Left Bundle Branch Block Detection in 12-Lead ECG Using End-to-End Deep Learning with Explainability

Junmo An[1], Richard E Gregg[1], Ben Bailey[1], Yu-He Zhang[1], Dillon J Dzikowicz[1,2]

[1]Philips, Cambridge, Massachusetts, USA
[2]University of Rochester, Rochester, New York, USA

Abstract

Cardiac resynchronization therapy (CRT) has shown potential in improving left ventricular function, particularly in patients with heart failure with reduced ejection fraction and left bundle branch block (LBBB). LBBB is diagnosed by a distinctive pattern on a 12-lead electrocardiogram (ECG) with a QRS duration exceeding 120 milliseconds and specific changes in chest leads. Despite advances in automated ECG interpretation, current rule-based algorithms for LBBB detection need refinement. The rise of machine learning (ML) and deep learning (DL) offers new possibilities for enhancing cardiac abnormality detection. This study evaluates the performance of rule-based algorithm, feature-based ML model, and end-to-end DL model in LBBB detection. Our findings reveal that the DL model outperforms both rule-based and ML approaches, demonstrating superior accuracy in LBBB detection. To address the 'black box' nature of DL, we employ Shapley additive explanations (SHAP), which enhance model transparency and visually highlight critical ECG segments relevant to LBBB diagnosis. Explainable AI-powered DL offers clinicians a robust tool for precise LBBB diagnosis, providing transparency into the algorithm's decision-making process for identifying those patients who may benefit from CRT.

1. Introduction

Cardiac resynchronization therapy (CRT) is currently recommended for patients with heart failure with reduced ejection fraction who remain symptomatic despite optimal medical therapy. Left Bundle Branch Block (LBBB), observed as a distinct QRS pattern on the 10-second 12-lead electrocardiogram (ECG), has been identified as an important predictor of a beneficial CRT response. However, over time, several definitions of LBBB have been proposed making it more challenging to identify which patients may benefit from CRT [1]. Despite advancements in automated ECG interpretation, the performance of rule-based algorithms for LBBB detection within 12-lead ECGs based on LBBB criteria still needs further significant enhancement.

The rise of machine learning (ML) and deep learning (DL) has transformed the detection of cardiac abnormalities, addressing its inherent challenges. Traditional methods relied on hand-crafted features based on morphology and rhythm extracted from ECGs, requiring complex pre-processing and clinician expertise before being input into various ML models. In contrast, end-to-end DL models offer a paradigm shift, automatically extracting relevant features and detecting cardiac abnormalities directly from raw ECG data. This eliminates the need for manual feature engineering and allows for streamlined, potentially more accurate diagnosis. Several studies have demonstrated the superiority of end-to-end DL models compared to feature-based ML approaches. In tasks like LBBB detection [2], atrial fibrillation detection [3], age prediction or classification [4], and multiclass classification [3], [5], DL models have consistently achieved higher accuracy and efficiency. This suggests that end-to-end DL could significantly improve ECG interpretation.

While the promise of end-to-end DL for automated ECG analysis is remarkable, concerns over its opaque 'black box' nature necessitate efforts to strengthen trust in these models. Particularly, clinicians, policymakers, and regulators require explainable and accountable DL models. These models should offer interpretable diagnosing reasoning to mitigate uncertainty and empower informed clinical decisions. To address the 'black box' issue of end-to-end DL models, researchers have developed interpretable techniques such as gradient-weighted class activation mapping (Grad-CAM) [6], local interpretable model-agnostic explanation (LIME) [7], and Shapley additive explanations (SHAP) [8].

This study makes two key contributions to the field of LBBB detection from 12-lead ECG. First, we develop and compare the performance of rule-based algorithm (standard practice), feature-based ML, and end-to-end DL models to determine their effectiveness. Second, we enhance LBBB detection by exploring ECG explainability

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.067

for deeper model insights. Our end-to-end DL model surpasses both rule-based algorithm and feature-based ML model for LBBB detection, demonstrating superior performance in ECG criteria selection and feature importance identification. Additionally, our model visually highlights specific morphological features associated with the LBBB in the ECG for better understanding of the reasons. This combination could lead to more precise diagnoses for clinicians, improved understanding of clinical evidence in cardiac diagnosis, and the possibility of better treatment decisions and outcomes for cardiac disease patients.

