ECG Quality Assessment from a PCG-Oriented Device Using a Minimal Fine-Tuning Strategy on a Pre-trained Convolutional Neural Network

Álvaro Huerta[1], Shib S. Banerjee[2], Francesco Renna[2], Pilar Escribano[1], Catia I. Costa[3], Cristina Oliveira[4], André Lobo[3], Ricardo Fontes-Carvalho[4], Óscar Ayo-Martín[5], José J. Rieta[6], Raúl Alcaraz[1]

[1] Research Group in Electronic, Biomedical and Telecom. Eng., Univ. of Castilla-La Mancha, Spain
[2] INESC TEC, Faculdade de Ciências da Universidade do Porto, Porto, Portugal
[3] ULSGE, Portugal
[4] UnIC@RISE, Portugal
[5] Department of Neurology, University Hospital of Albacete, Spain
[6] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain

Abstract

Deep learning methods have shown promising results for ECG signal quality assessment for standard ECG positions. However, their performance on ECG signals acquired with PCG-oriented devices, which present different morphologies, remains largely unexplored. This study evaluates a method based on the pre-trained convolutional neural network AlexNet on ECG signals from a PCG-oriented device and examines the impact of minimal fine-tuning using limited subject data. The algorithm, originally trained on a dataset of near-standard ECG segments, was tested on 1,465 5-second ECG intervals recorded simultaneously with PCG signals at the main auscultation sites. This experiment resulted in an accuracy of 71.5%, with a sensitivity of 56.6% and a specificity of 96.1%, revealing the impact of morphology variations on classification performance. To improve adaptation, a lightweight fine-tuning strategy was investigated using data from 1, 5, 10, 15, 20 and 25 patients from the target database. Fine-tuning with data from a single patient increased accuracy by about 14% and sensitivity by more than 30%, achieving the best performance with data from ten patients. These results demonstrate that the analyzed model was efficiently adapted to ECG signals acquired with PCG-oriented devices using a limited amount of target-specific data, supporting its deployment in new acquisition environments.

1. Introduction

Auscultation has traditionally been the primary method for assessing cardiac activity. Among the available techniques, PCG remains widely used due to its non-invasive nature, ease of implementation, and relatively low cost [1]. Furthermore, advances in modern technology have led to the development of novel devices that integrate PCG signal acquisition with ECG, thereby providing powerful diagnostic tools for use in primary care settings [2]. These ECG signals are recorded at common PCG auscultation sites rather than at standard ECG acquisition locations, providing complementary information about cardiac function and potentially enabling more accurate and robust diagnoses than conventional approaches [3]. Nevertheless, the portable nature of these devices can compromise ECG signal quality due to increased susceptibility to noise and recording artifacts, emphasizing the importance of robust signal quality assessment (SQA) procedures.

Several methods have been proposed for automated ECG quality assessment, ranging from traditional approaches based on handcrafted features and statistical metrics to more recent deep learning techniques [4]. Early methods typically relied on fiducial points or fixed decision thresholds. Although these approaches are often interpretable and computationally efficient, their performance may be limited when dealing with highly variable recording conditions. More recently, advances in artificial intelligence have enabled the development of data-driven solutions capable of automatically learning discriminative representations directly from raw signals. In particular, convolutional neural network (CNN)-based methods have attracted considerable attention and have demonstrated excellent performance in ECG quality assessment tasks when developed and validated using standard-lead recordings acquired under controlled conditions and morphologically homogeneous datasets [5]. Their ability to capture complex temporal and morphological patterns without extensive feature engineering has positioned CNNs as one of the most promising approaches for automated ECG quality evaluation [6].

Despite the remarkable progress achieved by CNN-based SQA methods, an important research gap remains in the literature. Most existing studies have focused on evaluating model performance using ECG recordings acquired using standard or near-standard leads, while the impact of morphology alterations associated with alternative sensing devices has received little attention. This issue is particularly relevant for ECG signals obtained from PCG-oriented devices, where the sensing configuration and measurement conditions may introduce noticeable changes in the typical morphology of the recorded ECG waveforms. Such modifications do not necessarily reflect a degradation of signal quality but may nevertheless influence the features learned by CNN-based models and affect their performance. To date, the impact of morphological ECG variations on SQA has not been systematically evaluated. Therefore, this study aims to evaluate the impact of these alterations observed in ECG signals acquired by a PCG-oriented device on the performance of a previously published CNN-based SQA model and examines the need and impact of minimal fine-tuning using limited subject data.

