M2SKD: Multi-to-Single Knowledge Distillation of Real-Time Epileptic Seizure Detection for Low-Power Wearable Systems

SALEH BAGHERSALIMI, École Polytechnique Fédérale de Lausanne, Switzerland, Switzerland
ALIREZA AMIRSHAHI, École Polytechnique Fédérale de Lausanne, Switzerland
FARNAZ FOROOGHIFAR, École Polytechnique Fédérale de Lausanne, Switzerland
TOMAS TEIJEIRO, BCAM - Basque Center for Applied Mathematics, Spain
AMIR AMINIFAR, Lund University, Sweden
DAVID ATIENZA, École Polytechnique Fédérale de Lausanne, Switzerland

Integrating low-power wearable systems into routine health monitoring is an ongoing challenge. Recent advances in the computation capabilities of wearables make it possible to target complex scenarios by exploiting multiple biosignals and using high-performance algorithms, such as Deep Neural Networks (DNNs). However, there is a trade-off between the algorithms' performance and the low-power requirements of platforms with limited resources. Besides, physically larger and multi-biosignal-based wearables bring significant discomfort to the patients. Consequently, reducing power consumption and discomfort is necessary for patients to use wearable devices continuously during everyday life. To overcome these challenges, in the context of epileptic seizure detection, we propose the M2SKD (Multi-to-Single Knowledge Distillation) approach targeting single-biosignal processing in wearable systems. The starting point is to train a highly-accurate multi-biosignal DNN, then apply M2SKD to develop a single-biosignal DNN solution for wearable systems that achieves an accuracy comparable to the original multi-biosignal DNN. To assess the practicality of our approach to real-life scenarios, we perform a comprehensive simulation experiment analysis on several edge computing platforms.

CCS Concepts: • **Mathematics of computing → Time series analysis**; **Dimensionality reduction**; • **Computing methodologies → Supervised learning by classification**.

Additional Key Words and Phrases: Edge computing, Deep learning, Electrocardiogram, Epilepsy, Knowledge distillation, Seizure detection, Multi-modal biosignal processing

ACM Reference Format:
Saleh Baghersalimi, Alireza Amirshahi, Farnaz Forooghifar, Tomas Teijeiro, Amir Aminifar, and David Atienza. 2024. M2SKD: Multi-to-Single Knowledge Distillation of Real-Time Epileptic Seizure Detection for Low-Power Wearable Systems. 1, 1 (June 2024), 30 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Epilepsy is amongst the five most common chronic diseases, and according to WHO, it is the most common chronic brain disease affecting more than 50 million people of all ages [1]. Besides suffering an associated stigma and discrimination, epilepsy represents the second neurological cause of years of potential life loss. The most life-threatening effect of

Authors' addresses: Saleh Baghersalimi, saleh.baghersalimi@epfl.ch, École Polytechnique Fédérale de Lausanne, Switzerland, Switzerland; Alireza Amirshahi, École Polytechnique Fédérale de Lausanne, Switzerland, alireza.amirshahi@epfl.ch; Farnaz Forooghifar, École Polytechnique Fédérale de Lausanne, Switzerland, farnaz.forooghifar@gmail.com; Tomas Teijeiro, BCAM - Basque Center for Applied Mathematics, Bilbao, Spain, tteijeiro@bcamath.org; Amir Aminifar, Lund University, Lund, Sweden, amir.aminifar@eit.lth.se; David Atienza, École Polytechnique Fédérale de Lausanne, Switzerland, david.atienza@epfl.ch.

This work has been partially supported by the PEDESITE Swiss NSF Sinergia project (GA No. CRSII5-193813 / 1), the RESoRT project (GA No. REG-19-019) from the Botnar Foundation, and the WASP Program funded by the Knut and Alice Wallenberg Foundation, the grant RYC2021-032853-I funded by MCIN/AEI/ 10.13039/501100011033 and by European Union NextGenerationEU/PRTR.

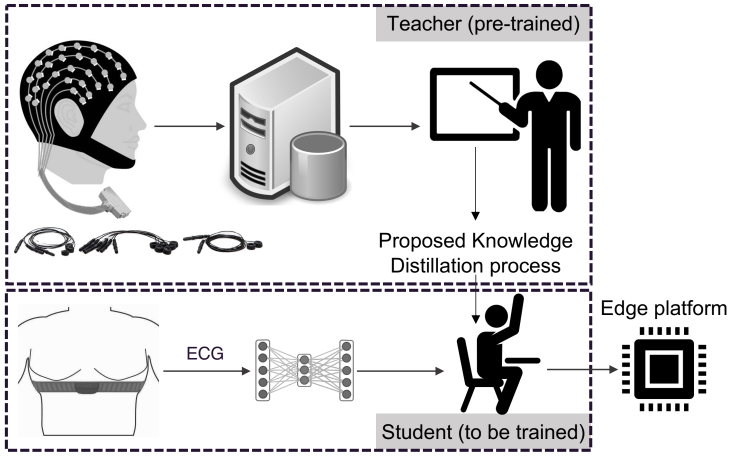

Fig. 1. Structure of our proposed M2SKD for wearable systems. Our student model requires only the ECG signal and can be applied on low-power wearable platforms to perform epileptic seizure detection. In the student model, the patient employs just one sensor, which is comfortable, causes no social stigma, and achieves the detection accuracy of the more complex teacher model.

seizure attacks is Sudden Unexpected Death from Epilepsy (SUDEP) [2]. To reduce the aforementioned effects of such attacks, we must continuously monitor these patients for long periods. With the widespread adoption of the Internet of Things (IoT) [3] in electronic medical care, low-power and easy-to-use embedded platforms are growing in popularity. Thus, we can perform long-term everyday patient monitoring [4] to inform their family members or caregivers in case of emergency situations.

The current trend in pathology detection systems is based on deep neural networks (DNNs) [5], which are capable of classifying large volumes of biomedical data. The ability of these networks to extract high-level and complex patterns from biomedical signals makes them an attractive tool for epilepsy monitoring [6]. Wearable devices typically own limited resources and cannot satisfy the computing requirement of complex DNNs. By addressing the challenges regarding comfortability, complexity, and energy consumption, these systems can perfectly fit the criteria of epileptic seizure detection systems [7]. This can be translated in using simpler and lighter networks, preferably with a smaller number of acquired biosignals, to make them more comfortable and improve their energy efficiency regarding data acquisition and processing.

According to the literature, monitoring the electroencephalogram (EEG) signal of the patients provides the standard detection accuracy in case of epileptic seizures [8]. However, by simultaneously monitoring other biosignals, the decision can become more accurate and robust, as each signal can reflect different effects of the seizures [9]. In this paper, we first introduce our multi-modal detection system, which uses separate residual neural networks [10] for different biosignals, namely, EEG and electrocardiogram (ECG) signals.

The overhead that comes with this multi-modal biosignal monitoring is an increase in the size of the detection system, as we should use a separate network for each signal. This bigger size results in higher energy consumption [11], which is problematic for wearable systems and long-term monitoring purposes. Besides, the synchronisation of signals from multiple devices is challenging in real deployments. Finally, connecting several wearable devices to the patient increases the stigma, especially in the case of EEG, where devices are usually bulky and draw a lot of attention [12].

To address the challenges of multi-modal monitoring, we turn to the concept of knowledge distillation [13], a popular technique in deep learning for model compression. Knowledge distillation involves training a smaller model, often referred to as the 'student', to emulate the behavior and predictions of a larger, more complex model, called the 'teacher'. Imagine having an expert teacher (a large model) that knows a lot but is resource-heavy and a student (a smaller model) that wants to learn as much as it can from this teacher but with fewer resources. In essence, the student tries to 'distill' or capture the knowledge of the teacher. Typically, this is done by training the student model on a dataset, but instead of using the original labels, the student aims to match the teacher's outputs [14]. The benefit of this approach is that the smaller student model can achieve performances comparable to the larger teacher model but at a fraction of the size and computational cost.

In our proposed approach, named M2SKD (Multi-to-Single Knowledge Distillation), we employ knowledge distillation for epileptic seizure detection. As illustrated in Fig. 1, we first create a teacher model—an ensemble deep neural network (DNN) trained on synchronized EEG and ECG signals. Subsequently, we train a student model using only the ECG biosignal, with the aim of achieving detection accuracy similar to that of the teacher but with reduced computational overhead, making it suitable for low-power wearable devices.

In the teacher model, which is a complex model with high energy consumption, the patient uses multiple sensors, which are not comfortable and cause stigma. Conversely, the student model contains a much simpler model resulting in less energy consumption and enabling more prolonged monitoring with wearable devices, as it requires only the ECG signal. This signal can be acquired using a chest strap, which is comfortable to use and removes the social stigma entirely. At the same time, the student model achieves the detection accuracy of the more complex teacher model. To the best of our knowledge, this work is different from the usual knowledge distillation frameworks that have been proposed in the sense that other works propose to change the system structure, but the inputs are the same. On the contrary, we are completely omitting the EEG in the student model and only require ECG inputs in our final system design.

Thus, the main contributions of our work are as follows:

- We develop a knowledge distillation-based approach named M2SKD to develop high-precision low-power wearable systems for epileptic seizure detection in real-time using only a single biosignal input. We show a process to create a neural network trained from the knowledge of a multi-modal DNN system relying on both EEG and ECG signals, while in the test phase, only the ECG signal is used as input. As a result, the synchronized EEG signal is only used in the initial (offline) training phase. This approach reduces the system's complexity and EEG acquisition power requirements to better fit the design constraints of low-power wearable devices. Moreover, we reduce the stigma and inconvenience of synchronizing multiple biosignals in real-life system deployments.
- As initial teacher network for our approach, we design a new multi-modal and multi-channel seizure detection DNN that increases the quality of seizure detection by acquiring information from both EEG and ECG. In this first step, to limit the design complexity and highlight how to target different biosignal inputs, we conceive independent 1-dimensional DNNs for each signal. Then, we use a weighted combination of the results to reach a final decision. Compared with a network using only the ECG signal for training, we show that the proposed multi-modal method increases the sensitivity by 6.12% and specificity by 16.07% when applied to the EPILEPSIAE dataset [15]. Our approach increases seizure detection accuracy and reduces false alarms at the same time, compared to using only the ECG signal.

- By using M2SKD, we reduce the complexity of the initial teacher DNN system to result in an implementable student network for a wearable device with considerably lower power capability. Our results show the distilled model can be implemented on current wearable platforms. Using an edge AI platform, we show that the required energy is reduced by 37.65%. This energy reduction is obtained while sensitivity and specificity are only reduced by 1.5% and 1.3% in comparison with the initial (and very uncomfortable for the patient) multi-modal teacher DNN.

The rest of this article is organized as follows. In Section 2, we review previous works on low-power embedded wearable platforms, multi-modal seizure detection, and knowledge distillation. Section 4 presents a general overview of our seizure detection systems and the different parts: preprocessing, initial multi-modal network architecture, and our M2SKD approach targeting low-power IoT wearable systems. Then, Section 5 presents our experimental setup, and in Section 6, we analyze the computational and energy consumption characteristics of the proposed M2SKD implementation and compare it with other similar wearable architectures. Finally, in Section 7, we summarize the main conclusions of this work.

2 RELATED WORK

2.1 EEG-based Seizure Detection

The gold standard for non-invasive seizure detection is EEG monitoring [16], which has been used for decades in highly specialized and costly hospital environments. EEG-based seizure detection has received noticeable attention in the literature as brain activity is significantly affected during seizure attacks. In [17], the authors proposed the use of spectral graphs to extract spatial-temporal patterns for seizure detection. In [18], they used the wavelet transform, which has been applied to the time-frequency domain for the detection of epileptic activity. In [19], the authors proposed an effective feature extraction algorithm named discrete short-time Fourier transform (DSTFT), which is an adaptive generalization of the classical short-time Fourier transform (STFT). The authors in [20] use discrete wavelet transform and statistical features to apply preprocessed EEG signals to a neural network classifier to detect epileptic seizures.

All the works mentioned above propose power-hungry algorithms that use a complete set of EEG channels. Therefore, if these methods are applied to real-life IoT wearable systems, patients would suffer from the social stigma of wearing a cap with electrodes on the head. Yet, our work presents a knowledge distillation to remove the most invasive signals used in continuous monitoring, leading to a more straightforward, more energy-efficient, and less stigmatizing setup applicable to IoT wearable systems.

