Paradigm Shift from Feature-Based Machine Learning to End-to-End Deep Residual Neural Networks for Pediatric Age Classification from 12-Lead ECG

Junmo An[1], Ben Bailey[1], Richard E Gregg[1]

[1]Philips Healthcare, Cambridge, Massachusetts, USA

Abstract

ECG criteria vary significantly by age in early years of childhood. Accurate knowledge of age is vital to select appropriate ECG criteria for the patient. Two machine learning and deep learning models were trained on three different inputs from a dataset of 151,725 recordings of 12-lead ECGs. The feature-based machine learning model utilizing multi-layer perceptron (MLP) was trained with ECG features, such as heart rate, T-wave amplitude relative to QRS amplitude, QRS peak-to-peak amplitude, biphasic QRS complexes, QRS duration, and negative T-waves on leads V2 and V3. For the end-to-end residual network (ResNet) model, the first input was a 10-second ECG (RawECG) signal, and the second input was the corresponding 1.2-second average representative beat (RepBeat) signal. The ResNet was trained with data augmentation, including baseline wander addition, dropout, scaling, and sigmoid compression using RandAugment. The feature-based MLP achieved an F1 score of 0.72 for pediatric detection (age < 16 years) and an average F1 score of 0.65 for classifying pediatric groups (neonate, infant, child, and adolescent) and adult. The end-to-end ResNet achieved an F1 score of 0.88 for pediatric detection from RawECG input and an F1 score of 0.82 from RepBeat input. The model also achieved an average F1 score of 0.78 for classifying pediatric groups and adult from RawECG input, and an F1 score of 0.74 from RepBeat input. The proposed end-to-end ResNet with RawECG input outperforms the feature-based MLP. The deep learning model provided the highest F1 score making it the best choice for pediatric age classification.

1. Introduction

Electrocardiogram (ECG) is a widely employed non-invasive diagnostic tool that measures the electrical activity of the heart by placing multiple electrodes on the torso and limbs. The standard 12-lead ECG, which typically lasts for 10 seconds using 10 electrodes (RA, LA, RL, LL, V1 to V6) and cables, has been extensively utilized for precise diagnosis and efficient monitoring of a wide range of cardiac abnormalities in clinical settings.

The accurate and timely interpretation of 12-lead ECGs is crucial for diagnosing cardiac arrhythmias and abnormalities, whether performed by clinicians or automated interpretation machines. However, automated, rule-based ECG analysis is often performed without entering age information, which can be a challenge in pediatric and adult cases. ECGs can vary significantly between pediatrics and adults, so age is a crucial determinant for establishing accurate ECG thresholds. This significantly impacts accurate diagnosis in patients of all age groups. Consequently, incorrect classification of ECGs due to the lack of age information poses a substantial risk of misdiagnosing cardiac diseases.

To address the challenges of age-insensitive automated ECG analysis, there is a need to develop methods for estimating age from ECGs. Traditional methods for age estimation from ECGs utilize signal processing techniques to extract various features from the ECG signal, including heart rate, QRS duration/amplitude, and T-wave amplitude. The extracted information is then provided as input features to a classifier or regressor [1]–[4]. On the other hand, end-to-end deep learning (DL) models for age estimation use the raw ECG signal as input to a classifier or regressor that learns to extract relevant features without any prior expert knowledge [5]–[13]. However, all of these previously proposed approaches have only been evaluated on adult ECG databases. Additionally, there is a lack of studies comparing the performance of feature-based ML and end-to-end DL models for age estimation or binning.

To our knowledge, no previous study has used a multi-site pediatric ECG dataset to classify pediatric age groups and adult. The primary objective of our study is to develop a deep residual neural network (ResNet) model to identify pediatric patients from 12-lead ECGs and classify them into five age groups: neonates, infants, children, adolescents, and adults. In this study, we outline the development and validation process of our end-to-end ResNet-based DL model compared to the traditional feature-based ML model for pediatric groups and adult age classification as well as pediatric detection. Our model's exceptional performance in pediatric age detection and age group classification from 12-lead diagnostic ECGs indicates its potential to improve rule-based performance by correct choice of age-based ECG criteria. This could result in improved diagnostic and treatment accuracy, leading to better patient outcomes.

