

Prior electrocardiograms not useful for predicting major adverse cardiac events with machine learning

Axel Nyström^a, Pontus Olsson de Capretz^{b,c}, Anders Björkelund^d, Jakob Lundager Forberg^{c,e}, Mattias Ohlsson^{d,f}, Jonas Björk^{a,g}, Ulf Ekelund^{b,c}

^a*Lund University, Department of Laboratory Medicine, Lund, Sweden*

^b*Skåne University Hospital, Department of Internal and Emergency Medicine, Lund, Sweden*

^c*Lund University, Department of Clinical Sciences, Lund, Sweden*

^d*Lund University, Department of Astronomy and Theoretical Physics, Lund, Sweden*

^e*Helsingborg Hospital, Department of Emergency Medicine, Helsingborg, Sweden*

^f*Halmstad University, Center for Applied Intelligent Systems Research (CAISR), Halmstad, Sweden*

^g*Clinical Studies Sweden, Forum South, Skåne University Hospital, Lund, Sweden*

Abstract

At the emergency department (ED), it is important to quickly and accurately determine which patients are likely to have a major adverse cardiac event (MACE). Machine learning (ML) models can be used to aid physicians in detecting MACE, and improving the performance of such models is an active area of research. In this study, we sought to determine if ML models can be improved by including a prior electrocardiogram (ECG) from each patient. To that end, we trained several models to predict MACE both with and without prior ECGs, using data collected from 19499 consecutive patients with chest pain from five EDs in southern Sweden, between the years 2017 and 2018. Our results indicate no improvement in AUC from prior ECGs. This was consistent across models, both with and without additional clinical input variables, for different patient subgroups, and for different subsets of the outcome. While contradicting current best practices for manual ECG analysis, the results are positive in the sense that ML models with fewer inputs are more easily and widely applicable in practice.

Keywords: Machine learning, Neural networks, Emergency department, Chest pain, Major adverse cardiac event, Electrocardiograms

1. Introduction

Chest pain is the second most common complaint at the emergency department (ED), and cardiovascular disease is the most common cause of death in Europe [1]. However, only one out of ten ED chest pain patients suffer from major adverse cardiac events (MACE) within 30 days, including acute myocardial infarction (AMI) and cardiac arrest [2]. It is therefore important to quickly and accurately identify the patients who require immediate care, and those who

can wait or be sent home. With current diagnostic methods however, many patients undergo lengthy assessments in the ED and less than one in four of those admitted to in-hospital care prove to have MACE [?]. Many admissions and investigations are thereby “unnecessary,” and cause a substantial health care burden.

To identify or rule out MACE, the ED physician evaluates the patient history, current symptoms, blood markers of myocardial necrosis such as high-sensitivity troponin T (hs-cTnT), and the electrocardiogram (ECG). If available, a previous ECG is often also assessed since acute myocardial ischemia or infarction induces new ECG changes and since guidelines state that diagnostic accuracy can be improved by comparison with a prior ECG [3, 4].

Computer-aided ECG analysis systems have existed since the 1960s, but advances in machine learning (ML) in combination with the availability of digitized ECGs has resulted in a rapidly growing number of publications on ML in ECG interpretation in the past years [5, 6, 7]. However, few studies have investigated the usefulness of a prior ECG when predicting cardiac events using ML.

Ohlsson et al. [8] in 2001 found that the addition of a previous ECG improved the area under the ROC curve of a neural network model from 0.85 to 0.88 when predicting AMI, but since then the definition of AMI has changed and we can now identify smaller AMIs with more sensitive blood tests (e.g. hs-cTnT). Sbrollini et al. [9] built a neural network to detect cardiac ischemia based on serial ECGs, but although it outperformed logistic regression, the performance gain from the addition of a previous ECG was unclear. ter Haar et al. [10] used features derived from serial ECGs as input to a logistic regression model predicting myocardial ischemia and did not find any improvement over the single ECG predictions from the Glasgow algorithm. All of these three studies used a small number of hand crafted input features from each ECG. To the best of our knowledge, the benefit of prior ECGs for ML classification of cardiac events using full ECG signals has not yet been explored.

In this study, we sought to determine if the prediction of MACE by ECG-based ML models in ED chest pain patients can be improved by the addition of a previous ECG. We compared ML models of varying complexity trained both with and without data on patient age, sex and hs-cTnT results.

2. Methods

2.1. Study design

This was a retrospective study based on data from the ESC-TROP study, which includes 26545 unique, consecutive adult (≥ 18 years) patients presenting with chest pain to five EDs in southern Sweden between Feb 1, 2017 and Nov 30, 2018 [11]. For each patient, the ESC-TROP database only includes the first ED visit during the two year study period. The database includes ECGs and blood sample data from the index visit, electronic health records up to five years prior to the index visit and all medical diagnoses made in the geographic region

(Skåne) up to 30 days after the index visit. The study was approved by the Regional Ethics Review Board in Lund, Sweden (Dnr 2017/831 and 2018/708).

2.2. Patient population

We included all patients in the ESC-TROP database where a hs-cTnT test as well as an ECG was recorded at the ED, and where a final diagnosis of STEMI was not made during the index visit. We excluded all patients with no ECG recorded within 2 hours after ED arrival, and all patients with no prior ECG of good technical quality after the year 1999.

The data was split chronologically into three parts: a training set, containing the first half of the patients, a validation set, containing the third quarter of the patients, and a test set, containing the last quarter of the patients. The training set was used to train the machine learning models, the validation set was used to perform model selection, and the test set was used to assess the performance of the final models. We chose to split the data chronologically rather than randomly, since it allows for a stronger indication of model generalizability [12].

2.3. Data processing

2.3.1. Outcome measures

The target outcome for the models was acute myocardial ischemia (AMI) or related complications in the form of major adverse cardiac events (MACE), including death from any cause within 30 days of the index visit. MACE was defined as unstable angina (UA), atrioventricular block type 2 or 3, ventricular arrhythmia requiring acute intervention, cardiac arrest, pulmonary edema, cardiogenic shock, coronary artery bypass grafting, percutaneous coronary intervention, transvenous pacemaker insertion, temporary cardiac pacing, or death from any cause. The complete list of ICD-10 diagnosis and intervention codes are listed in Appendix A, Table A.2.

