# EFFICIENT EDGE INTELLIGENCE IN THE ERA OF BIG DATA

by

Jun Hua Wong

**A Thesis**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Master of Science**

Department of Electrical and Computer Engineering at IUPUI

Indianapolis, Indiana

August 2021

## THE PURDUE UNIVERSITY GRADUATE SCHOOL
## STATEMENT OF COMMITTEE APPROVAL

**Dr. Qingxue Zhang, Chair**
Department of Electrical and Computer Engineering

**Dr. Brian King**
Department of Electrical and Computer Engineering

**Dr. Peter Schubert**
Department of Electrical and Computer Engineering

**Approved by:**
Dr.  Brian King

*Dedicated to My Parents: Ding Hook and Fang Mei,*
*My Godparents: Hock Kee and Sut Leng*
*My family, friends, and all well-wishers.*

# ACKNOWLEDGMENTS

I would like to begin by expressing my gratitude to my parents for their unconditional support in pursuing my degree. I would also like to thank my thesis advisor, Dr Qingxue Zhang for his guidance and instructions throughout the research. Without his expertise and valuable insights, this research would not have been possible. I am also grateful to have Dr. Brian King and Dr. Peter Schubert, for being part of the thesis committee, and Sherrie Tucker, who has assisted me tirelessly throughout my studies at IUPUI.

I am grateful to my brother, Jun Tat, and my sister, Poh Yee who have been encouraging and motivating me since day one. I would also like to honor my grandparents who has taught me perseverance and patience. My foster parents in America, Bridget Fultz is always supportive and make me feel welcomed in the states. Lastly, I would like to extend my sincere thanks to my friends and family who have supported me throughout my research.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

Smart wearables, known as emerging paradigms for vital big data capturing, have been attracting intensive attentions. However, one crucial problem is their power-hungriness, i.e., the continuous data streaming consumes energy dramatically and requires devices to be frequently charged. Targeting this obstacle, we propose to investigate the biodynamic patterns in the data and design a data-driven approach for intelligent data compression. We leverage Deep Learning (DL), more specifically, Convolutional Autoencoder (CAE), to learn a sparse representation of the vital big data. The minimized energy need, even taking into consideration the CAE-induced overhead, is tremendously lower than the original energy need. Further, compared with state-of-the-art wavelet compression-based method, our method can compress the data with a dramatically lower error for a similar energy budget. Our experiments and the validated approach are expected to boost the energy efficiency of wearables, and thus greatly advance ubiquitous big data applications in era of smart health.

In recent years, there has also been a growing interest in edge intelligence for emerging instantaneous big data inference. However, the inference algorithms, especially deep learning, usually require heavy computation requirements, thereby greatly limiting their deployment on the edge. We take special interest in the smart health wearable big data mining and inference. Targeting the deep learning's high computational complexity and large memory and energy requirements, new approaches are urged to make the deep learning algorithms ultra-efficient for wearable big data analysis. We propose to leverage knowledge distillation to achieve an ultra-efficient edge-deployable deep learning model. More specifically, through transferring the knowledge from a teacher model to the on-edge student model, the soft target distribution of the teacher model can be effectively learned by the student model. Besides, we propose to further introduce adversarial robustness to the student model, by stimulating the student model to correctly identify inputs that have adversarial perturbation.  Experiments demonstrate that the knowledge distillation student model has comparable performance to the heavy teacher model but owns a substantially smaller model size. With adversarial learning, the student model has effectively preserved its robustness. In such a way, we have demonstrated the framework with knowledge distillation and adversarial learning can, not only advance ultra-efficient edge inference, but also preserve the robustness facing the perturbed input.

Our research findings are expected to greatly advance the real-time, long-term wearable big data and mining applications. The proposed deep learning-empowered method, CAE can effectively extract important features and encode data in latent space representation. As a result, the reduced dimension of data could lower the total energy consumption in a smart wearable device. On the other hand, the proposed model compression framework can produce classifier models that are not only lightweight, but also robust against adversarial attacks. The framework leverages knowledge distillation and adversarial learning to enhance the robustness of the classifier models to perform ECG arrhythmia classification.

# 1. INTRODUCTION

Edge intelligence is a computing paradigm that pushes data storage, data analysis and communication of data insights in near real-time. With the emergence of Artificial Intelligence (AI), there has been significant advancements in smart health edge computing. For instance, ECG can be used by the medical professionals and/or machine learning (ML) models to understand the conditions of the heart and detect abnormalities if there is any. The early detection of abnormal conditions could then trigger the health wearables to notify the medical staff via cloud technologies to allow the user to receive medical care before the condition worsens.

Smart wearables, known as emerging paradigms for vital big data capturing, have been attracting intensive attentions. However, one crucial problem is their power-hungriness, i.e., the continuous data streaming consumes energy dramatically and requires devices to be frequently charged. Targeting this obstacle, we propose to investigate the biodynamic patterns in the data and design a data-driven approach for intelligent data compression. In Chapter 2, we will detail how we leverage Deep Learning (DL), more specifically, Convolutional Autoencoder (CAE), to learn a sparse representation of the vital big data. The minimized energy need, even taking into consideration the CAE-induced overhead, is tremendously lower than the original energy need.

Another core issue is the DL model size for edge inference. DL models can be utilized to detect arrhythmias from the ECG data. Nevertheless, these models can be computationally intensive and exhaust a huge portion of memory, which could worsen the battery life. In addition, there has been a rising attention to adversarial attacks as the vulnerability of the typical DL models have been heavily studied in recent works. The attacks could compromise the security and robustness of the DL models, leading to misclassification and error in health diagnosis that can cause serious repercussions. Therefore, in Chapter 3, we will propose to leverage knowledge distillation to achieve an ultra-efficient edge-deployable deep learning model. More specifically, through transferring the knowledge from a teacher model to the on-edge student model, the soft target distribution of the teacher model can be effectively learned by the student model. Besides, we propose to further introduce adversarial robustness to the student model, by stimulating the student model to correctly identify inputs that have adversarial perturbation.

From Chapter 2 and 3, methodologies driven by deep learning principles are proposed to address the challenges in the smart health edge computing. The experimental results have demonstrated the underlying potential of the proposed methodologies to reduce the energy consumption of smart wearable devices and thereby, advance the ubiquitous deployment of the edge ecosystem. The proposed framework to compress deep learning models can also reduce the size of the model while strengthening its defense against adversarial ECG signals.

The thesis is organized as below:

- Chapter 2: deep convolutional autoencoder for energy-efficient smart health wearables
- Chapter 3: ultra-efficient edge intelligence towards real-time, long-term wearable big data mining and inference
- Chapter 4: conclusion
- Chapter 5: future studies
- Chapter 6: references

In Chapter 2 and 3, detailed introduction will be further given, and corresponding methods and results will also be detailed.

# 2. DEEP CONVOLUTIONAL AUTOENCODER FOR ENERGY-EFFICIENT SMART HEALTH WEARABLES

## 2.1 Introduction

Nowadays, Internet of Things (IoT) is rapidly becoming an emerging big data technology for the society. It allows different kinds of network-enabled sensors in our daily life to communicate and interact with the cloud [1, 2]. With regards to that, there are large amounts of promising applications of IoT in the mobile healthcare industry, where body sensor network (BSN) can be used to monitor vital signs of humans. This study focuses on human sensing through wireless-enabled wearable IoT devices, such as smart-watches, activity trackers, or smart wristbands, which can be used to facilitate the health management and quality of lives.

For instance, wearable IoT devices can be utilized to capture and measure crucial dynamics such as the heart rate and other individualized metrics that are relevant to physiological, biomechanical, and mental health. This allows the users to monitor their health conditions continuously and be aware of their health. With the access to these data granted to the healthcare professionals, the doctors would be able to record and analyze data more accurately to prescribe personalized medical care plans for their patients. This allows for a significant improvement in healthcare efficiency and decreases the healthcare cost.

In this study, we consider Electrocardiogram (ECG) signals as our signal of interest. ECG signal is commonly considered as one of the most important bio-signals as it provides valuable information about a person’s cardiac dynamics. This allows healthcare professionals to detect heart problems, ensuring that their patients’ cardiac problems are diagnosed with precision.

A flowchart of IoT in the healthcare ecosystem is demonstrated in Fig. 2.1. The sensors in the wearable device can first capture and record the physiological parameters of humans. The signals are then sent to a smartphone and the cloud infrastructures over wireless connections. One crucial challenge in this big data application is that the wearable monitor consumes significant amounts of energy in long-term usage [3]. Clearly, this obstacle needs to be addressed towards unconstrained long-term wearable big data streaming.

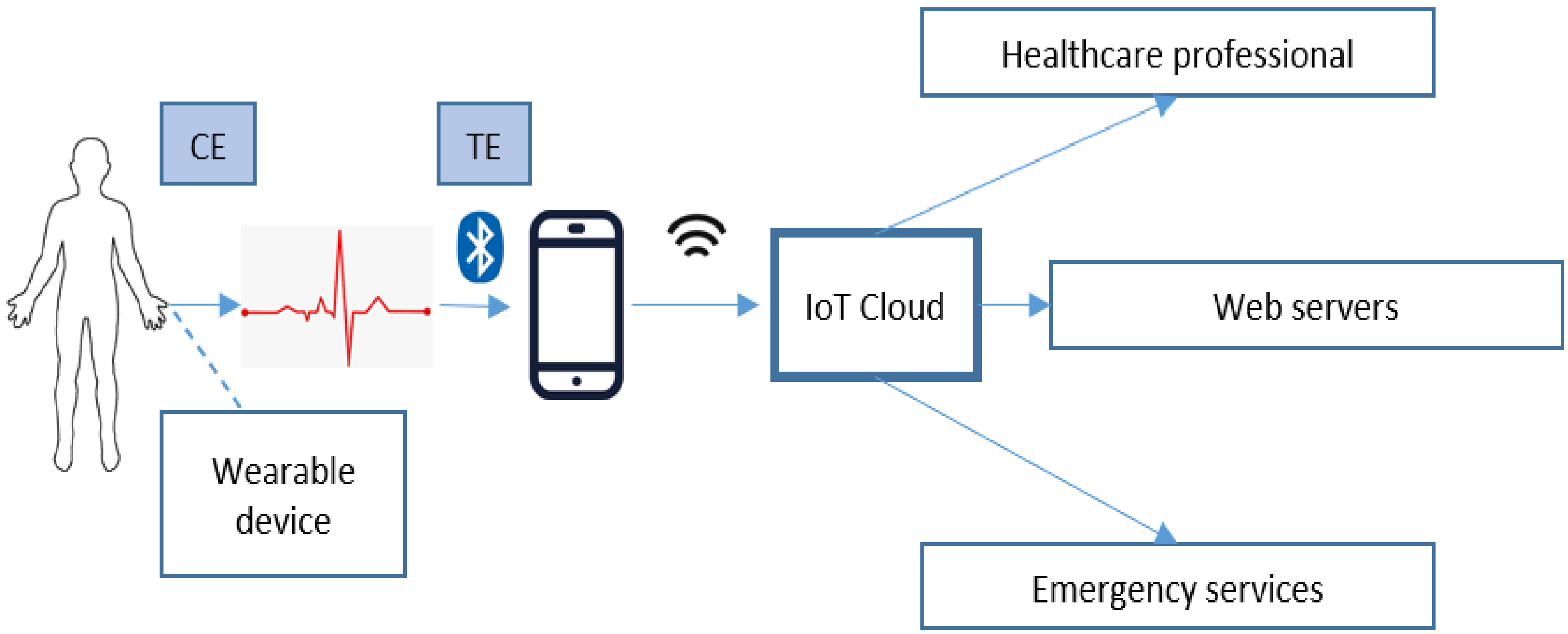


Figure 2.1 The demonstration of smart health wearables in era of big data. The wearable device continuously transmits Electrocardiogram (ECG) via Bluetooth connection to a smart phone, which then relays data to the cloud. The obstacle is how to enhance the energy efficiency of the wearable since streaming data continuously consumes significant amounts of energy and makes long-term usage very challenging. CE denotes compression energy and TE denotes transmission energy.

There are already some previous studies on lowering the energy consumption of wearables. For instance, Udaya et al. [4] proposed a Compressed Sensing (CS) method to compress the ECG signal. SC et al. [5] introduced a discrete wavelet transform (DWT) approach-based compression approach. Chompusri et al. [6] studied DWT with different types of mother wavelets. Bendifallah et al. proposed an ECG compression technique using Discrete Cosine Transform (DCT) [7]. Yildirim et al. [8] reported the autoencoder based method but did not study the energy numbers and overhead of the algorithm.

Overall, although these methods have obtained some extend of energy reduction, there are two crucial questions needing substantial further research efforts. They are: how to further leverage the complex dynamics in the data to maximize the compression ratio, and how to systematically evaluate the energy reduction and energy overhead of the compression algorithm. The former one is based on the observation that methods like DWT and DCT use predefined basis functions to analyze the sparsity of the data, which usually cannot extract very complex and nonlinear data dynamics. The latter one is because some studies only report data compression ratio, without further quantizing the algorithm overhead.

We emphasize the importance of the energy usage aspect in smart wearable devices because an energy-efficient compression algorithm would allow the users to use their device for an extensive period of time without recharging it between short spans of time. Most of these devices are operated with small batteries which have a limited capacity, and thus low energy usage should be given key consideration during the design process of a wearable device. More importantly, the quality of the ECG signals reconstruction should not be neglected.

In this study, we aim to address these gaps. We propose a deep learning-empowered energy-efficient wearable system to, not only learn how to compress the complex and nonlinear dynamics in the data via a deep neural network, but also take into account both energy reduction and algorithm's energy overhead. More specifically, first, we propose to leverage the convolutional autoencoder (CAE) – one of the main-stream deep learning algorithms [9], to analyze and compress the ECG signal. Compared with DWT, CAE is expected to be able to intelligently and effectively learn and extract the nonlinear dynamics in the data. Second, we propose to study how the topologies of CAE are contributing to factors like compression ratio, reconstruction error, and energy overhead, thereby enabling a systematic understanding of the total energy consumption and reconstruction error. This research will greatly advance smart health big data and/or IoT big data applications.