2.2 Energy Optimization in Wearable Systems

To optimize energy consumption, in [21] a hierarchical architecture with a wide set of near-sensor processing kernels is presented. They have shown that, due to their power management techniques, their architecture is suitable for IoT wearable systems. In [22], the authors have used optimization and parallelization techniques besides integration of domain-specific accelerators with the goal of improving the mapping and reducing the energy consumption of biomedical applications. The technique that we are targeting here is knowledge distillation, which is originally a model compression technique. This enables us to train smaller and less complex DNNs with comparable performance. To interpret DNNs for epileptic seizure detection on EEG signals, in [6], they associate certain properties of the model behavior with the expert medical knowledge. They have come up with online seizure event characterization able to handle inter-patient variability.

2.3 Addressing Social Stigma in Wearable Systems

To overcome the social stigma problem in IoT wearable systems using EEG caps, in [23], a wearable system based on the temporal EEG electrodes for the detection of epileptic seizures is presented. By reducing the acquired signals to four, the system could be implemented on glasses, which can be easily worn by patients. Although these glasses solve the social stigma issue, they use only EEG signals for the seizure detection task, which are hard to obtain compared to other biomedical signals. In contrast, our work uses the knowledge of EEG and ECG signals, while the final model uses only the ECG signal to detect seizures more effortlessly than with EEG signals.

2.4 Multi-modal Seizure Detection

Besides EEG, other biosignals can also get affected by epileptic seizures and provide additional information about the occurrence of such attacks. For example, seizures are often associated with cardiovascular alterations, and measures related to the heart rate are known to be useful clinical signs of an epileptic discharge [24–28]. Thus, combining other biosignals with EEG can result in better detection of different types of seizures, especially those that are not easily detected with just the EEG. Multi-modal seizure detection has been done previously in several works. In [29], a combination of electrodermal activity and accelerometer signals is used. In [30], the authors have used EEG, Electromyography (EMG) and ECG signals and have shown that the average sensitivity is improved in comparison with using each individual sensor separately. In [31], a combination of linear discriminant models extracted from EEG and ECG features is used to detect seizures in newborn infants. In [32], an SVM model is used with both EEG and ECG signals and could achieve high accuracy being tested on three patients. In [33], seizure detection is done using an SVM model on multi-channel EEG and single-channel ECG separately and then fusing them into one final decision. The effect of using this multi-modal method is then shown on the number of false alarms and detection delay. Recently, in [34], convolutional neural networks (CNN) are used in combination of EEG, ECG and respiration. They have observed that the multi-modal system outperforms systems with individual signals. The reduced deep convolutional stack autoencoder is used in another work using 18 channels of EEG signal, which has resulted in very high performance; but yet, the system cannot be implemented on IoT wearable devices and is intrusive [35].

2.5 Knowledge Distillation in Neural Networks

Knowledge distillation in neural networks was first introduced in 2015 [14], where a single model is trained from an ensemble of models (known as specialists or experts). The key aspect of knowledge distillation is learning not from the discrete labels of the dataset, but from the continuous output of the "expert" models. This concept has not only been tested on speech and images [36, 37] but has also shown satisfying results in certain healthcare applications. In that work, it is demonstrated that we can benefit from the knowledge distillation method to efficiently regularize the smaller network and achieve better generalization than directly training the network using the labels of the training set [38, 39]. In [40], authors have used gradient boosting trees as the experts to train their deep learning model on a real-world clinical time-series dataset, showing an improvement in the performance with respect to their initial deep learning model.

In [41], multi-modal segmentation of computerized tomography (CT) and magnetic resonance imaging (MRI) is compacted by distilling knowledge from cross-modal information of these images. Similarly, in [42], the student model learns from both labeled target data (e.g., CT), and unlabeled target data and labeled source data (e.g., MR) by two teacher models. They have shown that their approach can utilize unlabeled data and cross-modality data with superior performance, outperforming semi-supervised learning and domain adaptation methods with a large margin. Based on the reviewed state-of-the-art works, we hypothesise that distilling the knowledge from a multi-modal epileptic seizure detection system to a single-input system, can improve the performance of the single-input network. Thus, in the following section, we will leverage this concept and introduce first our multi-modal system and, after that, our distilled single-input system.

3 BACKGROUND AND MOTIVATION

3.1 Epileptic Seizures: Understanding Their Nature and Classifications

Epileptic seizures are spontaneous surges of electrical activity in the brain. They result from the brain's neurons firing in an abnormal, synchronized manner. The manifestations of these seizures vary significantly depending on their type and origin within the brain. From barely noticeable episodes, such as staring spells, to more severe forms like vigorous shaking and loss of consciousness, the range of symptoms is vast. This complexity not only affects an individual's motor functions and sensory perceptions but also has profound implications on behavior, emotions, and overall awareness. Understanding the distinct types of seizures is paramount for clinicians to diagnose and devise appropriate treatment strategies.

Focal onset seizures begin in a specific brain region and may remain confined or spread. Symptoms vary widely, from physical jerks to confusion and unusual feelings. Generalized onset seizures, involving both brain hemispheres, present more severe symptoms such as tonic-clonic convulsions or absence seizures. Seizures of unknown onset pose a diagnostic challenge; they lack identifiable origins, but advancements in diagnostics may allow reclassification as either focal or generalized. This summary clarifies the distinct origins and manifestations of seizure types, highlighting the potential for diagnostic evolution [43].

Epilepsy underscores a pressing healthcare challenge. Rapid and precise detection of seizures isn't just a medical imperative; it's a lifeline. Timely diagnosis paves the way for suitable interventions, potentially mitigating the severity of episodes and enhancing the overall quality of life for those affected.

3.2 Diagnostic Instruments: EEG and ECG Role in Seizure Detection

Historically, the diagnosis and assessment of epileptic seizures have been centered on neurological explorations. However, emerging research has spotlighted the complex relationship between brain neural activity and heart function during seizure episodes. This revelation underscores the value of a holistic diagnostic approach. Taking into account this, we examine how electroencephalography (EEG), a method of recording brain activity, has been used for a long time to directly check brain function. We also examine the growing importance of electrocardiography (ECG), a way of recording heart activity, to understand how seizures can affect the heart.

(a) **Electroencephalography (EEG):** In the domain of neurological diagnostics, Electroencephalography (EEG) is fundamental, recording the complex electrical activity of the brain. This tool provides detailed waveforms that allow for a deep understanding of a person's neurological state. Particularly in the treatment of epilepsy, EEG is crucial for diagnosis, monitoring seizures, and informing treatment strategies. However, interpreting the EEG data presents challenges. Specialists face difficulties such as interference from external noise, the ever-changing patterns of brain signals, and the complexity of the data, which is rich in detail and multi-layered.

(b) **Electrocardiography (ECG):** mainly used for cardiac diagnostics, carefully records the heart's electrical rhythm regularly. In the context of epilepsy, seizures can cause noticeable changes in ECG readings, highlighting the

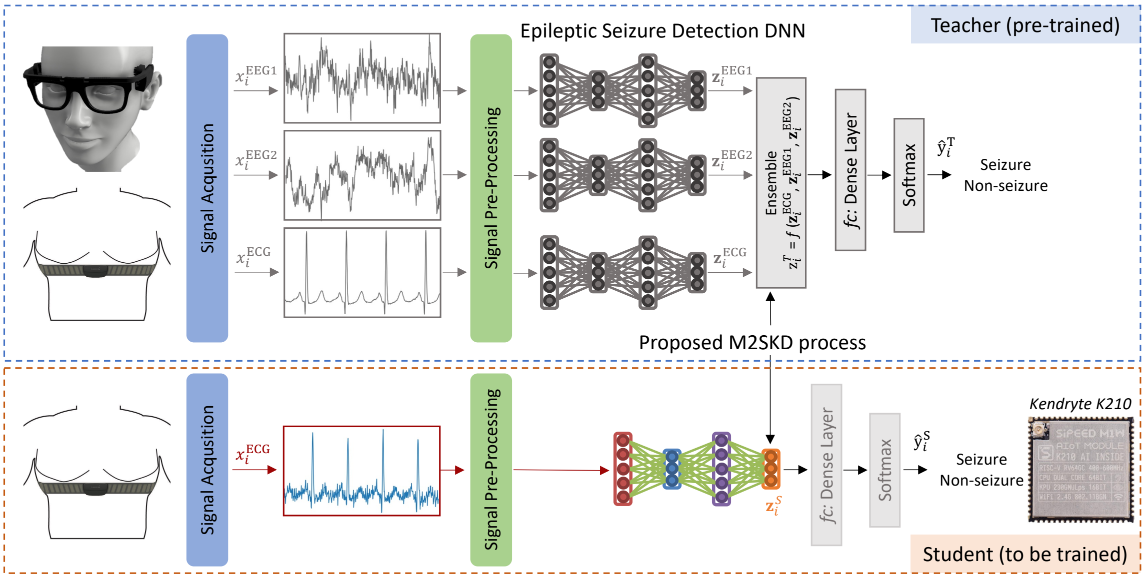

Fig. 2. The overall flow of the proposed M2SKD for epileptic seizure detection. The teacher DNN is an ensemble composed of three Res1DCNN to extract valuable information from the training set. Making epileptic seizure detection using the teacher DNN is heavy to compute as it requires both ECG and EEG signals and is not suitable for wearable systems. The student DNN is a single Res1DCNN that requires only an ECG signal. Thus, it is less computationally intensive. The Kendryte K210, Raspberry Pi Zero system, or the latest multi-core edge AI architectures (e.g., smartwatches or PULP SoC [22]) perform the epileptic seizure detection using the student DNN on ECG signal acquired by a chest strap.

interconnected behavior of brain neural activity and heart rhythms. It is vital to notice these changes in heart activity because doing so improves seizure detection and reveals the effects of seizures on cardiovascular health, providing a more complete view of the general well-being of a patient.

Epileptic seizures' complex nature, with its affects reaching both neurological and cardiac domains, necessitates a synthesized diagnostic methodology. By combining information from EEG and ECG, healthcare providers are in a better position to improve treatment plans. This combined approach is expected to lead to a better understanding, more accurate diagnosis, and a more detailed plan for treating epilepsy.

4 PROPOSED M2SKD APPROACH

This section describes the elements of our proposed feature-based M2SKD approach for epileptic seizure detection. The overall flow is presented in Fig. 2. Our proposed approach includes a pipeline divided into three phases: signal acquisition and pre-processing, multi-modal CNNs framework, and distilling the knowledge from the more extensive teacher network to the lighter student network. The teacher network requires both ECG and EEG signals and has a more complex network architecture than the student network, which results in more energy consumption. Thus, it is not convenient to be deployed on wearable devices or low-power platforms with limited resources for long-term monitoring. On the other hand, the student network receives only the ECG signal, has fewer parameters, requires lower computational costs, and performs real-time epileptic seizure detection without losing validity. Thus, it is suitable to be deployed in edge platforms and wearable devices.

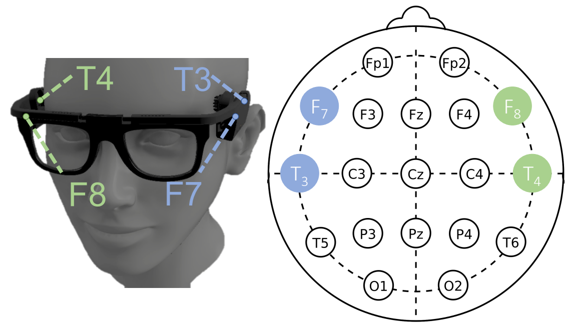

Fig. 3. Electrodes locations of F_7T_3 and F_8T_4 for EEG monitoring using the e-Glass wearable system for epileptic seizure detection [23].

4.1 Signal Acquisition and Pre-Processing

In our introduced M2SKD framework, both a teacher and a student network exist. The formation of the teacher network relies on EEG and ECG data sampled with a frequency of 256 Hz, while the student network utilizes solely the ECG signal. The EEG data for the teacher network is acquired through two channels, namely F_7T_3 and F_8T_4, using four specific electrodes, illustrated in Figure 3. These electrodes were purposefully chosen to interface with e-Glass, a wearable system tailored for real-time applications, harnessing the potential of these four EEG electrodes [23]. e-Glass aims to offer a discreet solution, allowing users to avoid the social challenges often associated with conventional EEG head caps. As detailed in section 6.2, we validate the capability of the teacher or student network in real-time epileptic seizure detection when implemented on edge AI platforms.