2. Methods

2.1. Dataset and Preprocessing

A dataset of 151,725 recordings of 12-lead ECGs was acquired at a sampling rate of 500 Hz at 5 uV/bit digitization at the Lucile Packard Children's Hospital at Stanford (Palo Alto, CA), University of Florida Medical Center (Jacksonville, FL), Vanderbilt University Medical Center (Nashville, TN), and a large teaching hospital (Minneapolis, MN). The dataset comprises 8,242 pediatric ECG recordings and 143,483 adult ECG recordings. The pediatric ECGs were classified into four groups: neonates (932 recordings: newborns up to one month of age), infants (1,297 recordings: one month to two years of age), children (2,574 recordings: two to twelve years of age), and adolescents (3,439 recordings: twelve to sixteen years of age).

Three different inputs from the dataset were used for training and testing the feature-based multilayer perceptron (MLP) network model and end-to-end deep learning model based on residual network (ResNet). The input to the feature-based machine learning model was ECG features, including heart rate, T-wave amplitude relative to QRS amplitude, QRS peak-to-peak amplitude, biphasic QRS complexes, QRS duration, and negative T-waves on leads V2 and V3 [1]. These custom features were extracted and provided to the MLP model. For the end-to-end ResNet model, the first input was a 10-second 12-lead ECG (RawECG) signal, and the second input was the corresponding 1.2-second average representative beat (RepBeat) signal. The RawECG signal was used to capture the overall shape of the ECG, while the RepBeat signal was used to capture the most important morphological features of the ECG that were automatically extracted by the Philips DXL algorithm.

2.2. Data Augmentation and Automation

The ResNet model was trained with the random combination of data augmentation techniques such as baseline wander addition, dropout, scaling, and sigmoid compression using the modified RandAugment automation method [14]–[16]. During the training phase, four data augmentation techniques were applied randomly using the RandAugment method with a probability of 0.3 for each technique. Figure 1 shows the representations of the four selected data augmentation schemes applied to lead II. To illustrate, baseline wander involved adding randomly generated sinusoidal wave signals at frequencies of 0.05, 0.1, 0.15, 0.2, and 0.5 Hz, along with random phase shifts, to the original signal across all leads. Dropout was employed by randomly setting 5% of signal values to zero for all leads. Scaling was applied by randomly compressing or stretching the amplitude of the signal using a random factor for all leads. Sigmoid compression was utilized by applying a sigmoidal activation function, defined as $(1 / (1 + e^{-x}))$, with $x = 2$ to the signal for all leads.

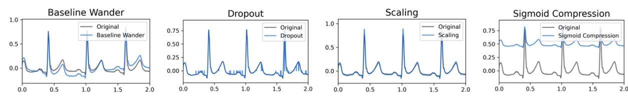

Figure 1. Data augmentation schemes (applied to lead II here).

2.3. Model Architecture

In this study, we developed and validated two machine learning (ML) and deep learning (DL) models to detect pediatric age by 12-lead ECGs and classify them into different pediatric groups and adult. The first model is a MLP network using the custom extracted features for feature-based ML, while the second model is a ResNet model using the RawECG and RepBeat for end-to-end DL.

