HHS Public Access

Author manuscript

J Neurosci Methods. Author manuscript; available in PMC 2025 May 26.

Published in final edited form as:

J Neurosci Methods. 2024 November ; 411: 110250. doi:10.1016/j.jneumeth.2024.110250.

Attention-based CNN-BiLSTM for sleep state classification of spatiotemporal wide-field calcium imaging data

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

[a]Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

[b]Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA

[c]Department of Radiology, Washington University School of Medicine, St. Louis, MO 63110, USA

[d]Department of Biomedical Engineering, Washington University School of Engineering, St. Louis, MO 63130, USA

[e]Department of Electrical and Systems Engineering, Washington University School of Engineering, St. Louis, MO 63130, USA

[f]Department of Physics, Washington University School of Arts and Science, St. Louis, Mo 63130, USA

[g]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Abstract

Background: Wide-field calcium imaging (WFCI) with genetically encoded calcium indicators allows for spatiotemporal recordings of neuronal activity 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, invasive and often suffers from low inter- and intra-rater reliability. Therefore, an automated sleep state classification method that operates on spatiotemporal WFCI data is desired.

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

[1]These authors contributed equally.

CRediT authorship contribution statement

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

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 A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jneumeth.2024.110250.

New method: A hybrid network architecture consisting of a convolutional neural network (CNN) to extract spatial features of image frames and a bidirectional long short-term memory network (BiLSTM) with attention mechanism to identify temporal dependencies among different time points was proposed to classify WFCI data into states of wakefulness, NREM and REM sleep.

Results: Sleep states were classified with an accuracy of 84 % and Cohen's κ of 0.64. Gradient-weighted class activation maps revealed that the frontal region of the cortex carries more importance when classifying WFCI data into NREM sleep while posterior area contributes most to the identification of wakefulness. The attention scores indicated that the proposed network focuses on short- and long-range temporal dependency in a state-specific manner.

Comparison with existing method: On a held out, repeated 3-hour WFCI recording, the CNN-BiLSTM achieved a κ of 0.67, comparable to a κ of 0.65 corresponding to the human EEG/EMG-based scoring.

Conclusions: The CNN-BiLSTM effectively classifies sleep states from spatiotemporal WFCI data and will enable broader application of WFCI in sleep research.

Keywords

Wide-field calcium imaging; Automated sleep state classification; CNN-BiLSTM; Deep learning; Local sleep

1. Introduction

Wide-field calcium imaging (WFCI) with genetically encoded calcium indicators is a powerful imaging technique that allows for simultaneous recordings of cortex-wide neuronal activity in mice with high signal-to-noise ratio (SNR) and spatiotemporal resolution (Ren and Komiyama, 2021; Nietz et al., 2022; Ma et al., 2016; Kozberg et al., 2016; Matsui et al., 2016; Ma et al., 2016). Given these capabilities WFCI has been extensively employed to study the functional organization and the development of mouse brain during various states of consciousness (Brier et al., 2019, 2021; Wright et al., 2017; Zhang et al., 2022; Ma et al., 2016), behaviors (Allen et al., 2017; West et al., 2022), and disease models (Balbi et al., 2019; Cramer et al., 2019). More recently, WFCI has also been applied to characterize calcium dynamics of neural activity during sleep and proven as a useful tool to uncover the neural correlates of sleep (Turner et al., 2020; Niethard et al., 2018, 2021). To study the mouse brain physiology during sleep, it is necessary to score the spatiotemporal WFCI recordings by assigning individual epoch into various sleep-wake states such as wakefulness, non-rapid eye movement (NREM) and REM sleep. This procedure is traditionally conducted by manual inspection of adjunct electroencephalogram (EEG) and electromyogram (EMG) signals. However, this process is time-intensive, often suffers from 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; Loredo et al., 1999; Norman et al., 2000; Rosenberg Richard and Steven, n.d; Silber et al., 2007; Whitney et al., 1998) and requires invasive electrophysiology. To overcome these limitations, an automated sleep state classification method that operates on spatiotemporal WFCI data alone is desired.

