Scientific Reports

https://doi.org/10.1038/s41598-025-34911-6

Article in Press

Prediction of left ventricular systolic dysfunction in left bundle branch block using a fine-tuned ECG foundation model

Received: 20 March 2025

Accepted: 31 December 2025

Published online: 17 January 2026

Cite this article as: Kim D.H., Bok Y., Lee S.H. *et al.* Prediction of left ventricular systolic dysfunction in left bundle branch block using a fine-tuned ECG foundation model. *Sci Rep* (2025). https://doi.org/10.1038/s41598-025-34911-6

Do Heon Kim, Youngnam Bok, Sun Hwa Lee, Jihye Heo, Seung Park, Dongheon Lee & Jae-Hyeong Park

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Prediction of Left Ventricular Systolic Dysfunction in Left Bundle Branch Block using a Fine-Tuned ECG Foundation Model

Do Heon Kim[1], Youngnam Bok, MD[2], Sun Hwa Lee, MD, PhD[3,4], Jihye Heo, MS[5], Seung Park, PhD[7], Dongheon Lee, PhD[1,5,6,†] and Jae-Hyeong Park, MD, Ph.D[1,2,†]

[1]Chungnam National University College of Medicine, Daejeon, Republic of Korea

[2]Department of Cardiology in Internal Medicine, Chungnam National University Hospital, Daejeon, Republic of Korea

[3]Division of Cardiology, Department of Internal Medicine, Jeonbuk National University Medical School, Jeonju, Republic of Korea

[4]Research Institute of Clinical Medicine of Jeonbuk National University-Biomedical Research Institute of Jeonbuk National University Hospital, Jeonju, Republic of Korea

[5]Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[6]Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea

[7]Chungbuk National University College of Medicine, Cheongju, Republic of Korea

ARTICLE IN PRESS

ARTICLE IN PRESS

Short running title: Fine-tuned ECG Foundation Model for LV Dysfunction in LBBB

†These authors contributed equally to this work.

*Co-corresponding author

Dongheon Lee, Ph.D

Department of Radiology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea

Institute of Medical and Biological Engineering, Seoul National University Medical Research Center, Seoul, Republic of Korea

Chungnam National University College of Medicine, Daejeon, Republic of Korea

E-mail: dhlee13@snu.ac.kr

Jae-Hyeong Park, MD, Ph.D

Department of Cardiology in Internal Medicine, Chungnam National University Hospital, Daejeon, Republic of Korea

Department of Biomedical Engineering, Chungnam National University College of Medicine, Chungnam National University Hospital, Daejeon, Korea

E-mail: jaehpark@cnu.ac.kr

ARTICLE IN PRESS

ARTICLE IN PRESS

Abstract

Left bundle branch block (LBBB) is an important electrocardiographic (ECG) finding strongly associated with left ventricular systolic dysfunction (LVSD), a condition linked to poor clinical outcomes. Although early LVSD detection is crucial, standard diagnosis via echocardiography may not always be immediately accessible. In this study, we propose a fine-tuned ECG foundation model (FM) to enhance LVSD detection specifically in patients with LBBB. We conducted a retrospective multicenter analysis of 2,031 paired ECG-echocardiographic datasets from 892 LBBB patients. The ECG-FM was fine-tuned for optimal LVSD prediction and compared against baseline models, which were conventional deep learning methods, including Fully Convolutional Network (FCN), LSTM-FCN, ResNet, and InceptionTime. The proposed ECG-FM with single-step full fine-tuning outperformed baseline models, achieving accuracy, sensitivity, and AUROC of 0.758, 0.771, and 0.807, respectively. Additionally, sequential partial fine-tuning exhibited the highest sensitivity (0.787), enhancing screening capability. DeepLIFT analysis identified QRS complex and T wave features in leads V1–V4 as critical predictive factors. Our results demonstrated that the recommended fine-tuned ECG-FM significantly improves LBBB patient LVSD detection, potentially enabling earlier clinical diagnosis in such cases when echocardiography is not readily available, thereby potentially improving patient outcomes and clinical management.

ARTICLE IN PRESS

ARTICLE IN PRESS

Keywords: Ventricular Dysfunction, Left Bundle-Branch Block, Electrocardiography, ECG-Foundation Model

ARTICLE IN PRESS

ARTICLE IN PRESS

I. Introduction

Left bundle branch block (LBBB) is an important electrocardiographic finding that arises from impaired conduction through the left branch of the cardiac conduction system.[1] Since its disturbance interferes with normal ventricular conduction, it results in asynchronous ventricular activation that diminishes left ventricular efficiency and can result in systolic dysfunction.[2] Most commonly associated with ischemic heart disease, hypertension, cardiomyopathies, or valvular dysfunctions, LBBB often indicates the presence of an underlying structural heart disorder.[3-5] Of particular concern among these is left ventricular systolic dysfunction (LVSD), which is closely associated with LBBB[6] as its relevance is known to adversely affect clinical outcomes.[7] So, the early detection of LVSD is crucial for implementing appropriate management and improving prognosis.

For diagnosing LVSD, typically defined as a left ventricular ejection fraction (LVEF) below 40%, echocardiographic evaluation is the gold standard.[8] In heart failure patients, the development of LBBB is clinically important because it could exacerbate LVSD, and severe LVSD becomes one of the major indications to consider cardiac resynchronization therapy (CRT).[9] For that, routine echocardiographic follow-up is advised even in patients with newly diagnosed LBBB who do not initially exhibit LVSD.[10] Although echocardiography is expensive, requires specialist imaging expertise, and is thus not easy to perform routinely in primary care or emergency practice.

ARTICLE IN PRESS

ARTICLE IN PRESS

Compared to echocardiography, electrocardiogram (ECG) is more accessible and cost-effective. Accurate ECG-based differentiation of patients with LVSD could facilitate early screening and reduce unnecessary echocardiographic tests.[11] Among conventional ECG morphological criteria, only QRS duration has demonstrated a strong correlation with LVSD[12]; however, reliance solely on visually identifiable morphologic markers may overlook subtle or nonspecific findings, limiting diagnostic accuracy.[13]

Over the last few years, deep learning models have been developed to predict LVSD from ECG data.[14-17] Attia et al.[15] proposed a convolutional neural network (CNN) that was able to predict LVSD from a 10-second 12-lead ECG in heterogenous populations, including specific disease subgroups and COVID-19 patients, with high robustness and demonstrating good generalizability.[18-20] Similarly, Kwon et al.[21] proposed a deep learning model based on a multi-layer perceptron (MLP) utilizing manually acquired features from raw ECG signal. This was further improved by adding a residual neural network architecture, externally validated in patients having atrial fibrillation with rapid ventricular responses.[22] These did not, however, study prediction of LVSD in LBBB patients particularly. Therefore, there remains a clinical need for specialized predictive models tailored explicitly for LVSD detection in the LBBB population.

In this study, we propose a deep learning-based model to predict LVSD using ECG, specifically tailored for patients with LBBB. While most existing models have been trained on general populations, our approach

ARTICLE IN PRESS

fine-tunes a foundation model (ECG-FM) pre-trained on large-scale public ECG datasets to better capture the unique characteristics of this high-risk subgroup. We applied and compared various fine-tuning strategies to overcome the limitations of a small patient dataset and demonstrated the generalizability of the model using multi-center data. The design not only enhances predictive performance compared to conventional approaches that train models from scratch, but also improves clinical interpretability, as we visualized the model's focus areas on ECG signals using DeepLIFT.

