boolean features (machine-generated reports). The total loss function represented the MSE loss added to the BCE loss multiplied by 9 (due to far more Boolean labels than numeric).

Prediction Model

Multimodal

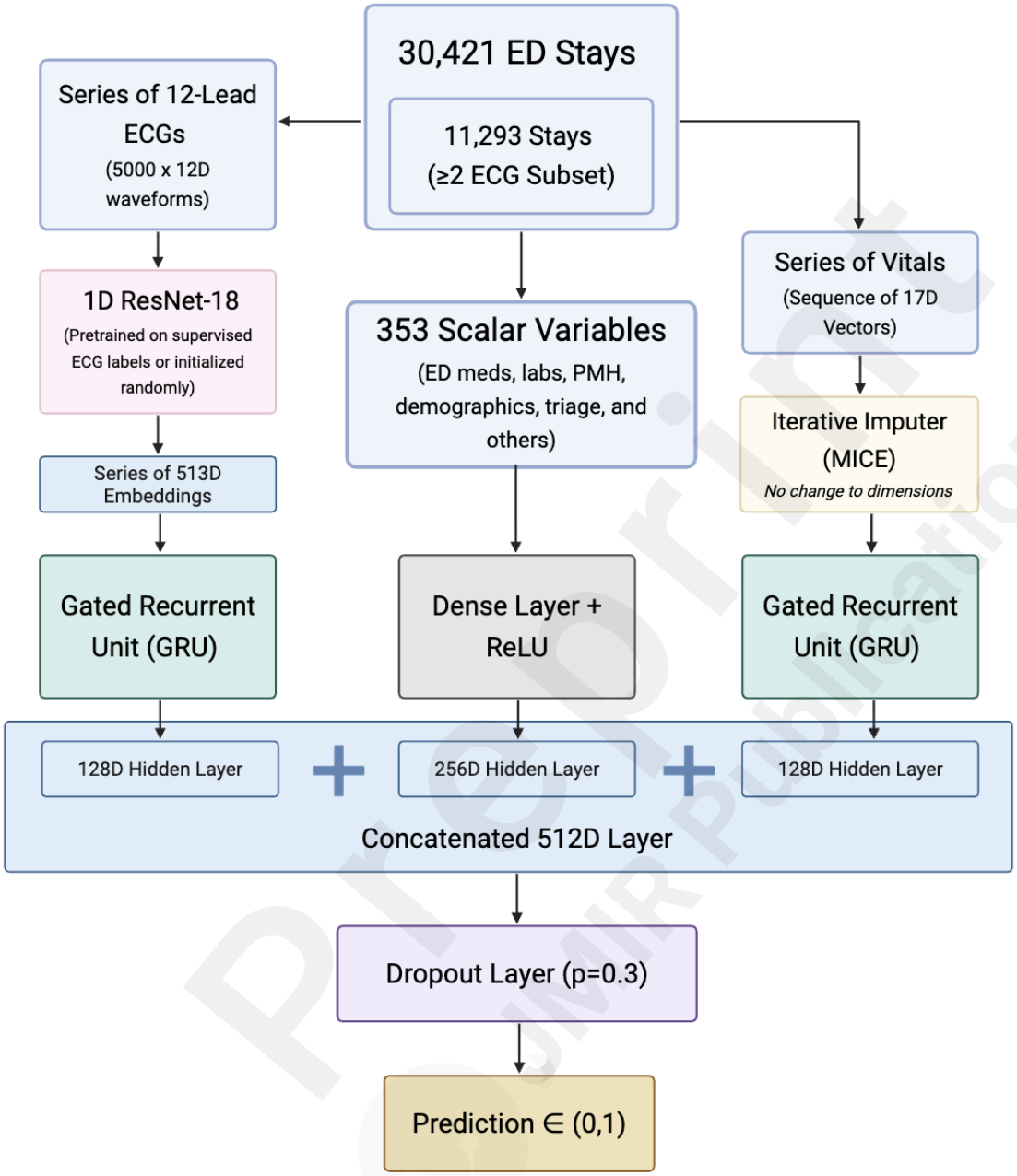


Figure 3. Diagram of the multimodal model architecture approach to predict hospitalization in the full and subset cohorts.

A multimodal model trained on sequential ECG waveforms, sequential vitals data, and 353 static variables was built to predict hospitalization in emergency department patients. The model utilized all data gathered in the stay before the final ECG. The model was trained on all stays as well as on a subset of stays containing at least two ECGs to quantify the importance of sequential ECGs and more available data within the time window.

Raw 12-lead electrocardiograms (ECGs) were first cleaned with NeuroKit2 and subsequently one-dimensional ResNet-18 initialized with random weights or the ResNet-18 that had been trained on

the larger MIMIC-IV-ECG dataset (757,754 recordings). In the transfer-learning approach, ResNet-18 layers through block layer3 were frozen: only layer4 and the linear adapter were trained during finetuning. The final global-average-pooling layer of the backbone was replaced with a linear adapter so that each waveform was mapped to a fixed-length, 512-dimensional embedding.

For every emergency department (ED) stay we retained, in temporal order, all ECGs recorded before the clinical disposition decision (maximum six per stay) and all vital-sign rows charted before the last ECG (maximum ten per stay). The ECG and vital embeddings were first stacked and padded to match the maximum length. As shown in Figure 3, two independent, single-layer gated recurrent units (GRUs) were used to summarize these sequences. Before being passed to the GRU, padded batches were converted to PackedSequence objects, ensuring the recurrent unit was unrolled only over the valid timesteps and ignored the artificial zero-padding. The time before the final ECG was added as an extra time-delta channel in the embeddings of each ECG and vitals sequence. The ECG GRU read the sequence of 513-dimensional embeddings and returned a 128-dimensional hidden state h_{ECG}; the vitals GRU processed a 17-variable vector at each timestep—seven z-scored physiological values, one time-delta channel, and nine binary mask indicators denoting whether the original measurement had been missing or unable/sleeping—and produced a similar 128-dimensional summary h_{VITALS}.
All 353 static features were concatenated, normalized with the mean and standard deviation calculated on the training set, and projected by a fully connected layer with ReLU activation to a 256-dimensional vector *s*.

