End-to-End Deep Learning Based, Automated Patient-Level Insomnia Classification

Arthur Sze, Mostafa Shahin, *Member, IEEE*, Beena Ahmed, *Member, IEEE*

***Abstract*— Insomnia is a sleep disorder that affects 30% of the adult population and is linked to various negative health effects. The diagnosis of insomnia is normally conducted by a sleep physician, but this process is time-consuming, tiresome, and requires expertise. Hence, there is a need for an automated process that simplifies the diagnosis process. In this paper we propose a novel deep learning based, end-to-end approach for patient-level insomnia classification with two distinct model architectures based off the popular computer vision Convolutional Neural Network (CNN), AlexNet. Our AlexNet one-dimensional (1D) models are trained on the patients' time-series electroencephalogram (EEG) data, while our AlexNet two-dimensional (2D) models are trained on spectrogram images converted from the EEG data via a multi-taper spectrogram process. Our results show that with the optimal combination of data augmentation and sleep stage results, we achieve up to 94.4% and 88.9% patient-level classification accuracy via our 1D and 2D models respectively.**

***Clinical Relevance*— Clinical diagnosis of insomnia is traditionally a time-consuming process requiring expertise to analyze polysomnographic (PSG) and other collected data. Automating the sleep EEG analysis process not only alleviates pressure on physicians, but provides an objective, standardized benchmark for clinics worldwide. The proposed work could be coupled with existing wearable technology to deliver patients with test results immediately, rendering a quick and effective pathway for insomnia diagnosis.**

I. Introduction

Insomnia is a sleep disorder in which an individual experiences difficulty initiating sleep, struggles to maintain sleep, and/or is dissatisfied with the sleep quantity and/or quality [1]. Affecting 30% of the adult population [1], it negatively impacts an individual's health and wellbeing, leading to obesity, diabetes, hypertension, depression and cardiovascular diseases [2].

Insomnia is diagnosed via a combination of interviews, questionnaires, surveys and overnight polysomnography (PSG). The PSG includes the electroencephalogram (EEG), a discrete, time-domain representation of the electrical activity of the brain which is divided into 30s frames[1] for analysis. Each frame is labelled as one of five sleep stages: Wake (W), Stage 1 (S1), Stage 2 (S2), Slow Wave Sleep (SWS) and Rapid Eye Movement (REM) [3]. They are then visually analyzed to identify the patient as healthy or insomniac. However, reviewing an entire night's sleep is time-consuming, tiresome, and requires expertise. Machine learning is increasingly being used to automate this process and assist clinicians in insomnia diagnosis.

Machine learning methods can be split into two approaches: feature-based and end-to-end. Feature-based approaches input a handcrafted EEG feature vector into a machine learning model. Features identified in previous work on automated insomnia diagnosis include spectral domain power, statistical measures and Hjorth parameters [4–6].

In end-to-end approaches, the data is inputted into the classifier without the need for manual feature engineering. CNNs (Convolutional Neural Networks) are the most popular architecture, accounting for 62% of all architectures used in EEG deep learning [7]. [8] proposed a one-dimensional (1D) CNN for sleep stage classification using a combination of electrooculogram (EOG) and EEG training data. [9] presented a 1D convolutional and attention mechanism neural network for single-channel sleep stage classification. [10] converted the raw EEG into multi-taper spectrogram images as input to their VGGNet-based CNN models for automated sleep stage scoring. The benefit of the multi-taper spectrum approach is it reduces bias and variance of the spectrum estimate, as opposed to the traditional single-taper estimate. [11] proposed a 1D variant of the popular computer vision CNN, AlexNet [12], for frame-level insomnia classification. Existing work for end-to-end approaches remain limited in EEG-related tasks compared to feature-based approaches, let alone for insomnia diagnosis.

Our work aims to address these deficiencies by using 1D and two-dimensional (2D) variants of AlexNet to perform frame-level insomnia classification. For the 2D variant we convert the EEG into a multi-taper spectrogram. Unlike previous work in [11] which only looked at frame-level classification, we make use of the frame-level results to produce patient-level insomnia classification and additional data augmentation techniques to improve performance.

