

# Data leakage in deep learning studies of translational EEG

Geoffrey Brookshire<sup>1,‡,\*</sup>, Jake Kasper<sup>1,‡</sup>, Nicholas Blauch<sup>1</sup>, Yunan “Charles” Wu<sup>1</sup>, Ryan Glatt<sup>2</sup>, David A. Merrill<sup>2,3,4</sup>, Spencer Gerrol<sup>1</sup>, Keith J. Yoder<sup>1</sup>, Colin Quirk<sup>1</sup>, and Ché Lucero<sup>1</sup>

<sup>1</sup>*SPARK Neuro Inc., New York, NY, USA*

<sup>2</sup>*Pacific Brain Health Center, Pacific Neuroscience Institute and Foundation, Santa Monica, CA, USA*

<sup>3</sup>*Saint John’s Cancer Institute at Providence Saint John’s Health Center, Santa Monica, CA, USA*

<sup>4</sup>*Psychiatry & Biobehavioral Sciences, Semel Institute for Neuroscience & Human Behavior, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*

<sup>‡</sup>*Equal contributions*

<sup>\*</sup>*Corresponding author: [Geoff.Brookshire@sparkneuro.com](mailto:Geoff.Brookshire@sparkneuro.com)*

## Abstract

A growing number of studies use deep neural networks (DNNs) to identify diseases from recordings of brain activity. DNN studies of electroencephalography (EEG) typically use cross-validation to test how accurately a model can predict the disease state of held-out test data. In these studies, segments of EEG data are often randomly assigned to the training or test sets. As a consequence, data from individual subjects appears in both training and test data. Could high test-set accuracy reflect leakage from subject-specific representations, rather than patterns that identify a disease? We address this question by testing the performance of DNN classifiers using segment-based holdout (where EEG segments from one subject can appear in both the training and test sets), and comparing this to their performance using subject-based holdout (where individual subjects' data appears exclusively in either the training set or the test set). We compare segment-based and subject-based holdout in two EEG datasets: one classifying Alzheimer's disease, and the other classifying epileptic seizures. In both datasets, we find that performance on previously-unseen subjects is strongly overestimated when models are trained using segment-based holdout. Next, we survey the literature and find that the majority of translational DNN-EEG studies use segment-based holdout, and therefore overestimate model performance on new subjects. In a hospital or doctor's office, clinicians need to diagnose new patients whose data was not used in training the model; segment-based holdout, therefore, does not reflect the real-world performance of a translational DNN model. When evaluating how DNNs could be used for medical diagnosis, models must be tested on subjects whose data was not included in the training set.

**Keywords:** Electroencephalography; deep neural networks; data leakage; Alzheimer's disease; epilepsy

## Introduction

Translational neuroscience studies increasingly turn to deep neural network (DNN) models to find structure in neural data. The power of DNN models comes from their ability to discover patterns in the data that researchers would not have been able to specify. In this literature, DNNs have been trained on a variety of imaging techniques to identify a wide range of clinical conditions. Many of these studies use DNNs to diagnose diseases based on anatomical neuroimaging. For example, DNN models can identify Alzheimer's disease (AD) using structural magnetic resonance imaging (MRI)<sup>74</sup>, and a variety of cancers and brain injuries using CT scans<sup>29,35</sup>. In addition to anatomical data, a large number of studies have used DNNs to identify diseases from functional neuroimaging data. For example, DNNs with functional MRI show promise for identifying AD, Autism spectrum disorders, attention-deficit/hyperactivity disorder (ADHD), and schizophrenia<sup>73</sup>. Furthermore, DNNs have been used with electroencephalography (EEG) to study a variety of different neural and cognitive disorders<sup>17</sup>.

Deep learning helps to reveal previously-unknown patterns in neuroimaging data, but it also presents researchers with subtle pitfalls. One set of challenges concerns how the data are split into separate training and test sets. The training set is used to fit the model’s parameters, and the test set is used to estimate the model’s performance on new data. (A third subset of the data is often held aside as a validation set, used to tune the model’s hyperparameters and to determine when to stop training the model.) In some cases, researchers train their model on one subset of the available data, and then evaluate the model’s performance on a separate test set. In other cases, researchers use cross-validation (CV) to train and test models on multiple subsets of the data. Under both of these approaches, researchers must be careful to avoid “data leakage” when splitting the data into training and test sets. Data leakage, which arises when information about the test set is present in the training set, results in a positively-biased estimate of the model’s performance<sup>36</sup>. For example, in a data-mining competition focused on identifying patients with breast cancer, one team of researchers found that the patient ID number carried predictive information about cancer risk<sup>59</sup>. These ID numbers may have appeared after compiling data from different medical institutions. Because the ID number was assigned based on patients’ diagnosis, it constitutes a source of data leakage<sup>59</sup>. In general, data leakage occurs when an experimenter handles the data in a way that artificially introduces correlations between the training and test sets.

DNN models typically require a large amount of training data to perform well, but neural datasets are usually expensive and difficult to obtain. To increase the number of observations available to train the model, these studies often split a single neural recording into multiple samples, and use each sample as a separate observation during training or testing. For example, a 3D structural MR volume could be split into multiple 2D slices, and an fMRI time-series could be split into multiple segments of time<sup>74</sup>. When multiple observations from a single subject are included in both the training and test sets, it constitutes data leakage: Instead of learning a generalizable pattern, these models could learn characteristics of the individual subjects in the training set, and then simply recognize those familiar subjects in the test set. As a result, these models perform well in the study’s test set, leading the researchers to believe they have a robust classifier. In new subjects, however, the model may fail to generalize. In fact, leakage of subject information does occur in a number of published MRI studies<sup>74</sup>. Furthermore, leakage of subject-specific information is widespread in translational studies using optical coherence tomography (OCT), and leads to strongly inflated estimates of test accuracy<sup>65</sup>.

Studies using DNNs with EEG are particularly susceptible to data leakage. In these studies, each subject’s full EEG time-series (lasting several minutes) is commonly divided up into brief segments (lasting several seconds)<sup>17</sup>. Each segment is then used as a separate observation during training or testing. This segmentation procedure is meant to ensure that DNN models have enough training data to learn robust rep-

representations of the patterns that characterize a disease, and to prepare the data for commonly-used model architectures. However, EEG segmentation leads to data leakage if the same subjects appear in both the training and test sets. Segments of EEG from one subject are more similar to each other than to segments from different subjects<sup>18</sup>. Instead of learning an abstract representation that would generalize to new subjects, a DNN model could therefore achieve high classification accuracy by associating a label with each subject’s idiosyncratic pattern of brain activity. As a consequence, randomly splitting EEG segments into training and test sets results in data leakage, and a biased estimate of test performance: accuracy is high on the researchers’ test set, but the classifier will generalize poorly to new subjects. In a clinical setting, this leads to an apparently-promising diagnostic tool that fails when applied to new patients. To avoid this kind of data leakage, all segments from a given subject must be assigned to only a single partition of the data (i.e. train *or* validation *or* test).

How does leakage of subject-specific information bias the results of translational DNN-EEG studies? Here we address this question by examining the effects of data leakage in two case studies, and then reviewing the published literature to gauge the prevalence of this leakage. In the case studies, we reproduce two convolutional neural network (CNN) architectures used by published studies – both of which used a train-test split that introduced data leakage. First, we use a CNN to classify subjects as either healthy or as having dementia due to Alzheimer’s disease. Second, we use a CNN to classify whether segments of time contain an epileptic seizure. In both datasets, we find that real-world performance is dramatically overestimated when data from individual subjects is included in both the training and test sets. In the literature review, we find that the majority of translational DNN-EEG studies suffer from data leakage due to data from individual subjects appearing in both the training and test sets.

