

Received 5 December 2023, accepted 20 December 2023, date of publication 25 December 2023,
date of current version 22 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346893

RESEARCH ARTICLE

Ultra-Efficient Edge Cardiac Disease Detection Towards Real-Time Precision Health

JUNHUA WONG¹, JEANNE NERBONNE², AND QINGXUE ZHANG^{1,3}, (Senior Member, IEEE)

¹Department of Electrical and Computer Engineering, School of Engineering and Technology, Purdue University, Indianapolis, IN 46202, USA

²School of Medicine, Washington University in St. Louis, St. Louis, MO 63110, USA

³Department of Biomedical Engineering, School of Engineering and Technology, Purdue University, Indianapolis, IN 46202, USA

Corresponding author: Qingxue Zhang (qxzhang@purdue.edu)

This work was supported in part by USA NSF CAREER Award under Grant 2047849.

ABSTRACT Nowadays, intensive interests are targeting the deep learning on edge precision health towards instantaneous disease measurements. However, edge inference usually has constrained computing resource, which poses a great challenge on running the heavy deep learning for real-time measurements. In this study, we propose to leverage a knowledge distillation methodology to enable ultra-efficient deep learning on edge. We take a special interest in Electrocardiogram (ECG)-based cardiac abnormality measurement. More specifically, we propose to train two deep learning models, including a heavy teacher model and a light-weight student model, and leverage the ‘soft target distribution’ learned by the teacher model to supervise the learning of the student model. So, the powerful teacher model can transfer learned knowledge to the student model and boost the latter’s accuracy. Further, to mitigate the vulnerability of the deep learning model under adversarial attacks, we further introduce preserving-robustness learning to the student model, without needing extra computing resources, through enhancing its loss function under adversarial perturbations. Our experiments on real-time heart disease measurement have demonstrated that, the learned lightweight student model, with a model reduction of 45x and under adversarial attacks, can still achieve comparable disease detection performance. The proposed robust knowledge distillation methodology has effectively enabled light-weight and secure cardiac measurement. Significance: This study is expected to contribute to on-edge deep learning-powered disease detection, for real-time, long term, and secured cardiac precision health.


INDEX TERMS Deep learning, edge inference, cardiac disease, real-time measurement.

I. INTRODUCTION

Technology have been greatly advancing emerging edge inference for physiological monitoring and data analytics [1], [2]. Wearable and edge devices have undergone rapid development in recent years and can measure important biomedical information about a person’s health condition. For instance, smart physical activity monitoring [3], [4], cardiac health monitoring, lifestyle management, and other practices, are attracting increasing attentions in health measurement and tracking [5], [6], [7], [8]. Meanwhile, advanced machine learning and deep learning algorithms can further analyze the captured biodynamics to generate medical insights and/or

send notifications or alerts to the users and the designated medical professionals if abnormalities are detected [9]. These physiological monitoring applications, if equipped with the deep learning ability, could save the medical professionals’ time through automatic data analytics and could also minimize the manual inspection errors [10], [11], [12]. So, it is promising to move the data analytics algorithms from the cloud or computer servers to be closer to the users, such as on smart phones, such that real-time, long-term big data mining and inference on edge can be achieved, as shown in Fig. 1.

Applied machine learning and deep learning algorithms have been attracting increasing attentions in biomedical detection, and interesting previous studies have been reported. For example, Song et al built a deep belief network-based approach for blood pressure estimation [13].

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang .

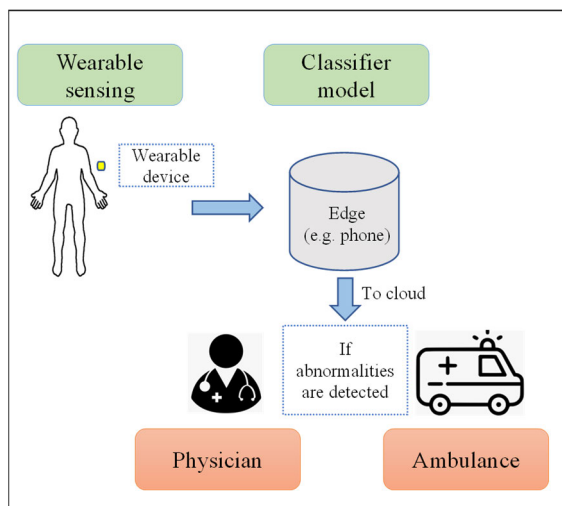


FIGURE 1. The application of AI on edge, e.g., executing classification tasks on smart phones, is expected to greatly advance real-time, long-term physiological monitoring applications. The abnormalities can be sent to the cloud, and the physician can check the data and/or the alarm will be sent to the emergency department.

Feng et al. developed an unsupervised feature learning method for sleep apnea detection [14]. Roy et al. have designed an AI approach for photoplethysmography processing with motion artifacts. Nuzzi et al. developed a deep learning-based hand gesture recognition system [15]. These studies have advanced applied artificial intelligence on biomedical applications.

Meanwhile, one challenge still lies in running the analytics algorithms, especially the heavy deep learning models, on resource-constrained edge inference, on the way towards real-time, long-term analytics of the biodynamics on the user side.

We take a special interest in the Electrocardiogram-based heart disease measurement system, to investigate the challenging edge-deployable deep learning algorithms. Cardiovascular diseases (CVDs) have become a leading cause of death in the world. One severe CVD is arrhythmia, an irregular rhythm of the heart that is coupled with many reasons like ventricular fibrillation, premature ventricular contraction, and tachycardia [16]. A person with arrhythmia may in the long-term face heart failure [17]. It is of vital importance to have early detection of the arrhythmia, such that the corresponding treatment can be prescribed to prevent deterioration of the condition. The arrhythmia condition disrupts the normal process of the heartbeat [18] as the heart undergoes depolarization and repolarization and thus could be detected by assessing the ECG signal. The procedure for measurement and diagnosis of arrhythmia primarily mainly consists of visual inspection of Electrocardiogram (ECG) signal by a clinical physician [19].

While the diagnosis procedure may have been in practice for decades, the manual inspection of ECG signals can take a substantial amount of time to diagnose arrhythmia and be

prone to human error. Especially, analyzing large amounts of ECG data can be very challenging for medical professionals. The traditional methods like the Holter monitor mainly provide ECG capturing [20], [21], [22]. Other methods use traditional signal processing methods to analyze the signal abnormalities [23], [24], [25] but may be limited to the data processing ability.

Machine learning and deep learning methods have also been reported for ECG signal processing on biomedical detection. Taji et al. has designed a deep belief network for atrial fibrillation detection [26]. Hammad et al. developed a combined approach with genetic algorithm-based feature selection and deep neural network-based classification for arrhythmia detection [27]. Acharya et al. presented a multi-layer model based on the deep convolutional neural network (CNN) for ECG-based heart disease classification [28]. Mostayed et al. proposed a long-short-term-memory (LSTM)-based model [29]. Teplizky et al. [30] proposed a deep neural network for automatic ECG annotation. Yeh et al. proposed a fuzzy c-means approach, a simple and reliable algorithm that performs classification of ECG signals by calculating the cluster centers for each class [31]. Li and Zhou introduced a method using wavelet packet decomposition (WPD) and random forests (RF) to perform classification task [32]. Sharma et al. presented a two-band optimal biorthogonal filter bank (FB) method and a k-nearest neighbor (KNN) classification algorithm [33]. Another approach was presented by Varatharajan et al. in which an enhanced Support Vector Machines (SVM) method with a weighted kernel function for ECG signals was proposed [34]. Another study [35] reported the CNN model designed and the pretrained deep learning models, which are used as ECG feature extractors for machine learning algorithms such as the support vector machine and decision tree. In one study [36], the CNN model with the empirical model decomposition-based ECG feature extractor was reported. These machine learning and deep learning methods have further advanced ECG-based heart disease detection. In our study, we take a special interest in the model efficiency towards edge inference.