ARTICLE IN PRESS

II. Methods

Study Design

This study proposes a deep learning-based model to predict LVSD (LVEF < 40%) in patients with LBBB using ECG data. LBBB was diagnosed using conventional LBBB criteria based on electrocardiograms measured by commercially available ECG machines. These diagnostic criteria included QRS duration greater than 120ms; the presence or absence of Q wave in lateral leads (I, aVL, V5, and V6); the morphology of R waves in lateral leads, morphology of QRS complex in right precordial leads including dominant S wave, QS or rS pattern; and the R wave peak time in V5/V6.[23] Conventional deep learning models for time-series data served as baseline models. These included FCN, LSTM-FCN, ResNet, and InceptionTime. Their performance was then compared with a fine-tuned version of ECG-FM[24], a publicly available foundation model pre-trained on large-scale ECG

ARTICLE IN PRESS

datasets. The performance of the ECG-FM backbone was evaluated by comparing single-step fine-tuning and sequential fine-tuning approaches (Figure 1b).

We used paired ECG-echocardiographic datasets of patients with LBBB, each dataset including 12-lead ECG and the associated echocardiographic test conducted within a one-month interval. Our datasets comprised 2,031 paired ECG-echocardiographic tests from three institutions: 1,657 datasets from 691 patients in Chungnam National University Hospital (CNUH), 259 datasets from 91 patients in Chungnam National University Sejong Hospital (CNUSH), and 115 datasets from 115 patients in Jeonbuk National University Hospital (JNUH). After accounting for five patients present in both CNUH and CNUSH datasets, the final cohort consisted of 892 unique patients.

This multicenter retrospective study was approved by the Institutional Review Boards (IRBs) of Chungnam National University Hospital (CNUH, Daejeon, Korea, CNUH IRB 2025-02-001) and Jeonbuk National University Hospital (JNUH, Jeonju, Korea, JNUH IRB 2025-02-030). All procedures were performed in accordance with relevant guidelines and regulations.

ARTICLE IN PRESS

Data preparation

We divided the dataset into four distinct subsets: training, tuning, internal validation, and external validation. Each dataset was a paired ECG-echocardiogram study. The training set comprised datasets from CNUH collected up to 2023 (n=1,200), while the tuning set included additional CNUH datasets from the same period (n=300). The internal validation set

ARTICLE IN PRESS

consisted of datasets collected from CNUH in 2024 (n=194). For external validation, we used two separate datasets: one from JNUH (n=115, recruited between 2008 and 2021) and another from CNUSH (n=222, recruited between 2022 and 2024). The sets were mutually exclusive to avoid data leakage. Table 1 shows comprehensive descriptions. No subgroup analyses (e.g., on age, sex, or institution) were planned in the present study.

Preprocessing

The ECG data used in this study consisted of two types. The first type included ECGs originally stored as digital images in electronic medical records, later converted to XML format (Mediv Co., Cheongju, Korea). The conversion was performed using a software system from Medical User Software Exchange (MUSE; General Electric Healthcare, Waukesha, WI) and an ECG management system (Medical Information System, Mediana Co., Ltd., Wonju, Korea). Samples were stored at various sampling rates, ranging from 99 Hz to 221 Hz. The second type comprised digitally acquired ECGs collected prospectively using commercial ECG machines (MAC5000 v1.0; General Electric Healthcare, Waukesha, WI) and stored directly in XML format within the MUSE system. For these samples, 12-lead resting ECG data were all recorded at a sampling rate of 500 Hz for 10 seconds. The detailed preprocessing is described in Supplementary Materials.

ARTICLE IN PRESS

ARTICLE IN PRESS

We converted both datasets into a standardized format of four channels (Figure 1a). There are 5,000 time points in each channel (500 Hz × 10 seconds). The converted final dataset consisted of three channels from a standard 12-lead ECG and a rhythm strip (lead II) that was recorded for 10 seconds. All ECG data underwent preprocessing with a high-pass Butterworth filter (0.5 Hz) and powerline noise removal (50 Hz). The detailed transformation and integration process for each ECG data type is presented in Supplementary Fig. S1.

Baseline Models for LVSD Prediction

The baseline models used in this study for predicting LVSD included Fully Convolutional Network (FCN)[25], Long Short-Term Memory Fully Convolutional Network (LSTM-FCN)[26], ResNet[25], and InceptionTime[27]. All baseline models were trained from scratch without pre-training. LVSD prediction was formulated as a binary classification task distinguishing patients based on an LVEF cutoff of 40%.

First, the FCN[25] model comprised three 1D convolutional blocks without local pooling layers to preserve original time-series length. The filter sizes of each convolutional block were 128, 256, and 128, in that order, and the kernel sizes were 8, 5, and 3, respectively. ReLU was used as the activation function in all three blocks. The convolutional architecture considers inter-lead spatial ECG characteristics but neglects long-range temporal dependencies. Second, the LSTM-FCN[26] paired a Long Short-Term Memory (LSTM) layer with the FCN architecture to learn temporal dynamics, overcoming the weakness of FCN to learn only spatial

ARTICLE IN PRESS

features. The number of convolutional layers, kernel sizes, and filters was identical to that used in the FCN. Third, the ResNet[25] model comprised three residual blocks. The initial two blocks consisted of four layers of Conv1D each, and the third block consisted of three Conv1D layers, totaling 11 layers within the network. Last, InceptionTime[27] utilized multi-scale pattern detection through iterative Inception modules with three convolutional filters of kernel sizes 10, 20, and 40. The module was repeated six times (depth=6), and two residual connections were included across the entire network to help stabilize training and capture ECG features at multiple temporal scales.

Fine-tuned ECG-FM for LVSD Prediction

In this study, a pre-trained ECG foundation model (ECG-FM) was fine-tuned using ECG data from LBBB patients to improve LVSD prediction. The ECG-FM effectively captured general ECG representations before being adapted specifically for LBBB. The pre-training method of ECG-FM, initially proposed by Oh et al.[28], is a lead-agnostic self-supervised approach that uses random lead masking and contrastive learning[29], combining CNN, transformer architecture, and temporal average pooling to capture latent ECG representations (Figure 2). McKeen et al.[24] adopted the technique proposed by Oh et al. and developed ECG-FM by pre-training it. The pre-training was conducted in two stages: first, the model was pre-trained on approximately 1.4 million ECG recordings (12-lead, 5 seconds each, sampled at 500 Hz). Second, this pre-trained model was further

ARTICLE IN PRESS

trained through multi-label diagnostic classification using general population ECG data from the PhysioNet 2021 dataset. It should be noted that all four linearly dependent leads (aVR, aVL, aVF, III) were retained during the pre-training phase without exclusion.

We performed the downstream task of predicting LVSD (LVEF < 40%) in a cohort of LBBB patients through single-step fine-tuning of the first pre-trained ECG-FM (Figure 1b). Due to differences in data format between our ECG inputs and the original backbone model, we modified the architecture accordingly. Our ECG data consisted of four channels - either four distinct leads recorded for 2.5 seconds each, or a single lead recorded continuously for 10 seconds (as a rhythm strip). To accommodate this, we added a projection layer using a 1D convolution at the front of the pre-trained ECG-FM and appended a linear classification head at the end, enabling binary classification for LVSD (Figure 2).