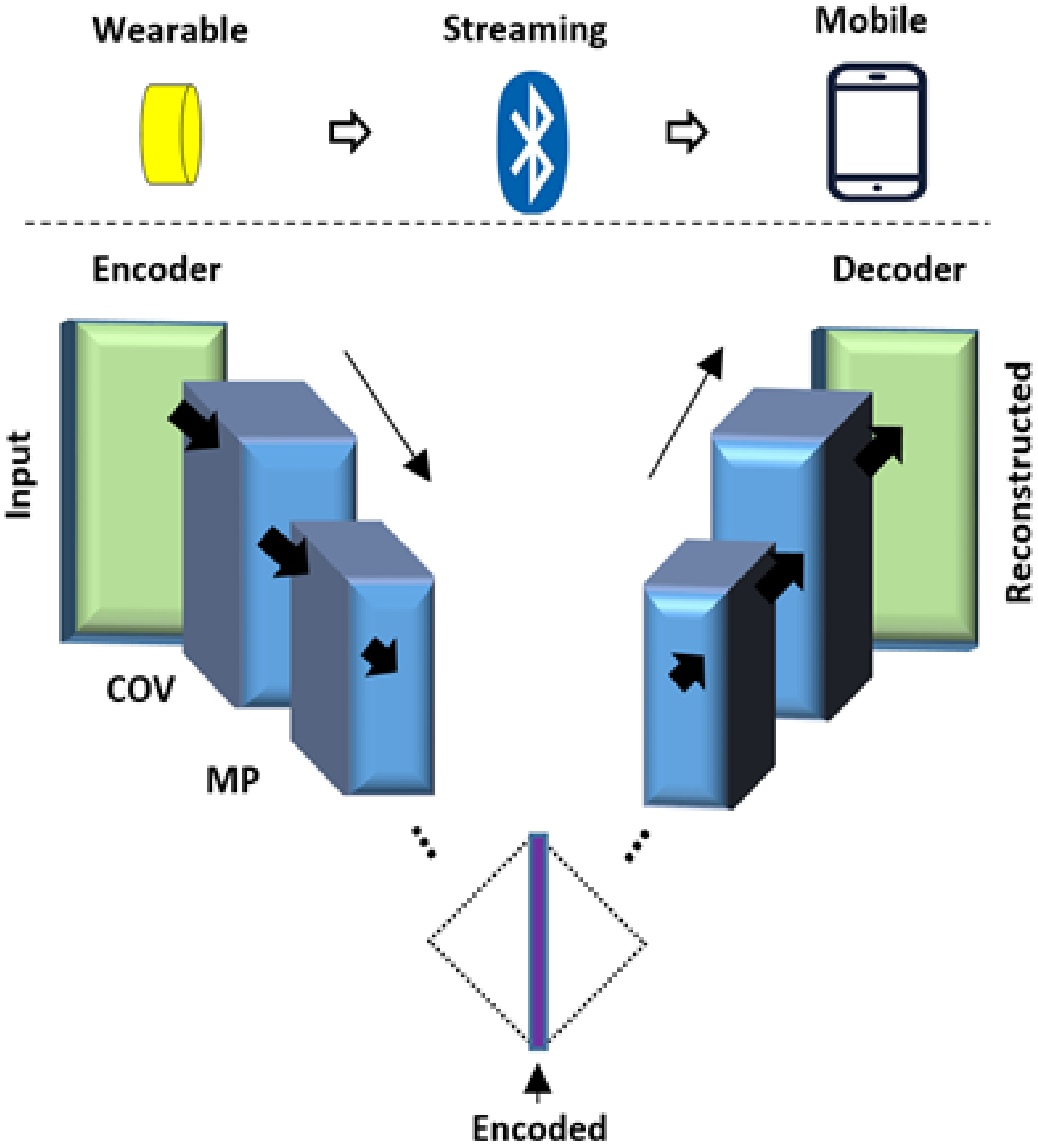


Figure 2.2 The system diagram of the wearable data compression and mobile data reconstruction, for minimizing the overall energy consumption of the wearable monitor. The encoded sparse representation of ECG is wirelessly streamed to the smart phone, which is then reconstructed on the phone or cloud. The encoder is based on convolutional autoencoder (CAE) and the decoder is inverse CAE. Notes. COV: convolutional filter; MP: max pooling

## 2.2 Approaches

### 2.2.1 System Diagram

The proposed deep learning framework is shown in Fig. 2.2, where the ECG signals are first compressed on the wearable device, and then transmitted to the smart phone or cloud for decompression. Below we will detail how to design the CAE for intelligent ECG compression and how to co-consider the total energy consumption and reconstruction error.

### 2.2.2 CAE Design

CAE, as shown in Fig. 2.2, is designed to firstly compress ECG on the wearable monitor and then decompress the encoded representation on the mobile phone or cloud depending on the computing resources and power budget of the whole big data system. In such a way, ECG can be converted to a sparse vector that encodes the major cardiac dynamics.

There are several key considerations on this deep learning-based data compression and decompression process. Firstly, CAE is different from the common Fully-connected Neural Networks (FNN) [9]. In CAE, the neurons in layer l are not fully connected to the subsequent layer, i.e. layer l + 1. Instead, the neurons in layer l are processed by convolutional filters to extract the spatial motifs. All neurons in the same feature map in layer l + 1 share the same neural weights. The computation complexity of CAE is greatly decreased in this way and is more suitable for wearable monitors, compared with FNN.

Secondly, the max pooling layers in the encoder (Fig. 2.2) gradually reduce the dimension of the input signal, thereby yielding the sparse representation. They also reduce the number of parameters and computation load of the network.

### 2.2.3 Topology Investigation

To design a CAE, there are a set of parameters, called hyperparameters, which include the number of layers L, input width W, number of feature map F, convolutional filter size C, and max pooling size P. The combinations of these parameters have a direct relationship with the compression ratio, and equally importantly, the energy overhead of the CAE algorithm.

To determine the optimal topologies for the CAE for minimization of both total energy and reconstruction error, we have extensively explored the above design parameters. Specifically, we have chosen the settings to be: $L \in [2, 4, 6], W \in [256, 512], F \in [2, 4, 6], C \in [2, 4, 8], P \in [2].$

By thoroughly evaluating different CAE topologies, we want to systematically understand how these design parameters impact the performance and energy. One thing to note is that the max pooling size is set to be 2, so that when adding a new max pooling layer, the compression ratio will be simply increased by 2 times.

### 2.2.4 Evaluation

To effectively evaluate the proposed methodology, we here introduce both evaluation criteria and comparison with state-of-the-art. Two evaluation criteria we choose include the reconstruction error (Root Mean Square Error: *RMSE*) and total energy consumption (*Etotal*). The former one is calculated as the error between the ground truth and the reconstructed signal. The latter one is the energy consumption of the signal transmission function and the signal compression algorithm. The signal transmission energy is calculated based on the number of bytes to be transferred per second [10]. The algorithm-consumed energy is calculated based on the number of operations [11]. As analyzed in the introduction section, many studies only report the compression ratio. This study, has already greatly advanced the study by the quantitative study of energy consumption.

Furthermore, DWT is also implemented for comparison purposes. The High Pass Filter (HPF) and Low Pass Filter (LPF) architecture have been used to determine the number of calculations required for different compression ratios. A detailed comparison is given in the next section.

## 2.3 Results

### 2.3.1 Experimental Setup

To effectively evaluate the proposed framework, we have chosen the well-known MIT-BIH Arrhythmia Database from Physionet. ECG signals with different abnormal morphologies and/or rhythms were collected from patients with heart diseases, which provide a good means to evaluate our system in the real-world case. Ten recordings without severe motion artifacts have been selected in this evaluation, and each recording is about 24-min long. The sampling rate is 360-Hz.

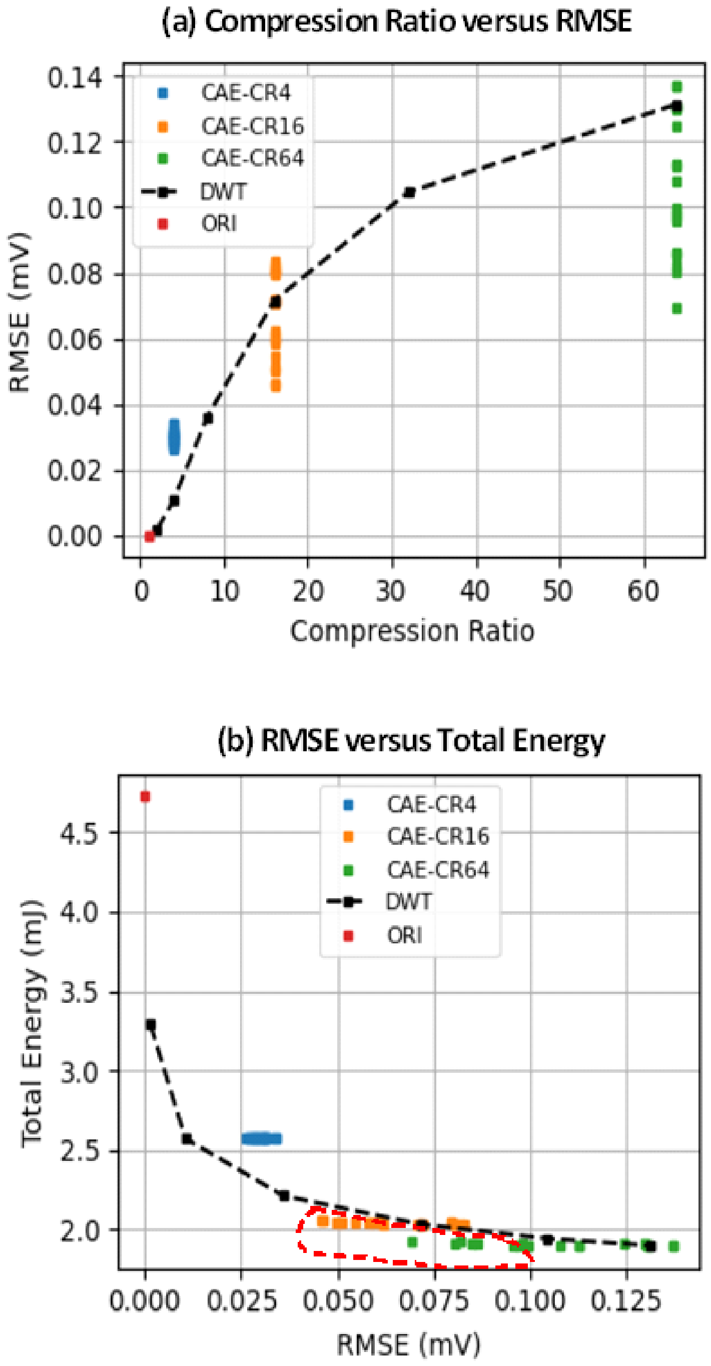


Figure 2.3. The comparison between proposed CAE and traditional DWT for energy-efficient wearable ECG monitoring. (a) shows that, when increase the compression ratio, CAE topologies generally provide a lower RMSE thanks to the nonlinear learning ability. (b) shows that, CAE topologies provide superior options (red cycle) compared with DWT, i.e., same energy but lower RMSE, or same RMSE but lower energy. Notes. CR: compression ratio; ORI: original, i.e., no compression.

### 2.3.2 CAE versus DWT

Fig. 2.3. gives the comparison between proposed CAE and traditional DWT for energy-efficient wearable ECG monitoring. Fig. 2.3(a) shows that, when increase the compression ratio (16 or 64), CAE topologies generally provide a lower RMSE. This is achieved by the powerful deep learning capability of CAE, which, compared with DWT, can more effectively learn the nonlinear ECG dynamics under high compression ratios. DWT, however, directly discards coefficients during data compression and thus impacts the performance dramatically when compression ratio is high. Fig. 2.3(b) shows that, CAE topologies provide superior options (within the red cycle) compared with DWT, i.e. same energy but lower RMSE, or same RMSE but lower energy.

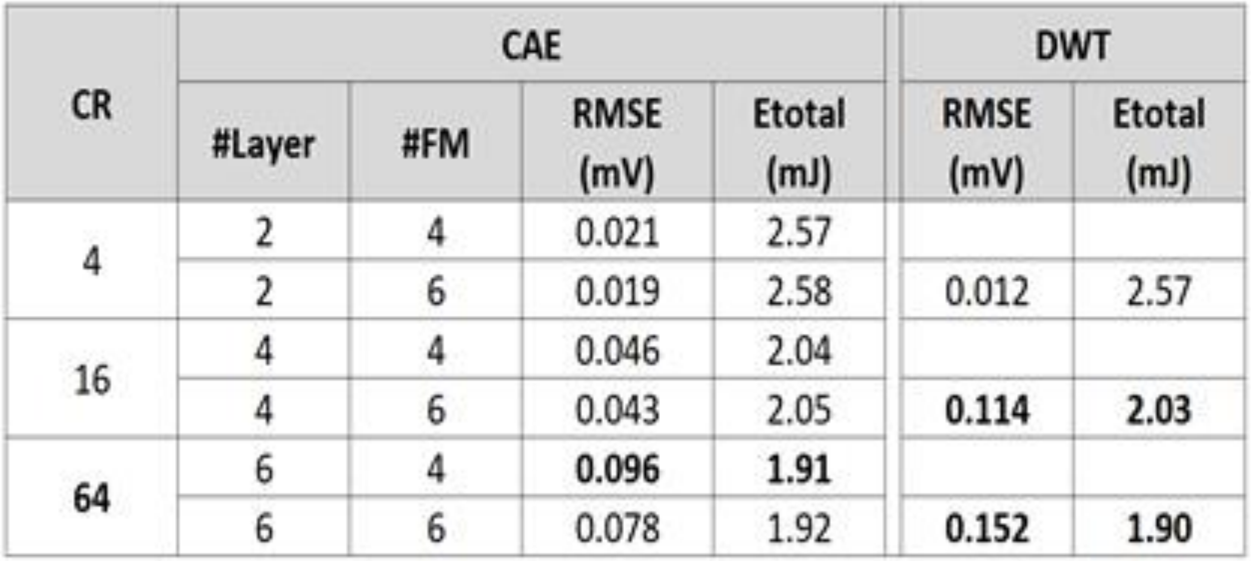
### 2.3.3 Performance Summary

We have further compared CAE and DWT in Table 2.1. There are several interesting observations. Firstly, as the compression ratio increases, the RMSE of CAE gradually increases but the RMSE of DWT increases quickly. This indicates again the superior learning ability of CAE. Secondly, under similar ‘Etotal’, e.g., 1.91 mJ for CAE-CR64-FM4 and 1.90 mJ for DWT-CR64, RMSE of CAE is only 0.096 mV, but it is as high as 0.152 mV for DWT. Thirdly, if we compare CAE-CR64-FM4 and DWT-CR32, we can find both RMSE and ‘Etotal’ of CAE are better than DWT significantly. This further comparison clearly demonstrates the superiority of our proposed CAE approach.