## Methods

### Deep neural network analysis overview

To investigate how segment-based holdout leads to data leakage, we reproduced the model architectures from two published studies<sup>56,58</sup>. The goal of these analyses was not to develop an optimal architecture, but rather to evaluate the impact of different cross-validation choices on the estimated model performance. We therefore re-used the published architectures and data processing pipelines without modification.

## Experiment 1: Alzheimer’s disease diagnosis

### EEG data

EEG recordings were provided by the Pacific Neuroscience Institute<sup>25</sup>. All procedures were approved by the St. John’s Cancer Institute Institutional Review Board (Protocol JWCI-19-1101) in accordance with the Helsinki Declaration of 1975. Patients were evaluated by a dementia specialist as part of their visit to a specialty memory clinic (Pacific Brain Health Center in Santa Monica, CA) for memory complaints. These evaluations included behavioral testing as well as EEG recordings. After these evaluations, subjects were selected by retrospectively reviewing charts for patients aged 55 and older seen between July 2018 and February 2021.

Patients received a consensus diagnosis from a panel of board-certified dementia specialists. Diagnoses were performed using standard clinical methods on the basis of neurological examinations, cognitive testing (MMSE<sup>23</sup> or MoCA<sup>54</sup>), clinical history (e.g. hypertension, diabetes, head injury, depression), and laboratory results (e.g. vitamin B-12 levels, thyroid stimulating hormone levels, and rapid plasma regain testing). These tests were used to rule out reversible causes of memory loss and to diagnose SCI, MCI, and dementia. EEG data was not included in the diagnostic process. Cognitive impairment was diagnosed on the basis of MMSE (or MoCA scores converted to MMSE<sup>8</sup>), with MCI diagnosed according to established criteria<sup>42</sup>. MCI was distinguished from dementia on the basis of preserved independence in functional abilities, and a lack of significant impairment in social or occupational functioning. SCI was diagnosed in patients with subjective complaints but without evidence of MCI. Diagnostic categorization was based on the clinical syndromes<sup>42</sup>, and did not consider disease etiology or subtypes within each stage.

EEG data were recorded at 250 Hz using the eVox System (Evoke Neuroscience), with a cap that included 19 electrodes following the International 10-20 system (FP1, FP2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, and O2). The full EEG session included a 5-minute block of eyes-open rest, a 5-minute block of eyes-closed rest, and a 15-minute go/no-go task. In this study, we analyzed only the eyes-open resting-state data. Recordings were low-pass filtered below 125 Hz, and split into non-overlapping segments of 2 s (500 samples) for model training. Channels were stacked to produce matrices of shape (500, 19) as model inputs.

We selected all 49 subjects in the dataset who were diagnosed with dementia due to Alzheimer’s disease (18 male, 31 female; age  $73.9 \pm 6.8$  years). As a comparison, we selected an equal number of subjects with subjective cognitive impairment (SCI;  $n = 49$ , 18 male, 31 female; age  $63.9 \pm 11.4$  years).

## Architecture

Based on previous work<sup>56</sup>, a basic 1D convolutional neural network was used to classify segments of time-series data as SCI or AD. Because our goal was to evaluate the effects of different cross-validation strategies, we re-used the published architecture without modification. This model learns temporal filters that are applied equivalently across each EEG channel. Progressing through the network, subsequent layers build more complex features that take into account a larger temporal receptive field, and some invariance is achieved through pooling over time. The model consisted of 4 convolutional layers, each followed by rectification, max pooling, and batch normalization; convolutional layers were followed by 2 dense fully-connected layers of 20 and 10 hidden units, respectively, each rectified, and finally a dense connectivity to the output layer with 2 units representing AD yes/no probability logits. All deep learning models were trained with Keras and Tensorflow. The exact Keras code used to specify the architecture can be found in the Appendix.

## Training

Models were trained for 70 epochs without any early stopping or hyperparameter tuning. A batch size of 32, initial learning rate of 0.0001, and the Adam optimizer were used to optimize models. Training accuracy was computed and stored online during each epoch, and averaged across batches to report the training accuracy for each epoch. To visualize how quickly the models reached their final performance, test set accuracy was also computed after each epoch, averaged across batches. Since we reused the model architecture from prior published work, no model selection was performed; performing ongoing validation on the test is therefore not a source of data leakage. For segment-based holdout, data were split using 10-fold cross-validation (see ‘Cross-validation’ for details).

## Experiment 2: seizure detection

### EEG data

We analyzed data from the Siena Scalp EEG Database<sup>19,20</sup> hosted on PhysioNet<sup>28</sup>. These recordings were collected in accordance with the Declaration of Helsinki, and approved by the Ethical Committee of the University of Siena. Participants provided written informed consent before beginning data collection. This dataset includes recordings from 14 epilepsy patients (age 20-71 years, 9 male) digitized at 512 Hz with electrodes arranged following the International 10-20 system. Seizures in the data were labeled by an expert clinician. This dataset contains 47 seizures in approximately 128 hours of recorded EEG. To ensure that the data were balanced between seizure and non-seizure epochs, we selected non-seizure data from the beginning of each subject’s recordings to match the duration of their seizure-labeled data. This led to 47 min 21 s of

data in each condition (1:34:42 s in total).

In contrast to the previous section where raw time series were used, EEG data were prepared for the classifier analysis in the frequency domain, following the approach used by Rashed-Al-Mahfouz and colleagues<sup>58</sup>. Spectrograms were computed with a window length of 256 samples (0.5 s) overlapping by 128 samples (0.25 s), using a Hann taper. Spectrograms were then divided into segments of 1.5 s. As in the original study, we used the RGB representation of the spectrogram (viridis color-map), and exported as 224 x 224 x 3 images for training and testing with the CNN models.

## Architecture

The goal of the simulations was primarily to evaluate the impact of different cross-validation choices, not to evaluate the architecture. Thus, the architecture of Rashed-Al-Mahfouz and colleagues<sup>58</sup> was used without modification. To handle 3D spectrogram data (vs. 2D time-series used in the previous section), a 2D convolutional neural network was used. This model learns 2D spectrotemporal features that are applied equivalently across the spectrogram. The model contains 4 convolutional layers, each followed by rectification, pooling, and batch normalization, followed by 2 hidden fully-connected layers of 256 and 512 units each, dropout, and a final classification layer of 2 units corresponding to seizure yes/no. The exact Keras code used to specify the architecture can be found in the Appendix.

## Training

Models were trained for 70 epochs with no early stopping. We used the RMSProp optimizer with a batch size of 32 and a learning rate of 0.00001. Training accuracy was computed and stored online during each epoch, and averaged across batches to report the training accuracy for each epoch. To visualize how quickly the models reached their final performance, test set accuracy was also computed after each epoch, averaged across batches. Since we reused the model architecture from prior published work, no model selection was performed; performing ongoing validation on the test is therefore not a source of data leakage.

