



Published in final edited form as:

J Neurosci Methods. 2022 January 15; 366: 109421. doi:10.1016/j.jneumeth.2021.109421.

Automated sleep state classification of wide-field calcium imaging data via multiplex visibility graphs and deep learning

Xiaohui Zhang^{a,1}, Eric C. Landsness^{b,1}, Wei Chen^b, Hanyang Miao^b, Michelle Tang^b, Lindsey M. Brier^c, Joseph P. Culver^{c,d,e,f}, Jin-Moo Lee^{b,c,d}, Mark A. Anastasio^{a,*}

^aDepartment of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

^bDepartment of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA

^cDepartment of Radiology, Washington University School of Medicine, St. Louis, MO 63110, USA

^dDepartment of Biomedical Engineering, Washington University School of Engineering, St. Louis, MO 63130, USA

^eDepartment of Electrical and Systems Engineering, Washington University School of Engineering, St. Louis, MO 63130, USA

^fDepartment of Physics, Washington University School of Arts and Science, St. Louis, MO 63130, USA

Abstract

Background: Wide-field calcium imaging (WFCI) allows for monitoring of cortex-wide neural dynamics in mice. When applied to the study of sleep, WFCI data are manually scored into the sleep states of wakefulness, non-REM (NREM) and REM by use of adjunct EEG and EMG recordings. However, this process is time-consuming and often suffers from low inter- and intra-rater reliability and invasiveness. Therefore, an automated sleep state classification method that operates on WFCI data alone is needed.

New method: A hybrid, two-step method is proposed. In the first step, spatial-temporal WFCI data is mapped to multiplex visibility graphs (MVGs). Subsequently, a two-dimensional

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Correspondence to: Department of Bioengineering, University of Illinois, Urbana-Champaign, IL, USA. maa@illinois.edu (M.A. Anastasio).

¹These authors contributed equally.

CRedit authorship contribution statement

Xiaohui Zhang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Eric C. Landsness:** Conceptualization, Resources, Data curation, Validation, Writing – original draft, Supervision, Funding acquisition, Software, Formal analysis. **Wei Chen:** Resources, Data Curation. **Hanyang Miao:** Resources, Writing – review & editing. **Michelle Tang:** Resources. **Lindsey M. Brier:** Resources, Data curation, Writing – review & editing. **Joseph P. Culver:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jin-Moo Lee:** Conceptualization, Writing – review & editing, Supervision, Project administration, funding acquisition. **Mark A. Anastasio:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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.

Appendix C. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jneumeth.2021.109421.

convolutional neural network (2D CNN) is employed on the MVGs to be classified as wakefulness, NREM and REM.

Results: Sleep states were classified with an accuracy of 84% and Cohen's κ of 0.67. The method was also effectively applied on a binary classification of wakefulness/sleep (accuracy=0.82, κ = 0.62) and a four-class wakefulness/sleep/anesthesia/movement classification (accuracy=0.74, κ = 0.66). Gradient-weighted class activation maps revealed that the CNN focused on short- and long-term temporal connections of MVGs in a sleep state-specific manner. Sleep state classification performance when using individual brain regions was highest for the posterior area of the cortex and when cortex-wide activity was considered.

Comparison with existing method: On a 3-hour WFCI recording, the MVG-CNN achieved a κ of 0.65, comparable to a κ of 0.60 corresponding to the human EEG/EMG-based scoring.

Conclusions: The hybrid MVG-CNN method accurately classifies sleep states from WFCI data and will enable future sleep-focused studies with WFCI.

Keywords

Wide-field calcium imaging; Automated sleep state classification; Multiplex visibility graph; Deep learning; 2D CNN; Local sleep

1. Introduction

Wide-field calcium imaging (WFCI) with genetically encoded calcium indicators enables recording of regional neuronal depolarization in mice across the entire cortex on a sub-second temporal scale with simultaneous examination of neurovascular coupling and cell type specificity (Ma et al., 2016a; Kozberg et al., 2016; Ma et al., 2016b; Matsui et al., 2016). Given these capabilities, WFCI has been employed to study mouse brain physiology during quiet wakefulness (Ma et al., 2016b), decision-making behavior (Allen et al., 2017), anesthesia (Wright et al., 2017) and under disease states (Balbi et al., 2019). Recently, it has also been employed to characterize the dynamics of neural activity during sleep (Brier et al., 2019). In these studies, WFCI has revealed several novel findings including sleep slow-oscillations during non-rapid eye movement (NREM) linked to changes in functional connectivity (Brier et al., 2019), selective increases in cerebral blood volume during NREM (Turner et al., 2020), and highly active neuronal subpopulations (Niethard et al., 2018; Niethard et al., 2021). Thus, WFCI is a powerful new tool to uncover the neural correlates of sleep.

To effectively apply WFCI to the study of sleep, brain neural activity must be classified into various sleep-wake states such as wakefulness, NREM and REM. Currently, sleep state classification of WFCI data relies on the simultaneous recording of electroencephalogram (EEG) and electromyogram (EMG) signals. These signals provide information on the electrical activity of the brain and on muscle tone that together allow for the unambiguous determination of sleep-wake states (Mang et al., 2014). Unfortunately, acquisition of the EEG signal requires meticulous placement of electrodes near the surface of the mouse cortex under deep anesthesia, which is invasive, increases the risk of infection, and obscures the imaging field of view. Additionally, sleep scoring by EEG/EMG is time-consuming because

it requires trained professionals to manually inspect EEG/EMG signals and subjectively assign brain states, with low inter- and intra-rater reliability (Bliwise et al., 1984; Collop, 2002; Danker-Hopfe et al., 2009; Drinnan et al., 1998; Lord et al., 1989; Loreda et al., 1999; Norman et al., 2000; Rosenberg et al., 2013; Silber et al., 2007; Whitney et al., 1998). As a result, it is highly desirable to eliminate the need for adjunct EEG/EMG data by developing a method that can classify sleep states from WFCI data alone. However, there are currently no established rules governing sleep state classification based on WFCI data, and human scoring of WFCI data is impractical due to the high number of measurements (up to thousands of pixels per time point). Therefore, an automated sleep state classification method for use with WFCI data is desired to advance sleep research with WFCI.

Sleep state scoring by human experts is based on the occurrence of discrete neuronal events such as K-complexes, spindles, theta rhythms and slow waves that are known to spread spatially and temporally across the cortex (Buzsaki, 2006; De Gennaro and Ferrara, 2003; Halász, 2016). As such, the establishment of a sleep state classification method that exploits discriminative features regarding spatiotemporal calcium dynamics obtained through WFCI and avoids the complexities of training deep spatial-temporal neural networks could be beneficial to accurate sleep state classification. One approach to describe the dynamics of multivariate time series is the multiplex visibility graph (MVG) introduced by Lacasa et al. (Lacasa et al., 2015; Lacasa et al., 2008). When applied to neural activity, a visibility graph focuses on visible peaks within time series and, therefore, can capture discrete neuronal events over time. As each layer of a visibility graph is combined across brain regions into a multiplex, the method effectively incorporates the spatial nature of the data. Several recent neuroimaging applications have employed MVGs within a feature extraction procedure to build feature vectors that can be used to accurately classify neurological disorders and brain states in modalities such as functional magnetic resonance imaging (fMRI), function near-infrared spectroscopy (fNIRS), and WFCI (Sannino et al., 2017; Zhu et al., 2018a; Zhu et al., 2018b).

Designing optimal MVG-based features for use in performing specific inferences that have been manually selected can be tedious and time-consuming. Fortunately, deep learning is capable of extracting important features from raw input data and eliminates the need for manual feature extraction. Deep learning methods such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks have been developed to automatically and adaptively learn hierarchical features of sleep EEG/EMG recordings to classify sleep states in mice (Barger et al., 2019; Cai, 2021; Svetnik et al., 2020; Yamabe et al., 2019). While deep learning-based methods hold promise for wide-spread application in sleep research, a deep learning-based sleep state classification method for use with WFCI data alone has not yet been implemented.

In this work, we propose a hybrid MVG-CNN method for automated sleep state classification from WFCI data that avoids the use of adjunct EEG/EMG data. The spatial-temporal WFCI data are first mapped to MVGs, where each layer of MVG corresponds to a single brain region. Subsequently, a two-dimensional (2D) CNN is employed with the MVGs to classify the sleep state as wakefulness, NREM or REM. To investigate the temporal characteristics of MVGs that are important for sleep state classification, gradient-

weighted class activation maps (Grad-CAM) (Selvaraju et al., 2020) were computed. Taking advantage of the spatial-temporal nature of WFCI, the effect of the epoch duration and the amount of spatial information on sleep state classification performance was evaluated. We find that the proposed hybrid MVG-CNN method accurately identifies sleep states by use of spatial-temporal information afforded by WFCI data.

2. Materials and methods

2.1. Mice

This study was approved by the Washington University School of Medicine Institutional Animal Care and Use Committee and performed in accordance with National Institutes of Health *Guide for the Care and Use of Laboratory Animals*. Transgenic mice (12–16 weeks of age) expressing GCaMP6f in excitatory neurons (driven by a Thy1 promotor) were acquired from Jackson Laboratories (JAX strain: C57BL/6J-Tg (Thy1-GCaMP6f)GP5.5Dkim; stock: 024276) and used for experimental studies (total n = 17, all males; n = 12 in Section 2.1.1, n = 5 in Section 2.1.2). Mice were housed in 12-hour light/dark cycles with lights on at 6:00 AM and given ad lib access to food and water.

2.1.1. Experiment 1: Wakefulness, NREM, and REM imaging—The first experiment recorded WFCI data with simultaneous EEG/EMG of head-fixed mice during wakefulness, NREM and REM. Similar to other experimental paradigms (Bojarskaite et al., 2020; Niethard et al., 2018, 2021; Turner et al., 2020; Yüzgeç et al., 2018), mice were acclimated to head fixation while secured in a black felt hammock for one to three sessions ranging from 30 to 180 min until the EEG/EMG signals showed the presence of sleep. Once sleep was established in the head-fixed position, the mouse then underwent a three-hour undisturbed WFCI session. All recordings occurred between 9:00 AM and 1:00 PM during the mice's normal sleeping hours in order to maximize the chance of recording sleep. After the recording, human experts scored WFCI recordings as wakefulness, NREM, or REM by use of adjunct EEG/EMG.