The MLP network consisted of three layers: an input layer, a hidden layer with 100 ReLU (rectified linear) activation units, and an output layer. Figure 2 shows the architecture of the end-to-end ResNet model designed for classifying pediatric and adult subjects, and further distinguishing between pediatric groups and adult classes. The ResNet inputs are the RawECG or RepBeat signals. The ResNet model architecture comprises a one-dimensional (1D) convolutional layer followed by four residual blocks. After the convolutional layer, there is batch normalization (BatchNorm), a leaky rectified linear unit (LeakyReLU) activation function, dropout (Dropout) with a probability of 0.2, and adaptive max-pooling (MaxPool1D). The output layer of the ResNet model performs the classification process, which can be either binary or multiclass classification. It is fed into a fully connected (dense) layer with a dropout probability of 0.2 and a sigmoid activation function to perform binary classification for distinguishing between pediatric and adult subjects. The sigmoid activation function outputs a probability between 0 and 1 for binary classification (pediatric or adult). In contrast, the softmax activation function in the output layer is used to output a probability distribution over five potential classes (neonate, infant, child, adolescent, and adult) for multiclass classification. The ResNet model minimizes two loss functions: binary cross-entropy loss for binary classification (such as classifying pediatric and adult subjects) and cross-entropy loss for multiclass classification (such as further distinguishing between pediatric groups and adult classes). The Python code for preprocessing, model building, binary and multiclass classifications, training, validation, testing, data augmentation, and workflow automation was implemented using the PyTorch, Scikit-Learn, SciPy, and WFDB libraries.

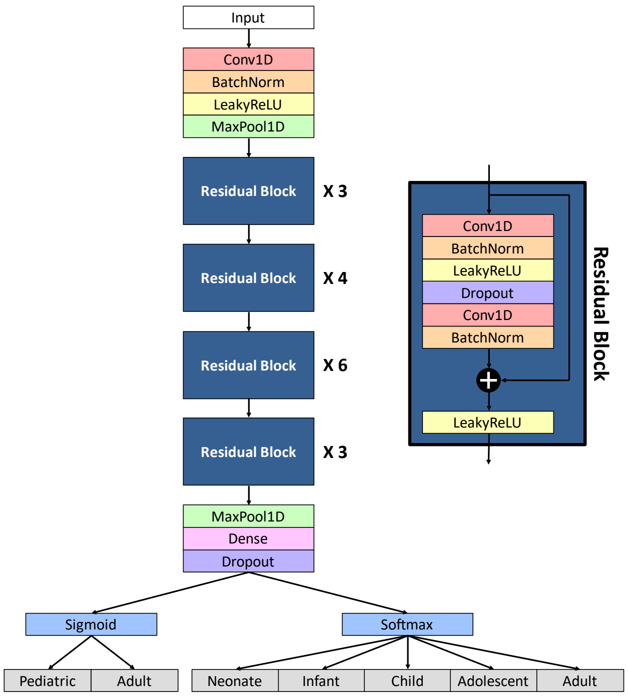

Figure 2. The proposed architecture of the end-to-end ResNet-based deep learning model.

2.4. Training and Performance Evaluation

The proposed ResNet model trained for 40 epochs with a batch size of 128. The adaptive moment estimation (Adam) was used to optimize losses with an initial learning rate of 0.0001. Additionally, the learning rate was decayed by a factor of 0.1 every 10 epochs during the training process.

The comprehensive evaluation of the feature-based MLP and end-to-end ResNet models involved diverse and extensive datasets containing pediatric and adult ECGs. The performance of the models was evaluated using 10-fold nested cross-validation in classifying two groups (pediatrics vs. adults) for pediatric detection and 5-fold nested cross-validation in classifying five groups (neonates, infants, children, adolescents, and adults) for pediatric groups and adult classification.

The classification performance, including sensitivity, specificity, positive predictive value (PPV), Area Under the Receiver Operating Characteristic curve (AUROC), and F1 score was calculated.

3. Results

The experiments were conducted on a desktop computer containing a Intel® Core™ i7-7700 processor running at a clock speed of 3.60 GHz. The system was equipped with 64.0 GB of RAM and utilized a Graphics Processing Unit (GPU) card from NVIDIA, specifically the GeForce RTX 3080 Ti.