## Cross-validation

This study is primarily concerned with the consequences of different approaches to splitting the data between training and test sets. We assess two types of train-test split: (1) holding out individual segments of EEG data without regard for subject ID (“segment-based holdout”), and (2) holding out entire subjects, ensuring that all segments for a given subject appear in only the training or the test set (“subject-based holdout”).

## Segment-based holdout

Segment-based cross-validation considers all EEG segments to be equivalent, and divides them into training and validation partitions without considering subject ID. This segment-holdout approach will lead to data leakage if there is statistical non-independence due to multiple EEG segments coming from each subject. Given  $n$  segments and  $m$  time-points per segments, we construct a matrix  $X$  of EEG segments of size  $(n, m)$ , and a vector  $\mathbf{y}$  of diagnostic label of length  $n$ . The cross-validation is a simple partition of the index vector  $\alpha = \{1, 2, \dots, n\}$  into disjoint subsets  $\alpha_{\text{train}}$  and  $\alpha_{\text{test}}$ . Where  $X_i$  gives the  $i$ th segments of  $X$ , we then have  $X_{\text{train}} = \{X_i\} \forall i \in \alpha_{\text{train}}$ ,  $X_{\text{val}} = \{X_i\} \forall i \in \alpha_{\text{test}}$ , and  $\mathbf{y}_{\text{train}} = \{\mathbf{y}_i\} \forall i \in \alpha_{\text{train}}$ ,  $\mathbf{y}_{\text{test}} = \{\mathbf{y}_i\} \forall i \in \alpha_{\text{test}}$ .

## Subject-based holdout

Subject-based cross-validation takes into account which subject each EEG segment comes from. This approach enforces that each subject appears in only one partition of the cross-validation, ensuring there is no leakage of subject-level information across training and test sets. To create this split, we consider an additional subject vector  $\mathbf{s}$ , which is used to constrain the partition of  $X$  and  $\mathbf{y}$ . Concretely, rather than partitioning the index vector  $\alpha$ , we partition the unique subject vector  $\mathbf{s}_u$ , which gives the unique entries of  $\mathbf{s}$ , and collect all corresponding segments from each subject contained in train and validation partitions into  $\alpha_{\text{train}}$  and  $\alpha_{\text{test}}$ . This enforces the constraint that  $\mathbf{s}_i \neq \mathbf{s}_j \forall i \in \alpha_{\text{train}}, j \in \alpha_{\text{test}}$ . To perform  $k$ -fold cross-validation, we first divide  $\mathbf{s}_u$  into  $k$  non-overlapping chunks, and each chunk to serve as the validation data in each fold of cross-validation, where the remaining  $k - 1$  chunks are reserved for training.

## Literature review

We searched the literature for studies that used deep learning with segments of EEG to classify a variety of diseases. We searched Google Scholar for papers investigating Alzheimer’s disease, Parkinson’s disease, attention-deficit/hyperactivity disorder (ADHD), depression, schizophrenia, and seizures. We then searched the references of these papers to identify additional appropriate publications for inclusion. This non-exhaustive search included 63 papers, all of which were published since 2018 and used deep learning to study one of the conditions named above.

Next, we examined how the training and test sets were determined in these studies. If a paper specified that the EEG recordings were split into segments, but did not specify that they used subjects as an organizing factor of the train-test split, we labeled that study as using ‘segment-based’ holdout. Some papers specifically stated that segments from individual subjects were included in both the training and test sets (for example, studies that trained separate models for each subject); these studies were also labeled as segment-based



holdout. If a paper specified that all the segments from a single subject were assigned to only the training or the test set, we labeled that study as using ‘subject-based’ holdout. If a study used both segment-based and subject-based holdout in different analyses, we labeled the study as ‘both’. We labeled studies as ‘unclear’ if we could not determine whether the models were trained on segments of EEG recordings, and it was not explicitly stated that subjects were used as a factor in the holdout procedure.

## Results

### Data leakage leads to biased test-set accuracy

We analyze two datasets to test how the estimated accuracy of a DNN classifier depends on the train-test split. First, we examine the effects of data leakage in a patient-level classifier by training a model to diagnose Alzheimer’s disease. Second, we examine the effects of data leakage in a segment-level classifier by training a model to identify periods of time that include an epileptic seizure. In each of these analyses, we reuse a published DNN architecture to analyze an existing dataset.

#### Identifying patients with Alzheimer’s disease

To determine whether segment-based holdout leads to a biased estimate of accuracy, we first trained a CNN to diagnose Alzheimer’s disease using segments of EEG. When the EEG segments were split into training and test sets without considering subject ID, the model showed nearly perfect test-set accuracy of 99.8% [99.1-100.0%] (Figure 1a). Performance quickly approached ceiling within the first 15 training epochs (Figure 2a). This high accuracy is consistent with prior studies that use segment-based holdout and report high accuracy for CNNs at identifying neurological disorders<sup>2,43,56</sup>. Could this pattern of high accuracy reflect data leakage, instead of a robust and generalizable classifier?

When we used subject-based holdout, ensuring that individual subjects’ data did not appear in both the training and test sets, test accuracy dropped to 53.0% [43.1-64.8%], with 95% confidence intervals that included chance performance of 50%. Performance remained low throughout the training epochs (Figure 2b). Compared with subject-based holdout, segment-based holdout significantly overestimates the model performance on previously-unseen subjects (Wilcoxon  $T = 0.0$ ,  $p = .002$ ).

#### Identifying segments containing epileptic seizures

In some cases, artificial neural network models have been used to identify time-limited events within ongoing brain activity, such as epileptic seizures. Does segment-based holdout also lead to data leakage when labeling

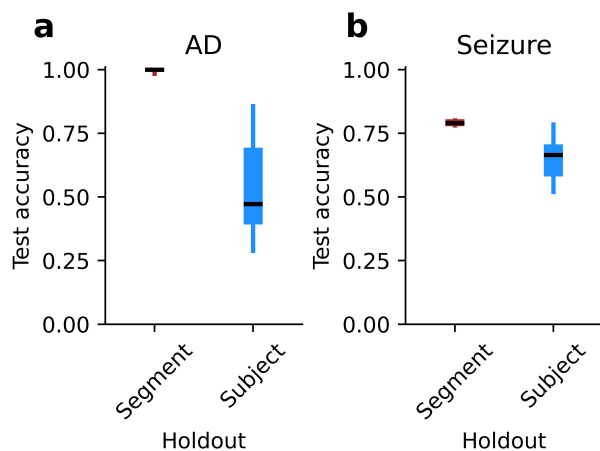


Figure 1: Test-set accuracy of CNN models predicting held-out data, plotted separately for segment-based holdout and subject-based holdout. (a) Accuracy for models trained to classify Alzheimer’s disease in individual subjects. Boxes show the inter-quartile range, dark lines show the median, and whiskers extend to the minimum and maximum points. (b) Accuracy for models trained to identify seizures in segments of EEG data. Details as in (a).

periods of time within subjects? To answer this question, we trained a CNN to classify segments of EEG data as containing an epileptic seizure or not.

When the EEG segments were split into training and test sets without considering subject ID, the model reached a high test-set accuracy of 79.1% [78.8-79.4%] (Figure 1b). Accuracy leveled out within 10 training epochs (Figure 2c). When individual subjects’ data segments were restricted to appear in only the training or test set, however, accuracy fell to 65.1% [61.3-69.1%]. Accuracy remained low throughout training epochs (Figure 2d). Even when the model is tasked with labeling periods of activity within subjects, segment-based holdout significantly overestimates performance on previously-unseen subjects (Wilcoxon  $T = 0.0$ ,  $p = 0.0001$ ).