The three modality-specific representations were then fused through simple concatenation,
$z = [\, h_{ECG} \| h_{VITALS} \| s \,] \in \mathbb{R}^{512}$,
followed by an additional dropout layer (rate = 0.30) and a single neuron that yielded the estimated probability of hospital admission $\hat{p} \in (0,1)$.

Model parameters introduced during the present study were optimized with AdamW (weight-decay 10^{-4} at a learning rate of $3*10^{-4}$ [32]; the adapter atop the ECG encoder was updated at 10^{-4} . Binary cross-entropy with class weighting was used as the optimization objective.

Baseline

To measure the effect of ECGs alone in predicting disposition, separate models were trained on only ECG features. The ECG-only model utilized the same parameters and architecture as the multimodal ECG branch.

Random forest is a machine-learning algorithm that combines predictions from an ensemble of decision trees to produce one final classification. A random forest was constructed solely on the 353 static tabular features as a baseline to the multimodal architecture. Both the ECG-only and tabular baseline models were evaluated in the full cohort (“all stays”) and in the ≥2-ECG subset using the exact stratified five-fold cross-validation splits from the multimodal pipeline.

Model Evaluation

Three models (multimodal, tabular, and ECG-only) were evaluated on the full cohort of all stays and on the ≥2-ECG subset. Across cross-validation folds, the mean AUROC, AUPRC, F1, precision, and recall were reported. Thresholds for F1, precision, and recall were selected using a nested procedure that maximized F1 on training folds and applied it to the validation fold. Pairwise AUROC differences between models were tested with the DeLong method.

Ethical Considerations

This study used the deidentified MIMIC-IV databases under the PhysioNet data use agreement and is exempt from IRB review and individual consent.

Results

Table 1. Baseline characteristics and measurements in all vs ≥2-ECG cohorts.[a,b]

Metric	All stays (N=30,421)	≥2 ECG stays (N=11,273)
Admit, n (%)	13,138 (43.2%)	4,070 (36.1%)
Acuity, mean (SD)	2.3 (0.6)	2.2 (0.6)
Age, mean (SD)	57.1 (19.8)	61.7 (16.3)
ECG metrics		
Number of ECGs, mean (SD)	1.48 (0.74)	2.30 (0.64)
≥2 ECGs, n (%)	11,273 (37.1%)	11,273 (100.0%)
≥3 ECGs, n (%)	2,538 (8.3%)	2,538 (22.5%)
LOS after final ECG, h, median (IQR)	4.6 (2.7–7.3)	3.4 (1.7–9.1)
LOS before final ECG, h, median (IQR)	0.3 (0.2–5.3)	6.5 (3.9–7.7)
Full LOS, h, median (IQR)	6.5 (4.3–9.8)	9.5 (6.9–17.8)
Laboratory values (first)		
Troponin T, ng/mL, median (IQR)	0.1 (0.0–0.2)	0.1 (0.0–0.2)
Missing, n (%)	28,376 (93.3%)	9,579 (85.0%)
Lactate, mmol/L, median (IQR)	1.7 (1.3–2.3)	1.7 (1.3–2.3)
Missing, n (%)	27,856 (91.6%)	9,206 (81.7%)
Creatinine, mg/dL, median (IQR)	0.9 (0.8–1.1)	0.9 (0.8–1.1)
Missing, n (%)	18,327 (60.2%)	1,177 (10.4%)
B-type natriuretic peptide, pg/mL, median (IQR)	1,092.5 (180.0–4,555.8)	948.5 (162.0–4,293.2)
Missing, n (%)	27,993 (92.0%)	9,363 (83.1%)
Potassium, mmol/L, median (IQR)	4.2 (3.9–4.6)	4.2 (3.9–4.6)

Missing, n (%)	18,340 (60.3%)	1,170 (10.4%)
Magnesium, mg/dL, median (IQR)	2.0 (1.9–2.2)	2.0 (1.9–2.2)
Missing, n (%)	26,702 (87.8%)	8,244 (73.1%)
Presenting symptoms		
Chest pain, n (%)	17,898 (58.8%)	8,519 (75.6%)
Dyspnea, n (%)	9,477 (31.2%)	2,575 (22.8%)
Syncope, n (%)	4,924 (16.2%)	983 (8.7%)
Summary counts		
No visible ED meds before final ECG, n (%)	21,368 (70.2%)	3,583 (31.8%)
No visible prior medications (past year), n (%)	22,074 (72.6%)	8,157 (72.4%)
No visible prior hospital diagnosis, n (%)	15,950 (52.4%)	5,275 (46.8%)
No visible prior ED stay, n (%)	16,594 (54.5%)	5,591 (49.6%)

[a]Abbreviations: LOS, length of stay

[b]Percentages for all rows are calculated using unique patients as the denominator (N). The values from the selected representative lab measurements are from tested patients only.

Detailed vitals summaries, including missingness counts by cohort, are shown in Table S1 (Multimedia Appendix 1).

ED Timing Analysis

Of the 30,421 ED stays included in our analysis, 11,273 (37.1%) involved stays with ≥ 2 ECGs (Table 1). In the full cohort, the model issued its prediction a median of 0.3 h after triage (IQR 0.1–5.3 h), leaving a median of 4.6h (IQR 2.7–7.3 h) before the patient physically left the ED. Note that these medians are calculated independently and therefore do not sum to the overall median length of stay. As a result, disposition was predicted in the first 18 minutes for most visits, but for the upper quartile, the model leveraged data gathered over 5 hours into the encounter. In the ≥2 ECG subset, the model predicted disposition far later, with a median of 6.5 hours after triage. The median time from prediction to disposition was also shorter than the full cohort, at 3.4 hours. However, the 75th-percentile (Q3) prediction lead time exceeded 9 hours, meaning the model anticipated disposition more than nine hours before the actual decision. This subset also had a longer overall ED length of stay (median 9.5 vs 6.5 hours in the full cohort). Because the ≥2 ECG cohort had more ECGs taken and the model predicts disposition after the final ECG, it often took more time to reach the final ECG.