### 2.3.4 Reconstructed Signals

Another demonstration is given in Fig. 2.4., which illustrates the reconstructed ECG signals using CAE and DWT for different users. The comparison clearly shows that CAE is more effective in ECG signal compression and reconstruction. Fig. 2.4(a1) and (b1) give the comparison of CAE and DWT for two different users, and the three selected cases correspond to the three cases in Table 1 emphasized with the bold font format. Fig. 2.4 (a2) and (a3) show the reconstruction error when increasing the compression ratio for CAE and DWT, respectively. Fig. 2.4 (b2) and (b3) show similar conditions for another user. Clearly, CAE is less sensitive to reconstruction error thanks to the deep learning ability.

Table 2.1: Performance Summary



| CR | CAE | | | | DWT | |
|---|---|---|---|---|---|---|
| | #Layer | #FM | RMSE (mV) | Etotal (mJ) | RMSE (mV) | Etotal (mJ) |
| 4 | 2 | 4 | 0.021 | 2.57 | | |
| | 2 | 6 | 0.019 | 2.58 | 0.012 | 2.57 |
| 16 | 4 | 4 | 0.046 | 2.04 | | |
| | 4 | 6 | 0.043 | 2.05 | **0.114** | **2.03** |
| **64** | 6 | 4 | **0.096** | **1.91** | | |
| | 6 | 6 | 0.078 | 1.92 | **0.152** | **1.90** |

Notes. CR: compression ratio; FM: feature map;
Etotal: total energy.

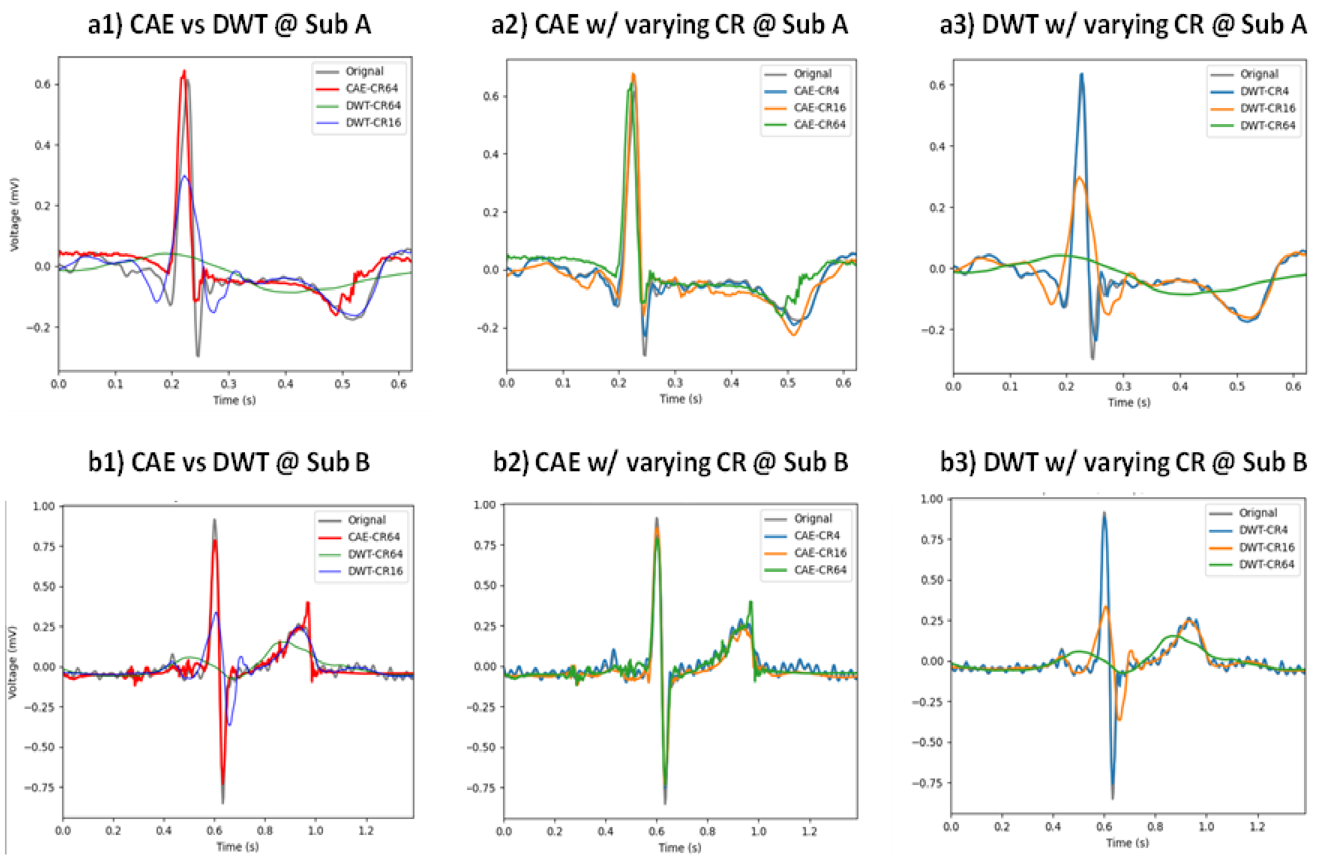


Figure 2.4. The reconstructed ECG signals using CAE and DWT, which further illustrates that the proposed CAE can more effectively compress and reconstruct the ECG signals. a1) and b1) give the comparison of CAE and DWT for two different users, and the three selected cases correspond to the three cases in Table 1 highlighted with bold font format. a2) and a3) show the reconstruction error when increasing the compression ratio for CAE and DWT, respectively. Clearly, CAE is less sensitive to reconstruction error thanks to the deep learning ability. b2) and b3) show similar conditions for another user. When the compression ratio is increased for DWT, the reconstruction error increases more quickly than CAE.

## 2.4 Conclusion

In this study, we have proposed deep learning-empowered energy-efficient wearable monitoring towards big data applications. To tackle the challenge faced by wearable monitors in long-term continuous monitoring, i.e., the energy-hungriness, we have researched and developed a data-driven approach for intelligent data compression. We leverage deep learning, more specifically, CAE, to learn a sparse representation of the vital big data. The minimized energy need, even taking into consideration the CAE-induced overhead, is tremendously lowered. The superiority of the CAE framework has also been demonstrated by our experimental results, in comparison with the state-of-the-art wavelet compression-based method. The validated approach is expected to greatly advance energy-efficient wearable monitoring in the era of big data.

# 3. ULTRA-EFFICIENT EDGE INTELLIGENCE TOWARDS REAL-TIME, LONG-TERM BIG DATA MINING

## 3.1 Introduction

Technologies have advanced the emerging smart health applications. The physical activity monitoring, cardiac health monitoring, lifestyle management, and many more practices, are paving the way of modern health monitoring [12-16]. However, the real-time, long-term big data mining and inference are still highly challenging, due to the fact that inference algorithms, especially deep learning, usually require heavy computation requirements, thereby greatly limiting their deployment on the edge. Targeting the deep learning's high computational complexity and large memory and energy requirements, new approaches are urged to make the deep learning algorithms ultra-efficient for wearable big data mining and inference.

Wearable computers have undergone rapid development in the recent years and can gather valuable biomedical information about the health condition of the users [17-19]. Advanced machine learning or deep learning algorithms can further analyze the big data to generate medical insights and/or send notifications or alerts to the users and the designated medical professionals if abnormalities are detected. This methodology could save the medical professionals' time to perform a manual inspection of the patient data and could also minimize the manual inspection errors. The wearable real-time, long-term big data mining and inference system is shown in Fig. 3.1, to illustrate the application scenarios.

Deep learning has demonstrated its amazing capabilities recently, in domains such as medical imaging, fraud detection, autonomous vehicles, language translation, and many others. Meanwhile, deep learning models typically consist of a tremendous number of parameters and have high computational complexity. For instance, the VGG16 model [20] that is typically used for image recognition tasks contains 138,357,544 (about 138 million) parameters. The number of floating-point operations (FLOPs) is approximately 15 billion, making it a model with extreme high complexity. Krizhevsky *et al.* [21] also proposed a model that has approximately 60 million parameters to perform multi-class image classification on a dataset that consists of 1000 different classes. The training of the final model was time-consuming as it took several days. It is very challenging or even impractical to deploy these computational-intensive models on devices that has limited hardware resources such as the health wearables [22].

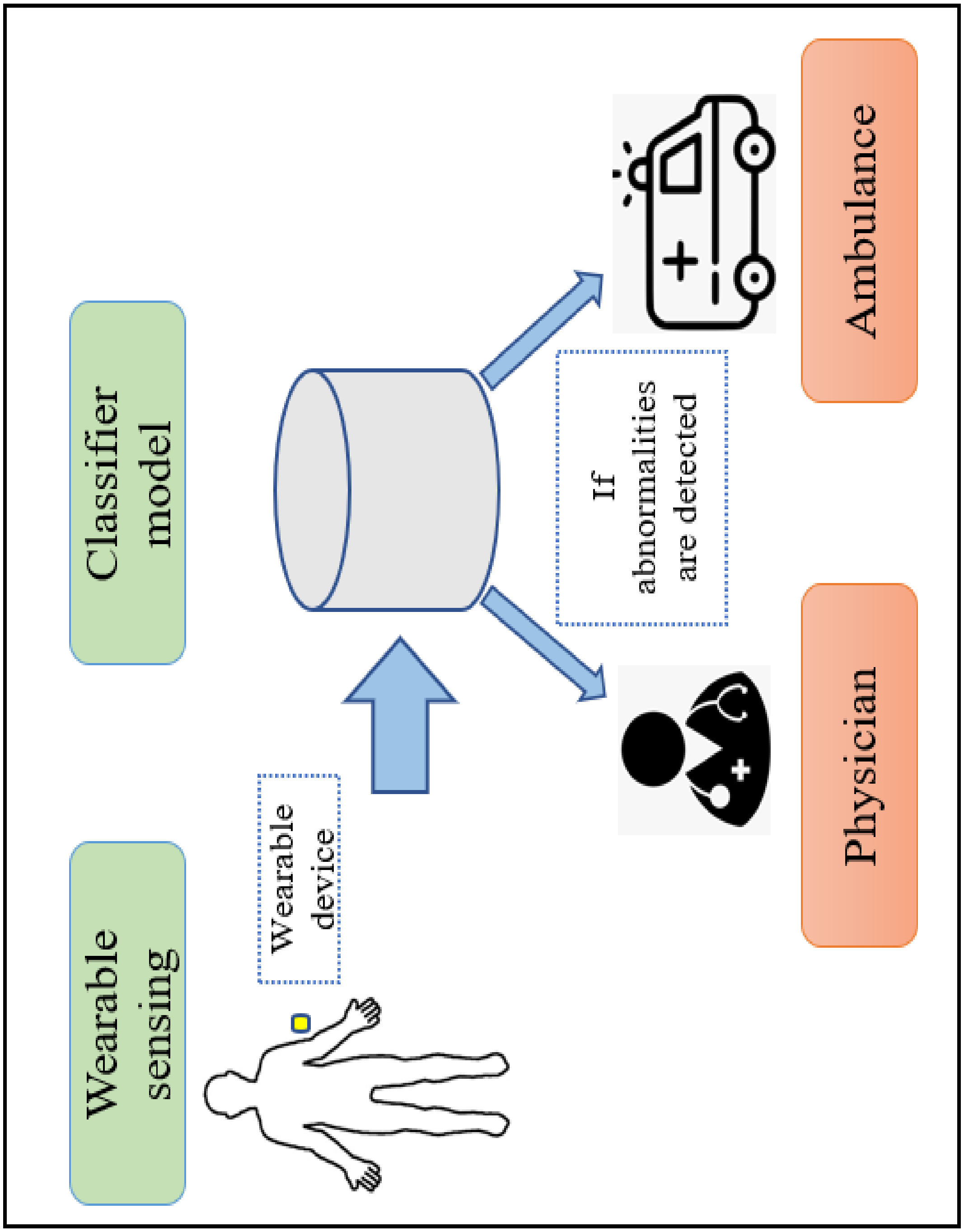


Figure 3.1. The wearable real-time, long-term big data mining and inference system is expected to greatly advance the smart health applications.

This study is focused on the Electrocardiogram-based cardiovascular monitoring application, to investigate the challenging edge-deployable deep learning algorithms. The primary cause of death around the world now is the cardiovascular disease (CVD). 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 [23]. These conditions may lead to heart failure and premature death [24]. A person with arrhythmia may face cardiac arrest as the impaired heart's pumping action is disrupted and unable to supply blood to the human body. While there are arrhythmia conditions that do not pose an immediate threat to a person's life, 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.

In the medical industry, the procedure for diagnosis of arrhythmia primarily consists of visual inspection of electrocardiogram (ECG) signal by a clinical physician [25]. It is a non-invasive measurement of the heart's rhythm and can be analyzed by the medical professionals to categorize it into normal and abnormal heartbeats [26]. The arrhythmia condition disrupts the electrical activity of the heart as the heart undergoes depolarization and repolarization and thus could be detected by assessing the ECG signal. 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 wearable data can be very challenging for medical professionals.

Various studies have been performed to study different traditional machine learning methodologies to perform ECG classification. Yeh *et. al.* proposed a fuzzy c-means method, a simple and reliable algorithm that performs classification of ECG signals by calculating the cluster centers for each class [27]. Taiyong *et. al.* introduced a method using wavelet packet entropy (WPE) to decompose the ECG signals and calculate the entropy, before implementing random forests (RF) to perform classification task [28]. Sharma *et al.* presented a filter bank (FB) method and a k-nearest neighbor (KNN) classification algorithm [29]. 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 [30].

In contrast with the typical machine learning algorithms, the recent development of deep learning models to perform arrhythmia detection has proven that it is an effective method for both feature extraction and classification. Deep learning methodologies are expected to further advance the current studies of ECG-based heart disease detection. Acharya et al. presented a deep convolutional neural network (CNN) that is comprised of multiple layers for ECG-based heart disease classification [31]. Mostayed et al. proposed a different neural network structure, recurrent neural network (RNN) that consists of 2 long-short-term-memory (LSTM) layers to detect abnormalities in the ECG data [32].

