PAPER

Multi-modal fusion model for predicting adverse cardiovascular outcome post percutaneous coronary intervention

IOP Publishing Physiol. Meas. 43 (2022) 124004 https://doi.org/10.1088/1361-6579/ac9e8a

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
9 September 2022

REVISED
18 October 2022

ACCEPTED FOR PUBLICATION
28 October 2022

PUBLISHED
22 December 2022

PAPER

Multi-modal fusion model for predicting adverse cardiovascular outcome post percutaneous coronary intervention

Amartya Bhattacharya[1], **Sudarsan Sadasivuni**[2], **Chieh-Ju Chao**[3], **Pradyumna Agasthi**[3], **Chadi Ayoub**[4], **David R Holmes**[3], **Reza Arsanjani**[4], **Arindam Sanyal**[5] **and Imon Banerjee**[4,5,*]

[1] Computer Science, University of Calcutta, India
[2] Electrical Engineering, University at Buffalo, Buffalo, United States of America
[3] Mayo Clinic Rochester, Rochester, Minnesota, United States of America
[4] Mayo Clinic Arizona, Scottsdale, Arizona, United States of America
[5] Arizona State University, Phoenix, Arizona, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: banerjee.imon@mayo.edu

Keywords: fusion model, multibranch network, cardiovascular outcome, percutaneous coronary intervention

Supplementary material for this article is available online

Abstract

Background. Clinical medicine relies heavily on the synthesis of information and data from multiple sources. However, often simple feature concatenation is used as a strategy for developing a multimodal machine learning model in the cardiovascular domain, and thus the models are often limited by pre-selected features and moderate accuracy. *Method.* We proposed a two-branched joint fusion model for fusing the 12-lead electrocardiogram (ECG) signal data with clinical variables from the electronic medical record (EMR) in an end-to-end deep learning architecture. The model follows the joint fusion scheme and learns complementary information from ECG and EMR. Retrospective data from the Mayo Clinic Health Systems across four sites for patients that underwent percutaneous coronary intervention (PCI) were obtained. Model performance was assessed by area under the receiver-operating characteristics (AUROC) and Delong's test. *Results.* The final cohort included 17,356 unique patients with a mean age of 67.2 $\pm$ 12.6 year (mean $\pm$ std) and 9,163 (52.7%) were male. The joint fusion model outperformed the ECG time-domain model with statistical margin. The model with clinical data obtained the highest AUROC for all-cause mortality (0.91 at 6 months) but the joint fusion model outperformed for cardiovascular outcomes - heart failure hospitalization and ischemic stroke with a significant margin (Delong's $p < 0.05$). *Conclusion.* To the best of our knowledge, this is the first study that developed a deep learning model with joint fusion architecture for the prediction of post-PCI prognosis and outperformed machine learning models developed using traditional single-source features (clinical variables or ECG features). Adding ECG data with clinical variables did not improve prediction of all-cause mortality as may be expected, but the improved performance of related cardiac outcomes shows that the fusion of ECG generates additional value.

1. Introduction

Percutaneous coronary intervention (PCI) is one of the most commonly utilized therapeutic procedures for patients with coronary artery disease (CAD) and acute coronary syndrome (ACS) (Peterson *et al* 2010, Alkhouli *et al* 2020, Tsao *et al* 2022), pp. 2003–2016]. There has been an increase in overall volume, clinical acuity and procedural complexity of PCI in the past decade (Inohara *et al* 2020, Kataruka *et al* 2020). Determining the post-PCI outcomes has gained significant research interest, hoping to enhance the decision making process and mitigate unwanted adverse prognosis. In the past, conventional risk score models have been developed to estimate short-term and long-term prognosis post-PCI (Chen *et al* 2010, Peterson *et al* 2010, McAllister *et al* 2016). With the advance of artificial intelligence (AI), many AI-based risk-prediction models have been

developed for the same task, and are reported to outperform conventional models (Peterson *et al* 2010). While many of these models achieved good performance, most of them primarily utilized data from a single category (Mortazavi *et al* 2019, Liu *et al* 2021), e.g., curated electronic medical records (EMR).