2. Methods

2.1. ECG quality assessment

In recent years, several studies have addressed the evaluation of ECG quality from portable and wearable devices in order to ensure reliable detection even in challenging contexts, such as recordings obtained in real-world monitoring environments or from patients with cardiac conduction abnormalities. Within this framework, a CNN-based method has been proposed for the robust quality assessment of single-lead ECG recordings obtained from portable and wearable monitoring systems, reporting high performance [6]. This algorithm converts ECG segments into time–frequency representations using continuous wavelet transform (CWT). This transformation converted each ECG segment into 2-D images, enabling the simultaneous analysis of temporal and spectral information. The generated images, known as scalograms, were fed into a CNN trained to automatically distinguish between high- and low-quality recordings without the need for signal preprocessing or manual feature extraction.

The algorithm was based on AlexNet, a widely used CNN architecture consisting of five convolutional layers followed by three fully connected layers [7]. Originally developed for large-scale image classification tasks, it has become one of the most commonly used architectures in deep learning applications due to its high feature extraction capability. In the referenced study [6], the pre-trained AlexNet model was fine-tuned using the wavelet scalograms generated from the ECG segments. The CNN automatically learns relevant patterns from the time–frequency representations through successive convolution and pooling operations, thereby avoiding the need for manual feature extraction. To fit the binary classification problem, the original fully connected layer was modified to discriminate between high- and low-quality ECG segments, and the network was fine-tuned using a stochastic gradient descent optimization procedure. This approach enabled robust ECG quality assessment while maintaining the advantages of deep learning-based automatic feature learning.

The proposed approach was fine-tuned using data from three different monitoring scenarios, including smartphone-based recordings, home telemonitoring systems, and wearable textile Holter devices. These signals contained sinus rhythm, atrial fibrillation (AF) episodes, other cardiac rhythms, and segments affected by noise and artifacts, thus representing a challenging and clinically relevant environment. The study reported a discriminative performance of approximately 94% for identifying high-quality ECG segments while successfully excluding noisy recordings by 90%. In addition, it correctly classified the 92% of the ECG intervals affected by AF. These results demonstrate the effectiveness and reliability of this algorithm using portable and wearable cardiac monitoring devices. Given this strong performance, the algorithm was selected in the present study to investigate its applicability and robustness when analyzing ECG signals acquired using PCG-oriented devices.

2.2. Dataset and experimental setup

To evaluate the performance, generalizability, and robustness of the described methodology [6], the approach was applied to the ULSGE (Unidade Local de Saúde de Gaia e Espinho) database. This dataset comprises multimodal recordings acquired using different sensing technologies and acquisition configurations, including standard 12-lead ECGs, single-lead ECGs, and PCG signals. For the present study, only the single-lead ECG recordings acquired simultaneously with PCG at the four main auscultation sites, i.e., tricuspid, pulmonary, mitral, and aortic valve positions were considered. These signals were recorded using a digital stethoscope (CardioSleeve, Rijuven) with a sampling rate of 500 Hz and recording durations ranging from 22 to 51 seconds and annotated by a single expert. This database involves 267 patients staying at the cardiology department of the hospital, with different heart conditions such as normal sinus rhythm and AF which constitute a challenging dataset due to the variability in signal quality and acquisition conditions. After segmenting the data into a 5-second length intervals without overlapping, a total of 1,465 ECG segments were analyzed, including 914 high-quality segments and 551 noisy excerpts.

The generalization capability of the selected CNN-based model for ECG quality assessment was evaluated in a first zero-shot experiment. Indeed, the pre-trained

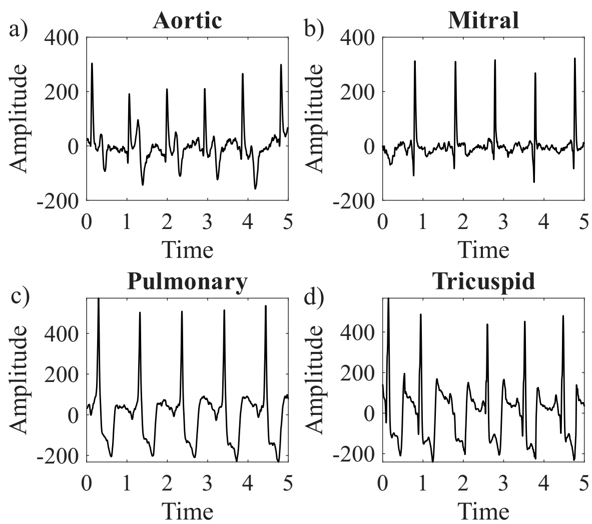

Figure 1. Different examples from ECG signals belonging to the same patient in the PCG positions aortic (a), mitral (b), pulmonary (c) and tricuspid (d).

