Multimodal Multiresolution Data Fusion Using Convolutional Neural Networks for IoT Wearable Sensing

Arlene John, *Student Member, IEEE*, Koushik Kumar Nundy, *Senior Member, IEEE*, Barry Cardiff, *Senior Member, IEEE*, and Deepu John, *Senior Member, IEEE*

***Abstract*—With advances in circuit design and sensing technology, the acquisition of data from a large number of Internet of Things (IoT) sensors simultaneously to enable more accurate inferences has become mainstream. In this work, we propose a novel convolutional neural network (CNN) model for the fusion of multimodal and multiresolution data obtained from several sensors. The proposed model enables the fusion of multiresolution sensor data, without having to resort to padding/ resampling to correct for frequency resolution differences even when carrying out temporal inferences like high-resolution event detection. The performance of the proposed model is evaluated for sleep apnea event detection, by fusing three different sensor signals obtained from UCD St. Vincent University Hospital's sleep apnea database. The proposed model is generalizable and this is demonstrated by incremental performance improvements, proportional to the number of sensors used for fusion. A selective dropout technique is used to prevent overfitting of the model to any specific high-resolution input, and increase the robustness of fusion to signal corruption from any sensor source. A fusion model with electrocardiogram (ECG), Peripheral oxygen saturation signal (SpO2), and abdominal movement signal achieved an accuracy of 99.72% and a sensitivity of 98.98%. Energy per classification of the proposed fusion model was estimated to be approximately 5.61 μJ for on-chip implementation. The feasibility of pruning to reduce the complexity of the fusion models was also studied.**

***Index Terms*—Sensor data fusion, wearable sensing, convolutional neural networks, sleep apnea detection, ECG, SpO2.**

I. Introduction

FUSION of data obtained from multiple sensors can improve detection performance, compared to that of using data from a single sensor source [1], [2]. It can also improve the quality and robustness of inferences when noise corrupts data from any of the input sensors. Fusion techniques have been demonstrated to improve the performance of a task without significant variations in the existing data acquisition setup and with minimal additional computational and power consumption costs. Therefore, data fusion has become popular in the design of wearable physiological monitoring systems.

With the advances made in 2-dimensional convolutional neural networks (2D-CNNs) for tasks such as object detection and image classification, a similar model for 1-dimensional (1D) time-series signals has been widely explored for biomedical applications [3]–[5]. Compact 1D-CNNs are less resource-intensive compared to their 2D counterparts and the filters used in these models are equivalent to a simple traditional 1D time-series filter [6], [7]. Due to its efficiency and low computational requirements, 1D-CNNs are suitable for event detection, classification tasks from time-series data and are well suited for fusion algorithms. Often, the different sensors used in this context have different sampling frequencies, and therefore extraction of temporal information from fused data often requires signal padding, re-sampling, etc [8].

Data fusion algorithms are most commonly classified based on the information abstraction level: signal level, feature level, or decision level [9]. Multiple signals are often combined at the signal level to generate one or more signals of the same form but of better quality. Alternatively, fusion can be done after feature extraction or in the final decision stage [10]. Decision-level fusion represents the highest level of abstraction and is commonly used when the signals provided are dissimilar [11], [12]. Fusion systems are also classified based on the relationship between sensor inputs as complementary, redundant, and cooperative [13]. In the context of deep learning, a deep network, namely CentralNet, was proposed for the fusion of information from different sensors that could automatically balance the tradeoff between early and late fusion i.e., between the fusion of low-level versus high-level information [14]. However, the issues arising due to different sampling frequencies of time series data were not discussed. Kim and Ghosh proposed a simple convolutional fusion layer, called the latent ensemble layer, that has a structural advantage in dealing with noise, with each source being allowed to provide a different number of channels, and admitting source-specific features to survive

Manuscript received August 30, 2021; revised October 20, 2021 and December 4, 2021; accepted December 5, 2021. Date of publication December 9, 2021; date of current version February 16, 2022. This work was supported in part by the University College Dublin and in part by the Microelectronic Circuits Centre Ireland. This paper was recommended by Associate Editor Tony Tae-Hyoung Kim. *(Corresponding author: Deepu John.)*

Arlene John is with the School of Electrical and Electronic Engineering, University College Dublin, D04 FX62 Dublin, Ireland (e-mail: arlene.john@ucdconnect.ie).

Koushik Kumar Nundy is with the Think Biosolution, D08 HKR9 Dublin, Ireland (e-mail: kknundy@thinkbiosolution.com).

Barry Cardiff and Deepu John are with the School of Electrical and Electronic Engineering, University College Dublin, D04 FX62 Dublin, Ireland (e-mail: barry.cardiff@ucd.ie; deepu.john@ucd.ie).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TBCAS.2021.3134043.

Digital Object Identifier 10.1109/TBCAS.2021.3134043

even after the fusion procedure [15], but the issues arising due to different sampling frequencies of time series data were not analyzed. A deep convolutional neural network (DCNN), namely SensorNet, was proposed to classify multimodal and multiresolution time series signals by generating a time-series image [8]. However, the data was resampled/ padded such that the input to the DCNN would be uniform, leading to the development of a model that does not address issues such as overfitting to signals with the larger sampling frequency. Undersampling the signal from a high sampling rate sensor can lead to the loss of important information. Oversampling/ padding the signal from a low sampling rate sensor could lead to the development of a larger model with more connections, weights, and additional computations. Therefore, traditional time series feature-based fusion algorithms are not suitable for multiresolution fusion tasks. This leads to the contributions of this work:

1) Development of a 1D-CNN based fusion framework for the data-driven fusion of multisensor, multimodal data at different sampling frequencies for temporal inferences, without having to resort to padding or resampling, by utilizing signal based selective dropout method that ensures that the fusion model does not overfit to the signals with higher sampling frequencies (or features from models with more samples in the input), which is applicable to biomedical Internet of Things (IoT) sensors.
2) A network development methodology that focuses on developing individual well-performing models for each sensor source prior to fusion so that the fused model developed uses the best features from each sensor source for fusion.
3) Development of a novel multimodal sleep apnea detection technique using the proposed fusion framework that is at a higher detection resolution compared to state-of-the-art and its performance evaluation.
4) Complexity analysis of the developed fusion models and their optimization using network pruning techniques to reduce computational costs.[1]

The core hypothesis focuses on analyzing whether an event detection model based on 1D-CNNs can fuse multimodal and multiresolution signals without resorting to resampling of the signals and that the fused models can improve detection accuracies compared to using 1D-CNNs for each signal individually.

The rest of the article is organized as follows: Section II discusses the methodology and experimental design, the fusion stage, and the method to prevent overfitting to signals with high sampling frequencies. Section III discusses the application of the proposed fusion architecture to the sleep apnea detection problem. Section IV includes the performance of the fusion models developed from three different sensor sources and discussions.

II. 1D-CNN Based Fusion for Multi-Sensor Data

This section presents the methodology for developing multimodal and multiresolution fusion models. Consider an event classification task from multi-sensor time-series data obtained from different sensor sources, where data from each source can be used to carry out the classification task with acceptable performance. At first, independent 1D-CNN models for achieving the task have to be developed based on each sensor input signal separately. Here, the hyperparameters of each model have to be tuned to design an optimal independent model based on a single sensor source data. These models are free to have any number of convolutional layers, filters, and filter size, so that the functions that the models learn are optimal. The number of nodes in the output layer of the model will be specified based on the type of classification. The performances of the independent models are tested on a test set to ensure that the model has acceptable performance when only data from a single sensor source is available.

A. Data Driven Fusion Approach

Once all the independent models are able to achieve acceptable performance, the output layers of the models have to be removed, and the flatten layers have to be concatenated together. This concatenated layer should then be attached to a fully connected dense network and then to a final output layer to form the multi-sensor fusion model. The structure of the multi-sensor fusion model will be as shown in Fig. 1, and it works based on a feature-level fusion approach. The same dataset with all different sensor signals is used to train the multi-sensor fusion model. All layers except the final fully connected layer are frozen during training to get the desired performance- equivalent to using the bottleneck features in transfer learning. The hyperparameters of the fully connected layers are then tuned for optimizing the fusion model's performance using the validation set. And finally, the performance of the fusion model is evaluated on the test set. The goal of the fusion algorithm is to improve the overall model performance, but in few scenarios combining a well-performing model with a not-so-well-performing model may cause the performance of the fused model to be lower than the individual models. Since the initial layers are frozen and training is carried out only on the layers after the concatenated flatten stage, the network should be able to learn the correct set of weights based on which of the signals are more reliable for improving detection accuracies. This is beneficial in scenarios where the individual models perform reasonably well and the model performance can be improved by combining information from the features learned by the individual models (which are optimal with regard to the learning algorithm for each individual model), leading to a data-driven approach to fusion and improvements in the performance of the fused model. This is advantageous in wearable devices as individual sensor features in the flatten stage (as they are already optimal for the single sensor network) can be used at any point in time to carry out individual sensor source inference by using the final weights after the flatten stage (maybe due to power constraints, fusion is not preferred and the wearable device can switch between using fusion methodology or using the best performing individual sensor depending on the battery levels easily due to the shared weights prior to the flatten stage) without significant additional memory or area requirements. Moreover, the fused model training time would also be lower with this approach with the confidence that the fused model

[1]Code and models available at [Online]. Available: https://github.com/arlenejohn/multi-resolution_fusion_sleep-apnea