The practice of contemporary medicine relies heavily on the synthesis of information and data from multiple sources by clinicians; this includes imaging pixel data, vital measures, structured laboratory data, unstructured narrative data, and in some cases, audio or observational data. A deep learning model with a fusion architecture can learn simultaneously complementary information from different data resources and boost overall performance of the model. However, only about 10% of published work adopts this approach (Huang *et al* 2020a). Our group previously reported multiple fusion models which combine multimodal data: the EMR-CT image fusion model for the detection of pulmonary embolism (Huang *et al* 2020b), the EMR and ECG fusion model for predicting sepsis (Sadasivuni *et al* 2022), and in every case, the fusion model outperformed the EMR-only model. However, existing fusion models in healthcare primarily apply *early fusion* where various features coming from different modalities are simply combined together which results in a very sparse high dimensional feature representation and thus, often pre-selection of clinical variables becomes necessary for training feasibility. For example, in the field of cardiovascular medicine, Hedeweg *et al* (Heldeweg *et al* 2016) designed a simple multivariable logistic regression model for 30 day major adverse cardiovascular event (MACE) prediction using an early fusion of variables computed from 12-lead ECG, vital and demographic data and achieved moderate accuracy using pre-feature selection. Similar techniques are also adopted by others (Runge Chen *et al* 2017, Mezzatesta *et al* 2019) for the prediction of various cardiac outcomes. A potential weakness of this approach is that early fusion assumes that the modalities are well aligned and describe similar semantics. Electrocardiograms (ECG) have been used as the first-line diagnostic tool for ACS (Amsterdam *et al* 2014), and are known to contain important prognostic information by stratifying acuity and suggesting vessels involved (Hersi 2003, Stebbins *et al* 2010). Existing models primarily compute a limited set of time/frequency domain features from the ECG data to be incorporated with EMR, which is also susceptible to noise and variations. However, to the best of our knowledge, a model that directly incorporates the full ECG signal data with the EMR is not established yet for detecting post-PCI prognosis. In this context, we propose an end-to-end deep learning model with a joint-fusion architecture that incorporates EMR and ECG signal to predict different endpoints in patients post PCI. Joint fusion is implemented with neural networks due to their ability to propagate loss from the prediction model directly to the feature extraction layers and optimize the co-learning between modalities. We hypothesize that the joint fusion model which directly reads the 12-lead ECG data and EMR can outperform the individual models based on a single data source in predicting different post PCI adverse endpoints.

2. Methods

2.1. Patient population

Retrospective data from the Mayo Health Clinic Systems across four sites (La Crosse, WI; Mankato, MN; Rochester, MN; Scottsdale, AZ) for patients that underwent percutaneous coronary intervention between January 2006 and December 2018 were obtained. Patients who were lost to follow-up after PCI were excluded. The study was approved by the Mayo Institutional Review Board (IRB). A total of 21,872 patients that underwent percutaneous coronary intervention (PCI) between January 2006 and December 2018 were initially identified. After excluding the patients without follow-up at our institution, the final cohort included 17,356 unique patients. The mean age of the cohort was 67.2 $\pm$ 12.6 year (mean $\pm$ std) and 9,163 (52.7%) were male. Table 1 summarizes patient characteristics.

2.2. Clinical variables

A total of 157 baseline demographic features after one-hot encoding of 60 variables are utilized in this model and recorded from the institutional registry (see *supplemental* table 1). We collected the past medical history (myocardial infarction (MI), coronary artery bypass graft (CABG), chronic kidney disease, number of tumors, long axis length of largest tumor), clinical procedural information, and inpatient and outpatient medication usage. These features were hand-curated by the physicians. There was only $<5\%$ missing data overall and we coded the missing data as -1. We converted the variables into categories and afterward represented them as one-hot encoding. These records are collected utilizing International Classification of Disease (ICD 9/10) structured diagnosis code, and Current Procedural Terminology (CPT) codes from the standard clinical database. Clinical endpoints include heart failure hospitalization (HFH), ischemic stroke, and all-cause mortality at 6-months after the index PCI procedure.

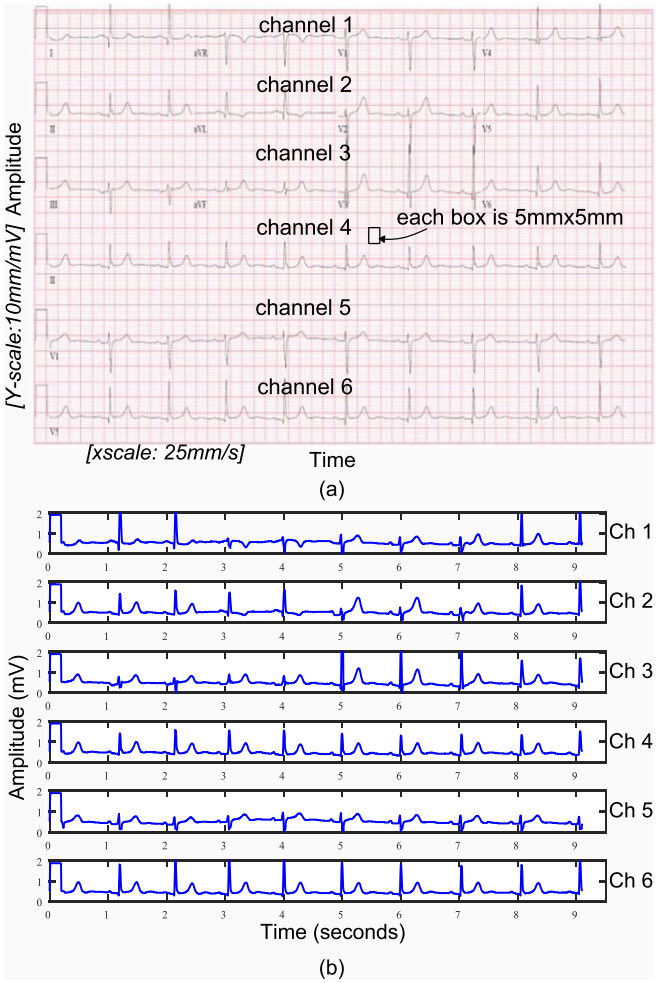

Figure 1. Standard ECG data processing pipeline. (a) 12 lead ECG strip images; (b) reconstructed 6 channel signal data.

Table 1. Distribution of patient cohort included in this study and breakdown into train and test sets.