2.3.2. ECG processing

The ESC-TROP database contains 536135 ECGs recorded in the included patients between 1970 and 2019 at health care institutions in region Skåne. Each ECG contains 12 leads and is 10 seconds long, sampled at 500 Hz. In each patient, the ECG recorded closest to the ED arrival was considered the index ECG. The previous ECG was chosen as the newest ECG recorded at least one week prior to ED arrival. This interval was chosen to minimize the risk of including ECGs recorded during the same acute episode.

2.3.3. Glasgow features

In order to compare models that use the raw input ECG signal directly with standard statistical methods such as logistic regression, we also extracted from each ECG a subset of 228 features derived from the Glasgow algorithm [13]. Specifically, for each of the 12 leads we used the Q , R , S , $T+$, $T-$, ST , $ST_{2/8}$, and $ST_{3/8}$ amplitudes as well as the Q , R , S , and QRS durations, the QRS area and ST slope. The ST amplitudes, QRS area and ST slopes were further

divided into positive and negative parts, so that the total number of features was $12 \cdot 19 = 228$. This set of features has previously been useful for predicting AMI [14].

2.3.4. Additional clinical variables

The ED physician considers more information than the ECG to decide patient management, and since this might affect the value of a previous ECG, we also included age, sex, time since the previous ECG, and the first hs-cTnT result as additional clinical variables. The distribution of the hs-cTnT results and the time between ECG recordings are heavily skewed and were therefore logarithmized before use in the models.

The hs-cTnT blood samples were those used in the care of the patient and collected in lithium heparin tubes and analyzed using Roche Cobas e602 (Roche Diagnostics). The assay has a limit of blank of 3 ng L^{-1} and a limit of detection of 5 ng L^{-1} . The 99th percentile cut-off is 14 ng L^{-1} and the coefficient of variation is $< 10\%$ at 13 ng L^{-1} [15].

2.4. Machine learning methods

Four types of models of increasing complexity were evaluated with four different sets of inputs: the index ECG with and without the previous ECG, and with and without the additional clinical variables. Each of the four model types were trained for each of the four sets of inputs, resulting in a total of 16 different models. The four models were, in order of complexity: logistic regression (LR), multilayer perceptron (MLP), convolutional neural networks (CNN), and a residual neural network model (RN) pre-trained by Ribeiro et al. [16] on two million ECGs.

The basic structure of the models and our strategy for combining different inputs is illustrated in Figure 1. The ECGs were first passed through a feature extraction function, which is the same for the index and prior ECGs, but can differ between the different models. The output from the feature extraction is combined in a combiner function, which is followed by a classification function that makes the final prediction. The hyper-parameters were primarily determined through random search, the details of which is described in further detail in the supplementary material ??.

2.4.1. Logistic regression

The LR models used as input the 228 features per ECG derived from the Glasgow algorithm as described in Section 2.3.3, pre-processed with principal component analysis (PCA) dimensionality reduction. The optimal target dimension for the PCA was determined through grid-search using the validation set. For the models using two ECGs, the features from the first ECG were concatenated with the difference between the features from the index and prior ECGs. The dimensionality reduction was applied on the Glasgow features before being concatenated with the additional clinical variables.

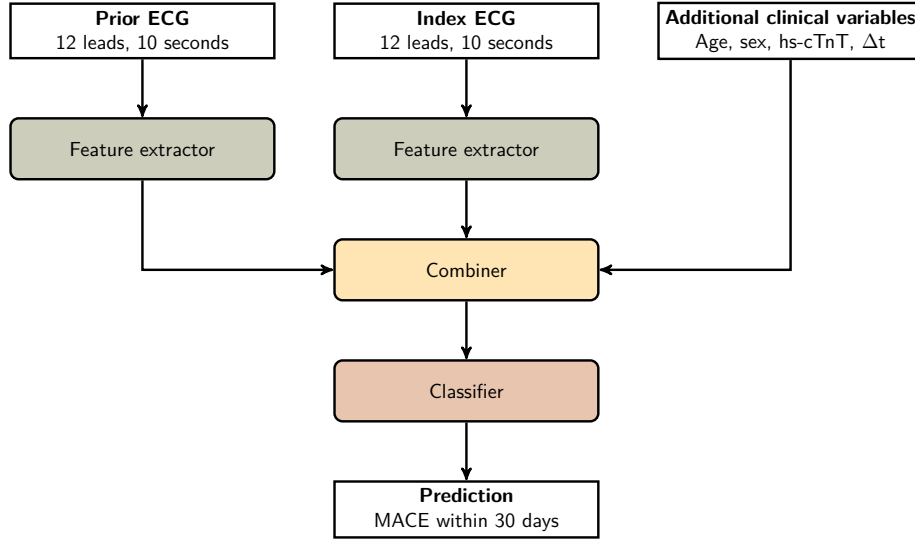


Figure 1: Model architecture. Abstract overview of the basic structure of the machine learning models. Each ECG is passed through a feature extractor function before being combined with additional clinical variables. The feature extractor is the same within a model, but can differ between models. All models make use of the index ECG (middle column), but the prior ECG (left column) and additional clinical variables (right column) are only used for some of the models. Δt is the time between index and prior ECG.

2.4.2. Multilayer perceptron

The MLPs were trained using the 228 Glasgow-features described in Section 2.3.3, but without dimensionality reduction. The hyper-parameters for each of the four sets of inputs (one or two ECGs, with or without additional clinical variables) were determined through random search using the validation set.

2.4.3. Convolutional neural networks

The CNNs used the full ECG signals, rather than Glasgow-features as inputs. Similarly to the MLPs, the hyper-parameters were determined through random search using the validation set.

2.4.4. Residual neural networks

The RN models were based on a pre-trained model developed by Ribeiro et al. [16], where approximately two million ECGs were used to train the network to classify 6 types of ECG abnormalities. In order to use the RN model, each input ECG was resampled and rescaled, and the leads were reordered to match the format expected by the network. The convolutional part of the RN model was kept fixed (but all weights remained trainable), the final dense output layer was removed and replaced by a small MLP. The final network, together with

other hyper-parameters, were determined by random search using the validation set.

2.4.5. Ensembles

Due to the stochastic nature of the training process, there was a small variance in the performance of the neural network models, depending on their initialization. We reduced this variance by forming ensembles of 10 identical models, each initialized with different random seeds. The output of the ensemble was taken as the arithmetic mean of each constituent model output.