At the same time, deep learning models may be susceptible to adversarial attacks [33-35] 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. These could have a detrimental effect on a patient's physical fitness as the classifier model installed in the wearable device could misclassify abnormal beats as normal beats, providing incorrect information to the patient.

In this paper, we have presented ideas to address these shortcomings. We propose to leverage knowledge distillation to achieve an ultra-efficient edge-deployable deep learning model. More specifically, through transferring the knowledge from a teacher model to the on-edge student model, the soft target distribution of the teacher model can be effectively learned by the student model. Besides, we propose to further introduce adversarial robustness to the student model, by stimulating the student model to correctly identify inputs that have adversarial perturbation. Our experiments have demonstrated the effectiveness of the proposed ultra-efficient and robust deep learning framework, towards real-time, and long-term wearable big data mining and inference applications. Below we will detail our approach and the experimental results.

## 3.2 Approaches

### 3.2.1 System Overview

The system overview is given in Fig. 3.2. Given the diverse biomedical abnormalities, the ECG signals are highly diverse for arrhythmia patients. The traditional deep learning methods usually need heavy models that have large amounts of parameters. With the proposed efficient dynamics learning framework, the robust knowledge distillation principles are investigated to generate edge-deployable and lightweight deep learning model.

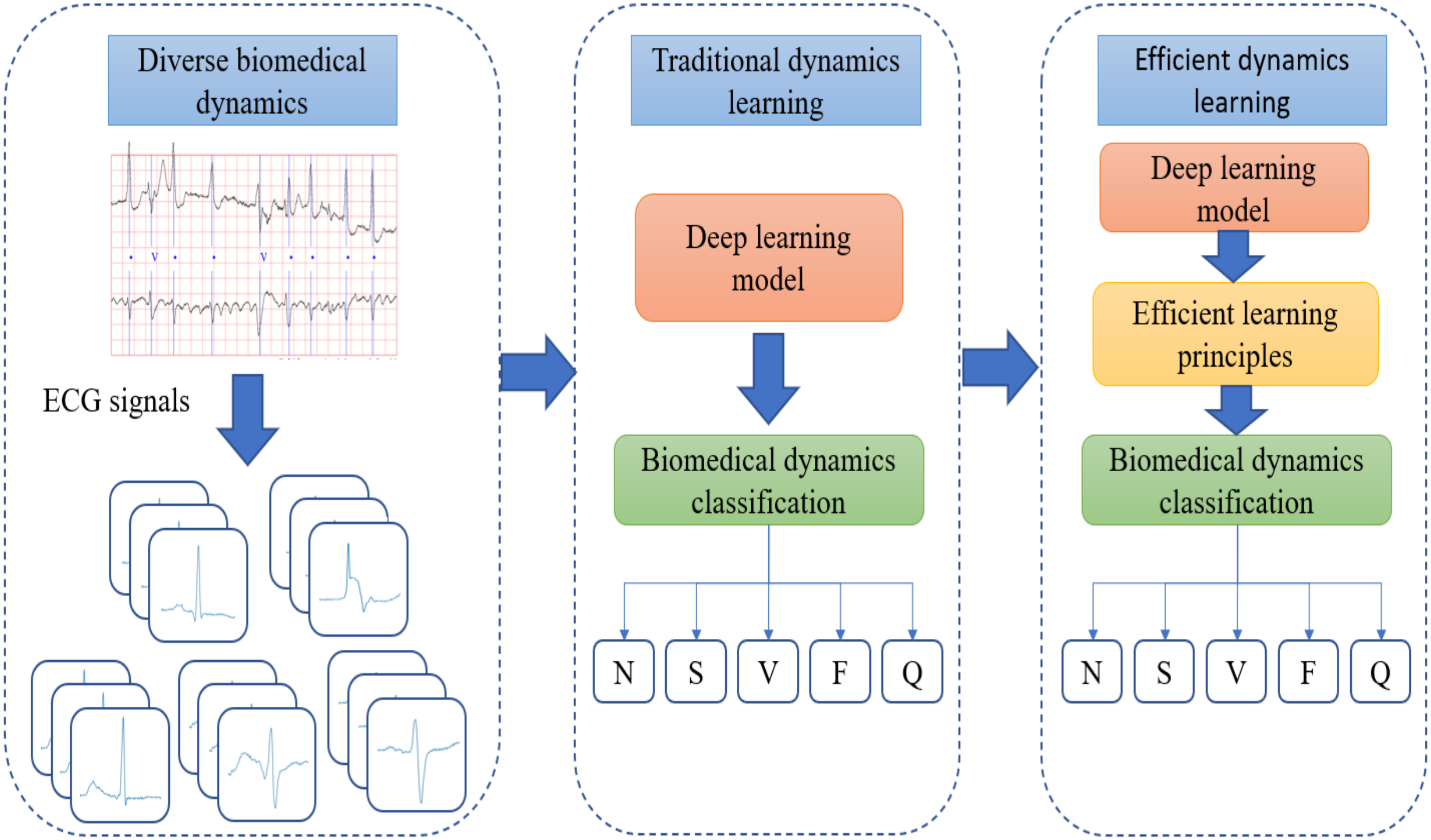

Diverse biomedical dynamics
ECG signals
Traditional dynamics learning
Deep learning model
Biomedical dynamics classification
N
S
V
F
Q
Efficient dynamics learning
Deep learning model
Efficient learning principles
Biomedical dynamics classification
N
S
V
F
Q

Figure 3.2. The system diagram of the Ultra-Efficient Edge Intelligence with efficient learning principles to compress the deep learning model. Note: N/S/V/F/Q – correspond to five categories of heartbeats to be detailed in the results section.

### 3.2.2 Residual Convolutional Neural Network (ResNet)

We have firstly built a residual CNN – ResNet [36] for ECG-based multi-class heart disease classification. CNN models typically consist of three categories of layers: a convolutional layer, a pooling layer, and a fully connected layer. ResNet builds on constructs inspired by neural cells with skip connections. As shown in Fig. 3.3, 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 is comprised of 15 layers: 8 convolutional layers, 4 max-pooling layers and 3 fully- connected layers. 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 final layer of the fully-connected layer is comprised of five neurons that correspond to the number of classes (to be detailed in the results section).

The ResNet model were implemented to classify the ECG signal.  Here the one-dimension raw ECG signal is directly fed into ResNet. This model will act as the teacher model to distill the knowledge to a much smaller model with knowledge distillation (given later). In both KDL and RP-KDL learning approaches, the output of this ResNet will be used as another source of ground truth to enhance the training process of the student model.

### 3.2.3 Knowledge Distillation Learning (KDL)

KDL is known as a model compression method in which a large and heavy model, commonly denoted as the teacher model, transfers knowledge to a light-weight model, denoted as the student model [37]. The process is implemented by training the student model to ‘imitate’ the teacher model without a significant loss of performance validity.

$$\theta_S = \underset{\theta_S}{argmin}\, \mathbb{E}_{(X,\mathbb{y})\sim\mathcal{D}} \begin{bmatrix} \alpha\tau^2 \mathcal{L}_{S\to T}\big(\Phi_S(X,\theta_S,\tau)\,,\Phi_T(X,\theta_T,\tau)\big) \\ +(1-\alpha)\mathcal{L}_{S\to\mathbb{y}}(\Phi_S(X,\theta_S),\mathbb{y}) \end{bmatrix} \tag{1}$$

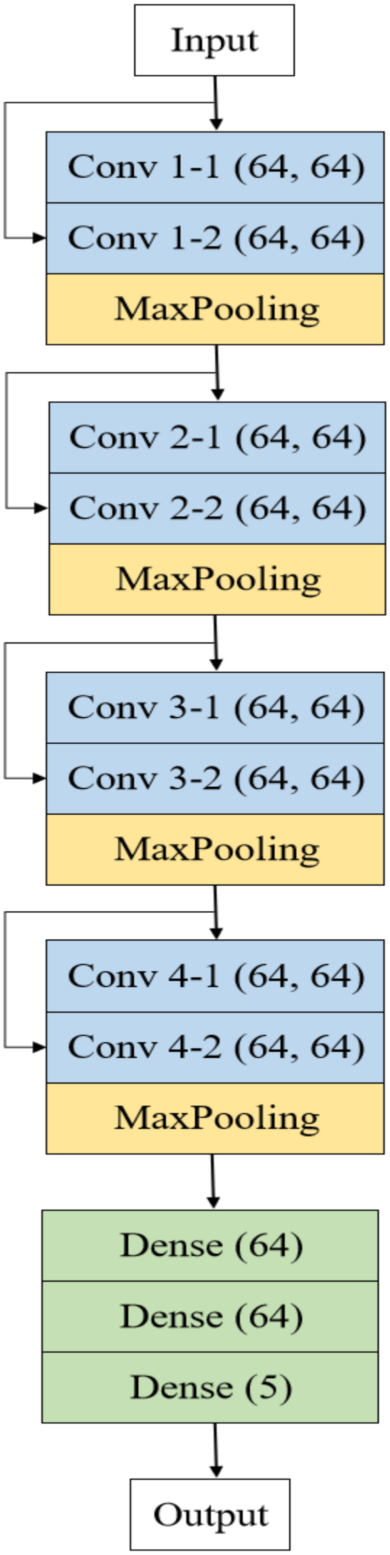


Figure 3.3. The ResNet model that is treated as the teacher model.

The teacher model is able to generalize and its learning is transferred to the student model using the class probabilities produced by the teacher model's soft target distribution as given in (1), where $\theta_S$ is the parameter set for the student model $\Phi_S$, and $\theta_T$ is the parameter set for the teacher model $\Phi_T$. The KDL learning process is searching $\theta_S$ by minimizing the loss across the dataset $(X, \mathbb{y}) \sim \mathcal{D}$, where $X$ and $\mathbb{y}$ are input and ground truth, respectively. $\mathcal{L}_{S \to T}$ is the loss for the student to emulate the teacher, and $\mathcal{L}_{S \to \mathbb{y}}$ is the loss for the student to generate ground truth-referred output. $\tau$ is the temperature to control the softness of the generated target distribution, and $\alpha$ 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.

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

### 3.2.4 Robustness-Preservation KD Learning (RP-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 [38] to KD, we here achieve RP-KDL for the ECG-based heart disease detection application, as shown in Fig. 3.4. It is expected to enhance the robustness of the KD model by learning how to identify inputs with perturbations.

The conception of defense against adversarial attacks can be represented in a natural saddle point (min-max) formulation. The min-max optimization problem is formulated in (2-3).

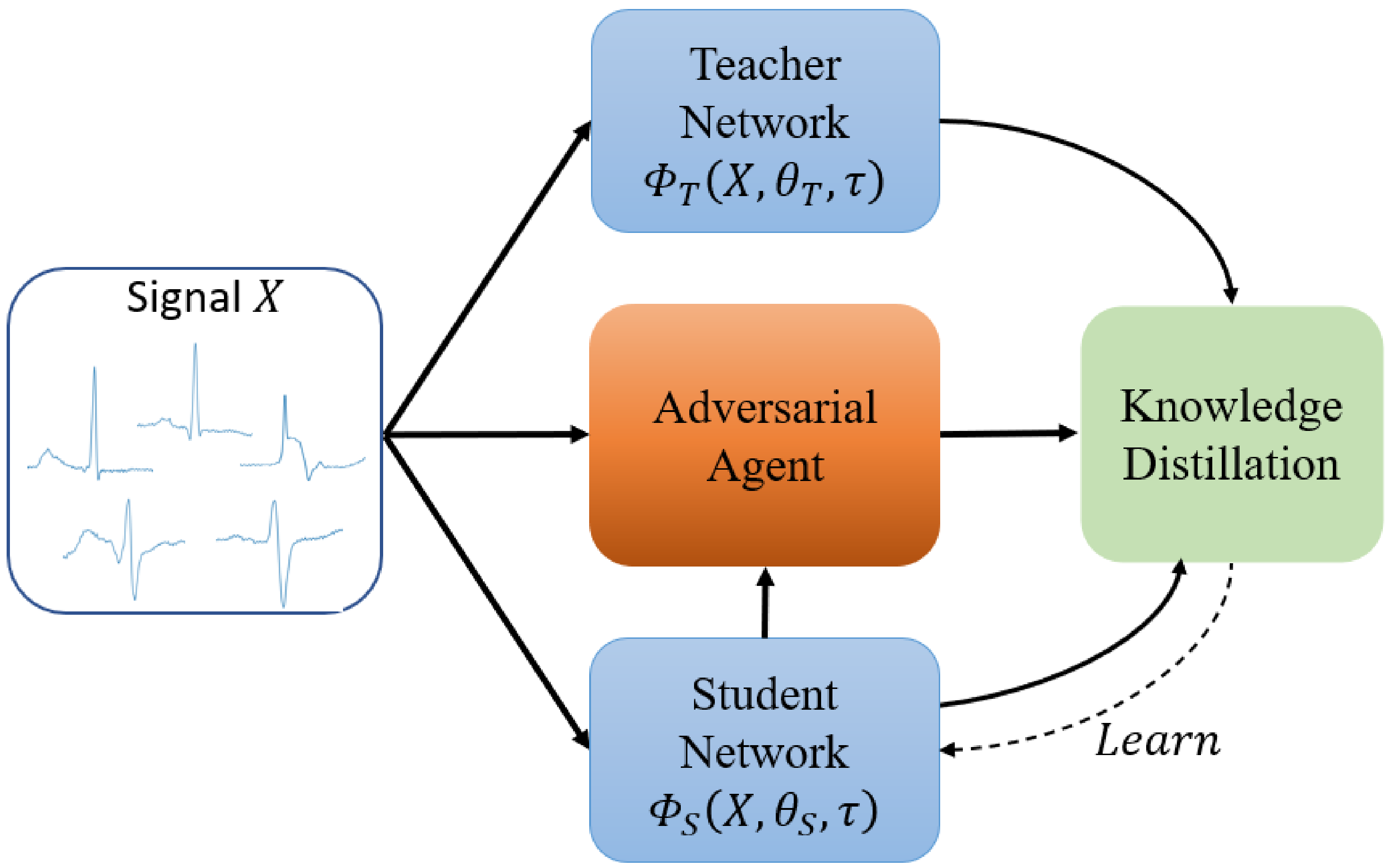

Signal $X$
Teacher Network $\Phi_T(X, \theta_T, \tau)$
Adversarial Agent
Student Network $\Phi_S(X, \theta_S, \tau)$
Knowledge Distillation
$Learn$

Figure 3.4. Robustness Preservation KD Learning to enhance the robustness of the student model.

$$\theta_S = \underset{\theta_S}{argmin}\, \mathbb{E}_{(X,\mathbb{y})\sim\mathcal{D}} \begin{bmatrix} \alpha\tau^2 \mathcal{L}_{S\to T}\left(\Phi_S(X',\theta_S,\tau)|_{X'=X+\,\delta_{X,\theta}}\,, \Phi_T(X,\theta_T,\tau)\right) \\ +(1-\alpha)\mathcal{L}_{S\to\mathbb{y}}(\Phi_S(X,\theta_S),\mathbb{y}) \end{bmatrix} \tag{2}$$

$$\delta_{X,\theta_S} = \underset{\|\delta\|_p < \,\varepsilon}{argmax}\, \mathcal{L}_{S\to\mathbb{y}}(\Phi_S(X',\theta_S),\mathbb{y})|_{X'=X+\,\delta} \tag{3}$$

The distillation loss is now calculated between the student model with the adversarial response and the teacher model. The input $X'$ is determined by pertubing the input $X$ by $\delta_{X,\theta_S}$, which is searched by the adversarial agent as shown in Fig. 3.4.