Characteristics	Subtypes	Total cohort	Train set	Test set
Age	67.2 ± 12.6 Years	68.1 ± 12.3 Years	66.2 ± 12.8 Years	
Gender	Male	9163 (70.13%)	4832 (70.32%)	4331 (69.93%)
	Female	3901 (29.86%)	2039 (29.67%)	1862 (30.06%)
Smoker	2529 (19.35%)	1164 (16.94%)	1365 (22.04%)	
Family history of CAD	2654 (20.31%)	1313 (19.11%)	1341 (21.65%)	
Comorbidities	Diabetes	3717 (28.45%)	1942 (28.26%)	1775 (26.66%)
	Hypertension	10039 (76.84%)	5278 (76.81%)	4761 (76.88%)
	Chronic Kidney Disease			
	Mild	2135 (16.34%)	1160 (16.88%)	975 (15.74%)
	Moderate	916 (7.01%)	515 (7.49%)	401 (6.47%)
	Severe	793 (6.07%)	395 (5.75%)	398 (6.43%)
	Any cancer	1672 (12.8%)	972 (14.15%)	700 (11.30%)

2.3. ECG data processing

12-lead standard 10 s ECG strips are recorded from the targeted patients in the clinic before the PCI (figure 1(a)). The ECG data are collected as standard images with a standard grid and do not contain any personal information about the patient. A standard 12-lead ECG displays signals in 6 rows. The top 3 rows contain the individual lead signal (I, II, III, aVR, aVF, aVL, and V1-V6), each span 2.5 s. The bottom 3 rows are full 'rhythm strip' of arbitrary leads, spanning the whole 10 s of the ECG(Supplemental table 2). We developed a python module to extract the 12-lead ECG signals into a 6-channel array, corresponding to the 6 rows (figure 1(b)) to directly feed

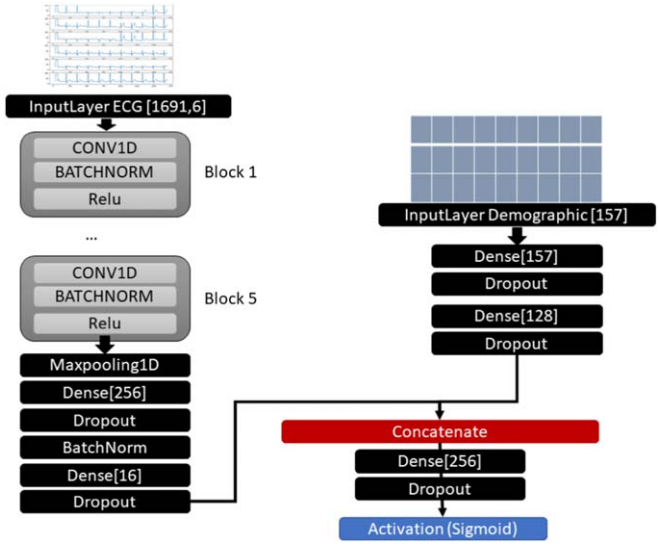

Figure 2. Block diagram of the proposed ECG and EMR joint fusion models. 1D convolutional blocks are used to process the ECG data and stacked layers with non-linear activation are used to process the EMR.

Table 2. Comparative performance of two baseline and fusion models. **Bold** represents optimal performance.

Outcome	Model	Precision	Recall	f1-score
Mortality 6 months	ECG only (time domain)	0.90 [0.89–0.90]	0.60 [0.59–0.60]	0.70 [0.70–0.71]
	EMR only	0.92 [0.91–0.94]	0.80[0.79–0.80]	0.84 [0.84–0.85]
	Fusion model	0.94 [0.93–0.94]	0.83[0.82–0.83]	0.87 [0.86–0.87]
Stroke 6 months	ECG only (time domain)	0.99 [0.99–0.99]	0.41[0.40–0.41]	0.58 [0.57–0.58]
	EMR only	0.99[0.98–0.99]	0.83[0.83–0.84]	0.90 [0.90–0.91]
	Fusion model	0.99 [0.99–0.99]	0.84[0.83–0.84]	0.91 [0.90–0.91]
Heart failure 6 months	ECG only (time domain)	0.99 [0.98–0.99]	0.52 [0.51–0.52]	0.68 [0.67–0.68]
	EMR only	0.96 [0.95–0.96]	0.67[0.66–0.67]	0.78 [0.77–0.77]
	Fusion model	0.96 [0.96–0.97]	0.77[0.76–0.77]	0.85 [0.84–0.85]

into the deep learning models. We manually evaluated 100 random samples to validate the data loss which is $<1\%$.

2.4. Joint fusion modelling using 12 leads ECG and clinical data

We proposed a two-branched deep learning model for fusing the 12 lead standard 10 s ECG strips with the clinical variables (see figure 1). The model follows the joint fusion scheme and learns an end-to-end representation of fusion. In other words, the proposed joint fusion (or intermediate fusion) model joins learned feature representations from intermediate layers of neural networks with features from other modalities as input to a final model and thus, optimizing the multimodal learning. Supplemental table 2 shows the correspondence between the 12 ECG leads and the 6 channels.