Nevertheless, these methods usually target the performance and apply standard deep learning models. The deep learning models are usually heavy and difficult to be deployed for real-time ECG analysis. Some previous studies have been reported on efficient deep learning for ECG analysis. Efficient deep learning studies are usually on image processing, natural language processing, and autonomous driving [37], [38]. However, the efficient deep learning on ECG analysis is still very limited. Mahmud et al. [39] used varying convolution filter sizes in the convolutional neural network to reduce the model size, which needs substantial efforts on model block optimization. We in this study, target the real-time, long-term deep learning of ECG for efficient edge inference, without substantial efforts on detailed engineering. Instead, we target the architecture level direct learning of efficient models. More specifically, we leverage theoretical

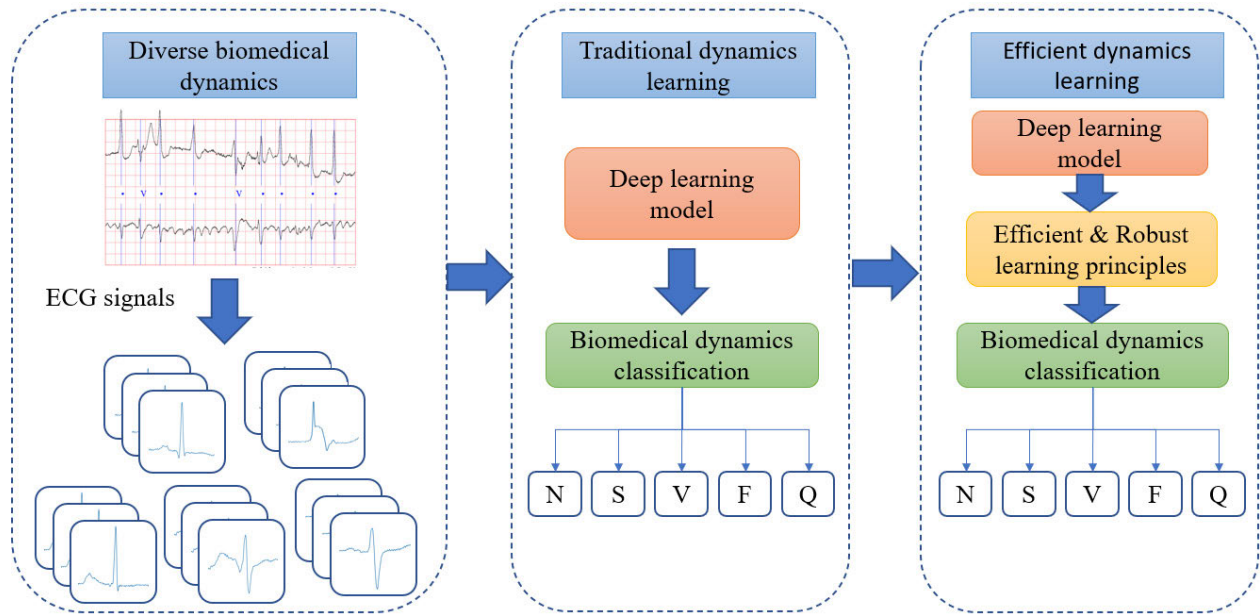


FIGURE 2. The design flow of the Edge Cardiac Inference System, with efficient and robust learning principles to achieve the edge-deployable and secure deep learning model. Note: N/S/V/F/Q – correspond to five categories of heartbeats to be detailed in results, according to the Advancement of Medical Instrumentation (AAMI) standard.

deep learning framework design and knowledge distillation, to achieve light-weight deep learning models with comparable accuracy, towards efficient deep learning models that are deployable on edge cardiac detection. One thing worth noting is that, the small models usually have lowered performance when implemented on the edge computing devices, due to the lower learning capability. But the knowledge distillation process aims to transfer the knowledge from the teacher to the student model, and thus the student model is expected to have similar performance as the heavy teacher model.

Besides, deep learning models on connected edge biomedical detection may be susceptible to adversarial attacks [40], [41], [42] that are in the form of small, imperceptible distortions. These attacks could fool the predictive models into misclassification, and lead to incorrect results in life-critical applications such as ECG classification. This may happen when the fake ECG data is sent to the edge device like the phone, and the phone just processes the data sent by the ECG device or the attacker. These could have a detrimental effect on a patient's health and well-being, due to the increased false negatives (the abnormality is misclassified as a normal condition) and increased false positives (the normal condition is misclassified as the abnormality). The former one may cause the missing severe heart abnormalities, and the latter one may increase the burden significantly to the medical resources. But current research in adversarial attacks on edge biomedical systems is still limited. In this study, we also investigate how to enhance the security without introducing additional hardware resources. This will be achieved by enhancing and securing the learning process which can see the adversarial

attacks [43], without changing the efficient deep learning architecture.

The precision medicine can thus be advanced with the efficient and secure edge biomedical detection, since the efficient inference can enable real-time and long-term health monitoring, and the secured inference without additional hardware resources can broaden the usage of edge inference towards pervasive deployment of the deep learning models on the edge.

This study is further motivated by the teacher and student learning strategy. The teacher model is a neural network, usually heavy, for complex pattern mining. The student model, usually a lightweight neural network, learns to follow the teacher output. It is unclear whether this strategy can facilitate the edge inference of biomedical signals like the ECG signal. Also, we are interested in effective deep learning model design, especially the deep network potentially enabled by the ResNet-like skip connections. Further, we are interested whether the algorithm framework can work effectively under both huge model size reduction and challenging adversarial attacks. Overall, to the best of our knowledge, this study aims to provide a first-time study on ECG-based heart disease inference on edge, considering the model size and robustness.

More specifically, in this study, targeting above two challenges – edge deployability and adversarial attacks, propose a knowledge distillation and robust learning methodology. It aims to enable both efficient and secure deep learning on edge biomedical applications, as Fig. 2, for real-time, and long-term cardiac abnormality detection, as well as other biomedical applications. More specifically, we propose to

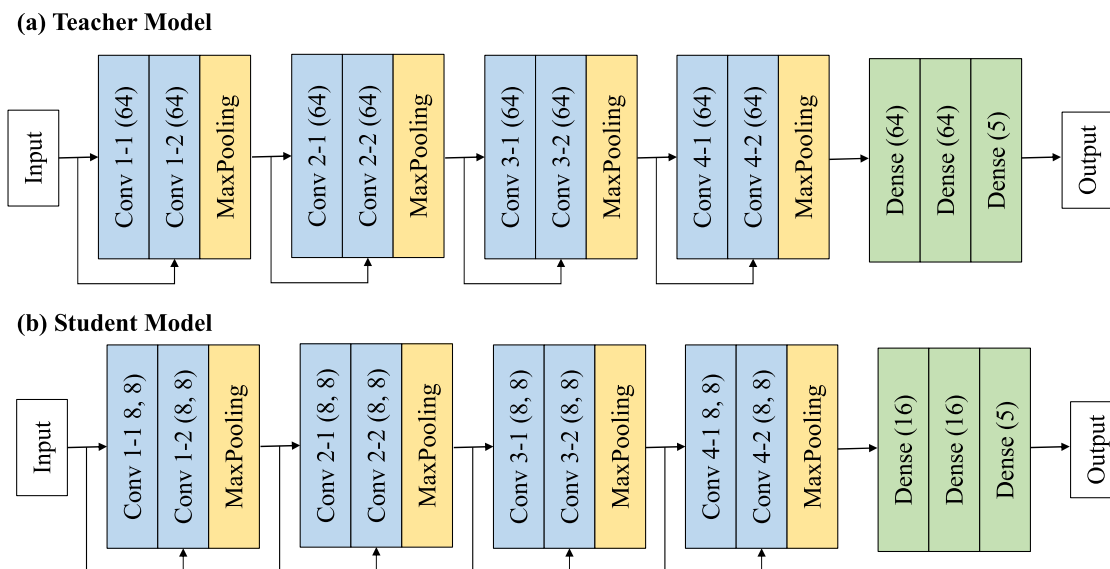


FIGURE 3. The ResNet teacher (a) and student (b) models. Note: the last layer has five output neurons, corresponding to the five ECG categories.