2.5. Methods for analyzing the results

2.5.1. Comparing models

In this study, we aimed to quantify the added value of a previous ECG in terms of AUC when predicting MACE within 30 days. There were two primary comparisons: Models using the index ECG versus the index and the old ECG, and models with versus without additional clinical variables. For each model, an approximate 95% confidence interval for the AUC score on the test set was constructed using the percentile bootstrap method with $B = 10000$ bootstrap samples [17].

2.5.2. Subgroup analyses

In clinical practice, previous ECGs are typically only reviewed if there are pathological changes in the index ECG, since changes also present in previous ECGs are believed to be less concerning. The added value of a previous ECG might thus be higher in patients with a pathological index ECG. To test this hypothesis, we used the Glasgow diagnoses that indicate myocardial ischemia or AMI, and classified all ECGs labeled with one or more of these as pathological. We then examined the performance of the algorithms on pathological and non-pathological ECGs separately.

It is also possible that the added value of a previous ECG for detecting MACE is higher in older patients because of a higher prevalence of comorbidities, and we therefore divided the patients in the test set into age quartiles; 18 - 50 years, 51 - 64 years, 65 - 75 years, and over 76 years.

Further, since AMI can result in persistent ECG changes and is associated with an increased risk of future cardiovascular events, it is possible that the benefit of a previous ECG differs in patients with or without a prior AMI [18, 19]. We evaluated the models separately on patients with and without AMI in the previous 5 years.

Finally, we analyzed the individual components of MACE. Since some diagnoses were rare, we used only three outcome groups: Unstable angina (UA, ICD-10: I20), AMI (ICD-10: I21 or I22), and all-cause death. The AUC of each model was evaluated in each group separately. It should be noted that all models were still only trained to predict the composite outcome MACE.

3. Results

3.1. Patient characteristics

As illustrated in Figure 2, the original ESC-TROP database contains 26545 distinct chest pain patients, of which 7046 were excluded because of a diagnosis of STEMI during the index visit, missing hs-cTnT lab value, or missing or low-quality ECGs. The remaining 19499 patients were split chronologically into a training set ($n = 9750$; 50%), a validation set ($n = 4875$; 25%) and a test set ($n = 4874$, 25%).

Table 1 shows the characteristics of the included patients. While the frequency of MACE in the training, validation and test sets is similar, we note that the disease history differs substantially, to the effect that patients in the training set are much more prone to have a history of AMI, congestive heart failure, and pulmonary disease, among others. The age of the prior ECG is also higher for patients in the validation and test sets than for the patients in the training set.

3.2. Main results

Figure 3 shows the AUC with confidence intervals for the different models. In general, there was no meaningful improvement in the prediction of MACE within 30 days with the addition of a prior ECG. Some models performed slightly better with an added prior ECG, but other models performed still better with only a single ECG. For example, the RN model performed slightly better with two ECGs than with one (AUC 0.766 versus 0.758), but the MLP model with one ECG was even better (AUC 0.774 versus 0.772). The intra-model variance of the AUC was typically an order of magnitude larger than any difference between models with and without prior ECGs, arguably rendering meaningless any statistically significant differences between such models. The complete results for all the models and comparisons are included in Appendix B.

3.3. Models with versus without additional clinical variables

As can be seen in Figure 3, all models improved considerably with the inclusion of additional clinical variables. Models using only the ECG had an AUC around 0.77 (95% CI for the best model: 0.754-0.796), while those including the additional variables had an AUC around 0.87 (95% CI for the best model: 0.864-0.893). The differences between models with and without a prior ECG were small in comparison. Of the four models, the MLP model performed the best, both with and without additional variables.

3.4. Patients with versus without pathological index ECGs

Figure 4 shows the results stratified by the presence or absence of a pathological index ECG, for the models including additional variables. In general, all models performed slightly better in patients with non-pathological index ECGs, but there was no improvement by the addition of a prior ECG. Models based on only the ECG performed slightly better with an added prior ECG, but the best performing model (MLP) did not improve.

Patient characteristics	Train	Validation	Test	Total
n (%)	9750 (50)	4875 (25)	4874 (25)	19499 (100)
Male, %	50.7 (50)	49.7	49.6	50.2
Age, mean (std)	63 (17.6)	62.2 (17.5)	62.4 (17.6)	62.6 (17.6)
MACE within 30 days, %	10.4 (50)	10.8	10.8	10.6
Disease history				
Acute Myocardial Infarction, %	9.1	5.5	5.9	7.4
Congestive Heart Failure, %	8.4	6.1	6.0	7.2
Peripheral Vascular Disease, %	1.9	1.3	1.7	1.7
Cerebral Vascular Accident, %	6.9	6.0	5.9	6.4
Dementia, %	0.7	0.5	0.5	0.6
Pulmonary Disease, %	7.9	5.4	6.6	6.9
Connective Tissue Disorder, %	1.0	0.9	1.1	1.0
Liver Disease, %	0.3	0.3	0.4	0.3
Severe Liver Disease, %	0.1	0.1	0.2	0.1
Diabetes, %	11.1	8.9	8.5	9.9
Diabetes Complications, %	1.2	0.8	1.1	1.1
Renal Disease, %	3.4	2.5	2.9	3.0
Cancer, %	5.8	5.2	5.4	5.5
Metastatic Cancer, %	1.0	1.2	1.3	1.1
Additional Clinical Variables				
hs-cTnT, median (IQR)	8 (4-17)	7 (4-16)	8 (4-17)	8 (4-17)
Minutes to hs-cTnT, median (IQR)	34 (20-60)	35 (20-62)	31 (18-54)	34 (19-59)
Days between ECGs, median (IQR)	285 (88-807)	372 (95-1006)	372 (122-984)	326 (97-909)
ECG				
Pathological diagnosis on ECG, %	42.4	40.2	40.2	41.3

Table 1: Patient characteristics

n, number; MACE, Major Adverse Cardiovascular Event; std, standard deviation; IQR, inter quartile range; hs-cTnT, high sensitivity cardiac troponin T; pathological diagnosis according to Glasgow on index ECG; Disease history recorded up to 5 years prior to the study event.