Drawing upon prior research[30], we employed three single-step fine-tuning methods: full fine-tuning, partial fine-tuning, and additive fine-tuning. Full fine-tuning updated all model parameters, including the added projection layer and classification head. Partial fine-tuning consisted of freezing some of the pre-trained ECG-FM layers and updating the others. More specifically, we froze the top 10, 8, 6, 4, and 2 transformer layers (out of 12) and generated five different partial fine-tuned models. This stepwise approach was based on the understanding that lower layers tend to capture general ECG representations, whereas upper layers encode more task-specific features; thus, selectively updating higher layers allows the model to retain broadly useful physiological representations while

ARTICLE IN PRESS

introducing only the minimal adaptation required for the downstream task. Additive fine-tuning froze all pre-trained ECG-FM layers but only updated the added projection layer and classification head. Thus, seven distinct models were evaluated through single-step fine-tuning.

For sequential fine-tuning, we utilized a second ECG-FM model previously fine-tuned for multilabel classification on other ECG datasets. Following the same approach as single-step fine-tuning, we added a projection layer at the input and a classification head at the output. Sequential fine-tuning was also performed using full, partial, and additive fine-tuning methods, producing another set of seven distinct models.

ARTICLE IN PRESS

Explainability using Deep Learning Important Features (DeepLIFT)

We applied DeepLIFT[31] to enhance the interpretability of ECG-based predictions. DeepLIFT is a backpropagation-based attribution method that quantifies feature importance by explaining the difference between the model's actual output and a reference output in relation to differences between actual inputs and reference inputs. This approach enables quantitative analysis of ECG signal features identified by the deep learning model and facilitates visualization of contributions from specific ECG segments, such as the P wave, QRS complex, and T wave, to the model's prediction.

DeepLIFT calculates contributions according to the following equation:

$$\sum_{i=1}^{n} C_{\Delta x_i \Delta t} = (1)$$

ARTICLE IN PRESS

where Δx_i denotes the difference between the input and its reference value, Δt represents the change in the output relative to the reference output, and $C_{\Delta x_i \Delta t}$ quantifies the contribution of each input feature x_i to the output difference. In this study, zero scalar values corresponding to each input tensor were used as reference inputs.

Implementation Details

The baseline models in this study were trained from scratch for 100 epochs with a learning rate of 5e-6. This setting was determined based on the observation that training accuracy had converged, and validation accuracy had reached its maximum around 100 epochs. The same approach was applied for fine-tuned ECG-FM models, which were trained for 60 epochs and had learning rates between 1e-3 and 5e-7, depending on the model. For every experiment, we employed the Adam optimizer and cross-entropy loss function.

ARTICLE IN PRESS

All calculations were carried out with Python (version 3.10). ECG preprocessing was carried out using the Neurokit2 toolbox. PyTorch (version 2.2) and sktime libraries were used to implement models, which were trained on an NVIDIA RTX A6000 GPU running CUDA version 12.3. DeepLIFT analysis for model interpretation was performed using the Captum library. The source code for this study is publicly available at: https://github.com/doheon114/ECG-FM-LVSD.git.

Evaluation Metrics

ARTICLE IN PRESS

We measured the predictive performance of the models using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1-score, area under the receiver operating characteristic curve (AUROC), and area under the receiver precision-recall curve (AUPRC) for one internal and two external validations. The performance metrics were calculated using macro-averages.

Statistical Analysis

In this study, statistical analyses were performed to assess variations in dataset characteristics and model performance. The normality assumption for parametric tests was assessed using the Shapiro-Wilk test, and homogeneity of variances was evaluated using Levene's test. To compare mean values of continuous variables between datasets, one-way ANOVA or Welch's ANOVA was performed depending on the results of the homogeneity test. Differences in categorical variables were evaluated using chi-square analysis. For comparison of model performance, both repeated-measures ANOVA and the Friedman test were applied to detect differences among all models. Post-hoc single model comparisons were conducted using Wilcoxon signed-rank tests with Holm's step-down correction. Statistical significance was defined at a level of 0.05. All analyses were performed using Python (version 3.10) and the statsmodels package (version 0.14.4).

ARTICLE IN PRESS

ARTICLE IN PRESS

III. Results

Comparative Performance of Models for LVSD Prediction

Table 2 summarizes the comparative performance of the baseline models (FCN, LSTM-FCN, ResNet, and InceptionTime) and ECG-FM models trained by single-step full fine-tuning (SS-FF) and sequential partial fine-tuning (Sq-PF). ECG-FM-based methods showed superior performance compared with baseline models across the majority of metrics for both internal and external validation sets. Specifically, the ECG-FM-based SS-FF model achieved the best overall performance, with an average accuracy of 0.758, specificity of 0.851, PPV of 0.651, F1-score of 0.703, AUROC of 0.807, and AUPRC of 0.678, surpassing the maximum performance values of all baseline models for these metrics. On the other hand, Sq-PF model derived using ECG-FM displayed the highest sensitivity (average = 0.787), crucial for effective LVSD screening, and highest NPV (average = 0.855).

Repeated-measures ANOVA indicated statistically significant performance differences among the models across the eight performance metrics ($F = 16.67$, $p < 0.001$). This finding was further confirmed by the Friedman test ($p < 0.001$). Post-hoc analysis demonstrated significant performance differences between the ECG-FM-based methods and baseline models. However, there were no statistically significant differences between the individual baseline models themselves (Figure 3).

Table 3 compares the accuracy and AUROC performance of the baseline model (InceptionTime) with consistent performance across metrics and the ECG-FM-based SS-FF method with overall best performance. The

ARTICLE IN PRESS

model trained from scratch showed the lowest overall performance, with a mean accuracy of 0.622 and an AUROC of 0.677, falling below the performance values of the baseline InceptionTime model, which achieved an accuracy of 0.691 and an AUROC of 0.746. This highlights that fine-tuning pre-trained ECG-FM weights—rather than altering the model architecture—significantly improved model performance.

Comparative Performance of ECG-FM-Based Models

We compared the performance of ECG-FM models trained using single-step fine-tuning (SS) and sequential fine-tuning (Sq) methods, varying the number of trainable (unfrozen) layers. For single-step fine-tuning, seven models were developed using additive fine-tuning (SS-AF), partial fine-tuning (SS-PF), and full fine-tuning (SS-FF). The SS-AF method, which updated parameters only in the projection layer and classification head while freezing all CNN and transformer encoder layers, achieved an F1-score of 0.598. In contrast, the SS-FF method, which allowed all layers to be trainable, reached a higher F1-score of 0.703. Partial fine-tuning methods, in which only some transformer encoder layers were trainable, showed intermediate performance between SS-AF and SS-FF (Supplementary Table S1). However, statistical analyses indicated no significant differences among the seven SS models based on repeated-measures ANOVA ($p=0.708$) and the Friedman test ($p=0.282$).

For sequential fine-tuning, we similarly developed seven models with additive fine-tuning (Sq-AF), partial fine-tuning (Sq-PF), and full fine-tuning (Sq-FF). Unlike single-step fine-tuning, sequential fine-tuning did

ARTICLE IN PRESS

not show a linear relationship between performance and the number of trainable layers. Notably, the Sq-PF model (8 trainable layers) outperformed the Sq-FF model (all 12 layers trainable) in terms of sensitivity, NPV, F1-score, AUROC, and AUPRC (Supplementary Table S2). Statistical analysis demonstrated significant differences among the sequential fine-tuning methods, and post-hoc tests revealed that the Sq-PF model (8 layers trainable) significantly differed from three of the other six models (Supplementary Figure S2).