build and learn two deep learning models, including a heavy teacher model and a light-weight student model, and leverage the ‘soft target distribution’ learned by the teacher model to supervise the learning of the student model. So, the powerful teacher model can transfer learned knowledge to the student model and boost the latter’s accuracy. Further, to mitigate the vulnerability of the deep learning model under adversarial attacks, we further introduce preserving-robustness learning to the student model through enhancing its loss function under adversarial perturbations. Our experiments on real-time heart disease measurement have demonstrated that the proposed robust knowledge distillation methodology has effectively enabled real-time and secure cardiac measurement on edge.

Our key contributions include:

(a) Leverage the knowledge distillation learning (KDL) methodology to transfer the ‘soft target distribution’ learned by a heavy teacher model to a light-weight student model. So, the latter one, with a 45x reduction in deep learning model size compared with the teacher, can still achieve comparable accuracy, thereby enabling edge-deployable deep inference on edge biomedical systems.

(b) Leverage the preserving-robustness KDL (PR-KDL) approach to enhance the student model’s loss function and make it insensitive to adversarial ECG with imperceptible perturbations, thereby enhancing the security of connected edge biomedical systems.

(c) Conduct thorough experiments to demonstrate the effectiveness of the proposed preserving-robustness knowledge distillation framework on cardiac abnormality measurement, for the first time. This will greatly advance real-time, long-term, and secure cardiac disease detection on edge.

II. APPROACHES

A. SYSTEM OVERVIEW

The ECG signals are highly diverse for arrhythmia patients, considering the diverse biomedical abnormalities. The traditional deep learning methods usually need heavy models that have large amounts of parameters. With the design flow that has embedded efficient and robust learning principles, as shown in Fig. 2, the proposed efficient and robust dynamics learning framework, *i.e.*, robust knowledge distillation, can generate edge-deployable, light-weight, and secured (insensitive to adversarial attacks) deep learning models. More specifically, the step “efficient & robust learning principles” aims to enhance the learning process by, not only boosting the efficiency by knowledge distillation learning, but also improving the security by adversarial learning. This step is achieved by the enhanced loss function that takes into account both the knowledge from the teacher model and the input perturbations from the adversarial operations.

B. RESIDUAL CONVOLUTIONAL NEURAL NETWORK

We have firstly built a residual CNN – ResNet [44], [45], [46] for ECG-based multi-class heart disease classification, as shown in Fig. 3(a). CNN models typically have convolutional, pooling, fully connected layers. ResNet is inspired by neural cells with skip connections. Also, the skip connection can jump over some layers and then connect with the target layer. This can help suppress the problem of vanishing gradients and facilitate the learning process.

The CNN architecture with skip connections consists of 15 layers in Fig. 3(a), which is defined as the heavy teacher model. There are four stages in the convolution-pooling part, and each stage includes two convolutional layers, one max-pooling layer, as well as a skip connection. The convolution

layers aim to mine the spatial motifs in the ECG signal, and the max-pooling lay reduces the dimensionality and abstracts the patterns. The skip connection enhances the information flow in the learning structures. The final layer of the last stage is comprised of five neurons that correspond to the number of classes (to be detailed in the results section). The last 3 dense layers fuse and the learned patterns and yield the final ECG signal classification results.

Here the one-dimension raw ECG signal is directly fed into the teacher model, which then transfers the knowledge to a much smaller model with knowledge distillation. The smaller model, also called the student model, is given in Fig. 3(b), with the reduced number of convolutional feature maps but with a same depth as the teacher model. The teacher model and the student model have 214,277 and 4,756 parameters, respectively. In both KDL and PR-KDL learning approaches, the output of the teacher model will be used as another source of ground truth to enhance the training process of the student model. We have designed and trained the models, without using existing ones or pre-trained already.

C. KNOWLEDGE DISTILLATION LEARNING (KDL)

KDL is known as a model compression method in which a large teacher model transfers knowledge to a light-weight model, known as the student model [47]. The process is implemented by training the student model to learn from the teacher model without a significant loss of performance validity.

The class probabilities generated by the teacher model, or called ‘soft target distribution’ has been used to distillate the knowledge:

$$\theta_S = \underset{\theta_S}{\operatorname{argmin}} \mathbb{E}_{(X,y) \sim \mathcal{D}} \left[\alpha \tau^2 \mathcal{L}_{S \rightarrow T}(\Phi_S(X, \theta_S, \tau), \Phi_T(X, \theta_T, \tau)) + (1 - \alpha) \mathcal{L}_{S \rightarrow Y}(\Phi_S(X, \theta_S), \mathbb{Y}) \right] \quad (1)$$

In the enhanced loss function as (1), θ_S is the parameter set for the student model Φ_S , and θ_T is the parameter set for the teacher model Φ_T . The KDL learning process is searching θ_S by minimizing the loss across the dataset $(X, \mathbb{Y}) \sim \mathcal{D}$, where X and y are input and ground truth, respectively. $\mathcal{L}_{S \rightarrow T}$ is the loss for the student to mimic the teacher, and $\mathcal{L}_{S \rightarrow Y}$ is the loss for the student to generate ground truth-referred output, τ is the temperature to control the softness of the generated target distribution, and α is the weighting factor to combine two loss criteria. By learning from the teacher, the student model tends to capture the ‘‘dark knowledge’’ which is information hidden in the tail end of the probability distribution of the teacher model.

More specifically, given an ECG segment to the teacher model, the teacher model will generate a 5-dimension vector, corresponding to the five output neurons in the last layer. This vector gives the probability of being each of the five categories. In a similar way, the student model also outputs a 5-dimension vector. Then an extra item $\mathcal{L}_{S \rightarrow T}$, i.e., the loss for the student model to mimic to teacher model, has been

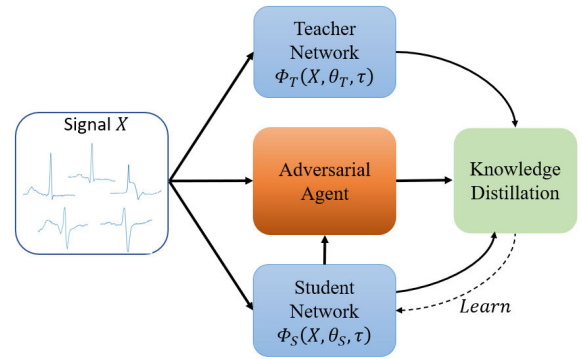


FIGURE 4. Preserving-robustness KDL to enhance the robustness of the student model.

introduced to (1) to enhance the original loss $\mathcal{L}_{S \rightarrow y}$, i.e., the loss between the student output and the ground truth.

By leveraging KDL, the student model is expected to yield comparable performance like the teacher model, with a significantly less parameters. The student model is shrunk from the teacher model. In the convolutional part, the number of feature maps has been halved. In the fully-connected layers, the number of neurons is reduced to a quarter, as Fig. 3.