### 3.2.5 Evaluation Method

We have applied multiple criteria to assess the performance of the deep learning models and learning approaches.

Accuracy is firstly used to evaluate the classification performance of the model as defined in (4), where *TP* represents true positives, *FP* represents false positives, *TN* represents true negatives, and *FN* represents false negatives.

$$A = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

Two types of accuracy are measured in this study: namely natural accuracy and robust accuracy. Natural accuracy refers to the accuracy when the classifier model is fed with clean ECG data, whereas robust accuracy refers to the accuracy when the model is dealing with adversarial samples.

Precision is used to measure the ratio between the true positives and all the positives. It can be expressed as (5).

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

Equation (6) indicates the recall which is a measure of the model correctly identifying the true positives among positive samples.

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

*F1* score is an overall performance score that combines precision and recall. It is formulated as (7):

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{7}$$

In this study, the number of parameters of a model is determined by calculating the total of non-zero parameters at each layer of the model. The dataset is split into 3 different sets: 80% training dataset, 10% validation dataset, and 10% testing dataset.

## 3.3 Results

### 3.3.1 Experimental Setup

PhysioNet MIT-BIH Arrhythmia Database is used as the data source for the paper [39]. The dataset is comprised of two-channel ECG recordings of 47 subjects, with each channel lasting a half hour. 23 out of the 48 recordings are arbitrarily selected from a set of 4000 ECG recordings (each lasting 24 hours) from a group of inpatients (accounting about 60%) and outpatients (accounting about 40%) at the Boston’s Beth Israel Hospital. The other 25 recordings consist of recordings that are comprised of clinically significant but rare arrhythmias from the same set.

The ECG recordings are sampled at a frequency of 360 Hz with a resolution of 11 bits. The total number of ECG beats extracted for the training, validation and testing process is 109, 441.

In this study, the ECG data are classified into five distinctive classes according to the American Association for the Advancement of Medical Instrumentation (AAMI) standard [40]. The 5 classes that the data samples can be divided into are: normal (N) beats, unknown (Q) beats, ventricular ectopic (V) beats, supraventricular ectopic beats (S) and fusion (F) beats [41].

The ECG samples are segmented into individual beats of a fixed length based on the annotation files. Examples of the ECG samples of the five classes are visualized in Fig. 3.5. The diverse morphologies indicate the complexity of the heart disease detection problem.

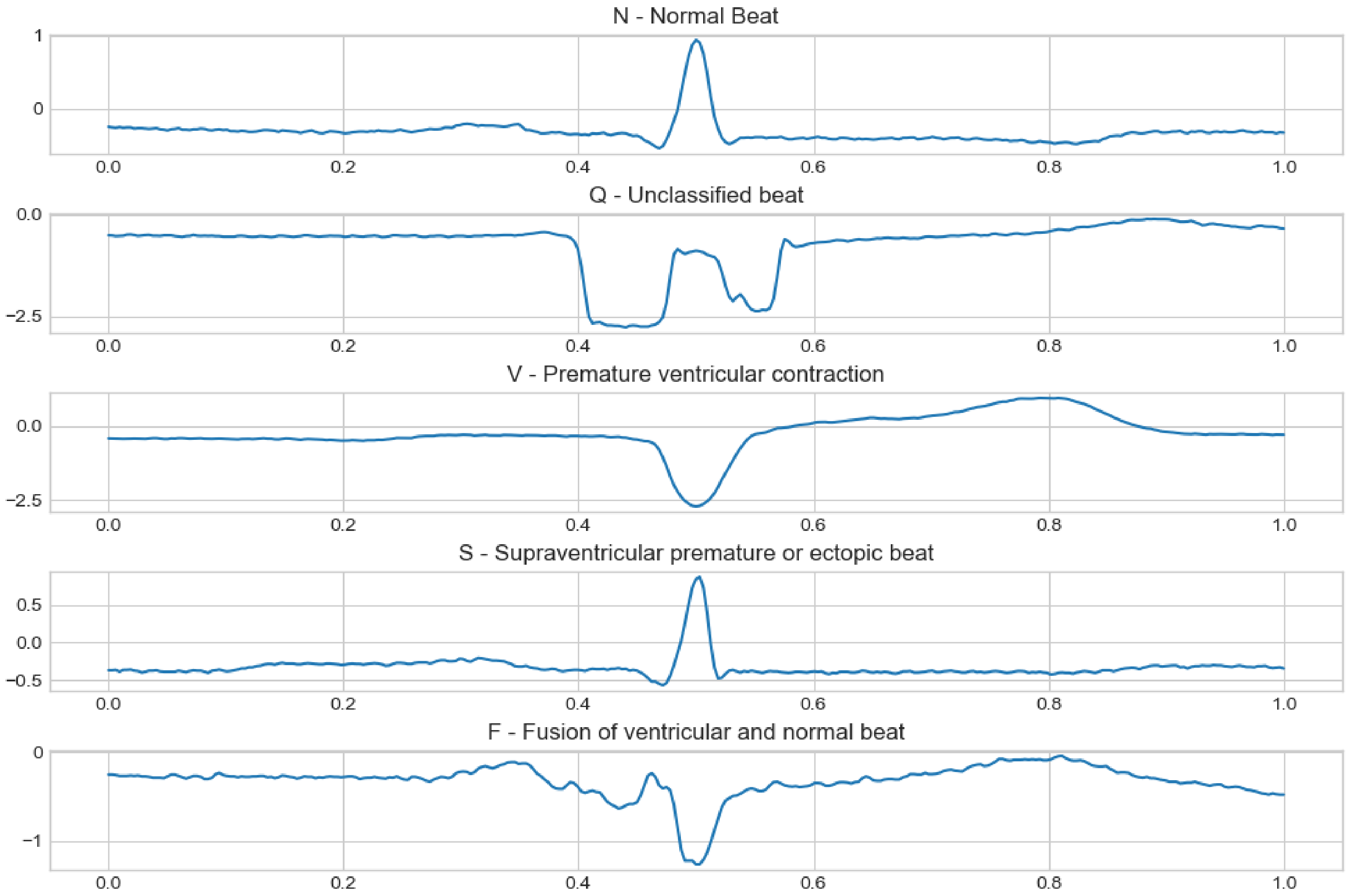


Figure 3.5. Visualization of ECG samples from each class.

### 3.3.2 Deep Residual CNN Learning

Fig. 3.6 and 3.7 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. 3.7 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.

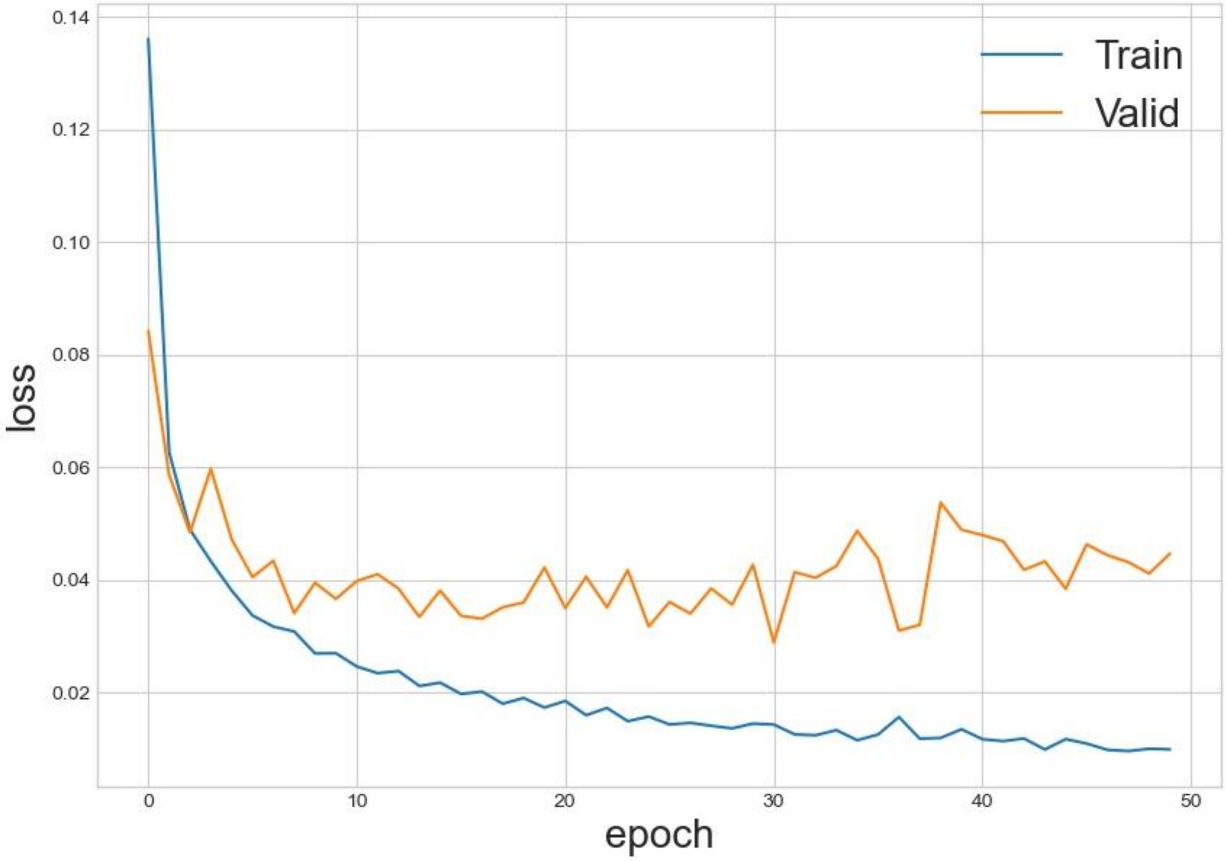


Figure 3.6. Learning curves of the teacher ResNet model.

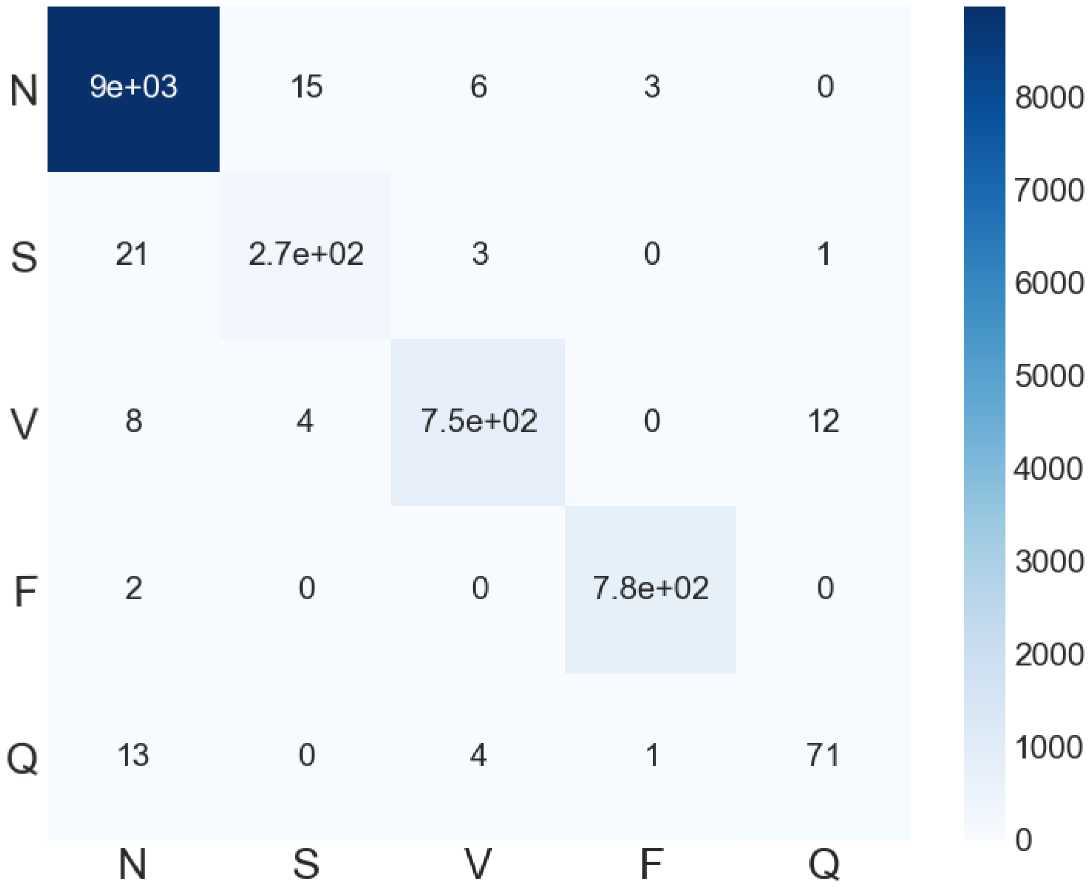


Figure 3.7. Confusion matrix @ the teacher deep model. Note: the vertical axis is the ground truth.

### 3.3.3 Knowledge Distillation Learning

Fig. 3.8 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.