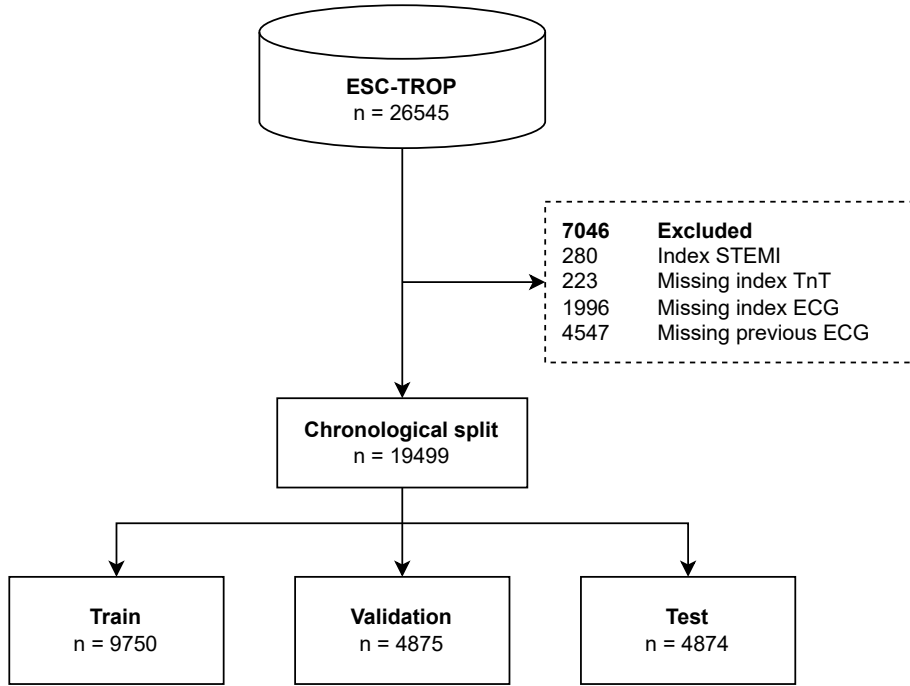


Figure 2: Inclusion criteria

The ESC-TROP database contains healthcare data from adult patients with chest pain as primary complaint at the ED between 2017 and 2018. Only the first chest-pain visit from each patient is included in ESC-TROP. The chronological split was done such that the patients in the first half of the study period were included in the training set, the patients in the third quarter were included in the validation set and the patients in the final quarter of the study period were included in the test set.

3.5. Patients in different age groups

The performance of the models in the different age groups is shown in Appendix B, Table B.5. Among the models using ECG only, the best model (MLP) improved by the addition of a previous ECG in patients up to 64 years, while the performance declined in those aged 65-75 years, suggesting that a previous ECG does not make a big difference. Models using additional clinical variables performed substantially better in younger patients, with e.g. a mean AUC of 0.92 with the MLP model in those 18-50 years and 0.82 in those over 76 years. This pattern was not present in the models using ECG only.

3.6. Patients with versus without previous AMI

The model performances stratified on previous AMI is shown in Appendix B, Table B.6. In patients with or without a previous AMI, adding a prior ECG generally did not improve the predictive ability of the models. The exception

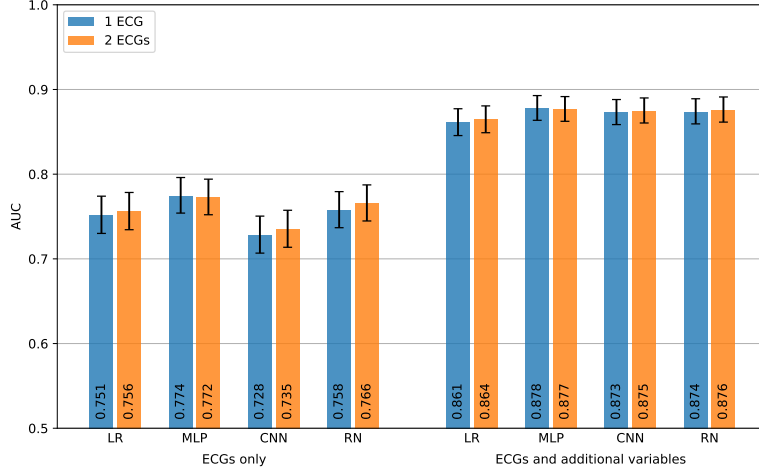


Figure 3: Main results

Barplot of the test-set AUC, including 95% confidence intervals approximated with bootstrapping. The blue bars correspond to models not using prior ECGs, and the orange bars correspond to models using both the index ECG and the prior ECG. The additional variables used by models on the right hand side of the figure are patient age, sex, troponin T and time since the previous ECG.

was the RN models using only ECG and the LR models with additional clinical variables, but both these models were outperformed by the corresponding MLPs. The AUC of all models, both with and without additional clinical variables, were substantially lower in patients with versus without a previous AMI.

3.7. MACE components as individual outcomes

As can be seen in Appendix B, Table B.7, there were no meaningful predictive improvements with the addition of a previous ECG when evaluating the models on UA, AMI and death separately. For models without additional clinical variables, UA was harder to predict (best AUC 0.71) than AMI (best AUC 0.80) and death within 30 days (AUC 0.83). The corresponding AUC for models including additional clinical variables were 0.77, 0.92 and 0.89, respectively.

4. Discussion

This study aimed to determine whether the addition of a previous ECG would improve the AUC of ML models predicting MACE within 30 days in ED chest pain patients. The purpose was thus not to identify the best predictive model for this outcome, but rather to give “equal opportunity” for models with different inputs, to see which performed the best. Our results consistently

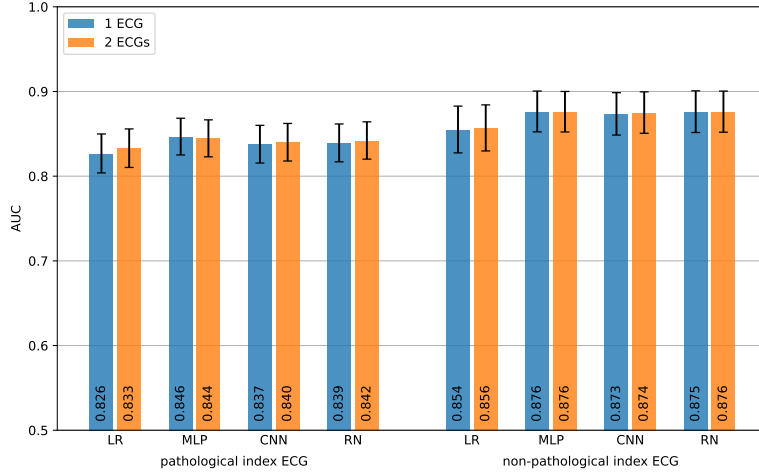


Figure 4: Results using additional clinical variables, stratified on pathological index ECG