The rest of the paper is organized as follows. Our proposed work and methods are described in Section II. Our results are presented in Section III and conclusions in Section IV.

II. Methods

Fig. 1 presents an overview of our 1D and 2D patient-level insomnia classifiers. In the 1D classifier, after pre-processing, the EEG data is augmented, then frame-level classification performed with the 1D version of the AlexNet classifier. In the 2D classifier, after pre-processing, the EEG is converted to a multi-taper spectrogram, the images augmented and frame-

A. Sze, M. Shahin and B. Ahmed are with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia (e-mail: z5165205@zmail.unsw.edu.au; m.shahin@unsw.edu.au, beena.ahmed@unsw.edu.au).

[1] In literature, this is commonly referred to as *sleep epochs*. To avoid confusion with *training epochs* used in deep learning, we use the terminology *sleep frames* or *frames*.

level classification performed with the 2D version of the AlexNet classifier. Finally, we utilize both frame-level results to obtain separate patient-level insomnia classification.

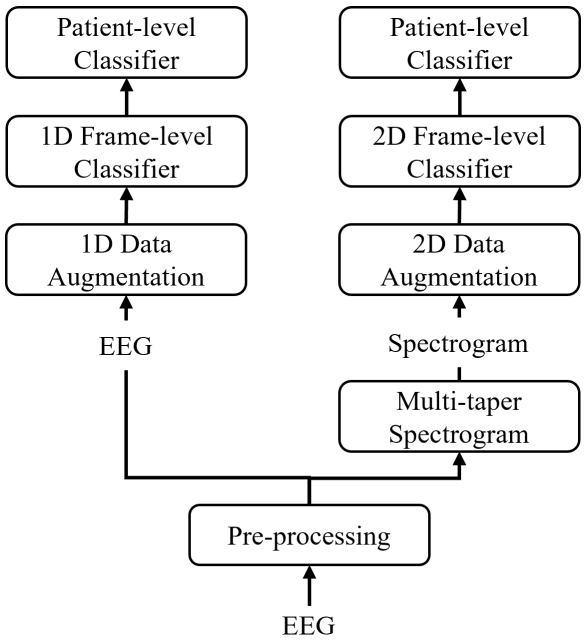

Figure 1. System Diagram

A. Dataset

For our work, we used the public CAP sleep database [13] which contains PSG recordings from 16 control (CL) and 9 insomniac (I) patients. We used the EEG channel with the most patient data, C4-A1, giving us 9 CL (labelled n1-5, n10-12, n14) and 9 I patients (labelled ins1-9).

B. Pre-processing

Patients' sleep EEG were first filtered via a FIR (Finite Impulse Response) filter with a 0.5-40 Hz bandpass and resampled to 128 Hz. The EEG were then divided into 30s frames. Each frame was classed as CL or I based off the patient's clinical class and assigned the corresponding sleep stage label. We kept frames labelled with sleep stages S1, S2, SWS or REM and with signal amplitude within the range [-250, 250 μV]; frames outside these criteria were discarded. DC offset removal was applied for each frame. See Table I for post-processed dataset statistics.

TABLE I. CAP DATASET STATISTICS

Frames	Sleep Stage				
	S1	***S2***	***SWS***	***REM***	***Total***
Control	308	3326	2376	2047	8057
Insomnia	390	3117	1386	1249	6142
Total	698	6443	3762	3296	14199

C. Multi-taper Spectrogram

A copy of each EEG frame was converted to a 256 × 227 pixel spectrogram frame, or image, via the multi-taper spectrogram conversion process [14]. The number of tapers, K, is dependent on the time-bandwidth product, TBW:

$$K = \lfloor TBW \rfloor - 1 \quad (1)$$

The TBW is determined based off the window size of which the signal is thought to be stationary, N, and the desired frequency resolution, Δf:

$$TBW = N \times \Delta f \quad (2)$$

We chose $N = 6$s and $\Delta f = 1$ Hz to obtain a time-bandwidth product of $TBW = 6$ and $K = 5$ tapers.

