Adversarial Multitask Learning Reduces the Correlation Between Age and Deep Learning Predictions of Myocardial Infarction from Electrocardiograms

Silvia Ibrahimi[1], Massimo W Rivolta[1], Roberto Sassi[1]

[1] Dipartimento di Informatica, Università degli Studi di Milano, Milan, Italy

Abstract

Deep Learning (DL) models have exhibited high performance in automated ECG diagnosis, yet some were developed on small datasets or biased towards a certain clinical condition. This study applies adversarial multitask learning (AML), a technique that trains a DL model by optimizing contrastive objectives, to identify myocardial infarction (MI) from ECG signals while mitigating the influence of age on model predictions. ECG recordings from healthy and MI subjects were extracted from the PTB-XL dataset and preprocessed to generate 12-lead average beats. Two DL models sharing the initial layers were trained. The first model was trained to identify MI, while the second to predict patient's age, with the parameters of the common layers frozen. Finally, the parameters of the common layers were trained to minimize the classification loss while maximizing the age prediction error, using two loss functions: i) mean squared error (MSE); and ii) negative squared covariance (NCOV). On the validation set, the first model achieved a classification accuracy of 0.87 while the second one had a Pearson's correlation coefficient (PCC) with age of 0.67. After adversarial training with MSE and NCOV, PCCs with age were -0.78 and -0.03, and accuracies were 0.82 and 0.85, respectively. The proposed AML reduced the correlation between true and predicted age while keeping a good performance for MI identification.

1. Introduction

In the last few years, many Deep Learning (DL) models have been implemented in the cardiovascular domain for the automated ECG diagnosis, reaching very high performances. Despite that, some of them were developed on small datasets or biased towards a certain clinical condition. DL models trained on such data often lack reliability and generalization to new data. There are very few studies that addressed this problem. In their work, Alday *et al.* [1] analyzed the effect of age, sex and race on ECG-based arrhythmia detection algorithms, highlighting the importance of fairness when implementing ECG classification models.

In the context of automatic myocardial infarction (MI) identification, various DL models have shown outstanding capabilities on identifying the disease [2]. However, in our previous study [3], we highlighted a critical aspect. By employing an explainable AI technique, we empirically proved that, despite achieving high test set performances, the model was relying on features that did not align with those recommended by the clinical guidelines for MI classification [4].

Adversarial multitask learning (AML) represents one approach to address a known bias in a DL algorithm. The method trains a DL model through two adversarial tasks. The first one optimizes the DL model towards the expected output (for instance, classification), while the second one minimizes the dependency of the model from the known bias. This framework has successfully been applied to speech recognition to extract age-independent features [5] and a speaker- and age-invariant features [6]. In another work [7], they adopted AML in the context of ECG classification to improve the generalizability of arrhythmia classification, trying to alleviate subject dependency.

Inspired by those works, we employed AML to identify MI from ECG signals trying to mitigate the impact of age on model predictions. The objective of our work is to enhance the reliability of MI classification models by proposing an alternative training technique. We selected to compensate for age since the possibility to estimate patient's age from ECG signals is well known [8].

2. Methods

2.1. Dataset

ECG signals were obtained from the freely available PTB-XL dataset [9]. It consists of 21799 clinical 12-lead ECG signals from 18885 patients (age range: 0 to 95 years). In this work, we considered the downsampled version (100 Hz) of the original signals. For the scope of our study, we considered only healthy control (HC) and MI patients. In particular, among the ECG signals, we excluded those which were also associated to the following abnor-

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.451

malities: bundle branch block, abnormal QRS, high or low QRS voltage, supraventricular tachycardia, sinus tachycardia, paroxysmal supraventricular tachycardia and left ventricular hypertrophy. We also excluded ECGs of patients whose age was not reliable (*i.e.*, set to 300). After ECG preprocessing (see sec. 2.2), we obtained 7735 HC subjects and 1695 MI patients.