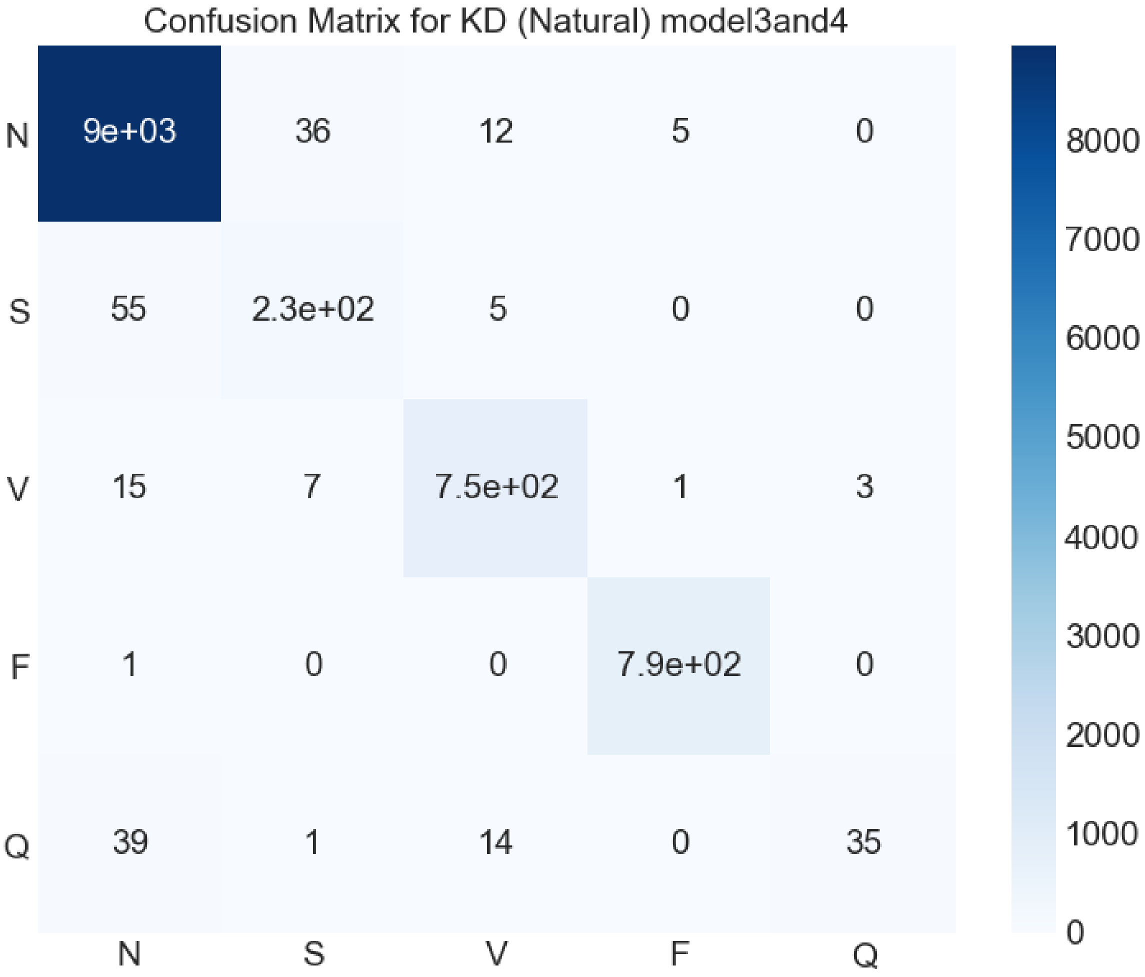


Figure 3.8. Confusion matrix @ the student deep model learned with KDL (standard ECG signal).

### 3.3.4 Adversarial Agent

Adversarial attacks are used to generate the perturbed samples, as shown in Fig. 3.9. It is a white-box attack and thus, it is under the assumption that the adversaries have full knowledge of the target model's parameters and structure. It is considered more malicious than black-box attacks (the attacker has no knowledge of the model's parameters and architecture) as the adversaries could design their attack to "fool" the classifier model by adding slight modifications to the data samples. The adversarial samples that are slightly altered by the perturbations should be considerably similar to the original data sample, such that the difference is imperceptible to humans' naked eyes but subtle enough to lead to misclassification by the classifier model. The visualization clearly shows that the adversarial samples are very similar to the standard ECG signal, indicating the changes posed to the heart disease detection task.

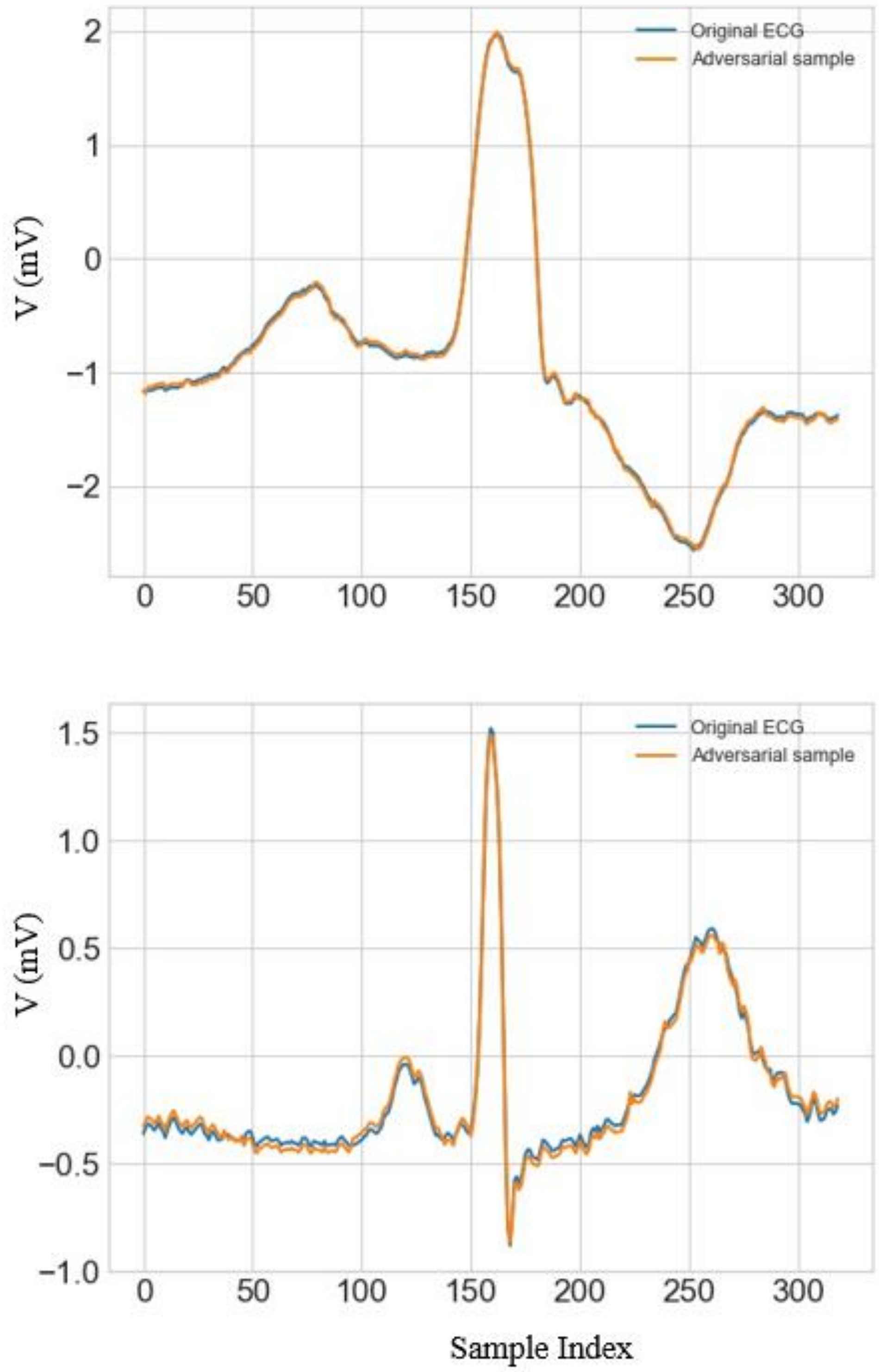


Figure 3.9. Attack signals generated by the adversarial agent.

### 3.3.5 KDL Under Adversarial Attacks

With adversarial attacks, KDL misclassifies a significant number of heartbeats, as shown in Fig. 3.10. A huge difference can be noticed when the model is facing adversarial attacks as compared to facing clean data in Fig. 3.8. 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.

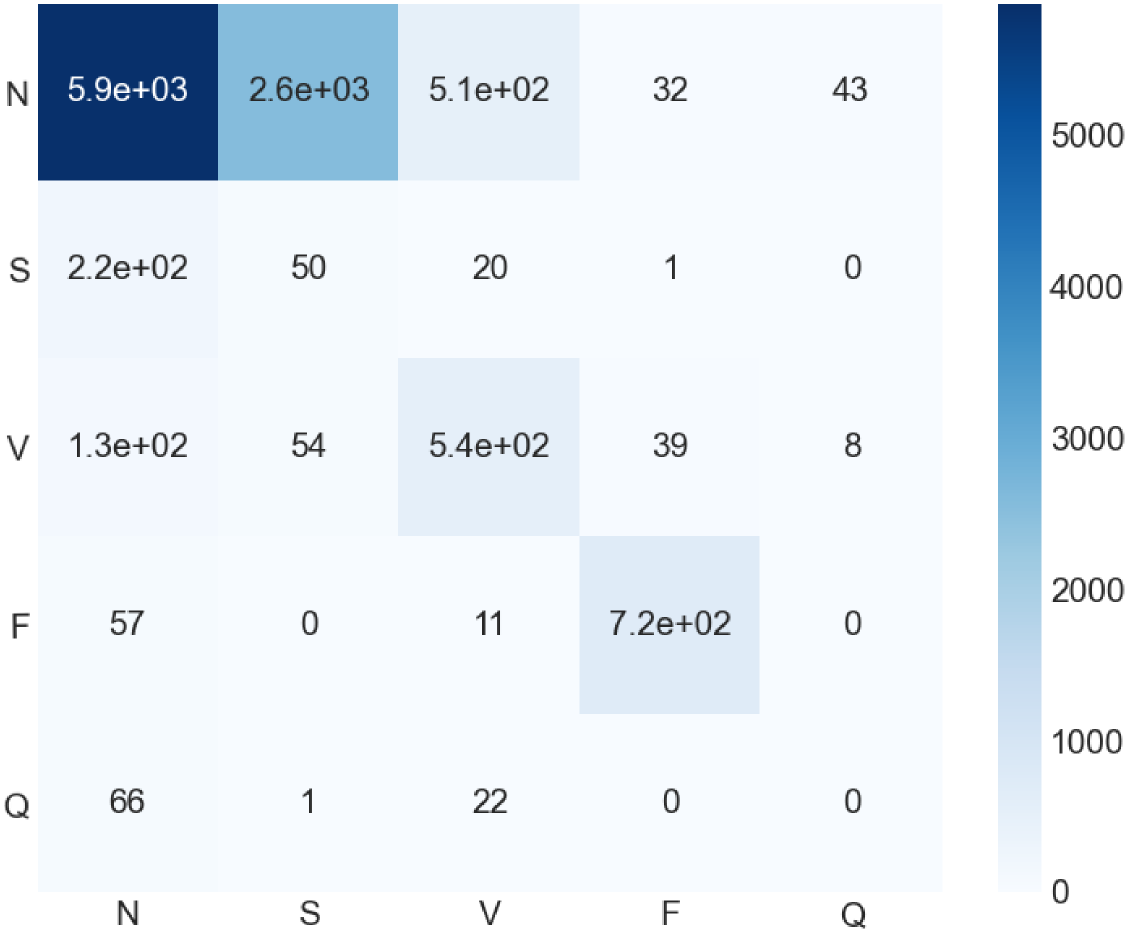


Figure 3.10. Confusion matrix @ the student deep model learned with KDL (adversarial ECG signal).

### 3.3.6 RP-KDL under Adversarial Attacks

The confusion matrix of the RP-KDL student model with adversarial samples is given in Fig. 3.11, which clearly demonstrate the enhanced robustness of the student model. For N/S/V/F classes, the performance has been all greatly improved, except for the Q class that has minimum number of heartbeats. Overall, our RP-KDL framework has greatly boosted the performance of the KDL framework, by introducing substantial robustness during the learning process.

For comparison purpose, the RP-KDL student model with standard samples is also given in Fig. 3.12. The performance is bit better, which is reasonable since the input is the standard and clean ECG signal.

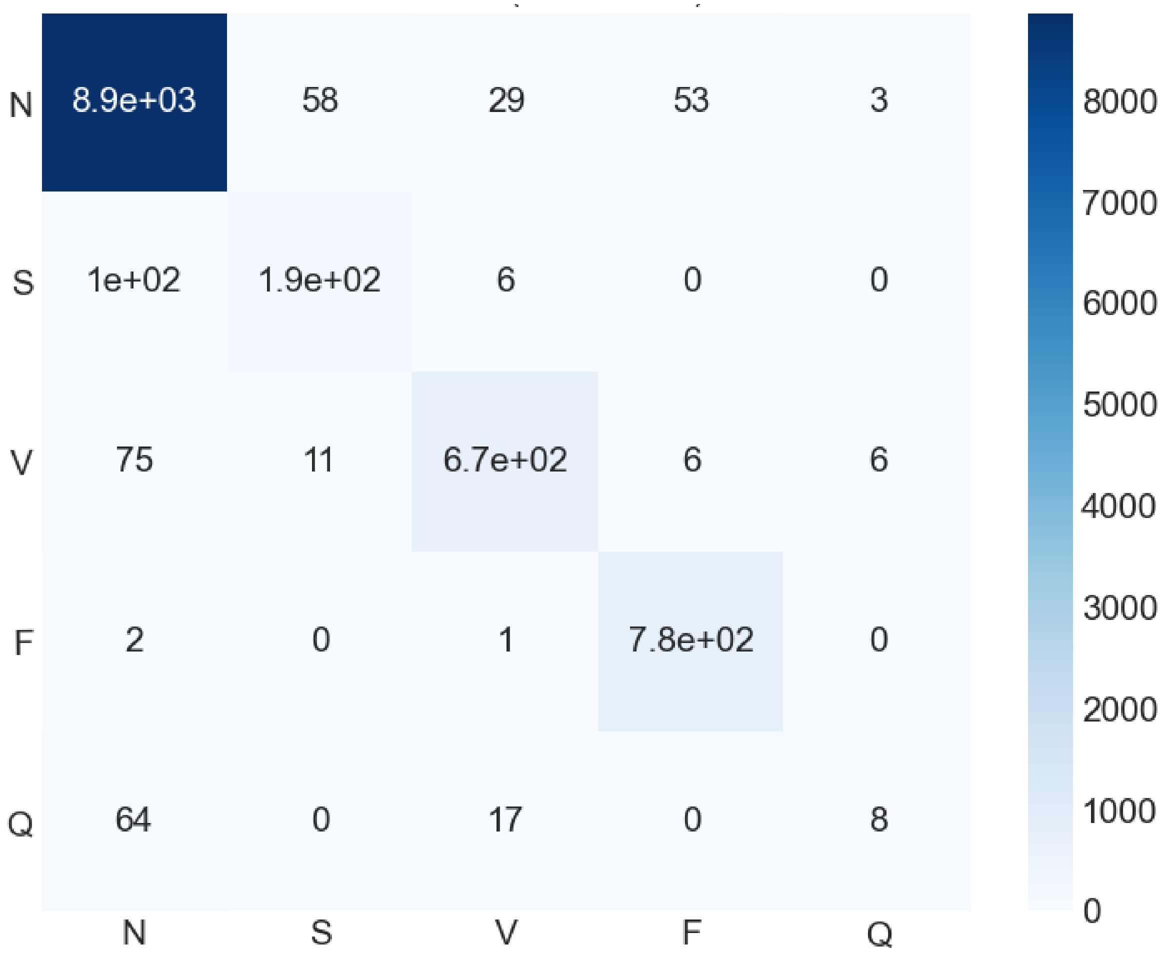


Figure 3.11. Confusion matrix @ the student deep model learned with RP-KDL (adversarial ECG signal).

