
Unravelling the neural signatures of dream recall in EEG: a deep learning approach

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Dreams and our ability to recall them are among the most puzzling questions
2 in sleep research. Specifically, putative differences in brain network dynamics
3 between individuals with high versus low dream recall rates, are still poorly under-
4 stood. In this study, we addressed this question as a classification problem where
5 we applied deep convolutional networks (CNN) to sleep EEG recordings to predict
6 whether subjects belonged to the high or low dream recall group (HDR and LDR
7 resp.). Our model achieves significant accuracy levels across all the sleep stages,
8 thereby indicating subtle signatures of dream recall in the sleep microstructure. We
9 also visualized the feature space to inspect the subject-specificity of the learned fea-
10 tures, thus ensuring that the network captured population level differences. Beyond
11 being the first study to apply deep learning to sleep EEG in order to classify HDR
12 and LDR, guided backpropagation allowed us to visualize the most discriminant
13 features in each sleep stage. The significance of these findings and future directions
14 are discussed.

15 1 Introduction

16 Dreams are arguably one of the most intriguing phenomena in cognitive science. Previous studies have
17 explored the differences in brain EEG patterns between HDR and LDR using different approaches
18 such as sleep spindles [1] and Event-Related Potentials [2] to identify the neural bases of dreaming.
19 Although these findings have advanced the field in understanding the neural bases of dreaming, neural
20 correlates of dream recall are still not completely understood.

21 In recent years, deep learning (DL) has proven to be one of the most successful techniques for
22 classification problems, with specific applications in computer vision, natural language understanding
23 as well as EEG signal decoding [3]. DL methods allow the identification of optimal discriminative
24 patterns in a given minimally pre-processed dataset, thus reducing the reliance on *a priori* selection
25 of features. However, the interpretability of such deep models has been a major roadblock in the
26 context of neuroimaging applications.

27 In this work, we used a convolutional neural network (CNN) for classification of sleep EEG recordings
28 into two groups (HDR vs LDR). Subsequently, we explored various techniques to visualize the
29 features learned by the network.

30 2 Materials and Methods

31 2.1 Data Collection

32 The sleep study consisted of 36 participants (18 male, mean age 23 ± 3 yrs). Brief interview was
33 conducted to determine which dream recall group the subject belonged. Those who reported more

34 that three dream recalls per week were classified as HDR and those who reported less than twice a
 35 month were categorized as LDR. Participants were made to sleep in an acoustically dampened and
 36 electrically shielded room.
 37 Twenty-one AG/AgCl scalp electrodes were manually positioned according to the extended Interna-
 38 tional 10-20 System (Fz,Cz,Pz,FP1,F3,FC1,C3,T3,CP1,P3,M1,O1). The electrophysiological data
 39 (EEG, EOG, and EMG) were continuously recorded via a BrainAmp system (Brain Products GmbH,
 40 Germany) with an amplification gain of 12,500, a high-pass filter of 0.1 Hz and a sampling rate of
 41 1000 Hz with an anti-aliasing low-pass filter.

42 2.2 Sleep stage scoring and Data Preprocessing

43 Sleep stages were scored visually by two sleep experts using
 44 Rechtschaffen and Kales guidelines. The data was then split into
 45 5s segments, each labeled as one of the 5 different vigilance state
 46 (wake, rapid eyes movements sleep – REM sleep, sleep stage 1 – S1,
 47 sleep stage 2 – S2 and slow wave sleep – SWS).
 48 For our analysis, the data was downsampled to 200Hz. It was
 49 also passed through a low pass filter to exclude frequencies greater
 50 than 40Hz. Each data point was considered as an array of
 51 $19(\text{eegchannels}) \times 1000(\text{timepoints})$ [5 second segment with
 52 200Hz]. The number of data points available for each sleep segment
 53 is mentioned in Table 1.

Table 1: Available Data

Sleep stage	Data points
S2	61860
SWS	52548
REM	35322
S1	15120
AWA	14568

54 2.3 Model Architecture and training

55 We developed a CNN architecture for classifying each data point into one of the two dream recall
 56 groups. The model architecture is shown in Figure 1. The network contains two parts – a feature
 extractor and a classifier.

