NeuroImage 272 (2023) 120054

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Deep learning-based EEG analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset

Min-jae Kim [a,1], Young Chul Youn [b,c,1], Joonki Paik [a,d,*]

[a] Department of Image, Chung-Ang University, Seoul, 06974, South Korea
[b] Department of Neurology, Chung-Ang University College of Medicine, Seoul, 06973, South Korea
[c] Biomedical Research Institute, Chung-Ang University Hospital, Seoul, 06973, South Korea
[d] Department of Artificial Intelligence, Chung-Ang University, Seoul, 06974, South Korea

ARTICLE INFO

Keywords:
Electroencephalography (EEG)
Dementia
Mild cognitive impairment (MCI)
Deep learning
Convolutional neural network (CNN)
Automatic diagnostic system

ABSTRACT

For automatic EEG diagnosis, this paper presents a new EEG data set with well-organized clinical annotations called Chung-Ang University Hospital EEG (CAUEEG), which has event history, patient's age, and corresponding diagnosis labels. We also designed two reliable evaluation tasks for the low-cost, non-invasive diagnosis to detect brain disorders: i) CAUEEG-Dementia with *normal, mci*, and *dementia* diagnostic labels and ii) CAUEEG-Abnormal with *normal* and *abnormal*. Based on the CAUEEG dataset, this paper proposes a new fully end-to-end deep learning model, called the **C**AUEEG **E**nd-to-end **D**eep neural **Net**work (CEEDNet). CEEDNet pursues to bring all the functional elements for the EEG analysis in a seamless learnable fashion while restraining non-essential human intervention. Extensive experiments showed that our CEEDNet significantly improves the accuracy compared with existing methods, such as machine learning methods and Ieracitano-CNN (Ieracitano et al., 2019), due to taking full advantage of end-to-end learning. The high ROC-AUC scores of 0.9 on CAUEEG-Dementia and 0.86 on CAUEEG-Abnormal recorded by our CEEDNet models demonstrate that our method can lead potential patients to early diagnosis through automatic screening.

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia (Cassani et al., 2018). Since there are no medications to cure or delay the symptom of dementia (Weller and Budson, 2018), it is urgent to obtain low-cost, reliable, and easy-accessible methods to detect dementia as early as possible (Farina et al., 2020). Mild cognitive impairment (MCI) is an intermediate stage between healthy aging and dementia, and is transformed to dementia at an increasing rate (Michaud et al., 2017). The annual rate of conversion from MCI to AD is 3–15 % compared to 1–2 % of the general population (Michaud et al., 2017). Especially, amnestic MCI (aMCI) is the most probable to convert to AD (Csukly et al., 2016).

For early detection of AD and MCI, the first research focused on structural magnetic resonance imaging (sMRI) (Park and Moon, 2016). Despite its acceptable accuracy of detection, the use of sMRI was limited by high cost and poor access in low-income countries (Musaeus et al., 2018). In this context, electroencephalography (EEG) is a good alternative for low-cost, noninvasive, and user-friendly detection tools (Cassani et al., 2018). As a biomarker for neurodegenerative diseases, EEG has recently received much attention (Al-Qazzaz et al., 2014; Cassani et al., 2018; Malek et al., 2017; Musaeus et al., 2018). However, clinical EEG diagnosis has several issues: i) diagnostic accuracy highly depends on trained EEG experts, ii) clinical EEG recording requires several years of pathological training, and iii) decoding EEGs is a time-consuming and exhausting process (Gemein et al., 2020).

Conventional EEG-based automatic systems for early diagnosis of MCI and dementia have made considerable efforts in feature engineering (Fiscon et al., 2018; McBride et al., 2014; Sharma et al., 2019). McBride et al. (2014) computed the relative spectral power, entropy, and complexity features from 48 resting EEG recordings (15 normal controls, 16 early MCI, and 17 early stage AD) during the cognitive and movement tasks. Using the support vector machine (SVM), they obtained an accuracy of 85.4% on three-way classification. Fiscon et al. (2018) analyzed the Fourier and wavelet features of EEG samples acquired from 23 healthy controls (HCs), 37 MCI, and 49 AD subjects, and then trained the decision tree classifiers. Their two-way classifiers achieved accuracies of 83.3% for HC *vs.* AD, 91.7% for HC *vs.* MCI, 79.1% for MCI *vs.* AD, and 73.4% for HC vs. MCI and AD, respectively. Sharma et al. (2019) extracted eight handcrafted features, including

* Corresponding author.
E-mail addresses: imkbsz@cau.ac.kr (M.-j. Kim), neudoc@cau.ac.kr (Y.C. Youn), paikj@cau.ac.kr (J. Paik).
[1] These authors equally contributed to this manuscript.

https://doi.org/10.1016/j.neuroimage.2023.120054.
Received 28 September 2022; Received in revised form 17 March 2023; Accepted 24 March 2023
Available online 29 March 2023.

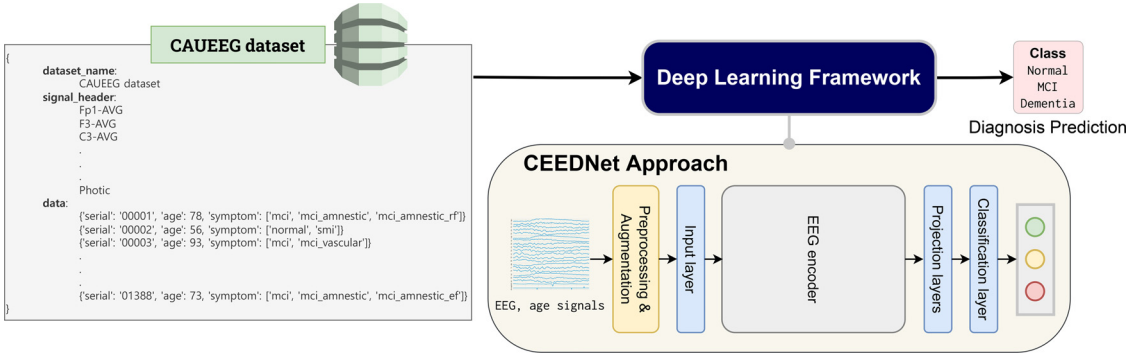


Fig. 1. Overview of the proposed approach and dataset for developing the EEG-based automated diagnostic system.

power spectral density, skewness, kurtosis, and entropy, from 44 EEG signals around the eye-open, eye-close, finger tapping test (FTT), and continuous performance test (CPT) events, and then trained the SVM classifiers. They achieved the binary classification accuracies of 89.8% for HC *vs*. MCI on FTT event, 87.9% for HC *vs*. dementia on CPT, and 88.0% for MCI *vs*. dementia on CPT.

As deep learning has recently experienced explosive success in various fields, including computer vision, natural language processing, and genetics (Brown et al., 2020; He et al., 2016; Senior et al., 2020), some recent works tried to use deep learning algorithms for the EEG-based early detection of dementia (Bi and Wang, 2019; Ieracitano et al., 2019; 2020). Ieracitano et al. (2019) computed power spectral density maps of size 19×159 from 19-channel, 1280-length EEG sequences collected from 63 AD, 63 MCI, and 63 HC EEGs. They then applied a convolutional neural network (CNN) to the power spectral density maps, and their best model showed a three-way classification accuracy of 83.3%. Ieracitano et al. (2020) extracted wavelet features and bispectrum representation from the same EEG dataset as Ieracitano et al. (2019). Their multi-layer perceptron (MLP) model showed a three-way classification accuracy of 89.2%. Bi and Wang trained a deep convolutional Boltzmann machine on 12000 spectral images extracted from 4 AD, 4 MCI, and 4 HC EEG samples via the multi-task learning strategy, resulting in a three-way classification accuracy of 95.0%. During preprocessing, they prepared 3-channel 32×32 spectral images from 64-channel EEG sequences as input to the network. To obtain more training data, these deep learning approaches segmented a single EEG recording into multiple epochs, and distributed them across the training and test sets (so-called "*epoch-based classification*" manner) (Bi and Wang, 2019; Ieracitano et al., 2019; 2020), which reduces the reliability of the experiments. Furthermore, as conventional methods, they still drastically reduced the amount of raw EEG signals using the handcrafted rules during preprocessing.

In summary, both machine learning and deep learning methods have common limitations: i) The amount of data used in their studies is very small (12–189 EEG recordings). ii) Although some models showed satisfactory classification performance, a fair comparison is difficult due to the lack of a common EEG data set and evaluation protocol; And iii) many handcrafted and manual processes, such as artifact rejection and feature extraction, hinder more chances to learn from data.

In this context, this paper aims to aid the active research on automatic EEG diagnosis for the early diagnosis of MCI and dementia. At first, this paper presents a new large EEG data set with well-organized clinical annotations called Chung-Ang University Hospital EEG (CAUEEG)[2] and two tasks related to neurological disorders. Secondly, based on the CAUEEG dataset, an effective, fully end-to-end baseline deep learning approach is proposed, as illustrated in Fig. 1. Extensive experiments demonstrate that our method helps to develop automatic screening for clinical diagnosis, leading the potential patient to early diagnosis.

The EEG-based automatic pathology classification is another research area related to EEG signal and machine/deep learning, which has flourished due to the Temple University Hospital Abnormal EEG Corpus (TUAB) dataset (Lopez de Diego, 2017). The latest TUAB v2.0.0 contains 2717 EEG recordings for training and 276 for testing. The goal of TUAB is to accelerate the reading process of EEG signals for neurologists, and its class labels were assigned according to whether the recorded EEG activity was normal. In contrast, the label decision for early diagnosis of MCI and dementia is made according to clinical criteria for the subject. Therefore, our dataset is more purpose-specified by providing a direct link to neurodegenerative diseases. This means that deep learning models can benefit from the end-to-end gradient backpropagation between the raw signal and specific neurological disorders. The pathology detection results can be used for more general-purpose, but whether the EEG activity is normal does not directly mean brain disease.

Further information on EEG analysis can be found in Roy et al. (2019), which proposed an in-depth review of deep learning-based EEG analysis. They reviewed 154 papers that apply deep learning to EEG, published between 2010 and 2018, for various applications. All the reviewed papers used CNNs or RNNs with a total of 3–10 layers. Perez-Valero et al. (2021) also published a review paper that provided a narrative review of state-of-the-art studies that combined signal processing and machine learning for the early detection of AD including an extensive comparison of machine learning and early deep learning methods until 2020.

The remainder of this paper is organized as follows: Section 2 introduces a new EEG dataset and evaluation tasks, and a new fully end-to-end deep learning approach is proposed in Section 3 for EEG-based diagnostic classification. An in-depth analysis on the proposed models is conducted in Section 4, and Section 5 concludes this paper.

2. Dataset and evaluation tasks

In this section, we describe the details of the CAUEEG dataset and two evaluation tasks including the structure, characteristics, and annotations of the collected data.

2.1. Signal acquisition and data preparation

The CAUEEG dataset contains 1379 EEG recording signals acquired from 1155 patients at Chung-Ang University Hospital from 24th August 2012 to 12th March 2020. Each recording in this dataset consists of 21 channels, of which the first 19 channels are EEG. The other two channels are electrocardiogram (EKG or ECG) and photic stimulation. The EEG recording was conducted according to the International 10-20 system (electrode locations: Fp1, F3, C3, P3, O1, Fp2, F4, C4, P4, O2, F7, T3, T5, F8, T4, T6, FZ, CZ, and PZ) with linked earlobe referencing. The EEG signals of subjects were measured while they were lying comfortably in

[2] https://github.com/ipis-mjkim/caueeg-dataset

Table 1
Convention of abbreviations for clinical diagnoses provided in the CAUEEG dataset. For all EEGs, labels were assigned by aggregating the diagnosis of each subject determined under neuropsychological examination and criteria (Ahn et al., 2010; Dubois et al., 2007; First and Pincus, 2002; Jahng et al., 2015). The boldface denotes the representative diagnosis of each group of diagnoses.

Abbreviation	Clinical diagnosis
dementia	dementia
ad	Alzheimer's disease dementia
load	late-onset Alzheimer's disease dementia
eoad	early-onset Alzheimer's disease dementia
vd	vascular dementia
sivd	subcortical ischemic vascular dementia
ad-vd-mixed	mix of Alzheimer's disease and vascular dementia
mci	mild cognitive impairment
mci-ad	mild cognitive impairment with amyloid PET positive
mci-amnestic	amnestic mild cognitive impairment
mci-amnestic-ef	amnestic mild cognitive impairment with encoding failure
mci-amnestic-rf	amnestic mild cognitive impairment with retrieval failure
mci-non-amnestic	nonamnestic mild cognitive impairment
mci-multi-domain	multi-domain mild cognitive impairment
mci-vascular	vascular mild cognitive impairment
normal	normal
cb-normal	community-based normal
smi	subjective memory impairment or cognitive decline
hc-normal	health care center normal
ftd	frontotemporal dementia
bvftd	behavioral variant frontotemporal dementia
semantic-aphasia	semantic aphasia
non-fluent-aphasia	non-fluent aphasia
parkinson-synd	Parkinson's syndrome
parkinson-disease	Parkinson's disease
parkinson-dementia	Parkinson's disease dementia
nph	normal pressure hydrocephalus
tga	transient global amnesia

bed but awake, and a qualified technician in the hospital supervised the recording process. After passing through an analog filter with 0.5–70 Hz band pass frequency, the signals were recorded with the sampling frequency of 200 Hz using a digital electroencephalograph system (Comet AS40 amplifier EEG GRASS; Telefactor, USA). The signals were then converted to the common average referencing and saved in European data format (EDF) on disks.

For each EEG recording, the CAUEEG dataset provides i) the event history that occurred while recording, ii) the patient's age at the time, and iii) the corresponding diagnosis label decided by neurologists based on neuropsychological examination and criteria. Table 1 shows the list of diagnostic labels provided in the CAUEEG dataset.

In the CAUEEG dataset, the labeling decision for *dementia* conformed to the probable dementia criteria by the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (Dubois et al., 2007) and the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV (First and Pincus, 2002). The subjects were considered with *mci* by the following criteria: i) intact function in activities of daily living, ii) the presence of memory complaints, iii) objective cognitive impairment ($\geq$ 1.0 SD below education- and age-adjusted norms) in more than one cognitive domain including memory on a comprehensive neuropsychological battery (Ahn et al., 2010; Jahng et al., 2015), iv) a clinical dementia rating of 0.5, and v) non-demented case according to the Diagnostic and DSM-IV criteria. The inclusion criteria for the *normal* subjects were as follows: i) intact activities of daily living and ii) no abnormality (within 1.0 SD of education- and age-adjusted norms) on a comprehensive neuropsychological battery (Ahn et al., 2010; Jahng et al., 2015).

We anonymized all the data in the CAUEEG to protect patients' privacy. For instance, a name of each EDF contains a random five-digit number. We removed the information on all unnecessary fields in the file header of all EDF files (*e.g.*, technician, patient name, gender, *etc*.). The record starting date was also altered as a meaningless date (1st January 2100) to prevent tracing the privacy. This study was approved by the Institutional Review Board (IRB) of Chung-Ang University Hospital (registration no. 2009-005-19331).

2.2. Dataset statistics

The mean and standard deviation of the recorded signal lengths in the CAUEEG dataset are 13.34 and 2.83 min, respectively. The histogram computed for the signal length of EEG recording in CAUEEG is provided in Fig. A.3. A histogram calculated for the patients' age when they conducted the EEG tests is shown in Fig. A.4, and their mean and standard deviation are 70.77 and 9.90. The gender ratio of subjects in the CAUEEG dataset is about 60 males per 100 females.

The CAUEEG dataset assigns at least one diagnostic label for each EDF recording among 28 types. Fig. 2 visualizes the occurrence and types of the diagnostic labels in the dataset. Note that the total count of occurrences for all diagnoses exceeds the number of EEGs since each recording can have multiple diagnoses in our dataset.