We opted for the F_7T_3 and F_8T_4 EEG channels in our seizure detection system due to several reasons, taking into account the diverse ways seizures can present in patients:

(1) Seizure Localization: The F_7T_3 and F_8T_4 channels are situated in the frontotemporal regions of the brain, which are frequently implicated in seizure phenomena. These channels are adept at recognizing a variety of seizure forms.
(2) Signal Quality: Signals from these channels are notably clear, with minimal noise or disruptions from muscular actions or ocular activities, thereby boost detection precision.
(3) Computational Efficacy: Limiting to two EEG channels, the computational demands are reduced, enabling real-time seizure detection, even on resource-constrained wearable devices.
(4) Societal Acceptance: Using the F_7T_3 and F_8T_4 channels in devices like e-Glass provides a less noticeable alternative compared to traditional EEG gear, reducing potential discomfort for patients.
(5) Generalizability: Despite the variations in seizure types, the combination of these channels with our system design ensures reliable detection across a wide array of patients.

Training a deep neural network (DNN) like the teacher DNN necessitates a significant amount of data. However, actual data pertaining to epileptic seizures for a specific patient is typically scarce. To address this, recent studies have employed data augmentation techniques to enhance both the volume and variety of epilepsy-related data [44, 45]. Our research leverages data augmentation by segmenting and overlapping synchronized ECG and EEG signals. Given the critical role of interpreting QRS complexes in ECG signal processing, we adopt 3-second intervals to guarantee the

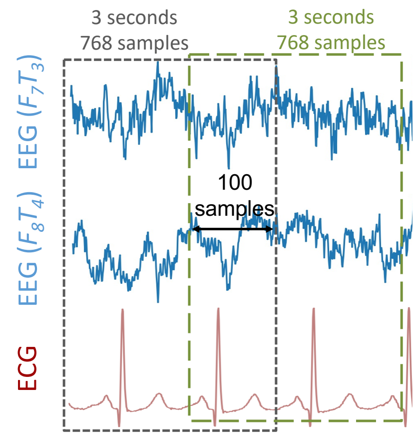

Fig. 4. Segmentation of ECG, $\text{EEG}_{F_7T_3}$ and $\text{EEG}_{F_8T_4}$ signals sampled with a frequency of 256 Hz using slots of 3-seconds with 100 samples overlap. In our framework, ECG, $\text{EEG}_{F_7T_3}$ and $\text{EEG}_{F_8T_4}$ signals are synchronized to train the teacher DNN. In other words, both ECG and EEG signals are acquired and measured in parallel. Therefore, we have the corresponding ECG signal when there is a seizure in the EEG signal. Moreover, note that in latest edge AI architectures deployed in the medical IoT ecosystem (e.g., multi-core PULP system [22]), it is possible to process the signal in less than one second, and during the remaining time the system can be in sleep mode.

inclusion of at least two QRS complexes. These intervals are extracted by sliding a fixed-length window—with an overlap of 100 samples—across the entire signal. The segmentation method for ECG and EEG signals is illustrated in Figure 4.

To train DNN models both effectively and efficiently, pre-processing techniques are indispensable and have been employed across various applications [46]. In this context, we introduce a straightforward pre-processing approach for each segment following signal segmentation. The pre-processing procedures for ECG and EEG segments are delineated in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1 Pre-processing of ECG Segment

Require: ECG segment x
Ensure: Standardized ECG segment $x_{\text{standardized}}$

1: **Step 1: Apply 10th-order Low-pass Butterworth Filter**
2: $x_{\text{filtered}} \leftarrow \text{ButterworthFilter}(x, \text{order} = 10, \text{cutoff} = 50)$

3: **Step 2: Perform Linear Detrending**
4: Fit a linear model $y = ax + b$ to x_{filtered}.
5: Subtract the fit from the initial data: $x_{\text{detrended}} \leftarrow x_{\text{filtered}} - (ax + b)$.

6: **Step 3: Standardize the Detrended ECG Segment**
7: Compute the mean μ and standard deviation σ of $x_{\text{detrended}}$.
8: $x_{\text{standardized}} \leftarrow \frac{x_{\text{detrended}} - \mu}{\sigma}$

In bioelectrical recordings, particularly ECG measurements, 50Hz noise is a dominant concern. Often dubbed "power-line interference" due to its association with standard electrical frequencies in many areas, this interference significantly

Algorithm 2 Pre-processing of EEG Segment

Require: EEG segment x
Ensure: Standardized EEG segment $x_{\text{standardized}}$

1: **Step 1: Apply 10th-order Low-pass Butterworth Filter**
2: $x_{\text{filtered}} \leftarrow \text{ButterworthFilter}(x, \text{order} = 10, \text{cutoff} = 50)$

3: **Step 2: Standardize the Filtered EEG Segment**
4: Compute the mean μ and standard deviation σ of x_{filtered}.
5: $x_{\text{standardized}} \leftarrow \frac{x_{\text{filtered}} - \mu}{\sigma}$

impedes the automatic detection and classification of arrhythmias [47]. Therefore, during ECG preprocessing, emphasis is placed on eliminating baseline wandering, a slow shift in the ECG trace. This wandering can result from respiratory effects, mismatches in electrode-skin impedance, and motion artifacts. Breathing affects the ECG baseline due to the movement of the heart and chest structures. Variations in electrode-skin contact and patient movements can introduce further disturbances [48]. Our decision to standardize ECG data on a per-window basis, rather than across the entire training set, acknowledges the variability in ECG baselines over time. This approach accommodates differing activity levels and mitigates the influence of outliers on data standardization.

4.2 Epileptic Seizure Detection DNN

In this section, we introduce the Deep Neural Network (DNN) adopted for seizure detection, dubbed Res1DCNN. This design stems from our preceding investigations, as elucidated in [49].

Let's consider a series of data samples represented as $\{x_1, ..., x_T\}$, where each x_i belongs to $\mathbb{R}^r$, and r typifies the count of samples within an individual window. For any designated data sample x_i, Res1DCNN processes it, yielding a feature map labeled as $\mathbf{z}_i \in \mathbb{R}^L$, where L characterizes the feature map's span. Subsequently, a simplistic fully-connected layer transforms this feature map to procure $\hat{y}_i$, which predicts whether it's a seizure or a non-seizure episode.

The decision to employ Res1DCNN was spurred by its proficiency in delivering consistent seizure detection outcomes, all the while being computationally efficient. A major benefit of the Res1DCNN is its low energy consumption, which is particularly effective on mobile devices where energy efficiency is critical. For instance, our detailed evaluations in Section 6 and the battery longevity assessments in Table 3 provide empirical validation of its energy-efficient stance. On devices like the Raspberry Pi Zero and Kendryte K210, the architecture showcased extended runtimes on a singular battery charge, underscoring its applicability for real-world scenarios with energy constraints.

The blueprint of our Res1DCNN is illustrated in Figure 5. Drawing inspiration from the Residual Neural Networks [10], it incorporates multiple skip connections. These connections ensure the norm is preserved within the residual blocks, which in turn aids in seamless information flow across the network's depth, thereby guaranteeing a stable training phase. Additionally, with its concise construct of merely 14 weight layers, Res1DCNN is a commendable candidate for battery-driven edge AI platforms with circumscribed computational capabilities.

4.3 Teacher Network: Multi-modal Res1DCNN

In this work, the teacher network uses the Res1DCNN architecture. The teacher takes synchronized ECG and 2-channel EEG segments as inputs and trains three analogous Res1DCNN models for every input. This method is an ensemble

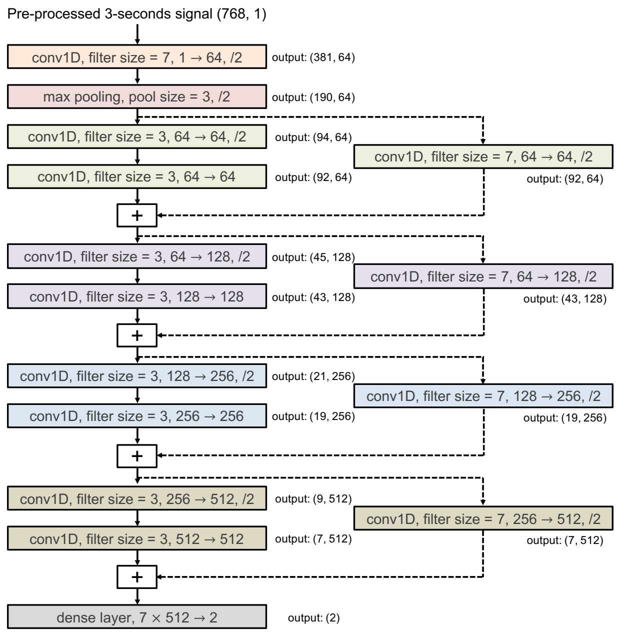

Fig. 5. The architecture of the Res1DCNN [49]. The network contains 14 convolutional layers with skip connections followed by a dense layer. Here, '/2' denotes the downsampling operator using a strided convolution with a factor of 2. '⟶' denotes the transition from the input to output channels.

learning technique where various DNN models are combined and trained to solve the same problem [50]. As shown in Section 6.1, by using our proposed ensemble learning method named Multi-modal Res1DCNN, we have a more accurate DNN model to perform epileptic seizure detection. Since we studied and experimented with a wearable system, we considered only two EEG channels where the signals can be acquired from e-Glass [23], as shown in Fig. 3.

In this network, the inputs are EEG and ECG segments $\{x_i^{\text{ECG}}, x_i^{\text{EEG1}}, x_i^{\text{EEG2}}\}$. We extract features from the inputs by passing each of them through Res1DCNN's feature extractor to obtain $\{\mathbf{z}_i^{\text{ECG}}, \mathbf{z}_i^{\text{EEG1}}, \mathbf{z}_i^{\text{EEG2}}\}$. We merge these feature maps into a single $\mathbf{z}_i^T \in \mathbb{R}^L$ using a linear combination of the features. More formally, $\mathbf{z}_i^T = f(\mathbf{z}_i^{\text{ECG}}, \mathbf{z}_i^{\text{EEG1}}, \mathbf{z}_i^{\text{EEG2}}; \theta)$, where θ is the trainable weight for the linear combination. When $\mathbf{z}_i^T$ is obtained, we train a simple fully-connected layer that predicts the output $\hat{y}_i$ from $\mathbf{z}_i^T$. Finally, a softmax layer outputs the predicted value. The softmax layer is a generalization of the logistic layer that highlights the largest values in a vector while suppressing the values significantly below the maximum.

4.4 Student Network: Distilling the Knowledge

In machine learning (ML) methods, training an ensemble of various models using the same data is a solution to improve performance [51]. However, making predictions using most DNNs and ensemble models requires significant storage and is too computationally expensive. Consequently, as our goal is to implement the epileptic seizure detection algorithm into a wearable medical platform that runs on a battery, the teacher network is not an appropriate model for implementation. Moreover, as described in Section 4.3, the teacher DNN uses different input data, such as EEG and ECG. However, in a real-world scenario, acquiring EEG signal is complex and uncomfortable for the patient. In this work, to address the abovementioned problems, we introduce a student network that gets only ECG signals as its input. As demonstrated in [52], it is achievable to compress the knowledge from an ensemble model into a single student model. This process enables the model to run on embedded devices, considering these devices' stringent energy and memory constraints.

As illustrated in Fig. 2, our approach leverages knowledge distillation—a technique where a compact student DNN is taught to mimic the behavior of a pre-trained, larger teacher DNN[14, 53, 54]. This methodology is often termed "Teacher-Student", with the teacher being the expansive DNN and the student as the more streamlined one. For our system, the student network, characterized by a single Res1DCNN, processes solely the x_i^{ECG} as its input, generating a feature map $\mathbf{z}_i^S \in \mathbb{R}^L$ as output. The primary aim during the distillation phase is to harmonize this feature map with that of the teacher network. To this end, our loss function is expressed as the distance L2 between $\mathbf{z}_i^S$ and $\mathbf{z}_i^T$. By minimizing this loss, we ensure that the student's feature map aligns closely with the teacher's. Due to the effective alignment of these feature maps, we can directly employ the fully-connected layer of the teacher within the student framework, obviating the need for additional training. It's worth noting that the convolutional component of the student is fine-tuned during the distillation to produce a matching feature map with the teacher. The teacher's feature map z_i^T is explicitly used in the loss definition due to its enhanced signal intensity and spatial correlation characteristics.

The L2 distance serves as the loss function in our framework to maintain feature fidelity, ensuring the student model accurately mirrors the teacher's features for better generalization. It also promotes stability and smoothness in model training, offering a gradual error landscape, unlike the more volatile cross-entropy loss. Lastly, it supports alignment with human perception, as it reflects human sensitivity to differences in images and audio.