Analysis of Explainability using DeepLIFT

Figure 4 illustrates an example of local explanations for LVSD classification results obtained after tuning the ECG-FM using the SS-FF method. The DeepLIFT analysis showed that, across most leads, the QRS complex contributed most significantly to the model's prediction, particularly in leads V1–V4, which reflect the septum and anterior wall of the left ventricle (LV). A notably strong contribution from the deep S wave was identified. Additionally, T wave contributions were prominent in leads V1–V4, indicating that ventricular repolarization characteristics (including shape and magnitude of the T wave) were important for LVSD prediction. Limb leads (I, II, III, aVR, aVL, aVF) occasionally showed significant QRS contributions, but the precordial leads (V1–V6), particularly V1–V4, consistently provided the highest contribution and thus had greater predictive importance.

ARTICLE IN PRESS

ARTICLE IN PRESS

IV. Discussion

Early detection of LVSD in patients with LBBB is crucial for improving clinical outcomes, but current deep learning models for predicting LVSD are not specifically designed for LBBB patients. In this study, we proposed a fine-tuned ECG foundation model (ECG-FM) for predicting LVSD in patients with LBBB. Unlike previous approaches that train models from scratch,[15,21] our method fine-tunes a publicly available ECG-FM pre-trained on large-scale ECG datasets using LBBB patient data, and we systematically compared their performance against conventional baseline models (Supplementary Table S3). Among the various fine-tuning strategies, the single-step full fine-tuning (SS-FF) method demonstrated the best overall performance, and sequential partial fine-tuning (Sq-PF) achieved the highest sensitivity. The superior overall performance of SS-FF likely reflects the benefit of fully adapting the pretrained ECG-FM representations to the downstream task, enabling the model to optimize across most evaluation metrics. In contrast, the higher sensitivity observed with Sq-PF suggests that freezing some of the layers allows the model to better preserve salient low-level ECG features relevant for detecting positive cases, even if this comes at the cost of reduced performance in other metrics.

By analyzing different layer-freezing strategies during fine-tuning, we identified optimal settings and confirmed that fine-tuning the ECG-FM significantly outperformed baseline models trained from scratch. It is important to note that the performance gain does not simply arise from

ARTICLE IN PRESS

architectural differences. Rather, the advantage primarily stems from the representation learning achieved during the large-scale pre-training stage, which enables the model to capture rich morphological and temporal patterns in ECG signals. Fine-tuning these pretrained representations for LVSD classification proved more effective than training models from scratch, ultimately leading to superior performance and robustness across multi-center validation cohorts. Important ECG regions associated with LVSD predictions were then visualized, enhancing interpretability.

Although several general ECG-based LVSD prediction models exist, this is the first study, to our knowledge, specifically developing an LVSD prediction model for LBBB patients. This specialized model can support personalized clinical decisions and early screening, offering timely opportunities for appropriate management.[32] In particular, the improved sensitivity is crucial for screening, as it maximizes the detection of true LVSD cases and minimizes missed diagnoses. Coupled with the enhanced AUROC, which reflects better overall discriminative ability, these improvements highlight the model's clinical relevance by enabling earlier identification and intervention in at-risk LBBB patients. While the proposed ECG-FM-based SS-FF model showed relatively low PPV (0.651), previous evidence indicates that patients with false-positive ECG predictions still exhibit a higher incidence of subsequent LVSD[15], suggesting clinical utility superior to routine echocardiography screening alone.

The explainability analysis revealed that significant predictive features were predominantly associated with the QRS complex and T wave, aligning

ARTICLE IN PRESS

with previous research findings.[12,33] The prominence of ECG signals from the septum and anterior LV wall in model predictions suggests these features play a critical role in LVSD onset. Although we applied various augmentation and oversampling techniques to mitigate overfitting from limited training data, their contribution to improved performance was minimal. Instead, fine-tuning an ECG-FM pre-trained on extensive ECG data consistently proved more effective than training conventional deep-learning models from scratch, especially for small, disease-specific datasets.

In addition, although performance on external validation datasets typically did not surpass that of internal validation, the proposed model showed better performance on the second external validation set compared to the internal set. This result likely occurred because the second external validation set exhibited a bimodal LVEF distribution around the 40% threshold, facilitating clearer classification (Supplementary Figure S2). Conversely, the internal validation set contained a larger proportion of patients with LVEF values between 40% and 50%, defined by the European Society of Cardiology (ESC) as heart failure with mildly reduced ejection fraction (HFmrEF).[8] Patients with HFmrEF often fall within a clinical "gray zone," where mild systolic dysfunction may or may not be present[34], complicating accurate classification by ECG-FM-based models.

This research has several limitations. First, the performance evaluation was constrained by differences in ECG data formats. To utilize the ECG-FM model, ECG data had to be converted from the 4-channel format used

ARTICLE IN PRESS

in this study to a 12-channel format. Because data format can significantly influence model performance, the results observed here may not generalize to ECGs in other formats. Future studies should systematically investigate different configurations while maintaining the 12-channel ECG format. Second, variations in sampling frequency posed a limitation. Some signals were up-sampled to 500Hz via linear interpolation (Supplementary Figure S1). Although this approach appears to preserve key waveforms, the impact of different sampling rates on model performance was not quantitatively evaluated, leaving uncertainty about the model's robustness to signals recorded at other frequencies.Third, patients with LBBB included in this study might have other cardiovascular conditions affecting LVSD prediction. Even though these comorbidities and histories of medication are helpful in defining the population and determining clinical utility for the model, these data were not consistently available, limiting the ability to fully assess the model's generalizability across heterogeneous populations. Nevertheless, since LBBB inherently associates with various cardiovascular diseases, imposing strict exclusion criteria for the existence of other LVSD-related illness might preclude the development of disease-specific models. Fourth, we used conventional criteria for LBBB detection. If we used strict criteria for detecting LBBB, it can detect LVSD more precisely and the results may be changed with more precise diagnostic power.[35] However, we aimed to perform this study to check the presence of LVSD when LBBB is diagnosed using currently available ECG machines commonly used in the usual clinical practice. Thus, the use of conventional criteria in the detection of LBBB may be

ARTICLE IN PRESS

appropriate in this study. Lastly, although DeepLIFT analysis was applied to provide local explanation for each sample, the direct relationship between specific ECG features and model predictions was not fully elucidated. This highlights a limitation of DeepLIFT in that its interpretability is primarily local rather than global, especially in a relatively small dataset. Nevertheless, these local insights can still support clinical understanding by providing clinicians with interpretative evidence of which ECG regions may contribute to LVSD predictions. Future research with larger datasets should aim to systematically investigate the influence of specific ECG characteristics (e.g., QRS complex, T wave) and perform global interpretability analyses.

In conclusion, this study proposes an ECG foundation model (FM)-based approach to predict LVSD in patients with LBBB. Compared to deep learning-based baseline and the scratch-trained models, the proposed model demonstrates significant performance improvement through fine-tuning the ECG-FM. This foundation model-based approach is expected to serve as a broadly applicable strategy for developing disease-specific predictive models across various clinical domains. Future research should focus on prospective validation, integration of the model into clinical workflows, and further investigate methods to enhance the interpretability of model predictions.