Each EDF file also has a paired event history file saved in JSON format. The event types and an exemplar history are shown in Table 2. Using this information, one can do an in-depth study by analyzing and comparing the signals around the specific events (*e.g.*, eye-opening, eye-closing, or photic stimulation).

2.3. Evaluation tasks

Based on the CAUEEG dataset, this paper presents two useful evaluation tasks: i) CAUEEG-Dementia and ii) CAUEEG-Abnormal.

2.3.1. CAUEEG-Dementia

For low-cost, non-invasive screening test of cognitive function to detect dementia, we designed an evaluation task to categorize the patients' EEG into normal, MCI, and dementia, *i.e.*, three-way classification. Due to the label diversity of the CAUEEG dataset, this task only uses a part of it, unlike CAUEEG-Abnormal utilizing it as a whole.

We started constructing this evaluation task by taking the EEG recordings with 459 *normal*, 417 *mci*, and 311 *dementia* diagnostic labels (marked in red, orange, and green boxes, respectively, in Fig. 2) in the CAUEEG dataset. The gathered 1187 data were randomly shuffled and then divided as training, validation, and test sets by class about 8:1:1 ratios. Table 3a shows the numbers of data consisting of this CAUEEG-Dementia task by the split and class. Some EEG recordings acquired from the same person on different dates flowed into different sets through this process. The number of subjects coexisting in the training and test sets was 28. These subjects caused the 39 and 28 EEG recordings with patient overlap in the training and test sets, respectively, which amount to $39/950 \approx 4.11\%$ of the training data and $28/118 \approx 23.73\%$ of the test data. Their min/average/median/max time differences were 10/370/178/1425 days. Among them, the diagnosis of $6/28 \approx 21.43\%$ of subjects shifted with time (*e.g.*, *normal* to *mci* or *mci* to *dementia*). There were three overlaps between validation and test sets, with one changed diagnosis.

Compared to the previous EEG-based dementia datasets (Bi and Wang, 2019; Ieracitano et al., 2019; 2020; Sharma et al., 2019), CAUEEG-Dementia has several advantages: i) the size of dataset is much larger than 12–189; ii) an individual EEG recording belongs to only one of the training, validation, and test splits, *i.e.*, there is no overlap; and ii) the data and annotation are so well-organized that it is smooth to connect to deep learning frameworks, including PyTorch (Paszke et al., 2019). The descriptions of the data amount and the annotation style of this task are expressed in Table 3a and Fig. A.2a, respectively.

2.3.2. CAUEEG-Abnormal

The second evaluation task of CAUEEG is to determine whether an owner of the given EEG recording has neurodegenerative diseases,

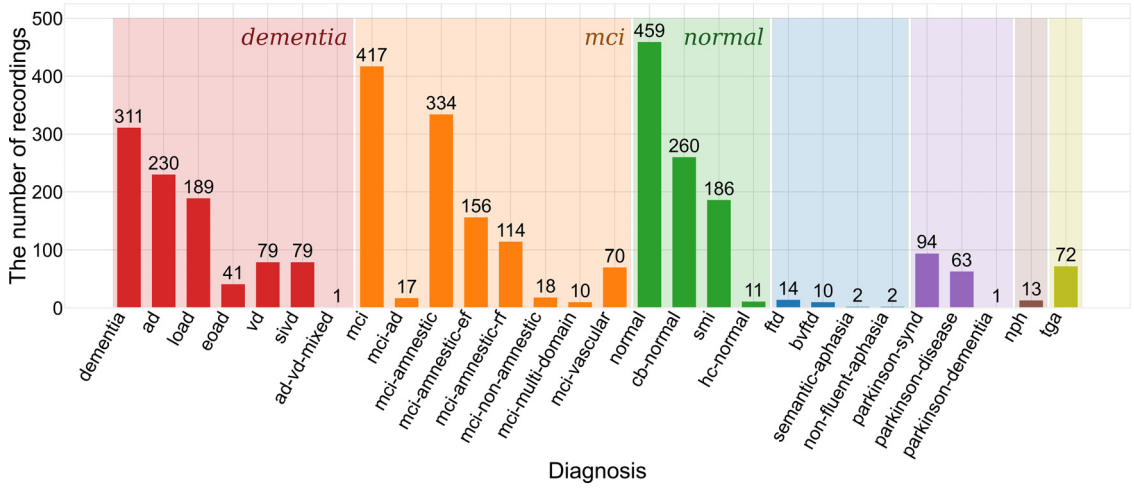


Fig. 2. Bar chart showing the number of EEGs with each symptom. The numbers above each bar denote the occurrence. The color boxes represent the groups of diagnoses. The CAUEEG dataset has no intergroup co-occurrence except for *parkinson-dementia*, which is simultaneously included in the *dementia* and *parkinson-synd* groups. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Event history information of the CAUEEG dataset. The timing is based on sequence number (*e.g.*, an event that occurred five seconds after starting is represented as $5 \times 200 = 1000$).

List of the most frequent twenty events		Example of a recorded event history	
Event keyword	Occurrence	Event keyword	Timing
Eyes Closed	16820	Start Recording	0
Eyes Open	16766	New Montage - Montage 002	0
Photic Off	10413	Eyes Open	36396
Paused	1859	Eyes Closed	72518
Move	1766	Eyes Open	73862
Start Recording	1379	Eyes Closed	75248
New Montage - Montage 002	1092	swallowing	76728
Photic On - 3.0 Hz	1067	Eyes Open	77978
Photic On - 6.0 Hz	1059	Eyes Closed	79406
Photic On - 9.0 Hz	1057	Photic On - 3.0 Hz	79996
Photic On - 12.0 Hz	1054	Eyes Open	80288
Photic On - 15.0 Hz	1048	Eyes Closed	81296
Photic On - 21.0 Hz	1045	Photic Off	82054
Photic On - 18.0 Hz	1042	Photic On - 6.0 Hz	84070
Photic On - 24.0 Hz	1029	Eyes Open	84488
Photic On - 27.0 Hz	1017	Eyes Closed	85538
Photic On - 30.0 Hz	975	Photic Off	86086
Recording Resumed	485	Photic On - 9.0 Hz	88144
artifact	366	⋮	⋮
swallowing	232	Paused	145000

Table 3
Number of train, validation, and test data by class in each evaluation task. The ratio of the amount of data between classes remains similar across all splits.

Data	Training	Validation	Test	Total
Normal	367	46	46	459
MCI	334	42	41	417
Dementia	249	31	31	311
Total	950	119	118	1187
(a) CAUEEG-Dementia evaluation task				
Normal	367	46	46	459
Abnormal	740	90	90	920
Total	1107	136	136	1379
(b) CAUEEG-Abnormal evaluation task				

which embraces all symptoms other than the *normal* group (see Fig. 2), as well as *dementia* and *mci*. In this task, we denote the union of all diagnostic labels but the *normal* group as *abnormal*. Although this CAUEEG-Abnormal may seem similar to TUAB (Lopez de Diego, 2017), the difference lies in the ultimate goal to seek. In other words, TUAB focuses on the abnormality of EEG activity and seeks to accelerate the reading process for neurologists. Its destination is to assist in writing medical reports by automating the determination of the normality of EEG activity. On the other hand, we made the *normal* and *abnormal* labels by considering the patients' comprehensive clinical diagnoses together with EEG activities. CAUEEG-Abnormal is, therefore, more purpose-oriented and aims at a more challenging goal: directing potential patients with brain disorders from local clinics to specialized hospitals for a thorough examination.

The CAUEEG-Abnormal task divides the EEG recordings into groups: the ones with *normal* labels and the others, producing 459 *normal* and 920 *abnormal* data (marked in green and the other colored boxes in Fig. 2), respectively. The training split for this task was generated by starting from the training set in CAUEEG-Dementia and adding random data until it reached 80% of the total. In other words, we organized the training split of this evaluation task to be a superset of the CAUEEG-Dementia training set. The validation and test sets were com-

posed by randomly shuffling and splitting the rest of the EEG recordings by class 1:1 ratios. We also treated the individual EEG recordings independently regardless of whether the same person recorded them. This resulted in the 27 coexisting subjects between the training and test sets. The data with the patient overlap are $39/1107 \approx 3.52\%$ of the training set and $27/136 \approx 19.85\%$ of the test set. Of these 27 patient overlaps, $6/27 \approx 22.22\%$ have different diagnoses between the two sets (*e.g.*, *normal* to *mci* or *mci* to *ftd*). The min/average/median/max time differences were 7/340/80/1080 days. There were three overlaps between validation and test sets, with one changed diagnosis. The numbers of EEG recordings belonging to classes and splits are shown in Table 3b. The information for training, validation, and test data list and their target labels of CAUEEG-Abnormal is described in a JSON file enclosed with the CAUEEG dataset, as shown in Fig. A.2b.

2.4. Evaluation methodology

Both the CAUEEG-Abnormal and CAUEEG-Dementia tasks adopt accuracy (%) as the primary performance metric. The class-wise sensitivity (%) and class-wise specificity (%) are calculated to analyze the performance by symptom and to reduce the bias owing to data imbalance between classes. These class-wise metrics are calculated by converting the classification results as multiple *one-vs-rest* outputs. The receiver operating characteristic (ROC) curve illustrates the diagnostic performance of the classifier at different classification thresholds. The ROC and its area under the curve (ROC-AUC) for each class would help interpret the classifier's behaviors. If a classifier generates the stochastic outputs, we strongly recommend averaging the classification results after multiple estimations to guarantee reproducibility.

The performance evaluation protocol of the CAUEEG tasks prohibits using the test data for training purposes. One should use all EEG signals for only one purpose if they are acquired from the same EDF file, similar to *patient-based classification*. Since there are some patient overlaps between training and test sets in our tasks, our evaluation protocol does not strictly follow *patient-based classification*. However, it is noteworthy that i) all overlapping EEGs were acquired on different days, ii) the patient overlaps account for 20–24% of the test data, and ii) about 21–22% of the overlaps experienced changes in diagnosis. Since Alzheimer's is a progressive neurodegenerative disorder associated with changes in brain function and structure over time (Caviness et al., 2015; Jelic et al., 2000), we believe that the changes in diagnoses over a long period would make our evaluation task more practical. For researchers who wish to circumvent this patient overlap, for any reason, we also offer the *no-overlap* versions of the task annotations, which are constructed by excluding all test and validation data with patient overlap with the training set. The *no-overlap* versions share the same training sets described in Sections 2.3.1 and 2.3.2. Section 4.4 compares our evaluation methodology with *epoch-based classification*, and Section 4.5 discusses the evaluation results on the *no-overlap* test sets.

3. Deep neural networks for diagnostic classification

Recent trends in the deep learning community are to grant the networks as much capacity, flexibility, and data as possible. However, most existing EEG-based dementia classification methods heavily depended on the manually designed preprocessing stages that confined the specific channels, bandwidth, connectivity, filter kernels, or clean usable intervals (Alhussein et al., 2019; Bajpai et al., 2021; Bi and Wang, 2019; Ieracitano et al., 2019; 2020; Sharma et al., 2019; Yıldırım et al., 2020). Since these preprocessing steps were typically done outside of deep learning frameworks, they impeded the information flow and efficiency in the overall process.

Based on the CAUEEG dataset, this paper presents a new fully end-to-end deep learning approach for screening EEG signals, called the **C**AUEEG **e**nd-to-end **d**eep neural **net**work (CEEDNet)[3]. The core idea of CEEDNet is to combine all the functional elements used to analyze EEG signals in a seamless learnable fashion. CEEDNet pursues to utilize the domain characteristics of EEG signals while minimizing unnecessary human intervention.

3.1. Model architecture

This subsection provides a detailed description of the model architecture of the CEEDNet approach.

Backbone networks As the backbone of CEEDNet, we considered VGG (Simonyan and Zisserman, 2014), ResNet (He et al., 2016), ResNeXt (Xie et al., 2017), and ViT (Dosovitskiy et al., 2020) milestone architectures that have proven excellent effectiveness in various fields. We used them with slight modifications since they were not tailored to the EEG data. Fig. 3 illustrates the proposed CEEDNet architecture and its atomic building blocks according to backbones.

Dimension of basic operation An EEG signal can be handled with one-dimensional (1D) and two-dimensional (2D) approaches. Many previous methods have performed spectral analysis after converting the EEG signals into 2D images through time-frequency representation (Bajpai et al., 2021; Bi and Wang, 2019; Gemein et al., 2020; Ieracitano et al., 2019; 2020; Sharma et al., 2019). Since an EEG signal is multi-channel time-series data, it is also smooth to deal with them via 1D models. We presented both 1D and 2D approaches. In Section 4, although each model showed comparable performances, we got the best accuracy using an ensemble of 1D and 2D models.

- **1D models**: This paper uses four 1D CNN models, including 1D-VGG-19, 1D-ResNet-18, 1D-ResNet-50, and 1D-ResNeXt-50-32x4d[3]. The atomic operation of these models is the 1D convolution layer, also known as temporal convolution and causal convolution. In this case, each 2D convolution layer of the backbones with $k \times k$ kernel size is replaced with the 1D convolution with the kernel size of k^2. The stride of each convolution stage is set evenly so that the output sequence length of the last convolution stage becomes between 4 and 8.
- **2D models**: We tested four 2D CNNs and a Vision Transformer (ViT), including 2D-VGG-19, 2D-ResNet-18, 2D-ResNet-50, 2D-ResNeXt-50-32x4d, and ViT-B-16[4], respectively. For these 2D models, an input image is prepared by the short-time Fourier transform (STFT), whose details are described in Section 3.2. Similar to 1D case, the stride of each convolution stage is set evenly so that the output feature map size of the last convolution stage becomes between 4×4 and 8×8.

Age signal Age is one of the most important risk factors for various symptoms, including dementia (Livingston et al., 2020). Thus, this paper proposes to interpret age as an input signal of the network. CEEDNet implements two ways to utilize the age signal:

- *age-conv*: The input EEG signal and a single channel filled with the age value are concatenated together in the input layer.
- *age-fc*: An age value is concatenated into the feature vector before the fully-connected layers (network head).

Projection layers Instead of directly predicting the classification score after the last convolution stage (attention stage for ViT case), we found that some additional feature projection layers improve the accuracy. Therefore, CEEDNet inserted the 1–4 fully-connected layers between the last convolution (or attention) stage and the classification layer. Each projection layer added halves the dimension of the feature vector.

[3] https://github.com/ipis-mjkim/caueeg-ceednet

[4] The suffix of a network name represents the depth of the network by naming convention. However, we do not strictly follow this convention in this paper. We sometimes appended 1–4 additional fully-connected layers at the end without changing the suffix to denote the pairing with the original network.