AlexNet architecture was used with the weights obtained in the original study, where the network had been trained using 20,000 ECG excerpts of 5-second length from two balanced classes [6]. This CNN model was externally evaluated with the ULSGE database as test set. The ECG signal morphologies provided by the PCG-based device exhibited significant differences compared with conventional ECG signals, as shown in Figure 1, which highlights the challenge of transferring models across acquisition systems. The scalograms derived from this dataset were directly provided as input to the network to be classified using the previously acquired knowledge. No additional retraining, adaptation, or calibration was performed using data from the target database. Therefore, this first experiment represented a strict external validation scenario, assessing the model's ability to classify ECG quality in previously unseen data with this specific morphology.

To determine whether the performance obtained in the previous experiment could be further improved, an additional study was conducted by incorporating a small amount of data from the target database into the training process. Specifically, ECG segments from a randomly selected single patient were used for a fine-tuning procedure, after which that patient was excluded from the testing set to avoid data leakage and ensure an unbiased evaluation. This lightweight fine-tuning strategy was implemented by retraining the entire network, with no layers frozen. Consequently, the weights of all layers were updated during training, including those of the final fully connected layer, which continued to provide the binary classification between low-quality and high-quality ECG excerpts. The model was trained for 30 epochs using a batch size of 16 and a stochastic gradient descent (SGD) optimizer with a fixed learning rate of 0.01. This methodology was then repeated using data from 5, 10, 15, 20 and 25 randomly selected patients. Even in the last case, the samples employed for fine-tuning of the initial algorithm represented less than 15% of the patients available in the target database. This experimental setup enabled the assessment of how limited target-domain data influences model performance while maintaining a lightweight adaptation strategy.

For each considered number of subjects, the fine-tuning procedure was repeated 10 times, randomly selecting the subjects in each repetition to account for sampling variability. In all cases, the performance of the algorithm was assessed through the mean and standard deviation of Sensitivity (Se), Specificity (Sp), and Accuracy (Acc) computed over the 10 conducted iterations. It should be noted that the class distribution in the test database was not balanced. However, this imbalance did not affect the fine-tuning procedure, as the dataset was used exclusively for evaluation purposes. Consequently, the reported metrics enabled a reliable assessment of the model's performance and generalization capability on unseen data.

3. Results

Table 1 summarizes the performance achieved by the CNN model under different fine-tuning configurations. In the zero-shot experiment, it reached an Acc of 71.5%, with relatively low Se (56.6%) but high Sp (96.1%), indicating a bias toward correct negative classification. The incorporation of ECG data from one subject of the target dataset for lightweight fine-tuning led to a substantial improvement, increasing Acc to 85.2% and Se to 88.2%. Further enlarging the fine-tuning dataset yielded moderate additional gains, achieving a maximum Acc of 87.6% and Se of 92.2% when data from ten patients were included, thus highlighting the progressive adaptation of the model to the target domain. Moreover, as the number of patients increased, the standard deviation decreased, indicating more stable estimates and reduced variability. Performance remained stable between 10 and 20 patients. However, a marked degradation was observed when 25 patients were included.

Table 1. Performance metrics reported by each experiment based on minimal fine-tuning strategy.

Fine-tuning	Se **(%)**	Sp **(%)**	Acc **(%)**
None	56.6	96.1	71.5
1 patient	88.2±8.3	80.1±6.7	85.2±5.4
5 patients	89.1±7.7	76.9±4.3	84.5±5.6
10 patients	92.2±5.3	79.8±3.1	87.6±4.3
15 patients	85.6±6.1	85.4±4.7	86.5±4.1
20 patients	88.8±5.1	82.6±4.6	86.6±3.8
25 patients	58.5±3.5	39.5±11.7	49.2±7.9