## Data leakage in published EEG studies

Do published translational EEG studies suffer from subject-specific data leakage, or do they avoid it by computing their test-set accuracy on held-out subjects? We examined the train-test split strategies in published studies that attempted to identify a clinical disorder using DNNs with EEG recordings. Out of the 63 relevant papers we found, only 17 (27.0%) unambiguously avoided this type of data leakage (Figure 3, Table 1). Leakage of subject-specific information is pervasive in the translational EEG literature.

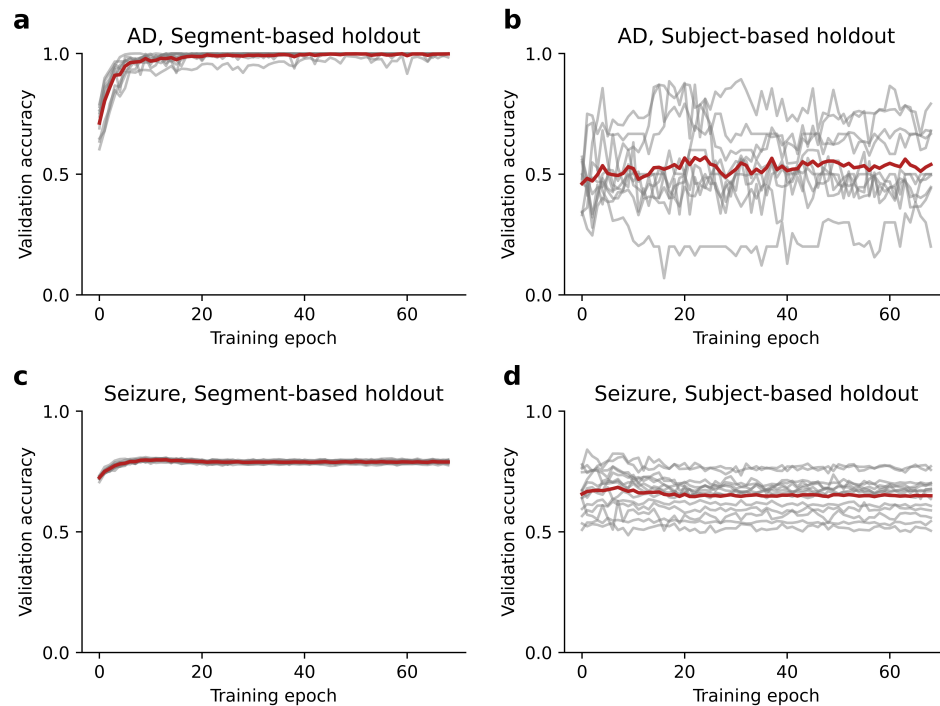


Figure 2: Test-set accuracy of CNN models plotted as a function of the training epoch. Grey lines show accuracy in individual cross-validation folds, and red lines show the average across folds. (a) Accuracy for models trained to classify Alzheimer’s disease using segment-based holdout. (b) Accuracy for models trained to classify Alzheimer’s disease using subject-based holdout. (c) Accuracy for models trained to identify seizures using segment-based holdout. (d) Accuracy for models trained to identify seizures using subject-based holdout.

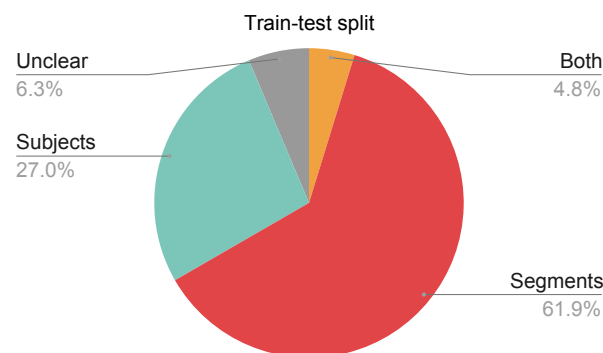


Figure 3: Number of studies using each type of test-split. “Segments”: Segments of EEG data were assigned to the training and test sets without regard to subject; this approach leads to data leakage. “Subjects”: Each subject’s data appeared in only the training set or the test set. “Both”: Both the Subjects and Segments approaches were used in different analyses. “Unclear”: We could not determine which approach was used for train-test splits.

## Discussion

In EEG studies using deep learning, data leakage can occur when segments of data from the same subjects are included in both the training and test sets. Here we demonstrate that leakage of subject-specific information can dramatically overestimate the real-world clinical performance of a DNN classifier. Our Alzheimer's CNN classifier appeared to have an accuracy of above 99% when using segment-based holdout, but its true performance on previously-unseen subjects was indistinguishable from chance. We found this bias in test-set performance both in a between-subjects task (identifying patients with Alzheimer's disease) and in a within-subjects task (identifying segments that contain a seizure). Next, we show that this type of data leakage appears in the majority of published translational DNN-EEG studies we examined. Together, these results illustrate how an improperly-designed training-test split can bias the results of DNN studies, and show that biased results are widespread in the published literature.

To be useful in a clinical setting, a diagnostic classifier must be able to identify a disease in new patients. Models trained using segment-based holdout, however, strongly overestimate their ability to perform this task. Instead, these models may learn patterns associated with individual subjects, and then associate those idiosyncratic patterns with a diagnosis. As a consequence, performance of these models drops precipitously when they are tested in new subjects. When training a translational DNN classifier, the model must be tested with subjects who were not included in the training set.

### Data leakage when identifying events within subjects

Instead of identifying a disease in each subject, some studies attempt to identify a diseased process in each segment of time (see Appendix Table 1). DNN models of epilepsy, for example, often aim to classify the segments of data that contain a seizure. We demonstrated that those studies are not immune to data leakage in training-test splits: the accuracy in novel subjects is strongly overestimated when the test set includes subjects who were also in the training set. This result could arise if the model uses different patterns to identify seizures in each subject.

Subject-specific studies indicate that a bespoke classifier could be trained to identify seizures in each new patient<sup>34,46,47</sup>. However, this would require every patient to have a large dataset of recordings that have already been labeled, which limits the clinical utility of this approach. A more realistic approach is to train DNN models to identify events in unseen patients.

## Data leakage in other methods

In studies which have only one observation per subject, cross-validation is trivial – single observations are simply assigned to the training or test set. However, in EEG and many other medical imaging methods, the data from each subject is routinely split into multiple segments. In this paper, we showed how data leakage can arise when a long recording is split into multiple shorter segments. However, the same principles apply to any other method that introduces statistical non-independence between the training and test sets. For example, some EEG-based DNNs treat every channel independently, and use information from each channel as a separate observation<sup>48</sup>. Those studies are likely to suffer from substantial data leakage, since physiological sources of electrical activity appear redundantly across multiple EEG scalp electrodes<sup>50</sup>.

These principles also apply to other medical imaging methods. Similar types of data leakage have been documented in studies using both functional<sup>73</sup> and anatomical<sup>74</sup> MRI, as well as in optical coherence tomography (OCT)<sup>65</sup>.

## Conclusion

Data leakage occurs when EEG segments from one subject appear in both the training and test sets. As a result, the test set accuracy dramatically overestimates the classifier's performance in new subjects. This type of data leakage is common in published studies using DNNs and translational EEG. To accurately estimate a model's performance, researchers must ensure that each subject's data is included in only the training or the test set, but not both.

