Artificial Intelligence in Pediatric Electrocardiogram Analysis: Sex and Age Estimation Across Puberty

Md Moklesur Rahman[1], Sara Battiston[1], Massimo Walter Rivolta[1], Antonio Sanzo[2], Irene Raso[3], Sara Santacesaria[3], Gianvincenzo Zuccotti[4,5], Savina Mannarino[3], Roberto Sassi[1]

[1] Computer Science Department, University of Milan, Milan, Italy
[2] Fondazione IRCCS Policlinico San Matteo, Pavia, Italy
[3] Pediatric Cardiology Unit, Buzzi Children's Hospital, Milan, Italy
[4] Pediatric Department, Buzzi Children's Hospital, Milan, Italy
[5] Department of Biomedical and Clinical Science, University of Milan, Milan, Italy

Abstract

Electrocardiograms (ECGs) are essential for evaluating electrical and structural heart problems, but pediatric ECG (pECG) interpretation remains a challenging area due to the dynamic physiological changes occurring throughout infancy and adolescence. Accurate interpretation of pECG is crucial for the diagnosis and management of various cardiac conditions in children, yet age and sex-related variations in ECG patterns complicate this task. Different from previous studies, which have typically focused on either age or sex predictions, this study aims to develop an artificial intelligence-based system that simultaneously predicts both age and sex from 12-lead pECGs. We employed a multitask deep learning model (DLM) trained on a curated dataset of 54,230 pediatric 12-lead ECG recordings collected at the Buzzi Children's Hospital in Milan, Italy, from 2011 to 2020. The DLM achieved a mean absolute error of 0.532 years for age prediction and an R^2 score of 0.932, indicating high accuracy in age prediction. For sex prediction, the model attained an accuracy of 0.712 on the test set. Overall, these results are consistent with prior studies and highlight the feasibility and novelty of applying multitask DLM to the pECG analysis.

1. Introduction

Cardiovascular disease is a leading cause of global mortality, responsible for around 17.9 million deaths annually [1]. Electrocardiograms (ECGs) are a common, non-invasive tool for assessing cardiac electrical function, and artificial intelligence (AI) has been increasingly applied to ECG analysis, particularly for arrhythmia detection and diagnosis [2, 3]. One notable application is the estimation of biological age from ECG signals. For example, Lima *et al.* [4] showed that AI-predicted age is a strong independent predictor of all-cause mortality, while Hirota *et al.* [5] found a significant link between AI-estimated ECG age and cardiovascular events, with larger age deviations correlating with higher myocardial infarction risk in individuals under 60 years.

While there has been considerable progress in AI applications to adult ECG datasets, pediatric applications remain comparatively underexplored, largely due to the limited availability of large, labeled pediatric ECG (pECG) datasets. However, recent studies have begun to address this gap, reflecting growing interest in the field [6, 7]. Given the unique physiological differences between pECGs and adult ECGs, pECG offers valuable insights into cardiac function across developmental stages, aiding in the detection of conditions such as congenital heart disease [8], arrhythmias [9], ventricular dysfunction [10], and predicting mortality in congenital heart defect patients [11]. Within this context, predicting age and sex from pECG adds clinical value by improving patient assessment, guiding personalized care, and aiding early detection of developmental or pathological abnormalities [12]. Age-related changes in ECG parameters, such as heart rate, wave morphology, and interval durations, provide insights into pediatric physiological development. Additionally, sex-based differences in ECG features, driven by hormonal and genetic factors, can further aid clinical interpretation [13].

Motivated by these clinical insights, recent studies have leveraged AI, particularly deep learning models (DLMs), to predict age and sex from pECG. For instance, Junmo An *et al.* [6] used a residual DLM to classify pediatric and adult groups, achieving an F1-score of 0.78 with 10 s standard 12-lead pECG. Similarly, Dutenhefner *et al.* [7] trained a residual DLM on 163,242 pECGs from 148,738 pediatric patients aged 0 to 18 years, reporting a mean ab-

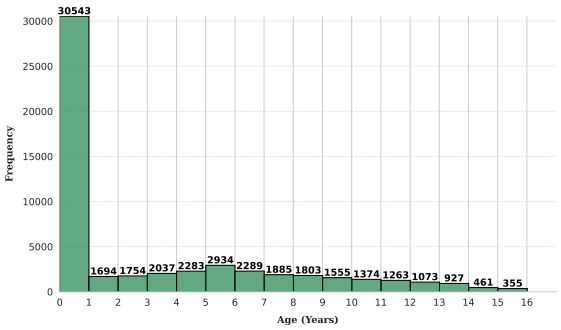

Figure 1: Age distribution in the pediatric dataset.

solute error (MAE) of 2.12 years in age prediction, despite a significant age group imbalance.