2.1.2. Experiment 2: Wakefulness, sleep and anesthesia imaging—The second experiment used simultaneous WFCI and EEG/EMG from a previously published study (Brier et al., 2019). Briefly, mice were placed in the black felt hammock with their heads fixed in place. The mice were left undisturbed for 30 min of simultaneous WFCI and EEG/EMG recordings followed by intraperitoneal injection of ketamine/xylazine anesthetic (86.9 mg/kg ketamine and 13.4 mg/kg xylazine) and recording for another 60 min. After the recordings, the mice were placed in their home cages and monitored until they resumed normal behavior (grooming, exploration, resting, eating). Offline, experts manually scored sleep states of WFCI data based off the EEG/EMG signal as either wakefulness, sleep, or anesthesia. With the same mice, but in a separate session a week apart, mice were sleep-deprived for three hours using a novel environment (Tobler and Borbély, 1990) and then placed in a black felt hammock with their heads fixed under the imaging system. Then the mice were left undisturbed and simultaneous WFCI and EEG/EMG data were recorded for 60 min. Offline, experts scored sleep-wake states of WFCI data based off the EEG/EMG signal as either wakefulness or sleep.

2.2. Surgical techniques

Following anesthesia induction with isoflurane (3% induction, 1.5% maintenance), mice were head-fixed in a stereotactic frame. The head of each mouse was shaven, and a midline incision was made to expose the skull. EEG and EMG electrodes were implanted (described below) and a Plexiglass head cap was fixed with a translucent adhesive cement (C&B-Metabond, Parkell Inc., Edgewood, New York) to allow for repeated imaging (Wright et al., 2017). Mice were allowed to recover for seven days in light-controlled conditions (12-hour light-dark schedule).

EEG and EMG electrode implantation was performed with different surgical techniques for the two experiments. For *Experiment 1*, copper EEG pins (Newark Electronics, catalog # 89H8939) were placed at the surface (0.7 mm cranial burr holes) of the brain overlying the lateral somatosensory cortex (−0.7 mm posterior to bregma, and +4.5 mm lateral to bregma) and fixed with Fusio dental cement. An EEG pin placed on the surface of the cerebellum served as a bipolar reference. To record muscle activity, two 23-gauge stainless steel needles were attached to the posterior aspect of the Plexiglass headcap and inserted bilaterally into the neck muscles. For *Experiment 2*, stainless steel EEG self-tapping screws (BASI Inc., West Lafayette, IN, USA) were fixed at the surface (1.0 mm cranial burr holes) of the brain at approximately −1 mm posterior to bregma, and +/− 4.5 mm lateral to bregma (near barrel/auditory cortex) and referenced to a cerebellum screw. To record muscle activity, a 203 micrometer Teflon coated EMG wire (A-M Systems, Sequim, Washington, catalog #792100) was threaded into the neck muscle and referenced to the cerebellum.

2.3. Wide field optical imaging acquisition and processing

GCaMP6f-expressing mice were placed in a black felt hammock with their heads secured in place with a small bracket (using the plexiglass' pre-tapped holes). During the imaging session, the mouse was able to move freely while its head was secured, preventing the awake or sleeping mouse from applying torque on their restrained head and optical window. The secured mouse was then placed approximately 8 cm (working distance 14 cm) under an overhead EMCCD camera (iXon 897, Andor Technologies, Belfast, Northern Ireland, United Kingdom) and four collimated LEDs, as previously described (Wright et al., 2017; Brier et al., 2019). Sequential illumination was provided by four LEDs: 454 nm (blue, GCaMP6 excitation), 523 nm (green), 595 nm (yellow), and 640 nm (red) for hyperspectral oximetric imaging. The LEDs were sequentially triggered at 16.8 Hz per LED. The CCD framerate (67.2 Hz) and exposure times were synchronized through MATLAB via a DAQ device (PCI-6733, National Instruments, Austin, TX, USA). To discard GCaMP6 excitation light and capture emission, a 515 nm longpass filter was used. The field of view was ~1 cm² and covered the dorsal surface of the brain (78 μm x 78 μm pixel size).

Image processing occurred offline using a custom MATLAB package (Brier and Culver, 2021) as follows. First, a binary mask was manually drawn around all brain tissue and affine transformed to Paxinos space using the positions of bregma and lambda (Paxinos and Franklin, 2019). Then, the signal was temporally and spatially detrended, smoothed, and global signal regressed. The modified Beer-Lambert law used reflectance changes in the 523 nm, 595 nm, and 640 nm LED channels to solve for relative fluctuations of oxygenated-

hemoglobin (HbO₂) and deoxygenated-hemoglobin (HbR). The recorded GCaMP6 emission was corrected for absorptions by HbO₂ and HbR using a ratiometric approach and the 523 nm reflectance channel (approximate GCaMP6 wavelength) as a reference:

$$y(t) = \frac{I^{em}(t)}{I^{ref}(t)} \cdot \frac{I_0^{ref}}{I_0^{em}}.$$

$y(t)$ is the final corrected GCaMP6 time series for a given pixel, I^{em} refers to the detected fluorescent emission intensity. I^{ref} describes the measured reflectance changes at the emission wavelength. The power spectrum of each pixel in the GCaMP6 signal (%dF/F) was computed using the MATLAB toolbox Chronux function 'mtspecgram' (Mitra and Bokil, 2007; Chronux Home, 2021) using a window size of 16.81, step-size of 10 s, and time-width product of 3 and 5 tapers.

2.4. Expert behavioral state scoring

Time-locked EEG and EMG signals were recorded at 1000 Hz using the AD Instruments (Dunedin, New Zealand) Dual Bio Amplifier (Catalog# FE232) and PowerLab data acquisition system (Catalog# PL2604). Offline, the EEG and EMG signals were band-pass filtered (0.5–35 Hz for EEG and 25–50 Hz for EMG) and the spectrogram (Hann window with cosine-bell and 50% overlap) was computed. Using the combination of the filtered EEG/EMG signal and spectrogram, behavioral states (wakefulness, NREM, REM, anesthesia and movement) were manually scored in 10-second (10-s) epochs by the author E.L., a certified sleep specialist with over 15 years of experience scoring sleep. Wakefulness was characterized by mixed frequencies in the EEG with increased EMG tone. NREM sleep was defined as having large amplitude 1–3 Hz (delta) activity in the EEG with relative attenuation of the EMG. REM sleep was defined as having 6–8 Hz predominance with EMG atonia. Supplemental Figure 1 shows examples of 10-s EEG/EMG signals scored as the different sleep states by a human annotator. Anesthesia was defined as the presence of uniform 1 Hz activity with absence of EMG activity. Movement artifact was defined by movement causing the inability to discern the EEG waveforms. If a single 10-s epoch contained a mixture of both states, then the predominant state was scored. The details of the data split and data source for various classification tasks conducted in this study are shown in Table 1.

2.5. Construction of multiplex visibility graphs

2.5.1. Multiplex visibility graphs—Visibility graphs were first proposed by Lacasa et al. as a method to map time series into networks, in which the underlying dynamics are inherited in the topology (Lacasa et al., 2008). Given a time series $p = f(t)$, two time points (t_i, p_i) , (t_j, p_j) are connected if any other time points (t_k, p_k) between them satisfies the natural visibility criterion:

$$p_k < p_i + (p_j - p_i) \frac{k - i}{j - i}.$$

The time series can be mapped into an undirect and unweighted natural visibility graph (NVG). The single-layer NVG was implemented by the open-source MATLAB software package Fast NVG (Iacobello et al., 2018; Iacobello et al., 2019). The NVG is represented by a two-dimensional (2D) binary adjacency matrix $D \in R^{n \times n}$, where $D_{ij} = 1$ if a connection exists between time points (t_i, p_i) and (t_j, p_j) according to the visibility criteria. Furthermore, a multivariate time series containing m components yields a multiplex visibility graph (MVG) that comprises m -layers (Lacasa et al., 2015), which can be represented by a concatenation of 2D adjacency matrices.

2.5.2. Mapping wide-field calcium imaging data to multiplex visibility graphs

—To map WFCI data to an MVG, the WFCI data were first split into 10-second epochs according to the manual EEG/EMG-based scoring. Let the tensor $X \in R^{n \times m \times k}$ denote a 10-s epoch of WFCI data, where $n = m = 128$ denotes the number of pixels in each spatial dimension and $k = 168$ is the number of frames per 10-s epoch. Next, using a total of 36 brain regions defined by the Paxinos atlas (Paxinos and Franklin, 2019) (Fig. 1a), the average time series for all pixels within each region was calculated (Fig. 1c). The average time series for the 36 brain regions will be represented as $T \in R^{168 \times 36}$. Select regions (olfactory, prelimbic, colliculi, and cerebellum) were excluded because they were outside the field of view. The average time series corresponding to a single brain region was mapped to an adjacency matrix D by the use of natural visibility criteria and then the concatenation of the 36 adjacency matrices led to the formation of the MVG $M \in R^{168 \times 168 \times 36}$ for a single 10-s epoch of WFCI data (Fig. 1d).

2.6. Classification with 2D convolutional neural network

After the WFCI data were mapped to MVG representations in 10-s increments, a 2D multi-channel CNN was employed to classify sleep states via supervised deep learning (Fig. 1e, Fig. 2). A compact 2D CNN consisting of three convolutional layers was employed, where each layer had a number of 32, 64 and 128 kernels, respectively, with a kernel size 3. Kernels were shifted with a stride of 1 in all three layers. Leaky Rectified Linear Units (Leaky ReLU) were used after each convolutional layer as nonlinearities. A max pooling operation with a pool size of 2 and a stride size of 2 was applied after the first two convolutional layers. The last convolutional layer was followed by a global average pooling (GAP) layer to minimize the risk of overfitting by reducing the number of parameters in the model. One densely connected layer with either softmax function for multi-class classification or sigmoid function for binary classification was applied to yield the classified states. The network was implemented in Python 3 with TensorFlow 2.2.0 using NVIDIA GPUs.