ARTICLE IN PRESS

References

1 Pujol-López, M., Tolosana, J. M., Upadhyay, G. A., Mont, L. & Tung, R. Left bundle branch block: characterization, definitions, and recent

ARTICLE IN PRESS

insights into conduction system physiology. *Cardiac Electrophysiology Clinics* **13**, 671-684 (2021).

2 Smiseth, O. A. & Aalen, J. M. Mechanism of harm from left bundle branch block. *Trends in cardiovascular medicine* **29**, 335-342 (2019).

3 Prinzen, F. W., Willemen, E. & Lumens, J. Vol. 12 978-980 (American College of Cardiology Foundation Washington DC, 2019).

4 Wang, X. *et al.* Beyond conduction impairment: unveiling the profound myocardial injury in left bundle branch block. *Heart Rhythm* **21**, 1370-1379 (2024).

5 Tan, N. Y., Witt, C. M., Oh, J. K. & Cha, Y.-M. Left bundle branch block: current and future perspectives. *Circulation: Arrhythmia and Electrophysiology* **13**, e008239 (2020).

6 Brunekreeft, J., Graauw, M., De Milliano, P. & Keijer, J. Influence of left bundle branch block on left ventricular volumes, ejection fraction and regional wall motion. *Netherlands Heart Journal* **15**, 89-94 (2007).

7 Egbe, A. C. *et al.* Left ventricular systolic dysfunction and cardiovascular outcomes in tetralogy of Fallot: systematic review and meta-analysis. *Canadian Journal of Cardiology* **35**, 1784-1790 (2019).

8 Members:, A. T. F. *et al.* 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). With the special contribution of the Heart Failure Association (HFA) of the ESC. *European journal of heart failure* **24**, 4-131 (2022).

9 Glikson, M. *et al.* 2021 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: Developed by the Task Force on cardiac pacing and cardiac resynchronization therapy of the European Society of Cardiology (ESC) With the special contribution of the European Heart Rhythm Association (EHRA). *European Heart Journal* **42**, 3427-3520, doi:10.1093/eurheartj/ehab364 (2021).

10 Sze, E. *et al.* Comparison of incidence of left ventricular systolic dysfunction among patients with left bundle branch block versus those

ARTICLE IN PRESS

with normal QRS duration. *The American journal of cardiology* **120**, 1990-1997 (2017).

11 Olesen, L. L. & Andersen, A. ECG as a first step in the detection of left ventricular systolic dysfunction in the elderly. *ESC heart failure* **3**, 44-52 (2016).

12 Sanna, G. *et al.* Left bundle branch block and ventricular dysfunction: the importance of QRS duration over strict morphological criteria; data from the RECOrd-LBBB registry. *European Heart Journal* **43**, ehac544. 825 (2022).

13 Attia, Z. I., Harmon, D. M., Behr, E. R. & Friedman, P. A. Application of artificial intelligence to the electrocardiogram. *European Heart Journal* **42**, 4717-4730, doi:10.1093/eurheartj/ehab649 (2021).

14 Bjerkén, L. V., Rønborg, S. N., Jensen, M. T., Ørting, S. N. & Nielsen, O. W. Artificial intelligence enabled ECG screening for left ventricular systolic dysfunction: a systematic review. *Heart Failure Reviews* **28**, 419-430 (2023).

15 Attia, Z. I. *et al.* Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nature medicine* **25**, 70-74 (2019).

16 Brito, B. O. d. F. *et al.* Left ventricular systolic dysfunction predicted by artificial intelligence using the electrocardiogram in Chagas disease patients–The SaMi-Trop cohort. *PLoS neglected tropical diseases* **15**, e0009974 (2021).

17 Cho, J. *et al.* Artificial intelligence algorithm for screening heart failure with reduced ejection fraction using electrocardiography. *ASAIO Journal* **67**, 314-321 (2021).

18 Attia, I. Z. *et al.* External validation of a deep learning electrocardiogram algorithm to detect ventricular dysfunction. *International journal of cardiology* **329**, 130-135 (2021).

19 Attia, Z. I. *et al.* Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular

ARTICLE IN PRESS

ARTICLE IN PRESS

systolic dysfunction. *Journal of cardiovascular electrophysiology* **30**, 668-674 (2019).

20 Attia, Z. I., Kapa, S., Noseworthy, P. A., Lopez-Jimenez, F. & Friedman, P. A. in *Mayo clinic proceedings.* 2464-2466 (Elsevier).

21 Kwon, J.-m. *et al.* Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean circulation journal* **49**, 629-639 (2019).

22 Jeong, J. H. *et al.* Deep learning algorithm for predicting left ventricular systolic dysfunction in atrial fibrillation with rapid ventricular response. *European Heart Journal-Digital Health* **5**, 683-691 (2024).

23 Bacharova, L., Szathmary, V. & Mateasik, A. Electrocardiographic patterns of left bundle-branch block caused by intraventricular conduction impairment in working myocardium: a model study. *J Electrocardiol* **44**, 768-778, doi:10.1016/j.jelectrocard.2011.03.007 (2011).

24 McKeen, K. *et al.* Ecg-fm: An open electrocardiogram foundation model. *arXiv preprint arXiv:2408.05178* (2024).

25 Wang, Z., Yan, W. & Oates, T. in *2017 International joint conference on neural networks (IJCNN).* 1578-1585 (IEEE).

26 Karim, F., Majumdar, S., Darabi, H. & Harford, S. Multivariate LSTM-FCNs for time series classification. *Neural networks* **116**, 237-245 (2019).

27 Ismail Fawaz, H. *et al.* Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery* **34**, 1936-1962 (2020).

28 Oh, J., Chung, H., Kwon, J.-m., Hong, D.-g. & Choi, E. in *Conference on Health, Inference, and Learning.* 338-353 (PMLR).

29 Le-Khac, P. H., Healy, G. & Smeaton, A. F. Contrastive representation learning: A framework and review. *Ieee Access* **8**, 193907-193934 (2020).

30 Han, Y., Liu, X., Zhang, X. & Ding, C. Foundation Models in Electrocardiogram: A Review. *arXiv preprint arXiv:2410.19877* (2024).

ARTICLE IN PRESS

31 Shrikumar, A., Greenside, P. & Kundaje, A. in *International conference on machine learning.* 3145-3153 (PMlR).

32 Goetzinger, K. R. & Odibo, A. O. Statistical analysis and interpretation of prenatal diagnostic imaging studies, Part 1: evaluating the efficiency of screening and diagnostic tests. *Journal of Ultrasound in Medicine* **30**, 1121-1127 (2011).

33 Kashani, A. & Barold, S. S. Significance of QRS Complex Duration in Patients With Heart Failure. *Journal of the American College of Cardiology* **46**, 2183-2192, doi:doi:10.1016/j.jacc.2005.01.071 (2005).

34 Hamdani, N. & El-Battrawy, I. Between the Beats: Unraveling Diagnostic, Therapeutic Challenges, and Sex Differences in Heart Failure's Gray Zone. *Journal of the American Heart Association* **14**, e038364, doi:doi:10.1161/JAHA.124.038364 (2025).