D. PRESERVING-ROBUSTNESS KDL (PR-KDL)

To defend against adversarial attacks, one of the main approaches to do so is adversarial learning. It is a rigorous process in which the neural models are retrained on adversarial samples to enhance its robustness against perturbations in the data.

By introducing adversarial learning [48] to KD, we here achieve PR-KDL for the ECG-based heart disease detection application, as shown in Fig. 4. It is expected to enhance the robustness of the KDL model by learning how to identify inputs with perturbations.

The robustness under adversarial attacks can be captured in a natural saddle point (min-max) formulation. The min-max optimization problem can be formulated as (2)-(3):

$$\theta_S = \underset{\theta_S}{\operatorname{argmin}} \mathbb{E}_{(X,y) \sim \mathcal{D}} [\alpha \tau^2 \mathcal{L}_{S \rightarrow T}(\Phi_S(X', \theta_S, \tau) |_{X'=X+\delta_{X,\theta}}, \Phi_T(X, \theta_T, \tau)) + (1 - \alpha) \mathcal{L}_{S \rightarrow Y}(\Phi_S(X, \theta_S), \mathbb{Y})] \quad (2)$$

$$\delta_{X,\theta_S} = \underset{\|\delta\|_p < \epsilon}{\operatorname{argmin}} \mathcal{L}_{S \rightarrow Y}(\Phi_S(X', \theta_S), \mathbb{Y}) |_{X'=X+\delta} \quad (3)$$

The distillation loss $\mathcal{L}_{S \rightarrow T}$ in (2) is now calculated between the student model with the adversarial response X' and the teacher model with the original input X . The perturbed input $X' = X + \delta_{X,\theta}$ is determined by adding $\delta_{X,\theta}$ to the original input X . θ_S and θ_T are corresponding to the parameter set for the student model Φ_S and the teacher model Φ_T , respectively. The perturbed input X' now includes the information from the adversarial attack defined by δ_{X,θ_S} , as (3). The adversarial

agent as shown in Fig. 4, searches the input space to determine this perturbation δ_{X, θ_S} by maximizing the standard student model loss $\mathcal{L}_{S \rightarrow y}$. The allowed input space under perturbation is defined by $\|\delta\|_p < \varepsilon$, meaning that the perturbation is within the ε ball and the ε is set as 0.03 so that the perturbation is difficult to be visually observable. The adversarial attack in (3) is now introduced to the student model, and the enhanced loss function in (2) can capture the induced performance drop so that the student model can learn to evolve and minimize the loss in the training process.

The adversarial agent does increase the training cost, due to the introduced data perturbation process and the enhanced learning loss function. The good thing is that, when using the trained model in the testing phase, there is no additional cost because of the adversarial agent because only the trained student model itself is working.

E. EVALUATION METHOD

We have applied multiple criteria to evaluate the performance of the deep learning models and learning approaches, including accuracy, precision, recall and F1 score. The dataset is split into 3 different sets: 80%, 10%, and 10% for training, validation and testing, respectively.

We have also given the confusion matrix to detail the classification performance under different experimental settings. Further, we have compared both the number of parameters and detection accuracy with visualization, to illustrate different models and different ECG inputs impact the model size the performance.

III. RESULTS

A. EXPERIMENTAL SETUP

To demonstrate the effectiveness of the proposed approaches, we have used the widely-applied PhysioNet MIT-BIH Arrhythmia Database in our experiments [49]. The dataset has ECG recordings from 47 subjects, with each channel lasting a half hour. 23 out of the 48 recordings are selected after checking the signal fidelity, considering some recordings are with noise or artifacts. The remaining recordings consist of recordings that are comprised of clinically significant but rare arrhythmias from the same set. The noisy conditions in the databased will further facilitate the algorithm evaluation and enhancement. In the current study, we make efforts to build the framework for efficient edge inference of heart diseases from relatively clean ECG signals, to demonstrate the possibility, considering the ECG data is scarce and no previous research has made this effort. In future, it will be promising and interesting to advance this research.

The ECG recordings have a sampling frequency of 360 Hz and a resolution of 11 bits. No signal filtering is applied. The ECG samples are segmented into individual beats of a fixed length based on the annotation files. A window size of 160 has been applied both before and after the ECG heartbeat location, corresponding to the indices from X-159 to X+159, where X is the index of the current heartbeat location.

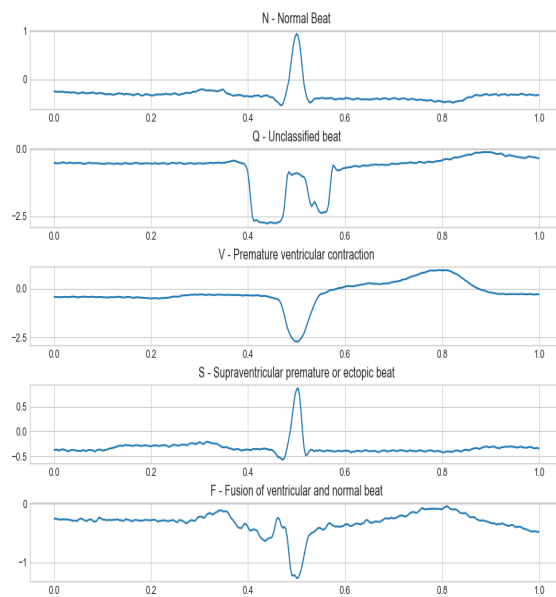


FIGURE 5. Visualization of ECG samples from each class.

Examples of the ECG samples of the five classes are visualized in Fig. 5. The diverse morphologies indicate the complexity of the heart disease detection problem.

In this study, the ECG data are classified into five distinctive classes based on the American Association for the Advancement of Medical Instrumentation (AAMI) standard [50]. The 5 classes that the data samples can be divided into are: normal (N) beats, unknown (Q) beats, ventricular ectopic (V) beats, supra ventricular ectopic beats (S) and fusion (F) beats [51]. The training, validation, and testing percentages are 80%, 10%, and 10%, respectively. The batch size, and number of epochs for the deep learning process, are selected as 32, and 50, respectively. Other hyperparameters like the learning rate are the default set by the PyTorch framework. Besides, the selected ECG segment can effectively cover the representative portion of a heartbeat, as shown in Fig. 5.

B. DEEP RESIDUAL CNN LEARNING

Fig. 6 and 7 show the performance of the teacher model. In the former one, the learning process has been given, indicating the convergence of both training and validation processes.

Fig. 7 further demonstrates the effective heart disease classification model using a confusion matrix. It is shown that most of heartbeats are distributed on the diagonal of the matrix, meaning that they are correctly classified.

The deep learning model, as shown in Fig. 6 and 7, has effectively captured the features in the ECG segments, through the residual convolutional blocks as given in Fig. 3. This automatically raw data-based feature learning process is highly effectively, without manual feature engineering. The learning curves in Fig. 6 demonstrates that, when the learning

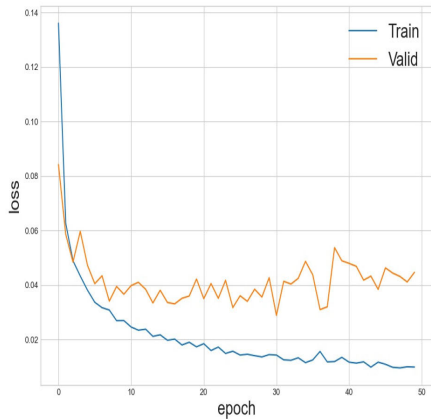


FIGURE 6. Learning curves of the teacher ResNet model.

