IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 February 2023, accepted 13 March 2023, date of publication 22 March 2023, date of current version 27 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3260215

RESEARCH ARTICLE

Distributed Neural Network System for Multimodal Sleep Stage Detection

YI-HSUAN CHENG, MARGARET LECH, (Member, IEEE), AND RICHARDT H. WILKINSON, (Senior Member, IEEE)
School of Engineering, RMIT University, Melbourne, VIC 3000, Australia

Corresponding author: Margaret Lech (margaret.lech@rmit.edu.au)

This work was supported in part by the Australian Government Research Training Program Scholarship, in part by the Engineering Top-Up Scholarship, and in part by the Royal Melbourne Institute of Technology (RMIT) Research Stipend.

ABSTRACT Existing automatic sleep stage detection methods predominantly use convolutional neural network classifiers (CNNs) trained on features extracted from single-modality signals such as electroencephalograms (EEG). On the other hand, multimodal approaches propose very complexly stacked network structures with multiple CNN branches merged by a fully connected layer. It leads to very high computational and data requirements. This study proposes replacing a stacked network with a distributed neural network system for multimodal sleep stage detection. It has relatively low computational and training data requirements while providing highly competitive results. The proposed multimodal classification and decision-making system (MM-DMS) method applies a fully connected shallow neural network, arbitrating between classification outcomes given by an assembly of independent convolutional neural networks (CNNs), each using a different single-modality signal. Experiments conducted on the CAP Sleep Database data, including the EEG-, ECG-, and EMG modalities representing six stages of sleep, show that the MM-DMS significantly outperforms each single-modality CNN. The fully-connected shallow network arbitration included in the MM-DMS outperforms the traditional majority voting-, average probability-, and maximum probability decision-making methods.

INDEX TERMS Machine learning, distributed networks, multimodal classification, sleep stage detection, decision-making networks, transfer learning.

I. INTRODUCTION

Sleep is a cyclic process. It progresses periodically through five stages: wake, light sleep, deep sleep, and rapid eye movement (REM). Each stage is characterized by a different manifestation of brain wave activity. These characteristics can be observed in the shape of the electroencephalogram (EEG) time waveforms recorded for a patient's diagnosis. Indications of different sleep stages can also be observed in other signals such as electrocardiograms (ECG), or electromyograms (EMG) recorded simultaneously to support the diagnostic decision. Traditional sleep diagnosis involves manual inspection of lengthy time waveforms of the diagnostic recordings captured over a few hours of sleep. Highly trained experts analyze features occurring at different stages of sleep that may be associated with specific sleep pathologies. This process is costly and time-consuming. The recent advent of modern machine learning technologies makes the possibility of automatic sleep scoring more realistic. While many machine learning techniques that recognize EEG patterns have been proposed, multimodal approaches combining different types of signals have not been widely reported. Although manual sleep scoring techniques show a clear advantage of using different modalities, mainly when the EEG exhibits some ambiguity, an apparent arbitration can only be achieved by examining the behavior of other simultaneously recorded signals. Depending on the design, a multimodal diagnostic system highlights issues related to the representation and modeling of different data modalities and the arbitration between diagnostic outcomes generated

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

by other modalities. Regarding the first problem, different modalities can be processed through parallel channels, and the fusion can occur at the final decision-making level. Each processing channel can use its independent data representation and modeling architectures in this case. Alternatively, the data can be fused at the feature extraction stage and passed to a single model making the final diagnosis. Both approaches have been investigated in the past and applied to medical diagnosis [1]. However, automatic sleep scoring still remains an open research area.

A. PAPER CONTRIBUTIONS

In this study, a multimodal approach to sleep stage detection is investigated. The paper offers the following contributions:

- A new multimodal classification and decision-making system (MM-DMS) is proposed that is comprised of distributed neural network modules. The diagnostic information is first derived separately for each modality by a parallel set of convolutional neural networks (CNNs) and then passed to a single feed-forward neural network to make the final decision. The majority of existing methods train a single CNN structure with either an early or late fusion of multimodal information. In the proposed case, each module can be trained (or retrained) separately with reduced time- and data requirements compared to single-model approaches.
- The use of RGB images of logarithmic amplitude spectrograms as a uniform modality-independent data representation of EEG-, ECG-, and EMG signals is proposed. Existing methods tend to apply different modality-dependent preprocessing and feature extraction techniques.
- Unlike other methods, using multimodal sensor data fused into a common feature array (early fusion) or a common fully connected layer (late fusion), the proposed approach applies a fusion of classification outcomes (soft probability vectors) from multiple single-modality modules passed to a separate shallow neural network (NN) trained to arbitrate between single-modality classification outcomes and make the final decision. The shallow NN decision-making process is compared against classical maximum probability, average probability, and majority voting methods.
- The proposed system was tested using the CAP Sleep Database data [2], [3], including the EEG-, ECG-, and EMG modalities representing the six stages of sleep.
- Finally, the achieved results are compared against related approaches reported in the literature.

B. PAPER STRUCTURE

The remaining parts of the paper are organized as follows: Section II provides a literature review of related works; Section III explains the proposed methodology; Section IV provides experimental validation of the proposed approach; Section V includes a discussion of the results presented, and the paper is concluded in Section VI.

II. RELATED WORKS

The beginning of the 21st Century has been marked by a rapid acceleration of machine learning (ML) techniques in various medical and biomedical applications. Machine learning methods are now commonly used in medical diagnosis, prediction, data labeling, and analysis. Sleep scoring is extensively researched to automate this process. It includes scoring based on EEG and other modalities – on their own or in combination. Two significant factors affect sleep scoring performance, i.e., the feature extraction methods and the classifier architectures used. Both have received a great deal of research attention in recent years. One of the early applications of ML in EEG classification was reported by [4]. A multi-class Support Vector Machine (SVM) model was trained to identify two different sleep stages. An average classification accuracy of 70.92% was reported. Deep Neural Networks (DNN) brought a wide range of opportunities for further improvements, which resulted in Convolutional Neural Networks (CNNs) quickly becoming the dominant technique for biomedical data classification. The concept of transfer learning with publicly available pretrained CNNs created an opportunity to achieve high-accuracy models at a relatively low computational cost and modest data requirements. It opened the possibility of conducting medical diagnoses based on single-modality data and combined multimodal information. Deep learning architectures proposed for multimodal classification included stacked deep neural networks containing separate branches for each modality and connected at the final decision-making layer.

Three data modalities (EEG, EOG, and EMG) from the Montreal Archive of Sleep Studies (MASS) dataset [5] were used to detect five sleep stages (W, N1, N2, N3, and R) [6]. The classification model was constructed as a stacked neural network containing two parallel CNN branches. The EEG- and EOG time waveforms were passed to the first branch, and the EMG signal to the second branch. The CNN branches were connected by a common SoftMax layer generating a single output label identifying the sleep stage. The classification accuracy reported was around 80%. In [7], the same three modalities were applied to detect five sleep stages. However, in this case, instead of depicting the time waveforms, the data was transformed into amplitude spectrograms using a linear frequency scale. A stacked deep neural network architecture included two separate CNN branches, one using EEG-, and the other using EOG spectrograms. The outputs from the CNN branches were concatenated with the EMG spectrograms and passed through Long Short-Term Memory (LSTM) layers to generate the final label. The results for healthy participants were reported as being in the low 80% range but dropped by 20% for clinical patients. In [8], a classifier architecture based on the time attention mechanism was investigated. It was applied to categorize time- and frequency domain feature parameters extracted from a few modalities of polysomnographs. A sleep stage detection accuracy of 90%-92% was reported for five stages.

