Computers in Biology and Medicine 166 (2023) 107532

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Enhancing the performance of premature ventricular contraction detection in unseen datasets through deep learning with denoise and contrast attention module

Keewon Shin [a,b], Hyunjung Kim [c], Woo-Young Seo [a], Hyun-Seok Kim [a], Jae-Man Shin [c], Dong-Kyu Kim [a], Yong-Seok Park [d], Sung-Hoon Kim [a,d,1,*], Namkug Kim [c,e,1,**]

[a] Laboratory for Biosignal Analysis and Perioperative Outcome Research, Biomedical Engineering Center, Asan Institute of Lifesciences, Asan Medical Center, Seoul, Korea
[b] Medical Device Research Platform, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Korea
[c] Department of Biomedical Engineering, University of Ulsan College of Medicine, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, Seoul, Korea
[d] Department of Anesthesiology and Pain Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea
[e] Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

ARTICLE INFO

Keywords:
Arrhythmia
Attention module
Deep learning
Denoise and contrast attention (DCAM)
Electrocardiogram (ECG)
Premature ventricular contraction (PVC)

ABSTRACT

Premature ventricular contraction (PVC) is a common and harmless cardiac arrhythmia that can be asymptomatic or cause palpitations and chest pain in rare instances. However, frequent PVCs can lead to more serious arrhythmias, such as atrial fibrillation. Several PVC detection models have been proposed to enable early diagnosis of arrhythmias; however, they lack reliability and generalizability due to the variability of electrocardiograms across different settings and noise levels. Such weaknesses are known to aggravate with new data. Therefore, we present a deep learning model with a novel attention mechanism that can detect PVC accurately, even on unseen electrocardiograms with various noise levels. Our method, called the Denoise and Contrast Attention Module (DCAM), is a two-step process that denoises signals with a convolutional neural network (CNN) in the frequency domain and attends to differences. It focuses on differences in the morphologies and intervals of the remaining beats, mimicking how trained clinicians identify PVCs. Using three different encoder types, we evaluated 1D U-Net with DCAM on six external test datasets. The results showed that DCAM significantly improved the *F1-score* of PVC detection performance on all six external datasets and enhanced the performance of balancing both the *sensitivity* and *precision* of the models, demonstrating its robustness and generalization ability regardless of the encoder type. This demonstrates the need for a trainable denoising process before applying the attention mechanism. Our DCAM could contribute to the development of a reliable algorithm for cardiac arrhythmia detection under real clinical electrocardiograms.

1. Introduction

Premature ventricular contractions (PVCs) are primarily asymptomatic and commonly observed in most individuals monitored for more than a few hours [1]. PVC is frequently observed in patients undergoing surgery regardless of general or regional anesthesia. Typically, intraoperative PVC tends to be negligible as its clinical significance is relatively limited. However, every PVC cannot be considered an innocent arrhythmia and should be monitored with caution in specific instances. Specifically, a new onset of PVC could be considered a potentially serious event as it may lead to fatal arrhythmias such as ventricular tachycardia or ventricular fibrillation [2]. Furthermore,

* Corresponding author. Department of Anesthesiology and Pain Medicine, University of Ulsan College of Medicine, Asan Medical Center, 88, Olympic-ro 43-Gil, Songpa-gu, Seoul, 05505, Republic of Korea.
** Corresponding author. Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, 88, Olympic-ro 43-Gil, Songpa-gu, Seoul, 05505, Republic of Korea.
E-mail addresses: shkimans@amc.seoul.kr (S.-H. Kim), nkkim@amc.seoul.kr (N. Kim).
[1] These corresponding authors contributed equally.

https://doi.org/10.1016/j.compbiomed.2023.107532
Received 13 May 2023; Received in revised form 31 August 2023; Accepted 27 September 2023
Available online 4 October 2023

PVCs may occasionally indicate underlying structural heart disease [3] or be a precursor of higher arrhythmogenic potential [4,5].

Therefore, considerable research has been conducted to develop algorithms to detect PVCs in electrocardiograms (ECGs); however, they have not reached the level of practical clinical application [6–12]. Technically, PVC can be detected in comprehensively as its typical characteristic is a single irregular ectopic beat. However, the identification of PVCs can be particularly difficult on several occasions if the signal is contaminated with noise or mixed with another arrhythmia, such as atrial fibrillation (AF). In addition, inappropriate detection of PVCs and false alarms may lead to significant fatigue, which negatively affects the attending physicians' vigilance. Furthermore, repeated false alarms may lead to a neglect of precautions and serious consequences [13]. To achieve clinically applicable level of technology, an accurate and sophisticated algorithm was required to reliably identify PVCs even in ECGs that contained noise.

Recent advances in smart biosignal measurement devices and machine learning analysis algorithms have led to PVC detection in ECG signals [14–16]. Zhou et al. [17] presented a method for detecting PVCs that combines convolutional neural networks (CNN), long short-term memory (LSTM), and rule inference. Their method achieved good performance considering that it was an early deep learning model for electrocardiogram (ECG) analysis. Wang et al. [18] proposed a 2D CNN-based method to extract ECG curves from scanned ECG images of clinical ECG reports without digital ECG data and segment and classify heartbeats. Naz et al. [19] and Ullah et al. [20] adopted a similar approach to Wang et al. and used a 2D CNN-based classification model pre-trained on ImageNet and achieved high diagnostic performance. Oh et al. [21] first modified the U-Net [22] and modified the model to perform beat-wise analysis on heterogeneously segmented ECGs of varying lengths extracted from the MIT-BIH arrhythmia dataset. Hu et al. [23] introduced a transformer-based model to achieve high detection performance for multiple arrhythmias other than PVCs on three public datasets. While the above studies have shown high performance for PVC detection, they only addressed internal test datasets, which limits true performance evaluation.

As high performance on external datasets can demonstrate real-world model performance, a few studies have recently evaluated PVC detection models in ECGs. Petryshak et al. [24] proposed a two-stage pipeline that first segments QRS complexes (QRS) and then classifies PVCs. They tested their model on two public datasets using cross-dataset paradigms and achieved a high *F1-score* for both tasks. Ivora et al. [25] presented a comprehensive model evaluation protocol using 12 public datasets for QRS detection and 4 databases for arrhythmia detection. Using a large private dataset, they trained their model and demonstrated excellent *micro-F1 scores*. Although these two studies achieved PVC detection performance on external test datasets, there is still a need to improve model performance before they can be considered for clinical application.

Moreover, artificial intelligence has experienced a surge in research endeavors to enhance the performance of deep learning models by employing attention mechanisms [26], a concept originating from language models that assign greater weight to the most pertinent vector among the neighboring vectors. Prominent attention mechanism modules include the Transformer [27], Convolutional Block Attention Module (CBAM) [28], Non-Local Neural Networks (NLNN) [29], and Squeeze-and-Excitation (SE) [30]. The attention module is also reported to improve the performance of deep learning models in time series analysis [31]. However, only a few studies have been interested in improving feature learning performance and not in evaluating the impact of attention mechanisms on performance on an unseen dataset [32,33]. Furthermore, attention modules may introduce dataset-specific biases and reduce the ability of models to generalize to unseen data [34]. To substantiate these claims, it is imperative to investigate whether the application of attention modules improves performance on unseen datasets. In this study, we propose and evaluate a novel Denoise and Contrast Attention Module (DCAM) to improve the PVC detection performance of deep learning models on an unseen ECG dataset.

2. Method

In this section, the model architecture, data collection, and model evaluation protocol are introduced. In particular, a brief overview of the seven datasets, including training and external test datasets. The appendix provides detailed information on these seven datasets.

2.1. Denoise and contrast attention module (f_{DCAM})

To improve the performance of a deep learning model for PVC detection, a CNN-based attention module could be applied to better capture the principles of how clinical experts detect PVCs. Although attention mechanisms are acknowledged to improve model performance, we assume a diagnostic approach similar to anesthesiologists' approach to detecting PVCs by paying more attention to a specific signal, the QRS. When diagnosing PVC, anesthesiologists begin by filtering out any irrelevant signal and then analyzing the shape and spacing of nearby sinus rhythms. To this end, DCAM is a two-step process that denoises a given ECG signal with CNN in the frequency domain and focuses on the contrast of morphologies of the remaining beats. DCAM consists of f_{DM} for frequency domain convolution-based denoising and f_{CAM} for contrast attention module, formulated by two procedures as

$$Y = f_{DCAM}(X) = f_{CAM} f_{DM}(X) = P(f_{DM}(X) + GELU(K - Q)) \quad (1)$$

where f_{DCAM} is a transformation mapping from input feature $X \in \mathbb{R}^{C \times S}$ to output feature $Y \in \mathbb{R}^{C \times S}$, where C represents the channel and S represents the signal length. $f_{DM}(X)$ is the denoised output from the original feature X. The subtraction between feature maps of $K \in \mathbb{R}^{C \times 1}$ and $Q \in \mathbb{R}^{C \times 1}$ extracted by each group CNN (GCNN) [35] is combined with $f_{DM}(X)$ after Gaussian error linear unit (GELU) activation is applied. Finally, the recalibration feature $P \in \mathbb{R}^{C \times S}$ is multiplied to obtain the output feature Y, as shown in Equation (1) and Fig. 1.