We recently reported a hybrid, two-step method where multiplex visibility graphs and a convolutional neural network (CNN) were combined for sleep state classification from WFCI data (Zhang et al., 2022). However, this approach requires an additional mapping of raw WFCI data to graph representations and thus fails to fully explore the spatiotemporal nature of the WFCI data. In addition, the interpretability of spatial and temporal features exploited by the network that are relevant to sleep state classification is limited. Thus, a sleep state classification model that directly utilizes the spatiotemporal calcium dynamics recorded by WFCI holds potential to further improve the performance and interpretability.

Sleep is a cyclical process and thus, temporal dependencies could be important when inferring sleep state. One widely adopted approach in deep learning that is capable of learning order dependence of sequential data is the recurrent neural network (RNN), such as the long short-term memory (LSTM) network. Several recent applications have employed the LSTM to infer sleep states or categorize heart disease on sequential physiological signals such as EEG (Zhuang et al., 2022; Michielli et al., 2019; Mousavi et al., 2019) and electrocardiogram (ECG) (Cheng et al., 2021; Saadatnejad et al., 2020). When assigning a sleep state to an individual epoch, it is typical for sleep experts to take into consideration adjacent time points and further locate the occurrence of discriminative features such as K-complexes, spindles and slow waves, which can guide them to make decisions (Buzsaki, 2006; De Gennaro and Ferrara, 2003; Halász, 2016). Therefore, a bidirectional LSTM (BiLSTM) that can leverage temporal information from both past and future events could potentially benefit the sleep state classification performance. In addition, incorporating an attention mechanism into the model enables selective emphasis on different parts of sequential input with varying degrees of importance. This could potentially better leverage temporal signatures in the data that aligns more effectively with the identification of sleep states (Bahdanau et al., 2016; Mousavi et al., 2019; Kwon et al., 2021).

In this work, we propose a hybrid, two-stage deep neural network architecture consisting of a convolutional neural network (CNN) and a BiLSTM with attention mechanism to jointly learn spatial and temporal information from WFCI data for automated sleep state classification is proposed and investigated. The CNN was employed to extract spatial information from WFCI image frames and a BiLSTM module was applied to capture temporal dependencies between features extracted from each time points. To identify brain regions that contribute strongly to the classification of different sleep states, gradient-weighted class activation maps (Grad-CAM) were computed. In addition, the temporal characteristic of WFCI data for sleep state classification were investigated. We found that the proposed CNN-BiLSTM effectively identifies sleep states by use of the spatial and temporal neural dynamics recorded by WFCI.

2. Methods

2.1. Mice and surgical preparations

A total of 16 transgenic mice (12–16 weeks of ages, all males) expressing GCaMP6f (JAX strain: C57BL/6 J-Tg (Thy1-GCaMP6f) GP5.5Dkim; stock: 024276) in excitatory neurons were acquired from Jackson Laboratories and used in the experiments in this study. All studies were approved by the Washington University School of Medicine Institutional

Animal Care and Use Committee and followed the guidelines of the National Institutes of Health *Guide for the Care and Use of Laboratory Animals*. 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. Prior to data collection, the head of each mouse was shaven, and a midline incision was made to expose the skill. A Pixelglass head cap was fixed with a translucent adhesive cement (C&B-Metabond, Parkell Inc., Edgewood, New York) to allow chronic, repeated imaging (Silasi et al., 2016). To implant EEG and EMG electrodes, 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. A referenced EEG screw was placed on the surface of the cerebellum. 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.

2.2. Wide-field calcium imaging of sleep

Mice expressing GCaMP6f were placed in a black felt hammock with their head secured for one to three adaptation sessions ranging from 30 to 180 minutes until the EEG/EMG signal showed the presence of sleep. Once sleep was established in the adaptation sessions, the mouse then underwent a 3-hour undisturbed WFCI session. All recordings occurred between 9:00 AM and 1:00 PM during the mice's normal sleeping hours to maximize the likelihood of recording sleep. A total of 16 sleep recordings were collected from 15 mice, with a single mouse undergoing a repeat recording 3 weeks after the initial recording.