In parallel, sex prediction from pECG has also shown promising results. O'Sullivan *et al.* [13] analyzed ECG data from 90,133 children (mean age: 12 years; 46.7% male), stratified into prepubertal (0–7 years), peripubertal (8–14 years), and postpubertal (15–18 years) groups. Performance was highest in the postpubertal (area under the curve or AUC = 0.980), moderate in the peripubertal group (AUC = 0.880), and lower in the prepubertal group (AUC = 0.650).

To the best of our knowledge, no prior studies have simultaneously addressed pediatric age and sex prediction from pECG. The main contributions of this study are twofold: i) the development of a multitask DLM for age and sex prediction from standard pECG, and ii) the analysis of its performance with reduced lead setting. Unlike previous studies that developed solely on 12-lead ECGs, we explore model performance under reduced-lead settings, which can enhance the clinical applicability and portability of the proposed method.

2. Materials and methods

2.1. Dataset and preprocessing

ECG recordings for this study were collected at Buzzi Children's Hospital in Milan, Italy, from 2011 to 2020 using GE Healthcare's MAC 5500 HD and MAC 2000 systems, with a 12-lead setting and 500 Hz sampling rate. The dataset includes 54,230 pediatric ECG recordings (ages 0–17 years), with a mean age of 3.16 years (SD: 4.17), and 46.5% females. The age distribution is shown in Figure 1.

Subjects were stratified into six pediatric groups based on age: neonates (0–6 months), infants (6 months–1 year), toddlers (1–3 years), preschoolers (3–6 years), children (6–12 years), and adolescents (12–17 years).

As for preprocessing, a 3^{rd}-order Butterworth filter with a cut-off frequency of 0.5–40 Hz was used on the pECG to remove baseline wander and broadband noise. Ages (in years, months, and days) were converted into continuous values in years for regression. For developing the DLM, subjects were split into 80% for train, 10% for validation, and 10% for test. The distribution of pediatric age groups across train, validation, and test sets is reported in Table 1.

Table 1: Distribution of pediatric age groups across train, validation, and test datasets.

Age Group	Train	Validation	Test
Neonates	22858	2858	2876
Infants	1517	196	188
Toddlers	2796	339	338
Preschoolers	5815	678	741
Children	8140	1057	995
Adolescents	2258	295	285
Total	43384	5423	5423

2.2. Deep learning model and experiments

We designed a multitask DLM, which is more efficient than training separate models as it shared learned features across tasks (age and sex prediction), reducing the need for separate resources for each task. We used a residual temporal attention (RTA)-based model, which is a modified version of an architecture previously proposed for arrhythmia detection [3].

To evaluate the model's adaptability to different lead settings, we developed different models for 12 leads, 8 leads, by excluding lead III and the three augmented limb leads, and a reduced 6-lead setting, composed solely of the precordial leads (V1–V6). A combination of binary cross-entropy loss for sex classification, and mean square error for age prediction was used. Finally, the following hyperparameters were used during training: i) the Adam optimizer with a learning rate of 0.001, ii) a batch size of 32, iii) a total of 100 epochs, and iv) early stopping with a patience of 6, which was employed to terminate training when no improvement was observed.

3. Results and discussions

The models were evaluated using various metrics, including MAE, R^2, sensitivity (Se), specificity (Sp), positive predictive value (PPV), and F1-score. For age prediction, the MAE and R^2 values for the 12-lead were 0.937 and 0.538, respectively. The scatter plot is presented in Figure 3. For the 8-lead, the MAE and R^2 values were 0.936 and 0.548, while for the 6-lead, they were 0.934 and 0.544, respectively. Overall, the differences in age prediction performance across lead settings were minimal, possibly due to the imbalanced dataset. We further assessed model performance across different pediatric age groups to obtain more insights. The classification results were summarized in Table 2.