2.2. Preprocessing

ECG signals were filtered with a Butterworth filter (3rd order, 0.5-40 Hz, zero-phase) to reduce baseline wandering, high-frequency noise and powerline interference. Beats were detected on the vector magnitude (i.e., square root of the sum of squares of the 12 leads) using the *gqrs* algorithm [10] and aligned using the Woody algorithm [11]. Signal quality was performed on each lead. Specifically, the mean Pearson's correlation coefficient was computed between each QRS (Q-20 ms; Q+100 ms) complex and an average QRS template. ECG signals presenting an average cross-correlation higher than 0.9 were considered of good quality. For the experiments, patients were further included if this condition was matched in at least 8 leads. We computed the average beat for each of the 12 leads, considering only beats whose inter beat-time interval did not vary more than 50 ms with respect to the median QQ value. We considered average beats lasting 580 ms (Q-250 ms; Q+330 ms).

2.3. Adversarial multitask learning

Adversarial multitask learning refers to techniques where a DL model is trained by minimizing certain losses, while maximizing others. These different losses are called primary and secondary tasks respectively, and they compete with each other as adversaries in a game. The training of the model is typically achieved through the gradient reversal layer (GRL) approach proposed by Ganin *et al.* [12]. Specifically, layers between the input and a certain hidden layer are trained by reversing the gradient with respect to secondary tasks, so that a worsening would occur on these tasks. Here, we defined MI identification as primary task and age prediction from ECGs as secondary one.

To implement AML, we followed the approach reported in [7]. Specifically, the DL model for the primary task, meant to detect MI vs HC, was defined as the composition of two functions $p = m_{\theta_M}(b_{\theta_B}(x))$, where x is the 12-lead average beat and p is the probability of detecting MI from the input ECG, m_{θ_M} and b_{θ_B} are neural networks parameterized by their respective parameter vectors θ_M and θ_B. The model for the secondary task was instead defined as $y = a_{\theta_A}(b_{\theta_B}(x))$, where y is the predicted age and a_{θ_A} is a neural network with parameters θ_A. It is worth noting that the two models shared the same layers defined in the

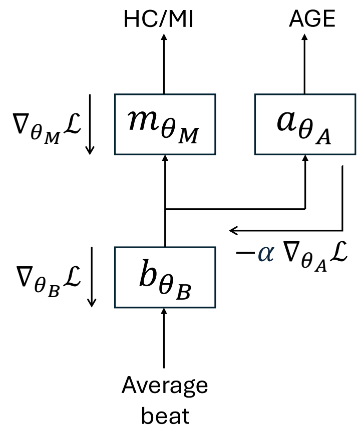


Figure 1: Diagram of AML.

function b_{θ_B}.

The training of the models worked as follows. First, the parameter vectors θ_B and θ_M were determined by minimizing the binary cross-entropy (BCE) loss for MI vs HC classification via stochastic gradient descent by the following update formulas

$$\begin{aligned} \theta_M &\leftarrow \theta_M - \eta \frac{\partial \mathcal{L}_{\mathrm{BCE}}}{\partial \theta_M} \\ \theta_B &\leftarrow \theta_B - \eta \frac{\partial \mathcal{L}_{\mathrm{BCE}}}{\partial \theta_B} \end{aligned} \tag{1}$$

where η is the learning rate.

Once the model was trained, the parameter vector θ_A was updated by minimizing the mean squared error (MSE) loss between the true and predicted age from the network a_{θ_A}, as follows

$$\theta_A \leftarrow \theta_A - \eta \frac{\partial \mathcal{L}_{\mathrm{MSE}}}{\partial \theta_A} \tag{2}$$

while keeping the parameter vector θ_B fixed. This allowed to create a model able to predict patient's age using the same b_{θ_B} model for MI identification.

Finally, the adversarial training to remove age prediction from the model b_{θ_B} was performed using GRL. Specifically, using the following update formula

$$\theta_B \leftarrow \theta_B - \eta \left(\frac{\partial \mathcal{L}_{\mathrm{BCE}}}{\partial \theta_B} - \alpha \frac{\partial \mathcal{L}_{\mathrm{MSE}}}{\partial \theta_B} \right) \tag{3}$$

where α controls the effect of the reverse gradient.

Figure 1 shows a diagram of the models described.