A cooled, frame-transfer EMCCD camera (iXon 897, Andor Technologies, Belfast, Northern Ireland, United Kingdom) overhead and four collimated LEDs were used for image acquisition, as previously described (Wright et al., 2017; Brier et al., 2019; 2021; Zhang et al., 2022). Sequential illumination was provided by the 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 channel. The field of view covered the dorsal surface of the brain resulting in an area of ~1 cm^2 with pixel resolution of approximately 78 μm x 78 μm.

2.3. Expert behavioral state scoring

Using the combination of the filtered EEG/EMG signals and spectrograms, sleep states of wakefulness, NREM and REM were manually assigned based on 10-second epochs by a certified sleep specialist with over 15 years of experience scoring sleep. In total, a total of 19,155 10-second epochs were collected that included 12,674, 5701 and 780 data epochs for wakefulness, NREM and REM, respectively. This dataset was randomly shuffled and split into 80 %, 10 % and 10 % for training, validation and testing of the sleep state classification model.

2.4. Data processing

Prior to analysis, the WFCI raw data underwent the following image processing steps. First, to control for session-to-session changes in ambient light levels, the mean light levels of non-illuminated frames for the session were subtracted from each image frame. To account

for the effect of photobleaching, all pixel time traces were detrended. The global signal was regressed from every pixel's time trace to remove global sources of variance, such as motion and cardiac pulsations. To correct for any fluorescent absorption-emission by hemoglobin and deoxyhemoglobin, a ratiometric correction was applied using the reflectance channels at the GCaMP6 emission wavelengths (523 nm LED) as a reference (Ma et al., 2016). Images were then spatially smoothed with a 5×5 pixel box Gaussian filter. To register images between mice to a common atlas space, images for each mouse were affine transformed to the Paxinos mouse atlas using the positions of bregma and lambda as reference. A white-light image of the skull for each mouse and region of interest was manually demarcated as either "brain" or "non-brain" mask. All subsequent analysis was performed on pixels labeled as "brain".

2.5. Attention-based CNN-BiLSTM network architecture and implementation details

2.5.1. 2D Convolutional neural network to extract spatial features: Both combined CNN-RNN models and 3D CNNs are commonly used in the field of deep learning for processing sequential data or spatiotemporal data, such as videos or time series data. However, 3D-CNNs are only able to capture local features rather than long-term dependencies in sequential data when both spatial and temporal information are important for a specific task such as sleep classification. In additional, 3D-CNNs generally have more trainable parameters and the risk overfitting is increased with limited amounts of data is increased. In our previous work (Zhang et al., 2022), the temporal information is encoded in image representations by use of visibility graph and therefore there is no temporal dependencies among different channels for the given adjacency matrix representation. In this study, we leverage the raw spatiotemporal information recorded by WFCI to classify sleep and better understand the contribution of the spatial and temporal information relevant to various sleep states, thus a CNN-LSTM is a more suitable choice to achieve this goal.

A 2D-CNN that contained five convolutional blocks was employed to extract spatial information from each image frame. Each CNN block comprised a 2D convolutional layer (Conv2D) with 64 kernels of size of 3, followed by a max pooling (MaxPool) operation with a pool size of 2 and a stride of 2 and activated by the use of leaky rectified linear unit (LeakyReLU). 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. The CNN blocks were wrapped by time-distributed layers so that they could be applied to every image frame of the input. Fig. 1 contains an illustration of the employed architecture.

2.5.2. Bidirectional long short-term memory with attention mechanism to learn temporal dependencies: Long short-term memory (LSTM) is one kind of recurrent neural network that consists of memory blocks for capturing long-term dependencies of sequential data. Three types of gates are used in a LSTM unit: input gate, output gate and forget gate, which control the flow of information (Hochreiter and Schmidhuber, 1997). Different from a uni-directional LSTM that is only able to leverage the information from previous events, the bi-directional LSTM contains an additional LSTM layer that allows for the input sequence to flow backward and the outputs from the forward and backward LSTM layers are combined. The use of a bidirectional LSTM recurrent layer

instead of a unidirectional one is motivated by the fact that both adjacent timepoints are inspected when human experts score sleep based on recorded neural activities. Therefore, temporal information from past and future time points could be beneficial to improve the classification performance. As illustrated in Fig. 1, given a WFCI data epoch as input to the spatial branch consisting of CNN blocks, feature vectors corresponding to each image frame are extracted. Subsequently, they are sent into the BiLSTM layer to learn the temporal dependency from both previous and following events. In this study, a total of 64 LSTM units were in each direction.