Table 1 compares the performance of pediatric age detection between the feature-based MLP and end-to-end ResNet models with RepBeat and RawECG. The ResNet-based model with RawECG achieved the highest F1 score for pediatric age detection, with a relative improvement of 22.8% over the feature-based MLP model (F1 score = 0.72). The ResNet model with RepBeat also had a higher F1 score, with a relative improvement of 14.4% over the MLP model.

Table 1. Performance comparison of binary classification for pediatric vs. adult.

Model	Sens.	Spec.	PPV	AUROC	F1
MLP with features	0.656	0.989	0.785	0.969	**0.715**
ResNet with RepBeat	0.791	0.992	0.847	0.985	**0.818**
ResNet with RawECG	0.826	0.997	0.937	0.991	**0.878**

Table 2 compares the classification performance of the MLP with extracted features and ResNet with RepBeat and RawECG inputs in classifying five groups (neonates, infants, children, adolescents, and adults) for pediatric groups and adult classification. Similar to the pediatric detection results, the ResNet-based model with RawECG achieved the highest average F1 score across all five classes for four pediatric groups and adult classification, with a relative improvement of 20.4% over the feature-based MLP model (average F1 score = 0.65). The ResNet model with RepBeat also had a higher average F1 score, with a relative improvement of 13.4% over the MLP.

Table 2. Performance comparison of multiclass classification for four pediatric groups and adult.

Model	Class	Sens.	Spec.	PPV	AUROC	F1
MLP with features	Neonate	0.690	0.998	0.642	0.996	0.665
	Infant	0.575	0.997	0.591	0.993	0.583
	Child	0.612	0.995	0.693	0.988	0.650
	Adolescent	0.284	0.994	0.522	0.977	0.368
	Adult	0.996	0.709	0.983	0.980	0.796
	Average	0.631	0.939	0.686	0.987	**0.651**
ResNet with RepBeat	Neonate	0.808	0.999	0.765	0.997	0.786
	Infant	0.692	0.998	0.724	0.995	0.708
	Child	0.693	0.996	0.777	0.991	0.733
	Adolescent	0.366	0.996	0.668	0.981	0.473
	Adult	0.996	0.709	0.983	0.980	0.990
	Average	0.711	0.940	0.783	0.989	**0.738**
ResNet with RawECG	Neonate	0.802	0.999	0.859	0.998	0.829
	Infant	0.730	0.998	0.768	0.996	0.749
	Child	0.829	0.994	0.692	0.991	0.754
	Adolescent	0.505	0.996	0.729	0.984	0.597
	Adult	0.995	0.808	0.989	0.985	0.992
	Average	0.772	0.959	0.807	0.991	**0.784**

4. Discussion and Future Work

We developed and evaluated an end-to-end ResNet model for automatic detection of pediatric patients and classification of pediatric age groups and adult from 12-lead ECGs. Our model outperformed a feature-based MLP model. Additionally, we found that RawECG signals using our ResNet model contained more valuable information for age estimation than RepBeat signals. Our findings suggest that the ResNet model is a powerful tool for accurately classifying all age groups from 12-lead ECG signals, potentially improving diagnostic accuracy in clinical practice.

This study has several limitations, and there are still opportunities for future work. First, the datasets utilized in this study are imbalanced between pediatric and adult ECGs, which suggests that the proposed model would benefit from a more extensive pediatric database and a larger multi-site database to enhance its generalization capabilities. Second, while the proposed model performed well in pediatric age detection and classification of pediatric age groups and adult, it may be necessary to explore the implementation of deep learning regression models for all ages estimation to enhance its overall capabilities. Third, although the binary and multiclass classification performance using the four selected data augmentation techniques yielded good results, there is still a need to investigate and identify the most effective data augmentation techniques for age estimation, aiming to improve the model's generalization. Finally, further research is needed to explore the explainability of the DL model using explainable AI (xAI) techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), SHapley Additive exPlanations (SHAP), and other relevant approaches. This would provide us with valuable insights into the model's decision-making process, discover new potential features to improve its performance, enhance its interpretability/explainability, and make it more appropriate for clinical applications/use.

Acknowledgments

The authors thank Yu-He Zhang, Sina Zarei, Wei Zong, Abhinay Pandya, Hyeonwoo Lee, Eric Helfenbein, and Milan Petkovic for their support.

References

[1] R. E. Gregg, *Automated detection of patient age by ecg*. Patent No. WO2015079387A1, 2015.

[2] V. Starc *et al.*, “Can functional cardiac age be predicted from the ECG in a normal healthy population?,” in *2012 Computing in Cardiology*, IEEE, 2012, pp. 101–104.

[3] R. L. Ball, A. H. Feiveson, T. T. Schlegel, V. Stare, and A. R. Dabney, “Predicting ‘heart age’ using electrocardiography,” *J. Pers. Med.*, vol. 4, no. 1, pp. 65–78, 2014.

[4] T. Lindow, I. Palencia-Lamela, T. T. Schlegel, and M. Ugander, “Heart age estimated using explainable advanced electrocardiography,” *Sci. Rep.*, vol. 12, no. 1, p. 9840, 2022.

[5] Z. I. Attia *et al.*, “Age and sex estimation using artificial intelligence from standard 12-lead ECGs,” *Circ. Arrhythm. Electrophysiol.*, vol. 12, no. 9, p. e007284, 2019.

[6] K. H. Lee and S. Byun, “Age Prediction in Healthy Subjects Using RR Intervals and Heart Rate Variability: A Pilot Study Based on Deep Learning,” *Appl. Sci.*, vol. 13, no. 5, p. 2932, 2023.

[7] E. M. Lima *et al.*, “Deep neural network-estimated electrocardiographic age as a mortality predictor,” *Nat. Commun.*, vol. 12, no. 1, p. 5117, 2021.

[8] A. O. Ladejobi *et al.*, “The 12-lead electrocardiogram as a biomarker of biological age,” *Eur. Heart J.-Digit. Health*, vol. 2, no. 3, pp. 379–389, 2021.

[9] H. E. van der Wall *et al.*, “Cardiac age detected by machine learning applied to the surface ECG of healthy subjects: Creation of a benchmark,” *J. Electrocardiol.*, vol. 72, pp. 49–55, 2022.

[10] N. Hirota *et al.*, “Cardiovascular events and artificial intelligence-predicted age using 12-lead electrocardiograms,” *IJC Heart Vasc.*, vol. 44, p. 101172, 2023.

[11] J. Libiseller-Egger *et al.*, “Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes,” *Sci. Rep.*, vol. 12, no. 1, p. 22625, 2022.

[12] E. D. Benavente *et al.*, “Studying accelerated cardiovascular ageing in Russian adults through a novel deep-learning ECG biomarker,” 2021.

[13] C.-H. Chang, C.-S. Lin, Y.-S. Luo, Y.-T. Lee, and C. Lin, “Electrocardiogram-based heart age estimation by a deep learning model provides more information on the incidence of cardiovascular disorders,” *Front. Cardiovasc. Med.*, vol. 9, p. 754909, 2022.

[14] J. An, R. E. Gregg, and S. Borhani, “Effective Data Augmentation, Filters, and Automation Techniques for Automatic 12-Lead ECG Classification Using Deep Residual Neural Networks,” in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2022, pp. 1283–1287.

[15] J. An, R. E. Gregg, and S. Babaeizadeh, “Effects of data augmentation and filters for automatic diagnosis of 12-lead ECG using residual network based deep learning,” in *2022 46th Annual Conference of the International Society for Computerized Electrocardiology (ISCE)*, ISCE, 2022.

[16] J. An, B. Bailey, and R. E. Gregg, “Automated classification of pediatric age from 12-lead ECG using deep residual convolutional neural networks,” *J. Electrocardiol.*, vol. 78, p. 21, 2023

Address for correspondence:

Junmo An, PhD
222 Jacobs St, Cambridge, MA 02141, USA
junmo.an@philips.com