2.1.1. Denoise module (f_{DM})

The f_{DM} is designed for denoising based on the fast Fourier transform (FFT) with convolutional operations [36,37], as an imaginary process by which an anesthesiologist ignores noise and artifacts before diagnosing an ECG by considering the time intervals between beats. As shown in Fig. 1(a), our implementation of f_{DM} is based on channel-wise FFT. f_{DM} consists of a three-step procedure to remove the noise from the input signal X. First, we performed an FFT on the signal to convert it from the time domain to the frequency domain. The FFT of a real signal is conjugate symmetric, the right half of the results can be derived from the left half. Therefore, we use the right FFT (rFFT) algorithms to reduce the complexity and calculate the FFT. Because the result of the signal in the frequency domain consists of real and imaginary parts, connecting these two doubles the number of channels. Second, we applied a convolutional layer followed by a normalization layer and a GELU activation function. This convolutional layer acted as a filter that reduced the noise level while preserving the signal features [36]. The normalization layer ensured that the output had a stable distribution and scale. The GELU activation function introduced non-linearity and improved the performance of the convolutional layer. Third, we performed an inverse rFFT on the output of the previous step to convert it back to the time domain. In the inverse rFFT process, we again divided the channel by 1/2 to separate the real and imaginary parts, and finally obtained the result of $f_{DM}(X) \in \mathbb{R}^{C \times S}$, which is a denoised signal that retains the original information and quality.

2.1.2. Contrast attention module (f_{CAM})

The f_{CAM} extracts the contrast features of the K and Q maps of the denoised signal $f_{DM}(X)$ and then recalibrates the signal. Fig. 1(b)

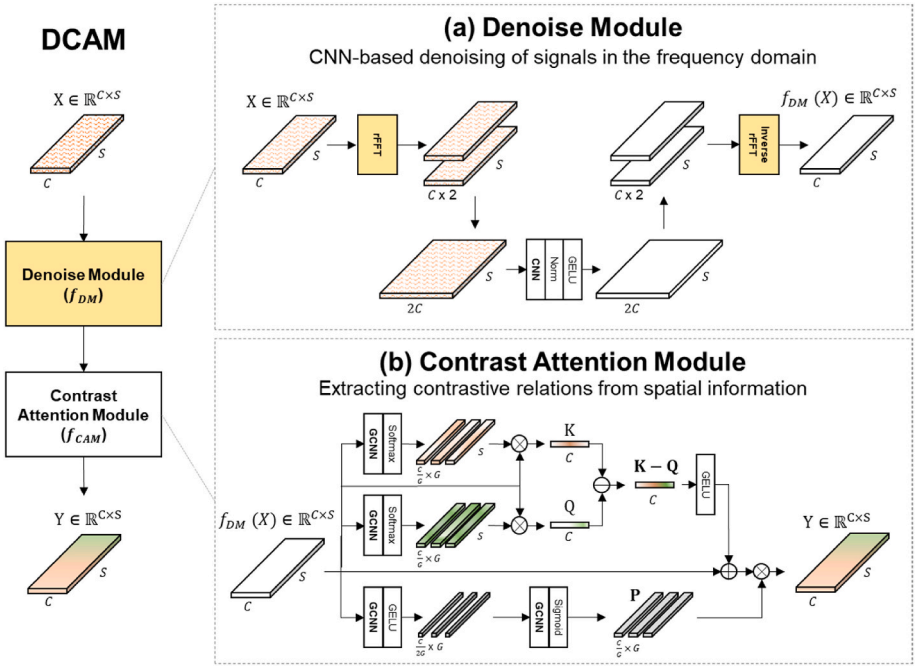

Fig. 1. Schematic of Denoise and Contrast Attention Module (DCAM). The DCAM has two components: (a) a denoising module, f_{DM} that is CNN-based denoising of signals in the frequency domain to filters out noise from the input signals and (b) a contrast attention module f_{CAM} that extracts contrastive relations from spatial information to focus on the QRS morphology difference between PVCs and normal sinus rhythms. Note: fast Fourier transform, FFT; group convolutional neural net, GCNN; Gaussian error linear unit, GELU; normalization, Norm.

illustrates f_{CAM}, which extracts objects of interest and their context for comparison and enhances the original signal feature. Our module was inspired by Attend-and-Compare (ACM) [38], a state-of-the-art method for difference modeling. Compared to ACM, our f_{CAM} extracts feature maps K and Q without normalizing the input signal. Moreover, recalibration P was experimentally determined while maintaining the input signal channel dimensions. To improve the discrimination between PVCs and regular beats, f_{CAM} employs GCNN to learn the contrast between K and Q. GCNN is a type of CNN that divides the input channels into G groups and performs convolution on each group separately. This allows GCNNs to emphasize different features of PVCs at K and normal sinus rhythms at Q in a given signal. Moreover, GCNNs can reduce the computational cost and increase model diversity by adjusting the G parameter. The input signal is split into G groups along the channel dimension, and each group undergoes convolution independently. The output of the GCNNs is passed through a SoftMax layer and batch-wise matrix production by $f_{DM}(X)$ and the matrix to obtain K and Q, respectively. Contrast attention is achieved by adding $K - Q \in \mathbb{R}^{C \times 1}$ to $f_{DM}(X)$ to obtain the contrast-enhanced feature $f_{DM}(X) + K{-}Q \in \mathbb{R}^{C \times S}$. Finally, this contrast attention is multiplied by the recalibrating feature P to obtain the final output Y. Our channel recalibrating feature P was defined as $P = GCNN \circ GELU \circ GCNN(f_{DM}(X))$. The resulting feature vector P will be multiplied by $f_{DM}(X) + K{-}Q$ to scale down certain channel information. P represents which channels to attend to for the task.

2.2. Experimental settings

2.2.1. Preprocessing

To test model on multiple datasets, consistent preprocessing was applied to the input ECG signals. First, we resampled all signals to 250 Hz, which was the best sampling rate in the initial experiment of the seven datasets, and used segments of 1280 samples (5.12 s) as input [39]. The segment length was also chosen depending on the performance of PVC detection models, and the results of our initial trials are shown in appendix Fig. A1. Then, we applied the fifth-order Butterworth bandpass filter at 0.5–40 Hz to remove baseline wandering [39,40]. Third, we normalized and standardized the ECG signals, as their amplitude varied depending on the variations of the patient, electrode placements, and measuring device. Finally, Z-score normalization was applied to standardize the voltage values. This is more resilient to outliers than min-max normalization, which can distort the shape of the signal if it contains artifacts [39].

2.2.2. Baseline PVC detection model

This session describes the baseline model architecture for PVC detection with DCAM. We used the 1D U-Net [22] architecture, a typical segmentation network, as a baseline model, as shown in Fig. 2. The encoder was given three variants, including VGG19 [41], ResNet34 [42], and DenseNet121 [43]. 1D U-Nets with these different encoders were used to examine whether any encoder structure with or without DCAM provides a balanced performance gain for DCAM. The default encoder was VGG19 with simply stacked convolutional layers. To model contextual information at various levels of feature representation, we inserted multiple DCAMs into the encoder network. When DCAM was applied to the encoder, it was inserted at the end of each encoder layer for a total of five. To compare the performances of DCAM and CNN-based attention modules on a 1D U-Net, models with CBAM, NLNN, SE, and ACM added to the same location were evaluated.

2.2.3. Postprocessing

Our proposed architecture takes a single lead ECG signal as input and outputs two channels of QRS and PVC results to simultaneously produce beat-wise QRS and PVC results. A sigmoid function was applied to the last output value of the model to have a value between 0 and 1 and binarized with a 0.5 threshold for both QRS and PVC. Post-processing reduced false positives for PVCs by removing cases smaller than a 0.6-s interval from the binarized segmentation results. When evaluating beat-wise PVC detection performance, the temporal threshold was 0.12 s, which is consistent with previous research [24,25,44].