M2SKD, in this work, enables us to use only ECG segments as the input for our wearable system in real-life operation. Thus the student network only processes the ECG inputs, and the Res1DCNN is instantiated only once, while in the teacher, we have three parallel Res1DCNN models. Consequently, the amount of computation is considerably decreased while the network's performance is maintained. Furthermore, using the student network, we can translate the proposed application into a real-life scenario that can benefit patients, clinicians, etc. This is done by removing the necessity to permanently wear a cap to monitor EEG outside the hospital, which causes social stigma and discomfort for patients.

5 EXPERIMENTAL SETUP

In this section, we present the experimental setup to evaluate our proposed feature-based M2SKD approach in terms of epileptic seizure detection performance and energy consumption.

5.1 Epileptic Seizures Dataset

In our research, we employ the renowned EPILEPSIAE dataset [15], a compilation of single-lead ECG and 19-channel EEG recordings from 30 patients, meticulously annotated by medical professionals for seizure detection and forecasting. This extensive dataset originates from a routine clinical setting, covering an array of seizure types such as complex partial (CP), simple partial (SP), secondarily generalized (SG), and Unclassified (UC). For a particular patient in this dataset, a disparity in the lengths of the ECG and EEG signals made them unsynchronizable. Consequently, we excluded this patient's signals and focused on the data from 29 patients, totaling 4603 hours with 277 seizures. This data boasts a sampling rate of 256 Hz and 16-bit resolution.

For our experiments, we used both the ECG and the EEG signals to create the teacher network, while the student network only used the ECG signal. Our framework acquired EEG signals for the teacher network from two channels, F_7T_3 and F_8T_4, as showcased in Figure 3. This specific channel choice enabled EEG signal capture through the e-Glass

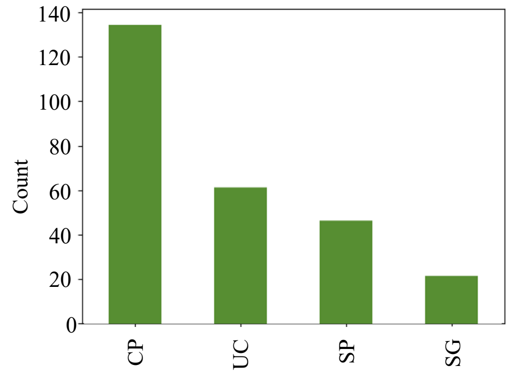

Fig. 6. EPILEPSIAE database

system, an innovative wearable based on four EEG electrodes, striving to alleviate the societal stigma associated with traditional EEG head caps [23]. The design emphasizes both inconspicuousness and real-time functionality.

The EPILEPSIAE dataset covers a diverse set of seizure types, illustrated in Figure 6, including:

(1) **Complex Partial (CP) Seizures**
- **EEG Electrodes**
  - Often originate from a specific region, frequently the temporal lobe.
  - Temporal electrodes, like T3 and T4, are particularly relevant. If originating from the frontal lobe, Fp1, Fp2, F3, F4, F7, and F8 may be informative.
- **ECG Effects**
  - Can affect the autonomic nervous system, leading to tachycardia or bradycardia.
  - ECG changes are more likely if the seizure generalizes or affects areas of the brain involved in autonomic regulation.

(2) **Simple Partial (SP) Seizures**
- **EEG Electrodes**
  - Start in a localized region of the brain.
  - Electrode relevance depends on the specific area affected. If it's temporal, T3 and T4 are essential; if frontal, the aforementioned F electrodes would be crucial.
- **ECG Effects**
  - Generally less common than with CP seizures due to the localized nature and lack of consciousness impairment.
  - Mild tachycardia or bradycardia might occur if the seizure focus influences the autonomic system.

(3) **Secondarily Generalized (SG) Seizures**
- **EEG Electrodes**
  - Begin as a focal seizure (either SP or CP) and then spread to involve broader brain regions.
  - In the focal phase, electrodes relevant to SP or CP seizures are crucial. As the seizure generalizes, more electrodes across the scalp become relevant to capture the widespread activity.

- **ECG Effects**
  - Tachycardia is common, especially during the generalized phase.
  - Rarely, significant bradycardia or ictal asystole can be seen, particularly if the seizure involves areas tied to autonomic function.

(4) **Unclassified (UC) Seizures**
- **EEG Electrodes**
  - Given their ambiguous nature, having a broad range of electrodes from various brain regions can assist in capturing and understanding these seizures.
- **ECG Effects**
  - Predicting ECG changes can be challenging due to the non-specific nature of these seizures.
  - Any changes would depend on the particular characteristics of the seizure and the brain regions affected.

Upon testing with the multi-modal DNN teacher network, we discerned the considerable significance of the ECG component for seizure detection. By refining this for the single-biosignal student network via M2SKD, we achieved a performance equivalent to the combined ECG+EEG system, underlining ECG's potential for standalone seizure detection in wearable technology.

In Section 6.2, we discerned the feasibility of epilepsy detection using the teacher network in real-time on various edge AI systems, such as the Raspberry PI Zero [55] or the PULP multi-core system [22]. However, the teacher network's complexity demands more energy and isn't viable on low-power platforms with limited resources. In contrast, the student network's reduced complexity makes it optimal for prolonged wearable usage, especially since it matches the teacher network's performance using only the ECG signal, simplifying signal acquisition for patients.

In our study, seizure and non-seizure events were segmented into overlapping 3-second windows, which were subsequently fed into the proposed architecture, with a particular emphasis on addressing the challenge of epileptic seizure detection in real-world applications [56]. Notably, the training set predominantly comprised non-seizure segments as compared to seizure segments. Addressing this significant class imbalance is of paramount importance in real-world datasets. To mitigate this, we applied undersampling, selecting from the majority class (non-seizure) an equivalent number of segments to the minority class (seizure). This method effectively maintained the class probability distribution during the training process.

We structured the dataset to facilitate both Global and Personalized Model Training Strategies, ensuring a comprehensive evaluation of our proposed models under different scenarios.

(1) Global Model Training Strategy: Under this approach, we designed the composition of the data set to encompass data from each patient in training, validation, and test sets. This strategy ensures that our model is exposed to a wide range of data variations, reflective of all participating individuals, enhancing its generalization capability. Notably, the test set comprises a mix of the latter half of both seizure and non-seizure recordings from each patient. A fundamental aspect of our approach is the strict division of data according to recording sessions. This separation guarantees that the data allocated to the training, validation and testing phases come from completely different recording sessions. This separation is vital in mitigating potential data leakage and ensuring that our evaluation metrics accurately reflect the true predictive performance of our models.
(2) Personalized Model Training Strategy: In alignment with personalized healthcare objectives, this strategy tailors the training process to individual patient profiles, enhancing model specificity and applicability to individual needs. In addition to the standard accuracy metric, we incorporated sensitivity, specificity, G-mean, F1-score,

and AUC, especially for the personalized model test set. This diverse set of metrics provided a thorough and detailed assessment in scenarios characterized by class imbalances. Further solidifying the robustness of our M2SKD methodology in diverse contexts, we have:

- Multiple Experiments for Each Patient: As stated, for each of the 29 patients from the EPILEPSIAE dataset, we conducted three distinct experiments:
  - Teacher Network (ECG+EEG): To establish a baseline performance.
  - Student Network (ECG without M2SKD): To benchmark the potential of ECG-only seizure detection without knowledge distillation.
  - Student Network (ECG with M2SKD): To validate the efficacy of our proposed M2SKD approach in harnessing information from ECG for seizure detection.
- Practical Real-world Scenario: For each patient, we emulated a real-world scenario by splitting the data such that half of the seizures and non-seizures recordings were used for testing, as delineated in Table 1. This split inherently creates a highly imbalanced dataset, resembling realistic conditions where seizures are infrequent events. The performance of our model under these conditions speaks to its utility in actual deployments.

In sum, our comprehensive approach underscores the model's theoretical soundness and its adaptability to authentic challenges, bolstering its prospective utility in the realm of epileptic seizure detection.

Table 1. Number of Seizure and Non-Seizure Segments per Patient in Test Sets

Patient	Test Set		Patient	Test Set	
	Non-seizures	Seizures		Non-seizures	Seizures
#1	111,540	155	#16	122,299	56
#2	184,930	93	#17	110,487	266
#3	114,024	83	#18	111,558	76
#4	97,606	110	#19	107,095	108
#5	124,352	55	#20	109,282	113
#6	95,092	74	#21	67,464	132
#7	65,713	77	#22	94,198	157
#8	80,132	77	#23	166,254	62
#9	80,052	23	#24	108,924	193
#10	93,756	104	#25	110,511	91
#11	94,743	175	#26	116,426	368
#12	110,427	58	#27	108,146	388
#13	104,364	152	#28	105,716	164
#14	80,232	77	#29	70,361	298
#15	65,306	38			

5.2 Detection Performance Metrics

To evaluate the detection performance of our proposed framework, we considered six different metrics. Sensitivity (Sen) (Eq. (1)) represents the percentage of ictal samples that are labeled correctly. Specificity (Spe) (Eq. (2)) shows the percentage of inter-ictal samples that are labeled correctly. These metrics are defined as follows:

$$Sen = \frac{TP}{TP + FN}, \tag{1}$$

$$Spe = \frac{TN}{FP + TN}, \tag{2}$$

where TP, TN, FP and FN are true positive, true negative, false positive and false negative, respectively.

Geometric mean (Gmean) (Eq. (3)) [57] reflects both sensitivity and specificity, and measures the balance between classification performance in both classes. A low geometric mean indicates poor performance in classifying the seizure cases, even if the non-seizures cases are correctly classified.

$$Gmean = \sqrt{Sensitivity \times Specificity}. \tag{3}$$

F_1 score (Eq. (4)), which is the harmonic mean of precision and recall. It gives a better measure of the incorrectly classified cases than the accuracy.

$$F_1\, score = \frac{2\, TP}{2\, TP + FP + FN}. \tag{4}$$

Finally, we use the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which is a metric that quantifies a model's ability to discriminate between positive and negative classes across all decision thresholds.

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(t)\, dt \tag{5}$$

where $\mathrm{TPR}(t)$ is the True Positive Rate (sensitivity) at threshold t.

5.3 Edge AI Evaluation Platform

Wearable devices have small batteries and low-power processors compared to desktop processors. In this work, we use the Kendryte K210 [58] and Raspberry Pi Zero [55] to analyze and compare the energy consumption and timing requirements for continuous execution of the proposed approach. Note that the proposed approach must be executed repeatedly in real-time. The Raspberry Pi Zero includes an ARM11 CPU running at 1 GHz, has 512MB RAM, and performs the inference process of a given DNN with power supplied via a micro USB connector. The Kendryte K210 is a chip system with specific circuits/components for machine vision and ML. This chip system employs advanced ultra-low processing with the help of a 64-bit dual-core processor equipped with a high-performance hardware accelerator of the CNN. It supports convolution kernels, any form of activation function, and neural network parameter size up to 6 MB for real-time application.

We used Otii Arc [59] as a power analyzer and power supply for the inference process of our proposed approach. Otii Arc is a measurement tool for designing highly energy-efficient algorithms. It is powered via USB from the laptop and records both current and voltage, and it displays them in real-time for analysis and comparison. It provides up to 5 V output voltage and runs high-resolution current measurements with a sample rate up to 4 kHz for the range of 1 μA-5 A. Figure 7 shows the hardware setup of our energy consumption measurement.

We considered the Kendryte K210 chip and Raspberry Pi Zero as they have comparable processing capabilities to modern wearable architectures [21]. We also discuss the potential benefit of using PULP-based ultra-low-power platforms and architectures proposed in [22] for wearable biomedical systems to further reduce power consumption.

Fig. 7. Hardware setup for the energy consumption measurement. Otii Arc is connected to the computer using a USB cable. The main output of the Otii Arc is connected to the voltage supply input of the wearable platform. The current provided by the computer's USB port is used to power both the Otii Arc and edge AI platform. The Otii desktop application enables us to measure and analyze the energy consumption of the edge AI platform. In our evaluation, we considered a development system, which uses AI technologies embedded on a Kendryte K210 chip used in different wearable systems [58] or Raspberry Pi Zero.

5.4 Learning Parameters

We trained our proposed networks from scratch using pre-processed 3-second EEG and/or ECG segments. The weights of the layers were initialized following a normal distribution with a zero mean and a standard deviation of 0.01. All biases were set to zero at initialization.

During training, the network adjusts its parameters to capture the correlation between the input segments and the output, which consists of two nodes. The aim is to minimize the cross-entropy loss. While it's possible for binary classification to use a DNN with a single output and a threshold, we opted for a configuration with two-node outputs. This approach aligns with a multi-class classification setup having two classes, facilitating easy extensibility to multi-class problems in potential future work.