We designed a joint fusion architecture where both the ECG data as well as the EMR data can be parsed simultaneously. For processing, two separate branches were created i.e. one branch focusing on extracting features from multiple channel ECG data and the other, on the EMR data (see figure 2 for the architecture). The ECG was fed in the form of six separate channels of time-series data to fetch the maximum information. For extraction of relevant features from these six leads, 5 stacked convolutional blocks were used, each block placed consecutively. Each of these blocks contained a convolutional 1D layer, followed by a batch normalization layer and a Rectified Linear Unit (ReLU) activation layer respectively. After passing the initial ECG data through these layers, the results were then downsampled using a Maxpooling1D layer. This downsampled information was then connected to a dense layer containing 256 neurons followed by a dropout layer having a dropout ratio of 0.25 and a batch-normalization layer, which normalized results coming from the 256 neurons. Finally, before merging the ECG branch, with the branch focusing on EMR data, the results coming from the previous layer were passed through a dense layer containing 16 neurons, followed by a dropout layer having a dropout ratio of 0.25. The result coming from the final dense layer was then stacked with the results coming from the branch

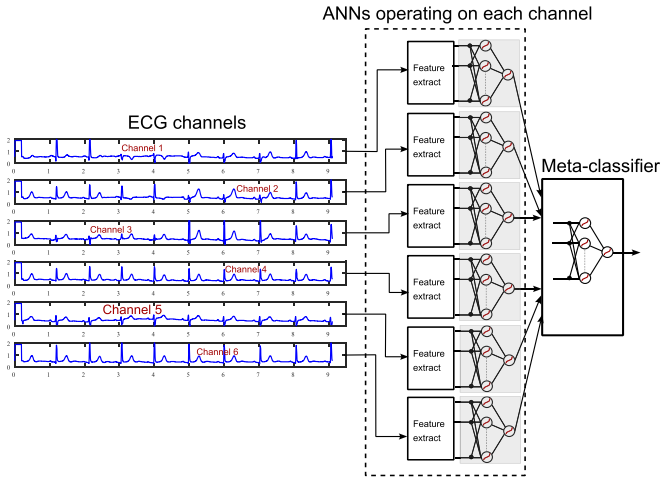

Figure 3. Late fusion model for prediction from ECG signal using time-domain features. Time-domain features are extracted from each ECG channel, and 6 ANNs are used for prediction using features from each ECG channel. The prediction scores from each of the 6 ANNs are sent to another ANN meta-classifier for fusion and final prediction.

containing information on the ECG data. After these two branches were merged, it passed through a dense layer having 256 neurons and a dropout layer respectively. The number of neurons and layers in the fusion model are selected empirically by observing the validation loss. The results from the dropout layer were connected to a single neuron which was responsible for giving a binary output.

2.5. Training strategy

One of the biggest challenges in the target task was the imbalanced dataset. Thus instead of the common binary cross-entropy loss function, the focal loss was used which applies a modulating term to the cross-entropy loss in order to focus learning on hard misclassified examples. In other words, it is a dynamically scaled cross-entropy loss, where the scaling factor decays to zero as confidence in the correct class increases. This strategy was applied to ensure that the model didn't get biased toward the negative samples which were present in huge proportions concerning the positive samples. The optimizer was chosen to be the Adam optimizer, having a learning rate of 1e-4, which was set after tuning the hyperparameter. Training was done by randomly splitting the training data into 60:20:20, train: validation: test split, and the performance was measured using a threshold agnostic metric, the ROC-AUC score. Validation loss was monitored throughout the whole training process and the early stopping criteria were designed to track validation loss; if no significant change in validation loss was observed for the 10 consecutive epochs, training was stopped.

2.5.1. Baseline 1. A predictive model using time-domain analysis of 12 leads ECG

The first baseline that we evaluated was the model developed using ECG time-domain features which is a popular way of extracting information from the ECG waveforms with low computational complexity (Chouvarda *et al* 2019). The ECG signal is normalized between [−1,1] before feature extraction and to remove baseline wander, we calculated the median value of the ECG signal window and subtracted it from all the samples in that window. For feature extraction, we consider only time-domain features based on first-order statistics of R-R peaks. We extract 14 time-domain features on ECG segments using Matlab and the extracted features are measures of central tendency, dispersion, the shape of the distribution of a window, R-R peaks, R-R intervals, the variance between R peaks, and average heart rate. The R peaks are identified in the time domain through thresholding. The threshold for peak detection is set to 30% of the difference between the maximum and minimum values for each segment. The extracted 14 time-domain features are used to train AI models for the prediction of various conditions. Considering computational efficiency, the 12-lead signals were combined into 6 channels, as summarized in Supplemental table 2. AI models are trained on each ECG channel and the prediction scores from each AI model are combined using a meta-classifier that gives the final prediction result through late fusion as shown in figure 3. A 2-layer ANN with 10 hidden neurons, tanh activation in the hidden layer, and softmax activation in the output layer is trained on each ECG channel and also used as the meta-classifier. Table 2 summarizes the precision, recall, and F1 score with a 95% confidence interval for the prediction of all-cause mortality, ischemic stroke, and heart failure at 6 months before the event.

2.5.2. Baseline 2. a predictive model using EMR branch deep learning model

To generate a strong baseline for clinical variables, we drop the ECG branch from the joint fusion model and use it only on the EMR branch. The EMR model was trained using only stacked dense layers since the variables are already coded into categorical space. The features were passed into a dense layer containing 157 neurons and a dropout layer having a dropout ratio of 0.25 simultaneously. The dense block was again repeated before passing the results coming from this layer to a final dense layer having 128 neurons which were then connected to a dropout layer. We used the same sigmoid activation and focal loss with early stopping to train the model.