Since sleep state scoring by human experts is typically based on the occurrence of temporal events such as K-complexes, spindles, theta rhythms and slow waves (Zhang et al., 2022; Buzsaki, 2006; De Gennaro and Ferrara, 2003; Halász, 2016), the sequential feature vectors from the LSTM layers could contribute differently for the classification of different sleep stages. To improve classification performance by allowing the LSTM network to focus on timesteps with more discriminative sleep-state related features, an attention module with Bahdanau's additive style (Bahdanau et al., 2016; Tezuka et al., 2021) was added to learn the importance of feature vectors of each time points. Given the hidden state vector $\boldsymbol{h}_i$ at every time point i, the importance score $\boldsymbol{s}_i$ can be calculated by the score function as:

$$\boldsymbol{s}_i = \tanh(\boldsymbol{W}_s \boldsymbol{h}_i + \boldsymbol{b}_s), \tag{1}$$

where $\boldsymbol{W}_s$ and $\boldsymbol{b}_s$ are trainable weights and bias, respectively.

The attention weight α_i was computed using the softmax function:

$$\alpha_i = softmax(s_i) = \frac{\exp(s_i)}{\sum_i s_i}. \tag{2}$$

The final output $\boldsymbol{v}$ was formed as

$$\boldsymbol{v} = \sum_i \alpha_i \boldsymbol{h}_i. \tag{3}$$

Finally, a fully connected layer with softmax activation was employed to form a decision variable for classifying the sleep state.

2.5.3. Implementation details: Each 10-s WFCI data epoch consisted of 168 image frames each of size of 128 ×128 and was given as input to the CNN-BiLSTM network. The network was trained by use of an Adam optimizer (β_1=0.9, β_2=0.999) (Kingma and Ba, 2017) with a learning rate of 0.0001 for 100 epochs to minimize the focal loss (Lin et al., 2020), which is used to address the class imbalance of the sleep states labels. To avoid overfitting, the model with the best validation accuracy was selected. The focusing parameter γ = 2 was considered. The network was implemented in Python 3 with

TensorFlow 2.2.0 using NVIDIA Quadro RTX 8000. An example of training curve is shown in Supplemental Figure 1.

2.6. Performance evaluation

The models were evaluated on test data consisting of unseen epochs from the same group of 15 training subjects as well as on a held-out subject that underwent repeat testing at a later date. Metrics including precision, accuracy and F1-score 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 CNN-BiLSTM classification results. 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 comprehensive description of the classification results.

2.7. Data and source code availability

A subset of the WFCI sleep data is available on PhysioNet (Landsness, Eric and Zhang, Xiaohui n.d.; Goldberger et al., 2000). All model training and testing code are available at https://github.com/comp-imaging-sci/attention-based-bilstm-sleep-scoring.

3. Results

3.1. Attention-based CNN-BiLSTM classifies sleep states as wakefulness, NREM and REM

To automatically classify sleep states as wakefulness, NREM, and REM, the CNN-BiLSTM was trained and tested on spatiotemporal WFCI data (Fig. 1). The sleep state classification results of the CNN-BiLSTM 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 pooled data of WFCI recordings (N=15), the precision (recall) was 0.89 (0.88) for wakefulness, 0.71 (0.74) for NREM and 0.83 (0.68) for REM (Table 1, Fig. 2). The CNN-BiLSTM achieved an F1-score of 0.84 and Cohen's κ value of 0.64, where a κ value > 0.60 is indicative of substantial agreement (Table 1). In addition, a previously reported method was implemented as a baseline. In that method, referred to as MVG-CNN, the WFCI data were mapped to multiplex visibility graphs (MVG) and a compact 2D CNN was trained on the MVG representations to classify sleep state (Zhang et al., 2022).