Barplot of the test-set AUC of the predictions stratified on pathological index ECGs. The 95% confidence intervals were approximated with bootstrapping on each strata separately. The blue bars correspond to models not using prior ECGs, and the orange bars correspond to models using both the index ECG and the prior ECG. Note that this figure only shows the results for the models including additional clinical variables.

showed no improvements from prior ECGs, across several different models, additional input features, and patient subgroups, including patients of different ages, with or without prior AMI, and with pathological or non-pathological index ECGs.

These findings are somewhat counterintuitive because they run against current guidelines and best practices for manual ECG analysis, with the implication that it might be a waste of time to consider prior ECGs. While one should be careful to draw hasty conclusions, it might be sensible to further scrutinize the evidence in favor of evaluating prior ECGs, especially when it is based on data that is several decades old.

The lack of model improvement from prior ECGs is an indication that future research on ML for predicting MACE might be better focused on adding variables other than prior ECGs to the models. This might turn out to be a good thing, since not all patients have prior ECGs, and not all health care environments have access to digital ECG databases, so to the extent that an ML model can maintain predictive performance absent prior ECGs, such models will have a wider applicability.

Our models consistently performed better for several subgroups (though not with respect to prior ECGs): patients with non-pathological index ECGs, younger patients, and those with no history of AMI, were all “easier” to pre-

dict than their counterparts. These patients are typically healthier and have fewer comorbidities, so one reason for the increased model performance could be that whenever there is no sign of MACE, such patients are more likely to in fact be healthy, and conversely, when there are signs of MACE, it counts for more because the signal is less likely to be “obscured” by comorbidities or other explanations.

Interestingly, the best models in our comparisons were not the most advanced – the MLP outperformed both the CNN and RN models despite markedly fewer trainable parameters. We speculate that this is because the MLP model used the Glasgow features, which are expert features known to be useful for ECG analysis, while the CNN and RN models learned everything from the raw signal. In this sense, the MLP models were “pre-trained” with human knowledge and the CNN models were not. Given enough data and computing power, we expect that the CNN and RN models can perform at least similar to the MLP models.

The MLP models used in this study are similar to that employed by Ohlsson et al. [8] who combined Glasgow features from serial ECGs in a small neural network. When adding a previous ECG, the AUC improved by three percentage points (0.85 to 0.88). In our study, we had almost three times more training data but were still unable to achieve a similar AUC, and we found no clinically significant improvement from a previous ECG. A possible explanation is that the patient populations were different; Ohlsson’s data were collected in 1990-1997 and ours in 2017-18. Indeed, our AMI prevalence was 7.4% compared to over 20% in the Ohlsson study. In addition, the characteristics of the AMI patients and the ECG findings are likely different due to new AMI definitions and highly sensitive troponin assays.

Although we only tested four different ML models, our results suggest that ML models predicting MACE in ED chest pain patients do not improve significantly by the addition of a prior ECG, and that adding other information might be more beneficial. However, it should be noted that ED physicians may still make better decisions by comparing new and old ECGs. Humans do not assess the ECG in the same way as ML models, and ED physicians are often interested in additional outcomes other than MACE within 30 days.

4.1. Limitations

This study has some limitations. First, we used AUC as the metric for model performance whereas a fixed classification threshold would be used for decision-making in clinical care. We used AUC since the chosen threshold might differ in various clinical settings, and since the AUC reflects the overall performance at all thresholds. Although we cannot exclude it, we find it unlikely that the results would have been qualitatively different with a fixed classification threshold.

Second, as observed in section 3.1, there were differences between the training, validation and test sets, particularly with respect to disease history. These differences likely occurred because the ESC-TROP database only includes the first visit for each patient during the study period, and because we split the dataset chronologically rather than randomly. The test set (and to a lesser degree the validation set) was therefore biased toward patients who rarely visit the

ED with chest pain. The models’ performances may have been worse as a result, and it is possible that splitting the dataset randomly might have improved performance. However, since this applied similarly to models with and without a previous ECG, and because the proportion of MACE was similar across the splits, we find it reasonable to believe that the qualitative effects on the results were negligible. Furthermore, randomly splitting the data comes with its own set of drawbacks, which are avoided by a chronological split [12].

Third, our study does not include STEMI patients, so it is plausible, if not likely, that our results would differ for that patient group. However, STEMI patients, when detected, have a specific pathway through the health-care system, which largely bypasses the ED. In fact, if STEMI is detected, the patient will immediately be transferred to the cardiac care unit, so from the perspective of the ED physician, STEMI patients are “someone else’s problem”.

4.2. Conclusion

In this study, we found no improvement by adding a prior ECG to ML models predicting MACE within 30 days in ED chest pain patients. Our results were consistent across model types, patient groups, and additional model inputs. These findings suggest that ML models predicting MACE will perform equally well for patients without prior ECGs, and in the many health care settings lacking a digital ECG database.

Data availability

The data used in this study is considered sensitive in accordance with GDPR and can therefore not be made publicly available. However, all the code used in the analysis, as well as the final, trained models are available at <https://github.com/Tipulidae/mim>.

Declaration of competing interest

The are no conflicts of interest to declare.

References

- [1] A. Timmis, P. Vardas, N. Townsend, A. Torbica, H. Katus, D. De Smedt, C. P. Gale, A. P. Maggioni, S. E. Petersen, R. Huculeci, D. Kaskiewicz, V. de Benito Rubio, B. Ignatiuk, Z. Raisi-Estabragh, A. Pawlak, E. Karagiannis, R. Treskes, D. Gaita, J. F. Beltrame, A. McConnachie, I. Bardinet, I. Graham, M. Flather, P. Elliott, E. A. Mossialos, F. Weidinger, S. Achenbach, European society of cardiology: Cardiovascular disease statistics 2021, *European Heart Journal* 43 (2022) 716–799. doi:10.1093/eurheartj/ehab892.