3. Results

3.1. Model performance

Figure 4 shows quantitative performance of the models in terms of the area under the receiver operating characteristic curves (AUROC). Each ROC represents a unique outcome assessed on the same test set. In order to interpret the importance of different data sources, we evaluated two baselines: *ECG only*- this model only considers time domain ECG features and *EMR only* - this model only considers categorical EMR features. Time-domain ECG model (ECG only) obtained low AUROC for all-cause mortality and heart failure at 6 months (0.6 and 0.62, respectively). However, the ECG model obtained moderate AUROC for stroke prediction for 6 months (0.72). The EMR-only model obtained the highest AUROC for all-cause mortality 0.91 for 6 months. However, the joint fusion model outperformed the EMR-only model for heart failure and stroke by a significant margin (Delong's $p < 0.05$) - +0.1 improvement in heart failure and +0.11 improvement in stroke for 6 months prediction.

Operating points were individually selected from the ROCs based on the optimal tradeoff between true positive (high) and false positive (low) rate and corresponding precision, recall, and f1-score are reported in table 2. Confidence intervals (95%) were calculated based on 100 bootstrapping on the hold-out test set. It can be observed that adding ECG data in the fusion model consistently improves (>0.1) the recall for both stroke and heart failure predictions, and provides a model with a higher true positive rate while maintaining similar low false-positive rates. For deployment in clinical use, it is crucial to maintain a higher trade-off between true positive and false-positive rates for the targeted prediction task.

3.2. Ablation study

Given the fact that the EMR-only model provides similar performance in terms of *precision*, we performed a detailed ablation study to understand the importance of EMR features and multi-channel ECG for each targeted end-points, and how noisy data from each branch may impact the performance of the fusion model. We designed two parallel ablation studies and compared the performance:

(i) *Randomized the top clinical variable* identified by the Local interpretable model-agnostic explanations (Lime). In Lime, surrogate models are trained to approximate the predictions of the underlying black-box model to provide an explanation. Lime identified prior stroke as the most important feature for the EMR-only model prediction (figure 5(a)). We randomized the feature between (0,1) and used the trained fusion model to predict on the same test set with randomized data.

(ii)*Randomized a single channel of the ECG data* where we selected a channel of the ECG and randomly manipulated the values within a certain range. The manipulated signals are shown in figures 5(b) and (c). We used the trained fusion model to predict on the same test set with randomized ECG data.

Figure 6 presents the outcome of the ablation study in terms of AUROC for all the 6 targeted outcomes. Interestingly, even though the EMR-only model produced a comparable performance with the fusion model, the performance drop of the fusion model was not the highest when randomizing the top EMR feature - prior stroke. The highest performance drop was observed when the channel 4 were ablated with random values—particularly for stroke and heart failure.

4. Discussion of results

In this retrospective study, we developed a joint fusion network architecture that directly reads the ECG signal and EMR data and learns an optimized feature space. The model outperformed the existing models in predicting clinical endpoints at 6 months in post-PCI patients. The core contributions of this work are – (1) designing an end-to-end joint fusion architecture for integrating 12 lead ECG waveform with high dimensional clinical data; (2) demonstrating the importance of ECG waveform over EMR for predicting cardiac outcomes. To the best of

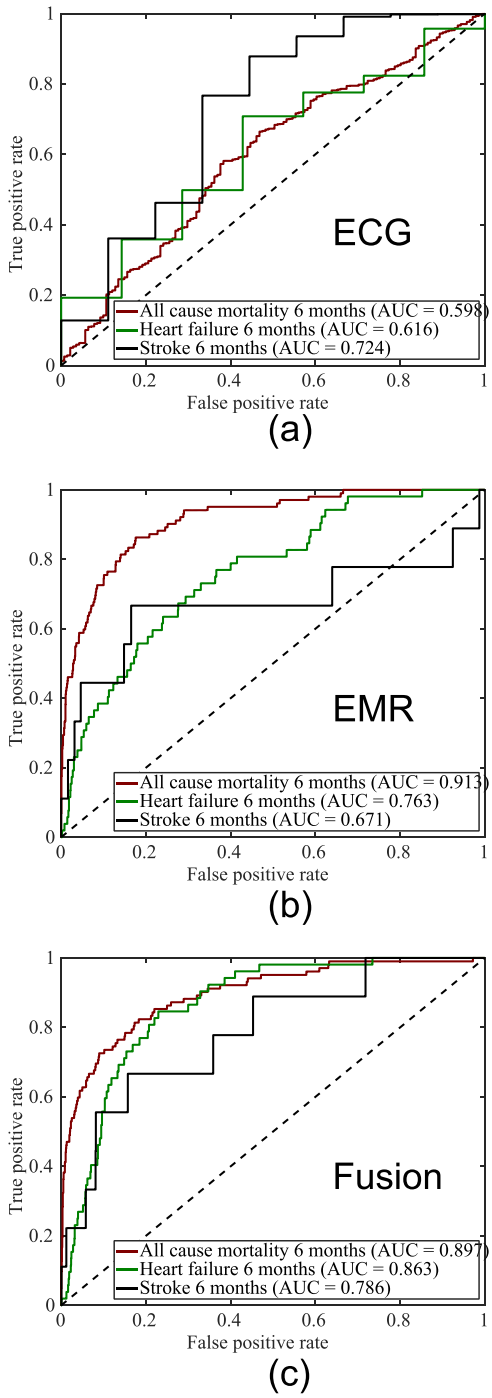

Figure 4. Receiver Operating characteristic curves (ROC) - legends show the area under the ROC. Panel (a) showed the prediction results using the time domain ECG, and Panel (b) showed the prediction results using the EMR model. Panel (c) showed the prediction results using the fusion model. Brown line: all-cause mortality, green line: heart failure hospitalization, black line: ischemic stroke.