Missing Values

When vitals were charted, only temperature was frequently missing within 41.8% (28,573/68,371) of vital chartings within all stays and 55% (25,376/46,098) in the ≥2 ECG subset (Table S1). All other vitals were <7% missing per charting, and many stays had multiple vitals chartings before prediction

cutoff. Of the nine lab values this study accounted for, six were entirely missing in over 90% (~27,400/30,421) of all patients before the prediction cutoff (Table 1). Because the subset included a longer median interval before the prediction cutoff, the proportion of missing laboratory results fell for every test. Notably, creatinine dropped from 60.2% (18,327/30,421) to 10.4% (1,177/11,273) missing between the cohorts. Past medical history within the dataset's visibility was very limited. Any prior diagnoses and ED stays were each completely missing in over half of the total patients. 72.6% (22,074/30,421) of patients had no visible medications prescribed to them in the past year in the dataset.

Electrocardiogram Feature Extractor

Compared with random initialization, supervised transfer-learning of the ResNet-18 did not yield statistically significant improvements in predictive performance. However, supervised pretrained models consistently converged faster and achieved marginally higher metrics in the ≥2 ECG subset. Accordingly, all reported results use transfer-learning approach.

Model Evaluation

Using a partially frozen pretrained ResNet-18 as an ECG encoder, a multimodal dual GRU fusion net was trained on the full cohort and ≥2 ECG subset to predict hospital admission. For comparison, an ECG-only variant (same encoder and GRU using only raw 12-lead waveforms) and a tabular-only random forest were trained for the same task. Performance is shown in Table 2 and Figure 4.

Table 2. Performance metrics for ECG-only, multimodal, and tabular random forest models. Evaluations use an identical time cut-off across models and are stratified by cohort (all stays; ≥2 ECGs). P-values are from paired DeLong tests versus the multimodal model. Metrics are mean values across the five-fold cross-validation.

Model	AUROC, mean (SD)	AUPRC, mean (SD)	F1, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	P value
All stays						
ECG only	0.852 (0.003)	0.813 (0.005)	0.745 (0.004)	0.698 (0.034)	0.803 (0.036)	<.001
Tabular (RF)	0.886 (0.003)	0.849 (0.006)	0.785 (0.003)	0.745 (0.017)	0.830 (0.014)	<.001
Multimodal	**0.911 (0.004)**	**0.889 (0.005)**	**0.809 (0.005)**	**0.784 (0.041)**	**0.839 (0.046)**	—
≥2 ECG stays						
ECG only	0.859 (0.011)	0.794 (0.017)	0.713 (0.010)	0.674 (0.029)	0.760 (0.038)	<.001
Tabular (RF)	0.911 (0.006)	0.865 (0.014)	0.793 (0.009)	0.774 (0.016)	**0.813 (0.013)**	<.001

Multimodal	**0.924 (0.009)**	**0.889 (0.016)**	**0.807 (0.006)**	**0.807 (0.030)**	0.808 (0.024)	—

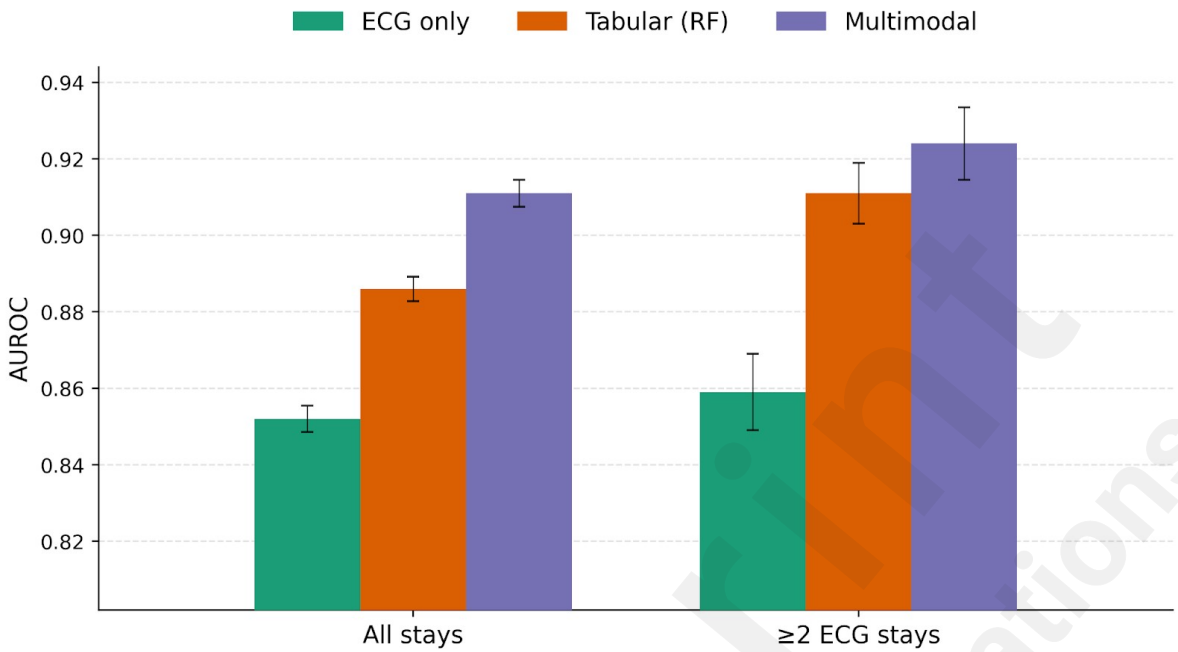


Figure 4. Mean validation AUROC of each model in each cohort plotted in a bar graph.

The multimodal model in the multiple ECG subset yielded the highest metrics, with AUROC 0.924. The tabular random forest model yielded slightly lower metrics (AUROC 0.91). Across cohorts, the AUROC of the ECG-only model differed by as little as 0.007 AUROC, suggesting that additional information from prior ECGs in the same stay does not improve prediction by a significant margin. In the all-stays cohort, where predictions were made much earlier in the ED course, the performance gap between the random forest and multimodal models was larger: 0.886 vs. 0.911 AUROC (P<.001). These findings indicate that combining ECG waveforms with vitals provides the largest benefit when static tabular data, such as lab results, are sparse early in the encounter.