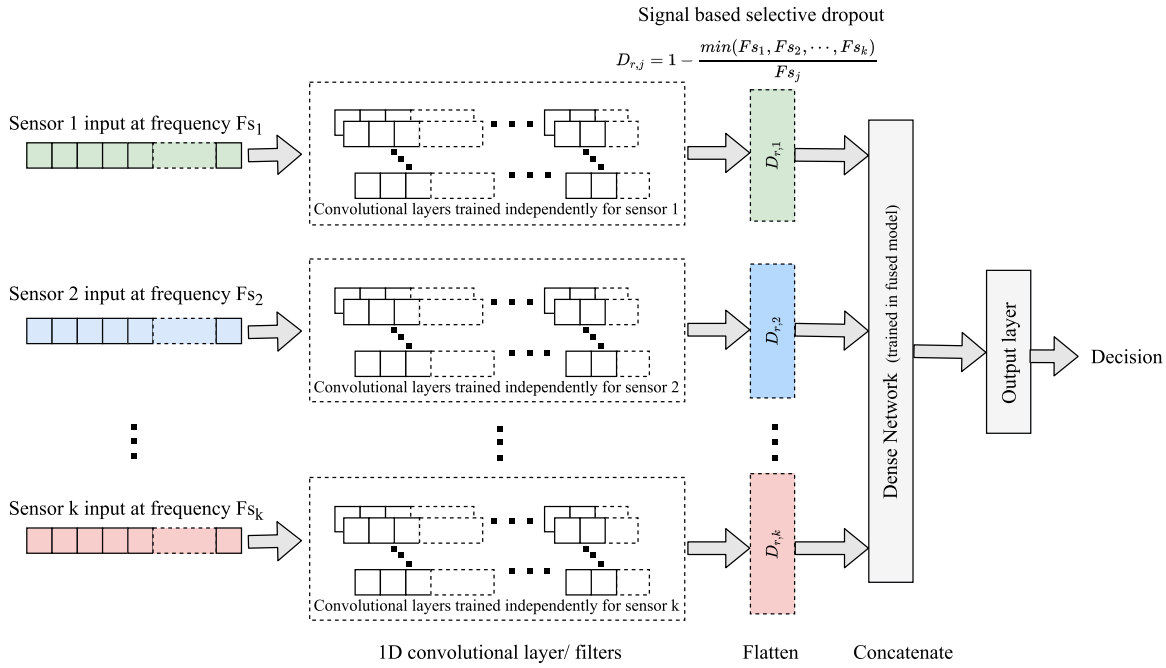

Fig. 1. Proposed 1D-CNN based fusion technique for multi-sensor data.

performance would be equal to or better than the individual model performances.

B. Signal Based Selective Dropout

Consider the scenario where some sensor sources have higher sampling rates than others, leading to more data samples per second. In this scenario, the decision-making of the fusion model will be dominated by the sensor source with a higher sampling frequency (Fs_{high}) than those with a lower sampling frequency (Fs_{low}). This essentially leads to overfitting of the model to the data from the sensor source at the high sampling frequency. Methods to prevent overfitting to training data, such as regularization and dropout, have been previously discussed in literature [16]. However, methods to prevent overfitting to inputs from a single source, in the case of a multi-source input scenario, have not been explored previously. Here, we propose to enable selective dropout during the training stage at the flatten layer for only the features from the signal source with higher sampling rates so that the fusion model would not overfit to the higher sampling rate inputs. The advantage of the signal-based selective dropout is that during the inference stage, all the features in the flatten stage can contribute to the fusion stage compared to undersampling the signal obtained at the higher sampling frequency which leads to the dropped signals or features not being able to contribute at all in the model. Moreover, dropout is random for each training cycle, thereby making sure that all the features at the various positions get to contribute during the training process, which is completely different from undersampling the obtained features (which often leads to the same samples/ samples in fixed positions being dropped and not contributing to the fusion stage). Moreover, this is the first method proposed to prevent overfitting to the model with larger number of neurons in the flatten stage as optimising individual model architectures can lead to different number of neurons in the flatten layers. For signals with different sampling frequencies but matching network structures, the dropout rate for the model for the signal with higher sampling frequency should be set to $1 - \frac{Fs_{\text{low}}}{Fs_{\text{high}}}$. Matched network structures indicate that for two different signals with sampling rates Fs_1 and Fs_2, and corresponding ratio of the input window lengths to the two networks being $\frac{Fs_1}{Fs_2}$, the kernel sizes and number of neurons in each layer of the two networks should have the same ratio of $\frac{Fs_1}{Fs_2}$. In case there is a mismatch in network structure such that the ratio of the neurons in the output layer does not match the ratio $\frac{Fs_{\text{low}}}{Fs_{\text{high}}}$, the dropout ratio D_r should be set as:

$$D_r = 1 - \frac{neurons_{\text{low}}}{neurons_{\text{high}}}, \tag{1}$$

where $neurons_{\text{low}}$ indicate the number of neurons in the flatten layer of model for signal with the lower sampling frequency Fs_{low}, and $neurons_{\text{high}}$ indicate the number of neurons in the flatten layer of model for the signal with the higher sampling frequency Fs_{high}. In the scenario where the fusion of k different sensors at k different sampling frequencies need to be carried out, the dropout ratio D_r for each of the j^{th} signal model can be calculated as:

$$D_{r,j} = 1 - \frac{min(Fs_1, Fs_2, \ldots, Fs_k)}{Fs_j} \forall j \in [1, \ldots, k]. \tag{2}$$

Alternatively, the dropout ratio for the j^{th} signal model in terms of the number of neurons in the flatten layer of each of the individual models can be calculated as:

$$D_{r,j} = 1 - \frac{min(neurons_1, neurons_2, \ldots, neurons_k)}{neurons_j}$$
$$\forall j \in [1, \ldots, k], \tag{3}$$

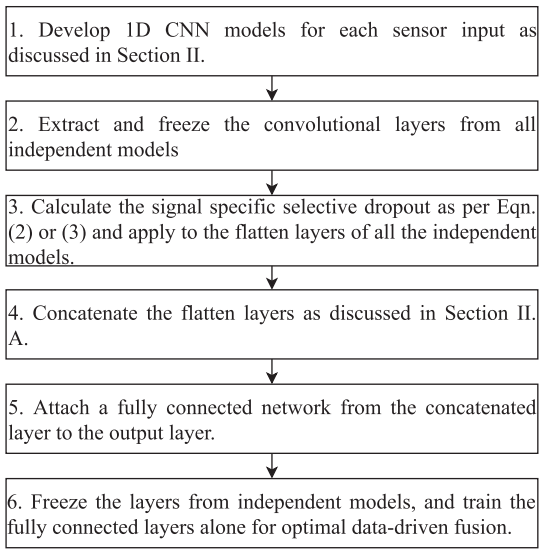

Fig. 2. Flow diagram of the proposed fusion methodology.

where $neurons_j$ corresponds to the number of neurons in the flatten layer of the j^{th} signal model. We call this signal based selective dropout. The flow of the proposed fusion methodology is as shown in Fig 2.

III. Application to Sleep Apnea Detection

The abnormal pause in breathing or lowering of the breath-rate during sleep is termed as the sleep apnea-hypopnea syndrome and is a disorder that affects 10% of middle-aged adults [17]. There is a complete pause in breathing during an apnea event, while a hypopnea event is characterized by a drop in oxygen saturation for at least 10 seconds due to the reduction in airflow [18]. Overnight polysomnography is the diagnostic tool used to diagnose sleep-related disorders under the supervision of a clinician. Recording polysomnograms for clinical evaluation is costly and involves wearing sensors that are uncomfortable for the patient during sleep like airflow thermistors at the nose. Therefore, the development of non-intrusive and automatic sleep-apnea detection methods for deployment in wearable devices like smartwatches and smartvests using physiological signals that can be easily acquired without causing discomfort to the patient is of paramount importance. However, in wearable sensing, these signals can be corrupted by electrode contact noise, power-line interference, motion artifacts, electromyographic noise, etc., leading to poor feature extraction, erroneous data interpretation, and false alarms [19]. This can lead to subsequent slowing down of response times due to alarm fatigue [20].

Electrocardiogram (ECG) is a record of the electrical activity of the heart and can be obtained through non-invasive wearable devices. Peripheral oxygen saturation (SpO2) is an estimation of blood oxygen levels and is usually measured with a pulse oximeter, which is non-invasive and is usually found in smartwatches [21]. Various studies have shown that sleep apnea events can be monitored using ECG and SpO2. An ECG based sleep apnea detection method using ECG derived respiration (EDR) that can detect sleep apnea occurrence during a minute with an accuracy of 94.12% was proposed in [22]. Heart Rate Variability (HRV) features based methods for apnea detection-from ECG signals with apnea event occurrences in 30-second windows- with an accuracy of 74.85% was proposed in [23]. An SpO2 based apnea detection method that could detect apnea event occurrences in a minute with an accuracy of 85.26% was proposed in [24]. A review of the methods discussed in the literature for sleep apnea detection using deep learning was carried out in [25]. In literature, it was observed that most studies carry out sleep-apnea detection on a minute-by-minute basis or for a window of 30 seconds and often use only a single sensor for inference. The highest resolution for sleep apnea detection from ECG signals (every 10 seconds) was studied by Urtnasan *et al.* [26]. The highest resolution of sleep apnea detection from SpO2 signals was explored in [27] with a resolution of 1 s and with a performance accuracy of 79.61%.