## Acknowledgements

This work was supported by SPARK Neuro, Inc.

## Conflict of interest

Geoffrey Brookshire, Keith J. Yoder, Spencer Gerrol, and Ché Lucero are employed at SPARK Neuro Inc., a medical technology company developing diagnostic aids to help clinicians identify and assess neurodegenerative disease.

## Appendix

Table 1: Prior translational studies using deep learning with EEG. Each line in the table describes one published translational study using a DNN with EEG data. The “Target” column holds the clinical condition being classified. The “Test split” column shows the approach used to determine how the data were divided into training and test sets. “Segments”: Segments of EEG data were assigned to the training and test sets without regard to subject; this approach leads to data leakage. “Subjects”: Each subject’s data appeared in only the training set or the test set. “Both”: Both the Subjects and Segments approaches were used in different analyses. “Unclear”: We could not determine which approach was used for train-test splits.

Article	Target	Test split
Ahmadi et al. (2021)	ADHD	Segments
Bakhtyari and Mirzaei (2022)	ADHD	Segments
Chang et al. (2022)	ADHD	Subjects
Chen et al. (2019)	ADHD	Segments
Chen et al. (2019)	ADHD	Segments
Dubreuil-Vall et al. (2020)	ADHD	Subjects
Mafi and Radfar (2022)	ADHD	Segments
Moghaddari et al. (2020)	ADHD	Segments
TaghiBeyglou et al. (2022)	ADHD	Subjects
Tosun (2021)	ADHD	Segments
Vahid et al. (2019)	ADHD	Subjects
Zhou et al. (2022)	ADHD	Unclear
Kim et al. (2018)	Alcoholism	Segments
Bi and Wang (2019)	Alzheimer’s	Segments
Gkenios et al. (2022)	Alzheimer’s	Both
Huggins et al. (2021)	Alzheimer’s	Segments
Ieracitano et al. (2019)	Alzheimer’s	Both
Kim and Kim (2018)	Alzheimer’s	Subjects

Morabito et al. (2016)	Alzheimer's	Subjects
You et al. (2020)	Alzheimer's	Segments
Zhao and He (2015)	Alzheimer's	Segments
Acharya et al. (2018)	Depression	Segments
Ay et al. (2019)	Depression	Segments
Kwon et al. (2019)	Depression	Subjects
Li et al. (2019)	Depression	Subjects
Li et al. (2020)	Depression	Subjects
Mumtaz and Qayyum (2019)	Depression	Segments
Uyulan et al. (2021)	Depression	Unclear
Xie et al. (2020)	Depression	Unclear
Zhang et al. (2020)	Depression	Segments
Khare et al. (2021)	Parkinson's	Segments
Lee et al. (2019)	Parkinson's	Segments
Loh et al. (2021)	Parkinson's	Segments
Oh et al. (2020)	Parkinson's	Segments
Shaban (2021)	Parkinson's	Segments
Shaban and Amara (2022)	Parkinson's	Subjects
Shi et al. (2019)	Parkinson's	Subjects
Ahmedt-Aristizabal et al. (2020)	Schizophrenia	Subjects
Chu et al. (2017)	Schizophrenia	Segments
Oh et al. (2019)	Schizophrenia	Both
Shalhaf et al. (2020)	Schizophrenia	Segments
Acharya et al. (2018)	Seizure	Segments
Avcu et al. (2019)	Seizure	Subjects
Choi et al. (2019)	Seizure	Subjects
Daoud and Bayoumi (2019)	Seizure	Segments
Emami et al. (2019)	Seizure	Subjects
Fürbass et al. (2020)	Seizure	Subjects
Gao et al. (2020)	Seizure	Segments
Hussein et al. (2019)	Seizure	Segments
Iešmantas and Alzbutas (2020)	Seizure	Subjects

Jana et al. (2020)	Seizure	Segments
Khan et al. (2017)	Seizure	Segments
Li et al. (2020)	Seizure	Segments
Liang et al. (2020)	Seizure	Segments
Raghu et al. (2020)	Seizure	Unclear
Rashed-Al-Mahfuz et al. (2021)	Seizure	Segments
Truong et al. (2018)	Seizure	Segments
Ullah et al. (2018)	Seizure	Segments
Wei et al. (2018)	Seizure	Segments
Wei et al. (2019)	Seizure	Segments
Zhao et al. (2020)	Seizure	Segments
Zhou et al. (2018)	Seizure	Segments
Bouallegue et al. (2020)	Seizure and autism	Segments



## AD model definition

Listing 1: AD model definition

```
AD_model = tf.keras.models.Sequential([
    tf.keras.layers.Conv1D(
        input_shape = (500, 19),
        filters = 5,
        kernel_size = (20),
        activation = 'relu',
    ),
    tf.keras.layers.MaxPool1D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Conv1D(
        filters = 10,
        kernel_size = (10),
        activation = 'relu',
    ),
    tf.keras.layers.MaxPool1D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Conv1D(
        filters = 10,
        kernel_size = (10),
        activation = 'relu',
    ),
    tf.keras.layers.MaxPool1D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Conv1D(
        filters = 15,
        kernel_size = (5),
        activation = 'relu',
    ),
    tf.keras.layers.MaxPool1D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(20, activation = 'relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(10, activation = 'relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(2, activation = 'softmax')
])
```

## Seizure model definition

Listing 2: Seizure model definition

```
seizure_model = tf.keras.models.Sequential([
    tf.keras.layers.Conv2D(
        input_shape = (224, 224, 3),
        filters = 32,
        kernel_size = 3,
        activation = 'relu',
    ),
    tf.keras.layers.MaxPooling2D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Conv2D(
        filters = 32,
        kernel_size = 3,
        activation = 'relu',
    ),
    tf.keras.layers.MaxPooling2D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Conv2D(
        filters = 64,
        kernel_size = 3,
        activation = 'relu',
    ),
    tf.keras.layers.MaxPooling2D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Conv2D(
        filters = 64,
        kernel_size = 3,
        activation = 'relu',
    ),
    tf.keras.layers.MaxPooling2D(pool_size = 2, strides = 2),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(256, activation = 'relu'),
    tf.keras.layers.Dense(512, activation = 'relu'),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(2, activation = 'softmax')
])
```

## References

- [1] U. Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Hojjat Adeli. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, 100:270–278, September 2018. ISSN 00104825. doi:10.1016/j.compbiomed.2017.09.017. URL <https://linkinghub.elsevier.com/retrieve/pii/S0010482517303153>.
- [2] U. Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Hojjat Adeli, and D. P Subha. Automated EEG-based screening of depression using deep convolutional neural network. *Computer Methods*