To further demonstrate the ability of the CNN-BiLSTM to classify sleep states from spatiotemporal WFCI data, epochs from a held-out, repeated 3-hour WFCI recording acquired on a different date were classified into the three states. On the repeated recording, the CNN-BiLSTM achieved a κ of 0.67, indicating a substantial agreement between EEG/EMG-based scoring and the CNN-BiLSTM classification, and comparative to our previous MVG-CNN approach. To further compare EEG/EMG-based scoring and CNN-BiLSTM classification, we analyzed measures of sleep fragmentation, sleep-state organization and spectral power. While the CNN-BiLSTM method resulted in shorter sleep state durations and an increased number of state transitions compared to the EEG/EMG-based method

(Table 2), it predicted fewer state transitions compared to the MVG-CNN. This suggests the potential of the LSTM to better capture temporal dependencies as compared to MVGs.

As depicted by the hypnogram (Fig. 3a), 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 CNN-BiLSTM 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. 3b), confirming the effective classification of sleep states by both methods. Further, this agreement between EEG/EMG-based human scoring and WFCI-based CNN-BiLSTM classification is comparable to the inter-rater reliability of two human experts with a κ of 0.65 (Supplemental Figure 2). These results confirm that sleep states classified by the CNN-BiLSTM using WFCI data alone are very similar to EEG/EMG-based human scoring.

3.2. Attention-based CNN-BiLSTM locates spatial characteristics for sleep states classification

There is emerging evidence of spatially unique patterns of neuronal activity under different states of sleep (Dong et al., 2022; Wang et al., 2022; Bréchet et al., 2020). Here, Grad-CAM (Selvaraju et al., 2020) was computed to employ class-specific gradient information flowing into the final convolutional layer of a CNN. Grad-CAM takes all kernels in the last convolutional layer into consideration when computed. This approach produced a coarse localization map of brain regions that contribute most to a classification decision and revealed different patterns of emphasis for various sleep states (Fig. 4a). For example, in NREM, the frontal brain regions such as motor and somatosensory cortex received more attention from the network while in REM the posterior regions including visual and retrosplenial cortex were emphasized more (Fig. 4b, Table 3). These observations also hold when only hemispheric data was given as training input (Supplemental Figure 3). Taken together, these results revealed that the CNN-BiLSTM classified NREM and REM sleep based on anterior and posterior areas, respectively.

3.3. Attention-based CNN-BiLSTM reveals temporal characteristics for sleep states classification

3.3.1. CNN-BiLSTM uses short- and long-range temporal information to classify sleep: Bi-directional LTSM leverages the temporal information of a signal to aid in classification. To gain insights into how the CNN-BiLSTM leverages the temporal information of WFCI to classify sleep-wake states, 5-s of data was added to the front and back each data epoch to make a 20-s epoch and the network trained again to classify the central 10-s of data within the epoch. We then investigated the importance of temporal information in various ranges within the WFCI data in different states. The weights from the attention layer were extracted for 20-s epochs and visualized for wakefulness, NREM and REM (Fig. 5). It was found that the network focuses on long-range temporal dependency when making decisions to classify the input as a REM state. In contrast, it was observed that NREM focuses on short-range temporal information and larger weights are potentially given to the slow oscillations in the recorded neural activity.

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. However, 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 (Wan et al., 2011). To investigate this, the epoch duration was varied from 1 to 20 s to examine the impact of temporal information incorporated from spatiotemporal WFCI data on sleep state classification performance (Fig. 6). As the epoch duration was increased from 1 s to 10 s, the classification accuracy and Cohen's κ improved. At an epoch duration of 10 s and higher, accuracy plateaued at ~0.84 with Cohen's of ~0.64. These results suggest that shortening epochs below 10 s or increasing beyond 10 s may not benefit sleep state classification performance by the CNN-BiLSTM for the considered WFCI dataset.

4. Discussion

In this study, we proposed a hybrid, two-stage deep neural network architecture consisting of a convolutional neural network and a bidirectional long short-term memory with attention mechanism to accurately classify sleep states from spatiotemporal WFCI data alone. Leveraging spatial information, we observed that frontal and posterior regions are important for classifying WFCI data as NREM or REM sleep, respectively. Using attentional weights, we showed a temporal dependency in network decision making with REM sleep relying on relatively long-scale dependencies. The improved interpretability of the proposed network allowed us to gain deeper insights into the spatiotemporal calcium dynamics that underlies sleep.