- [2] A. Mokhtari, C. Borna, P. Gilje, P. Tydén, B. Lindahl, H.-J. Nilsson, A. Khoshnood, J. Björk, U. Ekelund, A 1-h Combination Algorithm Allows Fast Rule-Out and Rule-In of Major Adverse Cardiac Events, *Journal of the American College of Cardiology* 67 (2016) 1531–1540. doi:10.1016/j.jacc.2016.01.059.
- [3] J. L. Anderson, C. D. Adams, E. M. Antman, C. R. Bridges, R. M. Califf, D. E. Casey, W. E. Chavey, F. M. Fesmire, J. S. Hochman, T. N. Levin, A. M. Lincoff, E. D. Peterson, P. Theroux, N. K. Wenger, R. S. Wright, 2012 accf/aha focused update incorporated into the accf/aha 2007 guidelines for the management of patients with unstable angina/non–st-elevation myocardial infarction: A report of the american college of cardiology foundation/american heart association task force on practice guidelines, *Journal of the American College of Cardiology* 61 (2013) e179–e347. doi:<https://doi.org/10.1016/j.jacc.2013.01.014>.
- [4] T. H. Lee, E. F. Cook, M. C. Weisberg, G. W. Rouan, D. A. Brand, L. Goldman, Impact of the availability of a prior electrocardiogram on the triage of the patient with acute chest pain, *Journal of General Internal Medicine* 5 (1990) 381–388. doi:10.1007/BF02599421.
- [5] H. V. Pipberger, R. J. Arms, F. W. Stallmann, Automatic Screening of Normal and Abnormal Electrocardiograms by Means of a Digital Electronic Computer., *Experimental Biology and Medicine* 106 (1961) 130–132. doi:10.3181/00379727-106-26260.
- [6] S. Ansari, N. Farzaneh, M. Duda, K. Horan, H. B. Andersson, Z. D. Goldberger, B. K. Nallamothu, K. Najarian, A Review of Automated Methods for Detection of Myocardial Ischemia and Infarction Using Electrocardiogram and Electronic Health Records, *IEEE Reviews in Biomedical Engineering* 10 (2017) 264–298. doi:10.1109/RBME.2017.2757953.
- [7] X. Liu, H. Wang, Z. Li, L. Qin, Deep learning in ECG diagnosis: A review, *Knowledge-Based Systems* 227 (2021) 107187. doi:10.1016/j.knosys.2021.107187.
- [8] M. Ohlsson, H. Öhlin, S. M. Wallerstedt, L. Edenbrandt, Usefulness of serial electrocardiograms for diagnosis of acute myocardial infarction, *The American Journal of Cardiology* 88 (2001) 478–481. doi:10.1016/S0002-9149(01)01722-2.
- [9] A. Sbröllini, M. C. De Jongh, C. C. Ter Haar, R. W. Treskes, S. Man, L. Burattini, C. A. Swenne, Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: A deep-learning approach, *BioMedical Engineering OnLine* 18 (2019) 15. doi:10.1186/s12938-019-0630-9.
- [10] C. C. ter Haar, R. J. G. Peters, J. Bosch, A. Sbröllini, S. Gripenstedt, R. Adams, E. Bleijenberg, C. J. H. J. Kirchhof, R. Alizadeh Dehnavi, L. Burattini, R. J. de Winter, P. W. Macfarlane, P. G. Postema, S. Man,

- R. W. C. Scherptong, M. J. Schalij, A. C. Maan, C. A. Swenne, An initial exploration of subtraction electrocardiography to detect myocardial ischemia in the prehospital setting, *Annals of Noninvasive Electrocardiology* 25 (2020) e12722. doi:10.1111/anec.12722.
- [11] A. Mokhtari, A. Khoshnood, J. Lundager Forberg, C. Hård af Segerstad, U. Ekström, T. Schyman, M. Akbarzadeh, B. Lindahl, U. Ekelund, Effectiveness and Safety of the European Society of Cardiology 0-/1-h Troponin Rule-Out Protocol: The Design of the ESC-TROP Multicenter Implementation Study, *Cardiology* 145 (2020) 685–692. doi:10.1159/000509390.
- [12] E. Steyerberg, *Validation of Prediction Models*, Springer New York, New York, NY, 2009, pp. 299–311. doi:10.1007/978-0-387-77244-8_17.
- [13] P. Macfarlane, B. Devine, E. Clark, The university of glasgow (Uni-G) ECG analysis program, in: *Computers in Cardiology*, 2005, IEEE, Lyon, France, 2005, pp. 451–454. doi:10.1109/CIC.2005.1588134.
- [14] J. L. Forberg, M. Green, J. Björk, M. Ohlsson, L. Edenbrandt, H. Öhlin, U. Ekelund, In search of the best method to predict acute coronary syndrome using only the electrocardiogram from the emergency department, *Journal of Electrocardiology* 42 (2009) 58–63. doi:10.1016/j.jelectrocard.2008.07.010.
- [15] E. Giannitsis, K. Kurz, K. Hallermayer, J. Jarausch, A. S. Jaffe, H. A. Katus, Analytical Validation of a High-Sensitivity Cardiac Troponin T Assay, *Clinical Chemistry* 56 (2010) 254–261. doi:10.1373/clinchem.2009.132654.
- [16] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. S. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira, T. B. Schön, A. L. P. Ribeiro, Automatic diagnosis of the 12-lead ECG using a deep neural network, *Nature Communications* 11 (2020) 1760. doi:10.1038/s41467-020-15432-4.
- [17] B. Efron, Nonparametric standard errors and confidence intervals, *Canadian Journal of Statistics* 9 (1981) 139–158. doi:10.2307/3314608.
- [18] C. Strom Moller, B. Zethelius, J. Sundstrom, L. Lind, Persistent ischaemic ECG abnormalities on repeated ECG examination have important prognostic value for cardiovascular disease beyond established risk factors: A population-based study in middle-aged men with up to 32 years of follow-up, *Heart* 93 (2007) 1104–1110. doi:10.1136/hrt.2006.109116.
- [19] T. Sawai, H. Imano, I. Muraki, M. Hayama-Terada, Y. Shimizu, R. Cui, A. Kitamura, M. Kiyama, T. Okada, T. Ohira, K. Yamagishi, M. Umesawa, T. Sankai, H. Iso, Changes in ischaemic ECG abnormalities and subsequent risk of cardiovascular disease, *Heart Asia* 9 (2017) 36–43. doi:10.1136/heartasia-2016-010846.