2.2.4. Implementation details

For training our PVC detection model, we used signal augmentation

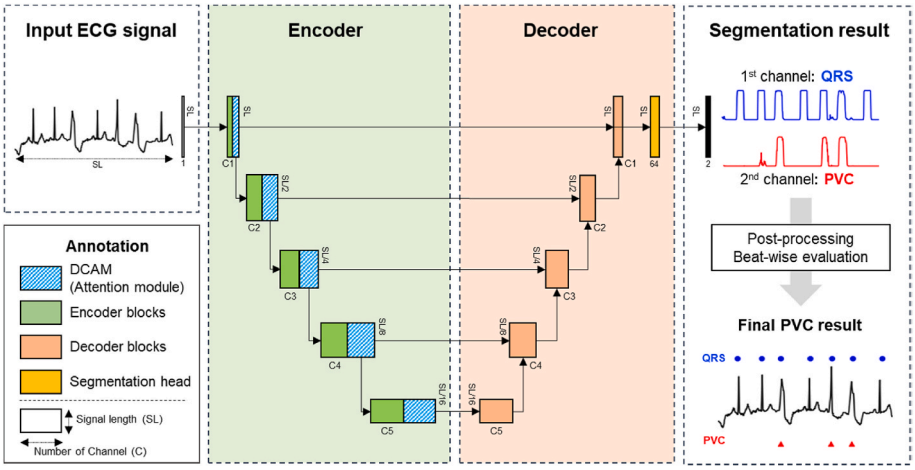

Fig. 2. Structure of a PVC detection model (1D U-Net with DCAM). Each attention module was attached to the end of the encoder layer to evaluate improvement of model performance. Three different types of encoders including VGG19, ResNet34, DenseNet121 were used. Note: electrocardiogram, ECG; premature ventricular contraction, PVC.

during training to increase the diversity and robustness of our model. To prevent degradation of the signal quality by applying excessive or inappropriate augmentation, we used the following augmentation methods: 1) adding Gaussian noise of a clean ECG signal to yield a 10 dB signal-to-noise ratio to simulate noisy environments [40]; 2) randomly resampling the signal within a 5% range from the original sampling rate to simulate variations in heart rate [39]; and 3) using NeuroKit2 [45] to add realistic artifacts and noise to the signal. We also balanced the training dataset between PVC and non-PVC classes using under-sampling non-PVC, which weights each sample based on the presence or absence of PVC at the segment level and uses these weights to balance the samples in batch operation.

A sliding window approach [46] was used for model inference to avoid performance degradation when the proposed model needs to evaluate an arbitrarily long ECG signal, which truncates the signal into segments with every 1280 sample lengths for training and inference. The hyperparameters for the sliding window method were constant weightings with a 0.25 overlap ratio [46].

Lastly, we used binary cross entropy for the training loss function, batch size of 256, and Adam as the optimizer. To ensure fast training of the model and repeatability of training, we used a step-learning rate scheduler. The initial training rate of 1e-3 was reduced by 10% every 30 epochs. In addition, early stopping of training was set when the *F1-score* of the validation dataset reached the maximum. We conducted training on a single GPU (24 GB) using Python 3.9, Pytorch 1.11, and Pytorch Lightning 1.5.0 environments. Training took 1–1.5 h depending on the model configuration.

2.3. Datasets and evaluation protocol

The seven datasets were used in this study to train and test the proposed algorithm for detecting PVCs. We used the Massachusetts Institute of Technology and Beth Israel Hospital Arrhythmia database (MIT-BIH) [47] for model development, as this dataset contains a variety of arrhythmia cases and the annotations per beat were accurately labeled. This data was split record-wise into three subsets: training, validation, and internal test sets, which were used to train our model, tune the hyperparameters, and assess the reference performance of the attention module on the seen dataset, respectively (Table A2).

2.3.1. Model development dataset (MIT-BIH)

To build our model development dataset, we went through several steps (Fig. A1). First, following the recommendations of the Association for the Advancement of Medical Instrumentation (AAMI), four records (102, 104, 107, and 217) containing paced beats were excluded from the experiments, and the remaining 44 recordings of the lead II signal were used [48]. Next, we treated fused beats as PVC beats and ignored unclassified and distorted beats from ground truth (GT) based on an extended definition of PVC in accordance with our anesthesiologist's opinions and previous studies [49,50]. Third, we removed beat-wise annotations that could not be confirmed by lead II alone. As seen in the bottom right corner of Fig. A1, QRS beats were annotated even though they were not visible due to artifacts in lead II. To discard GT that were incorrectly annotated on lead II, we used the Hamiltonian algorithm [51] to detect QRS, extracted cases where the GT indicated a QRS, but the algorithm did not detect it, and removed the mis-annotation by manual review. Table A3 shows how many samples we had before and after data curation.

2.3.2. Evaluation protocol

Our model evaluation protocol, based on Ivora et al. [25], is presented in Fig. 3. We also used six external datasets for external testing: 1) Asan Medical Center Liver Transplant database (AMC-LT), 2) the 3rd China Physiological Signal Challenge 2020 database (CPSC2020) [52], 3) European ST-T database (ES) [53], 4) St. Petersburg Institute of Cardiological Technics database (INCART) [54], 5) MIT-BIH Noise Stress Test Database (MIT-BIH-NS) [55], and 6) MIT-BIH Supraventricular Arrhythmia Database (MIT-BIH-SV) [56]. Table 1 summarizes the characteristics of these datasets, including the type, name, lead, number of recordings, sampling rate, total number of beats, and number of PVC beats. Since MIT-BIH was used for model development, we have detailed the dataset refinement process in the MIT-BIH dataset. For the other six external test datasets, detailed descriptions of each dataset are provided in the appendix material.

2.3.3. Evaluation metrics

To evaluate the PVC detection performance of the model, we used three metrics: *sensitivity*, *precision,* and *F1-score*. These metrics are suitable for imbalanced datasets and are commonly used in the literature on PVC detection. *Sensitivity* is the percentage of true positives in positive samples. Therefore, the recall rate also indicates the accuracy of available samples for examination. If there are false negatives, recall rate decreases. The formula is shown in Equation (2).

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. *Precision* indicates the accuracy of the available

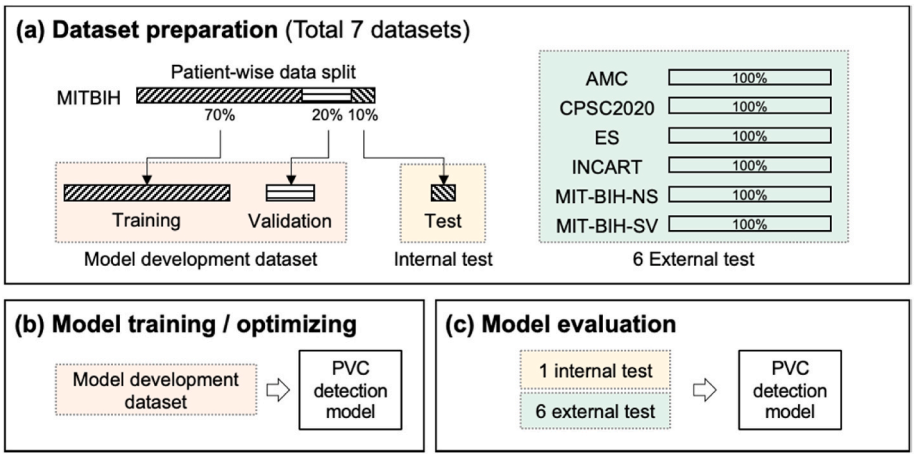

Fig. 3. Overview of experiment model evaluation protocols. Seven datasets were collected. Among them, the MIT-BIH dataset was used to training our model and for internal evaluation. The remaining six datasets were all used for external test evaluation.

Table 1
Descriptions of model development and external test datasets.

Type	Dataset	Lead	Recordings	Sampling rate	No. beats	No. PVC	PVC rate (%)
Training	MIT-BIH	II	30	360	69,764	5971	8.6
Validation		II	8	360	18,612	373	2.0
Internal Test		II	6	360	13,022	1362	10.5
Overall					101,398	7706	7.6
External Test	1) AMC-LT	II	436	125	12,535	784	6.3
	2) CPSC2020	II	10	400	895,711	42,075	4.7
	3) ES	V4	90	250	789,454	4821	0.6
	4) INCART	II	75	257	170,636	20,230	11.9
	5) MIT-BIH-NS	II	2	360	12,114	2766	22.8
	6) MIT-BIH-SV	II	78	128	172,304	9966	5.8
Overall					2,052,754	80,642	3.9

Note: Massachusetts Institute of Technology and Beth Israel Hospital Arrhythmia database, MIT-BIH; Asan Medical Center Liver Transplant database, AMC-LT; the 3rd China Physiological Signal Challenge 2020 database, CPSC2020; European ST-T database, ES; St. Petersburg Institute of Cardiological Technics database, INCART; MIT-BIH Noise Stress Test Database, MIT-BIH-NS; MIT-BIH Supraventricular Arrhythmia Database, MIT-BIH-SV.