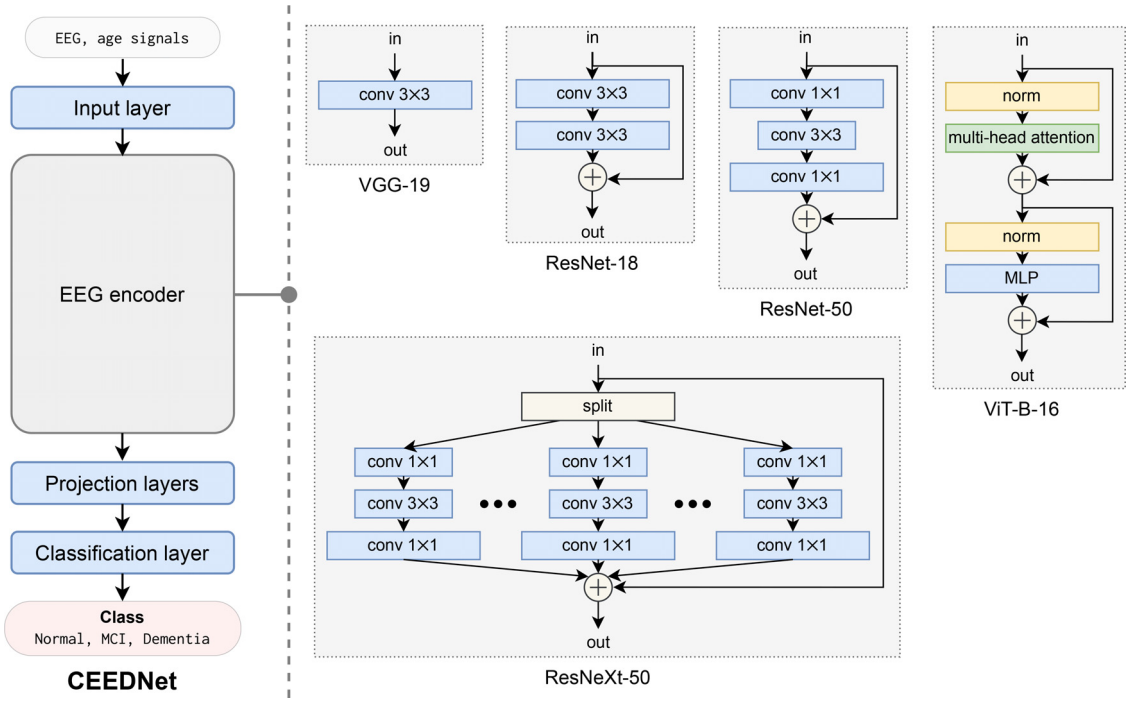


Fig. 3. CEEDNet architecture and its basic building blocks according to the backbones. For the 1D model cases, the 1D convolution with the kernel size of k^2 substitutes for each $k \times k$ convolution layer. Other details not described in this paper follow the original literature (Dosovitskiy et al., 2020; He et al., 2016; Simonyan and Zisserman, 2014; Xie et al., 2017).

Activation function The backbone networks mentioned above mostly adopt the rectified linear unit (ReLU; Nair and Hinton (2010)) as the default choice of nonlinearity between layers, but CEEDNet also takes into account the GELU (Hendrycks and Gimpel, 2016) and Mish (Misra, 2019) activation functions; both are similar to ReLU but bring a smoother decision boundary.

3.2. Preprocessing and data augmentation

Although CAUEEG evaluation tasks provide more data than previous EEG-based dementia datasets, it is still far less than those covered in other fields. For instance, ImageNet (Russakovsky et al., 2015) provides about 1 million image data, and LAION-400M (Schuhmann et al., 2021) about 400 million image-text pairs. For this reason, the data augmentation strategy is essential in training CEEDNet. At the same time, it is necessary to reduce unnecessary processes, which make the gradient flow noisy. Since some artifacts, such as motion and eye-blinking, may include helpful information for CEEDNet and the gradient flow cannot optimize the human-designed process, we did not employ the techniques such as explicit artifact correction or rejection. However, it does not mean our CEEDNet models are vulnerable to those artifacts. Our CEEDNet model learns how to correct or utilize the signals with artifacts through the training with a labeled dataset. For example, convolutional (or attention) layers in CEEDNet automatically reduce the effect of artifacts such as eye-blinking, and pooling layers select the crucial information such as artifact-free signals. Given that the neural networks are the universal function approximators (Sonoda and Murata, 2017), more complicated preprocessing, including independent component analysis, will also be trained on the networks in a data-driven manner if necessary.

In this paper, some augmentation techniques geared toward computer vision, including the horizontal flip, translation, and stretching, are not applied since they do not fit the nature of the EEG signal. The preprocessing and data augmentation steps employed in CEEDNet are described below in the order of application.

Signal crop augmentation Since the recording duration of EEG signals in CAUEEG differs from each other, CEEDNet truncates the input EEG signals to a fixed-length T to ease the difficulty of network design. Two possible options for this purpose are random cropping and event-driven cropping (*e.g.*, eye-opening and eye-closing). We chose random cropping as the preferred option in this paper since it showed better performance in an initial experiment, possibly due to the amount of augmentation.

Involving the random crop augmentation, the training objective for our model function f_w with the network parameters w can be written as:

$$\mathbb{E}^{train}_{crop}(f_w) = \int \frac{1}{N}\sum_{i=1}^{N} \ell\left(f_w(x^i_{t:t+T}), y^i\right) dp(t) \quad \text{for } t \sim \text{Uniform}\,(T_0, L^i - T), \tag{1}$$

where $x^i_{t:t+T}$ denotes the T-sized crop of ith EEG signal in the training set, y^i the label for $x^i_{t:t+T}$, ℓ the loss function, N the amount of training data, T_0 the startup transition time, and L^i the signal length of x^i, respectively. Here, we omitted the age signal for brevity. Contemplating the sensor transition time at startup, we excluded the first 10 s from cropping ($T_0 = 2000$). Note that $t \sim \text{Uniform}(T_0, L^i - T)$ brings randomness to Eq. (1), which would be helpful to the generalization (through data augmentation). Since Eq. (1) has the compute-expensive integral, we instead train our network using its approximated version:

$$\widetilde{\mathbb{E}}^{train}_{crop}(f_w) = \frac{1}{BN}\sum_{b=1}^{B}\sum_{i=1}^{N} \ell\left(f_w(x^i_{t:t+T}), y^i\right) \simeq \frac{1}{N}\sum_{i=1}^{N} \ell\left(\underbrace{\frac{1}{M}\sum^{M} f_w(x^i_{t:t+T})}_{\text{test-time augmentation}}, y^i\right), \tag{2}$$

where B denotes the number of epochs of training. Eq. (2) also shows that this random crop augmentation naturally accepts the test-time augmentation (*TTA*) with the $M \geq 1$ number of crops during the evaluation

time. When the *TTA* technique is applied, we used $M = 8$, otherwise $M = 1$ in this paper.

Although this random cropping reduces the manual process and allows more data diversity, it also introduces stochasticity to the evaluation results. However, even for past cases where random cropping was not applied (Bi and Wang, 2019; Ieracitano et al., 2019; 2020), the test accuracy varies if the test epochs are constructed with slightly different timings. In fact, given that brain activity always continues, the stochastic nature is already introduced when recording the part of it (*i.e.*, sampling), which is only invisible since it occurred outside the systems. Therefore, it can be seen that this random crop augmentation simulates the recording or segmenting process of the EEG signal. In Section 4, we repeatedly evaluate the test dataset 100 times and report the mean and standard deviation of the results. The reported standard deviation would imply stability for cropping timing, and guarantee practical interpretation and reproducibility in the real world.

Data normalization The cropped EEG signals are normalized by the mean and standard deviation calculated from the training set (z-score), as the common practice in ImageNet (Russakovsky et al., 2015). CEEDNet utilizes the averaged statistics calculated over multiple batches to relax the stochastic property caused by random signal cropping.

Random noise augmentation In order to simulate the sensor and circuit noise during training, CEEDNet augments the data by adding two kinds of random noise to the signal, both drawn from the Gaussian distributions with zero mean. The first is the multiplicative white Gaussian noise (MWGN) with standard deviation as σ_{mwgn}, which is proportional to the signal gain. Secondly, the additive white Gaussian noise (AWGN) with the standard deviation as σ_{awgn} is applied. For the age signal, we only added the AWGN with the standard deviation as σ_{awgn_age}. These random noise augmentations are considered only during the training phase.

Short-time Fourier transform For 2D models, CEEDNet prepares the input spectral image using the STFT. Starting from an input EEG sequence of length T, we chose the hyperparameters of STFT to make the resulting image as square as possible in space as follows: $win\_len = fft\_len = \left\lfloor 2\sqrt{2T} + 0.5 \right\rfloor$ and $hop\_len = \lfloor win\_len/4 + 0.5 \rfloor$. This produces an image of the spatial size of $H = \lfloor win\_len/2 \rfloor + 1$ and $W = \lfloor T/hop\_len \rfloor + 1$. For instance, this step generates an image of 64×63 spatial resolution from a 2000-length EEG sequence. This STFT preprocessing step is not applied to the 1D models.

3.3. Training

In this subsection we describe the training strategy of CEEDNet.

Regularization We retained the DropOut (Srivastava et al., 2014), BatchNorm (Ioffe and Szegedy, 2015), and LayerNorm (Ba et al., 2016) layers in their existing places from the original backbones. In order to form a smoother decision boundary and improve the generalization capability, we introduced the MixUp (Zhang et al., 2017) during training, which uses the virtual examples drawn from the vicinity distribution of the training data. The data augmentation steps described in Section 3.2 can also be considered part of the regularization.

Other details For all backbones, we trained our CEEDNet using the AdamW optimizer (Loshchilov and Hutter, 2017) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We trained the networks for 100 million samples (about 1053 epochs for CAUEEG-Dementia and 903 epochs for CAUEEG-Abnormal) unless otherwise stated. In this paper, we swept over some hyperparameters, such as p_{drop}, α_{mixup}, σ_{mwgn}, σ_{awgn}, σ_{awgn_age}, T, weight decay, peak learning rate, and learning rate schedule, via random grid search whose range was heuristically adjusted after several trials. The learning rate schedules considered are illustrated in Fig. B.1.

4. Results and discussion

This section reports the experimental results of our CEEDNet on the test sets of two CAUEEG evaluation tasks using the protocol described in Section 2.4. We then evaluate CEEDNet and other methods under the training and testing environment of *epoch-based classification*. The ablation experiments on the design choices of the proposed CEEDNet approach are also performed. Using the event information provided by the CAUEEG dataset, the impact of artifacts on performance is also investigated. Lastly, in order to understand better what the model learns, we conduct occlusion sensitivity experiments.

4.1. Experiments settings

All the models and preprocessing steps of CEEDNet were implemented with the PyTorch library (Paszke et al., 2019). We performed both training and testing using a desktop with AMD Ryzen-9 5900X CPU and a single NVIDIA GeForce RTX 3090 GPU. Taking into account the stochastic nature introduced by random cropping, this paper presents the performance of CEEDNet after averaging 100 runs.

For comparison, we trained and tested three shallow machine learning methods: K-nearest neighbors (KNN) (Fix and Hodges, 1989), Random forests (Ho, 1995), and linear SVM (Cortes and Vapnik, 1995). For KNN and Random forests, we pre-cropped each signal to multiple 2000-length epochs to train and test while not mixing the epochs between training and test sets. In this paper, we report the performance of linear SVM empowered by our random crop augmentation described in Section 3.2 since it performed better than other SVM-variants, including kernel SVMs in our experiments. The KNN and Random forests methods were implemented using scikit-learn (Pedregosa et al., 2011), and the linear SVM using PyTorch (Paszke et al., 2019). We also implemented the Ieracitano-CNN (Ieracitano et al., 2019) among the previous EEG-based diagnostic classification methods since they reported performance in both *epoch-based* and *patient-based* manners, while others only conducted the *epoch-based classification* (Bi and Wang, 2019; Ieracitano et al., 2020; Sharma et al., 2019). The Ieracitano-CNN is also relatively less reliant on cumbersome manual processes that depend on the device's characteristics.

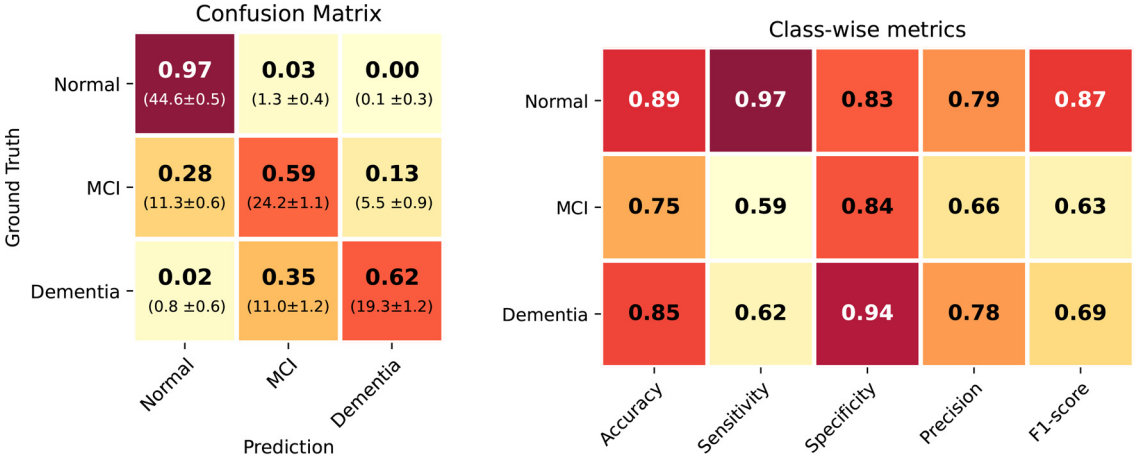
4.2. Evaluation on CAUEEG-Dementia

Table 4 summarizes the results of CEEDNet and other methods on the CAUEEG-Dementia task. Our CEEDNet models for all backbones have better accuracy than Ieracitano-CNN and shallow learning methods. Due to the random cropping augmentation applied during the training phase, our CEEDNet models get performance boosts using the *TTA* technique with multiple crops (Section 3.2). The last row is the largest CEEDNet model with 257 million parameters, composed of an ensemble of nine different CEEDNet models. The CEEDNet ensemble model shows the best accuracy of 74.66% with 22.74 EEGs per second, which is a near real-time process. The hyperparameter settings and more detailed experimental results on this task are given in Appendix (Tables B.1a and B.2, respectively). The throughput is measured as the number of EEGs processed per second. We used CPU to measure processing speed for KNN and Random forests models, whereas we used GPU for the other methods.

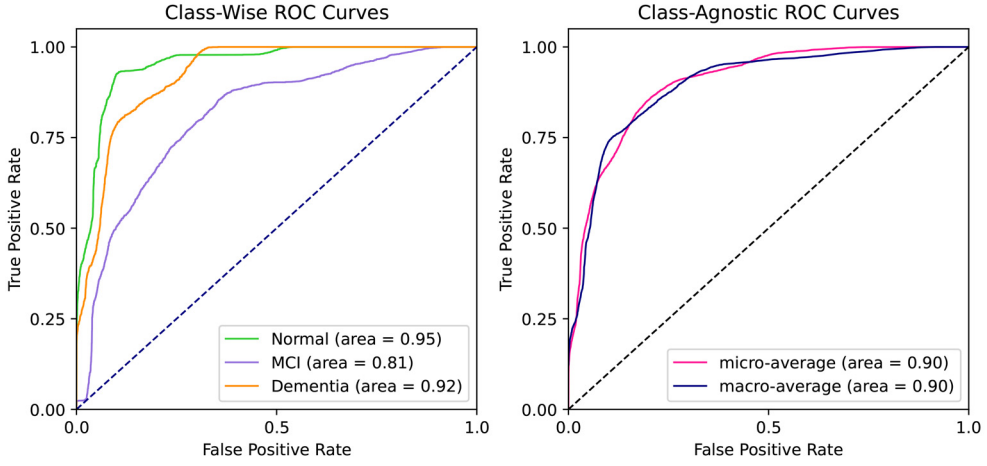
The confusion matrix, class-wise metrics, and ROC curves generated by the CEEDNet ensemble model are presented in Fig. 4. The confusion matrix of CEEDNet tells us that our model seldom misclassifies between *normal* and *dementia* and that most errors stem from concerning *mci*. It is noteworthy that our model effectively suppresses the two riskiest cases, which are in the $(1, 3)$ and $(3, 1)$ entries in the confusion matrix. Our models also in common have high specificity for *dementia* and high sensitivity for *normal*. For example, our ensemble model shows a high sensitivity of 96.98% to *normal*, which means that only 3.02% of healthy brain subjects have to afford the extra examination cost. Similarly, a specificity of 93.59% for *dementia* indicates that undemented subjects are less likely to be misdiagnosed as *dementia*. If interpreting the classification results as *normal-vs-others* outputs, our ensemble model nearly achieves 90% accuracy and 0.95 ROC-AUC score. Such a high ROC-AUC

Table 4
Experimental results on the CAUEEG-Dementia test set. *TTA* indicates the test-time augmentation technique with multiple random cropping. The best three scores are marked in bold. A More detailed version is given in Table B.2.