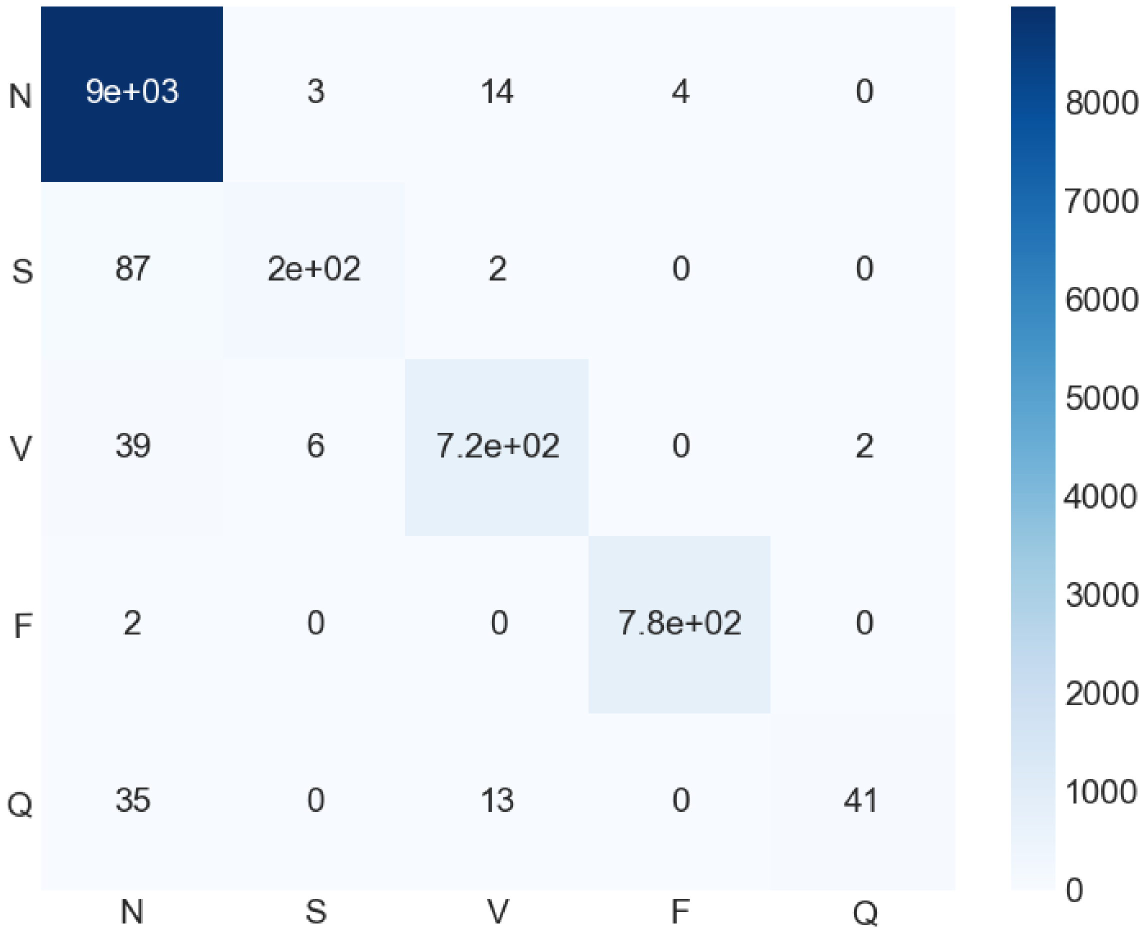


Figure 3.12. Confusion matrix @ the student deep model learned with RP-KDL (standard ECG signal).

### 3.3.7 Comparison between KDL and RP-KDL

To further compare KDL and RP-KDL, both natural accuracy and robust accuracy graphs are given.

In Fig. 3.13, the natural accuracy evaluated on the standard ECG signal is given for three models: teacher, KDL student model, RP-KDL student model. 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, the robustness introduced to RP-KDL is not reflected here but will be detailed in Fig. 3.14.

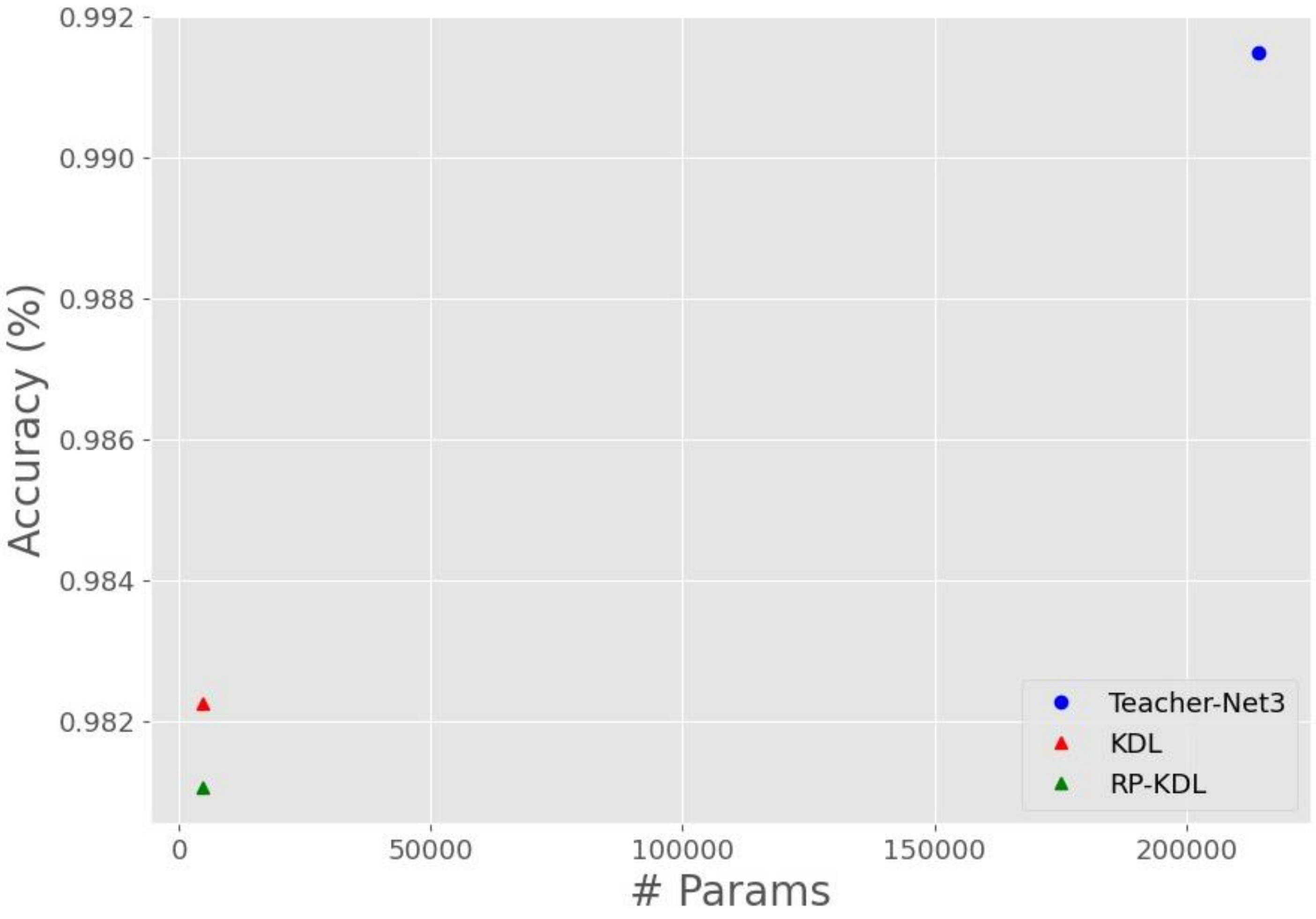


Figure 3.13. Natural accuracy of three models: teacher, KDL student model, and RP-KDL student model. The latter two show similar performance that is close to the teacher model.

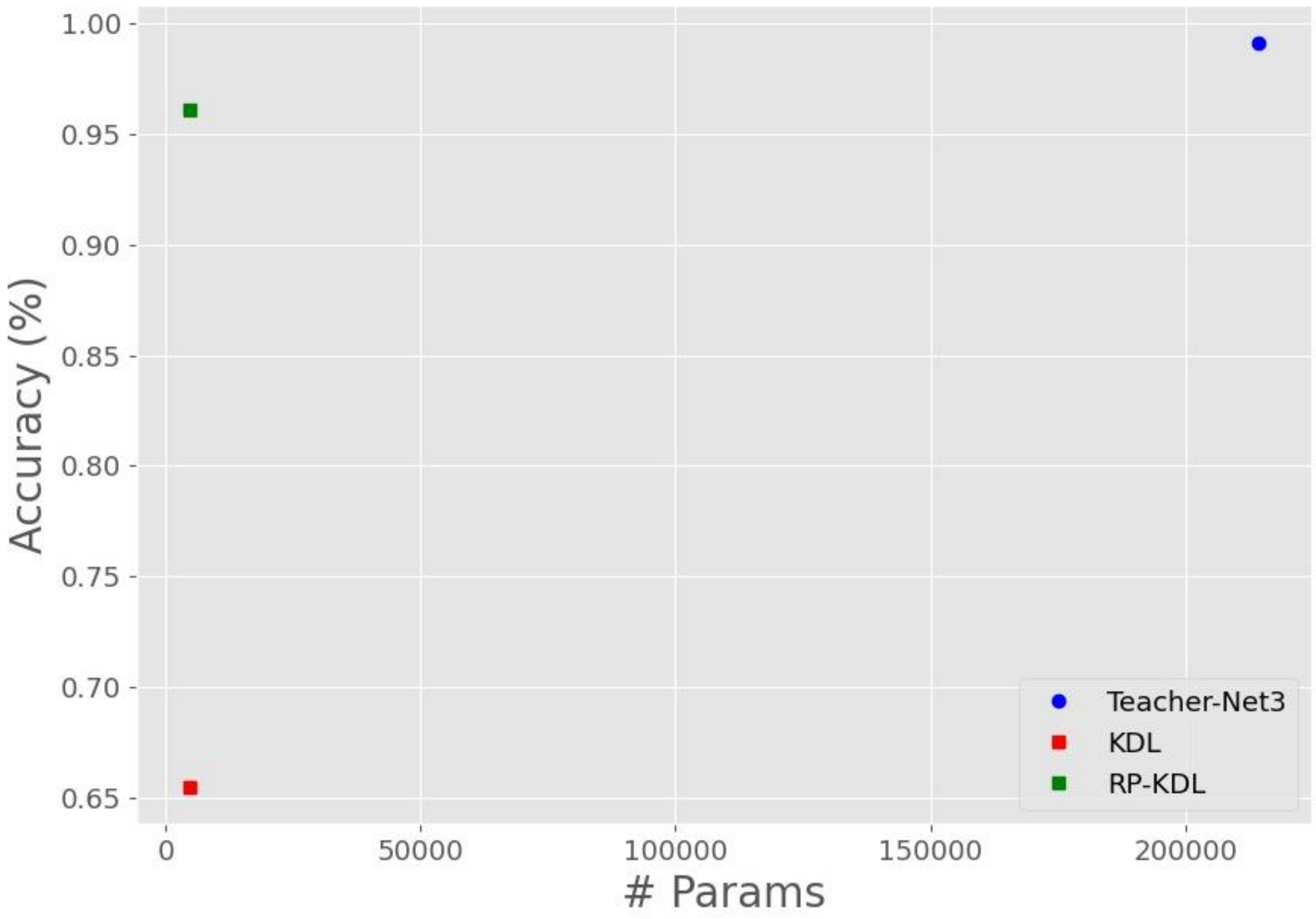


Figure 3.14. Robust accuracy of three models: teacher, KDL student model, and RP-KDL student model. The RP-KDL student model shows enhanced robustness under adversarial attacks.