2.4. DL model

The model b_{θ_B} included two 1D convolutional blocks with ReLU activation, batch normalization and max pooling. One fully connected layer was then concatenated. The classification head m_{θ_M} for MI identification consisted of 2 fully connected layers, with the output layer having a softmax function applied. The network a_{θ_A} for the secondary task, i.e. age estimation, comprised one fully connected layer. The input to the network was a 12x58 matrix representing the samples of the average beat for each of the 12 leads.

2.5. Compensation of age-related ECG changes

In the original AML framework, the same loss function was applied to both the secondary and adversarial tasks. In our study, the secondary task was to predict the age of the patients. To do so, we trained the model using the MSE loss function. However, maximizing it during the adversarial tasks does not necessarily discourage the model from relying on age-related features. Indeed, a perfect negative correlation between the true and predicted age worsen significantly the MSE, but keeping the network relying on such feature. To mitigate this, we proposed to modify the loss function only during the adversarial task, replacing the MSE loss in (3) with the negative squared covariance between the true and predicted age.

2.6. Experiments

The dataset was partitioned into training (80%) and validation sets (20%), ensuring that different recordings from the same patient were included in only one of the two sets. Since the classes MI/HC were highly imbalanced, the cross-entropy loss was adjusted to weight the two classes according to their number of ECGs. In our experiments, we choose a batch size of 32. Different number of epochs and learning rates during the three training stages were considered. Firstly, we trained a "baseline model" consisting of the shared backbone b_{θ_B} and the MI classification network m_{θ_M}. Epochs and learning rate were set to 100 and 0.0001, respectively. θ_M was updated according (1). In the second training phase, we trained the "age model". This model was composed by the pretrained network b_{θ_B} which was kept frozen and paired with a_{θ_A}. The learning rate was increased to 0.001 and the number of epochs was set to 200. In the final training stage, we considered AML. The pretrained network b_{θ_B} was updated using GRL according to (3), while keeping m_{θ_M} and a_{θ_A} frozen. We considered 100 epochs and $\alpha = 0.001$. To evaluate the efficacy of b_{θ_B} in generating features not correlated with age, we predicted the age considering m_{θ_M} and a_{θ_A} models together after AML. For this last stage, in addition to the MSE loss, we used the negative squared covariance loss as discussed before. We compared the original implementation of AML with the one proposed here and quantified the performance using accuracy for classification of MI and Pearson's correlation coefficient between the predicted and the true age of the patient.

3. Results

For the MI classification task, the baseline model achieved a validation accuracy of 0.87, a sensitivity of 0.88 and a specificity of 0.87. The age estimation task was assessed using Pearson's correlation coefficient. The age model produced a correlation of 0.67. When AML was implemented, this correlation increased in magnitude to -0.78 when using the MSE loss, but decreased to -0.03 with the negative squared covariance loss. The corresponding accuracies were 0.82 and 0.85, respectively. Figure 2 provides further insights, showing the correlation between true and predicted ages in the validation set for both models.

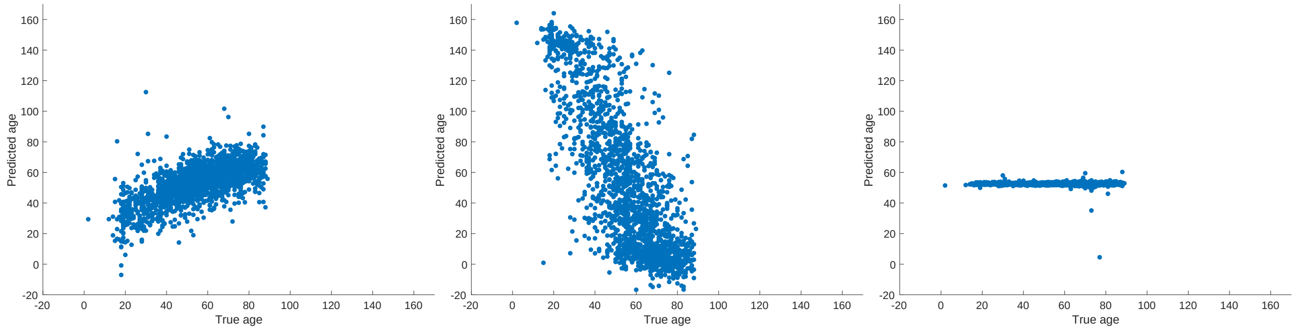
4. Discussion and conclusion

In this study, we employed the adversarial multitask learning approach to address the problem of age bias in MI identification from ECGs signals using AI models. Specifically, two neural networks, performing MI classification (primary task) and age estimation (secondary task), shared few input layers. On the one hand, the parameters of the common layers were trained by minimizing the classification loss, while on the other maximizing the age prediction error by implementing a GRL layer.