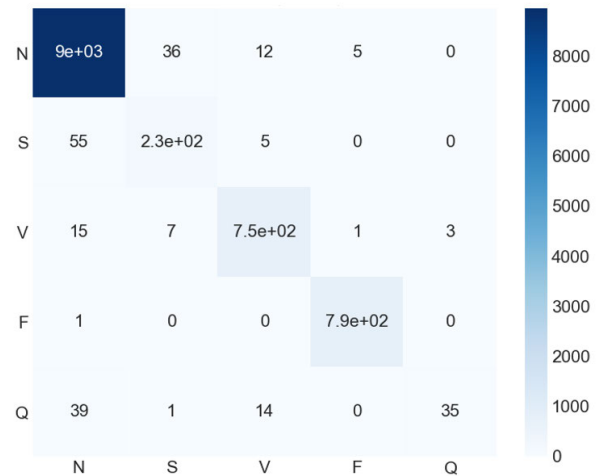


FIGURE 8. Confusion matrix for the testing data @ the student deep model learned with KDL. Note. Standard signal for both training and testing data.

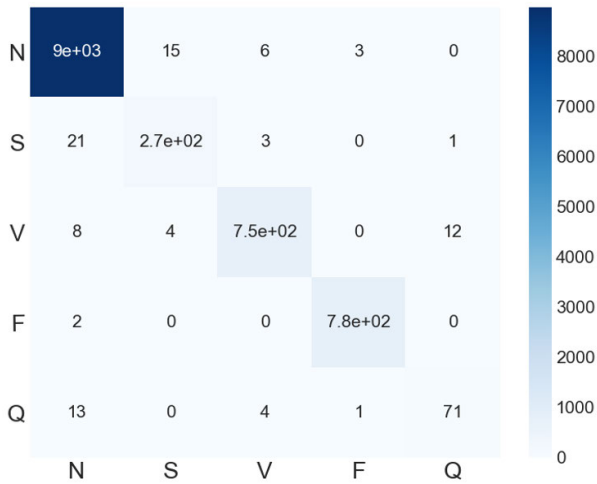


FIGURE 7. Confusion matrix for the testing data @ the teacher deep model. Notes. Standard signal for both training and testing data. The vertical axis is the ground truth.

loss decreases, the deep learning model has well-adjusted its neural parameters for the feature learning abstraction.

C. KNOWLEDGE DISTILLATION LEARNING

Fig. 8 gives the confusion matrix of the student model learned by KDL and evaluated on the standard ECG signal without adversarial perturbations. As observed, the student model is able to classify the majority of the ECG signals to their corresponding classes. The classification accuracy remains substantially high despite having significantly lower number of parameters as compared to the teacher model. One thing worth noting is that, for the class Q, the misclassification rate is a bit high, which is because of the minority class and imbalanced dataset. We have further used precision, recall, and F1 score later in the performance summary, to comprehensively report the experimental results, which still show attractive detection. In future, it will be interesting to further improve the performance for the minority class.

D. ADVERSARIAL AGENT

Adversarial attacks are used to generate the perturbed samples, as shown in Fig. 9. It is a white-box attack and thus, it is assumed that the adversaries have all information of the model. It is considered more malicious than black-box attacks which have no information about the model, as the adversaries could design their attack to “fool” the classifier model by perturbing the ECG signals. The adversarial samples that are slightly altered by the perturbations should be considerably similar to the actual signal, such that the difference is almost imperceptible to humans’ naked eyes but subtle enough to lead to misclassification by the classifier model. The visualization clearly shows that the adversarial sample s are very similar to the standard ECG signal, indicating the changes posed to the heart disease detection task.

E. KDL UNDER ADVERSARIAL ATTACKS

With adversarial attacks, KDL misclassifies a significant number of heartbeats, as shown in Fig. 10. The standard signal has been used as the training data, and the adversarial signal has been used as the testing data. A huge difference can be noticed when the model is facing adversarial attacks as compared to facing clean data in Fig. 8. The classification performance suffered a dip in accuracy as the adversarial samples can ‘trick’ the classifier model to make the wrong predictions. This indicates the necessities to enhance the robustness of the student model learned by KDL and proving that the traditional CNN classifier models have a severe lack of defense against these attacks.

F. PR-KDL UNDER ADVERSARIAL ATTACKS

The confusion matrix of the PR-KDL student model with adversarial samples is given in Fig. 11, which clearly demonstrate the enhanced robustness of the student model. Both training and testing are based on the adversarial signal.

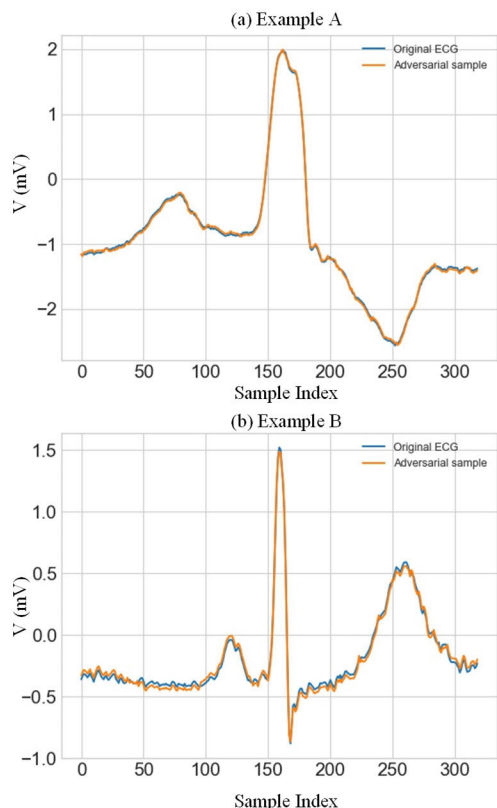


FIGURE 9. Two examples which include the perturbations generated by the adversarial agent.

For N/S/V/F classes, the performance has been all greatly improved, except for the Q class that has minimum number of heartbeats. The Q class, as a minority class, will be further studied in future. But overall, our PR-KDL framework has greatly boosted the performance of the KDL framework, by introducing substantial robustness during the learning process.

For comparison purpose, the PR-KDL student model with standard samples is also given in Fig. 12. The adversarial signal has been used as the training data, and the standard signal has been used as the testing data. The performance is bit better, which is reasonable since the input is the standard ECG signal.

One thing worth noting is that, the adversarial attack-based learning enhancement does not introduce additional computing overhead. Because it is achieved through enhancing the loss function, so only the training phase has more computing resources requirements. Once trained, the learned model has the same size as the efficient student model that has been used in the knowledge distillation framework. Therefore, the final student model has both high efficiency and high security, without needing additional computing resources.

G. COMPARISON BETWEEN KDL AND PR-KDL

To further compare KDL and PR-KDL, both natural accuracy and robust accuracy graphs are given in Fig. 13.

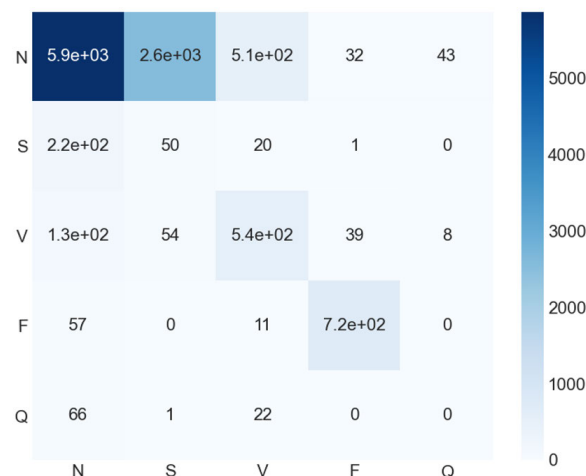


FIGURE 10. Confusion matrix for the testing data @ the KDL-based student model under attacks. Note. Standard signal on training data, and adversarial signal on testing data.