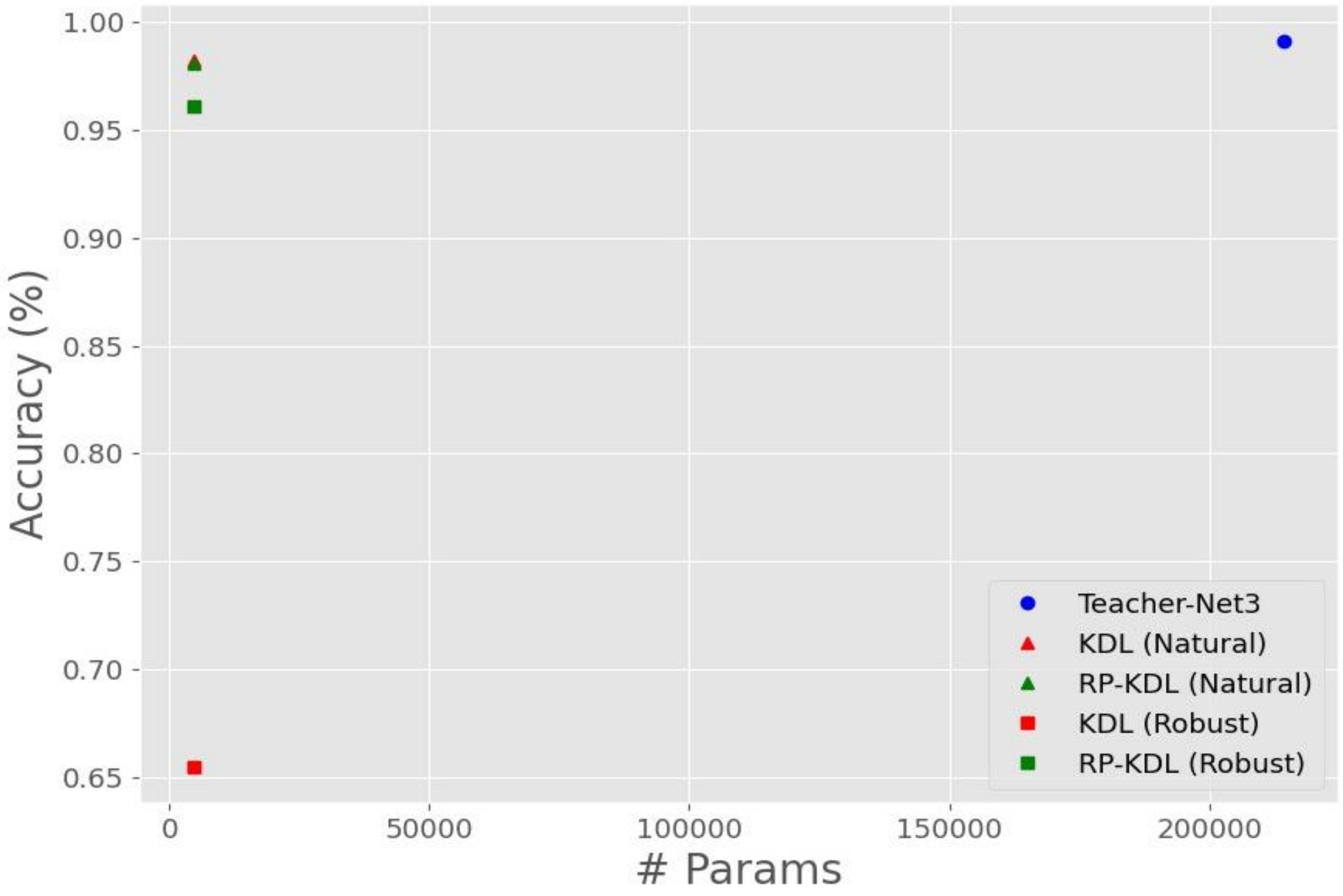


Figure 3.15. Co-visualization of natural and robust accuracy of three models: teacher, KDL student model, and RP-KDL student model.

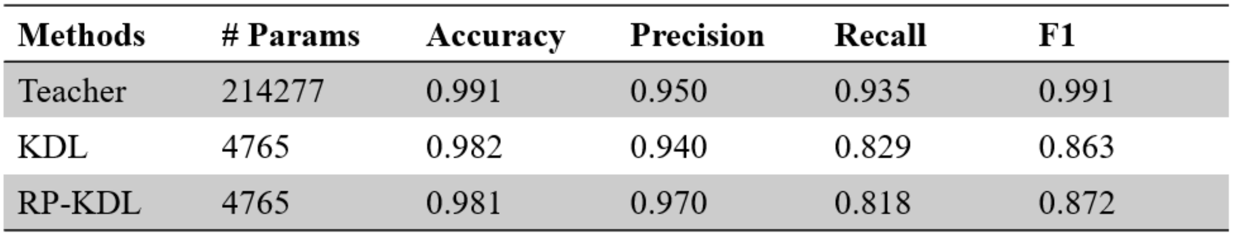
The robust accuracy of the models is shown in Fig. 3.14. As observed from the visualization, a huge percentage of adversarial samples are able to fool the KDL student model to generate incorrect detection results. However, the RP-KDL student model is able to defend against adversarial attacks. While facing adversarial noises, the RP-KDL student model remains strongly resilient to perturbations and the classification accuracy is still very high. Notice that the robust accuracy of the RP-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 of three models is also given in Fig. 3.15, to further demonstrate the performance drop of the KDL student model and the robustness of the RP-KDL model.

### 3.3.8 Performance Summary of KDL and RP-KDL

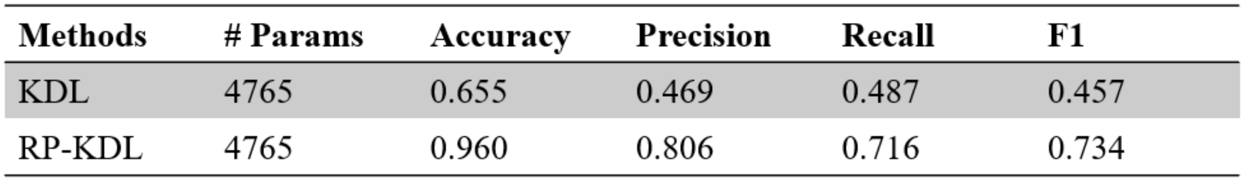
The performance of the teacher model, and KDL student model and RP-KDL student model is summarized in Table 3.1 and 3.2, corresponding to natural and robust performance, respectively. This summary further demonstrates: (1) the KDL student model can effectively learns the knowledge from the teacher model for standard ECG data, but with a much smaller model size – only 4765 parameters compared with the 214,277 parameters of the teacher model; (2) the RP-KDL model further enhanced the robustness of the student mode and achieves very promising performance under adversarial signals – the accuracy/precision/recall/F1 score are 0.960, 0.806, 0.716, and 0.734, respectively.

Table 3.1. 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 |

Table 3.2. 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 |

## 3.4 Conclusion

In this study, targeting the challenges faced by the edge big data inference practices, we have investigated both KDL and RP-KDL frameworks and demonstrated promising findings. The KDL framework can effectively reduce the complexity of the deep learning model while maintaining comparable performance, through transferring the knowledge from a teacher model to a student model. Further, under adversarial perturbations, RP-KDL has been proposed to enhance the robustness of the student model, thereby greatly boosting its performance facing adversarial inputs. In such a way, we have demonstrated the framework with knowledge distillation and adversarial learning can, not only advance ultra-efficient edge inference, but also preserve the robustness facing the perturbed input. This study is expected to greatly advance the real-time, long-term wearable big data and mining applications.

# 4. SUMMARY

We have studied the energy-efficient data compression methods for smart health wearable big data applications, powered by deep learning models, more specifically convolutional autoencoders. The data-driven method could encode the data before decoding it to reconstruct the data from the latent space representation. The proposed approach is evaluated based on the root mean square error of the reconstructed signal. The total energy consumed is also calculated based on the number of arithmetic operations involved. To determine the best design principles for the CAE, the architecture investigation of CAE has been implemented. The experimental results have demonstrated that CAE could reduce the power consumption of the modules in the wearable significantly by intelligently capturing important information of the data during the encoding process, even at a high compression ratio. The proposed method has been then compared with state-of-the-art wavelet compression method to demonstrate its effectiveness.

We have also investigated how to leverage knowledge distillation learning (KDL) to build a substantially small yet robust student model. The proposed approach utilizes the "teacher-student model" learning methodology to compress deep learning models. Meanwhile, recent works have suggested that DL models could be vulnerable to adversarial perturbations. We thus have also investigated the effects of adversarial ECG samples on the classification accuracy of the deep learning models. The model is also evaluated based on precision, recall and F1 score to fully indicate its effectiveness. The Robust Preservation – Knowledge Distillation Leaning (RP-KDL) has been developed and validated on the ECG-based heart disease detection application and demonstrated that the model is very light-weight yet can still yield robust detection results.

Our research is expected to greatly advance big data applications, including but not limited to wearable big data practices. By introducing a deep-learning empowered data compression method and a robust model compression framework, this research could contribute towards edge computing with energy efficiency and robustness.

# 5. FUTURE STUDIES

Future studies may include further investigating 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. Lastly, we would like to investigate other lossy and lossless data compression methods to provide a more comprehensive assessment of CAE.

# 6. REFERENCES

[1] C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Transactions on biomedical Engineering,* vol. 42, no. 1, pp. 21-28, 1995.

[2] 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 Transactions on Biomedical Engineering,* vol. 64, no. 11, pp. 2562-2574, 2016.

[3] F. J. Dian, R. Vahidnia, and A. Rahmati, "Wearables and the Internet of Things (IoT), applications, opportunities, and challenges: A Survey," *IEEE Access,* vol. 8, pp. 69200-69211, 2020.

[4] U. Pratap and R. K. Sunkaria, "ECG compression using Compressed Sensing with Lempel-Ziv-Welch Technique," in *2015 1st International Conference on Next Generation Computing Technologies (NGCT),* 2015: IEEE, pp. 863-867.

[5] S.-C. Tai, C. Sun, and W.-C. Yan, "A 2-D ECG compression method based on wavelet transform and modified SPIHT," *IEEE Transactions on Biomedical Engineering,* vol. 52, no. 6, pp. 999-1008, 2005.

[6] Y. Chompusri, K. Dejhan, and S. Yimman, "Mother wavelet selecting method for selective mapping technique ECG compression," in *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology,* 2012: IEEE, pp. 1-4.

[7] A. Bendifallah, R. Benzid, and M. Boulemden, "Improved ECG compression ethod using discrete cosine transform," *Electronics letters,* vol. 47, no. 2, pp. 87-89, 2011.

[8] O. Yildirim, R. San Tan, and U. R. Acharya, "An efficient compression of ECG signals using deep convolutional autoencoders," *Cognitive Systems Research,* vol. 52, pp. 198-211, 2018.

[9] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep convolutional autoencoder-based lossy image compression," in *2018 Picture Coding Symposium (PCS),* 2018: IEEE, pp. 253-257.

[10] R. Balani, "Energy consumption analysis for bluetooth, wifi and cellular networks," *Online Httpnesl Ee Ucla Edufwdocumentsreports2007PowerAnalysis Pdf,* 2007.

[11] S. Nikolaidis, "Instruction-level energy characterization of an ARM processor," in *Proceedings of the 2nd MARLOW workshop,* 2003.

[12] P.-C. Huang, C.-C. Lin, Y.-H. Wang, and H.-J. Hsieh, "Development of health care system based on wearable devices," in *2019 Prognostics and System Health Management Conference (PHM-Paris)*, 2019: IEEE, pp. 249-252.

[13] O. Amft, "How wearable computing is shaping digital health," *IEEE Pervasive Computing,* vol. 17, no. 1, pp. 92-98, 2018.

[14] M. Ma, M. Skubic, K. Ai, and J. Hubbard, "Angel-echo: a personalized health care application," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2017: IEEE, pp. 258-259.

[15] C. Kulkarni, H. Karhade, S. Gupta, P. Bhende, and S. Bhandare, "Health companion device using IoT and wearable computing," in *2016 International Conference on Internet of Things and Applications (IOTA)*, 2016: IEEE, pp. 152-156.

[16] G. You, L. Zhu, and J. He, "Research on Smart Health Services Based on Wearable Devices," in *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2019, vol. 1: IEEE, pp. 2644-2647.

[17] J. Zou and Q. Zhang, "eyeSay: Eye Electrooculography Decoding with Deep Learning," in *2021 IEEE International Conference on Consumer Electronics (ICCE)*, 2021: IEEE, pp. 1-3.

[18] Q. Zhang, "Deep Learning of Biomechanical Dynamics in Mobile Daily Activity and Fall Risk Monitoring," in *2019 IEEE Healthcare Innovations and Point of Care Technologies,(HI-POCT)*, 2019: IEEE, pp. 21-24.

[19] Q. Z. Jake Stauffer, A. C., C. J., "Deep Reconstruction Learning Towards Wearable Biomechanical Big Data.," *IEEE EMBS Conference on Biomedical Engineering and Sciences,* 2021.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, pp. 1097-1105, 2012.

[22] J.-H. Luo, J. Wu, and W. Lin, "Thinet: A filter level pruning method for deep neural network compression," in *Proceedings of the IEEE international conference on computer vision,* 2017, pp. 5058-5066.

[23] Y. Kutlu and D. Kuntalp, "A multi-stage automatic arrhythmia recognition and classification system," *Computers in Biology and Medicine,* vol. 41, no. 1, pp. 37-45, 2011.

[24] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Generation Computer Systems*, vol. 86, pp. 446-455, 2018.

[25] B. Murugesan *et al.*, "Ecgnet: Deep network for arrhythmia classification," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018: IEEE, pp. 1-6.

[26] W. Li and J. Li, "Local deep field for electrocardiogram beat classification," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1656-1664, 2017.

[27] 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, 2010.

[28] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, p. 285, 2016.

[29] M. Sharma, R. San Tan, and U. R. Acharya, "A novel automated diagnostic system for classification of myocardial infarction ECG signals using an optimal biorthogonal filter bank," *Computers in biology and medicine*, vol. 102, pp. 341-356, 2018.

[30] R. Varatharajan, G. Manogaran, and M. Priyan, "A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 10195-10215, 2018.

[31] U. R. Acharya *et al.*, "A deep convolutional neural network model to classify heartbeats," *Computers in biology and medicine*, vol. 89, pp. 389-396, 2017.

[32] A. Mostayed, J. Luo, X. Shu, and W. Wee, "Classification of 12-lead ECG signals with Bi-directional LSTM network," *arXiv preprint arXiv:1811.02090*, 2018.

[33] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[34] K. Xu *et al.*, "Topology attack and defense for graph neural networks: An optimization perspective," *arXiv preprint arXiv:1906.04214*, 2019.

[35] S. Ye *et al.*, "Adversarial Robustness vs. Model Compression, or Both?," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 111-120.

[36] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016.

[37] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[38] M. Goldblum, L. Fowl, S. Feizi, and T. Goldstein, "Adversarially robust distillation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 04, pp. 3996-4003.

[39] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215-e220, 2000.

[40] A. f. t. A. o. M. Instrumentation, "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," *ANSI/AAMI EC38*, vol. 1998, 1998.

[41] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45-50, 2001.