Discussion

Principal Findings

In this study, we developed a multimodal deep learning model that integrates sequential ECG waveforms and vital signs, along with static tabular features, to predict hospitalization in real time for ED patients with possible cardiac presentations. Our results show that ECGs can also be used alone to predict disposition, but are best used in combination with vitals and other commonly implemented static features such as lab results, medications administered, prior medical history, and triage data.

The multimodal model that required at least two ECGs achieved the highest performance of 0.924 AUROC in predicting hospitalization. This model predicted disposition 3.4 hours in advance on the median, but over 9 hours in advance for the top 25th percentile with high accuracy. However, a median of 6.5 hours was spent in the ED before this model reached a prediction. The random forest built from tabular data achieved a slightly lower 0.911 AUROC. As ECG-only model performance did not improve significantly with more ECGs, the higher AUROC in the ≥2 ECG subset for the multimodal and random forest models is most likely due to the more lenient cutoff to chart lab

results and ED medications before prediction.

When trained on all-stays (median prediction time less than 20 minutes after triage), the multimodal architecture yielded significantly higher metrics in comparison to tabular and ECG-only models. Incorporating additional data streams such as waveforms and vitals can lead to better and more reliable hospitalization predictions when traditional clinical data is sparse.

Comparison with Prior Work

Our ECG-based prediction models substantially outperform conventional triage tools like the nurse-assigned ESI level, which typically achieves AUROCs in the 0.69–0.70 range and is on par with or better than many prior machine learning models that used only triage-time data [9]. For example, Raita et al. reported an AUROC of 0.82 for a deep neural network predicting hospitalization using initial only triage information, and Hong et al. (2018) similarly found ~0.87 using triage data with XGBoost and a neural network [9,10]. Additionally, Hong et al also created XGBoost and Deep Neural Network models, which reached an AUROC of 0.92 implementing past medical history and triage data to predict hospital admission at the beginning of the ED stay [10]. Their approach is further discussed in the limitations section of this paper.

Some studies make use of data after triage to predict disposition. Barak-Corren et al. achieved high performance (AUROC 0.97 within 1 hour of triage) using logistic regression (LR) on standard demographic, triage, medication, lab, vital, and PMH data in a single-site Israeli ED [23]. However, they train their LR model to predict disposition on a per-visit basis, while we provide a per-patient approach, a more challenging task because there is no patient overlap within the train and validation cohort. Sezik et al. used similar data (lab results, history, and triage variables) in combination with vitals and other features extracted from text to reach an AUROC of 0.960 with a random forest Classifier [21]. However, they do not incorporate a cutoff time and use features conducted throughout the entire ED stay, including the length of stay. In contrast, our study stops feature collection after a given electrocardiogram in order to predict later changes in a cardiac-presenting patient’s condition and provide an early hospitalization prediction as opposed to serving as an aid to decide disposition at the end of the stay.

Importantly, our approach maintained high accuracy despite the broad outcome of general hospital admission, which includes many non-critical cases; this is notable because prior ML studies in chest pain often focused on narrower critical events. For instance, a recent study targeted only critical care outcomes such as ED cardiac arrest or ICU transfer, and achieved an AUROC of ~0.95 with a tailored LASSO logistic model [11]. The fact that we still attained strong performance (AUROC 0.911 and AUROC 0.924) for a broader outcome of hospitalization for any reason underlines the effectiveness of our multimodal fusion strategy. Using the MIMIC-IV-ED and MIMIC-IV-ECG modules (which include over 400,000 ED stays and 800,000 ECGs, respectively) ensured that our work is widely reproducible [29,30]. This reproducibility and open-source approach contrasts with many prior ED prediction studies that relied on proprietary data not available to outside researchers [10-14, 16-23]. By creating our model from a public database, we demonstrate the feasibility of developing advanced prediction tools using open data.

Limitations and Future Work

This study has several limitations. Firstly, all data used to train the model is specific to the MIMIC-IV database and the Beth Israel Deaconess Medical Center. Data from another format from another hospital would not be usable in the model as of currently. Due to exploring the promise of ECG

waveforms on ED disposition, our study relies on the patient presenting with a possible cardiac condition, which in this case only includes patients presenting with chest pain, dyspnea, and (pre)syncope. There is currently no external validation.

ED diagnoses were also not included because the dataset did not specify the time of diagnosis. In addition, many variables were largely missing, including over 90% of most labs tested and 41 to 55% of temperature readings when vitals were charted. As opposed to the 0.92 AUROC neural network proposed by Hong and colleagues, which contained full patient medical records and a comprehensive past medical history of each patient, we only include past history visible within the MIMIC-IV and MIMIC-IV-ED modules. 72.6% of patients in our study set had no prior medications, 52.4% had no prior diagnoses, and 54.5% had no prior ED stays. Past medical history recorded at another hospital was not visible. Considering the value shown by incorporating a comprehensive PMH in predicting disposition, our model performance could significantly benefit from having access to more prior diagnoses, medications, and ED stays. Finally, like most disposition models, our model treats the disposition recommended by the ED provider as a truth label. Thus, our model may be, to some extent, learning institutional decision policies which can vary across hospitals in addition to patient physiology.

Future work should explore including more data modalities in a time-series format and possibly leaving the cutoff for prediction variable throughout updates in lab testing, vitals, and diagnostic tools such as an ECG. Incorporating novel data streams like the ECG with a more complete PMH could lead to a far better prediction as well. Improving model interpretability, especially within deep-learning models, is also crucial to gain clinician trust and improve decision-making in the ED.

Conclusions

We have presented a novel deep learning approach that fuses multimodal ED data to predict hospital admission in patients with potential cardiac complaints. The model significantly outperforms traditional triage metrics and provides accurate predictions hours in advance of the final disposition decision. Through utilizing a supervised transfer-learning approach, we also present a model trained solely on ECG inputs, which can still predict hospitalization with strong accuracy. These findings suggest that advanced data-driven tools can enhance early risk stratification for patients with chest pain, syncope, and dyspnea, potentially improving ED resource allocation and patient outcomes. These findings also prove that non-traditional data modalities from open-source datasets can compete with or perform better than studies with extensive features and private data. Ultimately, by recognizing high-risk cardiac patients sooner and reducing unnecessary hospitalizations for low-risk individuals, our approach could contribute to more efficient and effective emergency care.