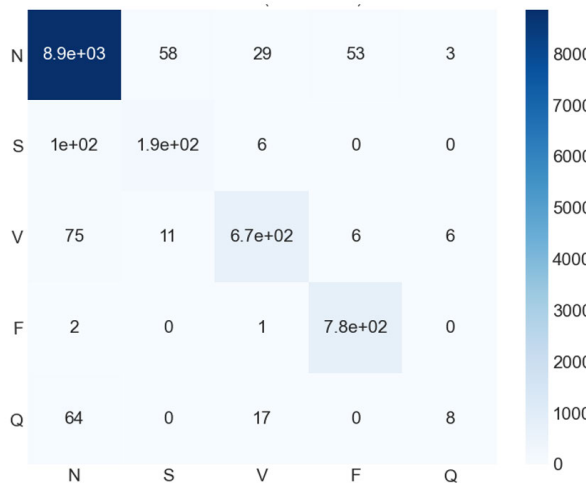


FIGURE 11. Confusion matrix for the testing data @ the RP-KDL-based student model under attacks. Note. Adversarial on both training and testing data.

First, the natural accuracy evaluated on the ECG signal is given for three models: teacher (natural), KDL student model(natural), PR-KDL student model (natural). Based on the visualization, it is observed that the classification accuracy is very similar for all the models. The attractive performance of the student model learned by KDL indicates the effectiveness of the KDL on learning the knowledge from the teacher model. Since there are no adversarial samples for these three models, more analysis under attacks is further given below.

The robust accuracy of the models is as shown in Fig. 13, including KDL (robust), and PR-KDL (robust). As observed from the visualization, a huge percentage of adversarial ECG signals are able to fool the KDL student model to generate incorrect detection results. However, the PR-KDL student

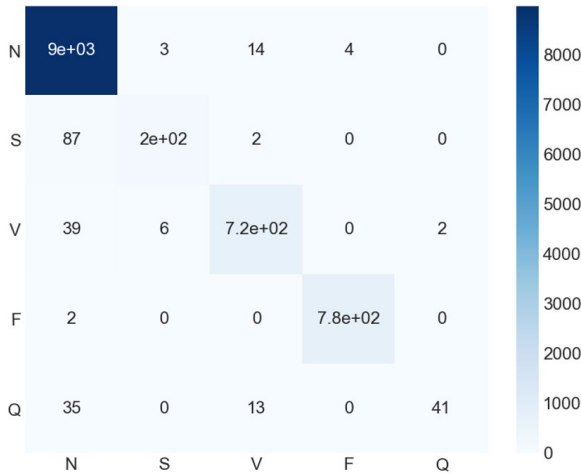


FIGURE 12. Confusion matrix for the testing data @ the RP-KDL-based student model without attacks. Note. Adversarial on training data, and standard signal on testing data.

TABLE 1. Summary of natural performance evaluated on the standard inputs.

Methods	# Params	Accuracy	Precision	Recall	F1
Teacher	214277	0.991	0.950	0.935	0.991
KDL	4765	0.982	0.940	0.829	0.863
RP-KDL	4765	0.981	0.970	0.818	0.872

model is able to defend against adversarial attacks. While facing adversarial noises, the PR-KDL student model remains strongly resilient to perturbations and the performance is still very good. Notice that the robust accuracy of the PR-KDL student model was slightly lower than the natural accuracy, which is an expected outcome as the purpose of adversarial attacks is to generate perturbations to the standard data to deceive the classifier.

A co-visualization of natural and robust accuracy in Fig. 13, can effectively demonstrate the performance drop of the KDL student model under attacks and the robustness of the PR-KDL model under attacks.

H. PERFORMANCE SUMMARY OF KDL AND PR-KDL

The performance of different models is summarized in Tables 1 and 2, corresponding to natural and robust performance, respectively. This summary further demonstrates: (1) the KDL student model can effectively learn the knowledge from the teacher model for standard ECG data, but with a much smaller model size – only 4765 parameters compared with the 214,277 parameters of the teacher model (50x reduction); (2) the PR-KDL model further enhanced the robustness of the student model and achieves very promising performance under adversarial signals – the accuracy, average precision, average recall and average F1 score are 0.960, 0.806, 0.716, and 0.734, respectively.

TABLE 2. Summary of robust performance evaluated on the adversarial inputs.

Methods	# Params	Accuracy	Precision	Recall	F1
KDL	4765	0.655	0.469	0.487	0.457
RP-KDL	4765	0.960	0.806	0.716	0.734

I. FUTURE STUDIES

In future, we will further investigate KDL for smaller model learning. Besides, the effect of different adversarial attacks on ECG Arrhythmia Classification can be further explored. Also, we are interested in further enhancing the performance for the class with limited number of samples. Besides, it will be interesting and proving more information if further experiments on the hardware implementation are added. We in this study, have mainly been targeting reducing the model size, i.e., the number of parameters. It is directly related to the computing resource needed, considering the reduced model size corresponds to less computations and less storage of parameters and intermediate data. The resulted student model has only 4,765 parameters, and can thus be easily stored in the arm context M4 processor chip. Also, the calculation is based on every one or two seconds, considering the low sampling rate of the ECG signal. So, the achieved model size is very promising for the edge deployment. We will in future further evaluate the model deployment on the hardware.

Further, edge computing has been reported in some interesting studies, such as the computation offloading on the mobile edge [52], distributed edge computing [53], compressed model [54], and hardware optimization [55]. We will further study the edge computing for the cardiac monitoring applications, which is currently in an early stage of investigations. And deployment of the deep learning models further on the hardware will provide interesting findings in the future research.

Besides, we have performed the comparison among different models, such as with/without KDL, and with/without adversarial learning. The model without both KDL and adversarial learning provides a good demonstration of the original model performance. Further, it will be interesting to evaluate more models and architectures, as well as more applications and databases, for framework exploration.

In this study, the real-world ECG-based heart disease application has been leveraged to demonstrate the effectiveness. Next, it will be very interesting to apply it in real-time on patients, after further efforts with physicians for more evaluations and experiments. Besides, the implementation of the algorithm on the mobile devices is also possible with the edge implementation tools. This study indicates the feasibility of knowledge distillation on ECG data learning and leverages the ResNet-based neural network architectures. It will be interesting to study further the model architectures and advance the knowledge distillation with other efforts like