D. Data Augmentation

We considered two data augmentation approaches in our work. In offline augmentation the entire dataset is expanded with additional samples prior to any model training. In online augmentation, samples in the training set are replaced with augmented samples on-the-fly during each training epoch to increase data variability. Table II and Table III summarize the augmentations applied to the 1D EEG and 2D spectrograms.

TABLE II. OFFLINE AUGMENTATIONS

Augmentation	Data	Description
Overlap	1D, 2D	A 25s overlapping window is applied between adjacent frames of the same sleep stage, expanding the dataset 6x.

TABLE III. ONLINE AUGMENTATIONS

Augmentation	Data	Description
Gaussian noise	1D	Each frame has a 50% chance of being augmented with Gaussian noise with zero mean and standard deviation, σ, of 0.0001, 0.001, or 0.01.
Signal reverse	1D	Each frame has a 50% chance of the signal being reversed along its time-axis.
Horizontal crop	2D	Each frame is cropped randomly along its horizontal axis from 256 × 227 down to 227 × 227 pixels. If this augmentation is disabled, images are simply resized to 227 × 227 pixels.
Horizontal flip	2D	Each frame has a 50% chance of being flipped along its horizontal axis.

E. Frame-level Classifiers

We implemented two architectures for our frame-level classifiers, AlexNet 1D and AlexNet 2D, based on the original AlexNet CNN [12]. Our architectures comprise of 5 convolutional layers followed by 3 dense layers. Our AlexNet 1D model parameters are based off existing work done in [11]. Our AlexNet 2D model parameters are similar to that of the original AlexNet architecture with the dense layer neuron counts adjusted to 4096, 1000 and 2 as per our binary classification task. Our frame-level classifiers were trained on a selected combination of sleep stage subsets and data augmentations as per Table IV. We applied the overlap offline augmentation for all our frame-level classifiers.

TABLE IV. COMBINATIONS OF SLEEP STAGE SUBSETS AND DATA AUGMENTATIONS FOR FRAME-LEVEL CLASSIFICATION

Sleep stage subsets	ALL (S1 + S2 + SWS + REM), S1, S2, SWS, REM
Offline data augmentation	Overlap
1D online data augmentations	All (Gaussian noise with σ + signal reverse), Gaussian noise with σ, signal reverse, none (no augmentation)
2D online data augmentations	All (horizontal crop + horizontal flip), horizontal crop, horizontal flip, none (no augmentation)

F. *Patient-level Classifiers*

We considered three approaches to obtain patient-level classification from the frame-level classification results:

- Single-subset – The patient is classified according to the majority class of their frames in the single chosen sleep stage subset. In ALL, the class of the majority of the total frames across S1, S2, SWS and REM is used.
- Ensemble-OR – The patient is classified as I if the majority of frames within one or more of the chosen sleep stage subsets are classed as I. Otherwise, the patient is classed as CL.
- Ensemble-AND – The patient is classified as I if the majority of frames within all of the chosen sleep stage subsets are classed as I. Otherwise, the patient is classed as CL.

Table V shows the different combinations of frame level sleep stage results we used for patient-level classification.

TABLE V. COMBINATIONS OF SLEEP STAGE SUBSETS USED FOR PATIENT-LEVEL CLASSIFICATION

Single-subset	ALL (S1 + S2 + SWS + REM), S1, S2, SWS, REM
Ensemble-OR	(S1, S2, SWS, REM), (N2, SWS, REM)
Ensemble-AND	(S1, S2, SWS, REM), (N2, SWS, REM)

III. RESULTS

A. *Testing Details*

9-Fold cross-validation was employed such that each iteration consisted of 7 folds for training and 1-fold each for validation and testing. Each fold contained frames belonging to one CL and one I patient. In the training set, frames from the minority class were oversampled to achieve a balanced number of CL to I frames. EEG signal amplitude and spectrogram pixel values were normalized to [0, 1], and standardized to zero mean and unit standard deviation. Each iteration was trained for 200 epochs and the model pertaining to the highest training epoch validation accuracy was selected for testing. Performance metrics were averaged across all iterations. Model training was completed on the UNSW computational cluster Katana [15] using Python and software packages Keras and TensorFlow.