For the optimization phase, we used the Adam optimizer [60] with a mini-batch size of 16 across all patients. A learning rate of 10^{-5} was chosen based on empirical observations, ensuring stable convergence for our setup.

Instead of a fixed iteration count, we implemented early stopping based on the validation set performance. While allowing for up to 10^4 iterations, this approach ensures halting the training once there is an observable plateau in the validation performance, thus preventing overfitting.

All models and training methodologies were crafted using Tensorflow 1.14.0 [61].

5.5 Baselines Description

- Teacher Network (ECG+EEG): The Teacher Network uses both ECG (heart data) and EEG (brain data). For the EEG, we took data from two specific channels: F_7T_3 and F_8T_4. Electrocardiograms (ECG) capture the heart's electrical activity, while Electroencephalograms (EEG) record the brain's electrical patterns. By integrating these two types of data, the Teacher Network aims to establish a robust and comprehensive baseline for seizure detection. This combined approach takes advantage of the rich information present in both cardiac and brain signals, potentially offering greater accuracy and specificity in identifying and distinguishing seizure events.
- Student Network (ECG without M2SKD): The Student Network exclusively utilizes ECG data, without incorporating the knowledge distillation process from M2SKD. This network's primary goal is to evaluate the inherent capabilities of ECG-only data in detecting seizures. By excluding EEG input and M2SKD, the model offers a simplified, yet

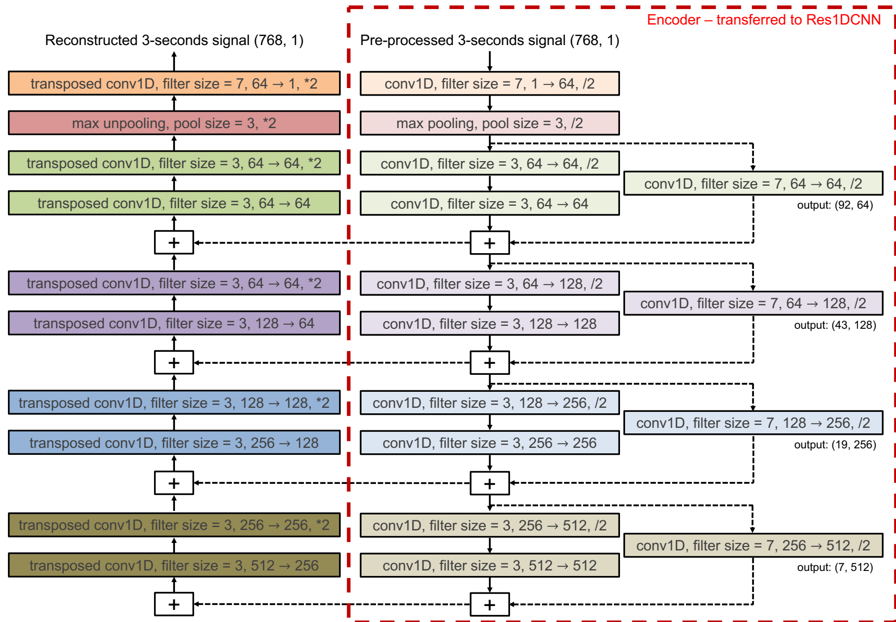

Fig. 8. Schematic of the Proposed Autoencoder Architecture with Integrated Res1DCNN. This figure illustrates the dual-section design, comprising an encoder that mirrors the structure of Res1DCNN for efficient knowledge transfer, and a decoder consisting of nine layers dedicated to reconstructing the data back to its original dimensions. Special emphasis is placed on the 'skip connections' between corresponding layers, underscoring their vital role in preserving details through the encoding process. The efficacy of this architecture is quantified using mean squared error for reconstruction accuracy and cross-entropy loss for Res1DCNN training.

insightful, perspective. It becomes imperative to understand the standalone value of cardiac signals in the realm of seizure detection and to contrast its performance with the combined ECG + EEG approach of the Teacher Network.

- SimCLR: In the context of our study with two types of signals—epileptic and non-epileptic seizures—we applied a contrastive learning framework [62]. This approach was used to train an encoder network on datasets from a single sensor (ECG). The framework's objective is to refine the encoder's ability to closely align augmented versions of the same type of signal—whether epileptic or non-epileptic—in the representation space. In contrast, it aims to separate the representations of the two different signal types. The goal is to develop a system that can identify and distinguish between epileptic and non-epileptic seizure signals.
- Autoencoder: Autoencoders are a type of neural network designed to learn a compact representation of data. They work by condensing the data into a simpler form and then using that condensed form to recreate the original input [63]. A well-known version of this technology is the U-Net, which maintains a close relationship between its encoding and decoding segments [64]. In U-Net, 'skip connections' are special links that join the initial, less complex layers directly to the later, more complex layers. These connections are crucial because they allow the network to preserve detailed information through the encoding process.

The autoencoder architecture we propose is illustrated in Fig. 8. This network is divided into two sections: an encoder and a decoder. The encoder is an exact copy of another network known as Res1DCNN. This design choice allows the knowledge acquired during the autoencoder's training to be transferred to the Res1DCNN. The decoder consists of

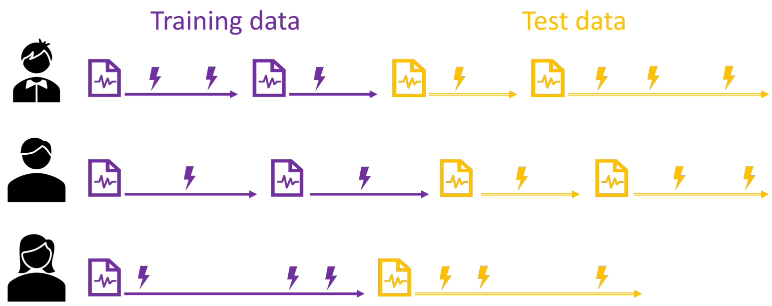

Fig. 9. Global Model Training Strategy: This figure illustrates the comprehensive approach to training the global model, where data from each patient is included across the training, validation, and test sets. Distinct recording sessions are utilized for each set to eliminate temporal bias, ensuring robust model performance across varied scenarios.

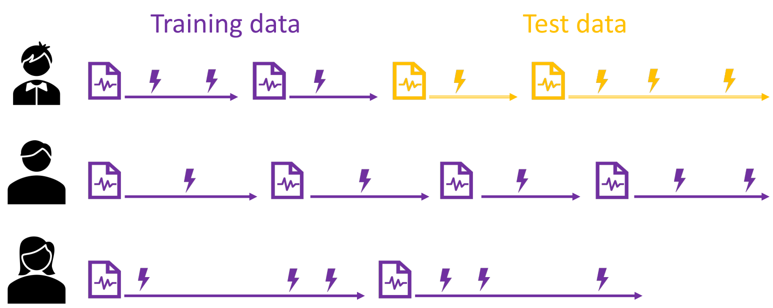

Fig. 10. Personalized Model Training Strategy: This figure shows the personalized approach to model training, focusing on tailoring the training process to individual patients. Using data specific to each patient and ensuring that there is no overlap between training, validation, and test sets, this strategy improves model specificity and adaptability to individual patient profiles.

nine layers that enlarge the condensed data back to its original size.

To evaluate the performance of our autoencoder, we use the mean squared error, which quantifies how close the reconstructed data is to the original input. For training Res1DCNN, we employ cross-entropy loss.

6 EVALUATION

In this section, the assessment of the seizure detection performance and energy consumption of the proposed M2SKD approach on the Kendryte K210 and the Raspberry Pi Zero units is presented. The approach incorporates two distinct training strategies, referred to as the Global Model Training Strategy and the Personalized Model Training Strategy, to enhance the robustness and applicability of the deep neural network (DNN) models. These strategies and their implications are shown in Fig. 9 and 10 for clarity.

6.1 Detection Performance Analysis

6.1.1 ***Global Model Training Strategy:*** In the Global Model Training Strategy, illustrated in Fig. 9, we considered the composition of the data set to include each patient in the training, validation and test sets. A critical aspect of our methodology is the careful separation of data based on recording sessions. This distinction ensures that the sets used for

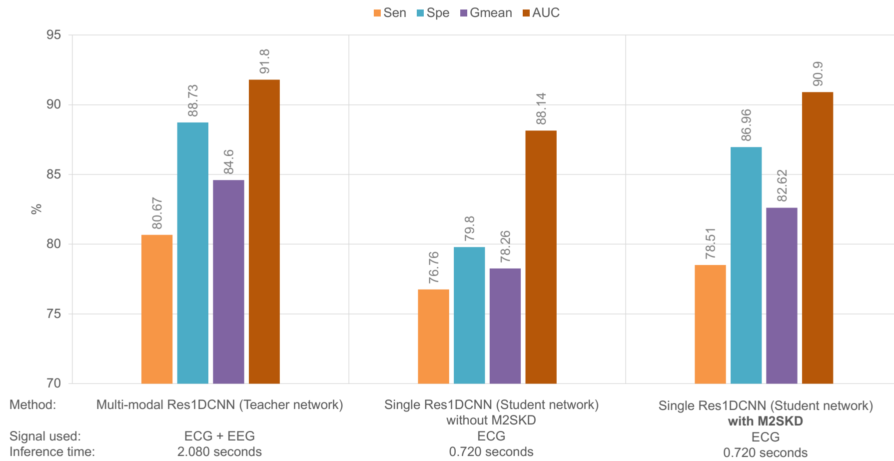

Fig. 11. Global Model Training Strategy: Performance Metrics for Imbalanced Test Set — This figure contrasts the capabilities of the multi-modal Res1DCNN (Teacher network) with single Res1DCNNs (Student networks) with and without M2SKD in the context of epileptic seizure detection. The metrics show how M2SKD effectively transfers critical knowledge from a larger ensemble DNN that uses both ECG and EEG, to a more compact DNN that operates solely on ECG data. Despite the imbalanced nature of the test set, the M2SKD-enhanced student network demonstrates a respectable preservation of detection sensitivity and specificity, affirming its practicality in real-world applications where rapid and efficient processing is paramount.

training, validation, and testing are extracted from entirely different recording sessions. Such an approach is critical to reduce potential biases present in time-series data, where sequential observations may not be statistically independent. By segregating the data according to recording sessions, we effectively eliminate the risk of temporal bias, ensuring that the performance of our model is not artificially affected by the sequential nature of the data.

To address the substantial data requirements necessary to train a robust DNN, we implemented a data enhancement strategy. This was achieved by segmenting the signals into windows of 300 samples, with an overlap of 100 samples between consecutive windows. Importantly, this augmentation process respects the boundaries set by our recording session-based data separation; as such, there is no overlap of segmented windows across the training, validation, and test sets. This strict separation helps us to make sure that our evaluation process is reliable. It prevents overfitting, ensuring that our model's performance truly shows how well it can work with new, unseen data.

Figure 11 illustrates the performance of epileptic seizure detection for the proposed multi-modal Res1DCNN (Teacher network) using both ECG and EEG segments. This is compared against the performance of the single Res1DCNN (Student network) without M2SKD, which utilizes only ECG segments. Our proposed multi-modal Res1DCNN achieves a geometric mean of 84.60%. This performance is competitive when compared to other state-of-the-art CNN results applied to EEG signals. The multi-modal Res1DCNN surpasses the single Res1DCNN without M2SKD—which is trained solely on ECG—by 6.34% in terms of the *Gmean*. This demonstrates that the use of additional features from EEG signals improves the efficiency of seizure detection.