Each layer of the MVG was considered as a channel of input to the 2D CNN. The kernels of the network share weight across all 36 channels corresponding to various brain regions. MVGs of different states were randomly shuffled during network training. The network was trained to minimize the focal loss (Appendix A.) by use of Adam optimizer (Kingma and Ba, 2017) with a learning rate of 0.0001 for 100 epochs and the CNN model with the best validation accuracy was selected.

2.7. Statistical analysis

The models were evaluated on test data consisting of unseen epochs from the same group of training subjects as well as on an independent subject. Metrics including $precision = \frac{TP}{TP + FP}$, $recall = \frac{TP}{TP + FN}$, $accuracy = \frac{TP + TN}{TP + TN + FP + FN}$, where TP , TN , FP , FN are the numbers representing true positive, true negative, false positive and false negative, respectively, were utilized to evaluate the model performance. The Cohen's kappa statistic, κ (Cohen, 1960), was computed to assess the inter-rater reliability between manual EEG/EMG-based scoring and the proposed automated WFCI-based MVG-CNN classification results. The kappa statistic is thought to be a more robust measure than percent agreement and a kappa magnitude between 0.61 and 0.80 indicates a substantial agreement between the two raters (McHugh, 2012). A confusion matrix for each classification task was formed to provide a direct interpretation of the classification results.

2.8. Data and source code availability

The WFCI data are available on PhysioNet (Goldberger et al., 2000; Landsness and Zhang, 2021) and the pre-processing code is available via https://github.com/brierl/Mouse_WOI (Brier and Culver, 2021). All model training and testing code are available at <https://github.com/comp-imaging-sci/MVG-CNN>.

3. Results

3.1. MVG-CNN classifies sleep states as wakefulness, NREM and REM

To automatically classify sleep states as wakefulness, NREM, and REM, WFCI data from mice ($n = 11$) were mapped to MVGs and a 2D CNN was trained on MVG representations (Fig. 1). The sleep state classification results of the MVG-CNN on WFCI alone were compared to human-scored EEG/EMG that were simultaneously collected with the WFCI data to assess the performance. For the individual sleep states in the test set consisting of 10% of the data in *Experiment 1* (Table 1), the precision (recall) was 0.87 (0.91) for wakefulness, 0.80 (0.72) for NREM and 0.76 (0.77) for REM (Table 2, Fig. 3). The MVG-CNN achieved an overall accuracy of 84% and Cohen's κ value of 0.67, where a κ value > 0.6 is indicative of substantial agreement (Table 2).

To further demonstrate the ability of the MVG-CNN to classify sleep states from WFCI data, sleep states of an unseen 3-hour WFCI recording were classified. The MVG-CNN achieved a κ of 0.65, indicating a substantial agreement between EEG/EMG-based scoring and the MVG-CNN classification. To further compare EEG/EMG-based scoring and MVG-CNN classification, we analyzed measures of sleep fragmentation, sleep-state organization and spectral power. The MVG-CNN method caused shorter sleep state durations and an increased number of state transitions (Table 3), suggesting increased sleep fragmentation. As depicted by the hypnogram (Fig. 4a, b), there was substantial agreement in the temporal pattern (sleep cycles) of transitions between wakefulness, NREM and REM. In addition, both EEG/EMG scored by a human and WFCI classified by the MVG-CNN showed an increase in delta (0.4–4.0 Hz) spectral power of the calcium signal exclusive to NREM and an increase in theta (6.0–8.0 Hz) exclusive to REM (Fig. 4c, d), confirming the

effective classification of sleep states by both methods. Further, this agreement between EEG/EMG-based human scoring and WFCI-based MVG-CNN classification is comparable to the inter-rater reliability of two human experts with a κ of 0.60 (Supplemental Figure 2). These results show that sleep states classified by the MVG-CNN using WFCI data alone are highly similar to EEG/EMG-based human scoring.

3.2. MVG-CNNs can also be effectively applied to different behavioral and experimental conditions

While the MVG-CNN was able to accurately classify sleep state when a single WFCI dataset was employed, applicability to other WFCI datasets with different animals, behavioral states or experimental conditions is desired for automated classification of WFCI data. Therefore, the MVG-CNN method was trained for two additional classification tasks that aimed to classify wakefulness versus sleep or wakefulness/sleep/anesthesia/movement by use of an additional dataset collected during *Experiment 2* (Table 1). In the binary classification, a pooled dataset consisting of epochs of wakefulness and sleep from *Experiment 1* and *2* (Table 1) was used. The MVG-CNN achieved a precision (recall) of 0.86 (0.85) for wakefulness and 0.76 (0.78) for sleep (Table 4, Fig. 5a). The accuracy was 0.82 with a Cohen's κ value of 0.62 (Table 4). In the four-class classification among wakefulness/sleep/anesthesia/movement, the MVG-CNN achieved an accuracy of 0.74 and a Cohen's κ value of 0.66 (Table 5), with precision (recall) of 0.63 (0.45) for wakefulness, 0.70 (0.86) for sleep, 0.95 (0.94) for anesthesia and 0.68 (0.75) for movement (Table 5, Fig. 5b). These results demonstrate that the MVG-CNN can be effectively applied to datasets with different experimental conditions and classification problems.

3.3. MVG-CNN reveals temporal characteristics for sleep state classification

3.3.1. MVG-CNN uses short- and long-range visibility connections to classify sleep—Understanding how the MVG-CNN makes a decision to classify WFCI data into various sleep-wake states could lead to a better understanding of the neural correlates of sleep. Therefore, a way to visualize how the CNN identifies the class-discriminative features of the MVG is needed. One method, Grad-CAM (Appendix B.) (Selvaraju et al., 2020), uses class-specific gradient information flowing into the final convolutional layer of a CNN to produce a coarse localization map of regions of emphasis. When applied to the adjacency matrices of MVG, Grad-CAM identified regions of interest contributing to a classification decision and showed different patterns of emphasis for various sleep states (Fig. 6). For instance, in the wakefulness state, the adjacency matrices have a nearly continuous focusing band along the main diagonal which indicates that the network focuses on the visible connections on a short time scale over the entire epoch. Similarly, NREM focuses on short visible connections, but in a more clustered pattern, which characterizes a local convexity property over a small range of time in a given epoch (Donner and Donges, 2012). In contrast, REM sleep shows a different characterization with the network emphasizing the off-diagonal elements that correspond to visible connections at a longer time scale. Taken together, these results show the MVG-CNN model classifies wakefulness and NREM sleep based on short-range visible connections, whereas REM is classified on longer time scale visibilities.

3.3.2. Varying epoch duration impacts sleep state classification performance

—Human experts conventionally score sleep EEG/EMG signals from mice with an arbitrary 10-second epoch duration. Often, human-defined sleep epochs contain a mixture of sleep states with the predominant state being classified. This mixture of states raises the question of whether shorter epoch durations would lead to better sleep state classification performance of WFCI data (Yan et al., 2011). Here, the epoch duration was varied from 1 to 20 s to investigate the effect of temporal information incorporated from WFCI epoch data on sleep state classification performance (Fig. 7). As the epoch duration was increased from 1 s to 7 s, the classification accuracy and Cohen's κ improved. At an epoch duration of 8 s and higher, accuracy plateaued at ~ 0.85 with Cohen's κ of ~ 0.70 . These results suggest that shortening epochs below 8 s or increasing beyond 10 s may not benefit sleep state classification performance by the MVG-CNN for the WFCI dataset being classified.

3.4. MVG-CNN identifies spatial characteristics for sleep state classification

The spatial-temporal nature of WFCI offers the possibility to alter the amount the spatial information in order to understand how individual brain regions affect the sleep state classification performance of the MVG-CNN. Therefore, visibility graphs built from each single brain region were given as input to CNNs, and the spatial distribution of sleep state classification accuracies and Cohen's kappa values were mapped to the Paxinos atlas of the left hemisphere of a mouse brain (Fig. 8). When classifying sleep by use of WFCI data from a single brain region alone, posterior regions including visual and retrosplenial cortex show the highest accuracy as compared to other regions such as somatosensory and motor cortex. Furthermore, MVGs consisting of a different number of layers from various brain regions were taken as input to the network to classify sleep states. As more brain regions from across the cortex were incorporated into MVGs, sleep classification performance improved (Table 6). Taken together, these results suggest that using larger amounts of spatial WFCI data improves sleep state classification accuracy.

4. Discussion

In this study, we proposed a hybrid method that combines a multiplex visibility graph and a 2D convolutional neural network (MVG-CNN) to accurately classify sleep states from WFCI data alone. Unique class activation patterns focusing on short- and long-range visibility were identified by the CNN when classifying WFCI data as wakefulness, NREM and REM. Additionally, regional information and epoch duration influenced sleep classification performance and accuracy. These results support that the spatial-temporal nature of the neuronal activity captured by WFCI plays an important role for characterizing sleep.

Accurate, automated sleep state classification methods are needed for sleep research. EEG/EMG-based methods have been successfully developed to automatically classify sleep in rodents (Barger et al., 2019; Yamabe et al., 2019). While classification of sleep based on other biosignals such as photoplethysmogram (PPG) (Korkalainen et al., 2020; Wu et al., 2020), heart rate and movement (Gaiduk et al., 2018; Sridhar et al., 2020) have been proposed, their performance is generally not as good as that based on EEG/EMG. Here,

we successfully applied an MVG-CNN model to a new imaging tool, WFCI, to classify sleep states in mice. Although the MVG-CNN did not perform as well as automated EEG/EMG-based methods (Barger et al., 2019; Yamabe et al., 2019), MVG-CNN classification of WFCI data compares favorably with the published gold standard of inter-rater reliability among human expert scorers of EEG/EMG (Rosenberg et al., 2013), and that of two human experts scoring the simultaneously acquired EEG/EMG of this dataset. Additionally, the MVG-CNN method is superior to other automated sleep state classification methods with non-EEG/EMG biosignals (Korkalainen et al., 2020; Wu et al., 2020; Gaiduk et al., 2018; Sridhar et al., 2020). Thus, the hybrid MVG-CNN method is an effective, accurate tool for automatically classifying sleep states in WFCI.