Automated accurate sleep state classification methods are desired for sleep research with WFCI. While recent research that addresses automated sleep scoring has attracted increasing attention, most of the proposed algorithms primarily work with biosignals such as photoplethysmograms (Korkalainen et al., 2020; Wu et al., 2020), heart rate and movement, as well as electrophysiological signals from EEG and EMG. Recent advances in WFCI enables researchers to record high-dimensional neural activities across various states such as spontaneous wakefulness, NREM, and REM sleep in mice (Niethard et al., 2016). In this context, we have successfully implemented a CNN-BiLSTM model to distinguish sleep states directly from spatiotemporal data acquired through WFCI. The proposed CNN-BiLSTM demonstrated performance on par with the established gold standard of inter-rater reliability among expert human EEG/EMG scorers (Rosenberg Richard and Steven, n.d). While it does not surpass the performance of our previous MVG-CNN method (Zhang et al., 2022), its ability to process raw spatiotemporal WFCI data enhances the interpretability of the temporal and spatial features of calcium dynamics during sleep. Thus, our hybrid CNN-BiLSTM method is an effective and accurate tool for automatic sleep scoring that also offers greater interpretability.

Sleep is a temporal process with a strong dependency between consecutive time points (Zhao et al., 2022). When human experts assess sleep patterns, they consider both the preceding and subsequent events to make decision regarding specific sleep states. In this

study, we sought to emulate this decision-making process by incorporating a bidirectional LSTM network and thus fully leverage temporal dependencies in both forward and backward directions of sleep dynamics captured by WFCI. In addition, when analyzing temporal neural activity, it is of interest to understand which time points possess enhanced significance, which facilitates the further investigation of the potential markers of a sleep event. Here, the attention weights extracted from our network reveal unique patterns for classifying different sleep states. For instance, the CNN-BiLSTM characterized REM sleep by focusing on long-scale dependency, aligning with observations from our prior MVG-CNN method (Zhang et al., 2022). In contrast, during NREM, the network exhibited a heightened emphasis on low frequency components, potentially corresponding to the slow waves observed in NREM (Niethard et al., 2021; Brier et al., 2019). Future studies investigating the signatures of EEG/EMG signals that align with high attention scores could enhance our understanding of the underlying mechanisms governing sleep events.

Sleep is not a unitary, homogeneous state but exhibits spatial diversity across the cortex. A major question within the field of sleep field is whether sleep can occur in confined regions of the cortex. Our findings indicate that frontal areas play a pivotal role in the identification of NREM sleep, aligning with the observation of strong inhibition of frontal pyramidal neurons (Li et al., 2023). Conversely, posterior regions seem to be more engaged during REM and wakefulness. This distinction in sleep states classification highlights the spatial heterogeneity of sleep, confirming that sleep is not a unitary, homogeneous state but is spatially diverse across the cortex(Zhang et al., 2022). Future WFCI studies that leverage techniques such as optogenetics manipulating sleep in these localized brain regions will allow for a better understanding of the underlying sleep mechanisms.

To our knowledge, this study is the first to develop a deep learning-based automated sleep state classification method that directly works on spatiotemporal WFCI data. The proposed method enhances the interpretability of the network’s decision-making process in assigning specific sleep state. However, there are possible iterative improvements for the classification of WFCI sleep-wake states. For example, the use of a transformer network that can model long-range dependencies and offer faster computational speed could further our understanding of the spatiotemporal characteristics in various sleep states (Wan et al., 2023; Du et al., 2022). Additionally, the application of diffusion models with high-quality image generation to augment imbalanced sleep data could potentially improve classification performance (Zhang et al., 2023). As with any deep learning algorithm, distribution shift caused by subject variability may degrade the model performance (Supplemental Figure 1). For any mice not included in the initial training that will be used for WFCI sleep studies and to be scored using the proposed method, it would be a good practice to annotate some epochs sampled from novel recordings and fine-tune the model so that it can be applied to future repeated recordings in longitudinal studies. To further improve the generalizability of this deep learning-based automated sleep classification approach, in future studies should incorporate advanced standardization methods such as mixture z-scoring (Barger et al., 2019) to correct distribution shift for WFCI data, or employ advanced training strategies such as contrastive learning (Xue et al., 2024; Garg et al., 2023).