The 12-lead setting demonstrated high performance

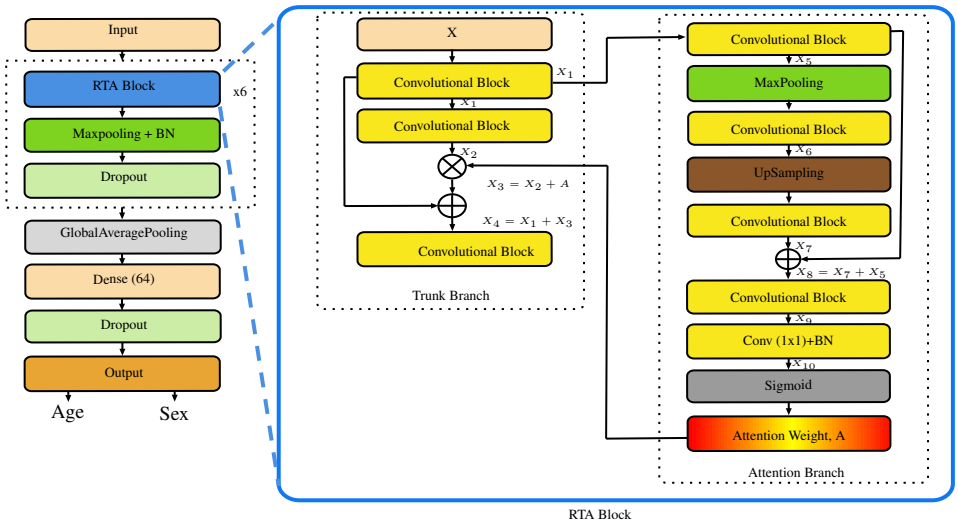

Figure 2: Architecture of the deep learning model for age and sex prediction from pECG.

Table 2: Performance for age group classification across different lead settings.

Lead	Age Group	Se	Sp	PPV	F1
12	Neonates	0.983	0.969	0.972	0.978
	Infants	0.551	0.983	0.534	0.542
	Toddlers	0.598	0.982	0.689	0.640
	Preschoolers	0.755	0.962	0.760	0.758
	Children	0.828	0.940	0.757	0.791
	Adolescents	0.535	0.990	0.745	0.623
8	Neonates	0.974	0.971	0.973	0.973
	Infants	0.503	0.982	0.505	0.504
	Toddlers	0.604	0.979	0.660	0.631
	Preschoolers	0.755	0.955	0.735	0.745
	Children	0.824	0.936	0.749	0.785
	Adolescents	0.507	0.990	0.746	0.604
6	Neonates	0.986	0.957	0.961	0.973
	Infants	0.342	0.987	0.496	0.405
	Toddlers	0.530	0.981	0.663	0.589
	Preschoolers	0.719	0.957	0.737	0.728
	Children	0.808	0.940	0.760	0.783
	Adolescents	0.694	0.981	0.672	0.683

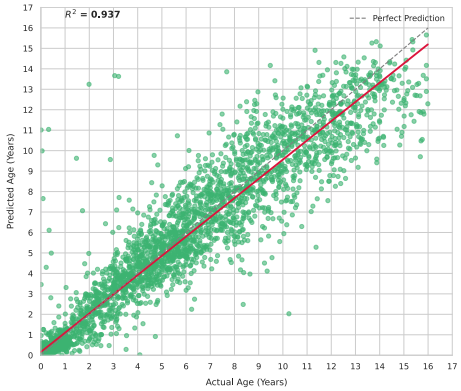

Figure 3: Scatter plot for the test set (12-lead).

for neonates (Se: 0.983) and preschoolers children (Se: 0.755), but lower performance for infants (Se: 0.551) and adolescents (Se: 0.535). The 8-lead setting also yielded strong results for neonates (Se: 0.974), but showed reduced performance for infants (Se: 0.503). The 6-lead setting performs well for neonates (Se: 0.986), but sensitivity decreased for infants (Se: 0.342) and toddlers (Se: 0.530). Overall, the 12-lead and 8-lead pECG settings provided the most consistent performance across age groups, whereas the 6-lead setting exhibited greater variability, particularly in younger age groups such as infants and toddlers.

In general, our results align with previous studies on pECG-based age and sex prediction, though direct comparisons are limited by differences in population distribution, dataset size, age groups, and model architecture. For instance, [13] reported a sex prediction AUC of 0.910, while we obtained 0.790, likely due to greater physiological variability and reduced ECG signal amplitude in children. Their dataset had 76% of samples from the 8–18 year age group, which may have made sex prediction easier. In age prediction, [7] reported a MAE of 2.12 years with a dataset skewed toward older groups, whereas our model achieved a significantly lower MAE of 0.532 years, despite a narrower age range and more younger samples. To make a thorough comparison, we computed the MAE separately for each age group using 12-lead pECG, yielding the following values: neonates (0.081), infants (0.581), toddlers (0.732), preschoolers (0.890), children (1.312), and adolescents (1.561). The MAE increases with age, while still remaining promising even for the older age groups. Regarding age-group prediction, [6] reported an average F1-score of 0.732 with four age groups (excluding adults), while our model achieved a comparable F1-score of 0.722

across six groups, demonstrating strong performance with the added complexity of age groups.