our knowledge, this is the first study that developed a deep learning model with fusion architecture for the prediction of post-PCI prognosis. We observed that different data sources contain different weights of information for the model to predict prognosis. For all-cause mortality, the EMR-only model had the best performance compared to the ECG-only and the fusion model. It is reasonable that the EMR data contains more relevant information for different causes of mortality, while the ECG information could be more cardiac-specific. Additionally, through the ablation study, we demonstrated that our fusion model is capable of compensating for noise or information loss in both modalities. While prior stroke is known to be the more important variable for all-cause mortality and heart failure, altering the input of prior stroke did not lead to a significant drop in the overall model performance.

Table 3 summarizes the modeling strategy and performance of our fusion model with state-of-the-art AI models that aim to predict different CVD endpoints. Multiple works use a combination of vitals, EHR, and lab

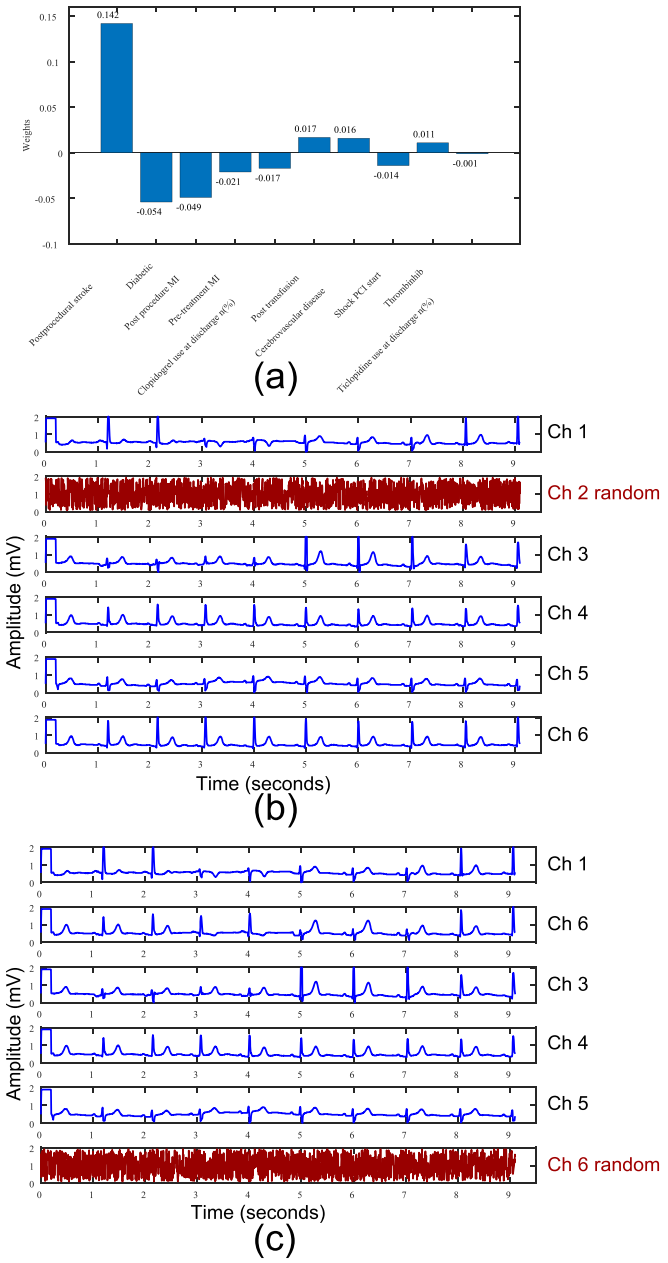

Figure 5. Visual explanation of the ablation study: (a) shows the local interpretable model-agnostic explanations (LIME) explanation of the electronic medical records (EMR) only model; (b) and (c) shows the ablation of the ECG data.

test results for the prediction of different cardiovascular endpoints. Previous work (Heldeweg *et al* 2016) predicted thirty-day MACE as the primary endpoint with AUROC of 0.78 and uses 10 clinical variables - age, gender, heart rate, 3 heart-rate variability parameters (average R-R interval, triangular interpolation of N-N intervals, and high-frequency power) and four 12-lead ECG variables (ST elevation, ST depression, Q wave, and QT prolongation). Other prior work (Siontis *et al* 2021) used only a 12-lead ECG and fusion CNN model to develop a point-of-care application for the prediction of at least one ECG showing AFib within 31 days after the sinus-rhythm ECG is classified as positive for AFib with an AUROC of 0.87. Another study (Runge Chen *et al* 2017) predicted the risk of coronary heart disease in patients with hypertension within 3 years after the first follow-up using EHR and ECG as clinical variables in a logistic regression model. Others (An *et al* 2021) have utilized a novel attention-based RNN model to predict the onset of cardiovascular diseases from high-dimensional EHR data with an AUROC of 0.77. Additionally, the risk of ischemic heart disease in patients on dialysis has been predicted over a time frame of 2.5 years using an SVM model with EHR predictors that result in an AUROC of 0.74 (Mezzatesta *et al* 2019). Compared to these previous models, our work achieves the highest performance metrics for predicting challenging long-term CVD outcomes (AUROC/precision/recall) using the fusion of EHR and ECG data. To the best of our knowledge, this is the first study to directly integrate the raw ECG waveform and EMR data in a single end-to-end deep learning model. Such a fusion model may have