Acknowledgements

The author thanks Ashwath Radhachandran and Amir Reza Vazifeh for providing valuable feedback on the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

None declared.

Data Availability

The study draws exclusively on the publicly available Medical Information Mart for Intensive Care

version IV (MIMIC-IV) databases—including the emergency-department (MIMIC-IV-ED) and electrocardiogram (MIMIC-IV-ECG) modules—hosted on PhysioNet. Because access to MIMIC-IV is restricted under a data-use agreement that protects patient privacy, raw or pre-processed patient-level data cannot be redistributed here. Interested researchers can obtain the same data free of charge by completing the data-use certification on PhysioNet. All scripts for data preprocessing, model training, and evaluation, as well as trained model weights, are openly available at GitHub [33].

Author Contributions

AA: Conceptualization, Methodology, Data Curation, Formal Analysis, Investigation, Software, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing. KBO: Supervision, Validation, Writing – Review & Editing.

Multimedia Appendix 1

Table S1. Vitals measurements in all vs ≥2-ECG cohorts.

References

1. Gupta M, Tabas JA, Kohn MA. Presenting complaint among patients with myocardial infarction who present to an urban, public hospital emergency department. Ann Emerg Med. 2002;40(2):180–186. doi:10.1067/mem.2002.125949
2. Yukselen Z, Majmundar V, Dasari M, Kumar PA, Singh Y. Chest pain risk stratification in the emergency department: current perspectives. Open Access Emerg Med. 2024;16:29–43. doi:10.2147/oaem.s419657
3. Patel PR, Quinn JV. Syncope: a review of emergency department management and disposition. Clin Exp Emerg Med. 2015;2(2):67–74. doi:10.15441/ceem.14.049
4. DeVos E, Jacobson L. Approach to adult patients with acute dyspnea. Emerg Med Clin North Am. 2015;34(1):129–149. doi:10.1016/j.emc.2015.08.008
5. McDevitt-Petrovic O, Kirby K, Shevlin M. The prevalence of non-cardiac chest pain (NCCP) using emergency department data: a Northern Ireland based study. BMC Health Serv Res. 2017;17(1):549. doi:10.1186/s12913-017-2493-8
6. Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med. 2002;347(3):161–167. doi:10.1056/nejmoa020233
7. Collins SP, Jenkins CA, Harrell FE, et al. Identification of emergency department patients with acute heart failure at low risk for 30-day adverse events. JACC Heart Fail. 2015;3(10):737–747. doi:10.1016/j.jchf.2015.05.007
8. Hatoum T, Sheldon RS. A rational evaluation of the syncope patient: optimizing the emergency department visit. Medicina. 2021;57(6):514. doi:10.3390/medicina57060514
9. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care. 2019;23(1):64. doi:10.1186/s13054-019-2351-7
10. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One. 2018;13(7):e0201016. doi:10.1371/journal.pone.0201016
11. Wu TT, Zheng RF, Lin ZZ, Gong HR, Li H. A machine learning model to predict critical care outcomes in patients with chest pain visiting the emergency department. BMC Emerg Med. 2021;21(1):112. doi:10.1186/s12873-021-00501-8
12. Gustafsson S, Gedon D, Lampa E, et al. Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients. Sci

Rep. 2022;12(1):19615. doi:10.1038/s41598-022-24254-x
13. Peck JS, Benneyan JC, Nightingale DJ, Gaehde SA. Predicting emergency department inpatient admissions to improve same-day patient flow. Acad Emerg Med. 2012;19(9):E1045–E1054. doi:10.1111/j.1553-2712.2012.01435.x
14. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the Emergency Severity Index. Ann Emerg Med. 2018;71(5):565–574.e2. PMID:28888332
15. Dugas AF, Kirsch TD, Toerper M, Korley F, Yenokyan G, France D, et al. An electronic emergency triage system to improve patient distribution by critical outcomes. J Emerg Med. 2016;50:910–918. PMID:27133736
16. Leegon J, Jones I, Lanaghan K, Aronsky D. Predicting hospital admission in a pediatric emergency department using an artificial neural network. AMIA Annu Symp Proc. 2006;2006:1004. PMID:17238623
17. Leegon J, Jones I, Lanaghan K, Aronsky D. Predicting hospital admission for emergency department patients using a Bayesian network. AMIA Annu Symp Proc. 2005;2005:1022. PMID:16779309
18. Lucini FR, Fogliatto FS, da Silveira GJC, Neyeloff JL, Anzanello MJ, Kuchenbecker RS, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. Int J Med Inform. 2017;100:1–8. doi:10.1016/j.ijmedinf.2017.01.001
19. Lucke JA, de Gelder J, Clarijs F, et al. Early prediction of hospital admission for emergency department patients: a comparison between patients younger or older than 70 years. Emerg Med J. 2018;35(1):18–27. doi:10.1136/emermed-2016-205846
20. Chang YH, Lin YC, Huang FW, et al. Using machine learning and natural language processing in triage for prediction of clinical disposition in the emergency department. BMC Emerg Med. 2024;24(1):237. doi:10.1186/s12873-024-01152-1
21. Sezik S, Cingiz MÖ, İbiş E. Machine learning-based model for emergency department disposition at a public hospital. Appl Sci. 2025;15(3):1628. doi:10.3390/app15031628
22. Barak-Corren Y, Israelit SH, Reis BY. Progressive prediction of hospitalisation in the emergency department: uncovering hidden patterns to improve patient flow. Emerg Med J. 2017;34:308–314. PMID:28188202
23. Barak-Corren Y, Fine AM, Reis BY. Early prediction model of patient hospitalization from the pediatric emergency department. Pediatrics. 2017;139:e20162785. PMID:28557729
24. Bouzid Z, Faramand Z, Martin-Gill C, Sereika SM, Callaway CW, Saba S, et al. Incorporation of serial 12-lead electrocardiogram with machine learning to augment the out-of-hospital diagnosis of non-ST elevation acute coronary syndrome. Ann Emerg Med. 2023;81(1):57–69. doi:10.1016/j.annemergmed.2022.08.005
25. Bernstein SL, Aronsky D, Duseja R, et al. The effect of emergency department crowding on clinically oriented outcomes. Acad Emerg Med. 2008;16(1):1–10. doi:10.1111/j.1553-2712.2008.00295.x
26. Johnson AEW, Bulgarelli L, Pollard TJ, et al. MIMIC-IV (version 3.0). PhysioNet; 2024. doi:10.13026/hxp0-hg59
27. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1–9. doi:10.1038/s41597-022-01899-x
28. Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215–e220. doi:10.1161/01.CIR.101.23.e215
29. Johnson AEW, Bulgarelli L, Pollard TJ, Celi LA, Mark RG, Horng S. MIMIC-IV-ED