In addition to age prediction, sex classification performance was evaluated across age groups using the AUC. Moderate discriminative ability was observed in neonates, infants, and toddlers (AUCs 0.765–0.788), which improved in preschoolers and children (AUCs 0.826–0.837) and reached excellent accuracy in adolescents (AUC = 0.960). This finding is consistent with O'Sullivan *et al.* [13], who reported increasing accuracy with pubertal development.

Moreover, we trained age-only and sex-only single-task DLM baselines with the same number of parameters and hyperparameters as the multitask DLM for the 12-lead setting. The multitask model achieved lower age MAE (0.532 vs. 0.574), higher R^2 (0.932 vs. 0.924), and similar sex accuracy (0.712 vs. 0.708), while halving the parameter count relative to deploying two single-task models. Therefore, in this regard, the multitask DLM is more efficient than training two different DLMs.

Overall, AI models predicting age and sex from pECG data can be combined with other clinical variables to assess future cardiovascular risk. Discrepancies between predicted and chronological age may reveal underlying health issues, such as accelerated aging or developmental anomalies, allowing for earlier, targeted clinical interventions.

The main limitation of this study is the imbalanced dataset and potential overlap in pECG features across age groups, as well as possible regional or population-specific characteristics. Future work will validate the model using diverse, multi-center datasets and incorporate longitudinal pECG data to capture growth trajectories.

4. Conclusions

In this study, we presented a multitask DLM capable of classifying pediatric age and sex from 12-lead pECGs. The results demonstrated that the DLM can capture age- and sex-specific patterns in pECGs, offering new avenues for personalized medicine and enabling better predictive tools for early intervention and risk assessment.

Acknowledgments

This study was supported by the project FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence).

References

[1] World Health Organization. Cardiovascular diseases (CVDs), 2021. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. Eur Heart J 2021;42(46):4717–4730.

[3] Rahman M, Rivolta MW, Maison-Blanche P, Badilini F, Sassi R. Residual-attention deep learning model for atrial fibrillation detection from Holter recordings. J Electrocardiol 2024;84:12. ISSN 0022-0736.

[4] Lima EM, Ribeiro AH, Paixão GM, Ribeiro MH, Pinto-Filho MM, et al. Deep neural network-estimated electrocardiographic age as a mortality predictor. Nat Commun 2021;12(1):5117.

[5] Hirota N, Suzuki S, Motogi J, Nakai H, Matsuzawa W, et al. Cardiovascular events and artificial intelligence-predicted age using 12-lead electrocardiograms. Int J Cardiol Heart Vasc 2023;44:101172.

[6] An J, Bailey B, Gregg RE. Paradigm shift from feature-based machine learning to end-to-end deep residual neural networks for pediatric age classification from 12-lead ECG. In Comput Cardiol, volume 50. 2023; 1–4.

[7] Dutenhefner PR, Lemos G, Rezende T, Fernandes JG, et al. ECG-ResNeXt: Age prediction in pediatric electrocardiograms and its correlations with comorbidities. In Encontro Nacional de Inteligência Artificial e Computacional (ENIAC). 2024; 49–60.

[8] Chen J, Huang S, Zhang Y, Chang Q, Zhang Y, Li D, Qiu, et al. Congenital heart disease detection by pediatric electrocardiogram based deep learning integrated with human concepts. Nat Commun 2024;15(1):976.

[9] Rahman J, Brankovic A, Khanna S. Machine learning model with output correction: Towards reliable bradycardia detection in neonates. Comput Biol Med 2024;177:108658.

[10] Anjewierden S, O'Sullivan D, Mangold KE, Greason G, Attia IZ, et al. Detection of right and left ventricular dysfunction in pediatric patients using artificial intelligence–enabled ECGs. J Am Heart Assoc 2024;13(21):e035201.

[11] Mayourian J, El-Bokl A, Lukyanenko P, La Cava WG, Geva T, Valente AM, Triedman JK, Ghelani SJ. Electrocardiogram-based deep learning to predict mortality in paediatric and adult congenital heart disease. Eur Heart J 2025;46(9):856–868.

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

[13] O'Sullivan D, Anjewierden S, Greason G, Attia IZ, Lopez-Jimenez, et al. Pediatric sex estimation using AI-enabled ECG analysis: influence of pubertal development. NPJ Digit Med 2024;7(1):176.

Address for correspondence:

Md Mokelsur Rahman
Computer Science Department, University of Milan,
Via Celoria 18, 20133, Milan, Italy
md.rahman@unimi.it