Alongside classifier architectures, the research investigated different ways of feature extraction from multimodal data. In [9], preprocessed time waveforms of EEG-, EMG-, and EOG signals from the Sleep Heart Health Study (SHHS) database [10], [11] were applied as inputs to a multi-channel CNN. The classification of five sleep stages resulted in an F_1-score of 76%. The SHHS database was used to detect five sleep stages in [12]. CNN embeddings of amplitude spectrogram features were generated and passed on to an LSTM classifier. In [13], the Physionet Sleep-EDFX [3], [14], and the MASS databases were used to detect five sleep stages. Each of the three data modalities, i.e., EEG, EMG, and EOG, was represented as linear spectrograms and passed to the CNN classifier. In [15], the Physionet Sleep-EDF [3], [14] and Physionet Sleep-EDFX databases were applied to detect six sleep stages using the EEG and EOG modalities. Separate CNNs were trained for each modality. The input time waveforms were split into segments, and each segment was passed to the CNN as a one-dimensional vector. The prediction accuracy achieved reached 91%. A similar approach was reported by [16] using the Physionet Sleep-EDF database with five sleep stages represented by EEG- and EOG modalities. In this case, a precision of 72% was achieved. In [17], a complex feature extraction and classification system was constructed. The features, extracted from spectrograms using a bidirectional recurrent neural network (BRNN), and from time waveforms using convolutional layers, were passed to the secondary BRNN to determine the final label. This approach was tested using the network classifiers used in the MASS database to detect five sleep stages simultaneously using three modalities, i.e., EEG, EMG, and EOG. The accuracy was reported as 86%-87%. The Physionet Sleep-EDF and Physionet Sleep-EDFx datasets were used in [18] to detect five sleep stages. A new method for fusing two sources of information, including EEG and EOG, was proposed. Features extracted from the EEG- and EOG waveforms were divided into two feature sets: the EEG features and fused features of EEG- and EOG data. Each feature set was transformed into images representing horizontal visibility graphs (HVGs). These images were used to train a CNN classifier. This algorithm achieved an accuracy of 93.58%. Clinical data from 33 healthy participants and 25 sleep disorder patients represented by two modalities, ECG and EMG, was used in [19]. The EEG features included entropy, statistical moments, and the synchrosqueezed wavelet transform (SSWT) coefficients. At the same time, the heart rate- and breathing-related features were extracted from the ECG data. An RNN was used to classify the EMG data, and a fully connected artificial neural network (ANN) was used to classify the ECG data. The outputs from both networks were passed to the final ANN to determine the sleep stage.

In [20], EEG-, EMG-, EOG-, and ECG features were extracted from the convolutional layers of the CNN. Integration blocks were added to the network structure to merge the modalities. The method was tested on the SHHS and Physionet Sleep-EDF databases to detect five sleep stages. Synchronized polysomnogram (PSG) and ECG recordings from 1743 participants were utilised to detect five sleep stages in [21]. The signal time waveforms were passed to an assembly of separate LSTM and CNN classifiers. The final label was derived by calculating either maximum- or mean output from these classifiers. An accuracy of 78.2% and an F_1-score of 69.8% were reported on the classification task for three sleep stages.

As shown in this brief review, sleep stage detection research is dominated by the design of complex stacked classification network architectures, with multimodal fusion occurring most often at the final decision levels. On the other hand, feature extraction tends towards parameters generated by convolutional layers of CNNs. However, there is no explicit agreement regarding the network input format for different modalities. Many studies use either time waveform segments or spectrograms. The extraction of engineered features appears to be a gradually declining trend.

III. METHODOLOGY

A. MULTIMODAL CLASSIFICATION AND DECISION-MAKING SYSTEM (MM-DMS)

As shown in Fig. 1, the MM-DMS system is a set of four interconnected neural network classifiers: EEG CNN, ECG CNN, EMG CNN, and Decision-Making NN. The final sleep stage label is derived based on the information flow through the system and algebraic connections between the component networks. At the first stage of the classification procedure, three parallel classifiers – EEG CNN, ECG CNN, and EMG CNN – are assembled, each making its own independent decision based on a single synchronized data modality (EEG, ECG or EMG). These three channels work as three independent assessors directly analyzing the physical data coming from the sensors. Since the assessment outcomes are likely to vary between these three assessors, a fourth decision-making network is employed to arbitrate between the assessors at the second stage of the classification process. The fourth network (Decision-Making NN) is not analyzing the physical data but the patterns of outcomes given by each first-stage assessor to derive the final label.

B. ADVANTAGES OF MM-DMS

One of the first questions often coming to mind when looking at the MM-DMS block diagram is how this method differs from a stacked network containing three CNN branches connected by a fully connected layer. Such structures have been previously proposed as multimodal classifiers [6], [7]. The stacked network is a highly complex and rigid structure. It is trained entirely on the sensor data. The training process is time-consuming, and a large amount of labeled- and balanced training data is required to achieve excellent performance. Any updates based on the newly collected data require the retraining of the whole network structure. On the contrary, the proposed MM-DMS consists of smaller

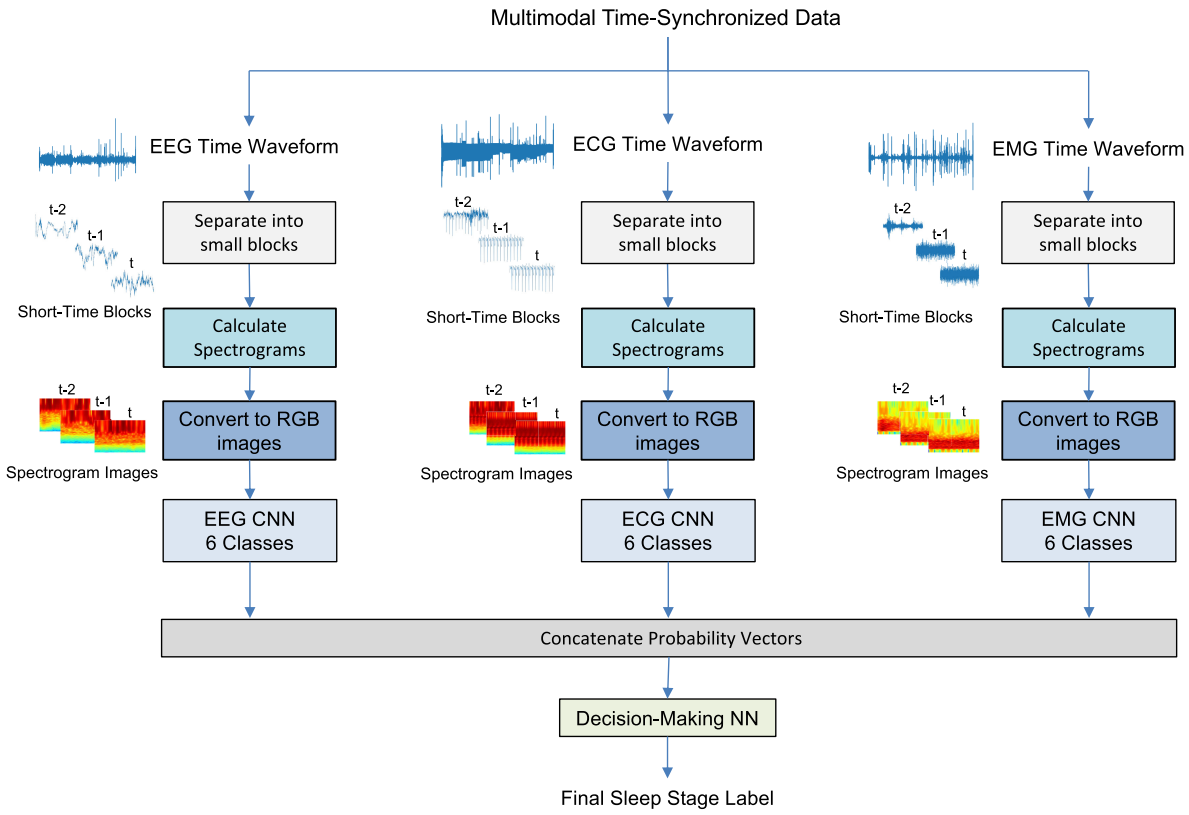


FIGURE 1. Block diagram of the proposed MM-DMS method.

independent classification units, with each unit trained independently on a relatively small amount of data. The first-level classification units are trained on the physical data, whereas the final second-level decision-making unit is trained on the probability vectors generated by the first-level classifiers. The sleep-stage recognition task is distributed between component units, and each unit provides its own contributions. If some of the first-level classification units do not perform well (e.g., due to a small training set size), the final arbitrating unit learns how to detect the classification errors, and compensate for them. Each unit can be retrained independently in a data- and time-efficient manner. Trained component units can also be shared between different classification systems designed to perform tasks other than sleep stage recognition. For example, the first-level sleep stage recognition units can be re-employed to provide supporting information in the process of detecting sleep disorders. Moreover, additional first-stage classifiers (using different modalities) can easily be added to the system to support the diagnoses without the need to retrain all system components. Such flexibility is not achievable with stacked network structures.