Method	#Params	Model size (MiB)	TTA	Throughput (EEG/s)	Test accuracy
K-Nearest Neighbors (K=5)	—	11848.7		52.42	36.80%
Random Forests (#trees=2000)	—	2932.8		808.76	46.62%
Linear SVM	0.1M	0.5		10393.40	52.33%
Ieracitano-CNN	3.5M	13.2		8172.36	54.27%
CEEDNet 1D-VGG-19	20.2M	77.2		6593.80	64.00%
CEEDNet 1D-VGG-19	20.2M	77.2	✓	931.64	67.11%
CEEDNet 1D-ResNet-18	11.4M	43.6	✓	1249.90	**68.75%**
CEEDNet 1D-ResNet-50	26.2M	100.2	✓	1013.18	67.00%
CEEDNet 1D-ResNeXt-50	25.7M	98.2	✓	702.64	68.54%
CEEDNet 2D-VGG-19	20.2M	77.1	✓	296.84	**70.18%**
CEEDNet 2D-ResNet-18	11.4M	43.7	✓	399.40	65.88%
CEEDNet 2D-ResNet-50	25.7M	98.5	✓	182.20	67.21%
CEEDNet 2D-ResNeXt-50	25.9M	99.1	✓	161.25	67.91%
CEEDNet ViT-B-16	90.1M	343.6	✓	47.31	66.18%
CEEDNet Ensemble	256.7M	981.1	✓	22.74	**74.66%**



(a) Confusion matrix (b) Class-wise metrics



(c) ROC curves and their AUC scores

Fig. 4. Performance evaluation of the CEEDNet ensemble model on the CAUEEG-Dementia test set. The confusion matrix is normalized in the row direction, and the brackets provide the mean ± std of 100 times evaluations. The class-wise sensitivity is the same as the recall. *Area* represents the ROC-AUC score. The *micro-average* becomes different from the *macro-average* due to class imbalance (Grandini et al., 2020). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Experimental results on the CAUEEG-Abnormal test set. *TTA* indicates the test-time augmentations technique of multiple random cropping. The best three scores are marked in bold. A more detailed version is given in Table B.3.

Method	#Params	Model size (MiB)	TTA	Throughput (EEG/s)	Test accuracy
K-Nearest Neighbors (K=7)	—	14015.3		41.19	51.42%
Random Forests (#trees=2000)	—	1930.5		830.80	72.63%
Linear SVM	0.1M	0.3		10363.76	68.00%
Ieracitano-CNN	3.5M	13.2		8293.08	65.98%
CEEDNet 1D-VGG-19	20.2M	77.2		7660.22	72.45%
CEEDNet 1D-VGG-19	20.2M	77.2	✓	998.54	74.28%
CEEDNet 1D-ResNet-18	11.4M	43.5	✓	844.65	74.85%
CEEDNet 1D-ResNet-50	26.3M	100.7	✓	837.66	**76.37%**
CEEDNet 1D-ResNeXt-50	25.7M	98.2	✓	800.49	**77.32%**
CEEDNet 2D-VGG-19	20.2M	77.2	✓	447.81	75.39%
CEEDNet 2D-ResNet-18	11.5M	43.8	✓	410.44	75.19%
CEEDNet 2D-ResNet-50	25.7M	98.5	✓	187.30	74.96%
CEEDNet 2D-ResNeXt-50	25.9M	99.1	✓	201.01	75.85%
CEEDNet ViT-B-16	86.9M	331.6	✓	63.99	72.70%
CEEDNet Ensemble	253.8M	969.9	✓	26.40	**79.16%**

Table 6
Experimental results under the *epoch-based classification* manner. The *epoch-based classification* was simulated using the CAUEEG-Dementia data. This experiment applies the same hyperparameters to all CEEDNet models without tuning. None of the *TTA* or ensemble techniques are adopted.

Experimental results			CEEDNet train settings	
Method	Train accruacy	Test accuracy	Hyperparmeter	Setting
k-Nearest Neighbors (k=5)	51.42%	38.26%	Crop length	2000
Random Forests (#trees=2000)	100.00%	46.28%	Age signal usage type	*age-conv*
Linear SVM	58.57%	44.08%	Photic channel usage	O
Ieracitano-CNN	100.00%	72.92%	EKG channel usage	O
CEEDNet 1D-VGG-19	100.00%	99.47%	MWGN	5.00E-02
CEEDNet 1D-ResNet-18	100.00%	99.38%	AWGN	5.00E-02
CEEDNet 1D-ResNet-50	99.97%	99.36%	AWGN (age)	5.00E-02
CEEDNet 1D-ResNeXt-50	100.00%	99.75%	MixUp	0.1
CEEDNet 2D-VGG-19	99.78%	99.37%	DropOut	0.3
CEEDNet 2D-ResNet-18	99.99%	99.82%	Weight decay	5.00E-02
CEEDNet 2D-ResNet-50	100.00%	99.83%	#FC layers	3
CEEDNet 2D-ResNeXt-50	100.00%	99.90%	Activation	GELU
CEEDNet ViT-B-16	98.12%	97.17%	Learning rate schedule	linear_decay_with_warmup

score for *normal* class means that our methods can effectively lead undiscovered patients to intensive examinations. On the other hand, except for the KNN, which has the lowest accuracy, all models struggle to distinguish *mci* compared to other classes (low class-wise accuracies for *mci* in Fig. 4 and Table B.2).

Note that this evaluation protocol does not adopt the *epoch-based classification* as explained in Section 2.4. The results produced under *epoch-based classification* will be described in Section 4.4.

4.3. Evaluation on CAUEEG-Abnormal

The task of the CAUEEG-Abnormal is a binary classification between *normal* and *abnormal*. Compared to CAUEEG-Dementia, this task's *abnormal* class includes diverse aspects beyond *mci* and *dementia*. The experimental results of this evaluation task are summarized in Table 5. Thanks to end-to-end learning, our CEEDNet models surpass in accuracy other models including Ieracitano-CNN. As in the previous case, the CEEDNet models benefit from the *TTA* and ensemble techniques. Provided only the EEG signal and age of a person, our best diagnostic model distinguishes between *normal* and *abnormal* with an accuracy of nearly 80% with 26.40 EEG/s and 253 million parameters. The overall performance of the CEEDNet ensemble model is illustrated in Fig. 5. The high sensitivity (90.39%) and low specificity (57.19%) to *abnormal* of the CEEDNet ensemble model show that the classifier prefers to minimize potential patients with undetected brain disorders even though the number of examinations rises. The hyperparameter settings for this evaluation task are listed in Table B.1b.

4.4. Epoch-based classification

As mentioned in Section 1, the previous EEG-based diagnostic systems for the dementia detection are evaluated under *epoch-based classification*. In order to compare our CEEDNet approach with previous methods, this subsection mimics the similar training and testing environment as *epoch-based classification*. From all the data in CAUEEG-Dementia, we chop each EEG sequence into multiple pieces of duration 30 s and then mix and scatter them as 8:1:1 ratios. This process produces 24417 EEG signals for training, 3052 for validation, and 3052 for testing.

Table 6 evaluates the CEEDNet and other methods under the simulated environment of *epoch-based classification*. Although the target task is identical to CAUEEG-Dementia, all CEEDNet models show extreme performance gaps between Tables 4 and 6. For instance, the CEEDNet 1D-VGG-19 model, without additional efforts, is evaluated as over 99% test accuracy under *epoch-based classification*. On the other hand, Ieracitano-CNN scored about 54% test accuracy in Table 4 (CAUEEG-Dementia) and 73% in Table 6 (*epoch-based classification*), which is almost consistent with the results reported in Ieracitano et al. (2019). Consider that our CAUEEG-Dementia is much larger in the data scale. Moreover, some techniques such as hyperparameter tuning, *TTA*, and the ensemble will bring improved results for CEEDNet.

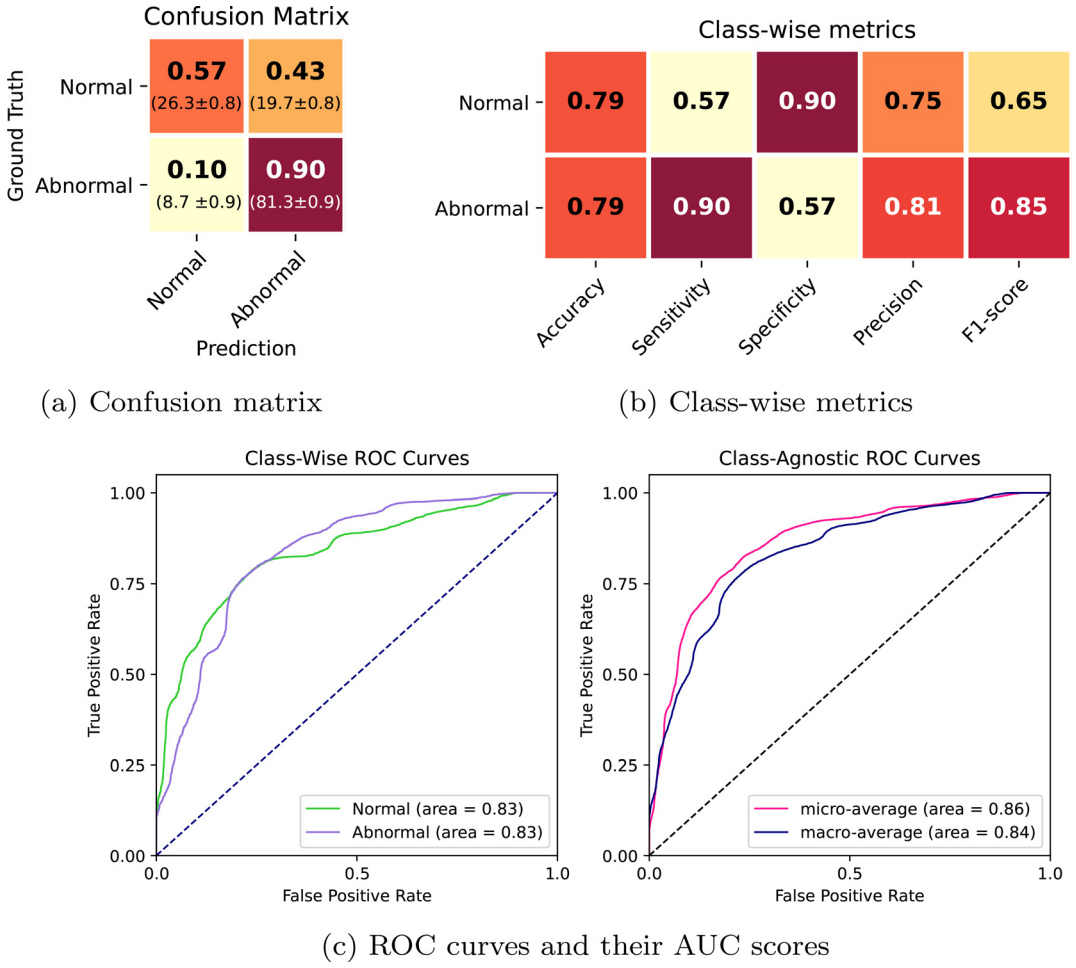


Fig. 5. Performance evaluation of the CEEDNet ensemble model on the CAUEEG-Abnormal test set. The confusion matrix is normalized in the row direction, and the brackets provide the mean ± std of 100 times evaluations. Since the CAUEEG-Abnormal is a binary classification task, the accuracies for two classes are the same. Sensitivity for *normal* is the same as the specificity for *abnormal* and vice versa. *Area* represents the ROC-AUC score. The class imbalance causes the difference between the class-wise and class-agnostic ROC curves, and *micro-* and *macro-averages* (Grandini et al., 2020). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

We conjecture that this large performance gap between our evaluation protocol and *epoch-based classification* is caused because *epoch-based classification* assumes an easy but impractical scenario. Under *epoch-based classification*, the test sample was always acquired from the same EEG recording in the training set with only slight timing differences, which means that other conditions are kept the same (not just the person but also the date, technician, device status, etc.). Therefore, the test data considered in the *epoch-based classification* differs substantially from the real-world test data, where data with all the same date, technician, diagnosis, and subject as the training data are unlikely to exist. Note that our evaluation protocol does not strictly follow the *patient-based classification* in that the number of test data with overlapping subjects with the training set ranges from 20–24%. However, we kept at least one condition changed among person, date, and even diagnosis between the training and test sets in CAUEEG-Dementia and CAUEEG-Abnormal.

Consequently, we argue that the proposed CAUEEG adopts a more trustworthy evaluation protocol than the *epoch-based classification* used in previous methods for EEG-based early detection of AD and MCI (Bi and Wang, 2019; Ieracitano et al., 2019; 2020; Sharma et al., 2019). Above all, this experiment reveals the effectiveness of deep-end-to-end learning compared with designing sophisticated signal processing analysis. The previous methods armed with manually designed feature extraction recorded about 83–95% test accuracies on much smaller data sets (12–189 EEG recordings) under *epoch-based classification*, whether their models are deep or shallow.

Table 7
Comparison of test accuracy with and without patient-overlapping data. (1) *Full* corresponds to the results in Tables 4 and 5. (2) *No-overlap* refers to the evaluation on the *no-overlap* test set (see Section 2.4). Here, the linear SVM model and the CEEDNet ensemble model with TTA used in Sections 4.2 and 4.3 are re-evaluated without modification or retraining.

		Test Accuracy		
Dataset	Model	(1) full	(2) *no-overlap*	(2) − (1)
CAUEEG-Dementia	Linear SVM	52.33%	53.50%	+1.17%
	CEEDNet	74.66%	76.66%	+2.00%
CAUEEG-Abnormal	Linear SVM	68.00%	66.02%	−1.98%
	CEEDNet	79.16%	75.77%	−3.39%

4.5. Patient overlap exclusion

As noted in Section 2.4, the CAUEEG dataset also provides the *no-overlap* version of the task annotations. This version of the dataset excludes any test data that overlaps with the patients in the training set. Table 7 compares the evaluation results between the *full* and *no-overlap* test sets. Excluding the overlapping patient data resulted in accuracy

Table 8
Comparison of test accuracy between 1D and 2D approaches. The detailed results for individual models can be seen in Tables B.2 and B.3.

Evaluation	CAUEEG-Dementia	CAUEEG-Abnormal
Average accuracy of 1D models	64.32%	73.13%
Average accuracy of 2D models	64.64%	72.45%
Accuracy of an ensemble of 1D models	70.90%	78.15%
Accuracy of an ensemble of 2D models	69.32%	77.38%
Accuracy of an ensemble of both 1D and 2D models	**73.29%**	**79.12%**

Table 9
CEEDNet ablation experiments. The CEEDNet 1D-ResNet-18 variants were evaluated on the validation set of CAUEEG-Dementia.

Elements	CEEDNet					
	Full	Ablations				
Gaussian noise	√					
MixUp	√	√				
DropOut	√	√	√			
Additional FC layers	√	√	√	√		
Age signal usage	√	√	√	√	√	
Validation accuracy	60.63%	59.64%	58.71%	57.42%	56.74%	52.46%
#Params	11.4M	11.4M	11.4M	11.4M	11.2M	11.2M

gains for CAUEEG-Dementia and losses for CAUEEG-Abnormal for both the CEEDNet ensemble and the linear SVM models. If we cautiously speculate on the reason, it is possible that the excluded data happened to be more difficult for CAUEEG-Dementia and easier for CAUEEG-Abnormal.

Despite this observed trend, it is important to note that the experiment demonstrates that patient overlap did not unfairly bias or inflate the results presented in Sections 4.2 and 4.3. Detailed experimental results for the *no-overlap* test sets can be found in Tables B.4 and B.5.