However, the deployability of large DNN models like the multi-modal Res1DCNN can be limited by the memory constraints of embedded medical platforms, as highlighted in [65]. In response, M2SKD serves as a compression technique, aiming to minimize the hardware footprint of a DNN model and, as a result, reduce its inference latency without significantly compromising inference accuracy. The efficiency of the M2SKD framework when applied to the

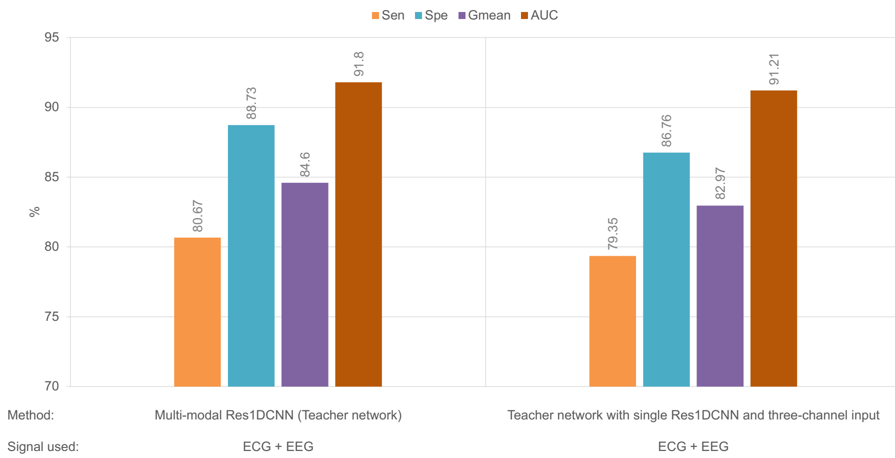

Fig. 12. Global Model Training Strategy for Imbalanced Test Set: Energy Consumption vs. Accuracy Trade-off in Teacher Networks — This figure illustrates a comparison between the original multi-modal Res1DCNN (Teacher network) and an optimized Teacher network utilizing a single Res1DCNN with a three-channel input configuration. The latter design, aimed at reducing energy consumption, processes combined ECG and EEG signals through a modified first layer to accommodate three distinct channels. While this innovative approach offers significant energy savings, it results in a slight decrease in detection accuracy, with sensitivity and specificity dropping by 1.32% and 1.97%, respectively, compared to the traditional multi-channel setup. This trade-off underscores our commitment to developing a high-precision teacher network for epileptic seizure detection.

multi-modal Res1DCNN is further detailed in Figure 11. Notably, this scheme successfully distills a collection of DNN models, such as the multi-modal Res1DCNN, into a solitary DNN model. This single model demonstrates superior performance compared to a DNN model of equivalent size that is trained directly using the same data.

Examining the results, the multi-modal Res1DCNN boasts a geometric mean of 84.6% when both ECG and EEG signals are employed in training and inference. On the other hand, the geometric mean for the single Res1DCNN (student network) with M2SKD, which undergoes training directly on the ECG, dips by only 1.98%.

We also considered two scenarios to evaluate the trade-off between energy consumption, social stigma, and accuracy.

(1) Teacher Network: A Trade-off Between Energy Consumption and Accuracy. Our primary teacher network diverges from the use of three identical Res1DCNNs for EEG1, EEG2, and ECG signals. Instead, it employs a singular DNN capable of processing a three-channel input. To accommodate this, we altered the first layer of the Res1DCNN to handle three distinct channels.
This configuration presents an interesting trade-off: a reduction in energy consumption at the cost of decreased accuracy. Specifically, the detection accuracy of this unified model (sensitivity and specificity) decreased by 1.32% and 1.97% respectively, when compared to the more complex model comprising three distinct Res1DCNNs. The results discussed above are summarized in Fig. 12. Despite the energy savings offered by the teacher model using single Res1DCNN, our primary objective was to craft an energy-efficient wearable system (student model) with high precision using only a single-biosignal input (ECG). To this end, we opted for a teacher model that employs separate 1D networks for each biosignal, ensuring optimal seizure detection.
(2) Student Network: Balancing Accuracy Against Social Considerations. In exploring alternative configurations, we considered a student network solely reliant on EEG signals (EEG1 and EEG2). A distilled DNN (student model) incorporating both EEG signals demonstrated a marginal increase in detection accuracy, specifically 0.71% and

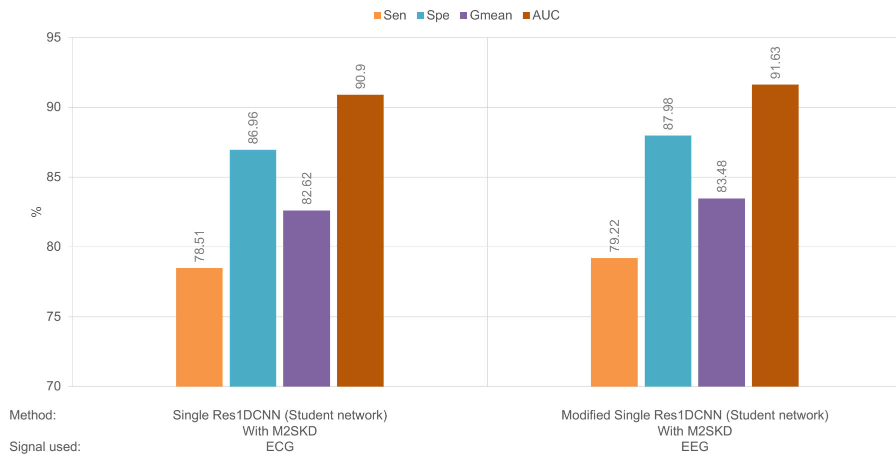

Fig. 13. Global Model Training Strategy for Imbalanced Test Set: ECG vs. EEG in Student Networks — Balancing Accuracy and Social Acceptability. This figure presents a comparison between student networks utilizing ECG and EEG signals, both enhanced with M2SKD, highlighting a detailed approach to epileptic seizure detection. Incorporating both EEG signals (EEG1 and EEG2) into the student model slightly improves detection accuracy, with increases of 0.71% in sensitivity and 1.02% in specificity over the ECG-only model. Despite these gains, the use of EEG signals is accompanied by concerns over social discomfort and stigma due to the necessity of wearing EEG head caps. In response, our research pivots towards ECG-based detection, striking a balance by offering accuracy on par with EEG systems while significantly enhancing user comfort and social acceptance. ECG's lower energy requirements and global availability further ensure broad applicability and the potential for prolonged monitoring across diverse patient populations.

1.02% in sensitivity and specificity, respectively, when compared against training solely on ECG. To facilitate a straightforward comparison, we have consolidated the discussed results in Fig. 13.

Yet, EEG-based approaches come with notable caveats. EEG head caps, while effective, often lead to social discomfort and stigma for patients. Recognizing these challenges, we developed a framework focused on epileptic seizure detection exclusively using ECG signals. This offers a dual advantage: maintaining the accuracy levels comparable to combined EEG and ECG systems while mitigating patient discomfort. The advantages of using ECG-based detection are that it uses less energy and ECG machines are found all over the world. This means that more people, including those in different areas, can use this technology to monitor epilepsy for longer periods.

Figure 14 shows the comparative performance of different training methodologies applied to a Single Res1DCNN (Student network) focusing solely on ECG signals within the context of epileptic seizure detection, especially under the constraints of an imbalanced test set in a Global Model Training Strategy. The methods evaluated include the baseline approach without any knowledge distillation, the utilization of SimCLR, an Autoencoder approach, and our proposed M2SKD.

The baseline student network, without the aid of knowledge distillation, achieved sensitivity (Sen) and specificity (Spe) rates of 76.76% and 79.8%, respectively. Incorporating SimCLR into the training process yielded a slight improvement, indicating the potential benefits of self-supervised learning methods in enhancing model performance. The Autoencoder method showed similar trends with modest gains, highlighting the utility of unsupervised feature learning in this domain. However, a significant leap in performance is observed with our proposed M2SKD framework. The Single Res1DCNN (Student network) with M2SKD demonstrated remarkable improvements, clearly outpacing the other

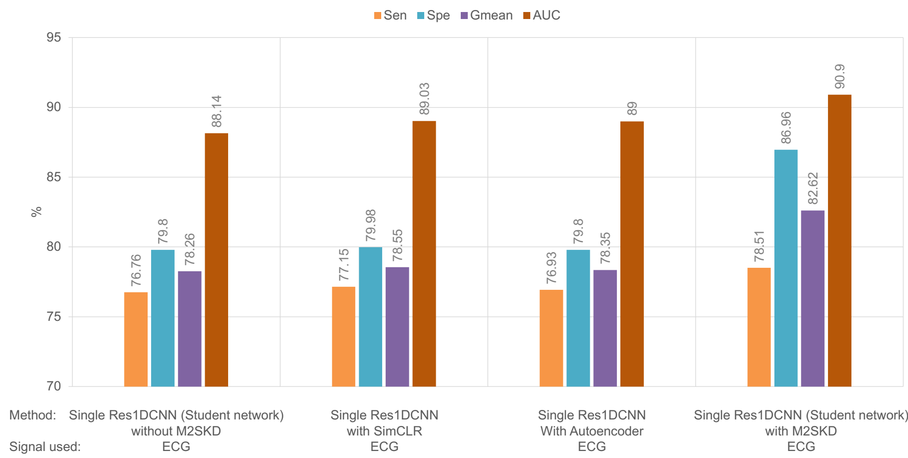

Fig. 14. Global Model Training Strategy: Performance Metrics for Imbalanced Test Set — This figure outlines a comprehensive comparison between four different training approaches applied to the Single Res1DCNN (Student network), focusing on ECG signal processing for epileptic seizure detection. The compared methods include the conventional approach without knowledge distillation, as well as those utilizing SimCLR and Autoencoder techniques, alongside our proposed M2SKD framework. The data illustrates that, while standard and self-supervised learning strategies yield some benefits, the M2SKD-enhanced Student network significantly outperforms them in all key metrics: sensitivity, specificity, geometric mean, and AUC. This notable performance leap underscores the effectiveness of the M2SKD method in improving and leveraging knowledge for the complex task of seizure detection, especially in the context of an imbalanced test set.

methods. This underscores the efficacy of the M2SKD framework in leveraging distilled knowledge for seizure detection, thereby enhancing the accuracy and reliability of the model.

6.1.2 ***Personalized Model Training Strategy:*** In our Personalized Model Training Strategy, illustrated in Fig. 10, we customized the training process for each patient to significantly enhance the model's specificity and practical relevance to individual health profiles. This approach allows for the development of models finely tuned to the fine distinction of each patient's data. For constructing a tailored model for a specific patient, we uniquely assembled the validation and test sets from the data exclusive to that individual. On the contrary, the training set was enriched by combining the data of the specific patient with the collective data of all other participants.

It's crucial to note that our dataset consisted of 29 patients. To demonstrate how effective our method is and to guarantee that it is highly customized for individuals, we carried out thorough and systematic training, conducting the training process separately for each patient. This resulted in a total of 29 unique training iterations for each of the three distinct models within our framework: the "teacher network," the "student network without M2SKD," and the "student network with M2SKD."

As before, a critical component of our methodology is the strategic separation of data based on distinct recording sessions for training, validation and test sets. Therefore, the segmented windows created for the training, validation, and test sets from these different sessions did not overlap. This separation is vital to mitigating biases in time-series medical data. By utilizing data from different recording sessions for each of the sets, we effectively eliminate the potential for temporal bias, ensuring a fair and unbiased evaluation of the model's capabilities.

As shown in Figure 15, we provide a thorough comparison of their detection abilities using pivotal metrics: Sensitivity (Sen), Specificity (Spe), Geometric Mean (Gmean), F1 score, and the Area Under the Curve (AUC).

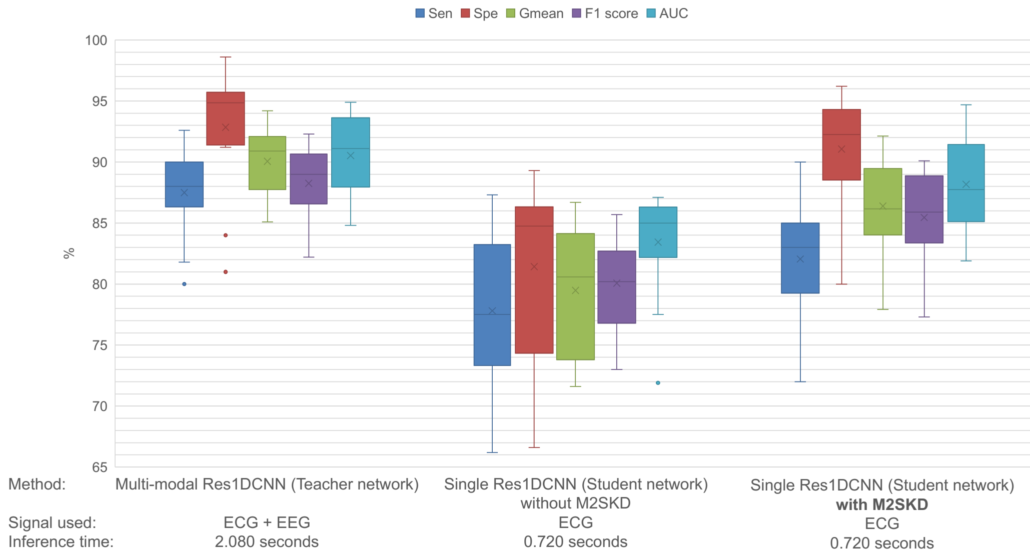

Fig. 15. Personalized Model Training Strategy: Performance Evaluation on an Imbalanced Test Set — This figure provides a detailed performance comparison for epileptic seizure detection using different neural network configurations within our personalized model framework: a multi-modal Res1DCNN (Teacher network) and single Res1DCNN (Student networks), both with and without the implementation of M2SKD. The evaluation is based on a real-world imbalanced dataset and reflects the cumulative results from 29 unique training iterations conducted for each of the three distinct models: the 'Teacher network,' the 'Student network without M2SKD,' and the 'Student network with M2SKD.' The analysis reveals that the M2SKD approach significantly enhances the performance of the Student network, enabling it to approach the effectiveness of the more complex Teacher network that utilizes both ECG and EEG signals, while benefiting from a substantially reduced inference time of 0.720 seconds. These findings highlight the effectiveness of M2SKD in distilling crucial knowledge into a more efficient model, proving its value in personalized, real-time medical monitoring applications.

- Multi-modal Res1DCNN Teacher Network (EEG and ECG signals):
  The Teacher DNN, when supplied with both EEG and ECG signals, showcased superior performance across all metrics. With median Sensitivity and Specificity values of 88% and 94.85%, respectively, this network attained a Gmean of approximately 90.24%, an F1 score of 88.95%, and an AUC of 91.35%. Such consistent performance emphasizes the advantages of integrating EEG and ECG signals in seizure detection.
- Single Res1DCNN Student Network without M2SKD (Only ECG signals):
  Transitioning to the Student Network without knowledge distillation revealed a noticeable performance decrease. Relying exclusively on ECG signals, this network registered median metrics of 78% for Sensitivity, 85.3% for Specificity, 79.89% for Gmean, 81.7% for the F1 score, and 84.65% for AUC. These outcomes illustrate the challenges of using ECG as the sole input without the guidance of a teacher network.
- Single Student Network with M2SKD (Only ECG signals):
  Implementing the M2SKD approach into the Student Network yielded significant improvements. With median values of 84.5% for Sensitivity and 91.9% for Specificity, this network surpassed its non-M2SKD counterpart. Moreover, the Gmean rose to approximately 87.1%, while the AUC climbed to 88.9%, marking notable gains over the Student

Table 2. Quantitative evaluation results of run-time for every 3-second segment in different network architectures.

Method	Platform	Run time (millisec.)
Multi-modal Res1DCNN (Teacher network)	Raspberry Pi Zero	2,080.56 ± 12.56
Single Res1DCNN (Student network) without M2SKD	Raspberry Pi Zero	720.15 ± 32.46
Single Res1DCNN (Student network) with M2SKD	Raspberry Pi Zero	720.15 ± 32.46
Single Res1DCNN (Student network) with M2SKD	Kendryte K210	1,040.64 ± 5.67

Network without M2SKD by 7.21% and 4.25%, respectively. Additionally, the F1 score settled at an average of 85.44%, showcasing an effective harmony between precision and recall.

Our analysis offers several key insights:

- The combination of EEG and ECG signals in the Teacher Network establishes a high standard for seizure detection.
- Relying solely on ECG signals, as observed in the Student Network without M2SKD, results in compromised performance, highlighting the limitations of using only ECG input.
- The M2SKD approach's potency is evident in the Student Network with M2SKD. This network not only recovers from the performance drop seen in the Student Network without M2SKD but also approaches the benchmarks set by the Teacher Network.

In summation, although the combination of EEG and ECG signals offers premier performance, employing the M2SKD method within a Student Network proves highly effective, particularly when limited to ECG signals. The knowledge distillation mechanism evidently harnesses crucial features, positioning it as a promising substitute when EEG data is unavailable.

6.2 Energy Consumption Analysis

A key challenge for low-energy embedded medical platforms with limited computational resources is designing and implementing an epileptic seizure detection algorithm based on DNNs for long-term patient monitoring. For instance, the e-Glass wearable system [23] shown in Fig. 3 contains a 570 mAh battery and features an ultra-low-power 32-bit microcontroller STM32L151 [66] with an ARM® Cortex®-M3 with 48 KB RAM and 384 KB Flash. In the case of an epileptic seizure, e-Glass communicates with a smartphone or smartwatch using Bluetooth low energy (nRF8001) [67] and sends a warning to the caregivers.

In this paper, to analyze the complexity, lifetime, and energy efficiency of our approach, we consider the Kendryte K210 and Raspberry Pi Zero platforms. In the implementation code, all the computations and storage are in 16-bit fixed-point. We chose 13 bits for the fractional part using the results of the validation set. We observed that dedicating more bits to the fractional part causes overflows in the computations. On the other hand, reducing the number of fraction bits gives rise to a considerable accuracy drop. Since we use the fixed-point representation of numbers, we save the amount of storage by a factor of 4 compared to 64-bit floating-point operations. This compression is crucial because

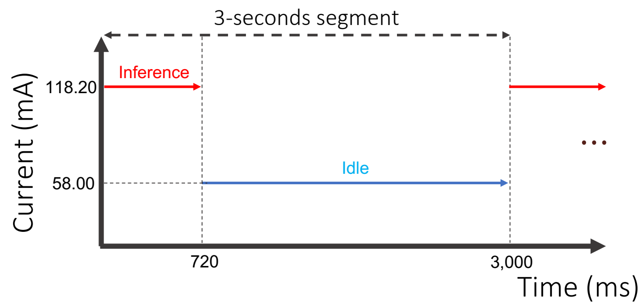

Fig. 16. Real-time energy consumption monitoring on Raspberry Pi Zero. Raspberry Pi Zero performs epileptic seizure detection using the single Res1DCNN (Student network) with M2SKD of a 3-seconds segment in 720 milliseconds and then goes to idle mode.

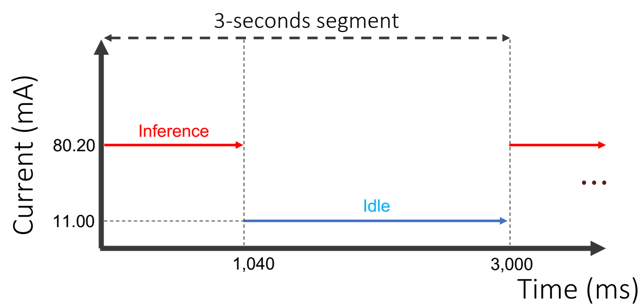

Fig. 17. Real-time energy consumption monitoring on the Kendryte K210. The Kendryte K210 performs epileptic seizure detection using the single Res1DCNN (Student network) with M2SKD of a 3-seconds segment in 1,036 milliseconds and then goes to idle mode with considerably lower energy consumption.

it enables our network to be applicable on various memory-limited embedded devices. However, the total accuracy of the model is reduced by 0.9% because of the quantization in 16-bit.

Table 2 shows the seizure detection execution time per 3-second segment for each DNN. The represented numbers are obtained by running the experiments for the whole test set, including 1568 segments of 3-second segments. We observe that due to fewer parameters, the network obtained by proposed M2SKD, which requires only the ECG signal, runs 2.9 times faster than the multi-modal Res1DCNN on the Raspberry Pi Zero. At the same time, the proposed network achieves a detection performance comparable to multi-modal Res1DCNN (see Fig. 15). In addition to the reduced memory and computational burden, the most relevant advantage of the proposed approach is that it avoids the acquisition and process of EEG data. We also observe that the network's end-to-end response time for a 3-second segment is only 720 milliseconds on the Raspberry Pi Zero and 1,040 milliseconds on the Kendryte K210; thus, we can continuously monitor the patients in real-time.

Epilepsy is characterized by unpredictable seizures and can cause other health problems; thus, the patients have to be monitored on a long-term basis. Table 3 evaluates the battery lifetime of DNN obtained by proposed M2SKD using the battery of the e-Glass [23], which is 570 mAh. We observe that our proposed M2SKD results in a model that operates for 7.86 hours on the Raspberry Pi Zero and 16.29 hours on the Kendryte K210 on a single charge to perform

Table 3. Battery life of an edge device using the e-Glass [23] battery for running each network architecture to perform patient monitoring.

Method	Platform	Battery life (hours)
Multi-modal Res1DCNN (Teacher network)	Raspberry Pi Zero	5.71 ± 0.01
Single Res1DCNN (Student network) without M2SKD	Raspberry Pi Zero	7.86 ± 0.09
Single Res1DCNN (Student network) with M2SKD	Raspberry Pi Zero	7.86 ± 0.09
Single Res1DCNN (Student network) with M2SKD	Kendryte K210	16.29 ± 0.06

real-time epileptic seizure detection. The proposed distilled network achieves a 37.65% energy reduction sacrificing just 1.5% of the accuracy. The battery life is measured considering the real-time energy consumption shown in Fig. 16 and 17. We observe that the Raspberry Pi Zero executes the inference of the 3-seconds segment in 720 ms, then goes to idle mode and waits for the next 3-seconds segment. As Fig. 16 shows, the Raspberry Pi Zero consumes a considerable amount of energy in idle mode. Therefore, as shown in Fig. 17 we considered other wearable platforms, such as the Kendryte K210, that consume less power in the idle mode which would be very beneficial for the patients to perform long-term monitoring. We can improve the battery life of the edge device by using the PULP-based ultra-low-power wearable platform proposed in [22], which consumes only 0.76 mA when the system is clock gated with respect to 23.58 mA when the system is running at 110 MHz, at the lowest energy point of the platform 0.8 V. In this scenario, assuming that the inference time will be similar for the PULP after parallelization [68], we can achieve a battery life of 91.33 hours with the same battery capacity.

7 CONCLUSION

The development of wearable systems that can accurately detect complex pathologies, such as brain disorders, in the long term and with minimal discomfort is still an open challenge. In this work, we have proposed a new knowledge distillation approach (M2SKD) to develop high-precision and low-power wearable systems using single-biosignal input for epileptic seizure detection. As the starting point for our teacher network, we have designed a multi-modal DNN, using information from both ECG and EEG signals, that relies on independent 1-dimensional networks for each biosignal to maximize the detection performance. Then, we used M2SKD to develop a compressed student network that relies exclusively on ECG data during the run-time operation of the wearable system. Besides reducing energy consumption, moving from multi-biosignal to single-biosignal has resulted in removing other major drawbacks in wearable devices when they need to be deployed in real-life setups, such as discomfort, stigma, and synchronization problems where multiple biosignals need to be acquired. Indeed, these benefits are achieved for our distilled network design considering wearable setups while achieving a comparable detection performance with respect to the multi-modal teacher DNN system. The results of our approach, implemented on different edge AI platforms for the wearable context, and tested on the EPILEPSIAE dataset, have shown a 37.65% reduction in energy consumption. In comparison, the resulted sensitivity and specificity are only 1.5% and 1.3% lower than in the initial multi-modal DNN teacher system. Thus, our proposed approach is ideal for the development of wearable setups as it removes the burden of acquiring and synchronizing multiple devices to make valid medical assessments.

REFERENCES

[1] Epilepsy. [Online]. Available:. URL http://www.who.int/mental_health/neurology/epilepsy/en/.

[2] Simon Shorvon and Torbjorn Tomson. Sudden unexpected death in epilepsy. *The Lancet*, 378(9808):2028–2038, 2011.

[3] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, 17(4):2347–2376, 2015.

[4] Philippe Ryvlin and Sandor Beniczky. Seizure detection and mobile health devices in epilepsy: Recent developments and future perspectives. *Epilepsia*, 61, 2020.

[5] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018.

[6] Valentin Gabeff, Tomas Teijeiro, Marina Zapater, Leila Cammoun, Sylvain Rheims, Philippe Ryvlin, and David Atienza. Interpreting deep learning models for epileptic seizure detection on eeg signals. *Artificial Intelligence in Medicine*, page 102084, 2021.

[7] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Hojjat Adeli. Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in biology and medicine*, 100:270–278, 2018.

[8] Lasitha S Vidyaratne and Khan M Iftekharuddin. Real-time epileptic seizure detection using eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):2146–2156, 2017.

[9] Frans SS Leijten, Dutch TeleEpilepsy Consortium, J van Andel, C Ungureanu, J Arends, F Tan, J van Dijk, G Petkov, S Kalitzin, T Gutter, et al. Multimodal seizure detection: a review. *Epilepsia*, 59:42–47, 2018.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. doi: 10.1109/CVPR.2016.90.