C. PRE-PROCESSING OF MULTIMODAL DATA

The pre-processing steps in Fig. 1 were performed consistently across all three data modalities (EEG, ECG, and EMG) in a time-synchronized way. Pre-recorded (or streamed in real-time) waveforms were divided into short-duration blocks to conduct block-by-block processing. The duration of each block was set to 10 seconds for all three modalities. Experiments comparing a range of different durations showed that 10 seconds led to the highest classification accuracy. A short stride of 1 second was applied between subsequent blocks, resulting in a 90% overlap between blocks. Using a short stride allowed us to generate a relatively large number of training data samples, thus ensuring adequate training of the classifiers. The amplitude levels were normalized within a range of -1 to 1. Since the recordings were labeled sample-by-sample, it was assumed that the label for a given signal block was the same as the corresponding data sample from which it was derived. A 2D amplitude spectrogram array was subsequently computed for each block.

D. CALCULATION OF AMPLITUDE SPECTROGRAMS

For all three modalities (EEG, ECG, and EMG), the raw time waveforms were sampled at 512 Hz, resulting in a 256 Hz signal bandwidth. This bandwidth was preserved when transforming the time waveforms into spectrograms. For each 10-second block of data, an amplitude spectrogram array was computed using the Short-Time Fourier Transform (STFT). For the purpose of comparison, spectrograms with two different types of frequency scales were tested, linear

and logarithmic. The time axis used in both cases was linear. Since the CNN models used in this study show excellent performance in image classification; therefore, any type of transformation that generates image-like representations can be expected to perform well. The signals were represented as amplitude spectrograms, as it is the dominant signal representation for real-time signal processing (speech, sound, EEG, ECG, and similar signals) with readily available software and hardware application platforms.

E. CONVERSION INTO RGB IMAGES

Spectral amplitude arrays were converted into RGB color images to create suitable inputs to the CNN classifiers. This conversion was done using the "jet" colormap. The dynamic range of the spectral amplitudes was normalized across the entire dataset with respect to the average maximum-, and minimum amplitude values for a given modality [22], [23]. Since the normalization was performed separately for EEG-, ECG-, and EMG samples, the dynamic range was different for each modality. For EEG, it was Min $= -0.0018$dB, Max $= 0.0019$dB, for ECG, Min $= -0.00083$dB, Max $= 0.00084$dB, and for EMG, Min $= -0.0047$dB, Max $= 0.0046$dB. As described in Section IV, the spectrogram-based classification for linear- and logarithmic scales was compared with classification performed using 10-second time waveform blocks as direct inputs to the classifier.

Fig. 2 shows examples of 10-second blocks of time waveforms and the corresponding RGB images of the linear- and logarithmic spectrograms for the EEG-, ECG-, and EMG signals. It can be observed that the images of logarithmic spectrograms show more details of the low-frequency part of the signal bandwidth compared to the linear spectrograms. It was anticipated that the presented experiments would reveal whether the choice of linear- vs logarithmic scale has an effect on the classification performance.

F. TRAINING CNN MODELS

Since all modalities were represented as RGB images, one of the existing CNN architectures designed for general image classification tasks could be adapted. A separate CNN model was trained for each modality (EEG CNN, ECG CNN, and EMG CNN). Each model learned to categorize six sleep stages independently: wake (W), light-to-deep sleep (S1, S2, S3, S4), and rapid eye movement (R). The CNN structure known as VGG16 [24], [25] was adapted and trained from scratch. It consisted of thirteen 2D convolutional layers and three fully connected layers, as shown in Fig. 3. The same CNN hyperparameters, listed in Table 1, were used for all three modalities. The dataset was split into two subsets: training (80%) and testing (20%). These subsets were mutually exclusive. The training and testing procedure was repeated three times, each time with different training and testing subsets. The reported results were calculated and averaged over three runs. The models were implemented in Python using the TensorFlow Keras library [26]. The hyperparameters in Table 1 were chosen experimentally; no automatic optimization was used.

TABLE 1. Hyperparameters for the VGG16 CNNs and the Shallow NNs.

Parameters	CNN (VGG16 trained from scratch)	Decision-Making Shallow NN (not pretrained)	Decision-Making Shallow NN pretrained
Optimization	SGD[a]	SGD[a]	SGD[a]
Initial learning rate	0.001	0.001	0.001
Batch size	10	3	3
Maximum epochs	100	10	10
Early Stopping	Yes	Yes	Yes

[a] SGD denotes Stochastic Gradient Descent.

G. CONCATENATION OF PROBABILITY VECTORS

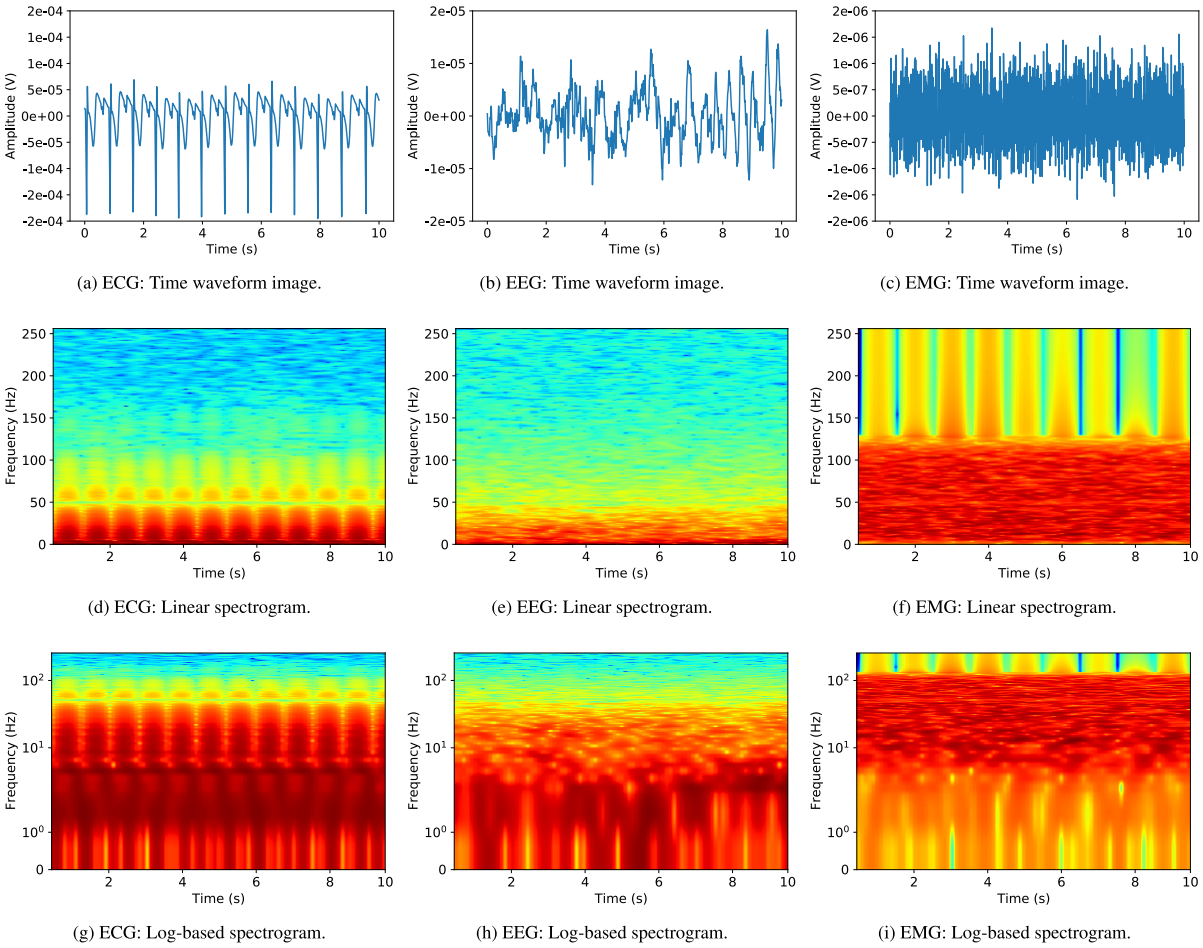
While the independent CNNs were trained directly on the sensor data converted to RGB images, the final decision-making neural network (NN) was trained on concatenated probability vectors generated by the CNNs. The CNNs were acting as parallel channels or an assembly of independent assessors.