4.6. Basic operation dimension and ensemble

In Figs. 4 and 5, we tried both 1D and 2D approaches to deal with EEG signals. Table 8 compares the performance of 1D models, 2D models, and their ensembles. The differences in average accuracy between the 1D and 2D approaches were 0.32% for CAUEEG-Dementia and 0.68% for CAUEEG-Abnormal, respectively. In other words, the individual models showed a slight difference in performance between the 1D and 2D approaches. The ensembles of 1D models or 2D models significantly improved the accuracy in two tasks, and the best performance was observed in the ensemble of both 1D and 2D models. In general, the effectiveness of ensembling is maximized when the variance in different models is high. The main difference between 1D and 2D models is how the receptive fields of the networks expand as the layer deepens. In other words, the order in which the networks look into the information on EEG signals differs. We conjecture that these two approaches' differences in information processing maximized the ensembling effect.

4.7. Ablation study

In deep learning research, an ablation study inspects performance changes by removing architectural or design choices to discover the impact of each component. The CEEDNet with the 1D-ResNet-18 backbone and its ablated variants were trained for 100M samples and evaluated on the validation set of CAUEEG-Dementia. Table 9 reports the experimental results. The usage of the age signal showed a critical effect on accuracy (57.74% with and 52.46% without the age signal), which is consistent with the report in Livingston et al. (2020). According to our observations, the *age-fc* and *age-conv* options had no noticeable difference in performance. The additional FC layers and other regularization methods bought gradual improvements.

In this paper, we implemented the CEEDNet using nine different backbones. These backbones can be divided into 1D CNN, 2D CNN, and Transformers. Combining the experimental results in Tables 4 to 6, some interesting points can be noticed. At first, 1D and 2D models are on par with each other in performance. This indicates that it is crucial to foster end-to-end learning conditions rather than extracting good input feature representations in drawing the power of deep models. Secondly, ViT (Transformers) showed relatively low accuracy and also underfitting tendency. We conjecture that longer training and more data are needed to derive the attention mechanism's capability. Lastly, an ensemble model surpassed all the single models with a noticeable difference, which indicates that there remains room to improve for individual models.

Table 10
Experimental results to inspect the effect of signal regions with artifacts to the classification models. The linear SVM and the CEEDNet 1D-ResNet-18 models trained using the original random signal crop variable t and the event-rejection signal crop variable t' were compared. Except for the signal crop variable, the other configurations were the same as Section 4.2 in this experiment.

Model	Crop variable	Validation accuracy	Rel. diff.
Linear SVM	Original t	50.37%	—
	Event-rejection t'	50.85%	+0.95%
CEEDNet	Original t	60.48%	—
	Event-rejection t'	58.59%	−3.13%

4.8. Impact of artifact events

In order to investigate the impact of artifacts caused by the subjects' movements on our CEEDNet model, we conducted an additional experiment. As shown in Table 2, the CAUEEG dataset includes the event history information during recording. For random sampling without movement events, we changed the random cropping variable t in Eqs. (1) and (2) to t' with zero probability of cropping around the region where the events occurred using that information, resulting in the event-rejection random sampling. The results of comparing the validation accuracy of the linear SVM and CEEDNet models using the original t and modified t' are shown in Table 10. When there is no movement, the shallow linear SVM's accuracy improves by about 1%. On the other hand, the accuracy of our deeper CEEDNet model is reduced by about 3%. From these results for our deep end-to-end model, we conjecture that the penalty for reducing the usable signal area during training out-

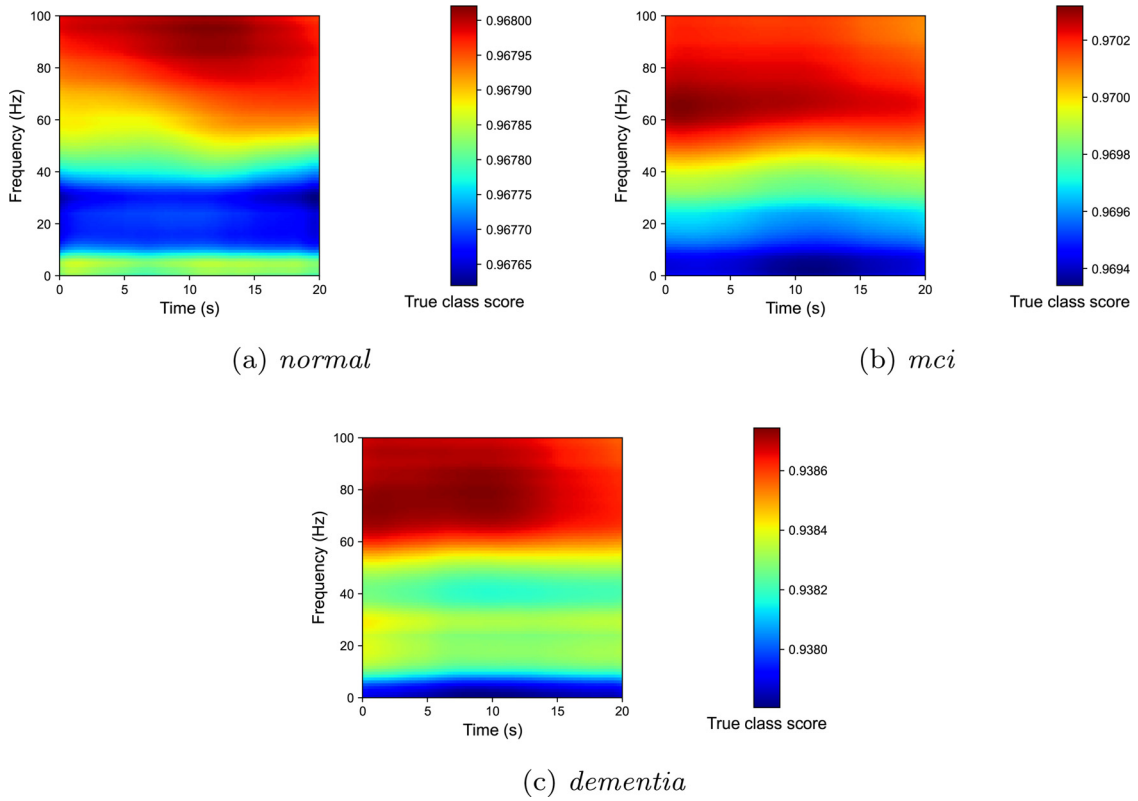


Fig. 6. Channel-agnostic occlusion sensitivity analysis of the CEEDNet 2D-VGG-19 model on the CAUEEG-Dementia training set. During classification, all channels in the input time-frequency feature map were simultaneously occluded by a sliding rectangle whose width and height are quarters of those of the input. These results were generated by averaging all results by class of the training set. The class score for the correct class is visualized in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

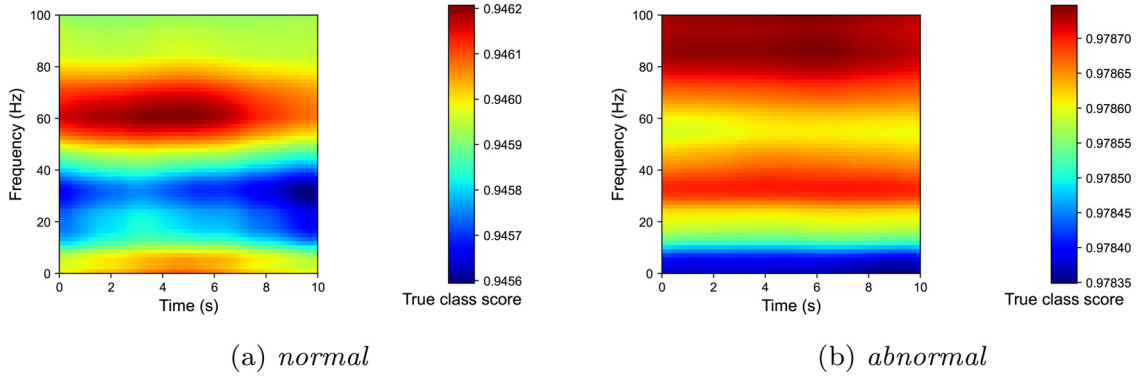


Fig. 7. Channel-agnostic occlusion sensitivity analysis of the CEEDNet 2D-VGG-19 model on the CAUEEG-Abnormal training set. During classification, all channels in the input time-frequency feature map were simultaneously occluded by a sliding rectangle whose width and height are quarters of those of the input. These results were generated by averaging all results by class of the training set. The class score for the correct class is visualized in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

weighs the benefit of lowering artifacts' impact. It seems that the CEEDNet learned from the entire range of EEG signals regardless of the presence or absence of artifacts, but at least it was not biased by artifacts per se.

4.9. Occlusion sensitivity analysis

The occlusion sensitivity was initially introduced in computer vision to identify whether the classifier network is sensitive to specific locations or whether the overall context is essential (Zeiler and Fergus, 2014). In recent years, occlusion sensitivity analysis has been used to inspect which part of the time-frequency representation of the EEG signal affects the classifier outputs (Roy et al., 2019). In this paper, we performed the occlusion sensitivity analysis with the 2D-VGG-19 models trained on CAUEEG-Dementia and CAUEEG-Abnormal. We systemically moved the occlusion box from the top left to the bottom right while estimating the classification scores of our model on the "occluded" input time-frequency map. The occlusion test was conducted by inserting a zero-out operation right after the input data normalization described in Section 3.2. We set the height and width of the occlusion box as a fourth of those of the input time-frequency feature map size.

Fig. 6 visualizes the experimental results of an occlusion sensitivity analysis performed channel-agnostic on the CAUEEG-Dementia task. In this experiment, the low-scoring (bluish) areas, which aggravate the

performance, indicate that they are critical to classify the input EEG correctly. No matter the class label, occluding a high frequency over 60 Hz rarely affected the class score estimation in CAUEEG-Dementia. The narrow area around the delta (< 4 Hz) and theta (4–7 Hz) bands showed the most remarkable influences on the classification results for *dementia*, while the *normal* class was sensitive to the wide region around alpha (8–12 Hz) and beta (13–30 Hz) waves. The occlusion sensitivity for *mci* seems to be intermediate from *normal* to *dementia* in this analysis. Note that this occlusion sensitivity experiment cannot reflect all nature of highly nonlinear deep CEEDNet models (*e.g.*, a single moving rectangle cannot simulate the combinatorial properties of distant feature relationships). Nonetheless, our results accord with previous findings that the EEG slow waves (theta and delta) activity of dementia patients is strengthened while the high-frequency waves (alpha, beta, and gamma) diminish (Cassani et al., 2018; Malek et al., 2017).

The results of the occlusion sensitivity test in CAUEEG-Abnormal are shown in Fig. 7. The low-scoring area for *normal* in CAUEEG-Abnormal seems narrower than the area for *normal* in CAUEEG-Dementia. Given that the training data for the *normal* classes in the two tasks are the same as each other, this difference is interesting. We conjecture that this is caused because the *abnormal* label in CAUEEG-Abnormal covers more diagnostic labels than the union of *dementia* and *mci* in CAUEEG-Dementia. In other words, the learned manifold for the *abnormal* class in CAUEEG-Abnormal would occupy more space or have multiple facets, which is plausible in Fig. 7b. Additional results for the channel-wise occlusion sensitivity analysis can be seen in Appendix C.

5. Conclusion

In this work, we present the CAUEEG dataset and its corresponding two evaluation tasks to aid the active research on automatic EEG diagnosis. Our tasks utilize a more practical evaluation protocol than the *epoch-based classification* used in previous works. Based on CAUEEG, we also propose a new end-to-end CEEDNet approach for the automatic early detection of dementia. To demonstrate the performance of the CEEDNet, we tested the shallow methods and Iercitano's CNN model. On the CAUEEG-Dementia and CAUEEG-Abnormal tasks, CEEDNet produced a significant improvement in accuracy and other metrics compared with existing methods. Our CEEDNet was tested with various variations to analyze the effect of model architecture and size. Extensive experiments suggest that, like in other research fields, deep models for automatic EEG diagnosis overwhelm the shallow models equipped with elaborately designed preprocessing when given sufficient end-to-end learning conditions. On the classification problems between *normal, mci*, and *dementia* and between *normal* and *abnormal*, the high ROC-AUC scores (0.9 and 0.86, respectively) recorded by the best CEEDNet model demonstrates that our method can lead potential patients to early diagnosis through automatic screening. In the future, we would like to advance the work in three directions: i) scaling up the CAUEEG dataset, ii) introducing the unsupervised/semi-supervised learning strategy, and iii) improving the performance of Transformer-based backbones.

Data and code availability statement

*Data Availability The CAUEEG dataset for this study is available in a public repository at https://github.com/ipis-mjkim/caueeg-dataset.

*Code Availability The codes for using the CAUEEG dataset, implementing CEEDNet, and generating the experimental results are available in a public repository at https://github.com/ipis-mjkim/caueeg-ceednet.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Min-jae Kim: Methodology, Software, Writing – original draft. **Young Chul Youn:** Conceptualization, Resources, Data curation, Validation. **Joonki Paik:** Project administration, Supervision, Writing – original draft, Funding acquisition.

Acknowledgments

This work was supported partly by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by Korea government (MSIT) [2021-0-01341, Artificial Intelligent Graduate School Program and 2021M3I1A1097911], partly by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No.2014-0-00077, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis], and partly by research grant from Biomedical Research Institute, Chung-Ang University Hospital (2021).

Appendix A. CAUEEG dataset and evaluation tasks details

Figs. A.1 and A.2 illustrate the preview of the JSON information files for the CAUEEG dataset and its two evaluation tasks, respectively. These information files are attached to the dataset.

CAUEEG dataset


```
{
    dataset_name:
        CAUEEG dataset
    signal_header:
        Fp1-AVG
        F3-AVG
        C3-AVG
        .
        .
        .
        Photic
    data:
        {'serial': '00001', 'age': 78, 'symptom': ['mci', 'mci_amnestic', 'mci_amnestic_rf']}
        {'serial': '00002', 'age': 56, 'symptom': ['normal', 'smi']}
        {'serial': '00003', 'age': 93, 'symptom': ['mci', 'mci_vascular']}
        .
        .
        .
        {'serial': '01388', 'age': 73, 'symptom': ['mci', 'mci_amnestic', 'mci_amnestic_ef']}
}
```


Fig. A.1. Preview of CAUEEG dataset annotation. All the necessary information, including task description and diagnostic labels, is saved in the JSON file format.