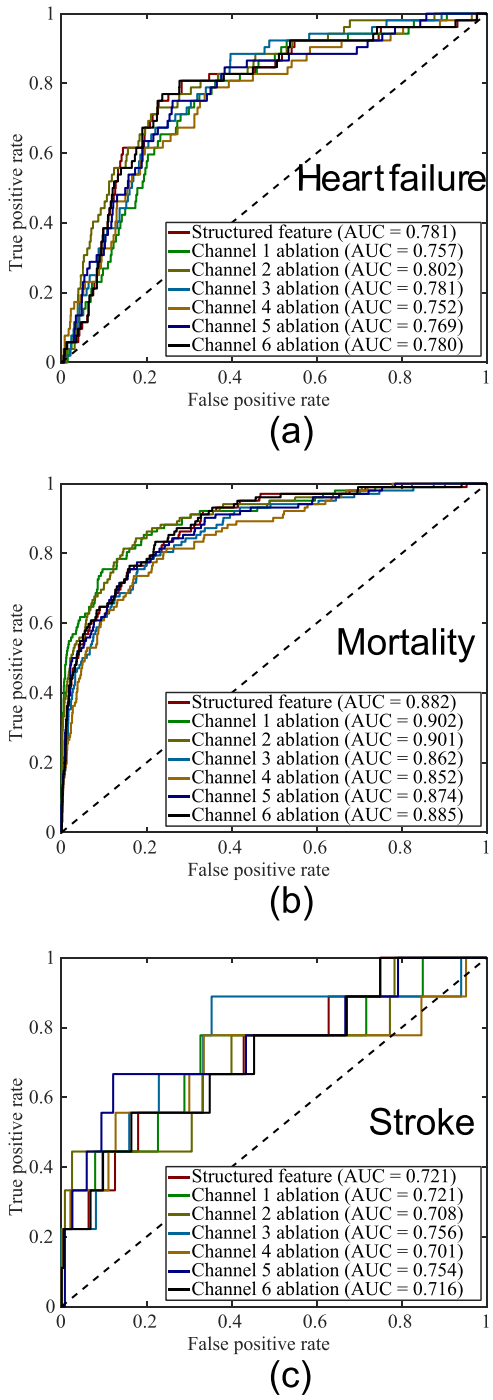

Figure 6. Ablation study results in terms of AUROC. The model performance: heart failure hospitalization, mortality, and Stroke at 6 months.

significant clinical impact, as it may identify patients at higher risk of adverse clinical events post PCI, who may benefit from closer clinical surveillance and potentially more aggressive medical therapies to help improve their prognosis.

Limitations

All-cause mortality was used as the endpoint instead of cardiovascular mortality, as part of our mortality data was based on the social security death index. Furthermore, the cardiovascular mortality endpoint has been considered less reliable in retrospective studies (Lauer *et al* 1999). Coronary artery anatomical information (SYNTAX score) was not incorporated in this model, which is known to have an effect on procedural complications and clinical outcomes. Given the complex multimodal and multidimensional (1D temporal data

Table 3. Comparative analysis of the state-of-the-art models for cardiovascular diseases (CVD) clinic end-point prediction.

Paper	Model	Prediction outcome	Performance metrics					Data sources			
			Time-to-event	AUROC	Precision	Recall	Vitals	ECG	Lab	Demographics	Clinical history
[13]	LR	MACE	1 month	0.78	0.52	0.82	Y	Y	Y	Y	Y
[20]	Fusion CNN	Silent AFib	1 month	0.92				Y			
[22]	Ensemble	CVD	1 month		0.93	0.89	Y	Y	Y	Y	Y
[23]	Cox	Stroke, myocardial infraction	2 years	0.73			Y		Y	Y	Y
[24]	RF	CVD	3 years	0.79			Y		Y	Y	Y
[25]	DNN	AFib, heart failure, stroke	2 years	0.78	0.52	0.08	Y		Y	Y	Y
[15]	LR	Coronary heart disease	3 years	0.84			Y		Y	Y	Y
[21]	DNN	CVD	2 years		0.67	0.81			Y	Y	Y
[14]	SVM	Ischemic heart disease	2 years		0.77	0.75	Y		Y	Y	Y
This work	CNN+ fusion	mortality, stroke, heart failure	6 months	0.89/0.86/0.79[1]	0.94/0.99/0.96[1]	0.83/0.84/0.77[1]	Y	Y	Y	Y	Y

and tabular data) nature of the study, none of the standard XAI approaches (such as LIME, SHAP) can be used for model interpretation. Thus, to evaluate the importance of the multimodal design, we performed a standard ablation study which targets to understand the contribution of the data components by ablating its value and recording the model performance. Our primary objective for this analysis was to show that even if the EMR-only model provides comparative performance to the fusion model, the final fusion model is able to retain the performance by consolidating multimodal data without considering the top weighted EMR features.