4. Discussion

The presented results suggest that the knowledge learned from standard ECG data for quality assessment can be effectively transferred to signals acquired with PCG-oriented devices through a remarkably small adaptation effort. The rapid performance improvement observed after incorporating data from only a few subjects indicates that the network can successfully adapt to domain-specific characteristics with limited exposure to the target distribution. Furthermore, the algorithm achieved the best performance when fine-tuned with 10 patients, highlighting the efficiency of the analyzed lightweight adaptation strategy and suggesting that the most relevant quality-related features learned from conventional ECG recordings remain informative in ECG signals acquired with PCG-oriented devices. Moreover, the model performance declined when more than 20 patients were included in the adaptation process, suggesting potential limitations in the model's ability to generalize across increasingly heterogeneous target populations. Since this degradation was consistently observed when fine-tuning with 25 patients, it may reflect overfitting to dataset-specific characteristics, reducing the model's ability to capture robust quality-related features. From a practical perspective, these findings suggest that only a limited amount of annotated target-domain data is required for effective deployment in new acquisition environments.

Despite the limited availability of studies investigating this particular scenario, a previous contribution also suggests that fine-tuned models can effectively adapt pretrained CNN schemes for ECG quality evaluation [8]. The results presented by Ribeiro et al. support our findings, showing that even a highly constrained adaptation process can achieve substantial performance improvements and robust discrimination of ECG signal quality.

5. Conclusions

The evaluated model, based on a well-established pretrained CNN for ECG quality assessment, demonstrated a strong capability to operate on ECG signals acquired using a PCG-oriented device, despite the substantial morphological differences between these signals and those recorded from standard ECG leads. Although the direct application of the model resulted in a considerable performance degradation, a lightweight fine-tuning strategy using data from only a small number of subjects was sufficient to markedly improve classification accuracy. These results underscore the robustness, transferability, and practical utility of the proposed CNN-based approach, highlighting its potential for efficient deployment in novel acquisition settings while requiring only minimal adaptation effort.

Acknowledgments

This research has received financial support from the company Daiichi Sankyo S.A.U. and from public grants PID2021-128525OB-I00, PID2021-123804OB-I00, and PID2025-173302OB-I00 funded by MICIU/AEI/10.13039/501100011033, and FSE+, grant SB-PLY/24/180225/000184 funded by Junta de Comunidades de Castilla-La Mancha, and grant AICO/2021/ 286 funded by Generalitat Valenciana, and the European Regional Development Fund (ERDF) through the Innovation and Digital Transition Programme (COMPETE 2030) under Portugal 2030 and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within project PULSE, with reference 17172 (COMPETE2030-FEDER-00864900).

References

[1] Calzoni A, Savardi M, Silvestri M, Benini S, Signoroni A. Bimodal ECG and PCG cardiovascular disease detection: Exploring the potential and modality contribution. Journal of Medical Systems 2025;49(1):113.

[2] De Fazio R, Cascella I, Yalçınkaya ŞE, De Vittorio M, Patrono L, Velazquez R, Visconti P. Synchronous acquisition and processing of electro-and phono-cardiogram signals for accurate systolic times' measurement in heart disease diagnosis and monitoring. Sensors 2025;25(13):4220.

[3] Karimi S, Shah AJ, Clifford GD, Sameni R. Cross-learning between ECG and PCG: Exploring common and exclusive characteristics of bimodal electromechanical cardiac waveforms. arXiv ;2506.10212, 2025.

[4] Satija U, Ramkumar B, Manikandan MS. A review of signal processing techniques for electrocardiogram signal quality assessment. IEEE reviews in biomedical engineering 2018; 11:36–52.

[5] Mondal A, Manikandan MS, Pachori RB. Automatic ECG signal quality determination using CNN with optimal hyperparameters for quality-aware deep ECG analysis systems. IEEE Sensors Journal 2024;24(11):17825–17833.

[6] Huerta Herraiz Á, Martínez-Rodrigo A, Bertomeu-González V, Quesada A, Rieta JJ, Alcaraz R. A deep learning approach for featureless robust quality assessment of intermittent atrial fibrillation recordings from portable and wearable devices. Entropy 2020;22(7):733.

[7] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems. 2012; 1097–1105.

[8] Ribeiro E, Almeida DA, Soares QB, Pereira JJ, GS R, Verardino TC, Samesima N, Monteiro R, et al. Deep Learning Signal Quality Assessment for Continuous Wearable ECG Monitoring. Computing in Cardiology 2025;1–4.

Address for correspondence:

Álvaro Huerta Herraiz
ITCT, Campus Universitario s/n, 16071, Cuenca, Spain
Phone: +34-969-179-100
e-mail: alvaro.huerta@uclm.es