CAUEEG-Dementia Benchmark

{
task_name:
CAUEEG-Dementia benchmark
task_description:
Classification of [Normal], [MCI], and [Dementia] symptoms.
class_label_to_name:
['Normal', 'MCI', 'Dementia']
class_name_to_label:
{'Normal': 0, 'MCI': 1, 'Dementia': 2}
train_split:
{'serial': '00587', 'age': 53, 'symptom': ['normal', 'cb_normal'], 'class_name': 'Normal', 'class_label': 0}
{'serial': '01301', 'age': 88, 'symptom': ['dementia', 'ad', 'load'], 'class_name': 'Dementia', 'class_label': 2}
{'serial': '00781', 'age': 68, 'symptom': ['mci', 'mci_amnestic', 'mci_amnestic_rf'], 'class_name': 'MCI', 'class_label': 1}
.
.
.
{'serial': '00153', 'age': 64, 'symptom': ['mci', 'mci_non_amnestic'], 'class_name': 'MCI', 'class_label': 1}
validation_split:
{'serial': '00341', 'age': 80, 'symptom': ['dementia', 'ad', 'load'], 'class_name': 'Dementia', 'class_label': 2}
{'serial': '00510', 'age': 82, 'symptom': ['dementia', 'vd', 'sivd'], 'class_name': 'Dementia', 'class_label': 2}
{'serial': '00172', 'age': 86, 'symptom': ['mci', 'mci_amnestic'], 'class_name': 'MCI', 'class_label': 1}
.
.
.
{'serial': '01303', 'age': 52, 'symptom': ['normal', 'smi'], 'class_name': 'Normal', 'class_label': 0}
test_split:
{'serial': '00789', 'age': 62, 'symptom': ['normal', 'cb_normal'], 'class_name': 'Normal', 'class_label': 0}
{'serial': '00934', 'age': 61, 'symptom': ['normal', 'cb_normal'], 'class_name': 'Normal', 'class_label': 0}
{'serial': '00142', 'age': 70, 'symptom': ['normal', 'cb_normal'], 'class_name': 'Normal', 'class_label': 0}
.
.
.
{'serial': '00520', 'age': 55, 'symptom': ['normal', 'smi'], 'class_name': 'Normal', 'class_label': 0}
}

(a) CAUEEG-Dementia evaluation task

CAUEEG-Abnormal Benchmark

{
task_name:
CAUEEG-Abnormal benchmark
task_description:
Classification of [Normal] and [Abnormal] symptoms
class_label_to_name:
['Normal', 'Abnormal']
class_name_to_label:
{'Normal': 0, 'Abnormal': 1}
train_split:
{'serial': '01258', 'age': 77, 'symptom': ['dementia', 'vd', 'sivd'], 'class_name': 'Abnormal', 'class_label': 1}
{'serial': '00836', 'age': 80, 'symptom': ['normal', 'smi'], 'class_name': 'Normal', 'class_label': 0}
{'serial': '00761', 'age': 75, 'symptom': ['dementia', 'ad', 'load'], 'class_name': 'Abnormal', 'class_label': 1}
.
.
.
{'serial': '00105', 'age': 71, 'symptom': ['normal', 'smi'], 'class_name': 'Normal', 'class_label': 0}
validation_split:
{'serial': '00152', 'age': 81, 'symptom': ['dementia', 'ad', 'load'], 'class_name': 'Abnormal', 'class_label': 1}
{'serial': '01034', 'age': 81, 'symptom': ['mci', 'mci_amnestic', 'mci_amnestic_ef'], 'class_name': 'Abnormal', 'class_label': 1}
{'serial': '00296', 'age': 71, 'symptom': ['tga'], 'class_name': 'Abnormal', 'class_label': 1}
.
.
.
{'serial': '00684', 'age': 65, 'symptom': ['normal', 'cb_normal'], 'class_name': 'Normal', 'class_label': 0}
test_split:
{'serial': '00560', 'age': 57, 'symptom': ['normal', 'cb_normal'], 'class_name': 'Normal', 'class_label': 0}
{'serial': '01156', 'age': 62, 'symptom': ['mci', 'mci_amnestic', 'mci_amnestic_rf'], 'class_name': 'Abnormal', 'class_label': 1}
{'serial': '00098', 'age': 62, 'symptom': ['normal', 'smi'], 'class_name': 'Normal', 'class_label': 0}
.
.
.
{'serial': '00888', 'age': 52, 'symptom': ['normal', 'cb_normal'], 'class_name': 'Normal', 'class_label': 0}
}

(b) CAUEEG-Abnormal evaluation task

Fig. A.2. Preview of evaluation task annotations. All the necessary information, including task descriptions, data splits, and labels, is saved in the JSON file format for each evaluation task.

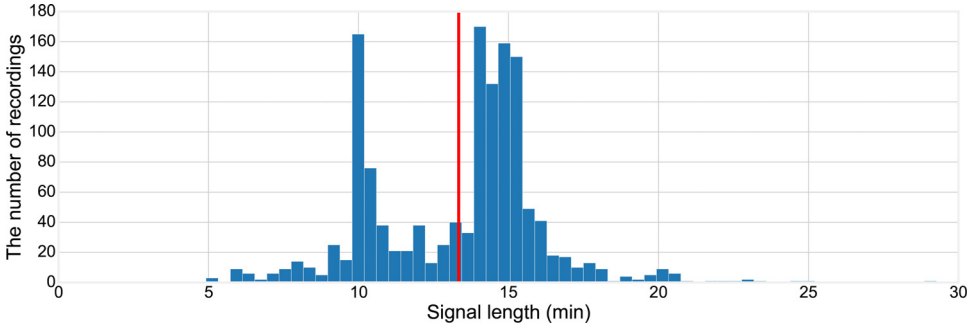


Fig. A.3. Histogram of sequence duration in minutes for all EEG recordings in the CAUEEG dataset. The red vertical line represents the mean length. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figs. A.3 and A.4 show the histograms of signal length and owner's age of all EEG recordings in the CAUEEG dataset, respectively.

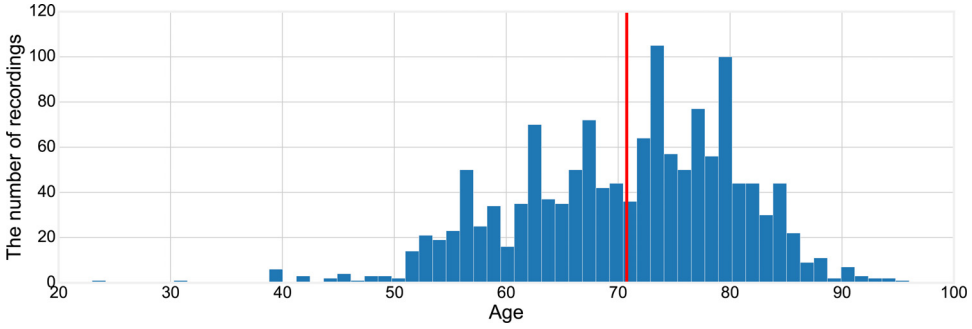


Fig. A.4. Histogram of patients' age for all EEG recordings in the CAUEEG dataset. The red vertical line denotes the mean age. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix B. Experimental details

This section supplements the training details and experimental results of CEEDNet. The detailed hyperparameter settings for CEEDNet are summarized in Table B.1. In this paper, we tried to train our CEEDNet using as various hyperparameter combinations as possible for each task and backbone. Thus, the hyperparameter settings shown in Table B.1 are not necessarily optimal. The learning rate schedules considered for the training of CEEDNet are depicted in Fig. B.1. Tables B.2 and B.3 provide the detailed experimental results of those in Sections 4.2 and 4.3, respectively. Tables B.4 and B.5 also show the detailed evaluation results on the *no-overlap* version of test sets, which are described in Section 2.4.

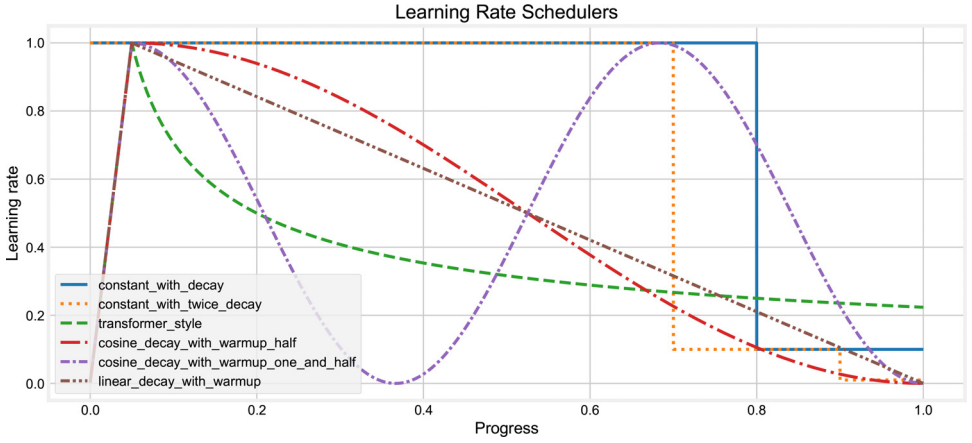


Fig. B.1. Learning rate schedules considered for training the CEEDNet models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table B.1
Hyperparameter settings for training the CEEDNet models on CAUEEG-Dementia and CAUEEG-Abnormal.

Hyperparmeters	1D-VGG-19	1D-ResNet-18	1D-ResNet-50	1D-ResNeXt-50	2D-VGG-19	2D-ResNet-18	2D-ResNet-50	2D-ResNeXt-50	ViT-B-16
Crop length	2000	2000	1000	4000	4000	2000	2000	4000	12000
Age signal usage type	*age-conv*	*age-conv*	*age-fc*	*age-fc*	*age-conv*	*age-fc*	*age-conv*	*age-fc*	*age-conv*
Photic channel usage	×	O	O	×	×	×	×	O	×
EKG channel usage	O	O	×	O	O	O	O	×	O
MWGN	1.10E-02	0.00E+00	4.30E-03	4.02E-02	3.68E-03	3.23E-02	1.10E-02	5.63E-02	8.39E-02
AWGN	8.12E-03	0.00E+00	1.25E-02	3.75E-02	4.79E-02	7.46E-02	8.12E-03	1.04E-01	2.96E-02
AWGN (age)	1.52E-03	0.00E+00	1.22E-02	1.15E-01	3.11E-02	2.78E-02	1.52E-03	1.80E-02	1.34E-01
MixUp	0.1	0.2	0.2	0.1	0.3	0.2	0.1	0.0	0.3
DropOut	0.2930	0.3000	0.2689	0.2722	0.2015	0.4970	0.2930	0.0420	0.4234
Weight decay	7.06E-02	4.39E-02	4.04E-05	4.82E-02	1.25E-02	1.34E-03	7.06E-02	1.59E-02	2.76E-02
#FC layers	3	3	3	3	2	2	3	5	2
Activation	ReLU	GELU	GELU	GELU	GELU	ReLU	ReLU	Mish	GELU
Minibatch size	256	512	192	256	128	192	256	192	48
#Total train samples	1.00E+08	1.00E+08	3.20E+07	5.00E+07	1.00E+08	3.20E+07	1.00E+08	3.20E+07	1.00E+08
Base learning rate	4.69E-04	4.69E-04	1.42E-03	1.90E-03	8.48E-05	1.07E-03	4.69E-04	2.53E-03	3.18E-05
Learning rate schedule	cosine_decay_with_warmup_half	cosine_decay_with_warmup_half	constant_with_twice_decay	cosine_decay_with_warmup_half	constant_with_decay	cosine_decay_with_warmup_one_and_half	cosine_decay_with_warmup_half	constant_with_twice_decay	constant_with_twice_decay

(a) CAUEEG-Dementia evaluation task

Crop length	2000	3000	4000	3000	2000	3000	4000	1000	4000
Age signal usage type	*age-conv*	*age-fc*	*age-fc*	*age-conv*	*age-conv*	*age-conv*	*age-fc*	*age-conv*	*age-conv*
Photic channel usage	×	O	O	O	×	O	×	O	O
EKG channel usage	×	O	O	×	O	×	O	×	O
MWGN	1.53E-02	1.32E-02	2.92E-02	7.95E-02	1.17E-02	2.85E-04	8.59E-02	8.42E-02	5.09E-02
AWGN	1.13E-01	9.95E-02	1.16E-01	8.05E-02	5.69E-02	9.88E-02	1.75E-02	1.96E-02	1.14E-01
AWGN (age)	1.30E-01	1.71E-01	9.64E-02	2.69E-01	1.78E-01	1.84E-01	1.35E-01	2.18E-01	9.66E-02
MixUp	0.0	0.1	0.3	0.1	0.2	0.2	0.3	0.3	0.2
DropOut	0.1393	0.4512	0.1559	0.0041	0.0049	0.3305	0.1144	0.1451	0.3886
Weight decay	3.39E-02	1.44E-04	4.71E-02	2.86E-02	1.17E-02	9.32E-03	3.77E-02	2.98E-05	3.47E-02
#FC layers	4	3	4	3	3	5	2	5	2
Activation	GELU	ReLU	GELU	GELU	ReLU	GELU	GELU	GELU	ReLU
Minibatch size	512	256	320	256	256	192	128	64	48
#Total train samples	4.00E+07	4.00E+07	4.00E+07	4.00E+07	1.00E+08	4.00E+07	1.00E+08	4.00E+07	1.00E+08
Base learning rate	6.75E-04	2.53E-03	1.33E-04	1.90E-03	1.70E-04	3.37E-03	3.29E-04	2.00E-04	8.80E-05
Learning rate schedule	constant_with_decay	transformer_style	constant_with_decay	constant_with_decay	linear_decay_with_warmup	transformer_style	transformer_style	constant_with_twice_decay	transformer_style

(b) CAUEEG-Abnormal evaluation task

Table B.2
Detailed experimental results on the CAUEEG-Dementia test set. *TTA* indicates the test-time augmentations technique of multiple random cropping. The best three scores are marked in bold.

Method	Model size (MiB)	TTA	Test accuracy	Class-wise accuracy			Sensitivity			Specificity		
				Normal	MCI	Dementia	Normal	MCI	Dementia	Normal	MCI	Dementia
k-Nearest Neighbors (k=5)	11848.7		36.80%	47.58%	58.46%	68.49%	74.00%	21.01%	11.99%	32.28%	79.96%	88.63%
Random Forests (#trees=2000)	2932.8		46.62%	71.12%	50.08%	72.34%	32.01%	**82.00%**	15.99%	**91.24%**	30.01%	93.28%
Linear SVM	0.5		52.33%	73.47%	59.06%	72.14%	77.15%	39.44%	32.55%	71.11%	69.50%	86.25%
Ieracitano-CNN	13.2		54.27%	70.86%	61.51%	76.17%	76.35%	37.14%	44.15%	67.35%	74.48%	87.58%
CEEDNet 1D-VGG-19	77.2		64.00%	81.47%	68.50%	78.03%	77.07%	49.26%	**64.09%**	84.28%	78.74%	83.00%
CEEDNet 1D-VGG-19	77.2	✓	67.11%	84.02%	70.67%	79.54%	79.66%	52.99%	**67.17%**	86.81%	80.08%	83.95%
CEEDNet 1D-ResNet-18	43.6		65.93%	82.64%	70.00%	79.22%	83.56%	51.10%	59.38%	82.06%	80.06%	86.29%
CEEDNet 1D-ResNet-18	43.6	✓	68.75%	84.26%	72.54%	80.69%	87.25%	54.26%	60.44%	82.34%	82.28%	87.91%
CEEDNet 1D-ResNet-50	100.2		61.83%	79.95%	65.95%	77.77%	85.87%	45.81%	47.35%	76.17%	76.66%	88.61%
CEEDNet 1D-ResNet-50	100.2	✓	67.00%	83.19%	70.00%	80.82%	93.36%	50.26%	50.04%	76.69%	80.51%	91.79%
CEEDNet 1D-ResNeXt-50	98.2		65.53%	83.66%	68.72%	78.68%	72.46%	62.03%	59.87%	**90.81%**	72.28%	85.39%
CEEDNet 1D-ResNeXt-50	98.2	✓	68.54%	**85.30%**	71.14%	80.63%	74.17%	**66.28%**	**63.15%**	**92.41%**	73.72%	86.86%
CEEDNet 2D-VGG-19	77.1		68.26%	83.13%	71.45%	81.95%	91.48%	59.69%	45.15%	77.80%	77.71%	**95.06%**
CEEDNet 2D-VGG-19	77.1	✓	**70.18%**	84.20%	**73.39%**	82.77%	92.61%	**65.15%**	43.56%	78.83%	77.78%	**96.74%**
CEEDNet 2D-ResNet-18	43.7		63.51%	79.78%	68.78%	78.46%	89.63%	50.46%	42.02%	73.49%	78.54%	91.44%
CEEDNet 2D-ResNet-18	43.7	✓	65.88%	80.98%	71.28%	79.51%	92.43%	54.32%	41.78%	73.66%	80.31%	92.96%
CEEDNet 2D-ResNet-50	98.5		65.67%	79.68%	69.03%	82.63%	84.65%	53.41%	53.73%	76.50%	77.35%	92.93%
CEEDNet 2D-ResNet-50	98.5	✓	67.21%	81.40%	69.18%	**83.84%**	87.24%	55.43%	53.07%	77.67%	76.50%	**94.81%**
CEEDNet 2D-ResNeXt-50	99.1		64.47%	81.13%	67.59%	80.21%	88.07%	48.46%	50.63%	76.70%	77.78%	90.75%
CEEDNet 2D-ResNeXt-50	99.1	✓	67.91%	83.35%	70.75%	81.73%	92.37%	51.38%	53.50%	77.59%	81.06%	91.79%
CEEDNet ViT-B-16	343.6		61.30%	79.09%	67.35%	76.15%	87.93%	37.62%	53.10%	73.45%	83.17%	84.37%
CEEDNet ViT-B-16	343.6	✓	66.18%	83.56%	69.55%	79.26%	**97.04%**	37.58%	58.22%	74.94%	**86.57%**	86.76%
CEEDNet Ensemble	981.1		**73.29%**	**87.73%**	**74.81%**	**84.04%**	**95.21%**	58.90%	59.81%	82.96%	**83.29%**	92.67%
CEEDNet Ensemble	981.1	✓	**74.66%**	**88.61%**	**75.38%**	**85.32%**	**96.98%**	59.09%	62.12%	83.27%	**84.06%**	93.59%

Table B.3
Detailed experimental results on the CAUEEG-Abnormal test set. *TTA* indicates the test-time augmentations technique of multiple random cropping. Since the CAUEEG-Abnormal is a binary classification task, sensitivity for *normal* is the same as the specificity for *abnormal* and vice versa. The best three scores are marked in bold.