Ethical statement

Retrospective data from the Mayo Health Clinic Systems across four sites (La Crosse, WI; Mankato, MN; Rochester, MN; Scottsdale, AZ) for patients that underwent percutaneous coronary intervention between January 2006 and December 2018 were obtained. The study was approved by the Mayo Institutional Review Board (IRB). All methods were carried out in accordance with relevant guidelines and regulations (Declaration of Helsinki). The study is exempt from patient consent, the IRB approval number is 19-004469.

ORCID iDs

Sudarsan Sadasivuni https://orcid.org/0000-0002-9008-8426
Imon Banerjee https://orcid.org/0000-0002-3327-8004

References

Alkhouli M *et al* 2020 Trends in characteristics and outcomes of patients undergoing coronary revascularization in the united states, 2003-2016 *JAMA Netw. Open* **3** e1921326

Amsterdam E A *et al* 2014 2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes: a report of the american college of cardiology/american heart association task force on practice guidelines *Circulation* **38** 121–23

An Y, Huang N, Chen X, Wu F and Wang J 2021 High-risk prediction of cardiovascular diseases via attention-based deep neural networks *IEEE/ACM Trans. Comput. Biol. Bioinform.* **18** 1093–105

Chen S-L *et al* 2010 Comparison between the NERS (New Risk Stratification) score and the SYNTAX (synergy between percutaneous coronary intervention with taxus and cardiac surgery) score in outcome prediction for unprotected left main stenting *JACC Cardiovasc. Interv.* **3** 632–41

Chouvarda I, Filos D and Maglaveras N 2019 Time-domain analysis of the electrocardiogram *Cardiovascular Computing—Methodologies and Clinical Applications* ed S Golemati and K S Nikita (Singapore: Springer Singapore) pp 81–102

Heldeweg M L A *et al* 2016 A novel cardiovascular risk stratification model incorporating ECG and heart rate variability for patients presenting to the emergency department with chest pain *Crit. Care* **20** 1–9

Hersi A 2003 Does the discharge ECG provide additional prognostic insight(s) in non-ST elevation ACS patients from that acquired on admission? *Eur. Heart J.* **24** 522–31

Huang S C, Pareek A, Seyyedi S, Banerjee I and Lungren M P 2020a Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines *Npj Digit. Med.* **3** 1–9

Huang S-C, Pareek A, Zamanian R, Banerjee I and Lungren M P 2020b Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection *Sci. Rep.* **10** 1–9

Inohara T *et al* 2020 Comparative trends in percutaneous coronary intervention in japan and the united states, 2013 to 2017 *J. Am. Coll. Cardiol.* **76** 1328–40

Kataruka A *et al* 2020 Temporal trends in percutaneous coronary intervention and coronary artery bypass grafting: insights from the washington cardiac care outcomes assessment program *J. Am. Heart Assoc.* **9** e015317

Lauer M S, Blackstone E H, Young J B and Topol E J 1999 Cause of death in clinical research: time for a reassessment? *J. Am. Coll. Cardiol.* **34** 618–20

Liu S *et al* 2021 Machine learning-based long-term outcome prediction in patients undergoing percutaneous coronary intervention *Cardiovasc. Diagn. Ther.* **11** 736–43

McAllister K S L *et al* 2016 A contemporary risk model for predicting 30-day mortality following percutaneous coronary intervention in England and Wales *Int. J. Cardiol.* **210** 125–32

Mezzatesta S, Torino C, Meo P D, Fiumara G and Vilasi A 2019 A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis *Comput. Methods Programs Biomed.* **177** 9–15

Mortazavi B J *et al* 2019 Comparison of machine learning methods with national cardiovascular data registry models for prediction of risk of bleeding after percutaneous coronary intervention *JAMA Netw. Open* **2** e196835

Peterson E D *et al* 2010 Contemporary mortality risk prediction for percutaneous coronary intervention: results from 588,398 procedures in the national cardiovascular data registry *J. Am. Coll. Cardiol.* **55** 1923–32

Runge C *et al* 2017 3-year risk prediction of coronary heart disease in hypertension patients: a preliminary study *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* **2017** 1182–5

Sadasivuni S, Saha M, Bhatia N, Banerjee I and Sanyal A 2022 Fusion of fully integrated analog machine learning classifier with electronic medical records for real-time prediction of sepsis onset *Sci. Rep.* **12** 1–11

Siontis K C, Noseworthy P A, Attia Z I and Friedman P A 2021 Artificial intelligence-enhanced electrocardiography in cardiovascular disease management *Nat. Rev. Cardiol.* **18** 465–78

Stebbins A *et al* 2010 A model for predicting mortality in acute ST-segment elevation myocardial infarction treated with primary percutaneous coronary intervention: results from the assessment of pexelizumab in acute myocardial infarction trial *Circ. Cardiovasc. Interv.* **3** 414–22

Tsao C W *et al* 2022 Heart disease and stroke statistics-2022 update: a report from the american heart association *Circulation* **145** e153–639