B. *Frame-level Classification Results*

Results comparing the use of different sleep stage subsets and data augmentations are shown in Fig 2 and Fig 3 for our 1D and 2D architectures respectively. On AlexNet 1D, the best performing classifier was trained on the REM subset with all augmentation ($\sigma = 0.01$) (accuracy = 74.4%); averaging across the sleep stage subsets, using all augmentation ($\sigma = 0.01$) also provided the best result (acc. = 66.9%). On AlexNet 2D, training on the S2 subset without online augmentation yielded the best results (acc. = 76.0%); averaging across the sleep stage subsets, the best classifier had no online augmentation as well (acc. = 67.1%). The effect of online data augmentation on frame-level classification appears to be minimal. The selection of the sleep stage subset is significant to model performance given the relatively poorer results associated with the S1 subset, however this could also be attributed to a lack of training data for this subset.

C. *Patient-level Classification Results*

1) *Single-subset Classification*

Fig. 4 displays the distribution of augmentations applied for each sleep stage subset and architecture; we list the best augmentations as follows. For the 1D data, the best classifier made use of the ALL sleep stage subset with all augmentation ($\sigma = 0.0001$) (accuracy = 94.4%, precision = 100%, recall = 88.9%). This was followed by REM with all augmentation ($\sigma = 0.001$) (acc. = 83.3%, pre. = 100%, rec. = 66.7%), S2 with all augmentation ($\sigma = 0.0001$) (acc. = 77.8%, pre. = 77.8%, rec. = 77.8%), SWS with all augmentation ($\sigma = 0.0001$) (acc. = 72.2%, pre. = 83.3%, rec. = 55.6%) and finally S1 with signal reverse augmentation (acc. = 61.1%, pre. = 66.7%, rec. = 44.4%). With the 2D data, the best performance was again achieved by using the ALL subset with no online augmentation (acc. = 88.9%, pre. = 100%, rec. = 77.8%) and S2 with flip augmentation (acc. = 88.9%, pre. = 81.8%, rec. = 100%). This was followed by REM with no online augmentation (acc. = 77.8%, pre. = 85.7%, rec. = 66.7%),