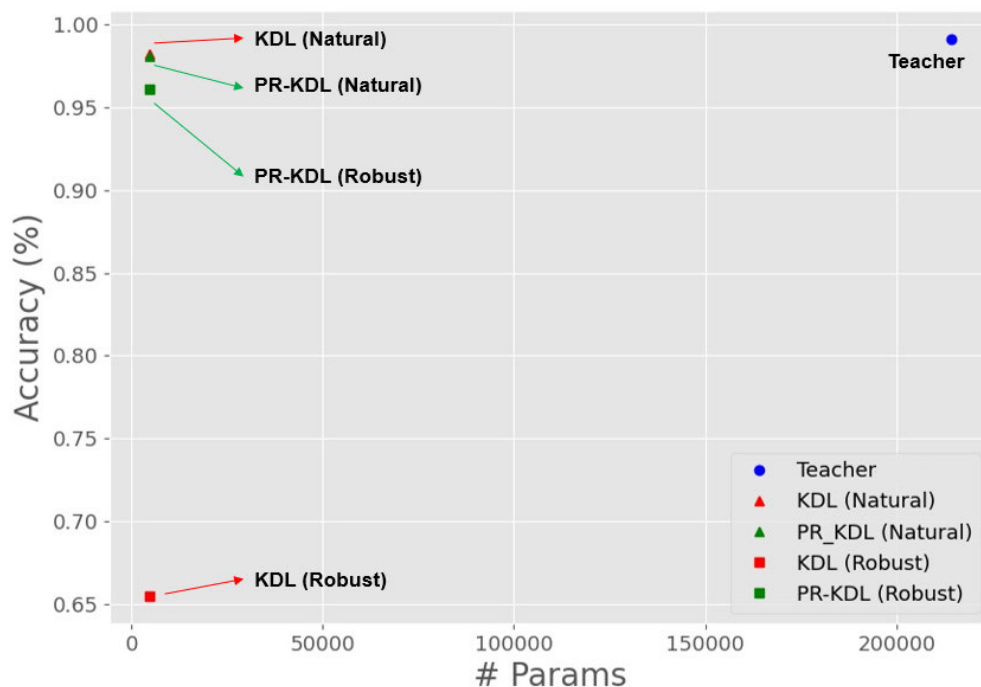


FIGURE 13. Co-visualization of natural and robust test of different models, indicating that KDL-based student model has comparable performance as the teacher model, and that PR-KDL model has boosted the performance when adversarial attacks exist. Notes. - Teacher: the teacher model trained and tested on natural standard ECG data; - KDL (Natural): the student model trained and tested on natural standard ECG data; - PR-KDL (Natural): the student model trained on adversarial data and tested on natural data; - KDL (Robust): the student model trained on natural data and tested on adversarial data; - PR-KDL (Robust): the student model trained on adversarial data and tested on adversarial data.

loss function enhancement. The adaptive student models can be selectively used for heterogeneous edge environments. The feasibility of the current algorithm framework has laid a good foundation for the potential future studies. The perturbation levels are also important factors for evaluating the algorithm framework. With different levels, it is interesting to see the ability of the algorithm and get the understanding of when it may work robustly and when it may not meet the requirements.

IV. CONCLUSION

In this study, targeting the challenges faced by the edge biomedical inference, we have investigated both KDL and PR-KDL frameworks and demonstrated their promising advancements. The KDL framework can significantly reduce the deep learning model size on the edge and still achieve comparable performance, through supervising the student model to learn the knowledge of the heavy teacher model. Further, under adversarial perturbations, PR-KDL has been proposed to enhance the robustness of the student model, thereby greatly boosting its performance facing adversarial inputs on the edge. With through experiments, we have shown that the proposed framework can enable light-weight and secure deep learning of cardiac abnormalities, on edge cardiac detection applications. This study is expected to greatly advance the real-time, long-term, and secure biomedical inference on edge.

ACKNOWLEDGMENT

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

- [1] B. Lobo, L. Farhy, M. Shafiei, and B. Kovatchev, "A data-driven approach to classifying daily continuous glucose monitoring (CGM) time series," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 654–665, Feb. 2022.
- [2] C. Qiu, F. Wu, W. Han, and M. R. Yuce, "A wearable bioimpedance chest patch for real-time ambulatory respiratory monitoring," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 9, pp. 2970–2981, Sep. 2022.
- [3] M. Sevil, M. Rashid, I. Hajizadeh, M. Park, L. Quinn, and A. Cinar, "Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 7, pp. 2251–2260, Jul. 2021.
- [4] J. Zou and Q. Zhang, "EyeSay: Eye electrooculography decoding with deep learning," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2021, pp. 1–3.
- [5] G. Guidoboni, L. Sala, M. Enayati, R. Sacco, M. Szopos, J. M. Keller, M. Popescu, L. Despina, V. H. Huxley, and M. Skubic, "Cardiovascular function and ballistocardiogram: A relationship interpreted via mathematical modeling," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2906–2917, Oct. 2019.
- [6] Q. Zhang and K. Frick, "All-ECG: A least-number of leads ECG monitor for standard 12-lead ECG tracking during motion," in *Proc. IEEE Healthcare Innov. Point Care Technol., (HI-POCT)*, DC, DC, USA, Nov. 2019, pp. 103–106.
- [7] A. Shahshahani, Z. Zilic, and S. Bhadra, "An ultrasound-based biomedical system for continuous cardiopulmonary monitoring: A single sensor for multiple information," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 268–276, Jan. 2020.