Table 2
Comparison of PVC detection performance with the attention module for a total of seven test datasets (one internal test dataset: MIT-BIH, six external test datasets: AMC-LT, CPSC2020, ES, INCART, MIT-BIH-NS, and MIT-BIH-SV). Average of five random runs are reported for each setting with standard deviation.

Metric	Encoder	Internal test	External test dataset					
		MIT-BIH	AMC-LT	CPSC2020	ES	INCART	MIT-BIH-NS	MIT-BIH-SV
F1-score (primary metric)	VGG19	96.1 ± 1.2	84.8 ± 1.0	83.8 ± 1.5	84.3 ± 3.4	93.3 ± 1.3	90.9 ± 0.4	78.7 ± 1.0
	VGG19+CBAM	98.4 ± 1.1	84.3 ± 1.1	**90.1** ± 1.2	86.3 ± 3.3	94.2 ± 1.3	89.5 ± 0.4	79.1 ± 1.0
	VGG19+NLNN	99.0 ± 1.1	86.0 ± 1.2	78.8 ± 1.3	84.3 ± 3.6	94.2 ± 1.4	89.4 ± 0.3	81.6 ± 1.1
	VGG19 + SE	**99.1** ± 1.1	87.9 ± 1.1	81.3 ± 1.2	88.6 ± 3.3	**95.0** ± 1.3	90.3 ± 0.4	81.4 ± 1.0
	VGG19 + ACM	98.7 ± 1.2	85.2 ± 1.2	88.6 ± 1.2	83.2 ± 3.5	94.0 ± 1.2	92.6 ± 0.5	81.2 ± 1.2
	VGG19+DCAM (Ours)	99.0 ± 1.2	**88.1** ± 0.9	89.6 ± 1.3	**89.6** ± 3.2	94.8 ± 1.1	**91.2** ± 0.3	**82.2** ± 1.1
Sensitivity	VGG19	92.6 ± 3.3	79.7 ± 2.4	78.7 ± 2.3	78.2 ± 4.5	88.5 ± 1.2	87.1 ± 1.7	77.2 ± 3.6
	VGG19+CBAM	97.0 ± 3.4	79.3 ± 2.3	88.2 ± 2.4	81.1 ± 4.3	89.9 ± 1.3	85.3 ± 1.6	69.2 ± 3.4
	VGG19+NLNN	98.1 ± 3.4	81.6 ± 2.3	75.1 ± 2.6	81.0 ± 4.4	90.0 ± 1.2	84.5 ± 1.6	75.9 ± 3.4
	VGG19 + SE	**98.3** ± 3.5	85.7 ± 2.3	77.1 ± 2.5	84.3 ± 4.2	**91.2** ± 1.3	86.6 ± 1.7	82.1 ± 3.5
	VGG19 + ACM	97.6 ± 3.4	**86.2** ± 2.4	**94.0** ± 2.5	**89.9** ± 4.5	89.9 ± 1.4	**89.7** ± 1.6	**87.0** ± 3.6
	VGG19+DCAM (Ours)	98.2 ± 3.4	85.3 ± 2.4	93.8 ± 2.4	84.8 ± 4.2	91.0 ± 1.3	88.8 ± 1.6	85.7 ± 3.5
Precision	VGG19	99.9 ± 0.1	90.8 ± 2.5	**89.5** ± 4.6	90.4 ± 2.1	98.6 ± 0.1	95.0 ± 3.4	82.4 ± 4.3
	VGG19+CBAM	99.9 ± 0.1	90.1 ± 2.5	92.1 ± 4.3	92.1 ± 2.0	98.9 ± 0.1	94.2 ± 3.4	**92.4** ± 4.3
	VGG19+NLNN	99.9 ± 0.1	90.8 ± 2.7	82.9 ± 4.5	87.8 ± 2.2	98.8 ± 0.1	94.9 ± 3.5	88.2 ± 4.6
	VGG19 + SE	99.9 ± 0.1	90.2 ± 2.6	85.4 ± 4.4	93.4 ± 2.1	**99.2** ± 0.1	94.3 ± 3.6	80.7 ± 4.4
	VGG19 + ACM	99.9 ± 0.1	84.2 ± 2.7	83.8 ± 4.7	80.4 ± 2.6	98.6 ± 0.1	95.7 ± 3.5	76.1 ± 4.6
	VGG19+DCAM (Ours)	**99.9** ± 0.1	**92.1** ± 2.7	85.7 ± 4.5	**95.0** ± 2.1	99.0 ± 0.1	**96.5** ± 3.8	78.8 ± 4.5

The best results are highlighted in bold.

samples for examination. If there are many false positives, the precision rate will decrease. The formula is given in Equation (3).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

We utilized the *F1-score*, which is the harmonic mean of *sensitivity* and *precision* in a single value, as a primary metric of model performance balance. The formula is shown in Equation (4).

$$F1 - score = \frac{TP + TN}{TP + FN + FP + TN} \tag{4}$$

To compare the performances of models impartially, we trained them with a fixed seed and configured our model to be deterministic so that it always outputs the same value for the same input. We repeated the experiment five times with variations of the random seed and recorded the mean and standard deviation of the performance metrics.

3. Result

3.1. Effect of DCAME

We compared the change in PVC detection performance with and without DCAM in the PVC detection model. Table 2 shows the comparison of PVC detection performance in a 1D U-Net based on the VGG19 encoder with various attention modules across seven test datasets. The evaluated convolution-based state-of-the-art attention modules were CBAM, SE, NLNN, and ACM. All the attention modules attached to the VGG19 encoder improved PVC detection performance in our internal test dataset. When evaluating the performance of the PVC detection model when five attention modules are attached, including the baseline VGG19 encoder, VGG19+DCAM achieved the highest *F1-score* on four out of six external datasets. Furthermore, VGG19+DCAM is the only attentional module to improve *F1-score* on all seven test datasets, including all internal and external datasets, compared to VGG19. VGG19+DCAM improved the *F1-score* of MIT-BIH (+2.9), AMC-LT (+3.3), CPSC2020 (+5.8), ES (+5.3), INCART (+1.5), MIT-BIH-NS (+0.2), and MIT-BIH-SV (+3.5) compared to VGG19. On the other hand, other attention modules, including ACM, tend to increase the *F1-score*, but do not necessarily guarantee improved *F1-score* performance. For example, VGG19 + ACM improved the F1-score on AMC-LT (+0.4) and CPSC2020 (+4.8), however, it decreased the *F1-score* on ES (−1.1). VGG19 + ACM had a lower *F1-score* than VGG + DCAM, as it had a relatively higher *sensitivity* but worse *precision*. Note that VGG19 + ACM has removed the denoising process from VGG19+DCAM, and the recalibration has been modified, but the improvement in PVC detection performance in external tests has not been generalized. VGG19 had the lowest performance on all metrics and datasets, demonstrating the effectiveness of adding attention modules to improve PVC detection. However, the attention modules could not guarantee performance improvements in the external test. With the exception of DCAM, the attention module did not improve *F1-score* performance on all external datasets and was inconsistent. However, DCAM showed an improvement in *F1-score* on all datasets and an improvement in *sensitivity* and *precision* compared to those without the attention modules.

3.2. Ablation study

3.2.1. DCAM architecture

We evaluated DCAM with different module architectures. Table 3 shows the *F1-score* performance of module combinations for PVC detection on test datasets. In our tests, the best module configuration for the attention module was $P(f_{DM}(X) + K - Q)$, which is a f_{DM} followed by f_{CAM}. The results show that DCAM improved the performance of PVC detection models on unseen datasets, particularly CPSC2020 and ES, where PVC frequency was low and inter-patient variability and artifacts were high. Whereas $P(X+K-Q)$ shows that the performance of the baseline is worse on ES. This test shows that denoising is essential before applying the attention module to improve PVC detection performance reliably.

Recalibration also had an impact on performance, but recalibration alone did not consistently improve performance in our tests. The $P(X+K-Q)$ module achieves superior performance on four out of seven datasets and is comparable to the $(X+K-Q)$ module on the other three datasets. The $P(X+K-Q)$ module improved *F1-scores* on MIT-BIH (+1.0), CPSC2020 (+2.3), INCART (+0.7), and MIT-BIH-SV (+3.2), but ES performed approximately 1.4 worse than the model without the module.