Many machine learning-based methods were discussed for sleep apnea detection by combining various features like ECG-derived respiration (EDR) and heart rate variability (HRV) derived from ECG signals [22], [28]–[30]. Convolutional neural networks have been used to generate features that can be used for sleep apnea detection in [26]. CNNs were used in combination with long-short-term-memory (LSTM) for sleep apnea detection from ECG signals [26], [31], [32]. Deep learning methods for apnea detection from SpO2 signals have been discussed in the literature [24]. Fusion methods that combine multiple sensor signals for detection of apnea events are also studied. In [33], fuzzy structural algorithms were utilized to identify and characterize apnea and hypopnea episodes using respiratory airflow and SpO2. Fusion methods based on combinations of features derived from ECG signal and SpO2 for a time segment with a resolution of 1 minute were proposed for apnea detection in [18]. A 2D-CNN based method for sleep apnea detection using SpO2, oronasal airflow, ribcage and abdominal movement was proposed in [27]. A method to fuse ECG and SpO2 signals through a combination of CNNs and LSTM was proposed in [32]. These works and the corresponding performance in sleep apnea detection helped in deciding which complementary signals were to be used for fusion to improve accuracy if any of the signals are corrupted or lost.

We propose a method for sleep apnea detection with a very high resolution of one second. The advantage of a high resolution model is that apnea events can be detected quickly and alert the patient as loss of oxygen for an extended period of time can lead to potential brain damage [34], [35] and faster detection of apnea events could help to reduce the impact. Additionally, depending on the power constraints of a wearable device (remaining battery power), the resolution can be adjusted in real-time by carrying out predictions once every few seconds. Since the model is trained to infer events on a per second basis, resolution can be reduced purely by skipping the signal windows to reduce power consumption, and high resolution processing can be resumed without a large wait time when needed. Apnea detection with a resolution of 1 s is carried out by considering a single signal window of 11 seconds and employing a 1D-CNN for apnea detection. An 11-second window is used as an apnea event is classified as sleep apnea if the patient does not breathe for at least 10 seconds. To achieve this, overlapping windows of duration 11 seconds and an overlap of 10 seconds are generated. A window

TABLE I
ECG SIGNAL MODEL PARAMETERS AS DISCUSSED IN [37]

Layer	Kernel		Output shape	# Parameters
	Size	# kernels		
Batch normalization	-	-	1408,1	4
Convolution	100	3	655,3	303
RELU Activation	-	-	655,3	0
Maxpooling	-	-	327,3	0
Convolution	10	50	318,50	1550
Maxpooling	-	-	159,50	0
RELU Activation	-	-	159,50	0
Convolution	30	30	130,30	45030
Maxpooling	-	-	65,30	0
RELU Activation	-	-	65,30	0
Batch normalization	-	-	65,30	120
Flatten	-	-	1950	0
Dropout	-	-	1950	0
Dense, Softmax	-	-	2	3902

TABLE II
SPO2 SIGNAL MODEL PARAMETERS AS DISCUSSED IN [38]

Layer	Kernel		Output shape	# Parameters
	Size	# kernels		
Batch normalization	-	-	88,1	4
Convolution	25	6	88,6	156
RELU Activation	-	-	88,6	0
Convolution	-	-	88,50	3050
Maxpooling	10	50	44,50	0
RELU Activation	-	-	44,50	0
Convolution	-	-	44,30	22530
Maxpooling	15	30	22,30	0
RELU Activation	-	-	22,30	0
Batch normalization	-	-	22,30	120
Flatten	-	-	660	0
Dropout	-	-	660	0
Dense, Softmax	-	-	2	1322

is assigned the label (apnea/ non-apnea) depending on whether the 2nd 1 s signal in that window is apneic or non-apneic. From a practical machine learning model training perspective, this way of window selection gave more data examples to train the model with, that probably lead to the high detection accuracies of the models obtained.

A. Dataset and Models

For this work, the St. Vincent's University Hospital's sleep apnea database is used [36]. The database contains polysomnogram data from 25 patients with annotations from a sleep expert. Signals recorded were: electroencephalogram, electrooculogram, submental electromyogram, oro-nasal airflow, ribcage movements, abdomen movements, snoring, body position, as well as the signals we are interested in- ECG signals at 128 Hz (Fs_{high}) and SpO2 (from finger pulse oximeter) at 8 Hz (Fs_{low}). The ECG and SpO2 signals are suitable for apnea detection as they can be easily incorporated into wearable devices. The labels provided by the sleep experts were used to mark sleep seconds as apneic or non-apneic. The dataset is split into training, validation, and test set in the ratio of 8:1:1. The class imbalance problem was addressed by oversampling the minority class. It is to be noted that the oversampling of the minority class (apnea events) signal examples is different to signal resampling to correct for resolution differences as discussed in Section I. The oversampling of the minority classes increases the number of examples used to train the 1D-CNN model, while signal resampling increases or decreases the window length of the data fed into the model. For the individual signal models, we use the model proposed in [37] for the ECG signal. The parameters of the model are as shown in Table I. For the SpO2 signal, we use the model proposed in [38], and the parameters of the model are as shown in Table II. The weights and biases were selected during training by using a validation callback method that stored the best parameters that exhibited the highest accuracy on the validation set. Since the number of neurons in the flatten layer of the ECG model was 1950 and that for the SpO2 model was 660, the dropout rate for the features extracted from the ECG model in the flatten layer was fixed at 0.67. The flattened layers from the two models are concatenated and batch-normalized prior to the output layer which uses a softmax activation.

IV. RESULTS

The total number of parameters from the fused model using ECG and SpO2 signals is 88,529, out of which 78,087 parameters are from the convolution layers and are frozen. The model was trained using the adam optimizer. Adam can be looked at as a combination of the RMSprop that uses the squared gradients to scale the learning rate individually for each parameter (2nd moment) and moving average of the gradient instead of gradient like in stochastic gradient descent with momentum. The advantage of the Adam optimizer is that it foregoes the need for a learning rate scheduler and is robust to any reasonably bounded learning rate for the problem at hand, and here the learning rate of 0.001 was chosen. The model was trained over the full training set and simultaneously validated on the validation set for each epoch, where a true positive stands for an accurately detected apnea event. Validation callback was used to determine the best weights during training based on which set of weights exhibited the highest validation accuracy. We evaluate the model performance based on the Sensitivity (Se), Specificity (Sp), and Accuracy (Ac) that are calculated as shown in Eqs. (4), (5), and (6).

$$\text{Se}\ (\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{4}$$

$$\text{Sp}\ (\%) = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \tag{5}$$

$$\text{Ac}\ (\%) = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \times 100 \tag{6}$$

However, due to the huge class imbalance with 50368 test set examples of non-apnea events and just 1368 apnea event samples, missing a few apnea events (FNs) lead to very low sensitivity values, although the overall accuracy doesn't get affected much due to a large number of true negatives. Therefore, a metric that can capture these two components is required. For this, we use the F_β- score- of which the F1-score is the most commonly used- such that both sensitivity (Se) and precision are taken into account with the right importance assigned to sensitivity and precision that we require. The precision/ Positive Predictive Value (PPV) is calculated as shown in Eqn (7).

$$\text{PPV}\ (\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{7}$$

The F_β- score is calculated from Se and PPV as shown in Eqn (8).

$$F_\beta - score(\%) = \frac{(1+\beta^2).\text{PPV.Se}}{\beta^2.\text{PPV} + \text{Se}} \times 100 \tag{8}$$

In this case, we are placing β times more importance in correctly identifying true positives (sensitivity) than reducing the number of false positives or false negatives (precision). Since there are approximately 36 times more non-apnea events than apnea events, we set $\beta = \sqrt{36} = 6$, and therefore we call this metric the F6-score (F6).

The ECG signal-based model ($M_1 1$ (1 in the subscript indicates single sensor)) was found to exhibit an F6-score of 95.82%, a sensitivity of 96.05%, and a specificity of 99.75% as discussed in [37]. The SpO2 signal-based model ($M_1 2$) was found to exhibit an F6-score of 82.86%, a sensitivity of 84.65%, and a specificity of 97.42% as discussed in [38]. The performance of the SpO2 model is not as good as the ECG signal-based model, however, the overall performance of the fused model should be better than that of the individual models. The fused model ($M_2 1$ (the 2 in the subscript indicates two sensors)) was found to exhibit an F6-score of 97.20%, a sensitivity of 97.44%, and a specificity of 99.68%. On comparing the performance of the fusion model with the SpO2 alone model, there is a significant performance improvement. However, when comparing the performance of the fusion with the ECG signal, there is only a slight improvement in overall performance. We also compare the performance of the fusion model ($M_2 1$) with a fusion model without the dropout layer ($\tilde{M}_2 1$) applied to the sensor model with the higher sampling frequency/ higher number of neurons in the flatten layer as discussed in II-B. In the case where we ignore the possibility of overfitting to the ECG based model, the ECG+Spo2 fused model ($\tilde{M}_2 1$) exhibited an F6-score of only 96.86%, which is poorer than the model that accounted for the possibility of overfitting to the ECG based model ($M_2 1$), thereby proving that the selective dropout method is suitable to prevent overfitting. However, this is not much evident in the case where both signals are clean. Overfitting to the sensor source with a higher sampling rate becomes more pronounced in the case where the signal sources are corrupted by noise, which is a major challenge in wearable monitoring. Severe motion artifacts, electrode contact noise, and electromyographic noise can affect data obtained from wearable devices. The main advantage of fusion methods becomes evident when any of the sensor sources are corrupted by noise, which is discussed in the next subsection.