- [8] Q. Zhang, D. Zhou, and X. Zeng, "A novel framework for motion-tolerant instantaneous heart rate estimation by phase-domain multiview dynamic time warping," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2562–2574, Nov. 2017.
- [9] G. Qu, W. Hu, L. Xiao, J. Wang, Y. Bai, B. Patel, K. Zhang, and Y.-P. Wang, "Brain functional connectivity analysis via graphical deep learning," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 5, pp. 1696–1706, May 2022.
- [10] X. Zhan, Y. Liu, S. J. Raymond, H. V. Alizadeh, A. G. Domel, O. Gevaert, M. M. Zeineh, G. A. Grant, and D. B. Camarillo, "Rapid estimation of entire brain strain using deep learning models," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 11, pp. 3424–3434, Nov. 2021.
- [11] C. D. Vente, P. Vos, M. Hosseinzadeh, J. Plum, and M. Veta, "Deep learning regression for prostate cancer detection and grading in bi-parametric MRI," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 374–383, Feb. 2021.
- [12] Q. Zhang, "Deep learning of electrocardiography dynamics for biometric human identification in era of IoT," in *Proc. 9th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, New York, NY, USA, Nov. 2018, pp. 885–888.
- [13] K. Song, K.-Y. Chung, and J.-H. Chang, "Cuffless deep learning-based blood pressure estimation for smart wristwatches," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4292–4302, Jul. 2020.
- [14] K. Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.
- [15] C. Nuzzi, S. Pasinetti, M. Lancini, F. Docchio, and G. Sansoni, "Deep learning-based hand gesture recognition for collaborative robots," *IEEE Instrum. Meas. Mag.*, vol. 22, no. 2, pp. 44–51, Apr. 2019.
- [16] Y. Kutlu and D. Kuntalp, "A multi-stage automatic arrhythmia recognition and classification system," *Comput. Biol. Med.*, vol. 41, no. 1, pp. 37–45, Jan. 2011.
- [17] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Gener. Comput. Syst.*, vol. 86, pp. 446–455, Sep. 2018.
- [18] W. Li and J. Li, "Local deep field for electrocardiogram beat classification," *IEEE Sensors J.*, vol. 18, no. 4, pp. 1656–1664, Feb. 2018.
- [19] B. Murugesan, V. Ravichandran, K. Ram, P. S. P., J. Joseph, S. M. Shankaranarayana, and M. Sivaprakasam, "ECGNet: Deep network for arrhythmia classification," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2018, pp. 1–6.
- [20] J. S. Healey and J. Wong, "Wearable and implantable diagnostic monitors in early assessment of atrial tachyarrhythmia burden," *EP Europace*, vol. 21, no. 3, pp. 377–382, Mar. 2019.
- [21] M. Dziubiński, "PocketECG: A new continuous and real-time ambulatory arrhythmia diagnostic method," *Cardiol. J.*, vol. 18, no. 4, pp. 454–460, 2011.
- [22] E. V. Platia and P. R. Reid, "Comparison of programmed electrical stimulation and ambulatory electrocardiographic (Holter) monitoring in the management of ventricular tachycardia and ventricular fibrillation," *J. Amer. College Cardiol.*, vol. 4, no. 3, pp. 493–500, Sep. 1984.
- [23] V. Shusterman, B. Aysin, G. B. Ermentrout, B. London, and D. Schwartzman, "Detecting instabilities of cardiac rhythm," *J. Electrocardiol.*, vol. 36, pp. 219–226, Dec. 2003.
- [24] C. Dussault, H. Toeg, M. Nathan, Z. J. Wang, J.-F. Roux, and E. Secemsky, "Electrocardiographic monitoring for detecting atrial fibrillation after ischemic stroke or transient ischemic attack: Systematic review and meta-analysis," *Circulat., Arrhythmia Electrophysiol.*, vol. 8, no. 2, pp. 263–269, Apr. 2015.
- [25] F. Enseleit and F. Duru, "Long-term continuous external electrocardiographic recording: A review," *EP Europace*, vol. 8, no. 4, pp. 255–266, Apr. 2006.
- [26] B. Taji, A. D. C. Chan, and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1124–1131, May 2018.
- [27] M. Hammad, A. M. Iliyasa, A. Subasi, E. S. L. Ho, and A. A. El-Latif, "A multitier deep learning model for arrhythmia detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
- [28] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.
- [29] A. Mostayed, J. Luo, X. Shu, and W. Wee, "Classification of 12-lead ECG signals with bi-directional LSTM network," 2018, *arXiv:1811.02090*.
- [30] B. A. Teplitzky, M. McRoberts, and H. Ghanbari, "Deep learning for comprehensive ECG annotation," *Heart Rhythm*, vol. 17, no. 5, pp. 881–888, May 2020.
- [31] Y.-C. Yeh, W.-J. Wang, and C. W. Chiou, "A novel fuzzy c-means method for classifying heartbeat cases from ECG signals," *Measurement*, vol. 43, no. 10, pp. 1542–1555, Dec. 2010.
- [32] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, p. 285, Aug. 2016.
- [33] M. Sharma, R. S. Tan, and U. R. Acharya, "A novel automated diagnostic system for classification of myocardial infarction ECG signals using an optimal biorthogonal filter bank," *Comput. Biol. Med.*, vol. 102, pp. 341–356, Nov. 2018.
- [34] R. Varatharajan, G. Manogaran, and M. K. Priyan, "A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing," *Multimedia Tools Appl.*, vol. 77, no. 8, pp. 10195–10215, Apr. 2018.
- [35] M. B. Abubaker and B. Babayigit, "Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods," *IEEE Trans. Artif. Intell.*, vol. 4, no. 2, pp. 373–382, Apr. 2023.
- [36] N. I. Hasan and A. Bhattacharjee, "Deep learning approach to cardiovascular disease classification employing modified ECG signal from empirical mode decomposition," *Biomed. Signal Process. Control*, vol. 52, pp. 128–140, Jul. 2019.
- [37] M. A. Abdou, "Literature review: Efficient deep neural networks techniques for medical image analysis," *Neural Comput. Appl.*, vol. 34, no. 8, pp. 5791–5812, Apr. 2022.
- [38] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.
- [39] T. Mahmud, S. A. Fattah, and M. Saquib, "DeepArrNet: An efficient deep CNN architecture for automatic arrhythmia detection and classification from denoised ECG beats," *IEEE Access*, vol. 8, pp. 104788–104800, 2020.
- [40] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, *arXiv:1412.6572*.
- [41] K. Xu, H. Chen, S. Liu, P.-Y. Chen, T.-W. Weng, M. Hong, and X. Lin, "Topology attack and defense for graph neural networks: An optimization perspective," 2019, *arXiv:1906.04214*.
- [42] S. Ye, K. Xu, S. Liu, H. Cheng, J.-H. Lambrechts, H. Zhang, A. Zhou, K. Ma, Y. Wang, and X. Lin, "Adversarial robustness vs. Model compression, or both?" in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 111–120.
- [43] Z. Bao, Y. Lin, S. Zhang, Z. Li, and S. Mao, "Threat of adversarial attacks on DL-based IoT device identification," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 9012–9024, Jun. 2022.
- [44] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," 2016, *arXiv:1603.08029*.
- [45] Q. Zhang and S. Zhu, "Real-time activity and fall risk detection for aging population using deep learning," in *Proc. 9th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, New York, NY, USA, Nov. 2018, pp. 1055–1059.
- [46] K. Gangadharan and Q. Zhang, "Deep transferable intelligence for spatial variability characterization and data-efficient learning in biomechanical measurement," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
- [47] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.
- [48] M. Goldblum, L. Fowl, S. Feizi, and T. Goldstein, "Adversarially robust distillation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 2020, vol. 34, no. 4, pp. 3996–4003.
- [49] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
- [50] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, ANSI/AAMI Standard EC 38, American National Standards Institute, Washington, DC, USA, 1998.
- [51] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.
- [52] A. Shakarami, A. Shahidinejad, and M. Ghobaei-Arani, "An autonomous computation offloading strategy in mobile edge computing: A deep learning-based hybrid approach," *J. Neww. Comput. Appl.*, vol. 178, Mar. 2021, Art. no. 102974.

[53] J. Chen, K. Li, Q. Deng, K. Li, and P. S. Yu, "Distributed deep learning model for intelligent video surveillance systems with edge computing," *IEEE Trans. Ind. Informat.*, pp. 1–9, Apr. 2019, doi: 10.1109/TII.2019.2909473.

[54] F. Wang, M. Zhang, X. Wang, X. Ma, and J. Liu, "Deep learning for edge computing applications: A state-of-the-art survey," *IEEE Access*, vol. 8, pp. 58322–58336, 2020.

[55] M. P. Véstias, R. P. Duarte, J. T. de Sousa, and H. C. Neto, "Moving deep learning to the edge," *Algorithms*, vol. 13, no. 5, p. 125, May 2020.



JEANNE NERBONNE received the Ph.D. degree from Georgetown University, Washington, DC, USA, in 1978.

She was a Researcher with the California Institute of Technology. She was a Faculty with the School of Medicine, Washington University in St. Louis, in 1985. She is currently an Alumni Endowed Professor and the Director of the Center for Cardiovascular Research. She is a Professor with the Medical School, Department of Developmental and Internal Medicine. Her publications have 51,656 citations and H-index is 78. Her lab has been continually funded by NIH&AHA. Her research interests include biochemical, electrophysiological, immunohistochemical and genetic approaches on the dynamic regulation, and dysregulation of neuronal and cardiac dynamics. She is an AHA Fellow. She served on numerous study sections as a member or the Chair for NIH&AHA for decades. She served on the Editorial Board for *Journal of Neuroscience* and *Journal of General Physiology*. She serves as an AE for *Scientific Reports* (Nature) and *Journal of General Physiology*.



QINGXUE ZHANG (Senior Member, IEEE) has over 15 years' of experience in both academia and industry, with his postdoctoral research with Harvard, products research and development in ICT, and Ph.D. research with the University of Texas at Dallas. He was a Faculty with the School of Engineering and Technology, Purdue University, Indianapolis, USA, in 2018. He is directing the Ubiquitous Intelligence Laboratory. His research interests include deep learning, biomedical instrumentation, edge computing, brain-inspired learning, targeting smart health, home, and world applications.

He was a recipient of the prestigious USA NSF CAREER Award. He received the Featured Journal Article Award in IEEE ACCESS, the Best Paper Award in UEMCON2017, the Early-Career Travel Award in AHA2020, the Favorite Professor Award in 2019, and the Google Cloud Award in 2019. He serves as a USA NSF/NIH/NIST Panelists and the IEEE CTSoc MDA Technical Committee. He serves as the AI Program Chair for the IEEE ICCE Conference, the Program Chair and the Workshop Chair for the IEEE IPCCC Conference, the Session Chair for the IEEE CCWC Conference, and the committee chair for multiple IEEE conferences. He serves as an AE for IEEE ACCESS and IEEE TRANSACTIONS ON CONSUMER ELECTRONICS.

...



JUNHUA WONG is currently pursuing the degree with the Department of Electrical and Computer Engineering, School of Engineering Technology, Purdue University, USA. His research interests include smart health, rehabilitation, deep learning, machine learning, big data, and efficient AI computing technologies.