3.2.2. Effect of G on GCNN in DCAM

In this study, we performed an ablation study to investigate the number of groups G in DCAM and the factors that influence the structure of DCAM. Table 4 shows the *F1-score* of PVC detection performance within the number of groups G on GCNN in DCAM with different encoder types. The results showed that when applying DCAM, the *F1-score* tended to increase and then decrease in performance as the G increased for both internal and external test datasets. The best performance is achieved by DCAM with the DenseNet121 encoder and G = 8, which has an *F1-score* of MIT-BIH (99.4), AMC-LT (86.5), CPSC2020 (89.9), ES (88.3), INCART (95.2), MIT-BIH-NS (92.4), and MIT-BIH-SV (81.6). For all encoders, the most optimal *F1-score* was obtained when G = 4 or G = 8, with no significant difference in performance between encoders. The optimal G value for DCAM that we found through experimentation was 8.

3.3. Qualitative result using attention map

To demonstrate the effectiveness of DCAM, we visualize an interest map comparing objects of interest and their context with CAM. CAM is a modification of the attention module based on ACM, which modified the recalibration structure for PVC detection tasks. The CAM learns to attend to different regions in a way that maximizes performance on the given task. Conversely, DCAM is an application of CAM with denoising based on FFT convolution. Rather than simply applying attention, DCAM is inspired by the process by which anesthesiologists first ignore artifacts in the sinus when reading a PVC. We visualized GTs and attention maps to see if DCAM produces an attention map that maximizes performance by highlighting interpretable beats.

Both DCAM and CAM learn to utilize the PVC region as the object of interest and the normal sinus rhythms region as its context; however,

Table 3
F1-score of PVC detection based on DCAM architectures.

Encoder	Module	Internal test			External test dataset			
		MIT-BIH	AMC-LT	CPSC2020	ES	INCART	MIT-BIH-NS	MIT-BIH-SV
VGG19	None	97.4 ± 1.2	86.2 ± 1.0	77.5 ± 1.4	85.1 ± 3.4	92.6 ± 1.3	90.9 ± 0.4	78.7 ± 1.0
	X + K − Q	98.2 ± 1.3	85.9 ± 0.9	79.6 ± 1.2	84.7 ± 3.2	93.2 ± 1.2	91.2 ± 0.5	81.6 ± 1.1
	P(X + K − Q) = f_{CAM}	98.4 ± 1.3	86.2 ± 1.1	79.8 ± 1.3	83.7 ± 3.3	93.1 ± 1.3	91.6 ± 0.4	81.9 ± 1.2
	f_{DM}(X) + K − Q	97.8 ± 1.2	86.1 ± 1.0	83.6 ± 1.3	84.9 ± 3.3	94.4 ± 1.1	**92.6** ± 0.3	82.2 ± 1.0
	P(f_{DM}(X) + K − Q) = f_{DCAM}	**99.0** ± 1.2	**88.1** ± 0.9	**89.6** ± 1.3	**89.6** ± 3.2	**94.8** ± 1.1	91.2 ± 0.3	**82.2** ± 1.1

The best results are highlighted in bold.

Table 4
F1-score of PVC detection within number of group convolution in DCAM with encoder type.

Encoder	No. GCNN (G)	Internal test	External test dataset					
		MIT-BIH	AMC-LT	CPSC2020	ES	INCART	MIT-BIH-NS	MIT-BIH-SV
VGG19+DCAM	None	97.4 ± 1.2	86.2 ± 1.0	77.5 ± 1.5	85.1 ± 3.4	92.6 ± 1.3	90.9 ± 0.4	78.7 ± 1.0
	4	98.2 ± 1.0	87.4 ± 0.9	84.5 ± 1.3	88.4 ± 3.7	94.9 ± 1.0	91.2 ± 0.4	81.9 ± 1.0
	8	**99.0** ± 1.2	**88.1** ± 0.9	**89.6** ± 1.3	**89.6** ± 3.2	94.8 ± 1.1	**91.2** ± 0.3	**82.2** ± 1.1
	16	99.0 ± 1.2	88.1 ± 0.9	89.3 ± 1.3	89.1 ± 3.2	**95.2** ± 1.1	91.2 ± 0.3	82.0 ± 1.1
	32	98.3 ± 1.1	85.5 ± 1.0	80.5 ± 1.4	86.5 ± 3.2	94.4 ± 1.2	90.6 ± 0.4	79.0 ± 1.2
ResNet34+DCAM	None	97.0 ± 1.3	84.3 ± 0.8	85.4 ± 1.1	80.9 ± 3.6	91.9 ± 1.0	91.0 ± 0.4	76.6 ± 1.1
	4	97.7 ± 1.3	86.3 ± 1.1	86.3 ± 1.2	84.6 ± 3.3	93.5 ± 1.0	91.5 ± 0.3	78.6 ± 1.2
	8	97.8 ± 1.2	**87.5** ± 0.9	**90.0** ± 1.2	84.9 ± 3.4	94.4 ± 1.0	**92.6** ± 0.4	**82.2** ± 1.1
	16	98.2 ± 1.3	85.7 ± 1.0	82.7 ± 1.4	**87.8** ± 3.5	92.8 ± 1.1	91.1 ± 0.5	79.1 ± 1.2
	32	**98.4** ± 1.3	86.2 ± 0.9	81.7 ± 1.3	87.4 ± 3.3	**94.8** ± 0.9	92.0 ± 0.4	80.5 ± 1.1
DenseNet121+DCAM	None	96.9 ± 1.1	85.1 ± 1.0	84.4 ± 1.2	86.8 ± 3.5	93.6 ± 1.2	90.8 ± 0.4	77.9 ± 1.1
	4	98.6 ± 1.2	86.0 ± 0.9	87.9 ± 1.2	86.9 ± 3.4	93.3 ± 1.1	92.2 ± 0.4	81.4 ± 1.0
	8	**99.4** ± 1.1	86.5 ± 1.0	**89.9** ± 1.1	88.3 ± 3.5	95.2 ± 1.2	**92.4** ± 0.3	**81.6** ± 1.1
	16	97.1 ± 1.1	**86.7** ± 1.1	79.9 ± 1.2	**88.4** ± 3.4	**95.8** ± 1.2	91.1 ± 0.3	73.3 ± 1.0
	32	99.3 ± 1.2	86.2 ± 1.2	87.8 ± 1.3	86.4 ± 3.6	95.3 ± 1.1	91.6 ± 0.4	78.9 ± 1.1

The best results are highlighted in bold.