A. Performance Evaluation With Noisy Data

To analyze the performance of fusion models with noisy data, we have simulated scenarios with noisy signal windows, where all the samples in a noisy signal window are corrupted by noise. This is achieved by adding -20 dB white Gaussian noise to the signal segments. The approach of using white Gaussian noise is to capture all noise frequencies, although during sleep, the low frequency noises would be dominant. However, as per the central limit theorem, the sum of multiple independent distributions tends to be a gaussian distribution, and therefore white gaussian noise is added to the signals to study the performance in noisy scenarios. For the training and validation set, all the signal samples in a window in 20% of the ECG window set are noisy while the corresponding SpO2 window set are clean, all the signal samples in a window in 20% of the SpO2 window set are noisy while the corresponding ECG window set are clean, and all the signal samples in a window in 20% of both SpO2 and ECG window set are noisy for both apnea and non-apnea events. For the test set, all the signal samples in a window in 11.42% of the ECG window set are noisy while corresponding SpO2 window set are clean, all the signal samples in a window in 11.42% of the SpO2 window set are noisy while corresponding ECG window set are clean, and all the signal samples in a window in 11.42% of both SpO2 and ECG window set are noisy for both apnea and non-apnea events. The individual models for ECG and SpO2 were retrained using this dataset with simulated noise and fused as discussed in Section III-A. The ECG signal-based model, $M_{1,n} 1$ (the 'n' in the subscript indicates that the model is generated for dataset containing noisy samples) was found to exhibit an F6-score of 74.80%, a sensitivity of 74.85%, and a specificity of 99.24%, and the SpO2 signal-based model ($M_{1,n} 2$) was found to exhibit an F6-score of 72.22%, a sensitivity of 77.85%, and a specificity of 91.56%. The fused model with selective dropout ($M_{2,n} 1$) was found to exhibit an F6-score of 82.3%, a sensitivity of 85.01%, and a specificity of 96.33%. From the results, it can be observed that the F6-score and sensitivity are significantly higher for the fusion model compared to ECG and SpO2 models. A fused model without the selective dropout stage ($\tilde{M}_{2,n} 1$) was found to exhibit an F6-score of 81.55%, a sensitivity of 83.41%, and a specificity of 97.26%. From the performance of the ECG and SpO2 models as well as the fused model without selective dropout, it can be seen that the models are highly biased towards the majority class, even with minority class oversampling. These observations indicate that when the signals are corrupted by noise, important features that can be learned on clean signals are not learned by the model, and in such scenarios, the fusion algorithm proposed in this article will be useful in tackling this issue.

B. Generalization to Multiple Sensors

To expand and generalize on the proposed method, we generalize the fusion model to include a third sensor source. Here, we consider the abdominal movement signals as the sensors used to obtain abdominal movement signals can be easily incorporated in a wearable smartvest. Abdominal movements are found to be highly reliable parameters for the sleep apnea detection problem [39]–[41]. Lin *et al.* proposed an adaptive nonharmonic model to model thoracic and abdominal movement signals to design features for sleep apnea events [39]. In [40], wavelet domain features from abdominal, chest, and nasal way signals are used to detect obstructive sleep apnea events using ensembles. In [41], a measuring module integrating abdominal and thoracic triaxial accelerometers, SpO2, and an ECG sensor was devised and a long short-term memory recurrent neural network model was proposed to classify four types of sleep breathing patterns. We use the abdomen movement signals recorded using strain gauges from the St. Vincent's University Hospital's sleep apnea database [36]. The dataset is split into training, validation, and

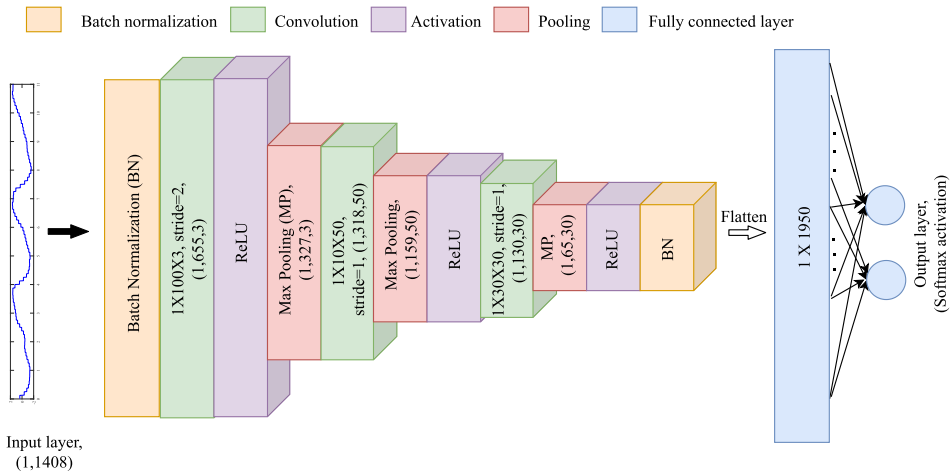

Fig. 3. Independent 1D-CNN fusion model for sleep apnea detection from abdominal movements.

TABLE III
THE MODEL PARAMETERS FOR THE ABDOMEN MOVEMENT SIGNAL BASED MODEL

Layer	Kernel		Output shape	# Parameters
	Size	# kernels		
Batch normalization	-	-	1408,1	4
Convolution	100	3	655,3	303
RELU Activation	-	-	655,3	0
Maxpooling	-	-	327,3	0
Convolution	10	50	318,50	1550
Maxpooling	-	-	159,50	0
RELU Activation	-	-	159,50	0
Convolution	30	30	130,30	45030
Maxpooling	-	-	65,30	0
RELU Activation	-	-	65,30	0
Batch normalization	-	-	65,30	120
Flatten	-	-	1950	0
Dropout	-	-	1950	0
Dense, Softmax	-	-	2	3902

TABLE IV
PERFORMANCE OF THE SINGLE SIGNAL MODELS

Model	**F6 (%)**	**Se (%)**	**Sp (%)**	**Ac (%)**
$M_1$1 (ECG)	95.82	96.05	99.75	99.56
$M_1$2 (SpO2)	82.86	84.65	97.42	97.08
$M_1$3 (Abdo)	98.34	98.76	99.54	99.52

test set in the ratio of 8:1:1. The class imbalance problem was addressed by oversampling the minority class. We developed a 1D-CNN model as discussed next for the abdominal signal.

1) 1D-CNN Model Based on Abdominal Signal: The 1D-CNN model developed for abdominal movement signals for sleep apnea detection is as shown in Fig. 3, and the model parameters are detailed in Table III. The model optimization used binary cross-entropy as the loss function with the adam optimizer and, the training procedure was similar to that used to generate the ECG- and SpO2-based models. The weights and biases were selected during training by using a validation callback method that stored the best parameters that exhibited the highest accuracy on the validation set. We refer to the abdominal movement-based single sensor model as $M_1$3. The results of the three independent models are summarised in Table IV.

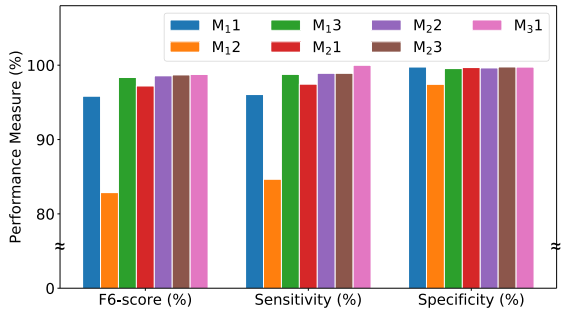

Fig. 4. Performance of the single signal models and fusion models employing selective dropout.

The model exhibited an F6-score of 98.34%, a sensitivity of 98.76%, and a specificity of 99.54% on the test set consisting of data from all the patients.

C. Fusion Algorithm With Abdominal Signal

The 1D-CNN model based on abdominal signal is fused with the SpO2 signal-based model (SpO2+Abdo $\rightarrow$ $M_2$2) and the ECG signal-based model (ECG+Abdo $\rightarrow$ $M_2$3). Since the number of samples used in the abdominal signal-based model is equal to the number of samples used in the ECG signal model, sampling frequency-based selective dropout is not required. However, when the abdominal signal-based model is fused with the SpO2 signal-based model, we need to account for the possible overfitting to the abdominal signal model features in the fusion model, and therefore signal features-based selective dropout needs to be carried out. For the abdominal signal-based model, the number of neurons in the flatten layer is 1950, and the number of neurons in the flatten layer of the SpO2 signal-based model is 660, and therefore the selective dropout ratio is set to 0.67. The performance of the fusion models in terms of F6-score, sensitivity, and specificity is as shown in Fig. 4, along with the three new fusion models, namely SpO2+Abdo ($M_2$2), ECG+Abdo ($M_2$3), and ECG+SpO2+Abdo ($M_3$1). On comparing with the ECG+SpO2 model ($M_2$1), it can be seen that

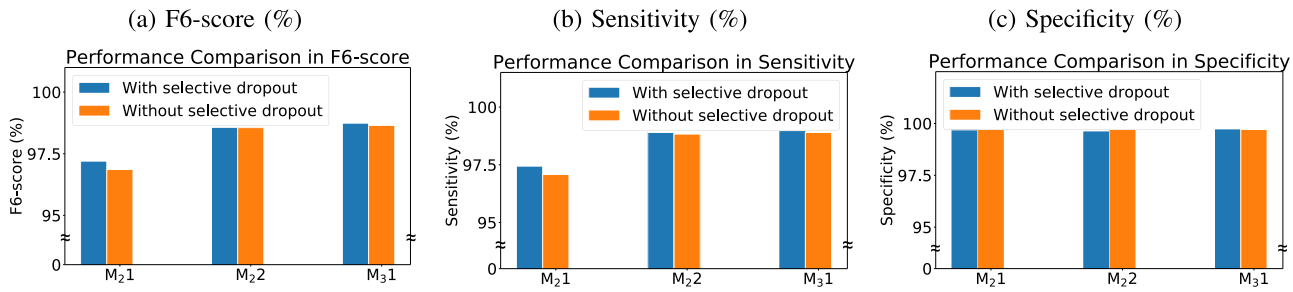

Fig. 5. Comparison of the fusion models with and without selective dropout.