Appendix A. Definition of MACE

	Description	30-day incidence (%)	Proportion of MACE (%)
Diagnoses (ICD-10)			
I200	Unstable angina	2.69	25.5
I21, I22	Acute myocardial infarction	6.14	58.1
I441, I442	Atrioventricular block	0.62	5.9
I46	Cardiac arrest	0.26	2.5
I470, I472, I490	Ventricular arrhythmia	0.33	3.2
J819	Pulmonary edema	0.19	1.8
R570	Cardiogenic shock	0.02	0.1
Interventions (KVA)			
DF005	Placement of aortic balloon pump	0.02	0.1
DF017, DF028	Cardiopulmonary resuscitation	0.04	0.3
FNA00, FNA10, FNC10, FNC20, FNC30, FNF96	Coronary artery graft	0.89	8.4
FNG00, FNG02, FNG05	Percutaneous coronary intervention	4.28	40.4
FPE00, FPE10, FPE20, FPE26	Transvenous pacemaker insertion	0.49	4.7
TFP00	Temporary cardiac pacing	0.01	0.0
Other			
Death	All causes	1.25	11.8

Table A.2: Definition of MACE

MACE was defined as any of the events occurring within 30 days of the ED visit. Multiple events can occur simultaneously, which is why the sum of percentages across the categories adds up to more than 100%.

Appendix B. Additional results

Model	AUC (95% Confidence interval)			
	Without additional clinical variables		With additional clinical variables	
	Index visit ECG only	Prior ECG added	Index visit ECG only	Prior ECG added
LR	0.751 (0.730 - 0.774)	0.756 (0.734 - 0.778)	0.861 (0.845 - 0.878)	0.864 (0.849 - 0.881)
MLP	0.774 (0.754 - 0.796)	0.772 (0.752 - 0.794)	0.878 (0.864 - 0.893)	0.877 (0.862 - 0.892)
CNN	0.728 (0.707 - 0.752)	0.735 (0.714 - 0.757)	0.873 (0.858 - 0.888)	0.875 (0.860 - 0.890)
RN	0.758 (0.737 - 0.779)	0.766 (0.745 - 0.787)	0.874 (0.859 - 0.889)	0.876 (0.861 - 0.891)

Table B.3: Main results

AUC of all the models on the test set, with 95% confidence interval approximated with bootstrapping.

		AUC (95% Confidence interval)			
		Without additional clinical variables		With additional clinical variables	
	Model	Index visit ECG only	Prior ECG added	Index visit ECG only	Prior ECG added
Non-pathological index visit ECG	LR	0.703 (0.664 - 0.744)	0.694 (0.654 - 0.735)	0.854 (0.828 - 0.883)	0.856 (0.830 - 0.884)
	MLP	0.733 (0.696 - 0.770)	0.731 (0.694 - 0.768)	0.876 (0.852 - 0.901)	0.876 (0.852 - 0.900)
	CNN	0.702 (0.664 - 0.742)	0.697 (0.660 - 0.736)	0.873 (0.849 - 0.899)	0.874 (0.851 - 0.900)
	RN	0.721 (0.684 - 0.760)	0.730 (0.693 - 0.770)	0.875 (0.852 - 0.901)	0.876 (0.852 - 0.900)
Pathological index visit ECG	LR	0.704 (0.675 - 0.735)	0.720 (0.692 - 0.749)	0.826 (0.804 - 0.850)	0.833 (0.810 - 0.856)
	MLP	0.730 (0.702 - 0.759)	0.732 (0.703 - 0.761)	0.846 (0.825 - 0.868)	0.844 (0.823 - 0.867)
	CNN	0.663 (0.632 - 0.694)	0.686 (0.657 - 0.717)	0.837 (0.815 - 0.860)	0.840 (0.818 - 0.862)
	RN	0.703 (0.674 - 0.732)	0.714 (0.686 - 0.743)	0.839 (0.817 - 0.862)	0.842 (0.820 - 0.864)

Table B.4: Results stratified on pathological index ECGs

AUC of all the models on the test set, stratified on pathological index ECGs, with 95% confidence interval approximated with bootstrapping. The models were evaluated separately on patients with and without pathological index ECGs, as determined by the Glasgow algorithm.

		AUC (95% Confidence interval)			
		Without additional clinical variables		With additional clinical variables	
	Model	Index visit ECG only	Prior ECG added	Index visit ECG only	Prior ECG added
18-50 years	LR	0.700 (0.582 - 0.828)	0.753 (0.658 - 0.859)	0.919 (0.883 - 0.962)	0.924 (0.890 - 0.963)
	MLP	0.764 (0.666 - 0.876)	0.814 (0.738 - 0.899)	0.924 (0.886 - 0.969)	0.933 (0.903 - 0.969)
	CNN	0.758 (0.664 - 0.864)	0.802 (0.724 - 0.895)	0.936 (0.907 - 0.972)	0.939 (0.910 - 0.975)
	RN	0.825 (0.739 - 0.925)	0.824 (0.742 - 0.918)	0.937 (0.908 - 0.972)	0.942 (0.916 - 0.974)
51-64 years	LR	0.667 (0.612 - 0.723)	0.707 (0.655 - 0.759)	0.852 (0.815 - 0.892)	0.858 (0.822 - 0.897)
	MLP	0.706 (0.653 - 0.758)	0.733 (0.684 - 0.782)	0.867 (0.834 - 0.904)	0.866 (0.833 - 0.902)
	CNN	0.672 (0.620 - 0.726)	0.676 (0.622 - 0.730)	0.860 (0.826 - 0.897)	0.861 (0.827 - 0.898)
	RN	0.695 (0.641 - 0.749)	0.700 (0.645 - 0.755)	0.863 (0.828 - 0.900)	0.863 (0.829 - 0.900)
65-75 years	LR	0.686 (0.643 - 0.731)	0.681 (0.638 - 0.728)	0.831 (0.800 - 0.864)	0.828 (0.797 - 0.861)
	MLP	0.710 (0.669 - 0.752)	0.682 (0.639 - 0.725)	0.846 (0.817 - 0.877)	0.843 (0.813 - 0.874)
	CNN	0.650 (0.606 - 0.695)	0.653 (0.608 - 0.698)	0.844 (0.815 - 0.875)	0.846 (0.816 - 0.876)
	RN	0.678 (0.636 - 0.724)	0.697 (0.656 - 0.740)	0.847 (0.818 - 0.879)	0.851 (0.822 - 0.881)
Over 76 years	LR	0.709 (0.670 - 0.749)	0.700 (0.659 - 0.741)	0.808 (0.775 - 0.843)	0.816 (0.783 - 0.850)
	MLP	0.726 (0.689 - 0.764)	0.727 (0.687 - 0.768)	0.823 (0.791 - 0.857)	0.818 (0.784 - 0.853)
	CNN	0.661 (0.620 - 0.702)	0.669 (0.627 - 0.710)	0.807 (0.772 - 0.843)	0.810 (0.775 - 0.846)
	RN	0.680 (0.640 - 0.720)	0.692 (0.653 - 0.732)	0.810 (0.776 - 0.845)	0.813 (0.779 - 0.849)