35 Zusterzeel, R. *et al.* The 43rd International Society for Computerized Electrocardiology ECG initiative for the automated detection of strict left bundle branch block. *J Electrocardiol* **51**, S25-s30, doi:10.1016/j.jelectrocard.2018.08.001 (2018).

ARTICLE IN PRESS

Author contributions

Study concept/design: J-H.P., D.L., Data collection: Y.B., S.H.L., Data processing: S.P, D.H.K, J.H.P, Data analysis and interpretation: D.H.K, D.L., J-H.P., Manuscript preparation: D.H.K., Manuscript editing: D.L., J-H.P., Final approval of this version to be submitted: all authors.

Funding

This work was supported by Chungnam National University Research Fund, 2023-0540-01 and by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT)(IITP-2025-RS-2023-00258971,

ARTICLE IN PRESS

RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)]

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. Due to the retrospective nature of the study, (Chungnam National University Hospital IRB 2025-02-001 and Jeonbuk National University Hospital IRB 2025-02-030) waived the need of obtaining informed consent.

Competing interests

The authors declare no competing interests.

ARTICLE IN PRESS

Additional information

Correspondence and requests for materials should be addressed to J-H.P. and D.L.

ARTICLE IN PRESS

Figure Legends

Figure 1. Flowchart of the left ventricular systolic dysfunction (LVSD) classification process using electrocardiographic (ECG) data and training approaches. (a) ECG scan images are converted into 4-channel ECG time series data through optical character recognition (OCR), followed by resampling. Separately, 12-channel native digital ECG data are spliced to match the same 4-channel format. (b) Each line represents baseline training models from scratch, single-step and sequential fine-tuning of the pre-trained ECG-FM. The preprocessed left bundle branch block (LBBB) ECG data from (a) are used to train and validate all models. Models can be loaded using custom frameworks or public checkpoints.

Figure 2. Architecture of the left bundle branch block (LBBB)-specific ECG-FM. It is used to fine-tune the ECG-FM with LBBB ECG data. The projection layer, which converts 4-channel ECG data into 12-channel data, feeds the transformed inputs into the convolutional encoder. After passing through the convolutional encoder, latent features are combined with positional embeddings and processed by the transformer encoder to refine local representations. * indicates that 7 individual models are developed by freezing the top 12, 10, 8, 6, 4, 2, and 0 layers of the transformer encoder, respectively.

Figure 3. Post-hoc pairwise comparison of all individual models between the baseline model and the ECG-FM using the Wilcoxon signed-rank test with Holm correction.
Sq-PF, sequential partial fine-tuning; SS-FF, single-step full fine-tuning.

Figure 4. Local explanation of the electrocardiographic (ECG) signal for left ventricular systolic dysfunction (LVSD) prediction using DeepLIFT. The areas highlighted in red indicate the parts of the ECG signal that

ARTICLE IN PRESS

contributed most to the model's prediction. Each lead (I, II, III, aVR, aVL, aVF, V1-V6, and the rhythm strip) is displayed with intensity proportional to the importance of the segment to the prediction. The color bar at the bottom represents the magnitude of the contribution, with higher values indicating greater importance.

ARTICLE IN PRESS

ARTICLE IN PRESS

Tables

Table 1. Dataset composition, including demographics and echocardiographic findings.

Characteristics	Training & Tuning	Internal Validation	External Validation 1	External Validation 2	P-value
Number of Patients	602	89	115	86	
Number of ECG-EchoCG pairs	1,500	194	115	222	
Demographics					
Age, mean ± SD (years)	72.6 ± 11.1	76.7 ± 9.9	71.8 ± 12.0	75.5 ± 10.8	< 0.001
Females, n (%)	738 (49.2)	88 (45.4)	81(70.4)	121 (54.5)	< 0.001
Echocardiographic findings					
LVEF, mean ± SD(%)	41.4 ± 14.0	46.0 ± 14.8	43.3 ± 14.4	46.2 ± 13.5	< 0.001
LVEF<40%, n (%)	680 (45.3)	68 (35.1)	45 (39.1)	80 (36.0)	0.004

ECG: electrocardiogram, EchoCG: echocardiography, LVEF, left ventricular ejection fraction

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 2. Performance of electrocardiographic-foundation model (ECG-FM) compared to conventional baseline models in the prediction of left ventricular systolic dysfunction.

Model (Tuning Method)	Dataset	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score	AUROC	AUPRC
FCN (Scratch)	Int. val	0.660	0.750	0.611	0.510	0.819	0.607	0.719	0.572
	Ext. val 1	0.617	0.804	0.493	0.514	0.791	0.627	0.681	0.567
	Ext. val 2	0.730	0.738	0.725	0.602	0.831	0.663	0.744	0.558
	Average	0.669	0.764	0.610	0.542	0.813	0.632	0.715	0.566
LSTM-FCN (Scratch)	Int. val	0.670	**0.765***	0.619	0.520	**0.830***	0.619	0.739	0.595
	Ext. val 1	0.626	0.739	0.551	0.523	0.760	0.613	0.682	0.569
	Ext. val 2	0.712	0.700	0.718	0.583	0.810	0.636	0.746	0.566
	Average	0.669	0.735	0.629	0.542	0.800	0.623	0.722	0.567
ResNet (Scratch)	Int. val	0.670	0.691	0.659	0.522	0.798	0.595	0.736	0.539
	Ext. val 1	0.652	**0.870***	0.507	0.541	0.854	0.667	0.672	0.536
	Ext. val 2	0.694	0.613	0.739	0.570	0.772	0.590	0.751	0.534
	Average	0.672	0.724	0.635	0.544	0.808	0.617	0.720	0.536
Inception Time (Scratch)	Int. val	0.701	0.603	0.754	0.569	0.779	0.586	0.739	0.505
	Ext. val 1	0.652	0.761	0.580	0.547	0.784	0.636	0.716	0.593
	Ext. val 2	0.721	0.575	0.803	0.622	0.770	0.597	0.783	0.586
	Average	0.691	0.646	0.712	0.579	0.778	0.606	0.746	0.561
ECG-FM (Sq-PF)	Int. val	0.742	0.691	0.770	0.618	0.822	0.653	0.776	0.581
	Ext. val 1	0.696	**0.870***	0.580	0.580	**0.870***	0.696	0.783	0.641
	Ext. val 2	0.784	**0.800***	0.775	0.667	**0.873***	**0.727***	0.844	0.740
	Average	0.741	**0.787***	0.708	0.622	**0.855***	0.692	0.801	0.654
ECG-FM (SS-FF)	Int. val	**0.773***	0.691	**0.831***	**0.671***	0.817	**0.681***	**0.786***	**0.619***
	Ext. val 1	**0.713***	0.848	**0.860***	**0.600***	0.623	**0.703***	**0.784***	**0.659***
	Ext. val 2	**0.788***	0.775	**0.863***	**0.681***	0.796	0.725	**0.850***	**0.755***
	Average	**0.758***	0.771	**0.851***	**0.651***	0.745	**0.703***	**0.807***	**0.678***

* indicates statistical significance

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the receiver precision-recall curve; ECG-FM, electrocardiographic-foundation model; Ext. val, External validation; FCN, Fully Convolutional Network; Int. val, Internal

ARTICLE IN PRESS

validation; LSTM-FCN, Long Short Term Memory Fully Convolutional Network; NPV, negative predictive value; PPV, positive predictive value; ResNet, Residual Network; SS-FF, single-step full fine-tuning; Sq-PF, sequential partial fine-tuning