- and Programs in Biomedicine*, 161:103–113, July 2018. ISSN 01692607. doi:10.1016/j.cmpb.2018.04.012. URL <https://linkinghub.elsevier.com/retrieve/pii/S0169260718301494>.
- [3] Amirmasoud Ahmadi, Mehrdad Kashefi, Hassan Shahrokhi, and Mohammad Ali Nazari. Computer aided diagnosis system using deep convolutional neural networks for adhd subtypes. *Biomedical Signal Processing and Control*, 63:102227, 2021.
- [4] David Ahmedt-Aristizabal, Tharindu Fernando, Simon Denman, Jonathan Edward Robinson, Sridha Sridharan, Patrick J Johnston, Kristin R Laurens, and Clinton Fookes. Identification of children at risk of schizophrenia via deep learning and eeg responses. *IEEE journal of biomedical and health informatics*, 25(1):69–76, 2020.
- [5] Mustafa Talha Avcu, Zhuo Zhang, and Derrick Wei Shih Chan. Seizure detection using least eeg channels by deep convolutional neural network. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1120–1124. IEEE, 2019.
- [6] Betul Ay, Ozal Yildirim, Muhammed Talo, Ulas Baran Baloglu, Galip Aydin, Subha D Puthankattil, and U Rajendra Acharya. Automated depression detection using deep representation and sequence learning with eeg signals. *Journal of Medical Systems*, 43(7):1–12, 2019.
- [7] Mohamadreza Bakhtyari and Sayeh Mirzaei. Adhd detection using dynamic connectivity patterns of eeg data and convlstm with attention framework. *Biomedical Signal Processing and Control*, 76:103708, 2022.
- [8] David Bergeron, Kelsey Flynn, Louis Verret, Stéphane Poulin, Rémi W Bouchard, Christian Bocti, Tamàs Fülöp, Guy Lacombe, Serge Gauthier, Ziad Nasreddine, et al. Multicenter validation of an mmse-mo ca conversion table. *Journal of the American Geriatrics Society*, 65(5):1067–1072, 2017.
- [9] Xiaojun Bi and Haibo Wang. Early Alzheimer’s disease diagnosis based on EEG spectral images using deep learning. *Neural Networks*, 114:119–135, June 2019. ISSN 08936080. doi:10.1016/j.neunet.2019.02.005. URL <https://linkinghub.elsevier.com/retrieve/pii/S0893608019300486>.
- [10] Ghaith Bouallegue, Ridha Djemal, Saleh A. Alshebeili, and Hesham Aldhalaan. A Dynamic Filtering DF-RNN Deep-Learning-Based Approach for EEG-Based Neurological Disorders Diagnosis. *IEEE Access*, 8:206992–207007, 2020. ISSN 2169-3536. doi:10.1109/ACCESS.2020.3037995. Conference Name: IEEE Access.

- [11] Yang Chang, Cory Stevenson, I-Chun Chen, Dar-Shong Lin, and Li-Wei Ko. Neurological state changes indicative of adhd in children learned via eeg-based lstm networks. *Journal of Neural Engineering*, 19(1):016021, 2022.
- [12] He Chen, Yan Song, and Xiaoli Li. A deep learning framework for identifying children with adhd using an eeg-based brain network. *Neurocomputing*, 356:83–96, 2019.
- [13] He Chen, Yan Song, and Xiaoli Li. Use of deep learning to detect personalized spatial-frequency abnormalities in eegs of children with adhd. *Journal of Neural Engineering*, 16(6):066046, 2019.
- [14] Gwangho Choi, Chulkyun Park, Junkyung Kim, Kyoungin Cho, Tae-Joon Kim, HwangSik Bae, Kyeongyuk Min, Ki-Young Jung, and Jongwha Chong. A novel multi-scale 3d cnn with deep neural network for epileptic seizure detection. In *2019 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–2. IEEE, 2019.
- [15] Lei Chu, Robert Qiu, Haichun Liu, Zenan Ling, Tianhong Zhang, and Jijun Wang. Individual recognition in schizophrenia using deep learning methods with random forest and voting classifiers: Insights from resting state eeg streams. *arXiv preprint arXiv:1707.03467*, 2017.
- [16] Hisham Daoud and Magdy A Bayoumi. Efficient epileptic seizure prediction based on deep learning. *IEEE transactions on biomedical circuits and systems*, 13(5):804–813, 2019.
- [17] Mateo de Bardeci, Cheng Teng Ip, and Sebastian Olbrich. Deep learning applied to electroencephalogram data in mental disorders: A systematic review. *Biological Psychology*, 162:108117, 2021.
- [18] Matteo Demuru and Matteo Frascini. EEG fingerprinting: Subject-specific signature based on the aperiodic component of power spectrum. *Computers in Biology and Medicine*, 120:103748, 2020.
- [19] Paolo Detti. *Siena Scalp EEG Database (version 1.0.0)*. PhysioNet, 2020. URL <https://doi.org/10.13026/5d4a-j060>.
- [20] Paolo Detti, Giampaolo Vatti, and Garazi Zabalo Manrique de Lara. Eeg synchronization analysis for seizure prediction: A study on data of noninvasive recordings. *Processes*, 8(7):846, 2020.
- [21] Laura Dubreuil-Vall, Giulio Ruffini, and Joan A Camprodon. Deep learning convolutional neural networks discriminate adult adhd from healthy individuals on the basis of event-related spectral eeg. *Frontiers in Neuroscience*, 14:251, 2020.

- [22] Ali Emami, Naoto Kunii, Takeshi Matsuo, Takashi Shinozaki, Kensuke Kawai, and Hirokazu Takahashi. Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images. *NeuroImage: Clinical*, 22:101684, 2019.
- [23] Marshal F Folstein, Susan E Folstein, and Paul R McHugh. “mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, 1975.
- [24] Franz Fürbass, Mustafa Aykut Kural, Gerhard Gritsch, Manfred Hartmann, Tilmann Kluge, and Sándor Beniczky. An artificial intelligence-based eeg algorithm for detection of epileptiform eeg discharges: validation against the diagnostic gold standard. *Clinical Neurophysiology*, 131(6):1174–1179, 2020.
- [25] Aarthi S Ganapathi, Ryan M Glatt, Tess H Bookheimer, Emily S Popa, Morgan L Ingemanson, Casey J Richards, John F Hodes, Kyron P Pierce, Colby B Slyapich, Fatima Iqbal, et al. Differentiation of subjective cognitive decline, mild cognitive impairment, and dementia using qEEG/ERP-based cognitive testing and volumetric MRI in an outpatient specialty memory clinic. *Journal of Alzheimer’s Disease*, pages 1–9, 2022 (preprint).
- [26] Yunyuan Gao, Bo Gao, Qiang Chen, Jia Liu, and Yingchun Zhang. Deep convolutional neural network-based epileptic electroencephalogram (eeg) signal classification. *Frontiers in Neurology*, 11:375, 2020.
- [27] Georgios Gkenios, Konstantina Latsiou, Konstantinos Diamantaras, Ioanna Chouvarda, and Magda Tsolaki. Diagnosis of alzheimer’s disease and mild cognitive impairment using eeg and recurrent neural networks. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3179–3182. IEEE, 2022.
- [28] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- [29] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz, and Hugo JWL Aerts. Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8):500–510, 2018.
- [30] Cameron J Huggins, Javier Escudero, Mario A Parra, Brian Scally, Renato Anghinah, Amanda Vitória Lacerda De Araújo, Luis F Basile, and Daniel Abasolo. Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer’s disease, mild cognitive impairment and healthy ageing. *Journal of Neural Engineering*, 18(4):046087, August 2021. ISSN 1741-2560, 1741-2552.

doi:10.1088/1741-2552/ac05d8. URL <https://iopscience.iop.org/article/10.1088/1741-2552/ac05d8>.