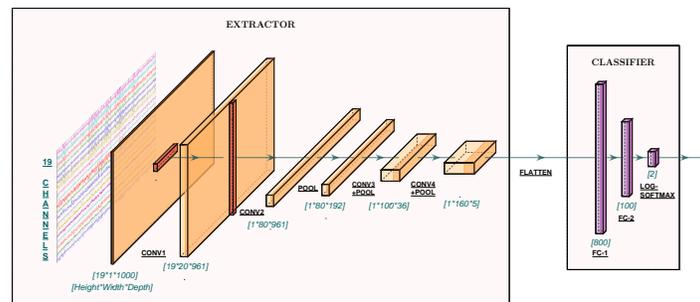


Figure 1: Model architecture

57
 58 The preprocessed data (Section 2.2) is passed through the extractor (architecture described in 2). The
 59 first convolutional layer has a filter size of 1×40 , which enables the network to capture temporal
 60 patterns of dream recall in each electrode. The second convolutional layer has a filter size of 19×1 ,
 61 which enables the network to capture spatial patterns across the recording electrodes. The output of
 62 the extractor is a 1-D feature vector of size 800, which is subsequently input to the classifier module
 63 (Figure 1). The classifier output represents log probabilities of the data sample to belong to the HDR
 64 or LDR classes.

65 The model was implemented using PyTorch 1.1 in Python. The negative log likelihood cost function
 66 was used to train the network and was then backpropogated through the classifier and extractor model.
 67 Adam optimizer was used for training and the learning rate was decreased using Step Scheduler.
 68 Early stopping was employed to prevent overfitting and the best model was stored for validation. The
 69 model hyperparameters used for training were Batch size: 100, Epochs: 150, Learning rate: 0.00005,
 70 Rate decay step size: 25 and Rate decay gamma: 0.7.

Table 2: Extractor network architecture

Layer	Type	Filter Size	Output Size
0	Input	-	[1,19,1000]
1	Conv+BatchNorm	[1,40]	[20,19,961]
2	Conv+BatchNorm+Relu	[19,1]	[80,1,961]
3	MaxPool	[1,5]	[80,1,192]
4	Conv+BatchNorm+Relu	[1,5]	[100,1,184]
5	MaxPool+Dropout	[1,5]	[100,1,36]
6	Conv+BatchNorm+Relu	[1,10]	[160,1,27]
7	MaxPool	[1,5]	[160,1,5]
8	Flatten	-	[800]

71 3 Results and Discussion

72 3.1 Decoding Accuracy

73 To test the model’s cross subject accuracy, we used a leave 2-subjects-out strategy, wherein the model
 74 was trained on 34 subjects and tested on 2 held-out subjects (1 each from both dream recall groups).
 75 18-Fold cross validation was done, and in each case 2 different subjects were a part of the testing set.

76 Figure 2 summarizes the decoding accuracy across the 5 sleep stages. We see that our deep model
 77 significantly outperforms the logistic regression model on all the sleep stages except S1 (might be
 78 attributed to the fact that less data was available for this stage). The model achieves decoding accuracy
 79 of around **70 to 75%** on the other 4 sleep stages.

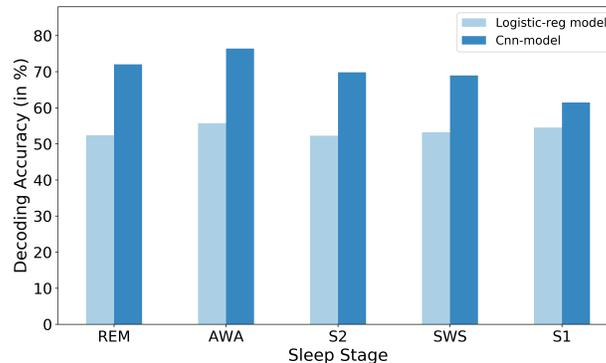
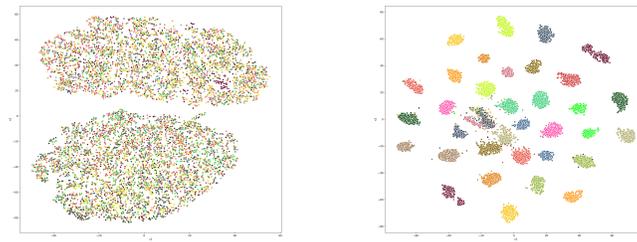