5. Conclusion

In this study, we described an automated sleep state classification method that operates directly on spatiotemporal WFCI data. Utilizing a convolutional neural network, our approach learns the spatial features of image frames while a bidirectional LSTM with an attention mechanism captures the temporal dependencies across different time points. The proposed CNN-BiLSTM model achieved substantial agreement with manual EEG/EMG-based scoring and provides enhanced interpretability of the networks' decision-making in categorizing various sleep states. We revealed the importance of spatial information across brain regions and temporal characteristics of calcium dynamic. Our findings advocate for the use of CNN-BiLSTM in elucidating the neural correlates of sleep via spatiotemporal WFCI and demonstrated its potential for broader research applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

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.), American Heart Association (20CDA35310607 to E.C.L.) and National Institute of Health (P41EB031772 to M.A.A.)

Data Availability

I have shared the link to the code and part of the available data.

References

Allen William E., Isaac V.Kauvar, Chen Michael Z., Richman Ethan B., Yang Samuel J., Chan Ken, Gradinaru Viviana, Deverman Benjamin E., Luo Liqun, Deisseroth Karl, 2017. Global representations of goal-directed behavior in distinct cell types of mouse neocortex. e6 Neuron 94 (4), 891–907. 10.1016/j.neuron.2017.04.017.

Bahdanau Dzmitry, Cho Kyunghyun, and Bengio Yoshua. 2016. Neural Machine Translation by Jointly Learning to Align and Translate. arXiv. 〈10.48550/arXiv.1409.0473〉.

Balbi Matilde, Vanni Matthieu P., Vega Max J., Silasi Gergely, Sekino Yuki, Boyd Jamie D., LeDue Jeffrey M., Murphy Timothy H., 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 Zeke, Frye Charles G., Liu Danqian, Dan Yang, Bouchard Kristofer E., 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]

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]

Bréchet Lucie, Brunet Denis, Perogamvros Lampros, Tononi Giulio, Michel Christoph M., 2020. EEG microstates of dreams. Sci. Rep 10 (1), 17069 10.1038/s41598-020-74075-z. [PubMed: 33051536]

Brier Lindsey M., Landsness Eric C., Snyder Abraham Z., Wright Patrick W., Baxter Grant A., Bauer Adam Q., Lee Jin-Moo, Culver Joseph P., 2019. Separability of calcium slow waves and functional connectivity during wake, sleep, and anesthesia. Neurophotonics 6 (03), 1. 10.1117/1.NPh.6.3.035002.

Brier Lindsey M., Zhang Xiaohui, Bice Annie R., Gaines Seana H., Landsness Eric C., Lee Jin-Moo, Anastasio Mark A., Culver Joseph P., 2021. A multivariate functional connectivity approach to mapping brain networks and imputing neural activity in mice. Cereb. Cortex no. bhab282 (September). 10.1093/cercor/bhab282.

Buzsaki Gyorgy, 2006. Rhythms of the Brain. Oxford University Press.

Cheng Jinyong, Zou Qingxu, Zhao Yunxiang, 2021. ECG signal classification based on deep CNN and BiLSTM. BMC Med. Inform. Decis. Mak 21 (1), 365. 10.1186/s12911-021-01736-y. [PubMed: 34963455]

Collop Nancy A., 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]

Cramer Julia V., Gesierich Benno, Roth Stefan, Dichgans Martin, Düring Marco, Liesz Arthur, 2019. In vivo widefield calcium imaging of the mouse cortex for analysis of network connectivity in health and brain disease. NeuroImage 199 (October), 570–584. 10.1016/j.neuroimage.2019.06.014. [PubMed: 31181333]

Danker-Hopfe Heidi, Anderer Peter, Zeitlhofer Josef, Boeck Marion, Dorn Hans, Gruber Georg, Heller Esther, et al. , 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 Luigi, Ferrara Michele, 2003. Sleep spindles: an overview. Sleep. Med. Rev 7 (5), 423–440. 10.1053/smrv.2002.0252. [PubMed: 14573378]