Given six data categories, M independent assessors, and N images, the probability vector generated by the j^{th} assessor ($j = 1, \ldots, M$) for image i ($i = 1, \ldots, N$) was $P_{i,j} = [p_{i,j,1}, p_{i,j,2}, p_{i,j,3}, p_{i,j,4}, p_{i,j,5}, p_{i,j,6}]$. In the presented case, M was set equal to 3. Therefore, the concatenated probability vectors C_i are given as:

$$C_i = [p_{i,1,1}, \ldots, p_{i,1,6}, p_{i,2,1}, \ldots, p_{i,2,6}, p_{i,3,1}, \ldots, p_{i,3,6}]. \quad (1)$$

The concatenated probability vectors and the corresponding "ground truth" data labels were passed to the decision-making NN. It was trained to provide the final sleep stage categorization label. The probability merging process required having the same numbers of representative images for each modality. Since the available data contained different numbers of spectrogram images for different modalities (see Table 2), to make the numbers even, 500 images per modality were randomly selected for the purpose of training the decision-making network. The training was repeated three times and average values of the performance parameters were calculated.

Note that the decision-making NN is trained separately to the first-stage single-modality CNNs, using different features (concatenated probability vectors) but using the same "ground truth" labels. The NN effectively learns how to weigh the first-level classifiers based on their response patterns. The MM-DMS assumes some level of training for



(a) ECG: Time waveform image. (b) EEG: Time waveform image. (c) EMG: Time waveform image.

(d) ECG: Linear spectrogram. (e) EEG: Linear spectrogram. (f) EMG: Linear spectrogram.

(g) ECG: Log-based spectrogram. (h) EEG: Log-based spectrogram. (i) EMG: Log-based spectrogram.

FIGURE 2. Examples of time waveforms (a,b,c), RGB images of linear spectrograms (d,e,f), and RGB images of logarithmic spectrograms (g,h,i) for EEG-, ECG- and EMG modalities.

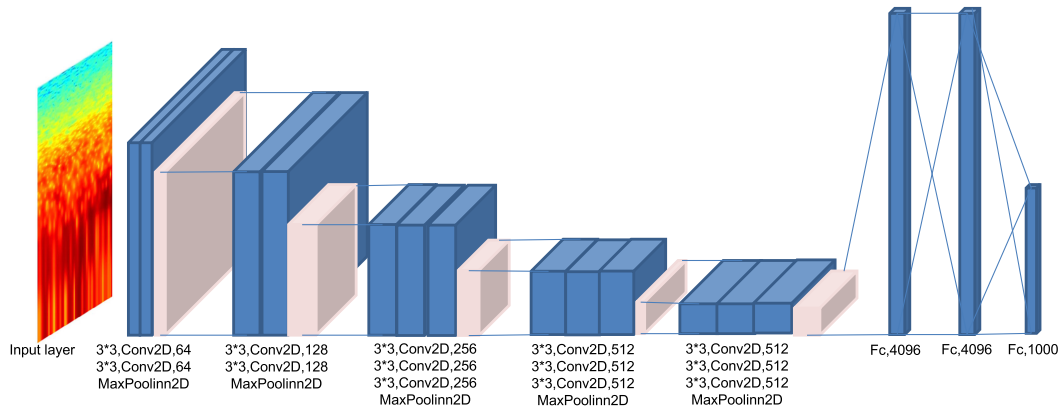


FIGURE 3. Structure of the VGG16 model.

the first level (single-modality) classifiers, but the training does not need to be perfect. In ideal cases where the data is balanced and all single-modality classifiers are perfectly trained, the outcomes from all classifiers should be the same and correct, thus eliminating the need for the NN. In reality, this never happens. Different modality classifiers often give different predictions based on the same input data instance. In such cases, an arbitration mechanism is needed to determine each modality's weight when making the multimodal decision. The research presented in this paper

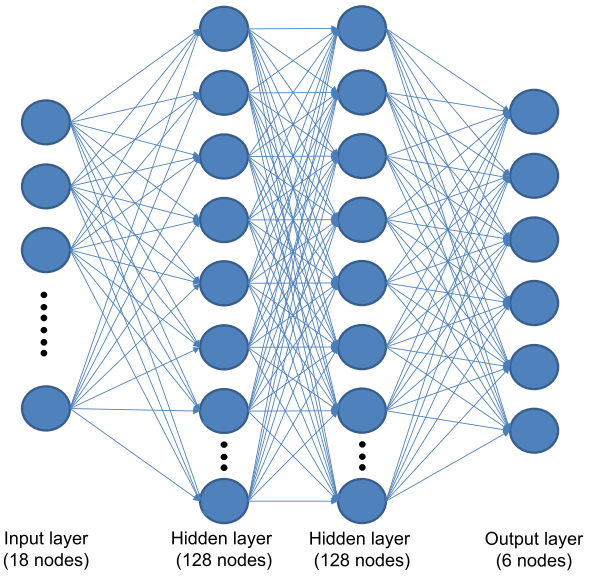


FIGURE 4. Structure of the decision-making shallow neural network.

shows that the shallow decision-making NN effectively learns how to weigh the outcomes from different single-modality classifiers to achieve optimal predictions. Most importantly, it is done without making any arbitrary selections or assumptions regarding the probability outcomes given by the single-modality classifiers.

H. DECISION-MAKING NEURAL NETWORK

The final decision-making mechanism in the proposed MM-DMS system was based on the shallow, fully connected multi-layer neural network. It consisted of an input layer containing 18 nodes, two hidden layers, each with 128 nodes, and the output layer with six nodes (Fig. 4). The Rectified Linear Unit (ReLu) function was applied to the activations from the input and hidden layers, and the SoftMax function to the activations from the output layer. To enhance the performance, the decision-making network was pretrained to recognise three sleep stages (W, S, and R) from EEG spectrograms using a single-modality decision-making system (SM-DMS). Details of the pretraining process can be found in [23]. The pretrained NN was then fine-tuned to identify six sleep stages. The fine-tuning parameters are listed in Table 1.

I. ALTERNATIVE DECISION-MAKING METHODS

To cross-validate the performance of the proposed decision-making NN, three other frequently used decision-making methods were tested, i.e., maximum probability, majority voting, and average probability. In the maximum probability method, the final label was assigned to the label indicated by the largest probability value across all three assessors. The majority voting approach would evaluate the categories suggested by each of the assessor CNNs and decide based on the category that achieved the highest vote. When all three assessors disagreed, the maximum probability criterion was used. The average probability method would average the voting provided by all three assessors for all categories and choose the category that scored the highest.

J. PERFORMANCE MEASURES

The performance of the MM-DMS system was assessed using standard measures, including the classification accuracy, F_1 score, and confusion matrices. Given the number of true positive (TP), true negative (TN), false-positive (FP), and false-negative (FN) classification outcomes, the accuracy A_{cc} was calculated using:

$$A_{\mathrm{cc}} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{2}$$

Due to the unbalanced numbers of data representing the classified categories of sleep, the F_1 score was calculated as,

$$F_1 = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}, \tag{3}$$

where $Recall$ was calculated as,

$$Recall = \frac{TP}{TP + FN}, \tag{4}$$

and, the $Precision$ parameter was given by,

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

Since the F_1 score can be viewed as the weighted average of recall and precision, an F_1 score closer to 0 indicates that there is still room to improve the model. However, when the F_1 score is close to 1, the trained model cannot achieve a higher performance.

IV. EXPERIMENTS AND RESULTS

A. DATABASE DESCRIPTION

The sleep stage classification methodology was tested using publicly available data from the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy, known as the CAP Sleep Database v1.0.0 - PhysioNet [2], [3]. It contains an extensive collection of multimodal recordings representing normal and pathological sleep conditions. This study only used the time-synchronized EEG-, ECG-, and EMG waveforms representing normal healthy sleep labeled with six sleep stage categories: wake (W), sleep sub-stages (S1, S2, S3, S4), and rapid eye movement (R). Each modality was represented by samples captured by groups of sensors. In the case of EEG, the recordings included signals recorded from 16 electrodes placed at different positions on the patient's head [3], [27], as illustrated in Fig. 5. The ECG signals were collected from two electrodes, ECG1 and ECG2, placed on the patient's chest [3], [28]. Finally, the EMG samples included EMG measurements of the submentalis muscle and bilateral anterior tibial EMG [3], [29].