- [31] Ramy Hussein, Hamid Palangi, Rabab K Ward, and Z Jane Wang. Optimized deep neural network architecture for robust detection of epileptic seizures using eeg signals. *Clinical Neurophysiology*, 130(1):25–37, 2019.
- [32] Cosimo Ieracitano, Nadia Mammone, Alessia Bramanti, Amir Hussain, and Francesco C. Morabito. A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. *Neurocomputing*, 323:96–107, January 2019. ISSN 09252312. doi:10.1016/j.neucom.2018.09.071. URL <https://linkinghub.elsevier.com/retrieve/pii/S0925231218311524>.
- [33] Tomas Iešmantas and Robertas Alzbutas. Convolutional neural network for detection and classification of seizures in clinical data. *Medical & Biological Engineering & Computing*, 58(9):1919–1932, 2020.
- [34] Ranjan Jana, Siddhartha Bhattacharyya, and Swagatam Das. Patient-specific seizure prediction using the convolutional neural networks. In *Intelligence Enabled Research*, pages 51–60. Springer, 2020.
- [35] Hussam Kaka, Euan Zhang, and Nazir Khan. Artificial intelligence and deep learning in neuroradiology: exploring the new frontier. *Canadian Association of Radiologists Journal*, 72(1):35–44, 2021.
- [36] Shachar Kaufman, Saharon Rosset, Claudia Perlich, and Ori Stitelman. Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(4):1–21, 2012.
- [37] Haidar Khan, Lara Marcuse, Madeline Fields, Kalina Swann, and Bülent Yener. Focal onset seizure prediction using convolutional networks. *IEEE Transactions on Biomedical Engineering*, 65(9):2109–2118, 2017.
- [38] Smith K. Khare, Varun Bajaj, and U. Rajendra Acharya. PDCNNNet: An Automatic Framework for the Detection of Parkinson’s Disease Using EEG Signals. *IEEE Sensors Journal*, 21(15):17017–17024, August 2021. ISSN 1558-1748. doi:10.1109/JSEN.2021.3080135. Conference Name: IEEE Sensors Journal.
- [39] Donghyeon Kim and Kiseon Kim. Detection of Early Stage Alzheimer’s Disease using EEG Relative Power with Deep Neural Network. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 352–355, July 2018. doi:10.1109/EMBC.2018.8512231. ISSN: 1558-4615.

- [40] Seonho Kim, Jungjoon Kim, and Hong-Woo Chun. Wave2vec: Vectorizing electroencephalography bio-signal for prediction of brain disease. *International Journal of Environmental Research and Public Health*, 15(8):1750, 2018.
- [41] Hyunjin Kwon, Seokhwan Kang, Wonjun Park, Jinhyeok Park, and Youngho Lee. Deep learning based pre-screening method for depression with imagery frontal eeg channels. In *2019 International conference on information and communication technology convergence (ICTC)*, pages 378–380. IEEE, 2019.
- [42] Kenneth M Langa and Deborah A Levine. The diagnosis and management of mild cognitive impairment: a clinical review. *Journal of the American Medical Association*, 312(23):2551–2561, 2014.
- [43] Soojin Lee, Ramy Hussein, and Martin J. McKeown. A Deep Convolutional-Recurrent Neural Network Architecture for Parkinson’s Disease EEG Classification. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–4, Ottawa, ON, Canada, November 2019. IEEE. ISBN 978-1-72812-723-1. doi:10.1109/GlobalSIP45357.2019.8969309. URL <https://ieeexplore.ieee.org/document/8969309/>.
- [44] Xiaowei Li, Rong La, Ying Wang, Junhong Niu, Shuai Zeng, Shuting Sun, and Jing Zhu. Eeg-based mild depression recognition using convolutional neural network. *Medical & Biological Engineering & Computing*, 57(6):1341–1352, 2019.
- [45] Xiaowei Li, Rong La, Ying Wang, Bin Hu, and Xuemin Zhang. A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography. *Frontiers in Neuroscience*, 14:192, 2020.
- [46] Yang Li, Yu Liu, Wei-Gang Cui, Yu-Zhu Guo, Hui Huang, and Zhong-Yi Hu. Epileptic seizure detection in eeg signals using a unified temporal-spectral squeeze-and-excitation network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(4):782–794, 2020.
- [47] Weixia Liang, Haijun Pei, Qingling Cai, and Yonghua Wang. Scalp eeg epileptogenic zone recognition and localization based on long-term recurrent convolutional network. *Neurocomputing*, 396:569–576, 2020.
- [48] Hui Wen Loh, Chui Ping Ooi, Elizabeth Palmer, Prabal Datta Barua, Sengul Dogan, Turker Tuncer, Mehmet Baygin, and U. Rajendra Acharya. GaborPDNet: Gabor Transformation and Deep Neural Network for Parkinson’s Disease Detection Using EEG Signals. *Electronics*, 10(14):1740, January 2021. ISSN 2079-9292. doi:10.3390/electronics10141740. URL <https://www.mdpi.com/2079-9292/10/14/1740>. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

- [49] Majid Mafi and Shokoufeh Radfar. High dimensional convolutional neural network for eeg connectivity-based diagnosis of adhd. *Journal of Biomedical Physics and Engineering*, 12(6):645–654, 2022.
- [50] Christoph M Michel and Bin He. Eeg source localization. *Handbook of Clinical Neurology*, 160:85–101, 2019.
- [51] Majid Moghaddari, Mina Zolfy Lighvan, and Sebelan Danishvar. Diagnose adhd disorder in children using convolutional neural network based on continuous mental task eeg. *Computer Methods and Programs in Biomedicine*, 197:105738, 2020.
- [52] Francesco Carlo Morabito, Maurizio Campolo, Cosimo Ieracitano, Javad Mohammad Ebadi, Lilla Bonanno, Alessia Bramanti, Simona Desalvo, Nadia Mammone, and Placido Bramanti. Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer’s disease patients from scalp EEG recordings. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pages 1–6, September 2016. doi:10.1109/RTSI.2016.7740576.
- [53] Wajid Mumtaz and Abdul Qayyum. A deep learning framework for automatic diagnosis of unipolar depression. *International journal of medical informatics*, 132:103983, 2019.
- [54] Ziad S Nasreddine, Natalie A Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L Cummings, and Howard Chertkow. The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4):695–699, 2005.
- [55] Shu Lih Oh, Jahmunah Vicnesh, Edward J Ciaccio, Rajamanickam Yuvaraj, and U Rajendra Acharya. Deep convolutional neural network model for automated diagnosis of schizophrenia using eeg signals. *Applied Sciences*, 9(14):2870, 2019.
- [56] Shu Lih Oh, Yuki Hagiwara, U. Raghavendra, Rajamanickam Yuvaraj, N. Arunkumar, M. Murugappan, and U. Rajendra Acharya. A deep learning approach for Parkinson’s disease diagnosis from EEG signals. *Neural Computing and Applications*, 32(15):10927–10933, August 2020. ISSN 0941-0643, 1433-3058. doi:10.1007/s00521-018-3689-5. URL <http://link.springer.com/10.1007/s00521-018-3689-5>.
- [57] Shivarudhrappa Raghu, Natarajan Sriraam, Yasin Temel, Shyam Vasudeva Rao, and Pieter L Kubben. Eeg based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Networks*, 124:202–212, 2020.