2. Methods

2.1. Dataset and Preprocessing

In 2018, the International Society for Computerized Electrocardiology (ISCE) and the Telemetric and Holter Warehouse Initiative (THEW: E-OTH-12-0602-024) organized a competition to develop automated algorithms that could detect strict LBBB in heart failure patients. The data used for this challenge stemmed from the MADIT-CRT trial at the University of Rochester (Rochester, NY), which focused on patients receiving CRT. The MADIT-CRT trial defined LBBB by a QRS duration of at least 130 milliseconds, a QS or rS pattern in lead V1, wide and often notched or slurred R waves in leads I, aVL, V5, and V6, and the absence of q waves in leads V5 and V6 [9]. The ECGs were recorded using 24-hour Holter recorders with a sampling frequency of 1 kHz and amplitude resolution of 3.75 microvolts per least significant bit. To ensure consistency, only 10-second ECG segments were extracted from the first 20 minutes of supine recordings. The challenge provided two datasets: a training set with 300 ECGs (174 with LBBB, 126 without) and a test set with 302 ECGs (156 with LBBB, 146 without).

To develop and evaluate the model, the original dataset (n = 300) was randomized and then stratified by gender. This created a training set (80%, n = 240; 144 with strict LBBB, 96 without) and a validation set (20%, n = 60; 30 with strict LBBB, 30 without). The test dataset (n = 302; 156 with strict LBBB, 146 without) remained unchanged for independent assessment.

Two models were trained and tested using distinct inputs: a feature-based ML model (Random Forest, RF) and an end-to-end DL model (Residual Network, ResNet). For the RF model, all ECG features were automatically extracted by the Philips DXL algorithm. A total of 17 relevant features, including QRS duration, LVH voltage, max spatial angle of QRST, QS configuration in V1 and V2, R wave duration in V6, I, V5, and AVL, R peak location in V6, V5, and AVL, S and R area of V6, global R peak location of V6, and Notch-Slur-Score and Notch-Slur-Sum-Score, were automatically selected based on feature importance using a RF classifier [2], [10]. For the end-to-end ResNet model, a 10-second 12-lead ECG (RawECG) signal and a 1.2-second average representative beat (RepBeat) signal, both sampled at 1 kHz, were used as inputs. The RawECG signal provides a comprehensive view of the ECG waveform, while the RepBeat signal, generated by the Philips DXL algorithm, offers a summarized view of key ECG features for diagnostics.

2.2. Data Augmentation and Automation

To enhance its robustness and performance, the ResNet model was trained using a random combination of data augmentation techniques through a modified RandAugment automation method [4], [5], [11]. In the training phase, three data augmentation techniques such as cutout, dropout, and scaling were randomly applied using the RandAugment method, with a probability of 0.333 assigned to each technique. Figure 1 illustrates the three chosen data augmentation techniques, showcasing their distinct effects on the ECG signal. In detail, cutout was used to zero out a random segment of the signal, with the segment length being within 10% of the total signal length, at a random time point, across all leads. Dropout involved randomly setting 5% of signal values to zero for all leads. Scaling was implemented by randomly compressing or stretching the amplitude of the signal using a random factor for all leads.

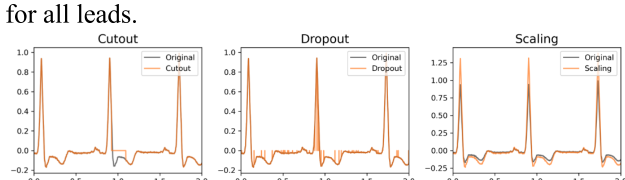

Figure 1. Data augmentation techniques incorporating cutout, dropout, and scaling applied to lead II.

2.3. Algorithm and Model Architecture

In this study, we embarked on a multifaceted approach to accurately detect LBBB from 12-lead ECGs, exploring rule-based, ML, and DL techniques. The first approach utilizes a rule-based algorithm, while the second employs a RF classifier with 17 extracted features for feature-based ML. The third approach utilizes a ResNet architecture, with RawECG and RepBeat as inputs for end-to-end DL.