One unique aspect of sleep is the sequential and temporal dependency of sleep stage transitions (wakefulness followed by NREM and then REM sleep). To account for this temporal dependency, human scorers sequentially interrogate EEG/EMG recordings integrating the evidence with each epoch into the broader context. In contrast, the MVG-CNN model classifies sleep states in a temporally independent manner, which likely leads to an overestimation of the number of state transitions and shorter sleep state lengths. In the future, either recurrent neural networks that use future and past states, such as a bidirectional LSTM (Yamabe et al., 2019), or a post-processing algorithm, such as a hidden Markov model, that uses contextual information to estimate the probability to change to a different sleep-wake state from one epoch to the next (Brodersen et al., 2021) could be employed to reduce sleep fragmentation and increase accuracy. In addition, different variations of visibility graphs, such as the weighted visibility graph to increase sensitivity to state changes (Cai et al., 2018) and limited penetrable visibility graph to reduce the influence of noise (Wang et al., 2018), could be employed to further improve accuracy.

Sleep has traditionally been defined based on the presence or absence of discrete features such as slow waves, spindles, K-complexes, rapid eye movements and muscle tone. One recent automated sleep state classification study using EEG found that the convolutional filters in a CNN use hierarchical feature formation to extract features that closely resemble the same discrete neuronal events used by human experts (Li and Guan, 2021). To better understand how the CNN assigned specific sleep states to an MVG, Grad-CAMs were computed to visualize the corresponding class activation patterns. Unique activation patterns on MVGs for classifying different sleep states were identified. For example, the CNN characterized NREM by focusing on short time-frame visibility corresponding to clustered patterns with high intensity on Grad-CAMs, indicating a local convexity on the spatial-temporal multivariate time series of WFCI data that could be consistent with slow waves observed in NREM and the relatively higher spectral power of the calcium delta oscillations seen in Fig. 4. In contrast, during REM, off-diagonal (long-scale) visibility was identified as a key discriminative feature which may be consistent with the characteristic higher frequency (theta), low amplitude activity observed in the spectral analysis of the calcium signal. These MVG findings affirm the presence of slow waves as the key defining feature of NREM, whereas REM was characterized by relatively higher frequency, uniform activity. Future studies will allow for the identification of learned MVG patterns that could lead to improvement in defining the neural correlates of sleep and their relation to disease.

A major question in the sleep field is whether sleep is a cortically global phenomenon or whether it can occur locally in confined regions of the cortex (Krueger et al., 2019). Traditionally, NREM is classified by the presence of slow waves and spindles throughout the cerebral cortex and REM as a homogenous “activated” low-voltage activity. However, recent research suggests that sleep slow waves can occur in a subset of brain regions rather than occurring synchronously across all cortical areas (Vyazovskiy et al., 2011; Nir et al., 2011; Siclari and Tononi, 2017), and can be found outside of NREM during wakefulness (Vyazovskiy et al., 2011; Hung et al., 2013; Bernardi et al., 2015; Quercia et al., 2018; Andrillon et al., 2021) or REM (Funk et al., 2016; Bernardi et al., 2019). In support of the idea of sleep as a global phenomenon (Sejnowski and Destexhe, 2000), we found that sleep state classification performance improved as wider areas of the cortex were incorporated into the sleep state classification model. However, taking advantage of the high spatial resolution afforded by WFCI, we found that individual areas of the cortex varied in the degree of accuracy to classify sleep states. For instance, posterior regions demonstrated relatively high accuracy versus anterior and lateral brain areas (motor and somatosensory) having lower accuracy. This spatial heterogeneity in sleep state classification accuracy may be the result of these subregions having isolated episodes of local slow waves while the rest of the cortex is in a state of wakefulness or REM sleep and is consistent with existing studies showing that local slow waves occur predominantly in both frontal and parietal regions (Vyazovskiy et al., 2011; Hung et al., 2013; Quercia et al., 2018). Taken together, the spatial information provided by WFCI confirms that sleep is not a unitary, homogeneous state but is spatially diverse across the cortex. Future studies utilizing WFCI to better characterize the presence of local slow wave sleep are needed not only for sleep state classification, but also in normal physiology and disease (Terzaghi et al., 2009; Dodet et al., 2015; Castelnovo et al., 2016; Riedner et al., 2016).

In sleep state classification of rodent EEG/EMG signals by humans, a conventional 10-second epoch duration is commonly used to inspect the signal and assign sleep states. Studies have suggested that the optimal choice of epoch duration should accurately illustrate animals’ sleep-wake profiles with epochs as short as 4 s being ideal for capturing state transitions (Yan et al., 2011) but 8–10 s being best for sleep classification algorithms (Branka k et al., 2010). Consistent with previous studies, by altering epoch duration of the WFCI data prior to input into the MVG-CNN, we found that accuracy significantly declined at 4 s, and performance was optimal with epoch durations of 8–10 s. The decrease in accuracy with shorter epoch durations might be explained by the fact that EEG/EMG-based scoring to train the network is fixed at 10 s while the MVG-CNN epoch duration can be varied from one second to 20 s. The difference in epoch duration between EEG/EMG scoring and WFCI classification could be resolved by manually re-scoring the entire EEG/EMG dataset at a shorter temporal resolution (e.g., four seconds) and re-training the MVG-CNN, but comes at the cost of tedious, labor-intensive work and may result in the inability to integrate longer timescale trends into the MVG-CNN model. Indeed, for the WFCI dataset considered in our study, the MVG-CNN emphasized long-range temporal connections when classifying REM sleep (Fig. 6), which may explain why at least 8 s of data is necessary for sleep state classification while shorter epoch duration could lead to worse performance. Moreover, recent findings using multiunit activity (MUA) recordings

suggest that the brain is able to “flicker” into different sleep states at the microsecond timescale (Schneider et al., 2021; Parks et al., 2021), and future studies combining the high spatial resolution of WFCI with high temporal resolution of MUA may elucidate how quickly sleep states change across the cortex and the optimal epoch duration for sleep state classification.

To the best of our knowledge, this study is the first to develop a deep learning-based automated sleep state classification method for WFCI. Beyond the inherent advantages of automated sleep state classification (decreased need for laborious human scoring, improved test, re-test reliability on repeated recordings), the proposed MVG-CNN method could be applied to a wide range of other uses to understand not just the neural correlates of sleep but other brain states as well. Some interesting topics remain to be investigated in future research. For instance, deep learning methods could use WFCI data to characterize functional brain networks in sleep (Li et al., 2021; Dong et al., 2020). Or, if larger (>1000) WFCI recordings become available, 3D CNNs (Tran et al., 2015) could be employed to characterize sleep states and gain a deeper insight of the spatiotemporal features in raw data space. Last but not least, it is also desirable to enhance the generalization of sleep state classification methods so that the method is more robust against experimental variability. For example, advanced standardization approaches such as mixture z-scoring that disentangles nuisances and class prevalence variability (Barger et al., 2019) can be applied on the data, and more generalized models trained on large numbers of cohorts could be options (Yamabe et al., 2019).

5. Conclusions

In this study, we describe an automated sleep state classification method using WFCI data alone by mapping spatial-temporal data to MVG representations and classifying sleep states with a CNN. The MVG-CNN achieved substantial agreement with manual EEG/EMG-based scoring and was effectively applied to different WFCI datasets and experimental conditions. The MVG-CNN model accurately distinguished wakefulness, NREM and REM by using short- and long-scale temporal features. Furthermore, temporal data was combined with spatial information provided by WFCI to identify a regionally diverse series of temporal events throughout the mouse cortex. This study supports the use of MVG-CNN to better understand the neural correlates of sleep with WFCI and holds promise for the application to other research fields.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the National Institute of Neurological Disorders and Stroke (R01NS099429 to J.P.C., R37NS110699 and R01NS094692 to J.M.L., K08NS109292-01A1 to E.C.L.), National Institute on Aging (F30AG061932 to L.M.B.), American Academy of Sleep Medicine Foundation (201-BS-19 to E.C.L.) and American Heart Association (20CDA35310607 to E.C.L.). We also thank Kathleen Schoch with the Washington University InPrint scientific communication network for help with editing of this manuscript.

Appendix A

Focal loss

An important issue to be considered when applying deep learning in classification problems is the class prevalence in the data (Johnson and Khoshgoftaar, 2019). In rodent sleep recordings, NREM and wakefulness states often account for over 90% of total recording time with relatively less REM sleep. In such circumstances, the major class overwhelms the cross-entropy loss in multi-class classification and dominates the gradient. Thus, a modified cross-entropy loss called focal loss (Lin et al., 2020) was used in training the neural network to classify sleep states. Focal loss introduces a tunable focusing parameter $\gamma \geq 0$ to allow hard-to-classify examples to be penalized more heavily relative to easy-to-classify examples (Lin et al., 2020):

$$L_{focal}(y, \hat{\mathbf{p}}) = -(1 - \hat{p}_y)^\gamma \log(\hat{p}_y)$$

where $y \in \{0, \dots, K-1\}$ is an integral class label in K classes, and $\hat{\mathbf{p}} = (\hat{p}_0, \hat{p}_1, \dots, \hat{p}_{K-1}) \in [0, 1]^K$ denotes a vector of the model estimated probability over K the classes. In our study, $\gamma = 2$ was considered.