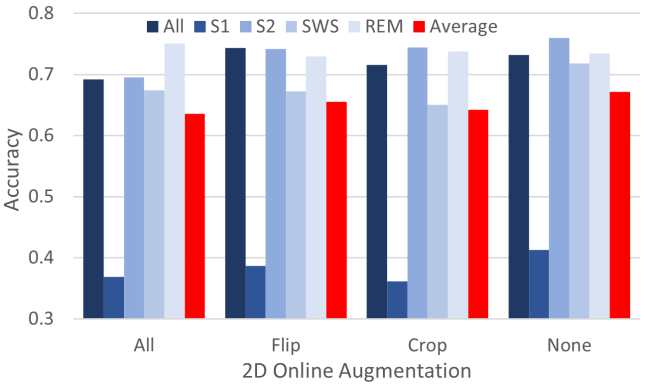

Figure 3. AlexNet 2D Frame-level Classifier Performance

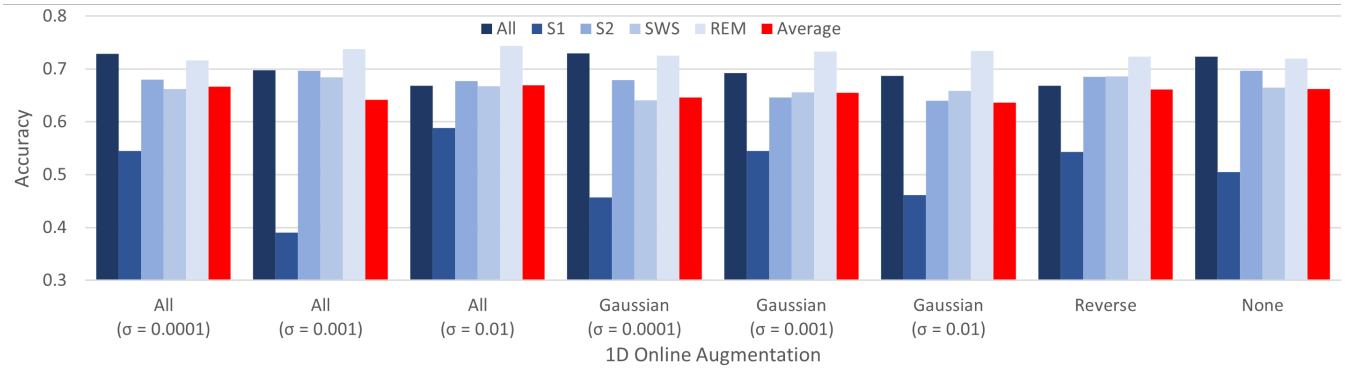

Figure 2. AlexNet 1D Frame-level Classifier Performance

SWS with crop augmentation (acc. = 77.2%, pre. = 75.0%, rec. = 66.7%) and lastly S1 with no online augmentation (acc. = 55.6%, pre. = 60.0%, rec. = 33.3%).

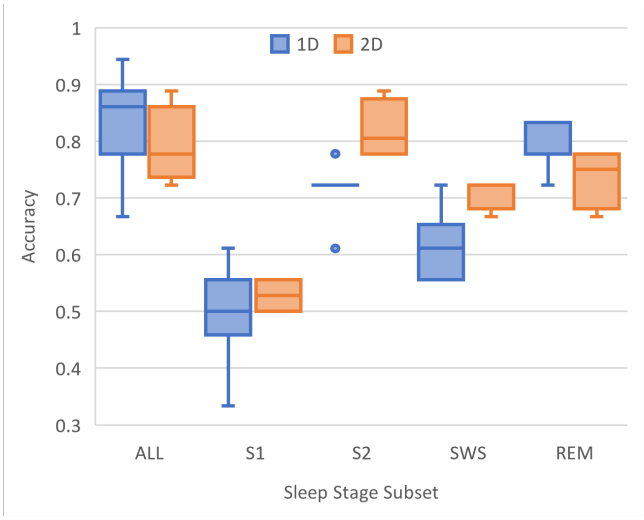

Figure 4. Single-subset Patient-level Classifier Performance

2) *Ensemble Classification*

The top-performing single-subset classifiers for each sleep stage subset and data types were combined for our ensemble patient-level classification method (results in Fig 5). For both 1D and 2D datasets, the S2, SWS, REM ensemble-OR classifier was the best performer (acc. = 83.3%, pre. = 75.0%, rec. = 100%), followed by the S1, S2, SWS, REM ensemble-OR classifier (acc. = 77.8%, pre. = 69.2%, rec. = 100%). On the 1D dataset, this was followed by the ensemble-AND classifiers trained on the S2, SWS, REM subsets (acc = 66.7%, pre. = 100%, rec. = 33.3%) and lastly the S1, S2, SWS, REM subsets (acc. = 50%, rec. = 100%). This was also same case for the 2D ensemble-AND classifiers (acc. = 72.2%, pre. = 100%, rec. = 44.4%) (acc. = 55.6%, pre. = 100%, rec. = 11.1%).

Overall, the patient-level classifier that used the ALL (S1+ S2 + SWS + REM) subset of frame-level results outperformed all other patient-level classifiers, including the ensemble classifiers which combined frame-level classifiers that were trained on individual sleep stages (Fig. 5). This suggests that a greater variety and quantity of training data is crucial for improved patient-level classification performance.