B. EXPERIMENTAL FRAMEWORK

The proposed MM-DMS system consists of three independent classifiers (assessors), each using different data

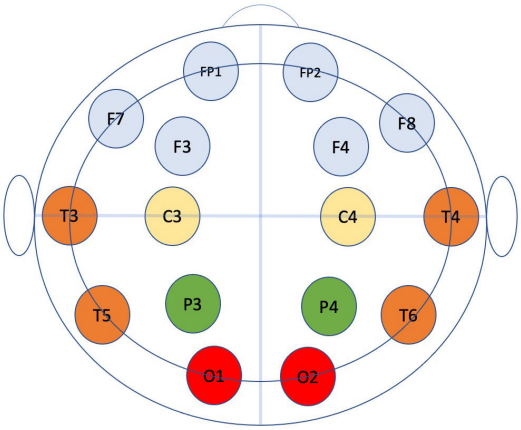


FIGURE 5. Positions of EEG electrodes according to the 10-20 electrode system (Redrawn from [27]).

TABLE 2. Numbers of spectrogram images for each sleep stage generated for 10-second intervals of EEG-, ECG-, and EMG waveforms.

Sleep Stage	EEG images	ECG images	EMG images
W	28510	3125	2072
S1	9561	1089	1002
S2	124738	12912	10587
S3	30606	2901	2343
S4	56526	5487	4794
R	75726	7464	6426
Total	325667	32978	27224

modalities when solving the same task of sleep stage detection. The proposed experimental schedule starts with testing each separate single-modality classifier and then progresses to testing the MM-DMS system. The single-modality tests provide a baseline comparison allowing us to determine which modality is the most efficient in sleep stage detection and if the combined multimodal MM-DMS approach improves the classification outcomes. Table 2 provides the total number of spectrogram images generated using the 10-second blocks of EEG-, ECG-, and EMG waveforms for each sleep category. While similar numbers of images represented the ECG- and EMG modalities, the number of images for the EEG modality was ten times larger. It can also be observed that there was a data imbalance in the sleep stage representation across modalities, with the S2 stage having the highest representation and the S1 stage having the lowest representation.

The numbers of images shown in Table 2 were used in all experiments, with 80% of the data assigned to training and 20% to testing the model. The training/testing procedure was repeated three times for each experiment, with the training- and testing subsets being mutually exclusive.

TABLE 3. Average classification accuracy (%) for EEG using a single CNN classifier.

EEG representation using a single CNN classifier	Average Classification Accuracy			
	First run	Second run	Third run	Average
Time Waveform	40.01%	39.49%	38.66%	39.39%
Linear Spectrogram	41.90%	43.38%	44.53%	43.27%
Logarithmic Spectrogram	53%	51.92%	52.88%	52.6%

TABLE 4. F_1 scores for EEG using a single CNN classifier.

Sleep Stage	Time Waveform	Linear Spectrogram	Logarithmic Spectrogram
R	0	0.31	0.54
S1	0	0	0.09
S2	0.55	0.56	0.60
S3	0	0.01	0.07
S4	0.18	0.23	0.52
W	0.19	0.50	0.52
All	0.15	0.27	0.41

C. RESULTS OF SLEEP STAGE DETECTION USING EEG

Sleep stage detection from EEG measurements alone was performed using three alternative types of inputs, i.e., 10-second duration intervals of time waveforms, RGB images of linear spectrograms generated from the 10-second EEG waveforms, and RGB images of logarithmic spectrograms also generated from the 10-second EEG waveforms. Table 3 summarizes the classification accuracy achieved across three training/testing runs as well as the average classification accuracy. Since the class representation was imbalanced across the six sleep stages, the F_1 scores were also calculated and listed in Table 4. To observe the distribution of true- and false positives and negatives across the sleep categories, the EEG confusion matrix shown in Fig. 6 was also generated. Based on the average accuracy; the spectrograms outperformed the time waveforms by 4% (Linear Spectrograms) and by 14% (Logarithmic Spectrograms). It is not surprising, given that CNNs have been designed to classify 2D image arrays rather than time sequence vectors.

Furthermore, the logarithmic spectrograms performed almost 10% better than the linear spectrograms. Given that the logarithmic scale emphasizes the lower frequency range of the EEG signals (Fig. 2), it indicates that low-frequency components of EEG may be particularly indicative of sleep stages. The F_1 score shows the same trend as the accuracy across the three data inputs. The highest F_1 score is 0.6 for class S2 and the lowest i.e., 0.09 for class S1, and 0.07 for class S3, respectively. For all six categories combined, the system's F_1 score is only 0.41, indicating that the learning was not optimal. This is most likely due to the imbalance in class representation. The effect of the imbalance

FIGURE 6. Confusion matrix for EEG classification using a single CNN and logarithmic spectrograms.

TABLE 5. Average classification accuracy (%) for ECG using a single CNN classifier.

ECG representation using a single CNN classifier	Average Classification Accuracy			
	First run	Second run	Third run	Average
Time Waveform	49.15%	48.39%	48.49%	48.68%
Linear Spectrogram	49.47%	49.02%	49.59%	49.36%
Logarithmic Spectrogram	53.79%	54.16%	53.54%	53.83%

is also visible in the confusion matrix shown in Fig. 6. Generally, the percentage of true positives increases with class representation. Hence the highest correct identification can be observed in Fig. 6 for the most highly represented stage S2, and the lowest for stage S3, which had the second-lowest representation. Interestingly, the S1 category, which had the lowest representation, performed slightly better than the S2 category.

D. RESULTS OF SLEEP STAGE DETECTION USING ECG

As for the EEG case, the ECG-based classification was performed using three alternative types of CNN inputs: 10-second intervals of time waveforms, RGB images of 10-second linear spectrograms, and RGB images of 10-second logarithmic spectrograms. Table 5 summarizes the ECG classification accuracy achieved across three training/testing runs, and the average classification accuracy. The F_1 scores are listed in Table 6, and Fig. 7 shows the confusion matrix for ECG classification.

The trends observed for the ECG-based accuracy are consistent with the EEG case, i.e., the spectrograms outperformed the time waveforms. The linear spectrograms performed better than the time waveforms by only 1% and the logarithmic spectrograms by 4%. As in the EEG case (Fig. 2), the logarithmic scale emphasizes the low-frequency range;

TABLE 6. F_1 scores for ECG using a single CNN classifier.

Sleep Stage	Time Waveform	Linear Spectrogram	Logarithmic Spectrogram
R	0.29	0.54	0.53
S1	0	0.11	0.13
S2	0.62	0.60	0.60
S3	0.03	0.10	0.21
S4	0.24	0.48	0.58
W	0.45	0.47	0.54
All	0.27	0.38	0.43

FIGURE 7. Confusion matrix for ECG classification using a single CNN and logarithmic spectrograms.

TABLE 7. Average classification accuracy (%) for EMG using a single CNN classifier.

EMG representation using a single CNN classifier	Average Classification Accuracy			
	First run	Second run	Third run	Average
Time Waveform	44.22%	43.56%	44.36%	44.05%
Linear Spectrogram	52.35%	52.08%	52.40%	52.28%
Logarithmic Spectrogram	57.88%	58.69%	59.95%	58.84%

therefore, the results indicate that low-frequency components of ECG are highly indicative of the sleep stage.

The largest F_1 score of 0.6 was achieved for the highly represented class S2 and the lowest, i.e., 0.13 for class S1, which had the lowest training data representation. For all six categories combined, the F_1 score was 0.43, again indicating that the learning was sub-optimal due to the imbalance in class representation. The confusion matrix shown in Fig. 7, shows an increase in the percentage of true positives with the class representation. The highest correct identification can be observed for the two highest represented stages, S2 scoring 66% and R scoring 63%, respectively. The lowest

TABLE 8. F_1 score for EMG using a single CNN classifier.

Sleep Stage	Time Waveform	Linear Spectrogram	Logarithmic Spectrogram
R	0.39	0.53	0.61
S1	0	0	0.19
S2	0.58	0.61	0.65
S3	0	0.09	0.26
S4	0.01	0.49	0.60
W	0.07	0.39	0.48
All	0.17	0.35	0.47

FIGURE 8. Confusion matrix for EMG classification using a single CNN and logarithmic spectrograms.

percentage of true positives was for the least represented stage, S1, scoring only 7.8%.

E. RESULTS OF SLEEP STAGE DETECTION USING EMG

To be consistent with the EEG- and ECG experiments, the EMG-based classification was also performed using three alternative types of CNN inputs: 10-second duration intervals of time waveforms, RGB images of 10-second linear spectrograms, and RGB images of 10-second logarithmic spectrograms. Table 7 summarizes the EMG classification accuracy achieved across three training/testing runs and the average classification accuracy. The F_1 scores are listed in Table 8, and Fig. 8 shows the confusion matrix for EMG classification.

As in the previous case of classification based on EEG and ECG, the EMG results show that spectrograms have higher accuracy than time waveforms. The linear spectrograms outperformed the waveforms by 8% and the logarithmic spectrograms by a staggering 14%. The logarithmic spectrograms performed better than the linear spectrograms by 6% and achieved the highest classification accuracy of 58.84%. This points to the importance of the low-frequency components of EMG to sleep stage recognition. The highest EMG F_1 score of 0.65 was achieved for the highest represented class S2. Whereas the lowest F_1 score of 0.19 was achieved for class S1. The second least represented class S3's F_1 score was 0.26. The combined F_1 score for all sleep stages was 0.47, confirming that there is space to improve the training model. The confusion matrix shows a clear bias towards the highly represented S2 class with 74% of correct identifications and only 12% for the least represented S1 category.

TABLE 9. Average classification accuracy (%) using the multimodal MM-DMS method.

Decision making methods	Average Classification Accuracy			
	First run	Second run	Third run	Average
Shallow NN (MM-DMS)	80.26%	82.42%	74.02%	78.9%
PT Shallow NN (MM-DMS)	95.68%	96.92%	93.73%	95.43%
Maximum Probability	65.73%	65.35%	66.02%	65.70%
Majority Voting	63.39%	63.56%	65.93%	64.29%
Average Probability	40.70%	41.21%	39.70%	40.54%

F. RESULTS OF MULTIMODAL SLEEP STAGE DETECTION USING MM-DMS

Based on the outcomes of the single-modality classification, the multimodal MM-DMS classification system was composed of the best-performing single-modality models. This means that the MM-DMS consisted of EEG-CNN, EEG-CNN, and EMG-CNN classifiers trained on RGB images representing logarithmic spectrograms calculated for 10-second intervals of EEG-, ECG-, and EMG signals, respectively. The three classifiers were fused at the final decision-making level using a decision-making module. Here, the proposed shallow decision-making NN module (trained from scratch and pretrained) is compared with three classical decision-making methods, maximum probability, majority voting, and average probability.

Table 9 shows the MM-DMS classification accuracy for each of the three runs and the average accuracy. The F_1 score is given in Table 10. Fig. 9 shows the MM-DMS confusion matrix when using the trained from scratch NN, and Fig. 10 when using the pretrained NN (denoted PT Shallow NN). While in the previous cases for single-modality classifiers, the performance of waveforms with spectrograms using different frequency scales was compared, this time, different decision-making methods will be compared.

The highest average accuracy of 95.43% was achieved for the MM-DMS using the pretrained shallow NN (denoted PT Shallow NN) as a decision-making method. It showed very significant improvements: 16.5% over the trained from scratch NN, 45% over the average probability, 31% over the maximum voting, and 30% over the maximum probability method. It indicates that the NN was able to outperform the traditional decision-making methods. It could be due to the fact that these methods make arbitrary assumptions that

TABLE 10. Comparison of F_1 scores achieved for different decision-making methods, including the proposed MM-DMS.

Sleep Stage	Shallow NN MM-DMS	PT Shallow NN (MM-DMS)	Max. Prob.	Majority Voting	Avg. Prob.
R	0.85	0.94	0.69	0.72	0.56
S1	0.61	0.99	0.18	0.16	0.05
S2	0.81	0.95	0.69	0.71	0.55
S3	0.61	0.97	0.24	0.23	0.20
S4	0.79	0.97	0.68	0.64	0.15
W	0.90	0.99	0.72	0.67	0.56
All	0.76	0.97	0.53	0.52	0.35

work in some cases, but not in general. On the other hand, the NN is free of such assumptions and learns the judgment based on the pattern of correct "ground truth" decisions learned during the supervised training process. Another critical factor contributing to such significant improvement of the classification accuracy was the fact that the decision-making network was pretrained on a related task, so it had a built-in relevant prerequisite knowledge.

Importantly, the overall values of accuracy for the MM-DMS are higher than for any of the single-modality classifiers tested in the previous experiments presented. Namely, for all sleep stages together, the MM-DMS (with the pretrained NN) (Table 9) outperformed the logarithmic versions of EEG CNN by 42.83% (Table 3), ECG CNN by 41.6% (Table 5), and EMG CNN by 36.59% (Table 7). It indicates that the combined multimodal information leads to a significant improvement in sleep stage prediction compared to the single-modality approaches (i.e., using EEG, ECG, or EMG alone).

The MM-DMS F_1 scores followed the same pattern as the average accuracy. Thus, the highest value of 0.97 for all sleep stages together was scored by the pretrained shallow NN, followed by the trained from scratch NN and then by the maximum probability with 0.53, majority voting with 0.52, and finally, the average probability with only 0.35. A similar pattern is shown across individual sleep stages. The high values of F_1 scores for the NN indicate that, unlike the classical methods, the decision-making NN allows a certain degree of compensation for the training biases of single-modality CNNs caused by the data imbalance in the sleep stage representation.

This ability is further confirmed by the pattern of individual sleep stage classification results given by the MM-DMS confusion matrices for the trained from scratch NN and pretrained NN. The MM-DMS with the pretrained NN matrix shows a "close to the ideal pattern" of high values along the diagonal with true positive cases ranging from 93% to 99%. In fact, the least represented S1 category of sleep achieved 99% accuracy, and the most represented category S2 achieved 93% accuracy.

FIGURE 9. Confusion matrix for the MM-DMS using a trained from scratch shallow NN.

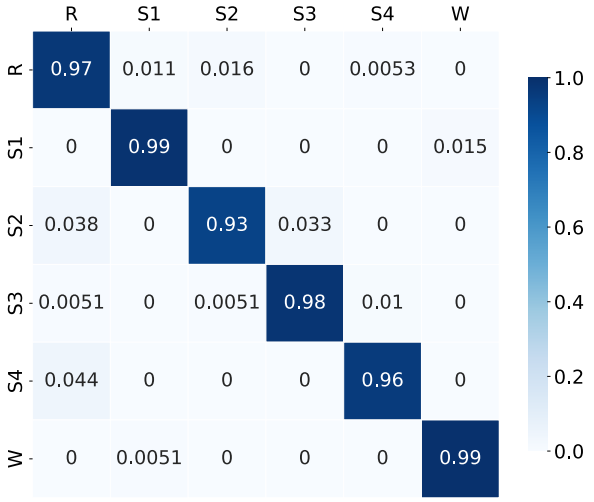


FIGURE 10. Confusion matrix for the MM-DMS using a pretrained shallow NN.

In spite of having the same unbalanced data as in the previous single-modality experiments, the proposed MM-DMS was able to eliminate the training bias to a large extent. Hence, the method is suitable not only to make decisions that are free of presumptions, but also that compensate for the training data imbalance.

V. DISCUSSION

In this section the most important observations arising from the automatic multimodal sleep stage detection experiments will be outlined.

A. COMPARISON BETWEEN SINGLE-MODALITY CLASSIFIERS

While the EMG signal was shown to provide the highest classification accuracy of up to 58.84%, the differences

TABLE 11. Average classification accuracy (%) across different approaches tested in this study.

Input Type	Modality Classification			Multimodal Classification				
	ECG	EEG	EMG	Max. prob.	Maj. voting	Avg. prob.	Shallow NN	PT Shallow NN
Time Waveform	48.68%	39.49%	44.05%	47.65%	43.44%	30.37%	53.14%	87.65%
Linear Spectrogram	49.36%	43.27%	52.28%	56.20%	51.03%	23.02%	66.61%	90.54%
Logarithmic Spectrogram	53.83%	52.6%	58.84%	65.70%	64.29%	40.54%	78.90%	95.43%

between EMG and other modalities (EEG, ECG) were relatively small (5% less for ECG and 6% less for EEG). It means that all three modalities can provide almost equally effective sleep stage classification.