Table B.5: Results stratified on age quartiles

AUC of all the models on the test set, stratified on age quartiles, with 95% confidence interval approximated with bootstrapping. The models were evaluated separately on patients in different age groups.

		AUC (95% Confidence interval)			
		Without additional clinical variables		With additional clinical variables	
	Model	Index visit ECG only	Prior ECG added	Index visit ECG only	Prior ECG added
No previous AMI	LR	0.749 (0.724 - 0.774)	0.759 (0.734 - 0.784)	0.868 (0.852 - 0.886)	0.871 (0.855 - 0.889)
	MLP	0.776 (0.753 - 0.799)	0.777 (0.754 - 0.801)	0.886 (0.871 - 0.902)	0.885 (0.870 - 0.902)
	CNN	0.724 (0.700 - 0.750)	0.732 (0.708 - 0.757)	0.882 (0.866 - 0.899)	0.883 (0.868 - 0.900)
	RN	0.760 (0.737 - 0.785)	0.768 (0.744 - 0.792)	0.883 (0.868 - 0.900)	0.884 (0.869 - 0.901)
Previous AMI	LR	0.672 (0.617 - 0.727)	0.661 (0.604 - 0.720)	0.752 (0.702 - 0.804)	0.765 (0.716 - 0.816)
	MLP	0.676 (0.621 - 0.731)	0.667 (0.611 - 0.724)	0.771 (0.724 - 0.821)	0.767 (0.718 - 0.817)
	CNN	0.658 (0.601 - 0.716)	0.660 (0.604 - 0.717)	0.755 (0.705 - 0.805)	0.762 (0.713 - 0.813)
	RN	0.645 (0.587 - 0.702)	0.660 (0.603 - 0.715)	0.757 (0.708 - 0.809)	0.766 (0.718 - 0.817)

Table B.6: Results stratified on previous AMI

AUC of all the models on the test set, stratified on previous AMI, with 95% confidence interval approximated with bootstrapping. The models were evaluated separately on patients with and without a previous AMI in the past five years.

		AUC (95% Confidence interval)			
		Without additional clinical variables		With additional clinical variables	
	Model	Index visit ECG only	Prior ECG added	Index visit ECG only	Prior ECG added
MACE	LR	0.751 (0.730 - 0.774)	0.756 (0.734 - 0.778)	0.861 (0.845 - 0.878)	0.864 (0.849 - 0.881)
	MLP	0.774 (0.754 - 0.796)	0.772 (0.752 - 0.794)	0.878 (0.864 - 0.893)	0.877 (0.862 - 0.892)
	CNN	0.728 (0.707 - 0.752)	0.735 (0.714 - 0.757)	0.873 (0.858 - 0.888)	0.875 (0.860 - 0.890)
	RN	0.758 (0.737 - 0.779)	0.766 (0.745 - 0.787)	0.874 (0.859 - 0.889)	0.876 (0.861 - 0.891)
Unstable Angina	LR	0.668 (0.624 - 0.715)	0.677 (0.633 - 0.721)	0.730 (0.694 - 0.767)	0.734 (0.699 - 0.770)
	MLP	0.701 (0.659 - 0.745)	0.712 (0.670 - 0.754)	0.765 (0.733 - 0.798)	0.763 (0.730 - 0.796)
	CNN	0.660 (0.615 - 0.704)	0.695 (0.652 - 0.739)	0.761 (0.728 - 0.795)	0.768 (0.735 - 0.802)
	RN	0.686 (0.642 - 0.731)	0.690 (0.647 - 0.736)	0.765 (0.732 - 0.798)	0.765 (0.733 - 0.798)
Acute Myocardial Infarction	LR	0.778 (0.752 - 0.804)	0.788 (0.764 - 0.814)	0.911 (0.897 - 0.925)	0.914 (0.901 - 0.929)
	MLP	0.800 (0.776 - 0.826)	0.804 (0.780 - 0.828)	0.922 (0.909 - 0.936)	0.920 (0.906 - 0.934)
	CNN	0.738 (0.711 - 0.765)	0.748 (0.722 - 0.775)	0.916 (0.902 - 0.931)	0.917 (0.903 - 0.932)
	RN	0.768 (0.742 - 0.794)	0.777 (0.752 - 0.804)	0.917 (0.903 - 0.931)	0.920 (0.907 - 0.935)
Death	LR	0.828 (0.783 - 0.878)	0.833 (0.792 - 0.877)	0.885 (0.853 - 0.922)	0.885 (0.855 - 0.922)
	MLP	0.821 (0.779 - 0.866)	0.795 (0.741 - 0.853)	0.873 (0.843 - 0.908)	0.870 (0.840 - 0.904)
	CNN	0.821 (0.777 - 0.869)	0.805 (0.757 - 0.856)	0.868 (0.838 - 0.902)	0.865 (0.834 - 0.901)
	RN	0.830 (0.785 - 0.881)	0.827 (0.780 - 0.880)	0.870 (0.841 - 0.904)	0.871 (0.841 - 0.907)

Table B.7: Results evaluated on different outcomes

AUC of all the models on the test set, evaluated on different outcomes, with 95% confidence interval approximated with bootstrapping. The models were evaluated against subsets of the MACE outcome, although they were optimized to predict MACE.