Appendix B

Visualizing CNN with gradient-weighted class activation maps (Grad-CAM)

To identify the features of the MVGs that were important for accurate classification of different brain states, gradient-weighted class activation maps (Grad-CAMs) (Selvaraju et al., 2020) were computed. The Grad-CAMs were created by using the gradient information flowing into the last convolutional layer of the CNN to identify the parts of an input image that most impact the classification score. First, for a class c , the gradient of the corresponding score y_c with respect to the feature map activations A_k of a convolutional layer was computed as $\frac{\partial y_c}{\partial A_k}$. A global average pooling operation over the spatial dimension indexed by i and j is applied on these gradients flowing back to obtain the importance weights of a feature map k of class c :

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A_{ij}^k}$$

Finally, Grad-CAM performed a weighted combinations of forward activation maps with ReLU:

$$L_{Grad-CAM}^c = ReLU \sum_k \alpha_k^c A_k$$

and upsampled to the input resolution to produce a heatmap of class-discriminative localization of the regions.

References

- Allen WE, Kauvar IV, Chen MZ, et al. , 2017. Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron* 94 (4), 891–907. 10.1016/j.neuron.2017.04.017. [PubMed: 28521139]
- Andrillon T, Burns A, Mackay T, Windt J, Tsuchiya N, 2021. Predicting lapses of attention with sleep-like slow waves. *Nat. Commun* 12 (1), 3657. 10.1038/s41467-021-23890-7. [PubMed: 34188023]
- Balbi M, Vanni MP, Vega MJ, et al. , 2019. Longitudinal monitoring of mesoscopic cortical activity in a mouse model of microinfarcts reveals dissociations with behavioral and motor function. *J. Cereb. Blood Flow. Metab* 39 (8), 1486–1500. 10.1177/0271678X18763428. [PubMed: 29521138]
- Barger Z, Frye CG, Liu D, Dan Y, Bouchard KE, 2019. Robust, automated sleep scoring by a compact neural network with distributional shift correction. *PLoS One* 14 (12), e0224642. 10.1371/journal.pone.0224642. [PubMed: 31834897]
- Bernardi G, Siclari F, Yu X, 2015. Neural and Behavioral Correlates of Extended Training during Sleep Deprivation in Humans: Evidence for Local, Task-Specific Effects. *J. Neurosci* 35 (11), 4487–4500. 10.1523/JNEUROSCI.4567-14.2015. [PubMed: 25788668]
- Bernardi G, Betta M, Ricciardi E, Pietrini P, Tononi G, Siclari F, 2019. Regional Delta Waves In Human Rapid Eye Movement Sleep. *J. Neurosci* 39 (14), 2686–2697. 10.1523/JNEUROSCI.2298-18.2019. [PubMed: 30737310]
- Bliwise D, Bliwise NG, Kraemer HC, Dement W, 1984. Measurement error in visually scored electrophysiological data: respiration during sleep. *J. Neurosci. Methods* 12 (1), 49–56. 10.1016/0165-0270(84)90047-5. [PubMed: 6513591]
- Bojarskaite L, Bjørnstad DM, Pettersen KH, 2020. Astrocytic Ca²⁺ signaling is reduced during sleep and is involved in the regulation of slow wave sleep. *Nat. Commun* 11 (1), 3240. 10.1038/s41467-020-17062-2. [PubMed: 32632168]
- Branka k J, Kukushka VI, Vyssotski AL, Draguhn A, 2010. EEG gamma frequency and sleep–wake scoring in mice: Comparing two types of supervised classifiers. *Brain Res.* 1322, 59–71. 10.1016/j.brainres.2010.01.069. [PubMed: 20123089]
- Brier LM, Culver JP, 2021. Open Source Stat. Data Process. Toolbox Wide-Field Opt. Imaging Mice, 10.1101/2021.04.07.438885.
- Brier LM, Landsness EC, Snyder AZ, et al. , 2019. Separability of calcium slow waves and functional connectivity during wake, sleep, and anesthesia. *Neurophoton* 6 (03), 1. 10.1117/1.NPh.6.3.035002.
- Brodersen, et al. , 2021. Somnotate: A robust automated sleep stage classifier that exceeds human performance and identifies ambiguous states in mice. *bioRxiv*. 10.1101/2021.10.06.463356.
- Buzsaki G, 2006. *Rhythms of the Brain*. Oxford University Press.
- Cai Qing, et al. , 2021. A multiplex visibility graph motif-based convolutional neural network for characterizing sleep stages using EEG signals. *Brain Science Advances*. 10.26599/BSA.2020.9050016.
- Cai L, Deng B, Wei X, Wang R, Wang J, 2018. Analysis of Spontaneous EEG Activity in Alzheimer’s Disease Using Weighted Visibility Graph. 2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC) 3100–3103. 10.1109/EMBC.2018.8513010.
- Castelnovo A, Riedner BA, Smith RF, Tononi G, Boly M, Benca RM, 2016. Scalp and Source Power Topography in Sleepwalking and Sleep Terrors: A High-Density EEG Study. *Sleep* 39 (10), 1815–1825. 10.5665/sleep.6162. [PubMed: 27568805]
- Chronux Home. (<http://chronux.org>). Accessed April 22, 2021.
- Cohen J, 1960. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas* 20 (1), 37–46. 10.1177/001316446002000104.)
- Collop NA, 2002. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep. Med* 3 (1), 43–47. 10.1016/s1389-9457(01)00115-0. [PubMed: 14592253]

- Danker-Hopf H, Anderer P, Zeitlhofer J, 2009. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *J. Sleep. Res* 18 (1), 74–84. 10.1111/j.1365-2869.2008.00700.x. [PubMed: 19250176]
- De Gennaro L, Ferrara M, 2003. Sleep spindles: an overview. *Sleep. Med Rev* 7 (5), 423–440. 10.1053/smr.2002.0252. [PubMed: 14573378]
- Dodet P, Chavez M, Leu-Semenescu S, Golmard J-L, Arnulf I, 2015. Lucid Dreaming in Narcolepsy. *Sleep* 38 (3), 487–497. 10.5665/sleep.4516. [PubMed: 25348131]
- Dong Q, Qiang N, Lv J, 2020. A Novel fMRI Representation Learning Framework with GAN (Lecture Notes in Computer Science). In: Liu M, Yan P, Lian C, Cao X (Eds.), *Machine Learning in Medical Imaging*, Vol 12436. Springer International Publishing, Cham, pp. 21–29. 10.1007/978-3-030-59861-7_3 (Lecture Notes in Computer Science).
- Donner RV, Donges JF, 2012. Visibility graph analysis of geophysical time series: Potentials and possible pitfalls. *Acta Geophys* 60 (3), 589–623. 10.2478/s11600-012-0032-x.
- Drinnan MJ, Murray A, Griffiths CJ, Gibson GJ, 1998. Interobserver variability in recognizing arousal in respiratory sleep disorders. *Am. J. Respir. Crit. Care Med.* 158 (2), 358–362. 10.1164/ajrcm.158.2.9705035. [PubMed: 9700107]
- Funk CM, Honjoh S, Rodriguez AV, Cirelli C, Tononi G, 2016. Local Slow Waves in Superficial Layers of Primary Cortical Areas during REM Sleep. *Curr. Biol* 26 (3), 396–403. 10.1016/j.cub.2015.11.062. [PubMed: 26804554]
- Gaiduk M, Penzel T, Ortega JA, Seepold R, 2018. Automatic sleep stages classification using respiratory, heart rate and movement signals. *Physiol. Meas* 39 (12), 124008 10.1088/1361-6579/aaf5d4. [PubMed: 30524059]
- Goldberger AL, Amaral LAN, Glass L, 2000. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* 101 (23), e215–e220. 10.1161/01.CIR.101.23.e215. [PubMed: 10851218]
- Halász P, 2016. The K-complex as a special reactive sleep slow wave - a theoretical update. *Sleep. Med. Rev* 29, 34–40. 10.1016/j.smr.2015.09.004. [PubMed: 26606317]
- Hung C-S, Sarasso S, Ferrarelli F, 2013. Local Experience-Dependent Changes in the Wake EEG after Prolonged Wakefulness. *Sleep* 36 (1), 59–72. 10.5665/sleep.2302. [PubMed: 23288972]
- Iacobello G, Scarsoglio S, Ridolfi L, 2018. Visibility graph analysis of wall turbulence time-series. *Phys. Lett. A* 382 (1), 1–11. 10.1016/j.physleta.2017.10.027.
- Iacobello G, Marro M, Ridolfi L, Salizzoni P, Scarsoglio S, 2019. Experimental investigation of vertical turbulent transport of a passive scalar in a boundary layer: Statistics and visibility graph analysis. *Phys. Rev. Fluids* 4 (10), 104501. 10.1103/PhysRevFluids.4.104501.
- Johnson JM, Khoshgoftaar TM, 2019. Survey on deep learning with class imbalance. *J. Big Data* 6 (1), 27. 10.1186/s40537-019-0192-5.
- Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs]. January 2017. <http://arxiv.org/abs/1412.6980>. Accessed March 18, 2021.
- Korkalainen H, Aakko J, Duce B, 2020. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea. *Sleep* 43. 10.1093/sleep/zsaa098.
- Kozberg MG, Ma Y, Shaik MA, Kim SH, Hillman EMC, 2016. Rapid Postnatal Expansion of Neural Networks Occurs in an Environment of Altered Neurovascular and Neurometabolic Coupling. *J. Neurosci* 36 (25), 6704–6717. 10.1523/JNEUROSCI.2363-15.2016. [PubMed: 27335402]
- Krueger JM, Nguyen JT, Dykstra-Aiello CJ, Taishi P, 2019. Local sleep. *Sleep. Med. Rev* 43, 14–21. 10.1016/j.smr.2018.10.001. [PubMed: 30502497]
- Lacasa L, Luque B, Ballesteros F, Luque J, Nuño JC, 2008. From time series to complex networks: the visibility graph. *PNAS* 105 (13), 4972–4975. 10.1073/pnas.0709247105.
- Lacasa L, Nicosia V, Latora V, 2015. Network structure of multivariate time series. *Sci. Rep* 5 (1), 15508. 10.1038/srep15508. [PubMed: 26487040]
- Li H, Guan Y, 2021. DeepSleep convolutional neural network allows accurate and fast detection of sleep arousal. *Commun. Biol* 4 (1), 1–11. 10.1038/s42003-020-01542-8. [PubMed: 33398033]
- Landsness Zhang, 2021. Wide-field calcium imaging sleep state database. *PhysioNet*. 10.13026/jzqa-j937.