Figure 2: 18-fold cross validation accuracy compared across five sleep stages

80 3.2 tSNE Visualization to check subject overfitting

81 To assess the subject specificity of the features extracted by the network, we visualized a low-
 82 dimensional representation of the feature space using t-Distributed Stochastic Neighbor Embedding
 83 (t-SNE) [4]. To this end, we generated the tSNE plots for the output of the extractor part of the
 84 network, as shown in Fig. 3.

85 The left image corresponds to the feature space learned by our trained model. There were no visible
 86 clusters specific to a subject (each subject being represented by a color). We compared this plot to the
 87 feature space learned by a network trained to identify subjects from the EEG recordings (the right
 88 image). The formation of subject-specific clusters corresponds to the extracted features containing
 89 subject-specific information. This analysis confirmed that our proposed network did not learn features
 90 based on subject-specific information. Therefore, the learned feature space corresponds to population
 91 level differences between the two groups.



(a) Feature space learned by proposed CNN model (b) Feature space learned by a model trained to identify subjects

Figure 3: tSNE embeddings calculated using outputs of feature extractor. Colours correspond to the subject label of that datapoint.

92 3.3 Electrode analysis using Guided Backpropagation

93 We used Guided backpropagation (GB) [5] technique to visualize the topological distribution of
 94 discriminative neural signatures of dream recall. Positive saliency map obtained using GB was used
 95 to infer the electrode importance. It was found that the electrode importance was similar across
 96 different validation sets for each sleep stage. The results obtained for various sleep stages are plotted
 97 in Figure 4.

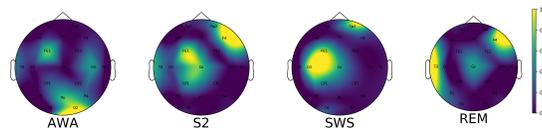


Figure 4: Electrode importance for dream recall classification as obtained from Guided-backpropagation. High intensity values correspond to higher importance given to that electrode.

98 4 Conclusions

99 This study illustrates how deep learning can be used to data-mine the neural activities of HDR and
 100 LDR. Specifically, we trained a deep CNN to classify between LDR and HDR and achieved significant
 101 decoding accuracies. Furthermore, we used tSNE to check for subject overfitting and GB to identify
 102 the brain regions carrying group-specific differences, thus illustrating the use of visualisation tools
 103 for deep models trained on neural data. Future work will involve explorations of other dimensions,
 104 including the frequency components of the data that contribute most to the classification.

105 References

- 106 [1] Tore Nielsen, Michelle Carr, Cloé Blanchette-Carrière, Louis-Philippe Marquis, Gaëlle Dumel, Elizaveta
 107 Solomonova, Sarah-Hélène Julien, Claudia Picard-Deland, and Tyna Paquette. Nrem sleep spindles are
 108 associated with dream recall. *Sleep Spindles & Cortical Up States*, 1(1):27–41, 2016.
- 109 [2] Jean-Baptiste Eichenlaub, Alain Nicolas, Jérôme Daltrozzo, Jérôme Redouté, Nicolas Costes, and Perrine
 110 Ruby. Resting brain activity varies with dream recall frequency between subjects. *Neuropsychopharmacol-*
 111 *ogy*, 39(7):1594, 2014.
- 112 [3] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn
 113 Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural*
 114 *engineering*, 2019.
- 115 [4] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning*
 116 *Research*, 9:2579–2605, 2008.
- 117 [5] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity:
 118 The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.