- [58] Md. Rashed-Al-Mahfuz, Mohammad Ali Moni, Shahadat Uddin, Salem A. Alyami, Matthew A. Summers, and Valsamma Eapen. A Deep Convolutional Neural Network Method to Detect Seizures and Characteristic Frequencies Using Epileptic Electroencephalogram (EEG) Data. *IEEE Journal of Translational Engineering in Health and Medicine*, 9:1–12, 2021. ISSN 2168-2372. doi:10.1109/JTEHM.2021.3050925. URL <https://ieeexplore.ieee.org/document/9319690/>.
- [59] Saharon Rosset, Claudia Perlich, Grzegorz Świrszcz, Prem Melville, and Yan Liu. Medical data mining: insights from winning two competitions. *Data Mining and Knowledge Discovery*, 20(3):439–468, 2010.
- [60] Mohamed Shaban. Automated Screening of Parkinson’s Disease Using Deep Learning Based Electroencephalography. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 158–161, May 2021. doi:10.1109/NER49283.2021.9441065. ISSN: 1948-3554.
- [61] Mohamed Shaban and Amy W. Amara. Resting-state electroencephalography based deep-learning for the detection of Parkinson’s disease. *PLOS ONE*, 17(2):e0263159, February 2022. ISSN 1932-6203. doi:10.1371/journal.pone.0263159. URL <https://dx.plos.org/10.1371/journal.pone.0263159>.
- [62] Ahmad Shalhaf, Sara Bagherzadeh, and Arash Maghsoudi. Transfer learning with deep convolutional neural network for automated detection of schizophrenia from eeg signals. *Physical and Engineering Sciences in Medicine*, 43(4):1229–1239, 2020.
- [63] Xinjie Shi, Tianqi Wang, Lan Wang, Hanjun Liu, and Nan Yan. Hybrid Convolutional Recurrent Neural Networks Outperform CNN and RNN in Task-state EEG Detection for Parkinson’s Disease. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 939–944, Lanzhou, China, November 2019. IEEE. ISBN 978-1-72813-248-8. doi:10.1109/APSIPAASC47483.2019.9023190. URL <https://ieeexplore.ieee.org/document/9023190/>.
- [64] Behrad TaghiBeyglou, Ashkan Shahbazi, Fatemeh Bagheri, Sina Akbarian, and Mehran Jahed. Detection of adhd cases using cnn and classical classifiers of raw eeg. *Computer Methods and Programs in Biomedicine Update*, page 100080, 2022.
- [65] Iulian Emil Tampu, Anders Eklund, and Neda Haj-Hosseini. Inflation of test accuracy due to data leakage in deep learning-based classification of oct images. *Scientific Data*, 9(580), 2022. doi:10.1038/s41597-022-01618-6.
- [66] Mustafa Tosun. Effects of spectral features of eeg signals recorded with different channels and recording

- statuses on adhd classification with deep learning. *Physical and Engineering Sciences in Medicine*, 44(3):693–702, 2021.
- [67] Nhan Duy Truong, Anh Duy Nguyen, Levin Kuhlmann, Mohammad Reza Bonyadi, Jiawei Yang, Samuel Ippolito, and Omid Kavehei. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105:104–111, 2018.
- [68] Ihsan Ullah, Muhammad Hussain, Hatim Aboalsamh, et al. An automated system for epilepsy detection using eeg brain signals based on deep learning approach. *Expert Systems with Applications*, 107:61–71, 2018.
- [69] Caglar Uyulan, Türker Tekin Ergüzel, Huseyin Unubol, Merve Cebi, Gokben Hizli Sayar, Mahdi Nezhad Asad, and Nevzat Tarhan. Major depressive disorder classification based on different convolutional neural network models: Deep learning approach. *Clinical EEG and Neuroscience*, 52(1):38–51, 2021.
- [70] Amirali Vahid, Annet Bluschke, Veit Roessner, Sebastian Stober, and Christian Beste. Deep learning based on event-related eeg differentiates children with adhd from healthy controls. *Journal of Clinical Medicine*, 8(7):1055, 2019.
- [71] Xiaoyan Wei, Lin Zhou, Ziyi Chen, Liangjun Zhang, and Yi Zhou. Automatic seizure detection using three-dimensional cnn based on multi-channel eeg. *BMC medical informatics and decision making*, 18(5):71–80, 2018.
- [72] Xiaoyan Wei, Lin Zhou, Zhen Zhang, Ziyi Chen, and Yi Zhou. Early prediction of epileptic seizures using a long-term recurrent convolutional network. *Journal of Neuroscience Methods*, 327:108395, 2019.
- [73] Dong Wen, Zhenhao Wei, Yanhong Zhou, Guolin Li, Xu Zhang, and Wei Han. Deep learning methods to process fMRI data and their application in the diagnosis of cognitive impairment: a brief overview and our opinion. *Frontiers in Neuroinformatics*, 12:23, 2018.
- [74] Junhao Wen, Elina Thibeau-Sutre, Mauricio Diaz-Melo, Jorge Samper-González, Alexandre Routier, Simona Bottani, Didier Dormont, Stanley Durrleman, Ninon Burgos, Olivier Colliot, et al. Convolutional neural networks for classification of Alzheimer’s disease: Overview and reproducible evaluation. *Medical Image Analysis*, 63:101694, 2020.
- [75] Yunlong Xie, Banghua Yang, Xi Lu, Minmin Zheng, Cunxiu Fan, Xiaoying Bi, Yingjie Li, et al. Anxiety and depression diagnosis method based on brain networks and convolutional neural networks. In *2020*

- 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1503–1506. IEEE, 2020.
- [76] Zeng You, Runhao Zeng, Xiaoyong Lan, Huixia Ren, Zhiyang You, Xue Shi, Shipeng Zhao, Yi Guo, Xin Jiang, and Xiping Hu. Alzheimer’s Disease Classification With a Cascade Neural Network. *Frontiers in Public Health*, 8:584387, November 2020. ISSN 2296-2565. doi:10.3389/fpubh.2020.584387. URL <https://www.frontiersin.org/articles/10.3389/fpubh.2020.584387/full>.
- [77] Xiaowei Zhang, Junlei Li, Kechen Hou, Bin Hu, Jian Shen, and Jing Pan. Eeg-based depression detection using convolutional neural network with demographic attention mechanism. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 128–133. IEEE, 2020.
- [78] Wei Zhao, Wenbing Zhao, Wenfeng Wang, Xiaolu Jiang, Xiaodong Zhang, Yonghong Peng, Baocan Zhang, and Guokai Zhang. A novel deep neural network for robust detection of seizures using eeg signals. *Computational and Mathematical Methods in Medicine*, 2020, 2020.
- [79] Yilu Zhao and Lianghua He. Deep Learning in the EEG Diagnosis of Alzheimer’s Disease. In C.V. Jawahar and Shiguang Shan, editors, *Computer Vision - ACCV 2014 Workshops*, Lecture Notes in Computer Science, pages 340–353, Cham, 2015. Springer International Publishing. ISBN 978-3-319-16628-5. doi:10.1007/978-3-319-16628-5\_25.
- [80] Dingfu Zhou, Zhihang Liao, and Rong Chen. Deep learning enabled diagnosis of children’s adhd based on the big data of video screen long-range eeg. *Journal of Healthcare Engineering*, 2022, 2022.
- [81] Mengni Zhou, Cheng Tian, Rui Cao, Bin Wang, Yan Niu, Ting Hu, Hao Guo, and Jie Xiang. Epileptic seizure detection based on eeg signals and cnn. *Frontiers in Neuroinformatics*, 12:95, 2018.