The rule-based algorithm adheres to specific LBBB criteria, encompassing considerations such as QRS duration, lead V1 morphology, and the presence of mid-QRS notching or slurring in specific leads [2], [10]. The RF classifier uses 100 trees and the Gini impurity criterion, employing custom-extracted features for feature-based ML [2], [10]. Figure 2 illuminates the architecture of the ResNet model, specifically designed to harness the power of DL for accurate LBBB identification. Its structure includes a one-dimensional (1D) convolutional layer followed by four residual blocks. Subsequent to the convolutional layer, there is the application of batch normalization (BatchNorm), Exponential Linear Unit (ELU) activation function with an alpha parameter set to 1.0, dropout (Dropout) with a probability of 0.2, and adaptive max-pooling (MaxPool1D). The model's output layer performs binary classification, feeding into a fully connected (dense) layer with a 0.2 dropout probability and a sigmoid activation function. This architecture is designed to differentiate strict-LBBB from non-LBBB using a sigmoid activation for probability output. The ResNet model uses cross-entropy loss for LBBB detection. Python libraries PyTorch, Scikit-Learn, SciPy, and Waveform Database (WFDB) powered the implementation of preprocessing, model construction, training, validation, testing, data augmentation, and workflow automation.

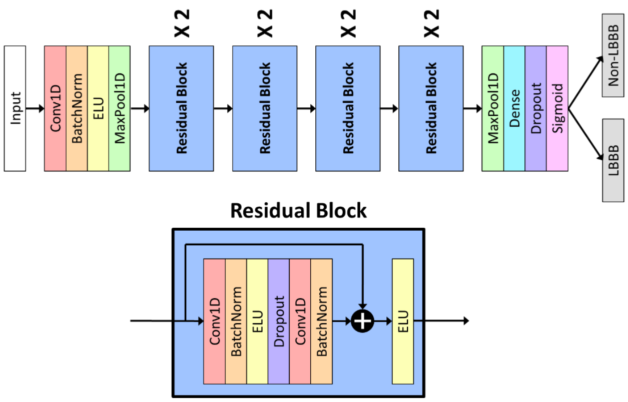

Figure 2. The proposed architecture of the end-to-end ResNet18-based deep learning model.

2.4. Training and Performance Evaluation

The ResNet model was trained for 100 epochs with a batch size of 16, utilizing the Adam optimizer with an initial learning rate of 0.0001 and a scheduled decay of 10% every 10 epochs. The model exhibiting the highest performance during training was then selected and employed for testing. Performance metrics for LBBB detection such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and accuracy were computed.

2.5. Explainability

While end-to-end DL models often achieve superior performance compared to traditional feature-based ML models, their complex architectures and non-linear decision-making processes can make them challenging for humans to understand [4]. This lack of transparency has led to their characterization 'black boxes'. To understand our ResNet model's inner workings and highlight the ECG features driving its LBBB detection, we used SHAP [8], a powerful interpretability technique that estimates each input signal's contribution to the model's output. In the context of ECG classification using a ResNet model, SHAP facilitates understanding which specific parts/segments or patterns within the ECG signals are most influential in the model's classification decisions.

3. Results

The experiments were conducted on a high-powered desktop with an Intel Core i9-12900K processor (3.5 GHz), 64 GB RAM, and a NVIDIA GeForce RTX 3080 Ti GPU for accelerated ML and DL model processing capabilities. Table 1 presents a comprehensive comparison of the performance metrics, including sensitivity, specificity, PPV, NPV, F1 score, and accuracy, for rule-based algorithm, feature-based RF, and end-to-end ResNet with RepBeat and RawECG in the context of LBBB detection. The ResNet model trained on RawECG input data achieved the highest performance in LBBB detection, with an accuracy of 0.931, sensitivity of 0.955, specificity of 0.849, PPV of 0.871, NPV of 0.947, and F1 score of 0.911. This represents a relative improvement of 20.60%, 12.03%, and 7.75% in accuracy compared to the rule-based algorithm (0.772), feature-based RF model (0.831), and ResNet model trained on RepBeat input data (0.864), respectively. These results indicate that the end-to-end DL model significantly outperforms both the rule-based algorithm and the feature-based ML model in LBBB detection. ResNet trained on RawECG outperformed RepBeat, indicating that longer ECG segments capture more clinically relevant information, like dynamic changes in cardiac electrical activity.

Table 1. Performance comparison of rule-based algorithm, feature-based ML, and end-to-end DL for LBBB detection.