[11] Yong-Deok Kim, Eunhyeok Park, Sungjoo Yoo, Taelim Choi, Lu Yang, and Dongjun Shin. Compression of deep convolutional neural networks for fast and low power mobile applications. *arXiv preprint arXiv:1511.06530*, 2015.

[12] Anouk Van de Vel, Kris Cuppens, Bert Bonroy, Milica Milosevic, Katrien Jansen, Sabine Van Huffel, Bart Vanrumste, Patrick Cras, Lieven Lagae, and Berten Ceulemans. Non-eeg seizure detection systems and potential sudep prevention: state of the art: review and update. *Seizure*, 41:141–153, 2016.

[13] Jianping Gou, Baosheng Yu, Stephen John Maybank, and Dacheng Tao. Knowledge distillation: A survey. *arXiv preprint arXiv:2006.05525*, 2020.

[14] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[15] Matthias Ihle, Hinnerk Feldwisch-Drentrup, César A. Teixeira, Adrien Witon, Björn Schelter, Jens Timmer, and Andreas Schulze-Bonhage. Epilepsiae - a european epilepsy database. *Comput. Methods Prog. Biomed.*, 106(3):127–138, June 2012. ISSN 0169-2607.

[16] Renée A Shellhaas. Continuous long-term electroencephalography: the gold standard for neonatal seizure diagnosis. In *Seminars in Fetal and Neonatal Medicine*, volume 20, pages 149–153. Elsevier, 2015.

[17] Miaolin Fan and Chun-An Chou. Detecting abnormal pattern of epileptic seizures via temporal synchronization of eeg signals. *IEEE Transactions on Biomedical Engineering*, 66(3):601–608, 2018.

[18] Dong Wang, Doutian Ren, Kuo Li, Yiming Feng, Dan Ma, Xiangguo Yan, and Gang Wang. Epileptic seizure detection in long-term eeg recordings by using wavelet-based directed transfer function. *IEEE Transactions on Biomedical Engineering*, 65(11):2591–2599, 2018.

[19] Kaveh Samiee, Peter Kovacs, and Moncef Gabbouj. Epileptic seizure classification of eeg time-series using rational discrete short-time fourier transform. *IEEE transactions on Biomedical Engineering*, 62(2):541–552, 2014.

[20] Md Abu Sayeed, Saraju P Mohanty, Elias Kougianos, and Hitten P Zaveri. Neuro-detect: a machine learning-based fast and accurate seizure detection system in the iomt. *IEEE Transactions on Consumer Electronics*, 65(3):359–368, 2019.

[21] Antonio Pullini, Davide Rossi, Igor Loi, Giuseppe Tagliavini, and Luca Benini. Mr. wolf: An energy-precision scalable parallel ultra low power soc for iot edge processing. *IEEE Journal of Solid-State Circuits*, 54(7):1970–1981, 2019.

[22] Elisabetta De Giovanni, Fabio Montagna, Benoît Walter Denkinger, Simone Machetti, Miguel Peon Quiros, Simone Benatti, Davide Rossi, Luca Benini, and David Atienza Alonso. Modular design and optimization of biomedical applications for ultra-low power heterogeneous platforms. Technical report, 2020.

[23] Dionisije Sopic, Amir Aminifar, and David Atienza. e-glass: A wearable system for real-time detection of epileptic seizures. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018.

[24] Carolina Varon, Katrien Jansen, Lieven Lagae, and Sabine Van Huffel. Detection of epileptic seizures by means of morphological changes in the ecg. In *Computing in Cardiology 2013*, pages 863–866. IEEE, 2013.

[25] Maeike Zijlmans, Danny Flanagan, and Jean Gotman. Heart rate changes and ecg abnormalities during epileptic seizures: prevalence and definition of an objective clinical sign. *Epilepsia*, 43(8):847–854, 2002.

[26] Fritz Leutmezer, Christiana Schernthaner, Stefanie Lurger, Klaus Pötzelberger, and Christoph Baumgartner. Electrocardiographic changes at the onset of epileptic seizures. *Epilepsia*, 44(3):348–354, 2003.

[27] Farnaz Forooghifar, Amir Aminifar, Leila Cammoun, Ilona Wisniewski, Carolina Ciumas, Philippe Ryvlin, and David Atienza. A self-aware epilepsy monitoring system for real-time epileptic seizure detection. *Mobile Networks and Applications*, pages 1–14, 2019.

[28] Farnaz Forooghifar, Amir Aminifar, and David Atienza. Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud. *IEEE transactions on biomedical circuits and systems*, 13(6):1338–1350, 2019.

[29] Francesco Onorati, Giulia Regalia, Chiara Caborni, Matteo Migliorini, Daniel Bender, Ming-Zher Poh, Cherise Frazier, Eliana Kovitch Thropp, Elizabeth D Mynatt, Jonathan Bidwell, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. *Epilepsia*, 58(11):1870–1879, 2017.

[30] F Fürbass, S Kampusch, E Kaniusas, J Koren, S Pirker, R Hopfengärtner, H Stefan, T Kluge, and C Baumgartner. Automatic multimodal detection for long-term seizure documentation in epilepsy. *Clinical Neurophysiology*, 128(8):1466–1472, 2017.

[31] Barry R Greene, Geraldine B Boylan, Richard B Reilly, Philip de Chazal, and Sean Connolly. Combination of eeg and ecg for improved automatic neonatal seizure detection. *Clinical neurophysiology*, 118(6):1348–1359, 2007.

[32] Iosif Mporas, Vasiliki Tsirka, Evangelia I Zacharaki, Michalis Koutroumanidis, Mark Richardson, and Vasileios Megalooikonomou. Seizure detection using eeg and ecg signals for computer-based monitoring, analysis and management of epileptic patients. *Expert systems with applications*, 42(6): 3227–3233, 2015.

[33] Marwa Qaraqe, Muhammad Ismail, Erchin Serpedin, and Haneef Zulfi. Epileptic seizure onset detection based on eeg and ecg data fusion. *Epilepsy & Behavior*, 58:48–60, 2016.

[34] Yu Liu, Shobi Sivathamboo, Peter Goodin, Paul Bonnington, Patrick Kwan, Levin Kuhlmann, Terence O'Brien, Piero Perucca, and Zongyuan Ge. Epileptic seizure detection using convolutional neural network: A multi-biosignal study. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–8, 2020.

[35] Mrutyunjaya Sahani, Susanta Kumar Rout, and Pradipta Kishor Dash. Epileptic seizure recognition using reduced deep convolutional stack autoencoder and improved kernel rvfln from eeg signals. *IEEE Transactions on Biomedical Circuits and Systems*, 15(3):595–605, 2021.

[36] Jianping Gou, Liyuan Sun, Baosheng Yu, Shaohua Wan, Weihua Ou, and Zhang Yi. Multilevel attention-based sample correlations for knowledge distillation. *IEEE Transactions on Industrial Informatics*, 19(5):7099–7109, 2022.

[37] Jianping Gou, Liyuan Sun, Baosheng Yu, Shaohua Wan, and Dacheng Tao. Hierarchical multi-attention transfer for knowledge distillation. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(2):1–20, 2023.

[38] Pradip Chitrakar, Chengcui Zhang, Gary Warner, and Xinpeng Liao. Social media image retrieval using distilled convolutional neural network for suspicious e-crime and terrorist account detection. In *2016 IEEE International Symposium on Multimedia (ISM)*, pages 493–498. IEEE, 2016.

[39] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[40] Zhengping Che, Sanjay Purushotham, Robinder Khemani, and Yan Liu. Distilling knowledge from deep networks with applications to healthcare domain. *arXiv preprint arXiv:1512.03542*, 2015.

[41] Qi Dou, Quande Liu, Pheng Ann Heng, and Ben Glocker. Unpaired multi-modal segmentation via knowledge distillation. *IEEE transactions on medical imaging*, 39(7):2415–2425, 2020.

[42] Kang Li, Shujun Wang, Lequan Yu, and Pheng-Ann Heng. Dual-teacher: Integrating intra-domain and inter-domain teachers for annotation-efficient cardiac segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 418–427. Springer, 2020.

[43] Raj D Sheth. Focal vs. generalized seizures: Why classify? *Clinical pediatrics*, 41(2):73–75, 2002.

[44] Damian Pascual, Alireza Amirshahi, Amir Aminifar, David Atienza, Philippe Ryvlin, and Roger Wattenhofer. Epilepsygan: Synthetic epileptic brain activities with privacy preservation. *IEEE Transactions on Biomedical Engineering*, 68(8):2435–2446, 2020.

[45] Xuyang Zhao, Jordi Solé-Casals, Binghua Li, Zihao Huang, Andong Wang, Jianting Cao, Toshihisa Tanaka, and Qibin Zhao. Classification of epileptic ieeg signals by cnn and data augmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 926–930. IEEE, 2020.

[46] Lei Huang, Jie Qin, Yi Zhou, Fan Zhu, Li Liu, and Ling Shao. Normalization techniques in training dnns: Methodology, analysis and application. *arXiv preprint arXiv:2009.12836*, 2020.

[47] Fuhai Huang, Zhenhai Wang, and Xu Bo. Design of 50hz notch filter based on virtual instrument. In *2015 2nd International Conference on Electrical, Computer Engineering and Electronics*, pages 453–456. Atlantis Press, 2015.

[48] Seung-Won Shin, Kyeong-Seop Kim, Chul-Gyu Song, Jeong-Whan Lee, Jeong-Hwan Kim, and Gyeo-Wun Jeung. Removal of baseline wandering in ecg signal by improved detrending method. *Bio-medical materials and engineering*, 26(s1):S1087–S1093, 2015.

[49] Saleh Baghersalimi, Tomas Teijeiro, David Atienza, and Amir Aminifar. Personalized real-time federated learning for epileptic seizure detection. *IEEE Journal of Biomedical and Health Informatics*, pages 1–1, 2021. doi: 10.1109/JBHI.2021.3096127.

[50] Cha Zhang and Yunqian Ma. *Ensemble machine learning: methods and applications*. Springer, 2012.

[51] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.

[52] Cristian Buciluǎ, Rich Caruana, and Alexandru Niculescu-Mizil. Model compression. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 535–541, 2006.

[53] Jangho Kim, SeoungUK Park, and Nojun Kwak. Paraphrasing complex network: Network compression via factor transfer. *arXiv preprint arXiv:1802.04977*, 2018.

[54] Lei Jimmy Ba and Rich Caruana. Do deep nets really need to be deep? *arXiv preprint arXiv:1312.6184*, 2013.

[55] Warren Gay. *Raspberry Pi hardware reference*. Apress, 2014.

[56] Sándor Beniczky, Philippa Karoly, Ewan Nurse, Philippe Ryvlin, and Mark Cook. Machine learning and wearable devices of the future. *Epilepsia*, 62: S116–S124, 2021.

[57] Philip J Fleming and John J Wallace. How not to lie with statistics: the correct way to summarize benchmark results. *Communications of the ACM*, 29(3):218–221, 1986.

[58] Kendryte K210:. URL https://canaan.io/product/kendryteai.

[59] Qoitech AB: Otii arc otii by qoitech (2020):. URL https://www.qoitech.com/otii/.

[60] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[61] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[62] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[63] Saleh Baghersalimi, Behzad Bozorgtabar, Philippe Schmid-Saugeon, Hazım Kemal Ekenel, and Jean-Philippe Thiran. Dermonet: densely linked convolutional neural network for efficient skin lesion segmentation. *EURASIP Journal on Image and Video Processing*, 2019(1):71, 2019.

[64] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[65] Greg Diamos, Shubho Sengupta, Bryan Catanzaro, Mike Chrzanowski, Adam Coates, Erich Elsen, Jesse Engel, Awni Hannun, and Sanjeev Satheesh. Persistent rnns: Stashing recurrent weights on-chip. In *International Conference on Machine Learning*, pages 2024–2033. PMLR, 2016.

[66] STM32L1 Series STMicroelectronics.

[67] Nordic Semiconductor. *nRF8001 Product Specification*, 2015. URL https://www.nordicsemi.com/Products/Low-power-short-range-wireless/nRF8000-series.

[68] Davide Rossi, Francesco Conti, Manuel Eggiman, Alfio Di Mauro, Giuseppe Tagliavini, Stefan Mach, Marco Guermandi, Antonio Pullini, Igor Loi, Jie Chen, Eric Flamand, and Luca Benini. Vega: A ten-core soc for iot endnodes with dnn acceleration and cognitive wake-up from mram-based state-retentive sleep mode. *IEEE Journal of Solid-State Circuits*, 57(1):127–139, 2022. doi: 10.1109/JSSC.2021.3114881.