Method	Model size (MiB)	TTA	Test accuracy	Sensitivity	
				Normal	Abnormal
K-Nearest Neighbors (K=7)	14015.3		51.42%	59.00%	48.00%
Random Forests (#trees=2000)	1930.5		72.63%	10.00%	**97.00%**
Linear SVM	0.3		68.00%	51.18%	76.60%
Ieracitano-CNN	13.2		65.98%	52.26%	72.98%
CEEDNet 1D-VGG-19	77.2		72.45%	51.65%	83.08%
CEEDNet 1D-VGG-19	77.2	✓	74.28%	50.75%	86.30%
CEEDNet 1D-ResNet-18	43.5		73.60%	46.78%	87.31%
CEEDNet 1D-ResNet-18	43.5	✓	74.85%	46.96%	89.11%
CEEDNet 1D-ResNet-50	100.7		74.11%	56.10%	83.32%
CEEDNet 1D-ResNet-50	100.7	✓	76.37%	56.09%	86.74%
CEEDNet 1D-ResNeXt-50	98.2		72.37%	46.65%	85.52%
CEEDNet 1D-ResNeXt-50	98.2	✓	**77.32%**	47.93%	**92.34%**
CEEDNet 2D-VGG-19	77.2		72.13%	57.76%	79.47%
CEEDNet 2D-VGG-19	77.2	✓	75.39%	**61.46%**	82.52%
CEEDNet 2D-ResNet-18	43.8		73.50%	**63.29%**	78.73%
CEEDNet 2D-ResNet-18	43.8	✓	75.19%	**65.95%**	79.91%
CEEDNet 2D-ResNet-50	98.5		73.06%	54.28%	82.66%
CEEDNet 2D-ResNet-50	98.5	✓	74.96%	55.11%	85.10%
CEEDNet 2D-ResNeXt-50	99.1		72.84%	57.73%	80.55%
CEEDNet 2D-ResNeXt-50	99.1	✓	75.85%	59.98%	83.96%
CEEDNet ViT-B-16	331.6		70.70%	36.31%	88.27%
CEEDNet ViT-B-16	331.6	✓	72.70%	33.91%	**92.52%**
CEEDNet Ensemble	969.9		**79.12%**	56.98%	90.44%
CEEDNet Ensemble	969.9	✓	**79.16%**	57.19%	90.39%

Table B.4
Detailed experimental results on the CAUEEG-Dementia *no-overlap* test set. The models used in Section 4.2 are re-evaluated on the *no-overlap* test set. *TTA* indicates the test-time augmentations technique of multiple random cropping. The best three scores are marked in bold.

Method	TTA	Test accuracy (no-overlap)	Class-wise accuracy			Sensitivity			Specificity		
			Normal	MCI	Dementia	Normal	MCI	Dementia	Normal	MCI	Dementia
Linear SVM		53.50%	74.41%	59.60%	72.98%	76.54%	41.84%	34.34%	73.05%	69.89%	85.49%
Ieracitano-CNN		52.64%	69.52%	59.78%	75.98%	74.75%	37.42%	40.28%	66.20%	72.72%	87.52%
CEEDNet 1D-VGG-19		61.79%	80.71%	66.83%	76.04%	74.34%	48.08%	**62.41%**	84.76%	77.69%	80.46%
CEEDNet 1D-VGG-19	✓	65.01%	83.74%	68.49%	77.79%	77.29%	50.73%	**66.91%**	**87.85%**	78.77%	81.31%
CEEDNet 1D-ResNet-18		66.51%	82.19%	71.08%	79.76%	83.88%	53.79%	57.97%	81.12%	81.09%	86.81%
CEEDNet 1D-ResNet-18	✓	70.02%	83.87%	74.13%	82.04%	86.97%	57.61%	61.68%	81.89%	83.70%	88.63%
CEEDNet 1D-ResNet-50		59.79%	78.43%	64.19%	76.95%	84.83%	44.73%	42.53%	74.36%	75.46%	88.08%
CEEDNet 1D-ResNet-50	✓	64.44%	82.68%	67.69%	78.52%	93.97%	49.94%	39.23%	75.49%	77.96%	91.24%
CEEDNet 1D-ResNeXt-50		64.81%	82.74%	68.42%	78.47%	68.79%	**63.85%**	59.94%	**91.61%**	71.06%	84.47%
CEEDNet 1D-ResNeXt-50	✓	67.11%	83.90%	70.52%	79.80%	69.91%	**67.45%**	**62.14%**	**92.80%**	72.30%	85.51%
CEEDNet 2D-VGG-19		67.32%	82.47%	71.40%	80.77%	91.41%	59.78%	40.29%	76.78%	78.12%	**93.86%**
CEEDNet 2D-VGG-19	✓	**70.13%**	**84.28%**	**74.16%**	81.83%	92.97%	**66.55%**	39.18%	78.75%	78.56%	**95.63%**
CEEDNet 2D-ResNet-18		61.71%	79.33%	67.93%	76.16%	90.03%	49.06%	35.64%	72.53%	78.85%	89.27%
CEEDNet 2D-ResNet-18	✓	64.51%	81.18%	71.06%	76.79%	93.54%	54.42%	33.45%	73.31%	80.68%	90.81%
CEEDNet 2D-ResNet-50		64.11%	78.58%	67.43%	82.21%	83.70%	51.63%	51.66%	75.32%	76.58%	92.09%
CEEDNet 2D-ResNet-50	✓	65.94%	80.21%	67.96%	**83.72%**	85.91%	54.18%	51.82%	76.58%	75.93%	**94.04%**
CEEDNet 2D-ResNeXt-50		62.50%	80.44%	65.25%	79.30%	88.26%	45.11%	47.59%	75.46%	76.92%	89.56%
CEEDNet 2D-ResNeXt-50	✓	65.00%	82.44%	67.37%	80.19%	92.60%	45.76%	49.95%	75.98%	79.88%	89.97%
CEEDNet ViT-B-16		60.34%	78.53%	66.99%	75.16%	86.59%	41.08%	47.47%	73.40%	81.99%	84.12%
CEEDNet ViT-B-16	✓	65.44%	83.07%	69.59%	78.23%	**96.43%**	41.64%	51.86%	74.56%	**85.77%**	86.76%
CEEDNet Ensemble		**73.96%**	**89.03%**	**75.40%**	**83.50%**	**96.73%**	60.97%	57.23%	84.13%	**83.75%**	91.99%
CEEDNet Ensemble	✓	**76.66%**	**91.21%**	**76.88%**	**85.22%**	**99.03%**	63.58%	60.68%	86.24%	**84.58%**	93.16%

Table B.5
Detailed experimental results on the CAUEEG-Abnormal *no-overlap* test set. The models used in Section 4.3 are re-evaluated on the *no-overlap* test set. *TTA* indicates the test-time augmentations technique of multiple random cropping. Since the CAUEEG-Abnormal is a binary classification task, sensitivity for *normal* is the same as the specificity for *abnormal* and vice versa. The best three scores are marked in bold.

Method	TTA	Test accuracy (no-overlap)	Sensitivity	
			Normal	Abnormal
Linear SVM		66.02%	43.99%	77.34%
Ieracitano-CNN		63.49%	49.14%	70.87%
CEEDNet 1D-VGG-19		68.97%	46.96%	80.28%
CEEDNet 1D-VGG-19	✓	69.61%	43.38%	83.10%
CEEDNet 1D-ResNet-18		71.40%	41.68%	86.68%
CEEDNet 1D-ResNet-18	✓	72.78%	41.32%	88.94%
CEEDNet 1D-ResNet-50		71.37%	51.50%	81.59%
CEEDNet 1D-ResNet-50	✓	73.65%	50.86%	85.36%
CEEDNet 1D-ResNeXt-50		68.58%	40.81%	82.86%
CEEDNet 1D-ResNeXt-50	✓	**73.71%**	41.22%	**90.40%**
CEEDNet 2D-VGG-19		68.74%	53.27%	76.69%
CEEDNet 2D-VGG-19	✓	71.79%	**56.51%**	79.64%
CEEDNet 2D-ResNet-18		68.99%	**58.73%**	74.27%
CEEDNet 2D-ResNet-18	✓	70.16%	**60.43%**	75.15%
CEEDNet 2D-ResNet-50		70.10%	46.24%	82.37%
CEEDNet 2D-ResNet-50	✓	71.72%	45.97%	84.94%
CEEDNet 2D-ResNeXt-50		69.44%	51.85%	78.48%
CEEDNet 2D-ResNeXt-50	✓	72.28%	53.78%	81.78%
CEEDNet ViT-B-16		69.32%	33.57%	87.69%
CEEDNet ViT-B-16	✓	71.40%	31.46%	**91.93%**
CEEDNet Ensemble		**75.46%**	47.40%	89.88%
CEEDNet Ensemble	✓	**75.77%**	47.03%	**90.54%**

Appendix C. Channel-wise occlusion sensitivity analysis

The results of the occlusion sensitivity tests performed in a channel-wise manner are shown in Figs. C.1–C.3 on CAUEEG-Dementia and in Figs. C.4 and C.5 on CAUEEG-Abnormal.

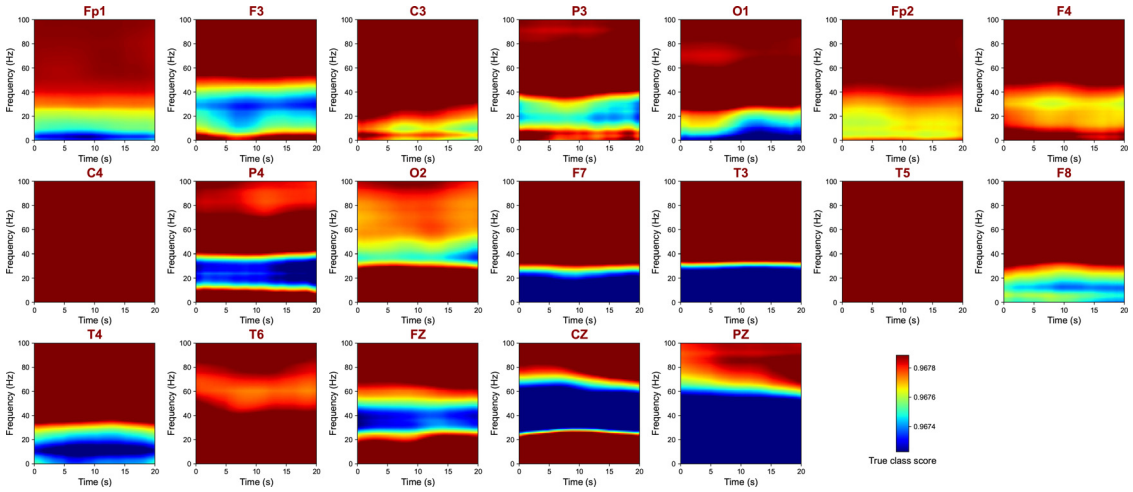


Fig. C.1. Channel-wise occlusion sensitivity analysis for the *normal* class of the CEEDNet 2D-VGG-19 model on the CAUEEG-Dementia training set. During classification, the input time-frequency feature maps were occluded in sequence by channel by putting a sliding rectangle whose width and height are quarters of those of the input. The titles of each subfigure (*e.g.*, Fp1) denote that the occlusion box was moved in that channel. These results were generated by averaging all results for *normal* in the training set. The class score for the correct class is visualized in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

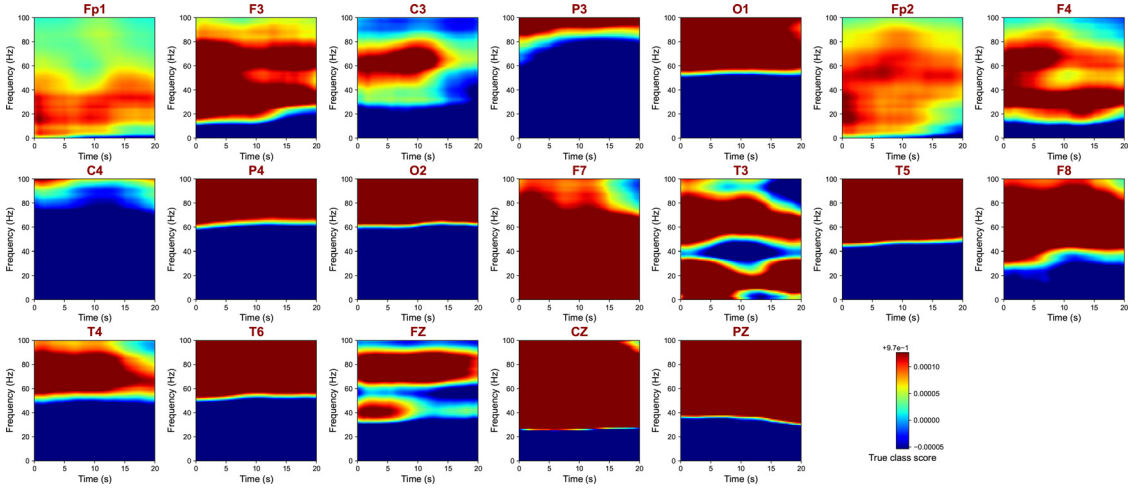


Fig. C.2. Channel-wise occlusion sensitivity analysis for the *mci* class of the CEEDNet 2D-VGG-19 model on the CAUEEG-Dementia training set. During classification, the input time-frequency feature maps were occluded in sequence by channel by putting a sliding rectangle whose width and height are quarters of those of the input. The titles of each subfigure (*e.g.*, Fp1) denote that the occlusion box was moved in that channel. These results were generated by averaging all results for *mci* in the training set. The class score for the correct class is visualized in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