TABLE V
PERFORMANCE OF THE FUSION MODELS AND WITHOUT WITHOUT SELECTIVE DROPOUT

Model	F6 (%)	Se (%)	Sp (%)	Ac (%)
With Selective dropout				
$M_2 1$ (ECG+SpO2)	97.20	97.44	99.68	99.62
$M_2 2$ (Abdo+SpO2)	98.57	98.90	99.63	99.61
$M_2 3$ (ECG+Abdo)	98.69	98.90	99.76	99.73
$M_3 1$ (ECG+SpO2+Abdo)	98.75	98.98	99.74	99.72
Without Selective dropout				
$\tilde{M}_2 1$ (ECG+SpO2)	96.87	97.08	99.70	99.63
$\tilde{M}_2 2$ (Abdo+SpO2)	98.56	98.83	99.70	99.68
$\tilde{M}_2 3$ (ECG+Abdo)[*]	98.69	98.90	99.76	99.73
$\tilde{M}_3 1$ (ECG+SpO2+Abdo)	98.65	98.90	99.71	99.70

*The individual ECG and Abdo models have the same number of neurons in the flatten layer, and therefore selective dropout ratio for both would be 0. Hence, $M_2 3$ and $\tilde{M}_2 3$ exhibits the same performance.

TABLE VI
PERFORMANCE OF THE SINGLE SIGNAL MODELS WHEN THE DATASET CONTAINS NOISY SAMPLES

Model	F6 (%)	Se (%)	Sp (%)	Ac (%)
$M_{1,n} 1$ (ECG)	74.80	74.85	99.24	98.60
$M_{1,n} 2$ (SpO2)	72.22	77.85	91.56	91.20
$M_{1,n} 3$ (Abdo)	83.05	88.45	93.15	93.03

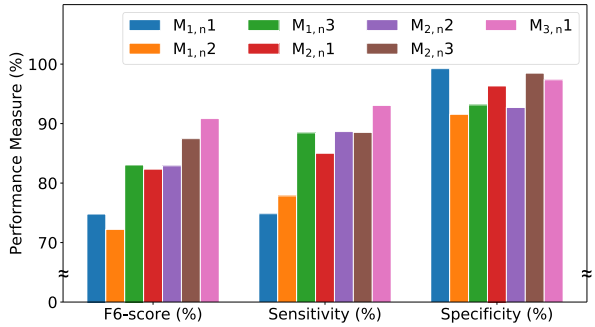

Fig. 6. Performance of the single signal models and fusion models with selective dropout in the presence of noise.

the sensitivity is higher when abdominal signal data is included (in models $M_2 2$, $M_2 3$, and $M_3 1$), and with higher F6-score for when ECG and abdominal models are fused ($M_2 3$). The results demonstrate that increasing the number of sensor sources generally improve the performance of the fusion models. We also check for the efficacy of the selective dropout method to prevent overfitting to the sensor source with a larger sampling rate. In the case of the SpO2+Abdo model with selective dropout ($M_2 2$) an F6-score of 98.57% is exhibited, while for a fused SpO2+Abdo model without selective dropout ($\tilde{M}_2 2$), the F6-score was just slightly lower at 98.56%. However, in the case of the ECG+SpO2+Abdo model, the difference is slightly more pronounced. For the ECG+SpO2+Abdo model with selective dropout ($M_3 1$), and F6-score of 98.75% is exhibited, while for a fused ECG+SpO2+Abdo model without selective dropout ($\tilde{M}_3 1$), the F6-score was lower at 98.65% as shown in Fig. 5. In the case of model $M_2 3$, since both ECG and abdominal movement signals have the same number of samples, selective dropout is not employed. The results of the fusion algorithms with and without selective dropout are as discussed in Table V.

We next analyze the performance of the fusion model with noisy input data.

1) Performance of Fusion With Three Inputs and Noisy Data: For the training and validation set, few windows in the dataset have been corrupted by noise in the following manner:

1) All the signal samples in a window in 10% of the ECG window set is noisy while corresponding SpO2 and abdomen window sets are clean, all the signal samples in a window in 10% of the SpO2 window set is noisy while corresponding ECG and abdomen window sets are clean, and all the signal samples in a window in 10% of the abdomen movement window set is noisy while corresponding ECG and SpO2 window sets are clean.
2) All the signal samples in a window in 10% of the ECG and SpO2 window sets are noisy while corresponding abdomen window set is clean, all the signal samples in a window in 10% of the SpO2 and abdomen window sets are noisy while corresponding ECG window set is clean, and all the signal samples in a window in 10% of the ECG and abdomen movement window sets are noisy while corresponding SpO2 window set is clean.
3) 10% of the ECG signal, SpO2 signal, and abdomen signal window sets are noisy.

For the test stage, the few samples in the test set have been corrupted by adding noise to the samples in a similar manner as discussed for the training and validation set, by adding noise to all the samples in a window for 5% of the signal window set from each of the 7 different noise combinations.

The individual models were retrained using this dataset with added noise and fused as discussed in Section III-A. The performances of the independent models are summarised in Table VI. The performance of the fusion models in terms of F6-score, sensitivity, and specificity is as shown in Fig. 6. From the results, it can be observed that F6-score is higher for the fusion models

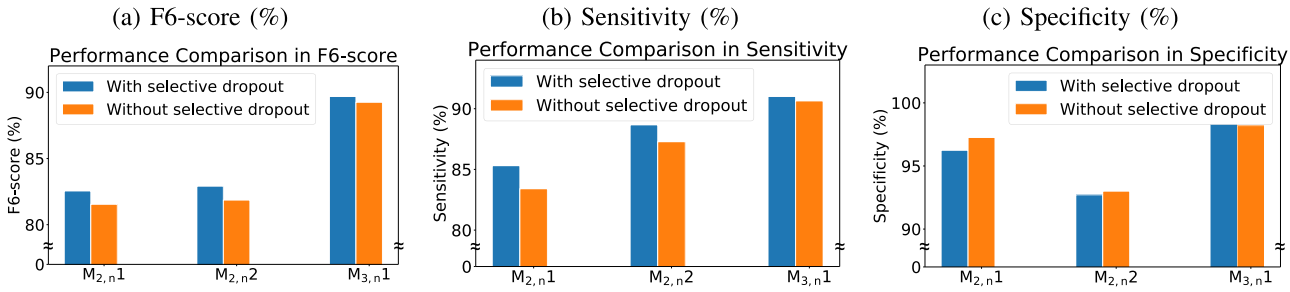

Fig. 7. Comparison of the fusion models with and without selective dropout when the dataset contains noisy signal samples.

compared to the individual models. From the performance of the ($M_{1,n}1$) and ($M_{1,n}2$) models, it can be seen that the models are highly biased towards the majority class, even with minority class oversampling. These observations indicate that when the signals are corrupted by noise, important features that can be learned on clean signals are not learned by the model. The performance of the fusion model is also dependant on the quality of the individual models, like in the case of model $M_{2,n}2$, the performance is poorer with an F6-score of 82.91%, compared to the single sensor model $M_{1,n}3$ with an F6-score of 83.05%, as when fusing with the SpO2 model with just an F6-score of 72.22%, reduced the specificity of the fused model.

The performance of the fused model $M_{2,n}2$ could have been improved upon by changing the training routine, reducing the regularization coefficients, etc, but for consistency across training routines across all models were maintained for comparison and we report these results here. From the results, it can be observed that in the noisy scenario, the three sensor model exhibits a significant improvement in F6-score and sensitivity compared to the single sensor or two-sensor fusion models, which was not as evident in the non-noisy scenario. From the results, it is noticeable that the fusion model with ECG+SpO2 model ($M_{2,n}1$ and $M_2 1$) does not outperform the Abdo model ($M_{1,n}3$ and $M_1 3$), leading to the conclusion that Abdo movement needs to be incorporated into the fusion algorithms for best performance, but the Abdo movement recordings is not a standard signal acquired from wearable devices. However, the ECG and SpO2 signals can be easily incorporated into the current wearable devices ecosystem, and the ECG+SpO2 ($M_2 1$) fusion algorithm is well suited for wearable devices as they exhibit significant performance improvement over the ECG alone ($M_1 1$) or SpO2 alone ($M_1 2$) model for both clean signal and noisy signal scenarios. However, in the scenario where abdominal movement signals can be obtained, the Abdo+ECG model ($M_2 3$) can be used to improve performance over the single sensor models, and the larger ECG+SpO2+Abdo model ($M_3 1$) can be used when the power constraints are not pressing. We also check for the efficacy of the selective dropout method to prevent overfitting to the sensor source with the larger sampling rate. In the case of the $M_{2,\ n}2$ model an F6-score of 82.91% is exhibited, while for a $M_{2,n}2$ model without selective dropout ($\tilde{M}_{2,n}2$), the F6-score was lower at 81.86% as shown in Fig. 7. For the model $M_{3,n}1$, and F6-score of 89.69% is exhibited, while for the fused model without selective dropout ($\tilde{M}_{3,n}1$), the F6-score was lower at 89.25%. The results of the fusion algorithms with and without selective dropout when the dataset contains noisy samples are as discussed in Table VII. The studies with and without selective dropout indicates that using selective dropout improves the model sensitivity with no additional computational cost during inference as selective dropout is applied only during training to prevent over fitting to the model that provides more features at the flatten stage of the fusion algorithm.