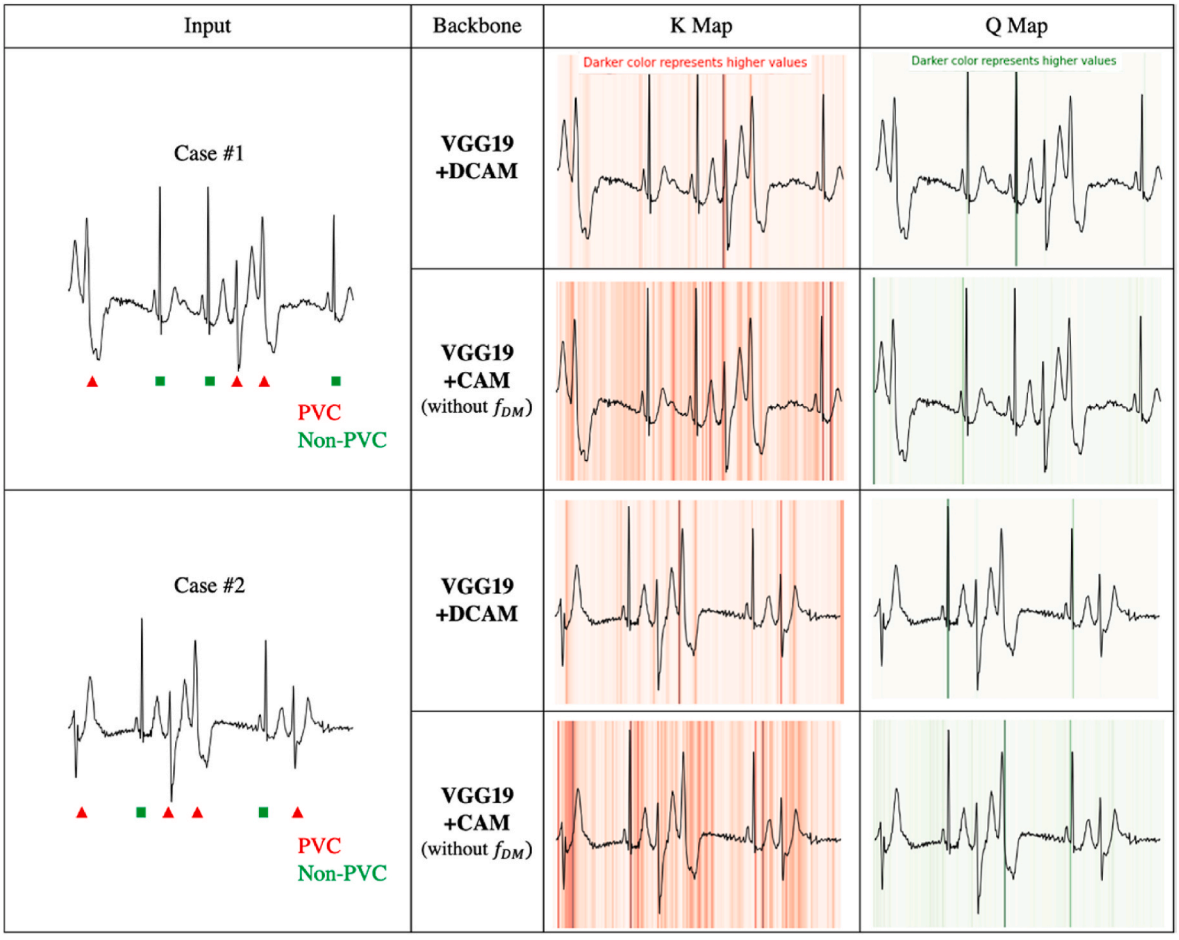

Fig. 4. Visualized attention maps for the PVC detection task. The 3rd layer of VGG19+DCAM was chosen. In GT, the red triangle indicates the location of the PVC. The attention map is darker with a larger gradient, the K map focuses on candidate PVCs, and the Q map focuses on normal sinus rhythms.

DCAM's attentional map is more precisely localized. The attention maps with DCAM and CAM applied for the two cases are visualized in Fig. 4. As shown in cases #1 and #2, CAM extracts PVC candidates from a wide variety of locations in K, while DCAM attaches K maps to relatively precise PVC locations. Furthermore, both CAM and DCAM detect normal sinus rhythms in the Q map, but CAM is less accurate than DCAM. This observation is consistent with our intuition that PVC detection can be more clearly distinguished by contrast after denoising, which DCAM learns automatically. This clearly demonstrates that the context of PVC and non-PVC is different after denoising, which aligns well with the design principles of DCAM.

Table 5
Comparing performance to previous studies that externally tested PVC detection.

External test dataset	Train dataset	Author	F1-score	Sensitivity	Precision
MIT-BIH-SV	MDT (private)	1	79.0	–	–
	MIT-BIH	2	85.0	85.4	84.50
		3 (ours)	82.2	85.7	78.8
CPSC2020	MIT-BIH	2	–	80.7	–
		3 (ours)	89.6	93.8	85.7
ES	MDT (private)	1	80.0	–	–
	MIT-BIH	3 (ours)	89.6	84.8	95.0
INCART	MDT (private)	1	69.0	–	–
	MIT-BIH	3 (ours)	94.8	91.0	99.0

The best results are highlighted in bold.
Note: [1]Ivora et al.; [2]Petryshak et al.; [3]Proposed model (VGG19+DCAM).

3.4. Performance comparison with previous studies

We compared three studies, including our own, in which an external model evaluation of PVC detection models was conducted, as shown in Table 5. These studies all used 1D U-Net to detect PVCs, and presented performance tests with datasets other than the trained data, with the datasets and evaluation metrics listed in each case for comparison. The proposed model (VGG19+DCAM) outperforms the other methods on most datasets and metrics. Our model achieves the highest *F1-score* on the CPSC2020 (89.1), ES (88.7), and INCART (94.8) datasets, and the second highest *F1-score* on the MIT-SV dataset (81.0). The proposed model also achieves the highest *sensitivity* on the CPSC2020 (90.6), ES (90.4), and MIT-SV (86.1) datasets and the second highest *sensitivity* on the INCART dataset (84.1). The proposed model also achieves the highest *precision* on the ES dataset (97.9), and the second highest precision on the CPSC2020 (85.2) and MIT-SV (81.9) datasets. Consequently, the proposed model (VGG19+DCAM) is a reliable and robust attention module for PVC detection in ECG signals and can generalize well to different datasets with varying characteristics and noise levels.

4. Discussion

We present a novel DCAM that can enhance the performance of CNN-based PVC detection algorithms on unseen ECG lead II signals. To demonstrate the effectiveness of DCAM, we evaluated the performance of the PVC detection model on six external test datasets. The results showed that DCAM improved the external test *F1-score* over baseline for all three encoder structures of the segmentation model used in the test. This can be interpreted as extracting representative features from signals where PVCs are difficult to detect, as an algorithm that mimics the principles by which experts detect PVCs.

The PVC detection model with DCAM has achieved *F1-scores* above 80 in all external test datasets, which is remarkably high considering the external evaluation. The datasets with low *F1-scores*, such as AMC-LT, ES, and MIT-BIH-SV, were challenging, the data contained many artifacts, and the PVC prevalence was very low. Particularly, the AMC-LT dataset contains several AFs that are morphologically similar to PVCs, as well as noise and artifacts that can occur in the operating room. It is very encouraging that even with these signals, DCAM was able to reliably increase the detection performance of the deep learning model. However, some of the state-of-the-art attention modules we evaluated did not show a trend toward improved performance on all external test datasets, although they did improve on internal test dataset. According to Table 2, one should exercise caution when applying the attention module to deep learning models since it can lead to overfitting internal dataset features.

The impact of the denoising module f_{DM} applied prior to difference attentions is substantial (Table 3). It is evident that our proposed DCAM surpasses the baseline model on unseen datasets. The performance of f_{DM} +K-Q was better than f_{CAM} with recalibration P, and additionally, the *F1-score* in external validation was improved in DCAM with recalibration P. Also, our ablation study found G = 8 to be optimal. It's a common observation in various tasks that increasing the Group parameter to a reasonable extent enhances performance, but any further increase may degrade it. Interestingly, regardless of the encoder type used among three different types tested, G = 8 was generally found optimal. Among these, VGG19 as an encoder yielded the most distinct results. This can be speculated as due to VGG19 not having skip connections in its encoder structure unlike ResNet34 and DenseNet121, which could lead to more consistent outcomes.

To demonstrate the effectiveness of the denoise process in DCAM, we utilized attention maps with and without f_{DM} to visualize where K and Q pay attention. We were able to get a more intuitive understanding by visualizing the attention map of DCAM in the fourth layer of the encoder (Fig. 4.). Without denoising, both K and Q show activity in multiple attentional regions. With DCAM, however, attention is concentrated in the R peak within the attentional map, resulting in a sharp contrast between K and Q. Based on these visualizations, it can be interpreted that performance can be improved if frequency-based convolution is applied before the attention mechanism is triggered.

The main goal of this research was to develop and validate an accurate and robust PVC detection model. Accurate detection of PVCs is crucial to proactively diagnose and prescribe medications for patients with cardiac arrhythmias [2]. Despite deep learning models used for PVC detection often demonstrating high performance on internal test datasets, they exhibit limited robustness and balance when applied to external datasets. For these models to be clinically viable, they must perform well on unseen datasets that encapsulate the true variability and complexity inherent in ECG signals [13]. This highlights the necessity for an evaluation protocol ensuring fair assessment across diverse data sources.

In the busy clinical setting of the operating theatre, early recognition and timely intervention for arrhythmia will enhance patient outcome no doubt. For example, we previously reported the incidence of atrioventricular block and atrial fibrillation as 5.0% and 1.3%, respectively, during liver transplantation surgery [57,58]. Although the incidence seems relatively low, their prognosis was significantly poor. In this occasion, AI powered surveillance systems may efficiently screen the potential vulnerable patients, eventually facilitating resource redistribution of hospitals and enhancing the patient's well-being. We believe this is a very important point of view because many healthcare providers complain about burnout who receive endless requests for vigilant attention and labor-intensive care from their clinical settings. Inspired by the decision method of trained clinicians, we propose a model that mimics the way human experts diagnose PVC, such as ignoring noise and context recognition. This denoising process differs from traditional denoising method with FFTs in that it is trainable. Recognizing the context implies focusing on differences in the morphologies and intervals of the remaining beats, which detect PVC with high accuracy even on new, unseen data. In this study, we discovered that DCAM is instrumental in developing a balanced PVC detection model. Simultaneously increasing *sensitivity* and *precision* is a challenging task due to their tradeoff, yet our DCAM enhanced the *F1-score* on unseen datasets, likely owing to generalized representation learning. The improvement of PVC detection *F1-score* using DCAM will significantly contribute to detecting cardiac diseases at their early stages.