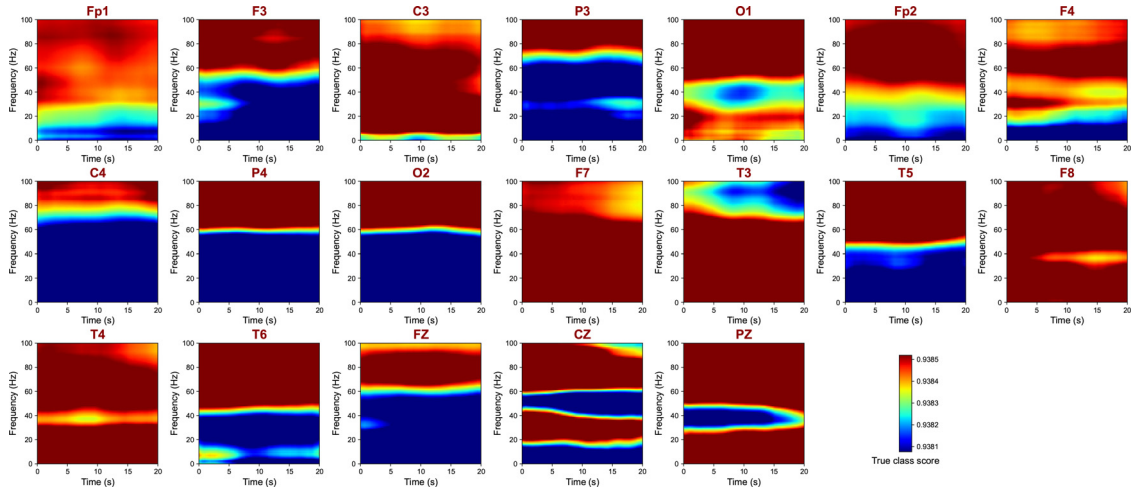


Fig. C.3. Channel-wise occlusion sensitivity analysis for the *dementia* class of the CEEDNet 2D-VGG-19 model on the CAUEEG-Dementia training set. During classification, the input time-frequency feature maps were occluded in sequence by channel by putting a sliding rectangle whose width and height are quarters of those of the input. The titles of each subfigure (*e.g.*, Fp1) denote that the occlusion box was moved in that channel. These results were generated by averaging all results for *dementia* in the training set. The class score for the correct class is visualized in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

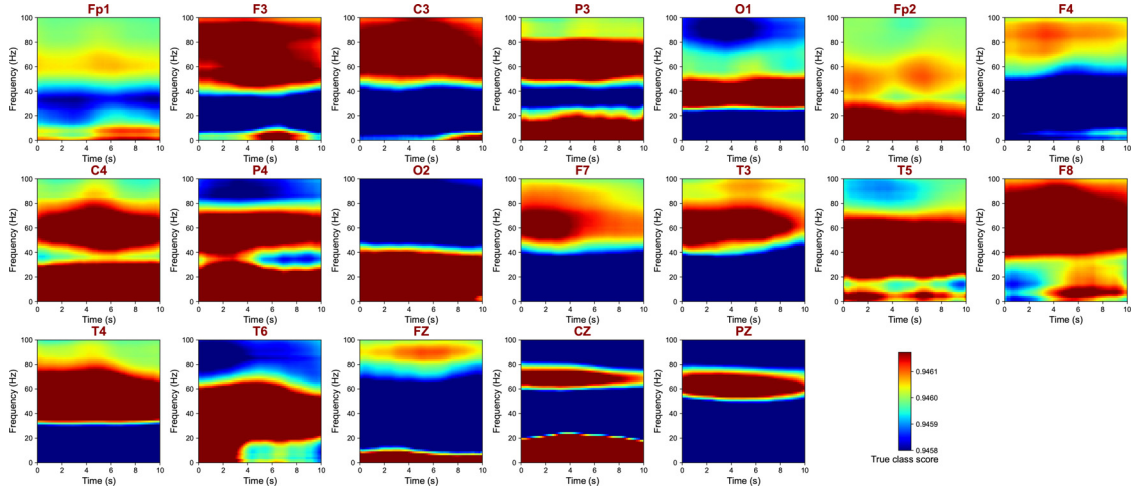


Fig. C.4. Channel-wise occlusion sensitivity analysis for the *normal* class of the CEEDNet 2D-VGG-19 model on the CAUEEG-Abnormal training set. During classification, the input time-frequency feature maps were occluded in sequence by channel by putting a sliding rectangle whose width and height are quarters of those of the input. The titles of each subfigure (*e.g.*, Fp1) denote that the occlusion box was moved in that channel. These results were generated by averaging all results for *normal* in the training set. The class score for the correct class is visualized in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

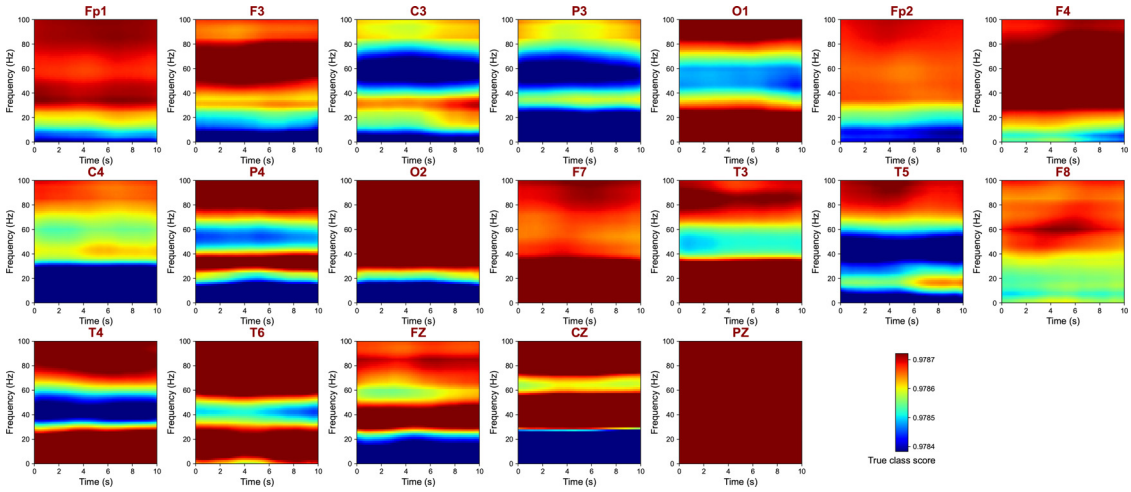


Fig. C.5. Channel-wise occlusion sensitivity analysis for the *abnormal* class of the CEEDNet 2D-VGG-19 model on the CAUEEG-Abnormal training set. During classification, the input time-frequency feature maps were occluded in sequence by channel by putting a sliding rectangle whose width and height are quarters of those of the input. The titles of each subfigure (*e.g.*, Fp1) denote that the occlusion box was moved in that channel. These results were generated by averaging all results for *abnormal* in the training set. The class score for the correct class is visualized in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Ahn, H.-J., Chin, J., Park, A., Lee, B.H., Suh, M.K., Seo, S.W., Na, D.L., 2010. Seoul neuropsychological screening battery-dementia version (SNSB-d): a useful tool for assessing and monitoring cognitive impairments in dementia patients. J. Korean Med. Sci. 25 (7), 1071–1076. doi:10.3346/jkms.2010.25.7.1071.

Al-Qazzaz, N.K., Ali, S.H.B., Ahmad, S.A., Chellappan, K., Islam, M., Escudero, J., et al., 2014. Role of EEG as biomarker in the early detection and classification of dementia. Sci. World J. 2014. doi:10.1155/2014/906038.

Alhussein, M., Muhammad, G., Hossain, M.S., 2019. Eeg pathology detection based on deep learning. IEEE Access 7, 27781–27788. doi:10.1109/ACCESS.2019.2901672.

Ba, J. L., Kiros, J. R., Hinton, G. E., 2016. Layer normalization. arXiv preprint arXiv:1607.06450. 10.48550/arXiv.1607.06450.

Bajpai, R., Yuvaraj, R., Prince, A.A., 2021. Automated EEG pathology detection based on different convolutional neural network models: Deep learning approach. Comput. Biol. Med. 133, 104434. doi:10.1016/j.compbiomed.2021.104434.

Bi, X., Wang, H., 2019. Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. Neural Netw. 114, 119–135. doi:10.1016/j.neunet.2019.02.005.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901. doi:10.48550/arXiv.2005.14165.

Cassani, R., Estarellas, M., San-Martin, R., Fraga, F.J., Falk, T.H., 2018. Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment. Dis. Markers 2018. doi:10.1155/2018/5174815.

Caviness, J.N., Hentz, J.G., Belden, C.M., Shill, H.A., Driver-Dunckley, E.D., Sabbagh, M.N., Powell, J.J., Adler, C.H., 2015. Longitudinal EEG changes correlate with cognitive measure deterioration in Parkinson's disease. J. Parkinson's Dis. 5 (1), 117–124. doi:10.3233/JPD-140480.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297. doi:10.1007/BF00994018.

Csukly, G., Sirály, E., Fodor, Z., Horváth, A., Salacz, P., Hidasi, Z., Csibri, E., Rudas, G., Szabó, A., 2016. The differentiation of amnestic type MCI from the non-amnestic types by structural MRI. Front. Aging Neurosci. 8, 52. doi:10.3389/fnagi.2016.00052.

Lopez de Diego, S.I., 2017. Automated interpretation of abnormal adult electroencephalograms. Temple University. Libraries Ph.D. thesis.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 10.48550/arXiv.2010.11929.

Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., et al., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. Lancet Neurol. 6 (8), 734–746. doi:10.1016/S1474-4422(07)70178-3.

Farina, F.R., Emek-Savaş, D.D., Rueda-Delgado, L., Boyle, R., Kiiski, H., Yener, G., Whelan, R., 2020. A comparison of resting state EEG and structural MRI for classifying Alzheimer's disease and mild cognitive impairment. NeuroImage 215, 116795. doi:10.1016/j.neuroimage.2020.116795.

First, M.B., Pincus, H.A., 2002. The DSM-IV text revision: rationale and potential impact on clinical practice. Psychiatr. Serv. 53 (3), 288–292. doi:10.1176/appi.ps.53.3.288.

Fiscon, G., Weitschek, E., Cialini, A., Felici, G., Bertolazzi, P., De Salvo, S., Bramanti, A., Bramanti, P., De Cola, M.C., 2018. Combining EEG signal processing with supervised methods for Alzheimer's patients classification. BMC Med. Inf. Decis. Mak. 18 (1), 1–10. doi:10.1186/s12911-018-0613-y.

Fix, E., Hodges, J.L., 1989. Discriminatory analysis. nonparametric discrimination: Consistency properties. Int. Stat. Rev./Revue Internationale de Statistique 57 (3), 238–247. doi:10.2307/1403797.

Gemein, L.A.W., Schirrmeister, R.T., Chrabaszcz, P., Wilson, D., Boedecker, J., Schulze-Bonhage, A., Hutter, F., Ball, T., 2020. Machine-learning-based diagnostics of EEG pathology. NeuroImage 220, 117021. doi:10.1016/j.neuroimage.2020.117021.

Grandini, M., Bagli, E., Visani, G., 2020. Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756. 10.48550/arXiv.2008.05756.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Springer, pp. 630–645. doi:10.1007/978-3-319-46493-0_38.

Hendrycks, D., Gimpel, K., 2016. Gaussian error linear units (GELUs). arXiv preprint arXiv:1606.08415.

Ho, T.K., 1995. Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition, Vol. 1. IEEE, pp. 278–282. doi:10.1109/ICDAR.1995.598994.

Ieracitano, C., Mammone, N., Bramanti, A., Hussain, A., Morabito, F.C., 2019. A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. Neurocomputing 323, 96–107. doi:10.1016/j.neucom.2018.09.071.

Ieracitano, C., Mammone, N., Hussain, A., Morabito, F.C., 2020. A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. Neural Netw. 123, 176–190. doi:10.1016/j.neunet.2019.12.006.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456. http://proceedings.mlr.press/v37/ioffe15.html

Jahng, S., Na, D.L., Kang, Y., 2015. Constructing a composite score for the seoul neuropsychological screening battery-core. Dementia Neurocognit. Disord. 14 (4), 137–142. doi:10.12779/dnd.2015.14.4.137.

Jelic, V., Johansson, S.E., Almkvist, O., Shigeta, M., Julin, P., Nordberg, A., Winblad, B., Wahlund, L.O., 2000. Quantitative electroencephalography in mild cognitive impairment: longitudinal changes and possible prediction of Alzheimer's disease. Neurobiol. Aging 21 (4), 533–540. doi:10.1016/S0197-4580(00)00153-6.

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., et al., 2020. Dementia prevention, intervention, and care: 2020 report of the lancet commission. Lancet 396 (10248), 413–446. doi:10.1016/S0140-6736(20)30367-6.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Malek, N., Baker, M.R., Mann, C., Greene, J., 2017. Electroencephalographic markers in dementia. Acta Neurol. Scand. 135 (4), 388–393. doi:10.1111/ane.12638.

McBride, J.C., Zhao, X., Munro, N.B., Smith, C.D., Jicha, G.A., Hively, L., Broster, L.S., Schmitt, F.A., Kryscio, R.J., Jiang, Y., 2014. Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. Comput. Methods Programs Biomed. 114 (2), 153–163. doi:10.1016/j.cmpb.2014.01.019.

Michaud, T.L., Su, D., Siahpush, M., Murman, D.L., 2017. The risk of incident mild cognitive impairment and progression to dementia considering mild cognitive impairment subtypes. Demen. Geriatr. Cognit. Disorders Extra 7 (1), 15–29. doi:10.1159/000452486.

Misra, D., 2019. Mish: a self regularized non-monotonic neural activation function. 4, 1048550. arXiv:1908.08681.

Musaeus, C.S., Engedal, K., Høgh, P., Jelic, V., Mørup, M., Naik, M., Oeksengaard, A.-R., Snaedal, J., Wahlund, L.-O., Waldemar, G., et al., 2018. Eeg theta power is an early

marker of cognitive decline in dementia due to Alzheimer's disease. J. Alzheimer's Dis. 64 (4), 1359–1371. doi:10.3233/JAD-180300.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: Icml. https://icml.cc/Conferences/2010/papers/432.pdf
Park, M., Moon, W.-J., 2016. Structural MR imaging in the diagnosis of Alzheimer's disease and other neurodegenerative dementia: current imaging approach and future perspectives. Korean J. Radiol. 17 (6), 827–845. doi:10.3348/kjr.2016.17.6.827.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. PyTorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32. doi:10.48550/arXiv.1912.01703.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830. doi:10.48550/arXiv.1201.0490.
Perez-Valero, E., Lopez-Gordo, M.A., Morillas, C., Pelayo, F., Vaquero-Blasco, M.A., 2021. A review of automated techniques for assisting the early detection of Alzheimer's disease with a focus on EEG. J. Alzheimer's Dis. 80 (4), 1363–1376. doi:10.3233/JAD-201455.
Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J., 2019. Deep learning-based electroencephalography analysis: a systematic review. J. Neural Eng. 16 (5), 051001. doi:10.1088/1741-2552/ab260c.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115 (3), 211–252. doi:10.1007/s11263-015-0816-y.
Schuhmann, C., Vencu, R., Beaumont, R., Kaczmarczyk, R., Mullis, C., Katta, A., Coombes, T., Jitsev, J., Komatsuzaki, A., 2021. LAION-400M: open dataset of clip-filtered 400 million image-text pairs. 10.48550/arXiv.2111.02114.
Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A.W.R., Bridgland, A., et al., 2020. Improved protein structure prediction using potentials from deep learning. Nature 577 (7792), 706–710. doi:10.1038/s41586-019-1923-7.
Sharma, N., Kolekar, M.H., Jha, K., Kumar, Y., 2019. Eeg and cognitive biomarkers based mild cognitive impairment diagnosis. Irbm 40 (2), 113–121. doi:10.1016/j.irbm.2018.11.007.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. 10.48550/arXiv.1409.1556.
Sonoda, S., Murata, N., 2017. Neural network with unbounded activation functions is universal approximator. Appl. Comput. Harmon. Anal. 43 (2), 233–268. doi:10.1016/j.acha.2015.12.005.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html
Weller, J., Budson, A., 2018. Current understanding of Alzheimer's disease diagnosis and treatment. F1000Research 7. doi:10.12688/f1000research.14506.1.
Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500. doi:10.1109/CVPR.2017.634.
Yıldırım, O., Baloglu, U.B., Acharya, U.R., 2020. A deep convolutional neural network model for automated identification of abnormal EEG signals. Neural Comput. Appl. 32 (20), 15857–15868. doi:10.1007/s00521-018-3889-z.
Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: European Conference on Computer Vision. Springer, pp. 818–833. doi:10.1007/978-3-319-10590-1_53.
Zhang, H., Cisse, M., Dauphin, Y. N., Lopez-Paz, D., 2017. mixup: Beyond empirical risk minimization. 10.48550/arXiv.1710.09412.