TABLE VII
PERFORMANCE OF THE FUSION MODELS WITH AND WITHOUT SELECTIVE DROPOUT WHEN THE DATASET CONTAINS NOISY SAMPLES

Model	F6 (%)	Se (%)	Sp (%)	Ac (%)
With Selective dropout				
$M_{2,n}1$ (ECG+SpO2)	82.34	85.01	96.33	96.04
$M_{2,n}2$ (Abdo+SpO2)	82.91	88.67	92.72	92.61
$M_{2,n}3$ (ECG+Abdo)	87.47	88.52	98.48	98.22
$M_{3,n}1$ (ECG+SpO2+Abdo)	90.85	93.05	97.37	97.25
Without Selective dropout				
$\tilde{M}_{2,n}1$ (ECG+SpO2)	81.55	83.41	97.26	96.89
$\tilde{M}_{2,n}2$ (Abdo+SpO2)	81.86	87.28	93.00	92.85
$\tilde{M}_{2,n}3$ (ECG+Abdo)*	87.47	88.52	98.48	98.22
$\tilde{M}_{3,n}1$ (ECG+SpO2+Abdo)	89.25	90.64	98.18	97.98

*The individual ECG and Abdo models have the same number of neurons in the flatten layer, and therefore selective dropout ratio for both would be 0. Hence, $M_{2,n}3$ and $\tilde{M}_{2,n}3$ exhibits the same performance.

D. Complexity Analysis

The computational complexities of the three networks were calculated in terms of the number of multiplications and additions required for each second [42]. The computations involved in the convolution layers were calculated based on simple filtering calculation count and without assuming a fast Fourier transform approach [43]. For the Max Pooling layers, the operation of selecting the maximum is approximated to an addition operation. For the dense layers, the calculations are carried out as discussed in [42]. Detection of sleep apnea events with ECG signal using $M_1 1$ requires 6534116 multiplications and 6546647 additions [37], and with SpO2 signal using $M_1 2$ requires 1270016 multiplications and 1272876 additions [38]. The fusion model $M_2 1$ requires 7809352 multiplications and 7824743 additions. Detection of sleep apnea events with abdomen movement signals using $M_1 3$ requires 6534116 multiplications and 6546647 additions like the ECG-based model $M_1 1$ due to the same network structure. The fusion model $M_2 2$ that fuses the SpO2 and abdomen movement signal requires 7809352 multiplications and 7824743 additions. The fusion model $M_2 3$ that fuses the ECG and abdomen movement signal requires 13076032 multiplications and 13101094 additions.

TABLE VIII
MODEL COMPLEXITY IN TERMS OF MULTIPLICATIONS AND ADDITIONS AND ESTIMATED ENERGY CONSUMPTION

Models	Multiplications	Additions	Energy (uJ)
$M_1 1$	6534116	6546647	2.55
$M_1 2$	1270016	1272876	0.50
$M_1 3$	6534116	6546647	2.55
$M_2 1$	7809352	7824743	3.05
$M_2 2$	7809352	7824743	3.05
$M_2 3$	13076032	13101094	5.10
$M_3 1$	14347368	14375290	5.61

The fusion model $M_3 1$ that fuses all the three signals requires 14347368 multiplications and 14375290 additions. The total energy consumption during prediction is estimated by assuming that the energy required for a 16-bit multiplication accumulation (MAC) operation is 0.39 pJ [44], [45], and for a 16-bit adder is around 20 fJ [46] in 28 nm FD-SOI technology. The energy consumption during prediction using $M_1 1$ is found to be 2.55 μJ, prediction using $M_1 2$ is found to be 0.50 μJ, and with $M_1 3$ is found to be 2.55 μJ. The total energy consumption during prediction using the fusion models $M_2 1$ is found to be 3.05 μJ, prediction using $M_2 2$ is found to be 3.05 μJ, and with $M_2 3$ is found to be 5.10 μJ. The energy consumption for prediction using the three-sensor fusion model $M_3 1$ is 5.61 μJ. The complexity in terms of multiplications and additions and the corresponding energy consumption are discussed in Table VIII. The corresponding models trained without selective dropout and the models generated for the dataset with noisy samples require the same number of computations as discussed above due to the same network structure. From performance and computational complexity analysis, abdominal signal-based models are found to be most suitable for the task. However, for ease of integration into the current generation of wearables, the ECG+SpO2 model is most ideal. For the dataset used in the clean signal scenario, the fusion model that combines all three signals does not significantly outperform the ECG+SpO2 model or the ECG+Abdo model. However, in the scenario where signals are noisy, the fusion model that combines all three signals significantly outperform the 2-signal fusion models, indicating that as the number of sensors used in fusion increase, the reliability of inferences also increases. As discussed in the previous section the ECG+SpO2 ($M_2 1$) fusion algorithm is well suited for integration into current wearable devices ecosystems. However, in the scenario where abdominal movement signals can be obtained, the Abdo+SpO2 model ($M_2 2$) which is the least computationally intensive of the abdominal movement based fusion models can be used. But the Abdo+SpO2 ($M_2 2$) model does not perform as well as the more computationally intensive ECG+Abdo model ($M_2 3$), and the larger ECG+SpO2+Abdo model ($M_3 1$) can be used when the power constraints are are not pressing. It can also be observed that the addition of the SpO2 signal to the ECG+Abdo ($M_2 3$) fusion does not lead to a very large increase in energy requirements due to the SpO2 ($M_1 2$) alone model being very small. The computational complexity of the fusion models can be reduced by investigating model pruning, which is discussed next.

1) Model Pruning: Pruning is the systematic removal of parameters from an existing network for reducing resource

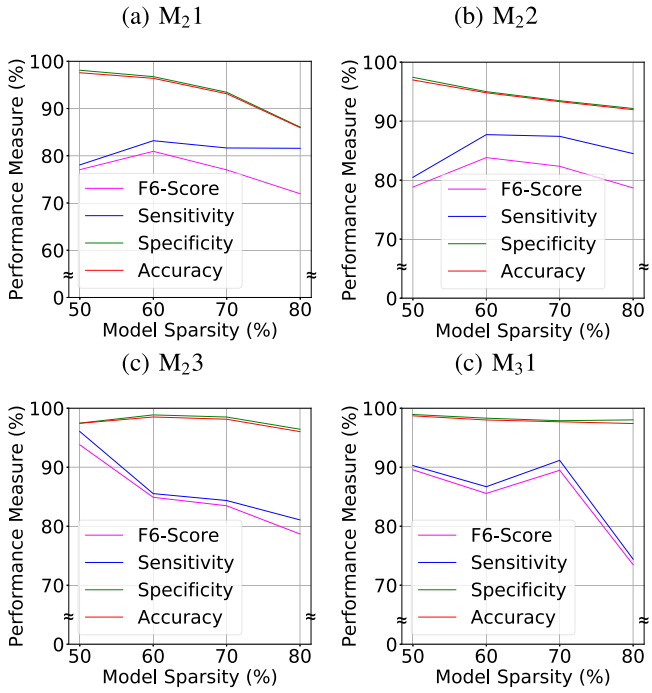

Fig. 8. Performance of the pruned models with sparsity varying from 50% to 80% when the (a) ECG and SpO2 signal is fused, (b) SpO2 and Abdomen movement signal is fused, (c) ECG and Abdomen movement signal is fused, and (d) SpO2, ECG, and Abdomen movement signal is fused.

TABLE IX
COMPUTATIONAL COMPLEXITY IN TERMS OF MULTIPLICATION AND ADDITION OPERATIONS AND THE ENERGY CONSUMPTION OF THE SPARSIFIED MODELS DURING PREDICTION

Models	Multiplications	Additions	Energy (μJ)
$M_2 1$ (50%)	3911656	3933763	1.53
$M_2 1$ (60%)	3131800	3147191	1.22
$M_2 1$ (70%)	2352472	2367863	0.92
$M_2 1$ (80%)	1572616	1587553	0.61
$M_2 2$ (50%)	3911656	3933763	1.53
$M_2 2$ (60%)	3131800	3147191	1.22
$M_2 2$ (70%)	2352472	2367863	0.92
$M_2 2$ (80%)	1572616	1587553	0.61
$M_2 3$ (50%)	6548632	6587126	2.55
$M_2 3$ (60%)	5243152	5268214	2.04
$M_2 3$ (70%)	3937672	3962734	1.54
$M_2 3$ (80%)	2632192	2656354	1.03
$M_3 1$ (50%)	7185972	7227326	2.80
$M_3 1$ (60%)	5753376	5781298	2.25
$M_3 1$ (70%)	4321308	4349230	1.68
$M_3 1$ (80%)	2888712	2915734	1.13

requirements at prediction time [47]. In this work, magnitude-based weight pruning is used, which gradually zero out model weights during the training process [48]. This enables compression of the model and is suitable for deployment in resource-constrained wearable devices. We consider the pruning of the fused models to reduce the computational complexity that results from the additional sensor sources used in fusion. For this, the developed fusion models are pruned with sparsity varying from 50% to 80% (aggressive pruning) using the weight pruning method. The performance of $M_2 1$ pruned models, $M_2 2$ pruned models, $M_2 3$ models, and $M_3 1$ models are as shown in Fig. 8(a), (b), (c), and (d) respectively, with model sparsity varying from

TABLE X
COMPARISON OF STATE-OF-THE-ART FUSION BASED SLEEP APNEA DETECTION ALGORITHMS WITH THE PERFORMANCE OF THE PROPOSED 1D-CNN BASED FUSION MODELS FOR APNEA DETECTION