Algorithm / Model	Sens.	Spec.	PPV	NPV	F1	Acc.
Rule-based algorithm	0.891	0.644	0.728	0.847	0.801	**0.772**
RF with features	0.878	0.781	0.811	0.857	0.843	**0.831**
ResNet with RepBeat	0.910	0.815	0.840	0.895	0.874	**0.864**
ResNet with RawECG	0.955	0.849	0.871	0.947	0.911	**0.931**

Using SHAP, we identified the specific morphological aspects of the ECG signals that the ResNet model relies on for LBBB detection. Figure 3 shows the representative example of feature importance based on SHAP value by using the end-to-end ResNet model with RepBeat input. The SHAP results for the RawECG input are not included due to word limitations, as LBBB is a morphology-based abnormality, making RepBeat input more appropriate. SHAP values were min-max normalized to a range of 0 to 1. The figure utilizes a color-coded heatmap to show the importance of different segments/areas superimposed on the ECG. Red indicates high importance (value of 1), while blue indicates low importance (value of 0). Larger dots represent higher importance. Furthermore, Figure 3 only displays the most impactful ECG leads, automatically selected based on their SHAP values. This figure demonstrates that SHAP analysis of the ResNet model highlighted several ECG features critical for LBBB detection. These features include a broad R wave in leads I and aVL, a QS pattern in leads V1 and V2, and a QRS notch and broad R wave in leads V5 and V6, all of which align with established LBBB diagnostic criteria.

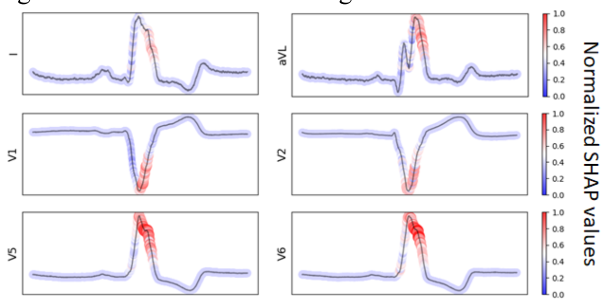

Figure 3. SHAP-based feature importance with RepBeat input for explainability.

4. Discussion and Future Work

We developed and evaluated three approaches for detecting LBBB from 12-lead ECGs: a rule-based algorithm, feature-based ML model, and an end-to-end DL model. Our experiments demonstrated that the end-to-end DL model outperformed both the rule-based algorithm and the feature-based ML approach. Additionally, we utilized AI explainability technique (i.e., SHAP) to highlight significant segments and patterns within the ECG signals, thereby aiding clinicians in making better-informed diagnoses and decisions.

This study has some limitations and presents opportunities for future work. First, the dataset used in this study was recorded from a single hospital. It is necessary to test the model on datasets from different hospitals located in various regions and countries. Second, other explainable AI (XAI) methods, such as Grad-CAM, LIME, and additional advanced techniques, should be explored and compared. Third, both qualitative and quantitative approaches should be employed to comprehensively evaluate the performance of the XAI methods. Future work will involve validating our end-to-end DL model with data from other hospitals, assessing model performance by age, sex, and ischemic/non-ischemic cardiomyopathy, and determining if combining this model with other clinical and ECG predictors can help clinicians identify heart failure patients who may benefit from CRT.

Acknowledgments

The authors thank Ting Yang, Abhinay Pandya, Wei Zong, Eric Helfenbein, Joost Mans, Milan Petkovic, Dieter Haase, and Mark Konrad for their support.

References

[1] S. Calle, F. Timmermans, and J. De Pooter, “Defining left bundle branch block according to the new 2021 European Society of Cardiology criteria,” *Neth. Heart J.*, vol. 30, no. 11, pp. 495–498, 2022.

[2] T. Yang, R. E. Gregg, and S. Babaeizadeh, “Detection of strict left bundle branch block by neural network and a method to test detection consistency,” *Physiol. Meas.*, vol. 41, no. 2, p. 025005, 2020.

[3] A. Y. Hannun *et al.*, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nat. Med.*, vol. 25, no. 1, pp. 65–69, 2019.

[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] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘ Why should i trust you?’ Explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

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

[9] W. Zareba *et al.*, “Effectiveness of cardiac resynchronization therapy by QRS morphology in the multicenter automatic defibrillator implantation trial–cardiac resynchronization therapy (MADIT-CRT),” *Circulation*, vol. 123, no. 10, pp. 1061–1072, 2011.

[10] T. Yang, R. E. Gregg, and S. Babaeizadeh, “Detection of Strict Left Bundle Branch Block Using Random Forest and Deep Neural Network,” *J. Electrocardiol.*, vol. 57, p. S105, 2019.

[11] 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