(version 2.2). PhysioNet; 2023. doi:10.13026/5ntk-km72
30. Gow B, Pollard T, Nathanson LA, Johnson A, Moody B, Fernandes C, Greenbaum N, Waks JW, Eslami P, Carbonati T, Chaudhari A, Herbst E, Moukheiber D, Berkowitz S, Mark RG, Horng S. MIMIC-IV-ECG: Diagnostic Electrocardiogram Matched Subset (version 1.0). PhysioNet; 2023. doi:10.13026/4nqg-sb35
31. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, et al. NeuroKit2: a Python toolbox for neurophysiological signal processing. Behav Res Methods. 2021;53(4):1689–1696. doi:10.3758/s13428-020-01516-y
32. Loshchilov I, Hutter F. Decoupled weight decay regularization. Proceedings of the International Conference on Learning Representations; 2019. URL: https://arxiv.org/abs/1711.05101
33. Altintepe A. *ECGhospitalization*. GitHub. URL: https://github.com/againh3x/ECGhospitalization

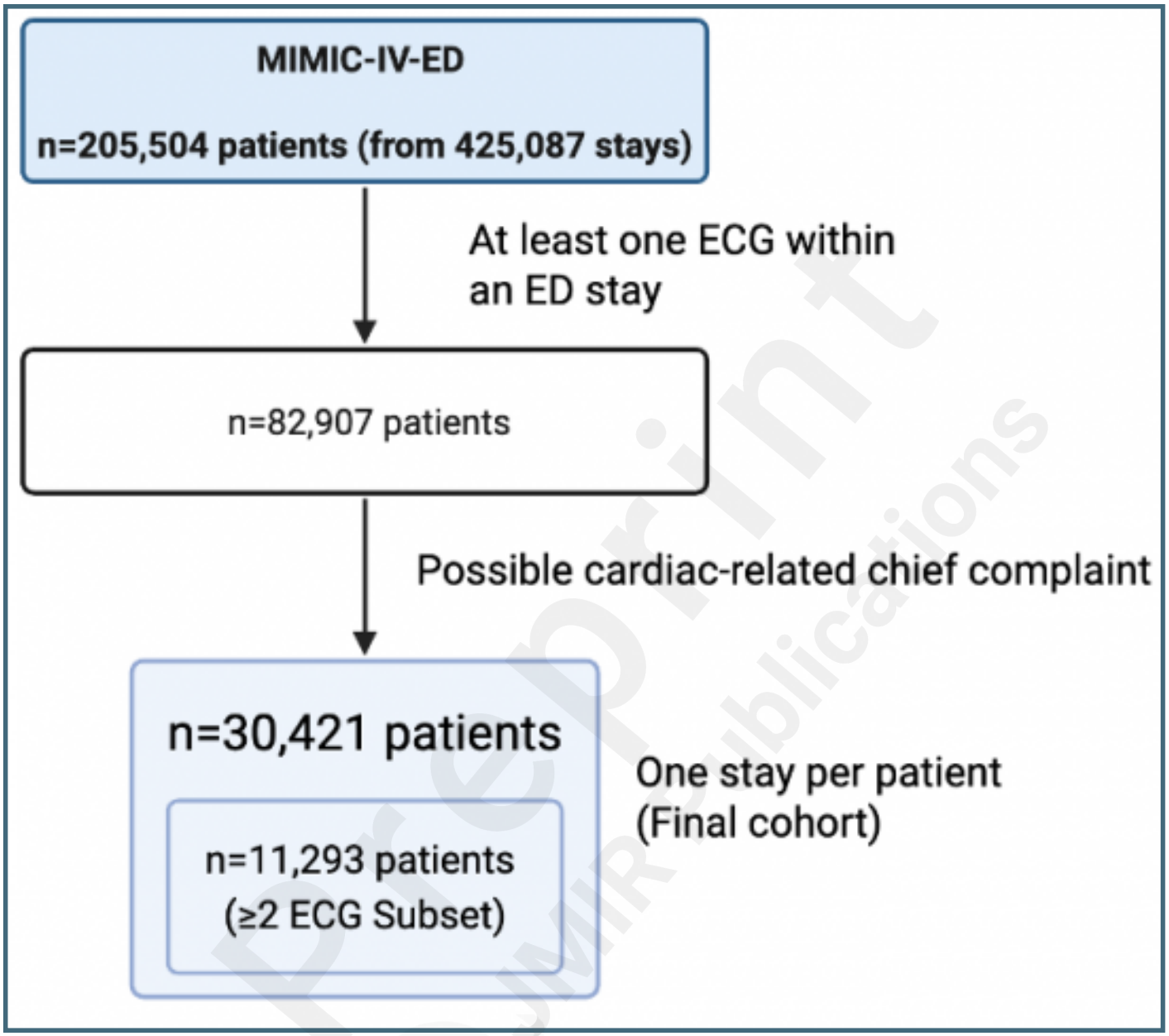
Abbreviations

AUPRC: area under the precision–recall curve
AUROC: area under the receiver operating characteristic curve
BCE: binary cross-entropy
BNP: B-type natriuretic peptide
BP: blood pressure
CRP: C-reactive protein
CV: cross-validation
DBP: diastolic blood pressure
ECG: electrocardiogram
ED: emergency department
EHR: electronic health record
ESI: Emergency Severity Index
F1: F1 score (harmonic mean of precision and recall)
GRU: gated recurrent unit
ICU: intensive care unit
IQR: interquartile range
IRB: institutional review board
LOS: length of stay
ML: machine learning
MIMIC-IV: Medical Information Mart for Intensive Care IV
MIMIC-IV-ED: MIMIC-IV Emergency Department module
MIMIC-IV-ECG: MIMIC-IV electrocardiogram module
PR: precision–recall
ResNet: residual neural network
RF: random forest
ROC: receiver operating characteristic
SBP: systolic blood pressure
SD: standard deviation

Supplementary Files

Figures

Flowchart of study population.