Article	Otero *et al.* [33]	Bernardini *et al.* [32]	Xie *et al.* [18]	Cen *et al.* [27]	**This work**			
Dataset	Proprietary polysomnogram data	Proprietary stroke unit database	UCD St. Vincent's database	UCD St. Vincent's database	UCD St. Vincent's database			
Resolution	2 minutes	3 minutes	1 minute	1 second	1 second			
Method	Fuzzy structural algorithm	CNN+ LSTM	Bagging with RepTree	2D-CNN	1D-CNN fusion			
Signal	Respiratory airflow and SpO2	ECG and SpO2	ECG and SpO2	SpO2, oronasal airflow, ribcage, and Abdo	ECG and SpO2	SpO2 and Abdo	ECG and Abdo	ECG, SpO2, and Abdo
Accuracy	-	81.50%	84.40%	79.61%	99.62%	99.61%	98.69%	99.71%
Sensitivity	96.00%	67.20%	79.75%	-	97.44%	98.90%	98.90%	98.98%

50% to 80%. From the figures, it can be observed that for models $M_2 3$, and $M_3 1$, as sparsity increases, the F6-score and sensitivity drops while accuracy and specificity stay above approximately 97%. However, for models $M_2 1$ and $M_2 2$, the F6-score and sensitivity behave unpredictably. The F6-score and sensitivity increases when sparsity is increased from 50% to 60%, and drops again after sparsity is increased to 70%. It can be observed that both $M_2 1$ and $M_2 2$ used the SpO2 signal and this trend is similar to that observed for the individually sparsified SpO2 model as discussed in [38], as even though the overall performance drops (in terms of accuracy) with an increase in sparsity, the weights that make the network highly specific are pruned away with an increase in sparsity, making the network more sensitive. Therefore, this trend is the contribution of the SpO2 based signal model. A similar trend is observed in the model $M_3 1$ that fuses all three signals, but with F6-score and sensitivity increasing when sparsity is increased from 60% to 70%, and drops again at a sparsity of 80%. The computational complexity and energy consumption of the pruned models are discussed in Table IX.

From the performance figures and model complexities, it can be concluded that an increase in model sparsity may lead to a drop in performance of the fused model, but the pruned fusion model performance can outperform individual model performances at lower computational complexities, which is promising. However, in this work, we have explored only magnitude-based aggressive pruning (sparsity above 50%) and, therefore, the performances of the pruned models are not as compelling. Hence, further investigation into different pruning methodologies at sparsity levels varying from 5% to 80% needs to be explored to derive a compelling argument for the benefits of pruned fusion models over single signal-based models.

The performance of the proposed fused networks is compared with that of state-of-the-art fusion algorithms for sleep apnea detection in Table X. It can be seen that all the fusion models outperform the fusion models discussed in [33], [32], [18], and [27] in terms of accuracy and sensitivity on the UCD St. Vincent's sleep apnea database. However, direct comparison with [33] and [32] is not recommended due to the studies using a proprietary dataset. The performance improvements observed with the proposed model in comparison to state-of-the-art can be attributed to the proposed methodology of developing high resolution well-performing individual sensor models, which is different from the methodology where all the data is fed to a single model and trained together as proposed in [33], [32], [18], and [27]. Moreover, the windowing approach with an overlap of 90.9% when generating the signal windows for training helped with training deep data-hungry models which improved performance. The pruned fusion models are also found to outperform the state-of-the-art methods.

V. CONCLUSION

In this work, we propose a 1D-CNN model for the fusion of multimodal and multiresolution signals to capture temporal information like event detection, without having to resort to re-sampling of the individual signals. Here, we explore whether the fusion model can improve accuracy over that of using 1D-CNNs for each signal individually. We also proposed an experimental study in which selective dropout of the features obtained from the sensor with a larger sampling rate is used to prevent the fusion model from overfitting to features from the sensor with the larger sampling rate. In the proposed architecture, the 1D-CNN network learns the optimal parameters for fusion - a data-driven approach instead of a heuristic method for fusion. We apply this model to the sleep apnea detection problem using ECG and SpO2 signals with the signals sampled at different sampling rates. Prior literature in the field indicates that these two signals can detect apnea events with reasonably good accuracy. The performance of the fusion model and individual signal models were analyzed over a common test set to confirm the hypothesis that the fusion model would perform better. We also study the advantage of including the selective dropout in the training process over models that do not use selective dropout. We then extend the model to include a third sensor source to exhibit the generalizability to multi-sensor fusion problems using abdomen movement signals for sleep apnea detection. The performance of the three signal fusion models and various combinations of two signal fusion models were analyzed to exhibit the merits of the fusion algorithm. In the case of the model that fuses ECG and SpO2 signal, the F6-score was found to be 1.6% higher on the test set compared to the ECG signal-based model and 18.03% higher than the F6-score of the SpO2 based model. In the case of the model that fuses ECG and Abdomen movement signal, the F6-score was found to be 2.99% higher on the test set compared to the ECG signal-based model and 0.35% higher than the F6-score of the abdomen movement-based model. We

also analyzed pruning methods to compensate for the increase in computational complexity arising from additional sensor sources. It was observed that although an increase in model sparsity may lead to a drop in performance of the fused model, the pruned fusion model performance can outperform individual model performances at lower computational complexities- indicating that pruned fusion models are a promising direction for sleep apnea detection in wearable devices. The performances of the fusion models were compared against other fusion models in literature and the proposed models outperform the state-of-the-art models discussed in the literature. The model complexities were also analyzed and it was observed that the fusion model that combines all three sensor sources does not significantly outperform models that combine two sensor sources in the scenario where the signal samples are noise free. However, in the scenario where signals are noisy, the fusion model that combines all three signals significantly outperform the 2-signal fusion models, indicating that as the number of sensors used in fusion increase, the reliability of inferences also increases, indicating that fusion algorithms for the detection of sleep apnea in wearable devices is a promising direction. Future iterations of the fusion model would focus on multi-stage automatic fusion level selection through a data-driven approach which uses attention mechanism after each layer to generate the attention weights. Future works would also focus on advanced CNN algorithms and aim to reduce the fusion model complexity while maintaining the model performance by employing smart hybrid pruning methods and quantization prior to deployment on a wearable device.

REFERENCES

[1] B. V. Dasarathy, "Sensor fusion potential exploitation-innovative architectures and illustrative applications," *Proc. IEEE IRE*, vol. 85, no. 1, pp. 24–38, Jan. 1997.

[2] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE IRE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[3] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 3, pp. 535–544, Jun. 2020.

[4] J. Zhang and Y. Wu, "A new method for automatic sleep stage classification," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 5, pp. 1097–1110, Oct. 2017.

[5] M. Zanghieri, S. Benatti, A. Burrello, V. Kartsch, F. Conti, and L. Benini, "Robust real-time embedded EMG recognition framework using temporal convolutional networks on a multicore IoT processor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 244–256, Apr. 2020.

[6] D. Biswas *et al.*, "CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 2, pp. 282–291, Apr. 2019.

[7] M. S. Roy, B. Roy, R. Gupta, and K. D. Sharma, "On-device reliability assessment and prediction of missing photoplethysmographic data using deep neural networks," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 6, pp. 1323–1332, Dec. 2020.

[8] A. Jafari, A. Ganesan, C. S. K. Thalisetty, V. Sivasubramanian, T. Oates, and T. Mohsenin, "SensorNet: A scalable and low-power deep convolutional neural network for multimodal data classification," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 66, no. 1, pp. 274–287, Jan. 2019.

[9] F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol 2013, Sep. 2013, Art. no. 704504.

[10] A. John, S. J. Redmond, B. Cardiff, and D. John, "A multimodal data fusion technique for heartbeat detection in wearable IoT sensors," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2021.3093112.

[11] M. A. Al-Jarrah, M. A. Yaseen, A. Al-Dweik, O. A. Dobre, and E. Alsusa, "Decision fusion for IoT-based wireless sensor networks," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1313–1326, Feb. 2020.

[12] R. C. Luo and M. G. Kay, "A tutorial on multisensor integration and fusion," in *Proc. 16th Annu. Conf. IEEE Ind. Electron. Soc.*, 1990, vol. 1, pp. 707–722.

[13] H. F. Durrant-Whyte, "Sensor models and multisensor integration," *Int. J. Robot. Res.*, vol. 7, no. 6, pp. 97–113, 1988.

[14] V. Vielzeuf, A. Lechervy, S. Pateux, and F. Jurie, "Multilevel sensor fusion with deep learning," *IEEE Sensors Lett.*, vol. 3, no. 1, pp. 1–4, Jan. 2019.

[15] T. Kim and J. Ghosh, "On single source robustness in deep fusion models," in *Proc. Adv. Neural Inf. Process. Syst. 32: Annu. Conf. Neural Inf. Process. Syst.*, 2019, pp. 4815–4826.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[17] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults." *Amer. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, May 2013.

[18] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 469–477, May 2012.

[19] U. Satija, B. Ramkumar, and M. S. Manikandan, "Real-time signal quality-aware ECG telemetry system for IoT-based health care monitoring," *IEEE Internet Things J.*, vol. 4, no. 3, pp. 815–823, Jun. 2017.

[20] M.-C. Chambrin, "Alarms in the intensive care unit: How can the number of false alarms be reduced?" *Crit. Care*, vol. 5, no. 4, pp. 184–188, 2001.

[21] D. L. T. Wong *et al.*, "An integrated wearable wireless vital signs biosensor for continuous inpatient monitoring," *IEEE Sensors J.*, vol. 20, no. 1, pp. 448–462, Jan. 2020.