This study has some limitations that need to be acknowledged. First, our DCAM did not sufficiently detect PVC in noise and artifacts

contained in ECGs. Our clinician confirmed that the *sensitivity* performance increased by 3.2 if noisy signals were ignored. The results also show that the performance varies across external test datasets, which have different characteristics such as PVC prevalence, noise level, and other arrhythmia symptoms. For example, CPSC2020 has the lowest *F1-score* among all datasets, which contains many artifacts as it is a 24-h Holter ECG dataset. However, INCART has the highest *F1-score* among the external datasets, because the dataset is very well refined. Second, we have demonstrated that DCAM improves PVC detection performance in six external test datasets, but we have not been able to validate it in other tasks. Relatively speaking, the detection of PVCs can be easily determined by comparing the surrounding signals, which needs to be verified in other more challenging tests in the future. Third, due to resource constraints, we were unable to perform a wide range of tests and extensively optimize model parameters. We used the 1D U-Net as the base model to evaluate DCAM and different attention modules under limited conditions. We were not able to study the optimization of PVC detection performance based on the location and number of attention modules. In addition, the state-of-the-art attention modules used the optimal values suggested by an existing study and are not optimized for our task. Fourth, our model cannot detect arrhythmic syndromes other than PVC in detail. The focus of our study was to accurately diagnose PVC and we did not address other disorders. However, it would be desirable to study non-PVC indicators along with PVC symptoms to detect PVC more accurately and comprehensively.

5. Conclusion

In this paper, we proposed a novel DCAM that enhances the performance of a 1D U-Net for detecting PVCs in unseen electrocardiogram signals. The DCAM consists of: a CNN-based trainable denoising module that filters out noise from the input signals, and a contrast attention module that focuses on the QRS morphologies and captures the contrast between PVCs and normal sinus rhythms. We evaluated our DCAM on six unseen datasets with different encoder types and compared it with existing attention methods. Our results showed that our DCAM achieved superior *F1-score* performance across all external test datasets and encoders and demonstrated that the attention module should be denoised prior to application to robustly improve detection performance. Our DCAM is expected to contribute to the advancement of deep learning models for cardiac arrhythmia detection and diagnosis, as well as other biomedical signal processing applications.

Funding

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR20C0026).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.107532.

References

[1] R.J. Simpson Jr., W.E. Cascio, P.J. Schreiner, R.S. Crow, P.M. Rautaharju, G. Heiss, Prevalence of premature ventricular contractions in a population of African American and white men and women: the Atherosclerosis Risk in Communities (ARIC) study, Am. Heart J. 143 (2002) 535–540.
[2] C.H. Kwon, S.-H. Kim, Intraoperative management of critical arrhythmia, Korean Journal of Anesthesiology 70 (2017) 120–126.
[3] D. Muser, P. Santangeli, S.A. Castro, R. Casado Arroyo, S. Maeda, D.A. Benhayon, I. Liuba, J.J. Liang, M.M. Sadek, A. Chahal, Risk stratification of patients with apparently idiopathic premature ventricular contractions: a multicenter international CMR registry, Clinical Electrophysiology 6 (2020) 722–735.
[4] U. Meyerfeldt, A. Schirdewan, M. Wiedemann, H. Schütt, F. Zimmerman, F. Luft, R. Dietz, The mode of onset of ventricular tachycardia: a patient-specific phenomenon, Eur. Heart J. 18 (1997) 1956–1965.
[5] R.J. Myerburg, Sudden cardiac death: epidemiology, causes, and mechanisms, Cardiology 74 (1987) 2–9.
[6] I.S. Murthy, M.R. Rangaraj, New concepts for PVC detection, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. (1979) 409–416.
[7] Y. Jung, H. Kim, Detection of PVC by using a wavelet-based statistical ECG monitoring procedure, Biomed. Signal Process Control 36 (2017) 176–182.
[8] M.L. Talbi, P. Ravier, Detection of PVC in ECG signals using fractional linear prediction, Biomed. Signal Process Control 23 (2016) 42–51.
[9] R.C.-H. Chang, C.-H. Lin, M.-F. Wei, K.-H. Lin, S.-R. Chen, High-precision real-time premature ventricular contraction (PVC) detection system based on wavelet transform, Journal of Signal Processing Systems 77 (2014) 289–296.
[10] K. Yasin, Classification of PVC beat in ECG using basic temporal features, Balkan Journal of Electrical and Computer Engineering 6 (2018) 78–82.
[11] W.H. Chang, K.-P. Lin, S.-Y. Tseng, ECG analysis based on Hilbert transform descriptor, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 1988, pp. 36–37.
[12] A. Pachauri, M. Bhuyan, Wavelet and energy based approach for PVC detection, in: 2009 International Conference on Emerging Trends in Electronic and Photonic Devices & Systems, IEEE, 2009, pp. 258–261.
[13] K.J. Ruskin, C. Corvin, S.C. Rice, S.R. Winter, Autopilots in the operating room: safe use of automated medical technology, Anesthesiology 133 (2020) 653–665.
[14] Y. Cao, W. Liu, S. Zhang, L. Xu, B. Zhu, H. Cui, N. Geng, H. Han, S.E. Greenwald, Detection and localization of myocardial infarction based on multi-scale ResNet and attention mechanism, Front. Physiol. (2022) 24.
[15] F. Murat, O. Yildirim, M. Talo, U.B. Baloglu, Y. Demir, U.R. Acharya, Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review, Comput. Biol. Med. 120 (2020), 103726.
[16] Z. Ebrahimi, M. Loni, M. Daneshtalab, A. Gharehbaghi, A review on deep learning methods for ECG arrhythmia classification, Expert Syst. Appl. X 7 (2020), 100033.
[17] F.-y. Zhou, L.-p. Jin, J. Dong, Premature ventricular contraction detection combining deep neural networks and rules inference, Artif. Intell. Med. 79 (2017) 42–51.
[18] L.-H. Wang, L.-J. Ding, C.-X. Xie, S.-Y. Jiang, I.-C. Kuo, X.-K. Wang, J. Gao, P.-C. Huang, P.A.R. Abu, Automated classification model with OTSU and CNN method for premature ventricular contraction detection, IEEE Access 9 (2021) 156581–156591.
[19] M. Naz, J.H. Shah, M.A. Khan, M. Sharif, M. Raza, R. Damaševičius, From ECG signals to images: a transformation based approach for deep learning, PeerJ Computer Science 7 (2021) e386.
[20] H. Ullah, M.B.B. Heyat, F. Akhtar, A.Y. Muaad, C.C. Ukwuoma, M. Bilal, M. H. Miraz, M.A.S. Bhuiyan, K. Wu, R. Damaševičius, An automatic premature ventricular contraction recognition system based on imbalanced dataset and pre-trained residual network using transfer learning on ECG signal, Diagnostics 13 (2022) 87.
[21] S.L. Oh, E.Y. Ng, R. San Tan, U.R. Acharya, Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types, Comput. Biol. Med. 105 (2019) 92–101.
[22] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[23] R. Hu, J. Chen, L.J.C.i.B. Zhou, Medicine, A transformer-based deep neural network for arrhythmia detection using continuous ECG signals 144 (2022), 105325.
[24] B. Petryshak, I. Kachko, M. Maksymenko, O.J.T. Dobosevych, H. Care, Robust deep learning pipeline for PVC beats localization 29 (2021) 475–486.
[25] A. Ivora, I. Viscor, P. Nejedly, R. Smisek, Z. Koscova, V. Bulkova, J. Halamek, P. Jurak, F.J.S.R. Plesinger, QRS detection and classification in Holter ECG data in one inference step 12 (2022), 12641.
[26] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.
[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[28] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[29] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.
[30] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[31] S.-A. Park, H.-C. Lee, C.-W. Jung, H.-L. Yang, Attention mechanisms for physiological signal deep learning: which attention should we take?, in: International Conference on Medical Image Computing and Computer-Assisted Intervention Springer, 2022, pp. 613–622.
[32] R. Gu, J. Zhang, R. Huang, W. Lei, G. Wang, S. Zhang, Domain composition and attention for unseen-domain generalizable medical image segmentation, in:

Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Springer, Strasbourg, France, 2021, pp. 241–250. September 27–October 1, 2021, Proceedings, Part III 24.
[33] H. Lai, Y. Luo, B. Li, G. Zhang, J. Lu, Domain-aware dual attention for generalized medical image segmentation on unseen domains, IEEE Journal of Biomedical and Health Informatics 27 (2023) 2239–2410.
[34] S.M. Ryu, K. Shin, S.W. Shin, S.H. Lee, S.M. Seo, S.-u. Cheon, S.-A. Ryu, J.-S. Kim, S. Ji, N. Kim, Automated landmark identification for diagnosis of the deformity using a cascade convolutional neural network (FlatNet) on weight-bearing lateral radiographs of the foot, Comput. Biol. Med. 148 (2022), 105914.
[35] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." *Communications of the ACM* 60.6 (2017): 84-90.A. Krizhevsky, I. Sutskever, G.E.J.C.o.t.A. Hinton, Imagenet classification with deep convolutional neural networks 60 (2017) 84–90.
[36] X. Mao, Y. Liu, F. Liu, Q. Li, W. Shen, Y.J.a.e.-p. Wang, Intriguing Findings of Frequency Selection for Image Deblurring, 2021 arXiv: 2111.11745.
[37] Chi, Lu, Borui Jiang, and Yadong Mu. "Fast fourier convolution." *Advances in Neural Information Processing Systems* 33 (2020): 4479-4488.L. Chi, B. Jiang, Y.J.A.i.N.I.P. S. Mu, Fast fourier convolution 33 (2020) 4479–4488.
[38] M. Kim, J. Park, S. Na, C.M. Park, D. Yoo, Learning Visual Context by Comparison, European Conference on Computer Vision, Springer, 2020, pp. 576–592.
[39] Berkaya, Selcan Kaplan, Alper Kursat Uysal, Efnan Sora Gunal, Semih Ergin, Serkan Gunal, Bilginer Gulmezoglu. "A survey on ECG analysis, Biomedical Signal Processing and Control 43 (2018) 216–235.
[40] Manuel Blanco-Velasco, Binwei Weng, E. Barner Kenneth, ECG signal denoising and baseline wander correction based on the empirical mode decomposition, Computers in biology and medicine 38 (1) (2008) 1–13.
[41] Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, Deep fisher networks for large-scale image classification, in: Advances in neural information processing systems, 26, 2013.
[42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[43] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[44] A.L. Goldberger, Z.D. Goldberger, A. Shvilkin, Clinical Electrocardiography: a Simplified Approach E-Book, Elsevier Health Sciences, 2017.
[45] D. Makowski, T. Pham, Z.J. Lau, J.C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, S. Chen, NeuroKit2: a Python toolbox for neurophysiological signal processing, Behav. Res. Methods 53 (2021) 1689–1696.
[46] C.-S.J. Chu, Time series segmentation: a sliding window approach, Inf. Sci. 85 (1995) 147–173.
[47] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (2001) 45–50.
[48] N. Association, For the Advancement of Medical Instrumentation, Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, ANSI/AAMI EC38 1998 (1998) 46.
[49] J. Oster, J. Behar, O. Sayadi, S. Nemati, A.E. Johnson, G.D. Clifford, Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 62 (2015) 2125–2134.
[50] M. Llamedo, J.P. Martínez, Heartbeat classification using feature selection driven by database generalization criteria, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 58 (2010) 616–625.
[51] P. Hamilton, Open source ECG analysis, in: Computers in Cardiology, IEEE, 2002, pp. 101–104.
[52] Z. Cai, C. Liu, H. Gao, X. Wang, L. Zhao, Q. Shen, E. Ng, J. Li, An open-access long-term wearable ECG database for premature ventricular contractions and supraventricular premature beat detection, J. Med. Imaging Health Inform. 10 (2020) 2663–2667.
[53] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. Moody, C. Zeelenberg, C. Marchesi, The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography, Eur. Heart J. 13 (1992) 1164–1172.
[54] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J. E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220.
[55] G. Moody, W. Muldrow, R. Mark, The MIT-BIH noise stress test database, Comput. Cardiol. (1984) 381–384.
[56] S.D. Greenwald, R.S. Patil, R.G. Mark, Improved Detection and Classification of Arrhythmias in Noise-Corrupted Electrocardiograms Using Contextual Information, IEEE, 1990.
[57] S.H. Kim, Y.J. Moon, S. Lee, S.M. Jeong, J.G. Song, G.S. Hwang, Atrioventricular conduction disturbances immediately after hepatic graft reperfusion and their outcomes in patients undergoing liver transplantation, Liver Transplant. 22 (2016) 956–967.
[58] Y.-J. Moon, H.-M. Kwon, Y.-S. Park, S.-H. Kim, G.-S. Hwang, Brief Episodes of Newly Developed Intraoperative Atrial Fibrillation Predicts Worse Outcomes in Adult Liver Transplantation, Transplantation Proceedings, Elsevier, 2018, pp. 1142–1146.

Keewon Shin is a research professor at the Medical Device Research Platform, Korea University Anam Hospital, Korea University College of Medicine, Seoul. He received his Ph.D. degree from Ulsan National University in 2023. His main research interests include the clinical application of machine learning techniques to medical imaging and biosignal analysis to improve the quality of clinical diagnosis.

Hyunjung Kim is a master's student in Biomedical Engineering at the University of Ulsan in Seoul, Korea. She received a bachelor's degree in English linguistics and literature and Communication sciences. Also, she completed Communication sciences' courses for master's degree. Her research interests include large language models, in particular ontology, electroencephalogram analysis and brain imaging.

Woo-Young Seo is a postdoctoral researcher at the Biomedical Engineering Research Center of the Asan Medical Center. He received his PhD in the Department of Physics and Astronomy at Seoul National University, specializing in numerical simulations and large-scale data processing. Currently, his research focuses on analyzing, modeling, and utilizing deep learning techniques to process high-quality biomedical signals.

Hyun-Seok Kim received the Ph.D. degree in bioengineering from Seoul National University, Seoul, South Korea, in 2018. Since 2021, he has been working as a Postdoctoral Researcher with the Biomedical Engineering Research Center, Asan Medical Center, Seoul. His research interests include machine learning, computer-aided disease diagnosis, and biosignal processing.

Jae-Man Shin received the M.S. degree from Sungkyunkwan University of Department of Mathematics, Suwon, Korea, in 2021. He is currently a researcher at the Department of Anesthesiology and Pain Medicine, Asan Medical Center, Seoul, Korea. He is also a Ph.D student majoring in Biomedical Engineering at University of Ulsan College of Medicine, Seoul, Korea.

Dong-Kyu Kim received the M.S. degree from the Department of Brain and Cognitive Engineering at Korea University, Seoul, Korea in 2022. He is currently an AI researcher at Department of Anesthesiology and Pain Medicine, Asan Medical Center, Seoul, Korea. His research area includes medical AI and bio-signal analysis.

Yong-Seok Park received the M.D. degree from University of Ulsan College of Medicine, Seoul, Korea in 2009, and M.Sc. degree from University of Ulsan College of Medicine, Seoul, Korea in 2013. He is currently a Clinical Assistant Professor at the Department of Anesthesiology and Pain Medicine, Asan Medical Center, Seoul. His research area includes general anesthesia, neuroanesthesia, and bio-signal analysis.

Sung-Hoon Kim received the M.D. degree from Hanyang University of School of Medicine, Seoul, Korea, in 2005, and Ph.D. degree from University of Ulsan College of Medicine, Seoul, Korea, in 2013. He also studied Statistical Informatics and received B.Sc. degree from Korea National Open University in 2017. He is currently Associate Professor at Department of Anesthesiology and Pain Medicine, Asan Medical Center, Seoul, Korea. His research area includes human bio-signal analysis and anesthesia outcome.

Namkug Kim is a tenured full professor at University of Ulsan College of Medicine, and also holds joint appointments at Asan Medical Center (http://eng.amc.seoul.kr/), one of leading hospitals in South Korea. He is currently a dual appointed professor at the Department of Convergence Medicine and Radiology. He received his BS, MS, PhD degrees from the Department of Industrial Engineering at Seoul National University and is the author of more than 300 peer-reviewed original articles and 100 patents (https://scholar.google.com/citations?user=namkugkim). He is a vice president of the Korean Society of Artificial Intelligence in Medicine (KoSAIM) and general director of the Korean Society of 3D Printing in Medicine. He received commendations from the President, the Minister of Health and Welfare and the Minister of Food and Drug Safety of South Korea, etc. His research interests are the areas of convergence medicine including deep learning, 3d printing, computer aided diagnosis, computer aided surgery, and robotic interventions, medical image processing, etc.