We assessed the effectiveness of the approach by testing whether the model could be unable to predict the patient's age, while correctly identifying the presence of MI. The results presented in Figure 2, along with Pearson's correlation coefficients, show that the adopted training strategy and the implemented negative squared covariance loss function were effective in reducing the correlation between predicted and actual ages. Unlike the approach used in [7], where the same loss function was applied to both the secondary task and the adversarial task, we specifically used a distinct loss for the adversarial task to decouple the learning of age-related features from the primary task. In fact, the near-zero Pearson's correlation coefficient suggests that the model was not capable of predicting age. However, in the case of adversarial training with the MSE loss function, the model exhibited a negative high correlation between predicted and true age, implying that it may have learned to encode age-related information in a manner that still influences its outputs, though inversely. This residual bias highlights the complexity of fully eliminating age-related features and the need for further investigation.

Regarding the MI classification task, the results showed no significant reduction in accuracy between the baseline model and the adversarial model.

As future works, it might be worth exploring the impact of other demographic confounding factors, such as gen-



(a) (b) (c)

Figure 2: Correlation between the real and the predicted age in the validation set for age model (a), and the AML framework (b, c). In b), the MSE loss was adopted during the training procedure, while in c) the negative squared covariance loss.

der and race, on MI identification, since they have already found relevant for arrhythmia detection [1]. The potential of AML for MI identification in mitigating these possible biases still needs investigation.

Acknowledgments

SI acknowledges support for her PhD fellowship by Novartis.

Part of this research was supported by the MUSA – Multilayered Urban Sustainability Action – project, funded by the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strengthening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D".

Finally, the authors acknowledge fruitful technical discussions with Kevin Drago during the preliminary experiments for this study.

References

[1] Alday EAP, Rad AB, Reyna MA, Sadr N, Gu A, et al. Age, sex and race bias in automated arrhythmia detectors. Journal of Electrocardiology 2022;74:5–9.

[2] Xiong P, Lee SMY, Chan G. Deep learning for detecting and locating myocardial infarction by electrocardiogram: A literature review. Frontiers in cardiovascular medicine 2022;9:860032.

[3] Bodini M, Rivolta MW, Sassi R. Opening the black box: interpretability of machine learning algorithms in electrocardiography. Philosophical Transactions of the Royal Society A Mathematical Physical and Engineering Sciences 2021; 379:20200253.

[4] Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, et al. Fourth universal definition of myocardial infarction (2018). J Am Coll Cardiol 2018;72(18):2231–64.

[5] Rumberg L, Ehlert H, Lüdtke U, Ostermann J. Age-invariant training for end-to-end child speech recognition using adversarial multi-task learning. In Proc. Interspeech 2021. 2021; 3850–3854.

[6] Shahin MA, Ahmed B, Epps J. Speaker- and age-invariant training for child acoustic modeling using adversarial multi-task learning. ArXiv 2022;abs/2210.10231.

[7] Shahin M, Oo E, Ahmed B. Adversarial multi-task learning for robust end-to-end ECG-based heartbeat classification. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020; 341–344.

[8] Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, et.al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. Circulation Arrhythmia and Electrophysiology 2019;12(9):e007284.

[9] Wagner P, Strodthoff N, Bousseljot R, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.1), Physionet (2020).

[10] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101:e215–e220.

[11] Woody CD. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical Biological Engineering Computing 1967;5(6):539–554.

[12] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks, 2016. URL `https://arxiv.org/abs/1505.07818`.

Address for correspondence:

Silvia Ibrahimi
Dipartimento di Informatica, Università degli Studi di Milano,
Via Celoria 18, Milan 20133, Italy
silvia.ibrahimi@unimi.it