- Li Q, Dong Q, Ge F, Qiang N, Wu X, Liu T, 2021. Simultaneous spatial-temporal decomposition for connectome-scale brain networks by deep sparse recurrent autoencoder. *Brain Imaging Behav.* 10.1007/s11682-021-00469-w.
- Lin T, Goyal P, Girshick R, He K, Dollár P, 2020. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell* 42 (2), 318–327. 10.1109/TPAMI.2018.2858826. [PubMed: 30040631]
- Lord S, Sawyer B, Pond D, 1989. Interrater reliability of computer-assisted scoring of breathing during sleep. *Sleep* 12 (6), 550–558. 10.1093/sleep/12.6.550. [PubMed: 2595177]
- Loredo JS, Clausen JL, Ancoli-Israel S, Dimsdale JE, 1999. Night-to-night arousal variability and interscorer reliability of arousal measurements. *Sleep* 22 (7), 916–920. 10.1093/sleep/22.7.916. [PubMed: 10566909]
- Ma Y, Shaik MA, Kim SH, 2016a. Wide-field optical mapping of neural activity and brain haemodynamics: considerations and novel approaches. *Philos. Trans. R. Soc. B Biol. Sci* 371 (1705), 20150360. 10.1098/rstb.2015.0360.
- Ma Y, Shaik MA, Kozberg MG, 2016b. Resting-state hemodynamics are spatiotemporally coupled to synchronized and symmetric neural activity in excitatory neurons. *Proc. Natl. Acad. Sci. USA* 113 (52), E8463–E8471. 10.1073/pnas.1525369113. [PubMed: 27974609]
- Mang GM, Nicod J, Emmenegger Y, Donohue KD, O’Hara BF, Franken P, 2014. Evaluation of a Piezoelectric System as an Alternative to Electroencephalogram/Electromyogram Recordings in Mouse Sleep Studies. *Sleep* 37 (8), 1383–1392. 10.5665/sleep.3936. [PubMed: 25083019]
- Matsui T, Murakami T, Ohki K, 2016. Transient neuronal coactivations embedded in globally propagating waves underlie resting-state functional connectivity. *PNAS* 113 (23), 6556–6561.
- McHugh ML, 2012. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 22 (3), 276–282. [PubMed: 23092060]
- Niethard N, Ngo H-VV, Ehrlich I, Born J, 2018. Cortical circuit activity underlying sleep slow oscillations and spindles. *PNAS* 115 (39), E9220–E9229. 10.1073/pnas.1805517115.
- Mitra P, Bokil, 2007M. *Observed Brain Dynamics*. Oxford University Press.
- Niethard N, Brodt S, Born J, 2021. Cell-Type-Specific Dynamics of Calcium Activity in Cortical Circuits over the Course of Slow-Wave Sleep and Rapid Eye Movement Sleep. *J. Neurosci* 41 (19), 4212–4222. 10.1523/JNEUROSCI.1957-20.2021. [PubMed: 33833082]
- Nir Y, Staba RJ, Andrillon T, et al. , 2011. Regional Slow Waves and Spindles in Human Sleep. *Neuron* 70 (1), 153–169. 10.1016/j.neuron.2011.02.043. [PubMed: 21482364]
- Norman RG, Pal I, Stewart C, Walsleben JA, Rapoport DM, 2000. Interobserver agreement among sleep scorers from different centers in a large dataset. *Sleep* 23 (7), 901–908. [PubMed: 11083599]
- Parks D, Schneider A, Haussler D, Hengen K, 2021. Probing sleep at the millisecond timescale. *Poster Presente Sleep*. 2021. June 10.
- Paxinos G, Franklin K.B.J. Paxinos, 2019. *and Franklin’s the Mouse Brain in Stereotaxic Coordinates*. Academic Press.
- Quercia A, Zappasodi F, Committed G, Ferrara M, 2018. Local Use-Dependent Sleep in Wakefulness Links Performance Errors to Learning. *Front Hum. Neurosci* 12. 10.3389/fnhum.2018.00122.
- Riedner BA, Goldstein MR, Plante DT, et al. , 2016. Regional Patterns of Elevated Alpha and High-Frequency Electroencephalographic Activity during Nonrapid Eye Movement Sleep in Chronic Insomnia: A Pilot Study. *Sleep* 39 (4), 801–812. 10.5665/sleep.5632. [PubMed: 26943465]
- Rosenberg, et al. , 2013. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. *Journal of clinical sleep medicine*. 10.5664/jcsm.2350.
- Sannino S, Stramaglia S, Lacasa L, Marinazzo D, 2017. Visibility graphs for fMRI data: multiplex temporal graphs and their modulations across resting-state networks. *Netw. Neurosci* 1 (3), 208–221. 10.1162/NETN_a_00012. [PubMed: 29911672]
- Schneider A, Parks D, Hengen K, 2021. Arousal States are Local and Unstable: Sleep and Wake States Fluctuate Transiently and in a Circuit-Specific Fashion in the Intact Brain of Freely Behaving Mice. *Poster Presente Sleep*. 2021. June 10.
- Sejnowski TJ, Destexhe A, 2000. Why do we sleep? *Brain Res.* 886 (1–2), 208–223. 10.1016/S0006-8993(00)03007-9. [PubMed: 11119697]

- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, 2020. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int J. Comput. Vis* 128 (2), 336–359. 10.1007/s11263-019-01228-7.
- Siclari F, Tononi G, 2017. Local aspects of sleep and wakefulness. *Curr. Opin. Neurobiol* 44, 222–227. 10.1016/j.conb.2017.05.008. [PubMed: 28575720]
- Silber MH, Ancoli-Israel S, Bonnet MH, 2007. The visual scoring of sleep in adults. *J. Clin. Sleep. Med* 3 (2), 121–131. [PubMed: 17557422]
- Sridhar N, Shoeb A, Stephens P, et al. , 2020. Deep learning for automated sleep staging using instantaneous heart rate. *npj Digital Med.* 3 (1), 1–10. 10.1038/s41746-020-0291-x.
- Svetnik V, Wang T-C, Xu Y, Hansen BJ, Fox S V, 2020. A Deep Learning Approach for Automated Sleep-Wake Scoring in Pre-Clinical Animal Models. *J. Neurosci. Methods* 337, 108668. 10.1016/j.jneumeth.2020.108668. [PubMed: 32135210]
- Terzaghi M, Sartori I, Tassi L, et al. , 2009. Evidence of Dissociated Arousal States During NREM Parasomnia from an Intracerebral Neurophysiological Study. *Sleep* 32 (3), 409–412. [PubMed: 19294961]
- Tobler I, Borbély AA, 1990. The effect of 3-h and 6-h sleep deprivation on sleep and EEG spectra of the rat. *Behav. Brain Res.* 36 (1), 73–78. 10.1016/0166-4328(90)90161-7. [PubMed: 2302323]
- Tran D, Bourdev L, Fergus R, Torresani L, Paluri M Learning Spatiotemporal Features with 3D Convolutional Networks. In: 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE; 2015:4489–4497. (doi:10.1109/ICCV.2015.510).
- Turner K, et al. , 2020. Neurovascular coupling and bilateral connectivity during NREM and REM sleep. *Elife.* 10.7554/eLife.62071.
- Vyazovskiy VV, Olcese U, Hanlon EC, Nir Y, Cirelli C, Tononi G, 2011. Local sleep in awake rats. *Nature* 472 (7344), 443–447. 10.1038/nature10009. [PubMed: 21525926]
- Wang R, Yang Z, Wang J, Shi L, 2018. An Improved Visibility Graph Analysis of EEG Signals of Alzheimer Brain. In: 2018 11th International Congress on Image and Signal Processing. Biomed. Eng. Inform. (CISP-BMEI) 1–5. 10.1109/CISP-BMEI.2018.8633052.
- Whitney CW, Gottlieb DJ, Redline S, 1998. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep* 21 (7), 749–757. 10.1093/sleep/21.7.749. [PubMed: 11286351]
- Wright PW, Brier LM, Bauer AQ, 2017. Functional connectivity structure of cortical calcium dynamics in anesthetized and awake mice. *PLoS One*, e0185759. 10.1371/journal.pone.0185759. [PubMed: 29049297]
- Wu X, Yang J, Pan Y, Zhang X, Luo Y, 2020. Automatic sleep-stage scoring based on photoplethysmographic signals. *Physiol. Meas* 41 (6), 065008 10.1088/1361-6579/ab921d. [PubMed: 32392540]
- Yamabe M, Horie K, Shiokawa H, Funato H, Yanagisawa M, Kitagawa H, 2019. MC-SleepNet: Large-scale Sleep Stage Scoring in Mice by Deep Neural Networks. *Sci. Rep* 9 (1), 15793. 10.1038/s41598-019-51269-8. [PubMed: 31672998]
- Yan M-M, Xu X-H, Huang Z-L, Yao M-H, Urade Y, Qu W-M, 2011. Selection of optimal epoch duration in assessment of rodent sleep-wake profiles: Epoch duration selection in EEG analysis. *Sleep. Biol. Rhythms* 9 (1), 46–55. 10.1111/j.1479-8425.2010.00484.x.
- Yüzgeç Ö, Prsa M, Zimmermann R, Huber D, 2018. Pupil Size Coupling to Cortical States Protects the Stability of Deep Sleep via Parasympathetic Modulation. *e3 Curr. Biol.* 28 (3), 392–400. 10.1016/j.cub.2017.12.049.
- Zhu L, Haghani S, Najafizadeh L, 2018a. Spatiotemporal Characterization of Brain Function Via Multiplex Visibility Graph. *Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS)*. Hollywood, Florida: OSA. 10.1364/TRANSLATIONAL.2018.JTh3A.54.
- Zhu L, Lee CR, Margolis DJ, Najafizadeh L, 2018b. Decoding cortical brain states from widefield calcium imaging data using visibility graph. *Biomed. Opt. Express* 9 (7), 3017. 10.1364/BOE.9.003017. [PubMed: 29984080]