ARTICLE IN PRESS

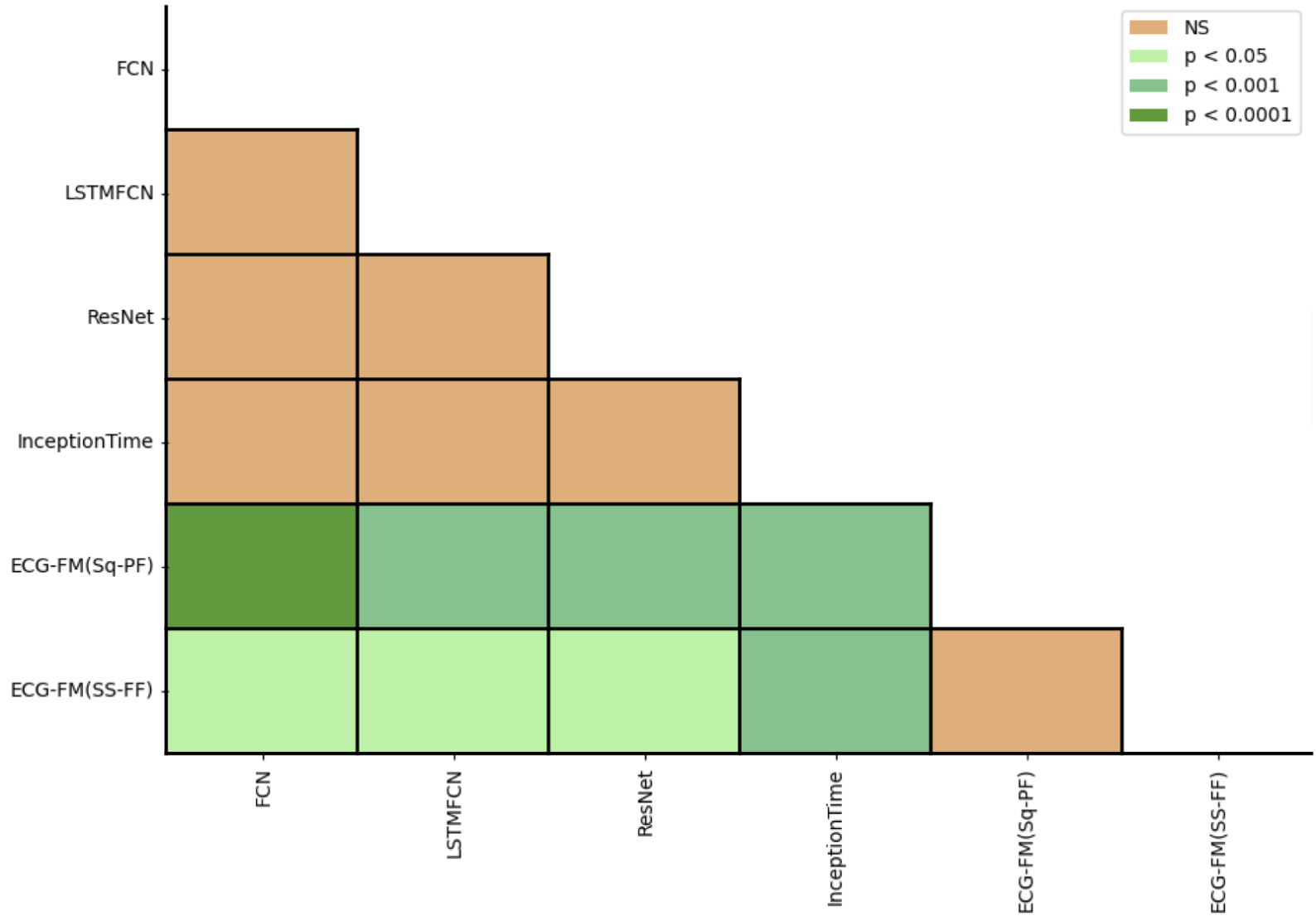
ARTICLE IN PRESS

Table 3. Comparison of performance of electrocardiographic-foundation model (ECG-FM) and InceptionTime model. ECG-FM includes single-step full fine-tuning (SS-FF) and scratch training methods.

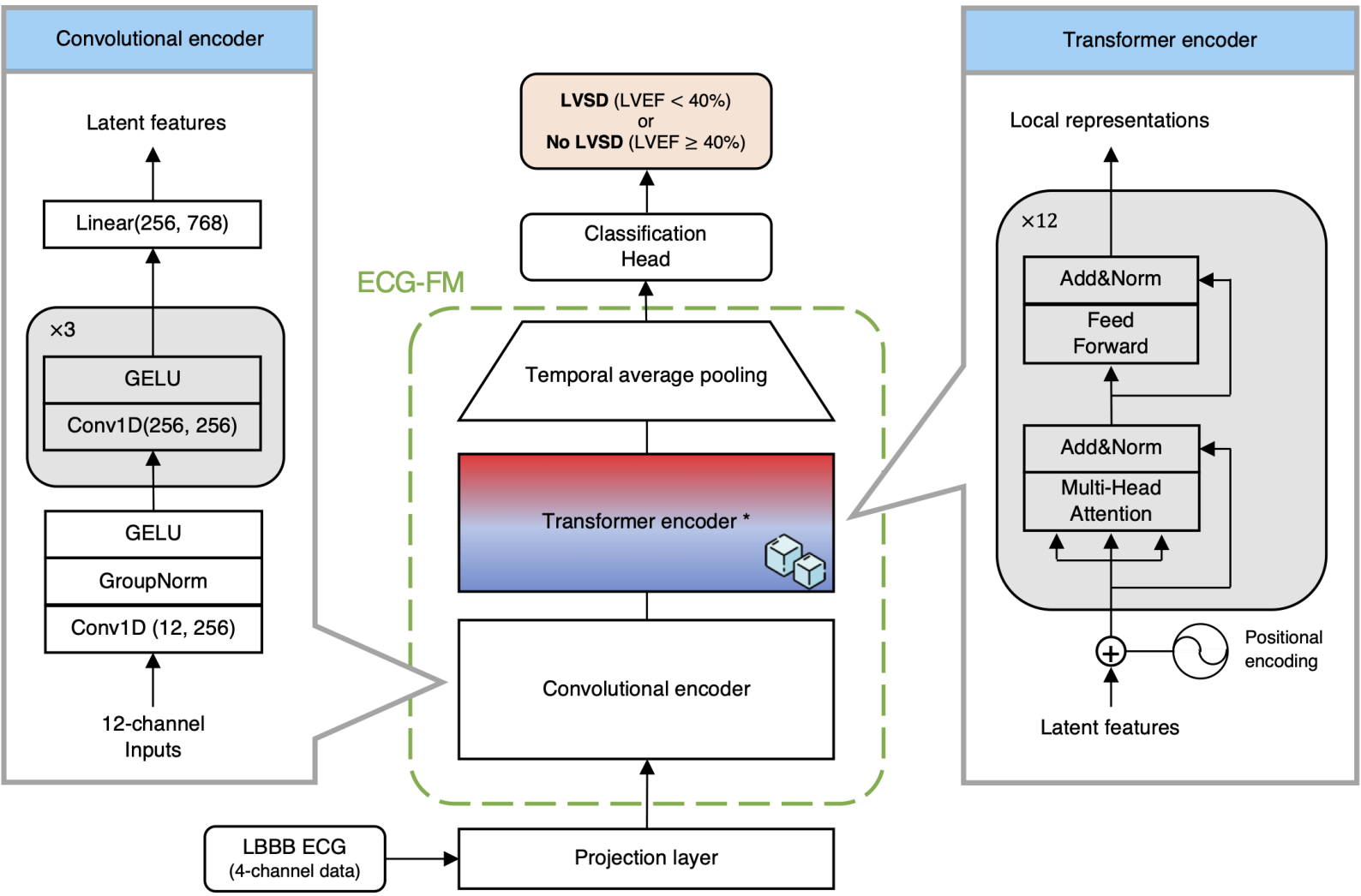
Average Score	InceptionTime	ECG-FM	
	Scratch	Scratch	SS-FF
Accuracy	0.691	0.622	**0.758***
AUROC	0.746	0.677	**0.807***

* indicates statistical significance; AUROC, area under the receiver operating characteristic curve; ECG-FM, electrocardiographic-foundation model; SS-FF, single-step full fine-tuning

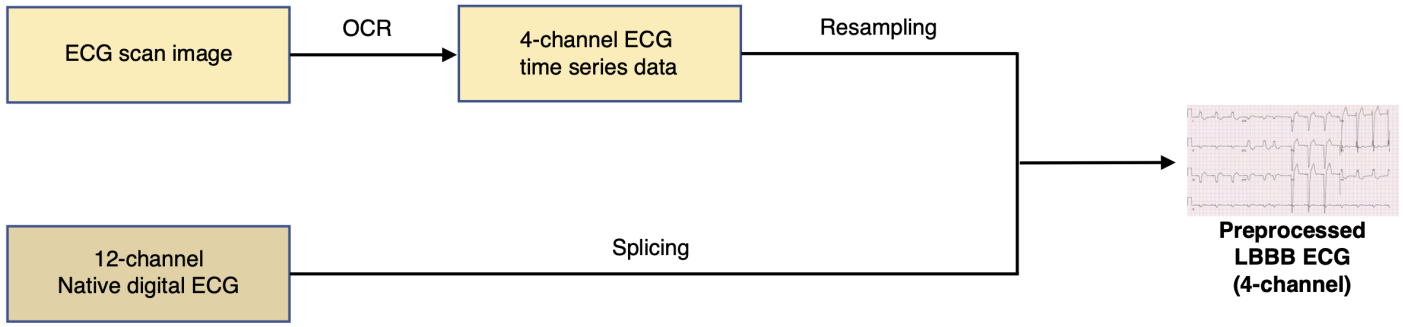
ARTICLE IN PRESS


NS
p < 0.05
p < 0.001
p < 0.0001
FCN
LSTMFCN
ResNet
InceptionTime
ECG-FM(Sq-PF)
ECG-FM(SS-FF)
FCN
LSTMFCN
ResNet
InceptionTime
ECG-FM(Sq-PF)
ECG-FM(SS-FF)

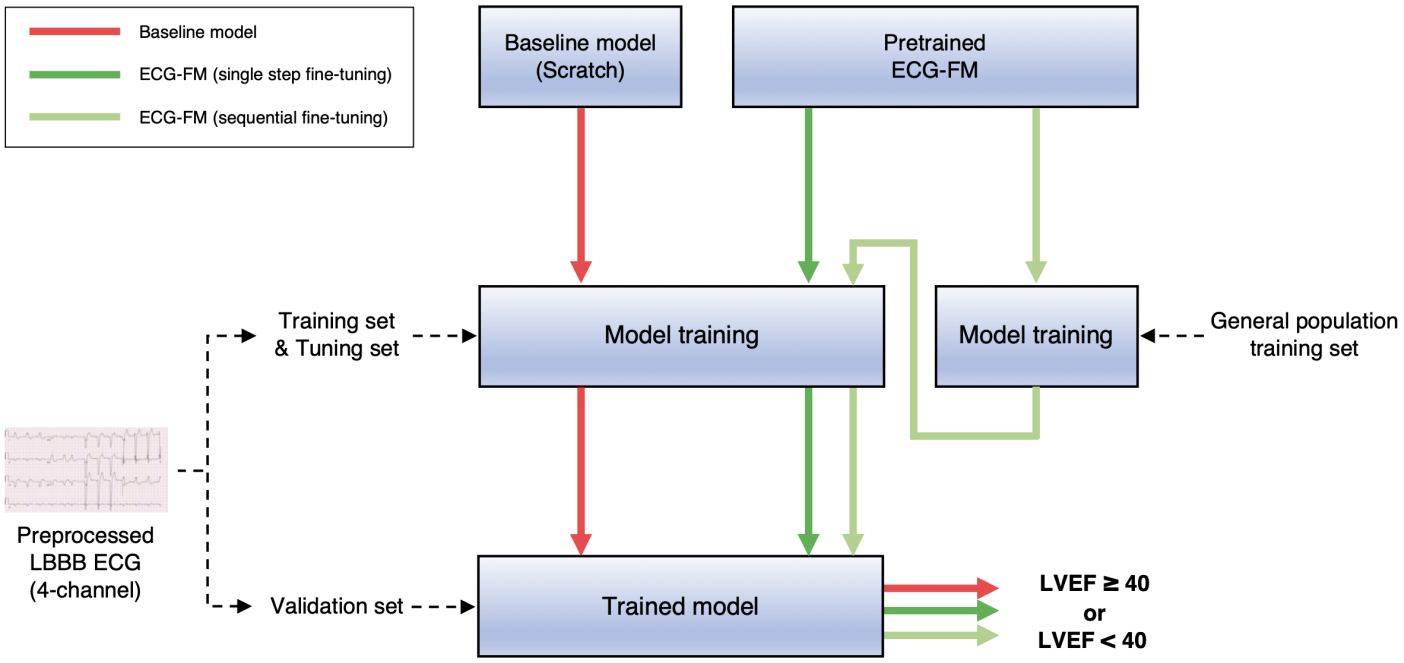
I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6
II (Rhythm Strip)
0.0
0.2
0.4
0.6
0.8
1.0


Convolutional encoder
Latent features
Linear(256, 768)
×3
GELU
Conv1D(256, 256)
GELU
GroupNorm
Conv1D (12, 256)
12-channel
Inputs
LVSD (LVEF < 40%)
or
No LVSD (LVEF ≥ 40%)
Classification
Head
ECG-FM
Temporal average pooling
Transformer encoder *
Convolutional encoder
LBBB ECG
(4-channel data)
Projection layer
Transformer encoder
Local representations
×12
Add&Norm
Feed
Forward
Add&Norm
Multi-Head
Attention
Positional
encoding
Latent features

(a)


ECG scan image
OCR
4-channel ECG
time series data
Resampling
12-channel
Native digital ECG
Splicing
Preprocessed
LBBB ECG
(4-channel)

(b)


Baseline model
ECG-FM (single step fine-tuning)
ECG-FM (sequential fine-tuning)
Baseline model
(Scratch)
Pretrained
ECG-FM
Training set
& Tuning set
Model training
Model training
General population
training set
Preprocessed
LBBB ECG
(4-channel)
Validation set
Trained model
LVEF ≥ 40
or
LVEF < 40