B. COMPARISON BETWEEN CNN INPUTS

A clear impact of the CNN input format on the classification outcomes was observed. The logarithmic spectrogram images led to the highest performance in single-modality tests. The second-best performance was observed for the linear spectrogram images, while the time waveforms were the worst performers. This observation was consistent across all three modalities (EEG, ECG, and EMG). The superior performance of spectrogram images can be largely attributed to the use of CNN models that were designed to perform a general image object recognition task. Whereas the higher performance of logarithmic spectrograms indicates that the sleep stage information is likely to be encoded within the low-frequency range of the signal bandwidth. This applies to all three modalities.

C. COMPARISON BETWEEN SINGLE-MODALITY AND MULTIMODAL CLASSIFICATION

It was shown that the multimodal MM-DMS approach using the shallow NN as a final decision-making mechanism significantly outperformed multimodal approaches using traditional decision-making methods, as well as each of the three single-modality classifiers. It applies to both improvement of accuracy (Fig. 11a) and F_1 score (Fig. 11b). This outcome was to be expected, considering that the information used to make the multimodal decision was much richer compared to what was available to single-modality classifiers. Different sleep stages are likely to be characterized by events associated with specific modalities. Therefore, the more modalities are used, the more informed the decision-making process. Most significantly, the multimodal MM-DMS improved not only the overall classification accuracy, but also the confusion matrix (Fig. 10). It effectively canceled out the detrimental effect of class imbalance that crippled the single-modality performance.

D. COMPARISON BETWEEN SLEEP STAGES

As revealed by the confusion arrays and F_1 scores, when using the single-modality CNNs, the accuracy of sleep stage detection varied significantly across different stages.

TABLE 12. F_1 scores across different approaches tested in this study.

Input Type	Modality Classification			Multimodal Classification				
	ECG	EEG	EMG	Max. prob.	Maj. voting	Avg. prob.	Shallow NN	PT Shallow NN
Time Waveform	0.27	0.15	0.17	0.29	0.15	0.1	0.46	0.84
Linear Spectrogram	0.38	0.27	0.35	0.41	0.33	0.14	0.60	0.9
Logarithmic Spectrogram	0.43	0.41	0.47	0.53	0.52	0.35	0.76	0.97

Generally, stages represented by larger numbers of training samples were identified with higher accuracy than stages represented by the smaller number of samples. Hence, the S2 stage had the highest prediction accuracy across all modalities (EEG, ECG, and EMG), whereas the least represented stage S1 - was the lowest. This situation changed dramatically when the single-modality CNN models were replaced by the multimodal MM-DMS combined with the shallow decision-making NN. It led to a uniform distribution of classification accuracy across all six sleep stages (Fig. 10).

E. COMPARISON BETWEEN DECISION-MAKING METHODS

The multimodal MM-DMS approach combines three single-modality models at the final decision-making stage. A comparison between traditional decision-making mechanisms such as the maximum accuracy, majority voting, and average probability with the proposed shallow decision-making NN, showed that the shallow NN (with the pretrained NN) led to the best performance. The improvement was observed in the significantly higher average classification accuracy (95.43%) and well-balanced confusion arrays with a very even distribution of correct classification across all six sleep stages (93%-99%). This outcome was observed in spite of an uneven class representation. It shows that the proposed method not only improves the overall accuracy but also compensates for the imbalanced training conditions.

F. COMPARISON WITH RELATED STUDIES

Table 13 lists a few examples of closely related works to show how the current study fits into the existing body of knowledge in the field. As already explained in Section III-B, the proposed MM-DMS method differs significantly from previous studies as it proposes a distributed system of independent networks rather than a single multimodal stacked network structure [6], [7].

The top performance amongst the studies listed in Table 13 belongs to [18] who identified five sleep categories using two data modalities, EEG and EOG, and achieved an average accuracy of 94.34%. That approach applied a fusion of features extracted from different modalities. In contrast, this study has identified six sleep stages using three modalities, EEG, ECG, and EMG, achieving a slightly higher accuracy

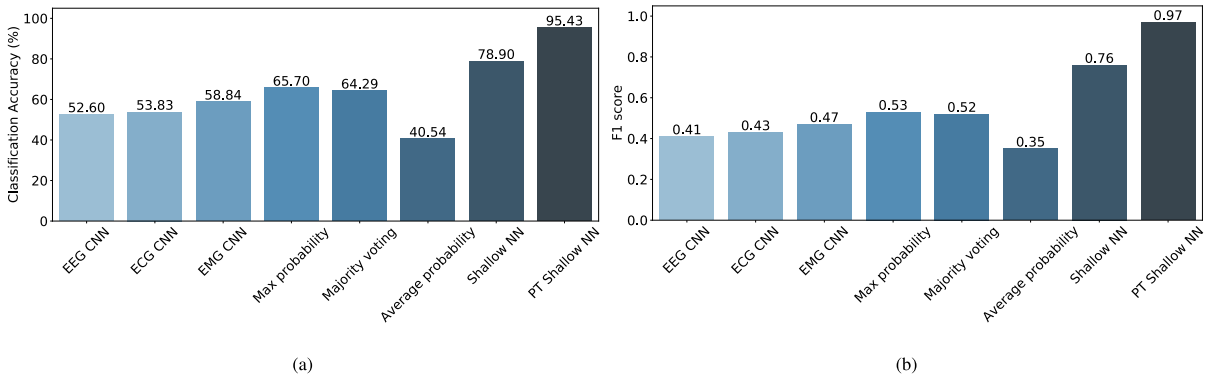


FIGURE 11. Comparison between sleep classification methods; (a) Average accuracy (%), (b) F_1 score.

TABLE 13. Comparison with related studies.

Authors	Database	Modalities	Classes	Method	Accuracy(%)
Zhang, *et al.* [12]	SHHS	EEG, EMG,EOG	5	CNN + LSTM	87%
Abdollahpour, *et al.* [18]	Sleep-EDFx datasets	EEG, EOG	5	Transfer learning + Multi-CNN	94.34%
Zhai, *et al.* [21]	Synchronised PSG	EEG, EMG,EOG	5	CNN+LSTM+ Maximum/Mean operator	78.2%
Yan, *et al.* [20]	SHHS	EEG,ECG, EMG,EOG	5	CNN	85.2%
Jia, *et al.* [30]	Sleep-EDF-153 dataset	EEG, EOG	5	SalientSleepNet (2 U-Net)	84.1%
Pathak, *et al.* [31]	MST	EEG, EMG,EOG	5	CNN + Bi-LSTM	77%
Proposed Method 1	CAP sleep database	EEG, ECG, EMG	6	MM-DMS - Distributed CNNs + Shallow NN	78.9%
Proposed Method 2	CAP sleep database	EEG, ECG, EMG	6	MM-DMS - Distributed CNNs + PT Shallow NN	95.43%

of 95.43%. The proposed approach applied a fusion of classification labels rather than features. Although, a direct comparison of this method with examples listed in Table 13 is not possible due to different data and experimental conditions. In conclusion, the approach presented in this paper matches, if not outperforms the state-of-the-art, as well as all other examples shown in Table 13.

VI. CONCLUSION

The study investigated a multimodal approach to sleep stage detection. A new multimodal classification and decision-making (MM-DMS) system was proposed and validated using six sleep stages (W, S1, S2, S3, S4, and R) and three data modalities (EEG, ECG, and EMG). The classification outcomes derived separately for each modality by a parallel set of CNNs were fused and passed to a shallow NN to make the final diagnosis. The results showed a high average classification accuracy of up to 95.43%, as well as a uniformly distributed confusion array with the accuracy for individual sleep stages ranging from 93% to 99% in spite of unbalanced class representation. Future research will investigate a combined multimodal and multi-label classification of sleep data. Different feature representations of EEG-, ECG- and EMG signals will also be investigated, e.g., the wavelet transform.

ACKNOWLEDGMENT

The CAP Sleep database from the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy, was downloaded via physionet.org.

REFERENCES

[1] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.

[2] M. G. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, and A. Walters, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Med.*, vol. 2, no. 6, pp. 537–553, Nov. 2001, doi: 10.1016/S1389-9457(01)00149-6.

[3] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, Jun. 2000, doi: 10.1161/01.cir.101.23.e215.

[4] C.-S. Huang, C.-L. Lin, W.-Y. Yang, L.-W. Ko, S.-Y. Liu, and C.-T. Lin, "Applying the fuzzy C-means based dimension reduction to improve the sleep classification system," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2013, pp. 1–5, doi: 10.1109/FUZZ-IEEE.2013.6622495.

[5] C. O'Reilly, N. Gosselin, J. Carrier, and T. Nielsen, "Montreal archive of sleep studies: An open-access resource for instrument benchmarking and exploratory research," *J. Sleep Res.*, vol. 23, no. 6, pp. 628–635, Dec. 2014, doi: 10.1111/jsr.12169.

[6] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018, doi: 10.1109/TNSRE.2018.2813138.

[7] A. Malafeev, D. Laptev, S. Bauer, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, J. Buhmann, and P. Achermann, "Automatic human sleep stage scoring using deep neural networks," *Frontiers Neurosci.*, vol. 12, p. 781, Nov. 2018, doi: 10.3389/fnins.2018.00781.

[8] Y. Fang, H. Yang, X. Zhang, H. Liu, and B. Tao, "Multi-feature input deep forest for EEG-based emotion recognition," *Frontiers Neurorobotics*, vol. 14, pp. 1–11, Jan. 2021, doi: 10.3389/fnbot.2020.617531.

[9] I. Fernández-Varela, E. Hernández-Pereira, D. Alvarez-Estevez, and V. Moret-Bonillo, "A convolutional network for sleep stages classification," 2019, *arXiv:1902.05748*.

[10] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, "The national sleep research resource: Towards a sleep data commons," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1351–1358, Oct. 2018, doi: 10.1093/jamia/ocy064.

[11] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, F. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, J. M. Samet, and P. W. Wahl, "The sleep heart health study: Design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, Dec. 1997, doi: 10.1093/sleep/20.12.1077.

[12] L. Zhang, D. Fabbri, R. Upender, and D. Kent, "Automated sleep stage scoring of the sleep heart health study using deep neural networks," *Sleep*, vol. 42, no. 11, pp. 1–10, Oct. 2019, doi: 10.1093/sleep/zsz159.

[13] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Joint classification and prediction CNN framework for automatic sleep stage classification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1285–1296, May 2019, doi: 10.1109/TBME.2018.2872652.

[14] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, Sep. 2000, doi: 10.1109/10.867928.

[15] O. Yildirim, U. Baloglu, and U. Acharya, "A deep learning model for automated sleep stages classification using PSG signals," *Int. J. Environ. Res. Public Health*, vol. 16, no. 4, p. 599, Feb. 2019, doi: 10.3390/ijerph16040599.

[16] M. Sokolovsky, F. Guerrero, S. Paisarnsrisomsuk, C. Ruiz, and S. A. Alvarez, "Deep learning for automated feature discovery and classification of sleep stages," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 6, pp. 1835–1845, Dec. 2020, doi: 10.1109/TCBB.2019.2912955.

[17] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 400–410, Mar. 2019, doi: 10.1109/TNSRE.2019.2896659.

[18] M. Abdollahpour, T. Y. Rezaii, A. Farzamnia, and I. Saad, "Transfer learning convolutional neural network for sleep stage classification using two-stage data fusion framework," *IEEE Access*, vol. 8, pp. 180618–180632, 2020, doi: 10.1109/ACCESS.2020.3027289.

[19] D. Jarchi, J. Andreu-Perez, M. Kiani, O. Vysata, J. Kuchynka, A. Prochazka, and S. Sanei, "Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning," *Sensors*, vol. 20, no. 9, p. 2594, May 2020, doi: 10.3390/s20092594.

[20] R. Yan, F. Li, D. Zhou, T. Ristaniemi, and F. Cong, "A deep learning model for automatic sleep scoring using multimodality time series," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Amsterdam, The Netherlands, Jan. 2021, pp. 1090–1094, doi: 10.23919/Eusipco47968.2020.9287518.

[21] B. Zhai, I. Perez-Pozuelo, E. A. D. Clifton, J. Palotti, and Y. Guan, "Making sense of sleep: Multimodal sleep stage classification in a large, diverse population using movement and cardiac sensing," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–33, Jun. 2020, doi: 10.1145/3397325.

[22] M. Lech, M. Stolar, C. Best, and R. Bolia, "Real-time speech emotion recognition using a pre-trained image classification network: Effects of bandwidth reduction and companding," *Frontiers Comput. Sci.*, vol. 2, pp. 1–14, May 2020, doi: 10.3389/fcomp.2020.00014.

[23] Y.-H. Cheng, M. Lech, and R. Wilkinson, "Sleep stage recognition from EEG using a distributed multi-channel decision-making system," in *Proc. 15th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Sydney, Australia, Dec. 2021, pp. 1–7, doi: 10.1109/ICSPCS53099.2021.9660265.

[24] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Las Vegas, NV, USA, Jan. 2018, pp. 169–175, doi: 10.1109/CCWC.2018.8301729.

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

[26] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, and S. Ghemawat. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/

[27] G. H. Klem, H. O. Luders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation. The international federation of clinical neurophysiology," *Electroencephalogr. Clin. Neurophysiol.*, vol. 52, pp. 3–6, Jan. 1999.

[28] M. Mlynczak, M. Zylinski, W. Niewiadomski, and G. Cybulski, "Ambulatory devices measuring cardiorespiratory activity with motion," in *Proc. 10th Int. Joint Conf. Biomed. Eng. Syst. Technol.* Porto, Portugal: SciTePress, 2017, pp. 91–97, doi: 10.5220/0006111700910097.

[29] B. Frauscher, A. Iranzo, B. Högl, J. Casanova-Molla, M. Salamero, V. Gschliesser, E. Tolosa, W. Poewe, and J. Santamaria, "Quantification of electromyographic activity during REM sleep in multiple muscles in REM sleep behavior disorder," *Sleep*, vol. 31, no. 5, pp. 724–731, May 2008, doi: 10.1093/sleep/31.5.724.

[30] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," 2021, *arXiv:2105.13864*.

[31] S. Pathak, C. Lu, S. B. Nagaraj, M. van Putten, and C. Seifert, "STQS: Interpretable multi-modal spatial–temporal-sequential model for automatic sleep scoring," *Artif. Intell. Med.*, vol. 114, Apr. 2021, Art. no. 102038, doi: 10.1016/j.artmed.2021.102038.

YI-HSUAN CHENG received the M.S. degree in electronics engineering from RMIT University, Australia, where he is currently pursuing the Ph.D. degree in electrical and electronics engineering with the School of Engineering. His research interests include data integration and data cleaning, information retrieval, small-samples machine learning, deep learning, and artificial neural networks.

MARGARET LECH (Member, IEEE) received the M.S. degree in physics from Maria Curie-Skłodowska University, Poland, the M.S. degree in biomedical engineering from the Warsaw University of Technology, Poland, and the Ph.D. degree in electrical engineering from the University of Melbourne, Australia. She is currently a Professor of signal processing and artificial intelligence with the School of Engineering, RMIT University, Australia. Her research interests include machine learning applications in speech and image processing, system modeling, and optimization.

RICHARDT H. WILKINSON (Senior Member, IEEE) was born in Vereeniging, South Africa. He received the B.Eng. degree in electrical and electronic engineering and the M.Eng. and Ph.D. degrees in electrical engineering from the University of Stellenbosch, Stellenbosch, South Africa, in 1994, 1998, and 2004, respectively. He joined the Cape Peninsula University of Technology, Cape Town, South Africa, as a Postdoctoral Researcher, in 2005, after which he was appointed as a Senior Researcher, in 2007. In 2008, he was appointed as the Head of the Centre for Instrumentation Research and promoted to Associate Professor, in 2011. He joined RMIT University, Melbourne, Australia, in 2012, where he is currently a Senior Lecturer with the School of Engineering. His research interests include multilevel power electronic converters, modulation techniques, Fourier techniques, digital signal processing, artificial intelligence, digital audio amplifiers, FPGA development, and embedded controller design. He is a Senior Member of the IEEE Signal Processing Society, the IEEE Power Electronics Society, the IEEE Industry Applications Society, the IEEE Industrial Electronics Society, and the IEEE Power Engineering Society. He served as the Chapter Chair for the South Africa Section's Joint Industry Applications/Industrial Electronics/Power Electronics Chapter, from 2007 to 2012.