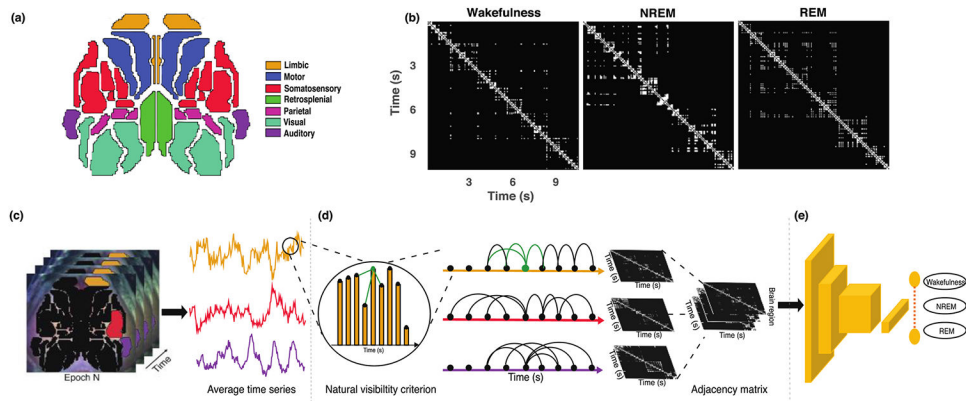


Fig. 1. The proposed hybrid framework using MVG and 2D CNN to automatically classify sleep states with WFCI data in mice. (a) The 36 brain regions defined by the Paxinos atlas (Paxinos and Franklin, 2019) within the field of view used in constructing the MVG. (b) Examples of single-layer visibility graphs from visual cortex during wakefulness, NREM and REM, represented in a binary adjacency matrix (white=visible, black=non-visible). (c)-(e) Schematic using 3 sample brain regions showing the construction of the MVG for a 10-s epoch of WFCI data. (c) First, an average time series is created for each brain region. (d) Then, a visibility graph associated with each average time series is constructed and represented in its adjacency matrices. The MVG is constructed by stacking the adjacency matrices of each brain region. (e) The adjacency matrices for MVGs are taken as input to a 2D CNN to classify sleep states.

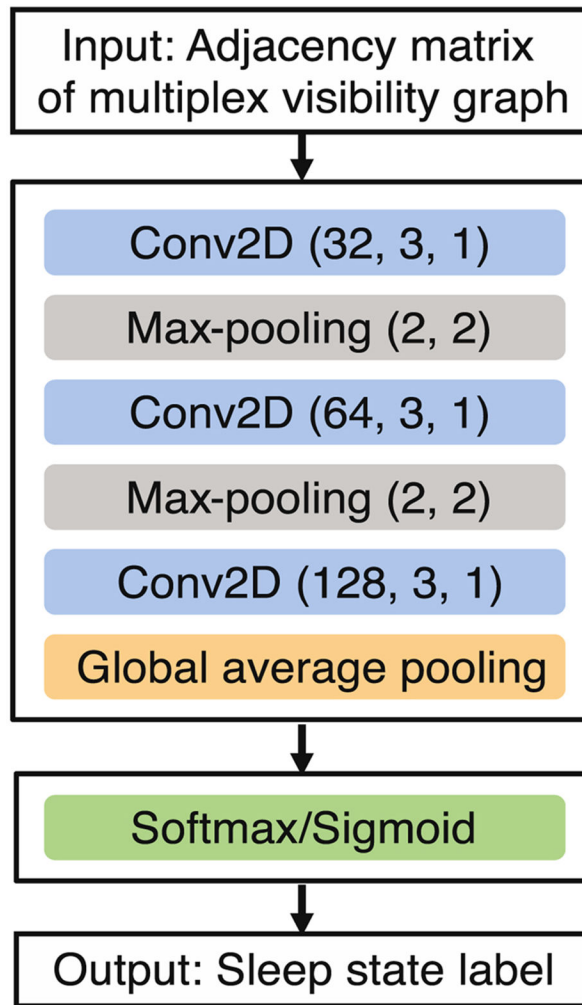


Fig. 2. Illustration of the architecture of the 2D CNN. Three convolutional layers (blue) with parameters (number of filters, kernel size and stride size) are employed with Leaky ReLU. Max-pooling layers (gray) with parameters (pool size, stride size) were included after the first two convolutional layers. A global average pooling layer (orange) is placed after the final convolutional layers. Following the convolutional blocks, a softmax/sigmoid function (green) is applied to classify sleep states.

Predictions	Wakefulness	630 (86.8%)	56 (19.6%)	9 (14.3%)
	NREM	84 (11.6%)	228 (79.7%)	6 (9.5%)
	REM	12 (1.7%)	2 (0.7%)	48 (76.2%)
		Wakefulness	NREM	REM
		True Labels		

Fig. 3. Confusion matrix for the MVG-CNN on three-state sleep classification of wakefulness, NREM, and REM in test set (n = 1075 epochs). Manual EEG/EMG-based scoring is on the x-axis and MVG-CNN predictions are on y-axis. The diagonal cells in blue correspond to the numbers of correctly classified epochs and precision rate (%) across wakefulness, NREM and REM states. Non-diagonal cells indicate misclassified epochs for each state.

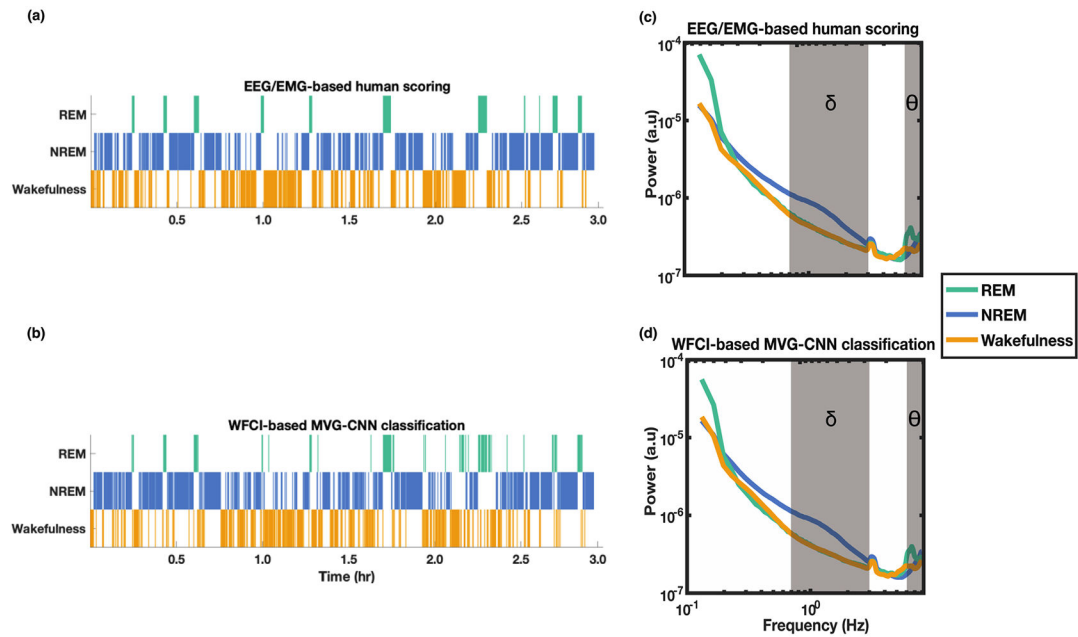


Fig. 4.

Comparison of sleep state classification between human annotator and MVG-CNN on a 3 h recording of a mouse. Hypnograms corresponding to (a) human EEG/EMG-based scoring and (b) MVG-CNN classification based on WFCI recording. Average power spectra of the calcium signal plotted for wakefulness, NREM and REM based on the (c) true scoring produced by a human annotator or (d) predictions from MVG-CNN. Shaded gray areas represent the delta (δ , 0.4–4.0 Hz) and theta (θ , 6.0–8.0 Hz) frequency ranges.

Predictions	Wakefulness	715 (86.4%)	130 (24.9%)		
	Sleep	113 (11.6%)	393 (75.1%)		
		Wakefulness	Sleep		
		True Labels			

Predictions	Wakefulness	69 (63.3%)	36 (21.6%)	5 (3.3%)	44 (27.3%)
	Sleep	14 (12.8%)	117 (70.1%)	1 (0.7%)	4 (2.5%)
	K/X	4.6 (5.0%)	0 (0.0%)	143 (94.7%)	4 (2.5%)
	Movement	21 (19.2%)	14 (8.4%)	2 (1.3%)	109 (67.7%)
		Wakefulness	Sleep	K/X	Movement
		True Labels			

Fig. 5. Confusion matrix that summarizes the MVG-CNN classification performance on (a) binary classification for distinguishing wakefulness and sleep, and (b) four-state classification for distinguishing among wakefulness, sleep, anesthesia (ketamine/xylazine, K/X) and movement. Manual EEG/EMG-based scoring on x-axis and MVG-CNN predictions on y-axis. The diagonal cells in blue correspond to the numbers of correctly classified epochs across different states, with the precision percentages in the parentheses. Non-diagonal cells indicate misclassified epochs for each state.