[22] R. Atri and M. Mohebbi, "Obstructive sleep apnea detection using spectrum and bispectrum analysis of single-lead ECG signal," *Physiol. Meas.*, vol. 36, no. 9, pp. 1963–1980, 2015.

[23] B. Sulistyo, N. Surantha, and S. M. Isa, "Sleep apnea identification using HRV features of ECG signals," *Int. J. Elect. Comput. Eng.*, vol. 8, pp. 3940–3948, 2018.

[24] S. S. Mostafa, F. Mendonça, F. Morgado-Dias, and A. Ravelo-García, "SpO2 based sleep apnea detection using deep learning," in *Proc. IEEE 21st Int. Conf. Intell. Eng. Syst.*, 2017, pp. 91–96.

[25] S. S. Mostafa, F. Mendonça, A. G. Ravelo-García, and F. Morgado-Dias, "A systematic review of detecting sleep apnea using deep learning," *Sensors*, vol. 19, no. 22, 2019, Art. no. 4934.

[26] E. Urtnasan, Y.-J. Kim, J.-U. Park, E.-Y. Joo, and K.-J. Lee, "Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram," *Comput. Methods Prog. Biomed.*, vol. 180, 2019, Art. no. 105001.

[27] L. Cen, Z. L. Yu, T. Kluge, and W. Ser, "Automatic system for obstructive sleep apnea events detection using convolutional neural network," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 3975–3978.

[28] L. Almazaydeh, K. Elleithy, and M. Faezipour, "Detection of obstructive sleep apnea through ECG signal features," in *Proc. IEEE Int. Conf. Electro/Inf. Technol.*, 2012, pp. 1–6.

[29] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, and S. V. Huffel, "A novel algorithm for the automatic detection of sleep apnea from single-lead ECG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2269–2278, Sep. 2015.

[30] H. D. Nguyen, B. A. Wilkins, Q. Cheng, and B. A. Benjamin, "An online sleep apnea detection method based on recurrence quantification analysis," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1285–1293, Jul. 2014.

[31] X. Liang, X. Qiao, and Y. Li, "Obstructive sleep apnea detection using combination of CNN and LSTM techniques," in *Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf.*, 2019, pp. 1733–1736.

[32] A. Bernardini, A. Brunello, G. L. Gigli, A. Montanari, and N. Saccomanno, "AIOSA: An approach to the automatic identification of obstructive sleep apnea events based on deep learning," *Artif. Intell. Med.*, vol. 118, 2021, Art. no. 102133. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365721001263

[33] A. Otero, P. Felix, M. R. Alvarez, and C. Zamarron, "Fuzzy structural algorithms to identify and characterize apnea and hypopnea episodes," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2008, pp. 5242–5245.

[34] P. M. Macey, "Damage to the hippocampus in obstructive sleep apnea: A link no longer missing," *Sleep*, vol. 42, no. 1, 2019, Art. no. zsy266. [Online]. Available: https://doi.org/10.1093/sleep/zsy266

[35] M. J. Morrell *et al.*, "Changes in brain morphology in patients with obstructive sleep apnoea," *Thorax*, vol. 65, no. 10, pp. 908–914, 2010. [Online]. Available: https://thorax.bmj.com/content/65/10/908

[36] "St Vincents University Hospital/ University College Dublin sleep apnea database," 2011. [Online]. Available: http://physionet.org/physiobank/database/ucddb/
[37] A. John, B. Cardiff, and D. John, "A 1D-CNN based deep learning technique for sleep apnea detection in IoT sensors," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2021, pp. 1–5.
[38] A. John, K. K. Nundy, B. Cardiff, and D. John, "SomnNET: An SpO2 based deep learning network for sleep apnea detection in smartwatches," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2021, pp. 1961–1964
[39] Y. Lin, H. Wu, C. Hsu, P. Huang, Y. Huang, and Y. Lo, "Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1533–1545, Nov. 2017.
[40] C. Avcı and A. Akbaş, "Sleep apnea classification based on respiration signals by using ensemble methods," *Bio-Med. Mater. Eng.*, vol. 26, pp. S 1703–S 1710, 2015.
[41] H.-C. Chang, H.-T. Wu, P.-C. Huang, H.-P. Ma, Y.-L. Lo, and Y.-H. Huang, "Portable sleep apnea syndrome screening and event detection using long short-term memory recurrent neural network," *Sensors*, vol. 20, no. 21, 2020, Art. no. 6067.
[42] A. John, R. C. Panicker, B. Cardiff, Y. Lian, and D. John, "Binary classifiers for data integrity detection in wearable IoT edge devices," *IEEE Open J. Circuits Syst.*, vol. 1, pp. 88–99, 2020.
[43] K. Abdelouahab, M. Pelcat, J. Serot, and F. Berry, "Accelerating CNN inference on FPGAs: A Survey," 2018, *arXiv:1806.01683*.
[44] H. Reyserhove, N. Reynders, and W. Dehaene, "Ultra-low voltage datapath blocks in 28 nm UTBB FD-SOI," in *Proc. IEEE Asian Solid-State Circuits Conf.*, 2014, pp. 49–52.
[45] R. Taco, I. Levi, M. Lanuzza, and A. Fish, "An 88-fJ/40-MHz [0.4 v]-0.61-pJ/1-GHz [0.9 v] dual-mode logic 8×8 bit multiplier accumulator with a self-adjustment mechanism in 28-nm FD-SOI," *IEEE J. Solid-State Circuits*, vol. 54, no. 2, pp. 560–568, Feb. 2019.
[46] R. Taco, I. Levi, M. Lanuzza, and A. Fish, "Evaluation of dual mode logic in 28 nm FD-SOI technology," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2017, pp. 1–4.
[47] G. Castellano, A. Fanelli, and M. Pelillo, "An iterative pruning algorithm for feedforward neural networks," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 519–531, May 1997.
[48] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Guttag, "What is the state of neural network pruning?" 2020, *arXiv:2003.03033*.

Arlene John (Student Member, IEEE) received the Bachelor of Technology degree in electrical and electronics engineering from the National Institute of Technology, Calicut, India, in 2017. She is currently working toward the Ph.D. degree with the School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland. In the summer of 2016, she was a Research Intern with the Indian Institute of Science, Bangalore, India. During June 2017–June 2018, she was with Bosch India Ltd., as a Project Manager and in engineering and strategy development for hybrid electric vehicles. During March–June, 2019, she was a Senior Visiting Researcher with the Beijing University of Technology, Beijing, China. Her research interests include multisensor data fusion, biomedical signal processing, and machine learning. Her research focuses on the development of data fusion frameworks for wearable health monitoring devices. In 2020, she was a finalist in the young female STEM pioneer category of the Diversity in tech awards. She was the recipient of the prestigious University College Dublin President's Award 2021.

Koushik Kumar Nundy (Senior Member, IEEE) received the Bachelor of Technology degree in electronics and communication engineering from the National Institute of Technology, Durgapur, India, in 2010, and the Master of Science degree in telecommunication from the Hong Kong University of Science and Technology, Hong Kong, in 2011. He is currently the CTO and a Co-Founder of Think Biosolution Limited, a medical devices and healthcare IT company in Dublin, Ireland. He was with Multiple Research and Academic Institutes, including the National University of Singapore and the Indian Institute of Science, and has presented his work in multiple conferences and journals. He was also with the R&D Team, Altai Technologies Ltd, a pioneer in wireless communication systems. He is the Ireland Area Chair for the IEEE Young Professionals. He has won multiple awards, including the Roche Unicorn Champion 2020, Innovation of the Year 2017, Start-up Weekend 2015, and has been featured in publications, such as the Irish Times, Rochester Business Journal, and Silicon Republic. He was the recipient of the NTSE Scholarship in 2004 and NGS scholarship in 2013.

Barry Cardiff (Senior Member, IEEE) received the B.Eng., M.Eng. Sc., and Ph.D. degrees in electronic engineering from University College Dublin, Dublin, Ireland, in 1992, 1995, and 2011, respectively. From 1993 to 2001, he was a Senior Design Engineer or Systems Architect for Nokia, moving to Silicon & Software Systems (S3 group) thereafter as a Systems Architect in their R&D division, focused on wireless communications and digitally assisted circuit design. Since 2013, he has been an Assistant Professor with University College Dublin. He holds several U.S. patents related to wireless communication. His research interests include digitally assisted circuit design and signal processing for wireless and optical communication systems.

Deepu John (Senior Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the University of Kerala, Thiruvananthapuram, India, in 2002, and the M.Sc. and Ph.D. degrees in electrical engineering from National University Singapore, Singapore, in 2008 and 2014, respectively. He is currently an Assistant Professor with the School of Electrical and Electronics Engineering, University College Dublin, Dublin, Ireland. From 2014 to 2017, he was a Postdoctoral Researcher with Bio-Electronics Lab, National University Singapore. Previously, he was a Senior Engineer with Sanyo Semiconductors, Gifu, Japan. He is the recipient of the Institution of Engineers Singapore Prestigious Engineering Achievement Award in 2011, the Best Design Award at the Asian Solid-State Circuit Conference in 2013, and the IEEE Young Professionals, Region 10 Individual Award in 2013. He was a Member of the organizing committee or technical program committee for several IEEE conferences, including TENCON, ASICON, ISCAS, BioCAS, and ICTA. He was a Guest Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I and IEEE OPEN JOURNAL OF CIRCUITS AND SYSTEMS. He is an Associate Editor for IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS and *Wiley International Journal of Circuit Theory and Applications*. His research interests include low-power biomedical circuit design, energy-efficient signal processing, and edge computing.