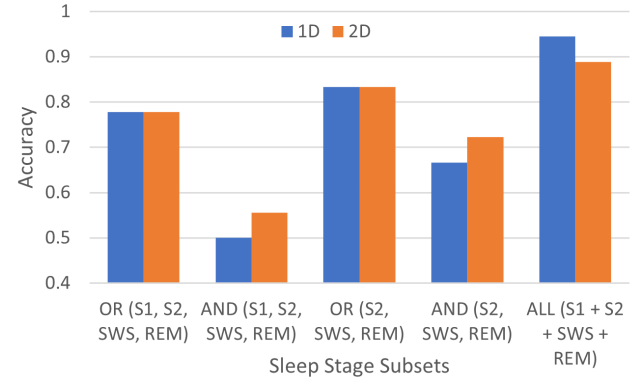

Figure 5. Ensemble Patient-level Classifier Performance

IV. Conclusion

In this paper we propose a novel, end-to-end method for automated patient-level insomnia diagnosis. We show how patient-level classification can be obtained from CL/I frame-level 1D and 2D variants of the AlexNet CNN classifiers. We also show how combinations of frame-level classifiers applied on different sleep stages and data augmentation techniques impact patient-level classification. Results indicate that it is possible to accurately classify patients with insomnia (acc. = 94.4%), but only by maximizing data variability and quantity by using frames from all stages and data augmentation. Our work demonstrates the merit of further development of not only automated insomnia diagnosis, but also diagnosis of other EEG-related disorders.

References

[1] "Insomnia." The American Academy of Sleep Medicine, 2008. [Online]. Available: https://aasm.org/resources/factsheets/insomnia.pdf

[2] O. M. Buxton and E. Marcelli, "Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States," *Soc. Sci. Med.*, vol. 71, no. 5, pp. 1027–1036, Sep. 2010, doi: 10.1016/j.socscimed.2010.05.041.

[3] C. Iber, "The AASM Manual for the Scoring of Sleep and Associated Events : Rules," *Terminol. Tech. Specif.*, 2007, Accessed: Nov. 22, 2021. [Online]. Available: https://ci.nii.ac.jp/naid/10024500923/

[4] M. Shahin, B. Ahmed, S. T. Hamida, F. L. Mulaffer, M. Glos, and T. Penzel, "Deep Learning and Insomnia: Assisting Clinicians With Their Diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1546–1553, Nov. 2017, doi: 10.1109/JBHI.2017.2650199.

[5] S. T. Hamida, T. Penzel, and B. Ahmed, "EEG time and frequency domain analyses of primary insomnia," in *2015 37th Annual of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, pp. 6206–6209. doi: 10.1109/EMBC.2015.7319810.

[6] C. Dissanayaka, H. Abdullah, B. Ahmed, T. Penzel, and D. Cvetkovic, "Classification of Healthy Subjects and Insomniac Patients Based on Automated Sleep Onset Detection," in *for Innovation in Biomedical Engineering and Life Sciences*, Singapore, 2016, pp. 188–192. doi: 10.1007/978-981-10-0266-3_39.

[7] E. Lashgari, D. Liang, and U. Maoz, "Data Augmentation for Deep-Learning-Based Electroencephalography," p. 55.

[8] O. Yildirim, U. B. Baloglu, and U. R. Acharya, "A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals," *Int. J. Environ. Res. Public. Health*, vol. 16, no. 4, Art. no. 4, Jan. 2019, doi: 10.3390/ijerph16040599.

[9] T. Zhu, W. Luo, and F. Yu, "Convolution- and Attention-Based Neural Network for Automated Sleep Stage Classification," *Int. J. Environ. Res. Public. Health*, vol. 17, no. 11, Art. no. 11, Jan. 2020, doi: 10.3390/ijerph17114152.

[10] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring," in *2017 IEEE 27th Workshop on Machine Learning for Signal Processing (MLSP)*, Sep. 2017, pp. 1–6. doi: 10.1109/MLSP.2017.8168133.

[11] B. Yang and H. Liu, "Automatic Identification of Insomnia Based on Single-Channel EEG Labelled With Sleep Stage Annotations," *IEEE Access*, vol. 8, pp. 104281–104291, 2020, doi: 10.1109/ACCESS.2020.2999915.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[13] M. G. Terzano *et al.*, "CAP Sleep Database." physionet.org, 2001. doi: 10.13026/C2VC79.

[14] M. J. Prerau, R. E. Brown, M. T. Bianchi, J. M. Ellenbogen, and P. L. Purdon, "Sleep Neurophysiological Dynamics Through the Lens of Multitaper Spectral Analysis," *Physiology*, vol. 32, no. 1, pp. 60–92, Jan. 2017, doi: 10.1152/physiol.00062.2015.

[15] *Katana*. UNSW. Accessed: Dec. 02, 2021. [Online]. Available: https://researchdata.edu.au/katana/1733007