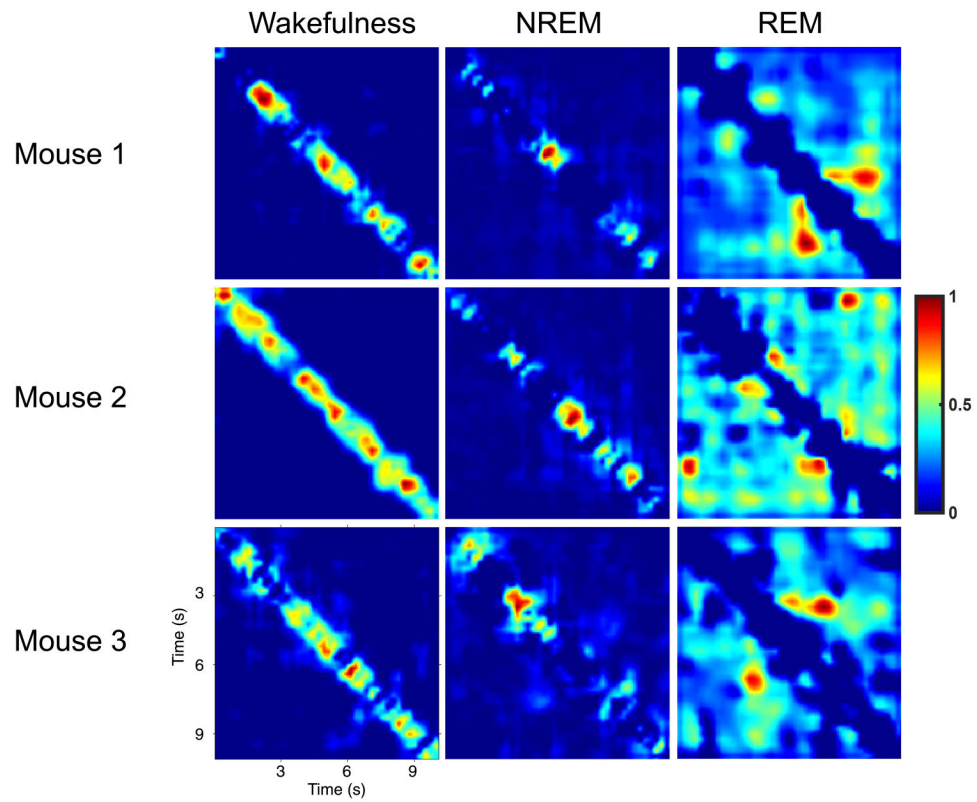


Fig. 6. Representative Grad-CAM examples of wakefulness, NREM and REM from three mice. A higher intensity with the color gradients (i.e., red, value 1) reveals that the 2D CNN focuses more on such regions of interest of adjacency matrices when making corresponding decisions.

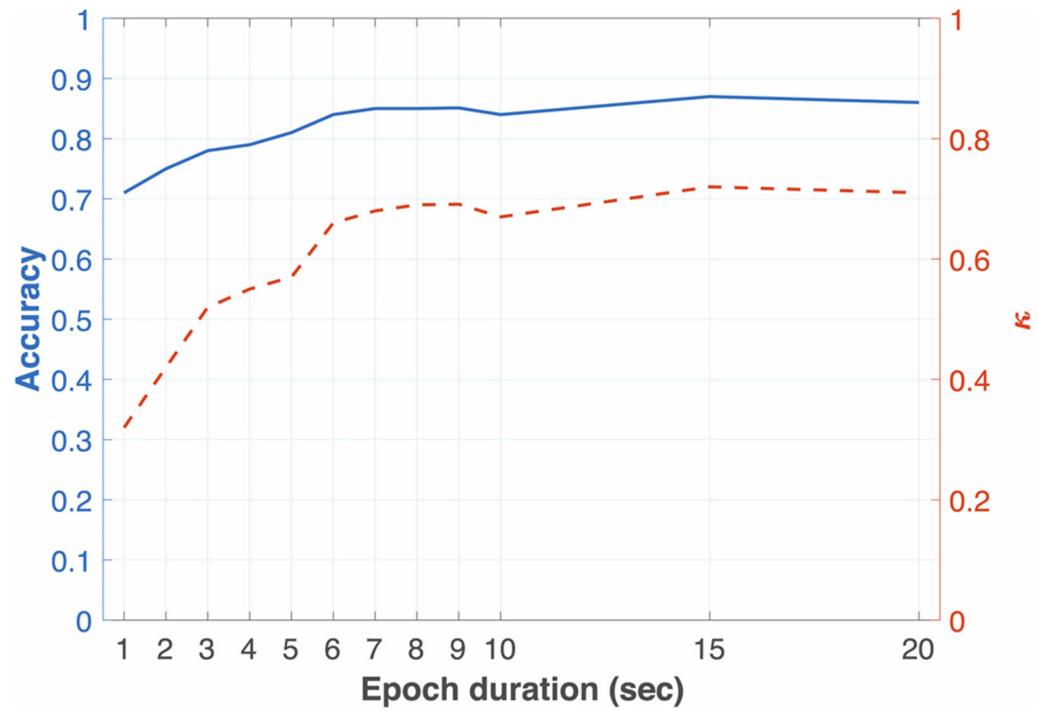


Fig. 7. The sleep state classification performance with respect to the epoch durations used in the MVG-CNN. As epoch duration was varied, accuracy (left y-axis, solid blue line) and the Cohen's Kappa statistic (right y-axis, dotted red line) were compared.

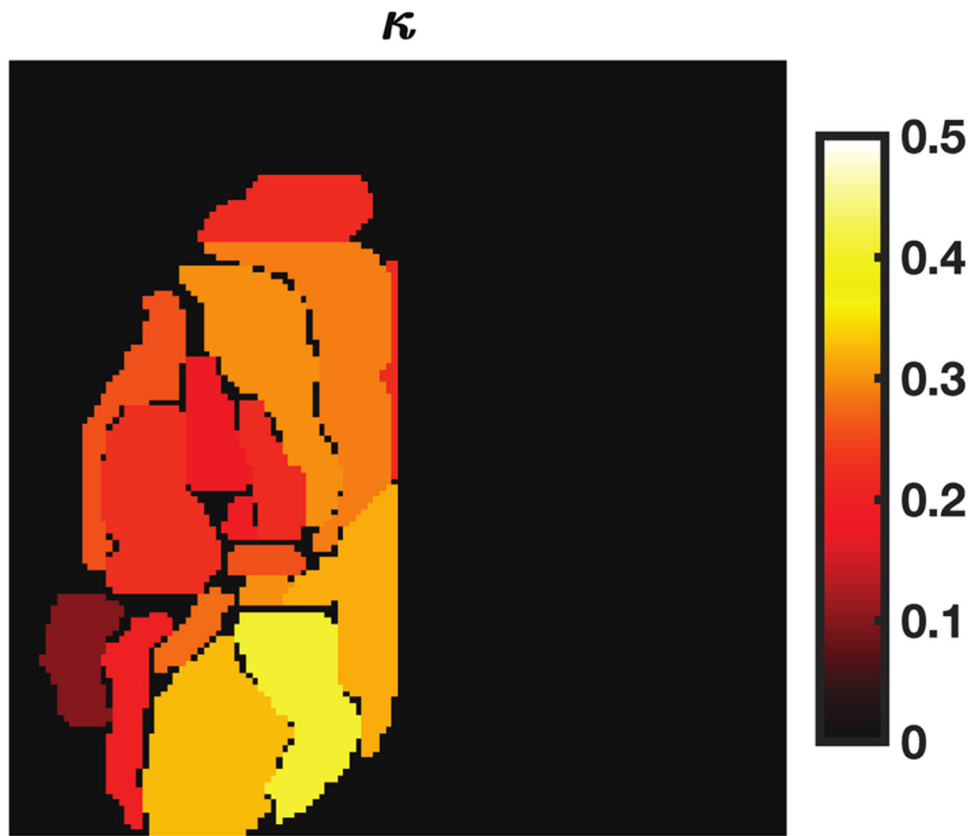


Fig. 8. The sleep state classification performance with respect to using single-layer visibility graphs from individual brain regions in the left hemisphere. The Cohen's κ value of each brain region is mapped to the atlas (Fig. 1a) to reveal the spatial importance in classifying WFCI data as wakefulness, NREM and REM.

Table 1

Data source of different classification tasks and the corresponding number of epochs for training, validation and testing.

Classification task	Data source	Training	Validation	Testing
Wakefulness/NREM/REM	Experiment 1	8592	1074	1075
Wakefulness/sleep/anesthesia/movement	Experiment 2	4704	588	588
Wakefulness/sleep	Experiment 1 + 2	10800	1350	1351

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Metrics to evaluate the three-state classification performance on test data (n = 1075 epochs). MVG-CNN achieved substantial agreement of $\kappa = 0.67$ compared to manual EEG/EMG-based scoring. Prec., precision; Rec., recall; κ , Cohen's Kappa statistic.

<u>Wakefulness</u>		<u>NREM</u>		<u>REM</u>		Accuracy	κ
Prec.	Rec.	Prec.	Rec.	Prec.	Rec.		
0.87	0.91	0.80	0.72	0.76	0.77	0.84	0.67

Table 3

Comparison of sleep fragmentation between WFCI-based MVG-CNN classification and EEG/EMG-based human scoring.

Scoring method	Average sleep state length (s)			Number of state transitions
	Wakefulness	NREM	REM	
WFCI-based MVG-CNN classification	28	54	30	272
EEG/EMG-based human scoring	47	66	81	180

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Metrics to evaluate the binary wakefulness/sleep classification performance with MVG-CNN on test data (n = 1351 epochs). Prec., precision; Rec., recall; κ , Cohen's Kappa statistic.

<u>Wakefulness</u>		<u>Sleep</u>		Accuracy	κ
Prec.	Rec.	Prec.	Rec.		
0.86	0.85	0.76	0.78	0.82	0.62

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Metrics to evaluate the four-state classification of wakefulness, sleep, anesthesia, and movement with MYG-CNN on test data (n = 588 epochs). Prec., precision; Rec., recall; κ , Cohen's Kappa statistic.

<u>Wakefulness</u>		<u>Sleep</u>		<u>Anesthesia</u>		<u>Movement</u>		<u>Accuracy</u>	<u>κ</u>
Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.		
0.63	0.45	0.70	0.86	0.95	0.94	0.68	0.75	0.74	0.66

Table 6

The sleep state classification performance using different amounts of spatial information in MVG-CNN. L: left hemisphere, V1: primary visual cortex, Barrel: somatosensory barrel cortex, M1: primary motor cortex.

Brain regions (n)	Accuracy	κ
L-V1 (1)	0.72	0.33
L-Barrel (1)	0.69	0.23
L-M1 (1)	0.71	0.30
L-M1 +L-Barrel+L-V1 (3)	0.78	0.49
Left hemisphere (18)	0.83	0.64
Whole brain (36)	0.84	0.67

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript