Automated Pediatric Ventricular Pre-Excitation (Wolff-Parkinson-White Syndrome) Detection in 12-Lead ECG Using Deep Learning and SHAP-Based Interpretability

Junmo An[1], Richard E Gregg[1], Ben Bailey[1], Yu-He Zhang[1], Abhinay Pandya[1], Wei Zong[1], Hyeonwoo Lee[1], Dillon J Dzikowicz[1], Anne M. Dubin[2]

[1]Philips, Cambridge, Massachusetts, USA
[2]Stanford University School of Medicine, Palo Alto, California, USA

Abstract

Wolff Parkinson White (WPW) syndrome, a significant cause of ventricular pre-excitation, presents diagnostic challenges in patients where early detection is crucial to prevent life threatening arrhythmias. The condition is characterized by electrocardiogram (ECG) features such as a short PR interval and a delta wave. Although traditional rule-based interpretation algorithms and feature-based machine learning (ML) models exist, they often lack the accuracy required for consistent and reliable WPW diagnosis. This study evaluates an end-to-end deep learning (DL) model for automated WPW detection from 10-second 12-lead ECGs in patients under 17 years old, a presentation that is relatively rare, compared to a ML model. Our results show that the DL model substantially outperforms the feature-based ML model, demonstrating superior accuracy in identifying pediatric WPW. To address the "black box" nature of DL, we applied Shapley additive explanations (SHAP) to improve model transparency. SHAP analysis visually highlights critical ECG segments, such as the delta wave, that most influence the model's decisions. This explainable AI-powered DL model provides clinicians with a powerful, transparent tool for precise WPW diagnosis, supporting improved clinical risk stratification and management.

1. Introduction

Wolff Parkinson White (WPW) syndrome, characterized by ventricular pre-excitation (VPE) due to the presence of an accessory atrioventricular pathway, is a recognized cause of supraventricular tachycardia and, in some cases, sudden cardiac death in young individuals under 17 years of age [1]. The prevalence of pediatric WPW in the general population is estimated at 0.1 to 0.3 percent, making it comparatively rare and challenging for the development of automated detection algorithms [1], [2]. Early and accurate identification of WPW from a standard 10-second 12-lead electrocardiogram (ECG) is critical for risk stratification and treatment planning, particularly in pediatric patients, where subtle variations in ECG morphology, age-related changes in conduction, and accessory pathway location can complicate diagnosis.

Pediatric WPW diagnosis relies on detecting characteristic ECG features such as a short PR interval, prolonged QRS duration, abnormal QRS morphology, and the presence of a delta wave. However, children often display normal developmental variations on their ECGs, which can obscure these features and lead to false-positive interpretations. Previous studies report substantial interobserver variability in interpreting pediatric ECGs for pre-excitation, highlighting the need for robust automated approaches [1]. Automated rule-based algorithms and feature-based machine learning (ML) models have been developed for pediatric WPW detection [1], [3], but these methods may often be limited by fixed thresholds, handcrafted features, and difficulties in generalizing across diverse pediatric ECG presentations, particularly in the presence of noisy ECGs or atypical morphologies.

Recent advances in ML and deep learning (DL) have transformed automated ECG interpretation [4], [5], [6]. While feature-based ML methods depend on manual extraction of morphological and rhythm parameters, end-to-end DL models learn relevant features directly from raw ECG waveforms, bypassing the need for extensive preprocessing. This paradigm shift has led to significant improvements in tasks such as arrhythmia detection, conduction block classification, age group classification, and ECG-based patient phenotyping [4], [6]. Despite these successes, the "black box" nature of DL models remains a barrier to clinical adoption. For pediatric WPW detection, where treatment decisions can involve invasive electrophysiological studies or ablation, explainability and transparency are crucial for clinician trust.

To address this, interpretable AI techniques such as Shapley Additive Explanations (SHAP) have emerged [7],

[8], providing quantitative insights into feature importance and local decision reasoning. Such approaches not only enhance model transparency but also enable validation against established ECG criteria, bridging the gap between data-driven predictions and clinical interpretability.

In this study, we present and evaluate an end-to-end DL model for automated detection of pediatric WPW from 10-second 12-lead ECGs. Our work makes two primary contributions: (1) we compare the diagnostic performance of a feature-based ML model and our proposed end-to-end DL model; and (2) we integrate SHAP-based interpretability to identify and highlight ECG segment importance driving the model's predictions. Our results demonstrate that the DL model outperforms feature-based ML model. The SHAP analysis highlights physiologically relevant regions, such as the QRS upstroke and delta wave morphology. This enhances both the accuracy and explainability of automated pediatric WPW detection.

2. Methods

2.1. Dataset and Preprocessing

Digital 10-second 12-lead ECGs were retrospectively gathered from pediatric patients (<17 years old) across multiple institutions as part of a cohort study. All ECGs were annotated by board-certified pediatric cardiologists for the presence or absence of VPE consistent with WPW syndrome. The diagnosis was based on established criteria, including a short PR interval, widened QRS complex, slurred upstroke of the QRS, delta wave morphology, and absence of normal septal Q waves in lateral leads [9]. Cases with uncertain interpretation were excluded to ensure a high-confidence reference standard.

The pediatric training database (n = 772) included 36 ECGs with WPW, 492 normal ECGs, and conduction defect cases with right bundle branch block (RBBB, n = 232) and left bundle branch block (LBBB, n = 11). The independent test set (n = 764), including 18 WPW ECGs, was randomly sampled from the same source and contained WPW cases not present in the training set.

All ECGs were sampled at 500 Hz with 5 μV resolution. For the end-to-end DL model, two types of inputs were used: (1) the raw 10-second 12-lead ECG waveform (RawECG) and (2) the 1.2-second 12-lead averaged representative beat (RepBeat). The RepBeat was generated using the Philips DXL algorithm, a commercially available and widely used method, by averaging beats of similar morphology, resulting in a noise-reduced waveform that highlights key diagnostic features. Both RawECG and RepBeat signals were normalized lead-wise prior to model training.

For comparison, the performance of both the ML and DL models was systematically evaluated by measuring their agreement with annotations provided by pediatric cardiologists on an independent test set, allowing for a direct assessment of their diagnostic accuracy.

2.2. Data Augmentation and Automation

To improve robustness and generalization, the DL model was trained with a random combination of three data augmentation techniques: cutout, dropout, and scaling, applied using a modified RandAugment automation framework [4], [5], [8], [10]. During each training iteration, one of the three techniques was selected with equal probability (0.333). Cutout set a contiguous segment of up to 10% of the signal to zero at a random time across all leads. Dropout randomly set 5% of signal values to zero, simulating transient loss. Scaling adjusted the amplitude by randomly compressing or stretching the signal uniformly across all leads. Figure 1 shows the visual effect of each augmentation on the ECG waveform.

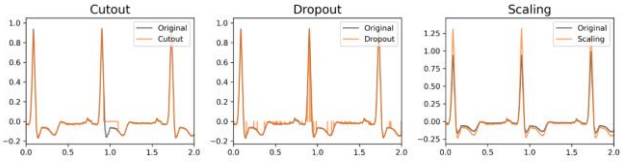

Figure 1. Data augmentation techniques incorporating cutout, dropout, and scaling applied to lead II.

2.3. Algorithm and Model Architecture

In this study, we adopted a comprehensive strategy to accurately detect WPW from pediatric 12-lead ECGs by evaluating both feature-based ML and end-to-end DL methods. The first approach uses a feature-based ML model, while the second employs an Attention Residual Network (AResNet) architecture, taking RawECG and RepBeat as inputs for end-to-end DL.

The feature-based ML model was implemented using a decision tree for pediatric WPW detection, incorporating the following features: delta wave score; R-peak time of the Frank lead vector magnitude signal; frontal plane QRS axis; QRS duration; PR segment; maximum frontal plane P-wave vector magnitude; spatial QRS-T angle at maximal vector magnitude; logarithm of age; Q-wave area in lead V6; spatial QRS-T angle at the mean point of the vector loops; and PJ interval [1]. As shown in Figure 2, the AResNet-based architecture is designed for accurate WPW detection using DL. The network begins with a 1D convolutional layer followed by batch normalization, Exponential Linear Unit (ELU) activation, and max pooling. It then processes the input through four sequential residual stages. Within each residual block, convolutional layers, batch normalization, ELU activation, dropout (p = 0.3), and an attention mechanism are applied to enhance informative features. Downsampling layers are included when required to match dimensions. After the final

residual stage, the model applies both adaptive average pooling and adaptive max pooling, concatenating their outputs. The concatenated features are flattened and passed to a fully connected layer to produce the final output. For binary WPW classification, the model outputs a single logit, which can be converted to a probability using the sigmoid function during inference. Binary cross-entropy is used as the training loss to optimize classification performance. The development process, including data preparation, model building, training, evaluation, testing, augmentation, and process orchestration, was facilitated by Python packages like PyTorch, Scikit-Learn, SciPy, and the Waveform Database (WFDB) toolkit.

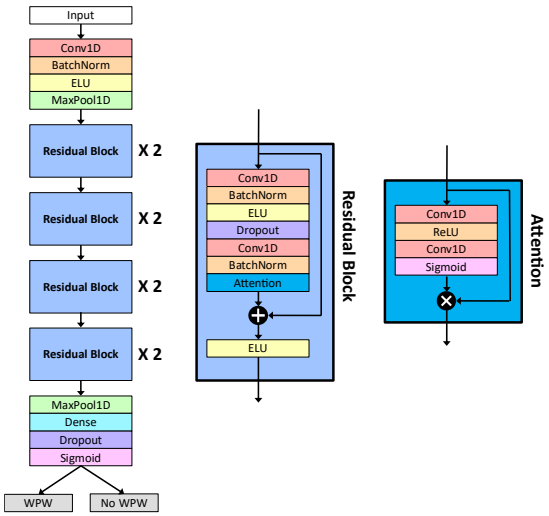

Figure 2. The proposed architecture of the end-to-end Attention ResNet-based deep learning model.

2.4. Training and Performance Evaluation

The AResNet model underwent training for 50 epochs with a batch size of 16, employing the Adam optimizer with an initial learning rate of 0.0001 and a 10% decay scheduled every 10 epochs. The model that achieved the highest F1 score during training was selected for testing. For WPW detection, performance metrics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and F1 score were calculated.

2.5. Explainability

Although end-to-end DL models typically outperform traditional rule-based algorithms and feature-based ML approaches, their intricate architectures and non-linear decision processes often render them difficult for human interpretation [4]. This opacity has earned them the label of "black box". To gain insight into the internal mechanisms of our AResNet model and identify the ECG features critical for WPW detection, we applied SHAP [7], [8], a robust interpretability method that quantifies the contribution of each input ECG signal to the model's output. In the context of ECG binary classification using the AResNet model, SHAP helps reveal which specific segments or patterns in the ECG signals most significantly influence the model's detection decisions.

3. Results

Our experiments were performed on a powerful desktop computer equipped with an Intel Core i9-12900K processor clocked at 3.5 GHz, 64 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU to accelerate DL model processing. Table 1 details the performance metrics for the feature-based ML model and the end-to-end AResNet model using both RepBeat and RawECG inputs for pediatric WPW detection. The AResNet model trained on RawECG data achieved the best results for WPW detection. It posted an F1 score of 0.743, a sensitivity of 0.813, a specificity of 0.992, a PPV of 0.684, an NPV of 0.996, and an accuracy of 0.988. This represents a significant improvement in the F1 score: a 26.15% relative increase over the feature-based ML model (0.589) and a 5.24% relative increase over the AResNet model trained with RepBeat data (0.706). These results confirm that the end-to-end DL approach substantially outperforms the conventional feature-based ML model for pediatric WPW detection. Furthermore, the superior performance of the AResNet model using RawECG compared to the one using RepBeat suggests that longer raw ECG segments contain more critical clinical information than averaged beats. This information is essential for capturing the dynamic changes in cardiac electrical activity characteristic of WPW.

Table 1. Performance comparison of feature-based machine learning and end-to-end deep learning models for pediatric WPW detection.

Algorithm / Model	Sens.	Spec.	PPV	NPV	Acc.	F1
Feature-based ML	0.625	0.989	0.556	0.992	0.982	**0.589**
AResNet with RepBeat	0.750	0.992	0.667	0.995	0.987	**0.706**
AResNet with RawECG	0.813	0.992	0.684	0.996	0.988	**0.743**

SHAP analysis was applied to determine which morphological characteristics of the ECG the AResNet model prioritizes when identifying WPW syndrome. Figure 3 presents an example of the resulting feature importance map generated from the AResNet model using the RepBeat input. Given the morphology focused nature of WPW and to enhance visual clarity, only the RepBeat

based SHAP results are shown; RawECG results were omitted due to space constraints. SHAP scores were normalized to a 0 to 1 scale and visualized as a heatmap overlay on the ECG trace, where red highlights areas of highest relevance and blue represents minimal influence. Dot size corresponds to the magnitude of feature contribution. For clarity, the figure includes only the leads with the highest overall SHAP scores, selected automatically. The analysis shows that the model consistently focuses on hallmark WPW indicators, most notably delta wave morphology and prolonged QRS duration, closely matching recognized diagnostic guidelines. These results support the model's clinical relevance and its consistency with established ECG interpretation standards.

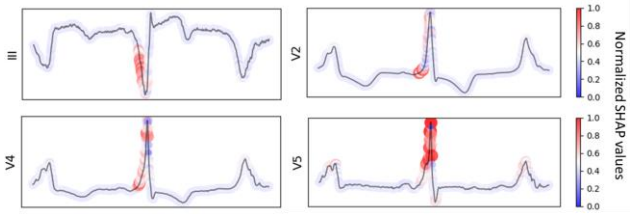

Figure 3. Feature importance visualization using SHAP with RepBeat input for model interpretability.

4. Discussion and Future Work

We designed and assessed two methods for identifying pediatric WPW syndrome using standard 10-second 12-lead ECGs: a feature-based ML model and an end-to-end deep learning model. Our findings indicate that the DL model based on AResNet achieved significantly higher performance in detecting pediatric WPW related ECG patterns compared to the feature-based ML approach. To improve model transparency and support clinical decision making, we utilized the SHAP explainability method to highlight important ECG features and waveform regions such as delta wave morphology and prolonged QRS duration that the model relies on for its predictions.

This study has some limitations and offers several opportunities for future research. First, because the dataset comes only from the United States, it is important to evaluate the model on datasets from multiple centers that include diverse populations and countries to confirm its general applicability. Second, further investigation and systematic comparison of additional explainable AI methods such as Grad-CAM, LIME, and other advanced techniques could help improve interpretability. Third, combining both qualitative expert assessments and quantitative measures will provide a more complete evaluation of these explainable AI approaches. Future studies will focus on validating the ML and DL models on larger and more diverse populations and geographic regions, examining performance differences across demographic and clinical categories, integrating the DL model with clinical information to support personalized care for pediatric WPW patients, and applying explainable AI to identify novel features for more robust feature-based ML models.

Acknowledgments

The authors thank Eric Helfenbein, Joost Mans, Ziyu Li, Naredos Almaw, Milan Petkovic, and Dieter Haase for their support.

References

[1] R. E. Gregg, S. H. Zhou, and A. M. Dubin, "Automated detection of ventricular pre-excitation in pediatric 12-lead ECG," *J. Electrocardiol.*, vol. 49, no. 1, pp. 37–41, 2016.

[2] R. Kobza *et al.*, "Prevalence of preexcitation in a young population of male Swiss conscripts," *Pacing Clin. Electrophysiol.*, vol. 34, no. 8, pp. 949–953, 2011.

[3] S. M. Baek, M. K. Song, J.-S. Uhm, G. B. Kim, and E. J. Bae, "New algorithm for accessory pathway localization focused on screening septal pathways in pediatric patients with Wolff-Parkinson-White syndrome," *Heart Rhythm*, vol. 17, no. 12, pp. 2172–2179, 2020.

[4] J. An, B. Bailey, and R. E. Gregg, "Paradigm Shift from Feature-Based Machine Learning to End-to-End Deep Residual Neural Networks for Pediatric Age Classification from 12-Lead ECG," in *2023 Computing in Cardiology (CinC)*, IEEE, 2023, pp. 1–4.

[5] 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.

[6] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Syst. Appl. X*, vol. 7, p. 100033, 2020.

[7] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[8] J. An, R. E. Gregg, B. Bailey, Y.-H. Zhang, and D. J. Dzikowicz, "Left Bundle Branch Block Detection in 12-Lead ECG Using End-to-End Deep Learning with Explainability," in *2024 Computing in Cardiology (CinC)*, IEEE, 2024.

[9] B. Surawicz and T. Knilans, *Chou's electrocardiography in clinical practice: adult and pediatric*. Elsevier Health Sciences, 2008.

[10] 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," *J. Electrocardiol.*, vol. 73, pp. 16–17, 2022.

Address for correspondence:

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