Flow chart of supervised pretraining workflow for the transfer-learning approach.

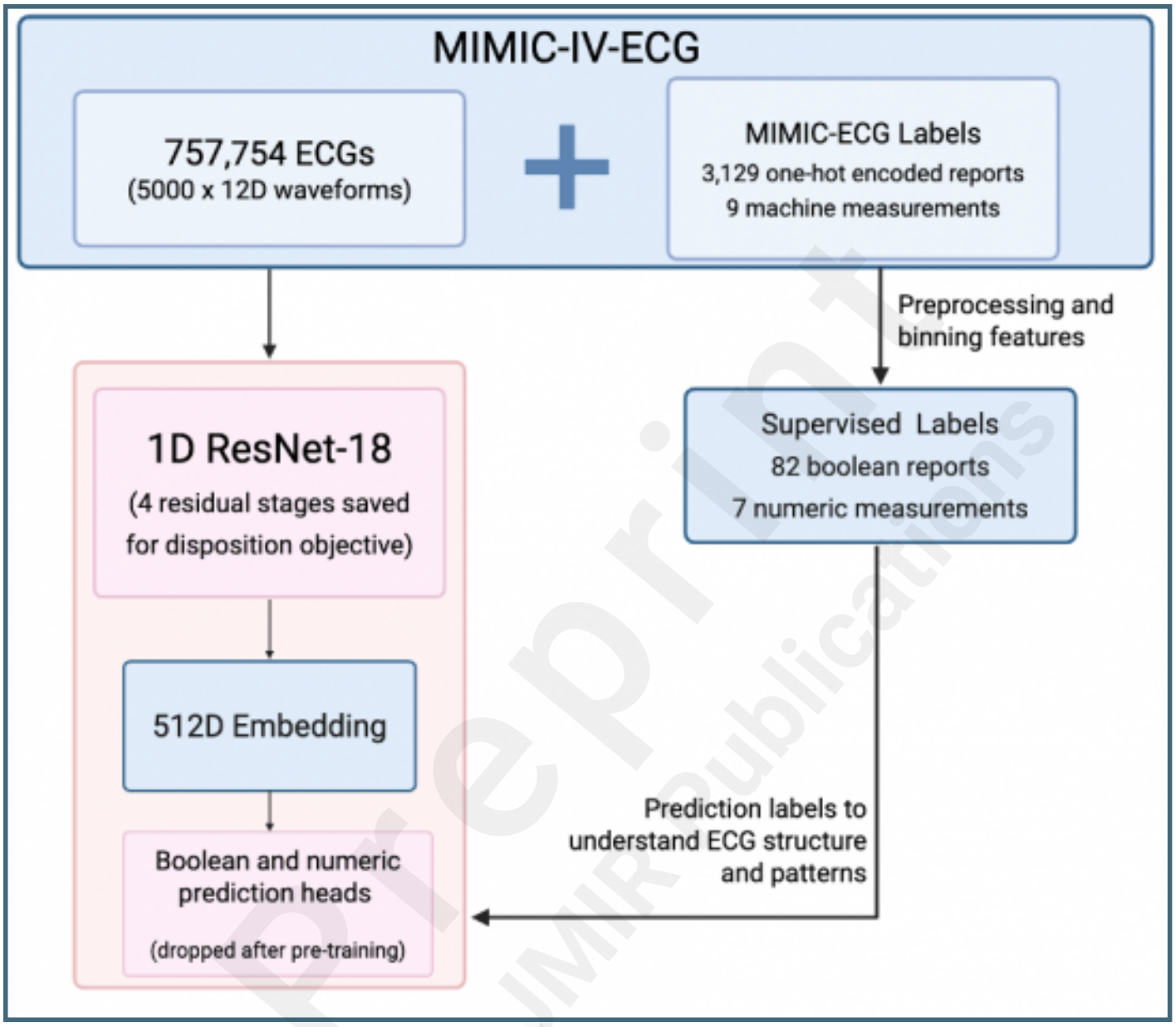
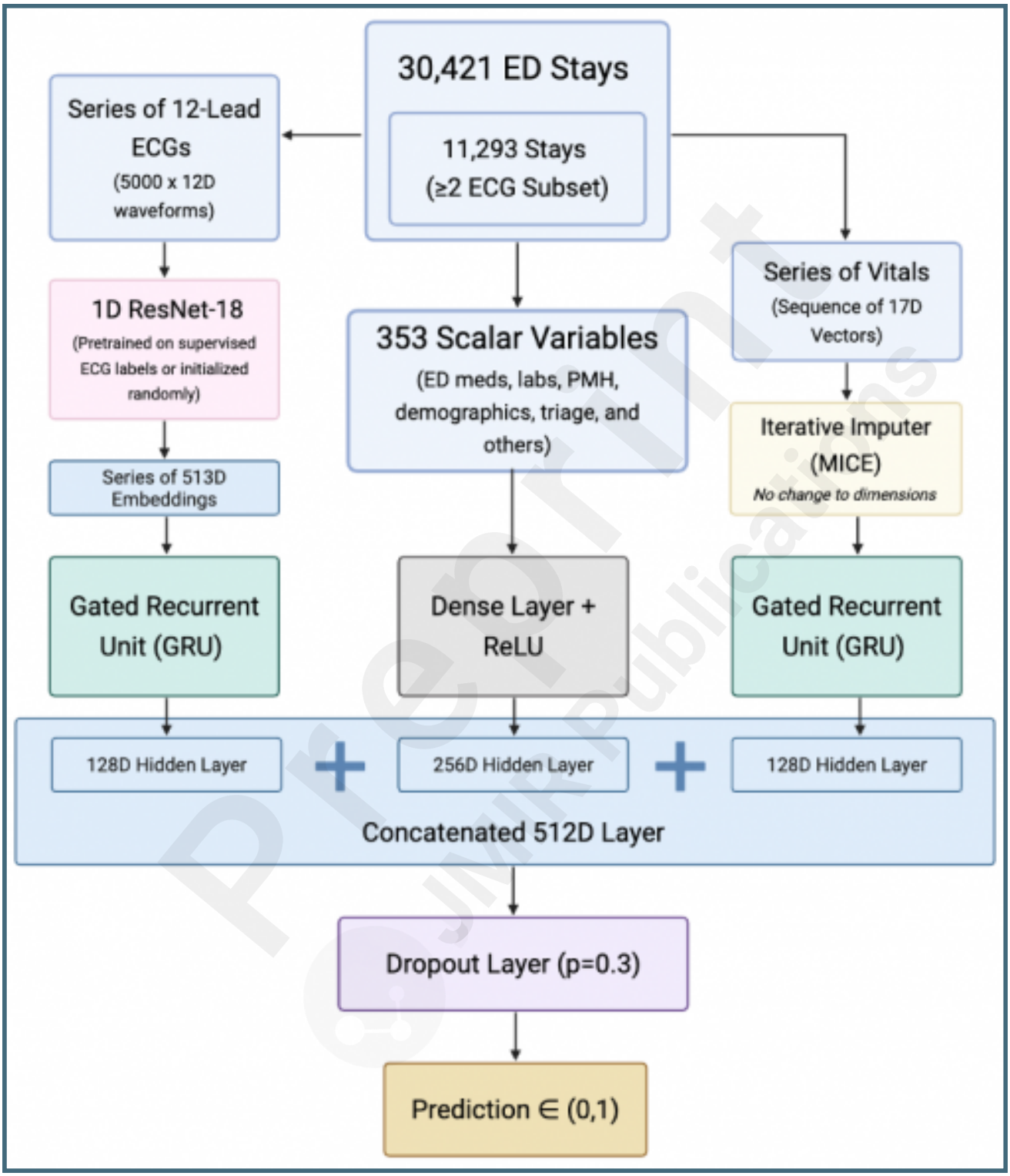
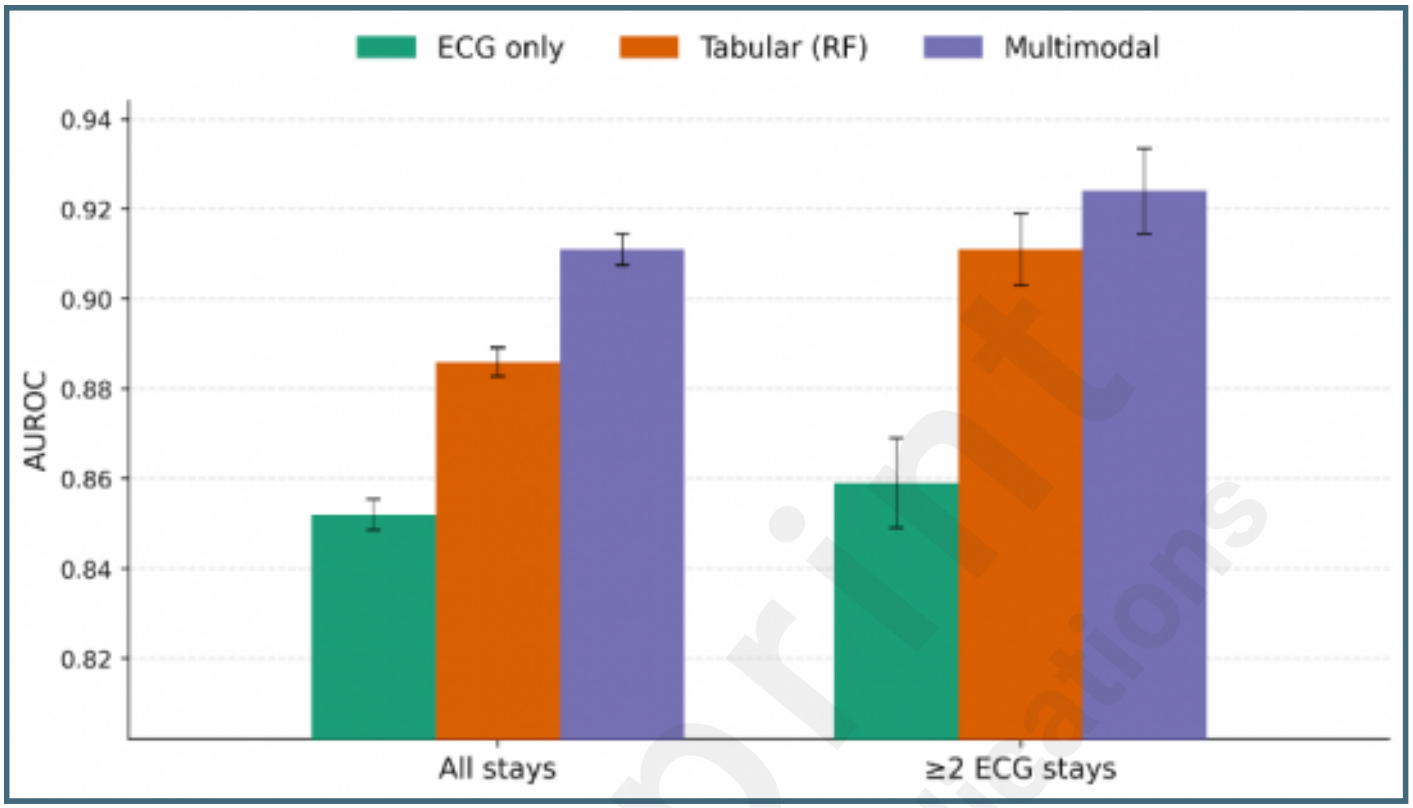


Diagram of the multimodal model architecture approach to predict hospitalization in the full and subset cohorts.



Mean validation AUROC of each model in each cohort plotted in a bar graph.



Multimedia Appendixes

Vitals measurements in all vs ?2-ECG cohorts.
URL: http://asset.jmir.pub/assets/ae66412be39a3635c550b85be7262ce7.docx

TOC/Feature image for homepages

An AI model on a hospital computer screen predicting patient admission risk in the emergency department.