Dong Yufan, Li Jiaqi, Zhou Min, Du Yihui, Liu Danqian, 2022. Cortical Regulation of Two-Stage Rapid Eye Movement Sleep. Nat. Neurosci 25 (12), 1675–1682. 10.1038/s41593-022-01195-2. [PubMed: 36396977]

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/ajrccm.158.2.9705035. [PubMed: 9700107]

Du Yang, Xu Yongling, Wang Xiaoan, Liu Li, Ma Pengcheng, 2022. EEG temporal–spatial transformer for person identification. Sci. Rep 12 (1), 14378 10.1038/s41598-022-18502-3. [PubMed: 35999245]

Garg Saurabh, Setlur Amrith, Lipton Zachary, Balakrishnan Sivaraman, Smith Virginia, Raghunathan Aditi, 2023. Complementary benefits of contrastive learning and self-training under distribution shift. Adv. Neural Inf. Process. Syst 36 (December), 11621–11673.

Halász P.éter, 2016. The K-complex as a special reactive sleep slow wave - a theoretical update. Sleep. Med. Rev 29 (October), 34–40. 10.1016/j.smrv.2015.09.004. [PubMed: 26606317]

Hochreiter Sepp, Schmidhuber J.ürgen, 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. 10.1162/neco.1997.9.8.1735. [PubMed: 9377276]

Kingma Diederik P., and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [Cs], January. 〈http://arxiv.org/abs/1412.6980〉 .

Korkalainen Henri, Aakko Juhani, Duce Brett, Kainulainen Samu, Leino Akseli, Nikkonen Sami, Afara Isaac O., Myllymaa Sami, Töyräs Juha, Leppänen Timo, 2020. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea. Sleep 43 (zsaa098). 10.1093/sleep/zsaa098. [PubMed: 32436942]

Kozberg Mariel G., Ma Ying, Shaik Mohammed A., Kim Sharon H., Hillman Elizabeth M.C., 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]

Kwon Hyun Bin, Choi Sang Ho, Lee Dongseok, Son Dongyeon, Yoon Heenam, Lee Mi. Hyun Lee, Yu Jin, Park Kwang Suk, 2021. Attention-Based LSTM for non-contact sleep stage classification using IR-UWB radar. IEEE J. Biomed. Health Inform 25 (10), 3844–3853. 10.1109/JBHI.2021.3072644. [PubMed: 33848253]

Li Bing, Ma Chenyan, Huang Yun-An, Ding Xinlu, Silverman Daniel, Chen Changwan, Darmohray Dana, et al. , 2023. Circuit mechanism for suppression of frontal cortical ignition during NREM sleep. Cell December. 10.1016/j.cell.2023.11.012.

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, O'Connell D, Eyland A, Mant A, Hensley MJ, Saunders NA, 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 Ying, Shaik Mohammed A., Kozberg Mariel G., Kim Sharon H., Portes Jacob P., Timerman Dmitriy, Hillman Elizabeth M.C., 2016. Resting-state hemodynamics are spatiotemporally coupled to synchronized and symmetric neural activity in excitatory neurons. Proc. Natl. Acad. Sci 113 (52), E8463–E8471. 10.1073/pnas.1525369113. [PubMed: 27974609]

Ma Ying, Shaik Mohammed A., Kim Sharon H., Kozberg Mariel G., Thibodeaux David N., Zhao Hanzhi T., Yu Hang, Hillman Elizabeth M.C., 2016. 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.

Matsui Teppei, Murakami Tomonari, Ohki Kenichi, 2016. Transient neuronal coactivations embedded in globally propagating waves underlie resting-state functional connectivity. Proc. Natl. Acad. Sci 113 (23), 6556–6561. [PubMed: 27185944]

Michielli Nicola, Acharya U.Rajendra, Molinari Filippo, 2019. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. Comput. Biol. Med 106 (March), 71–81. 10.1016/j.compbiomed.2019.01.013. [PubMed: 30685634]

Mousavi Sajad, Afghah Fatemeh, Rajendra Acharya, U., 2019. SleepEEGNet: automated sleep stage scoring with sequence to sequence deep learning approach. PLOS ONE 14 (5), e0216456. 10.1371/journal.pone.0216456. [PubMed: 31063501]

Niethard Niels, Hasegawa Masashi, Itokazu Takahide, Oyanedel Carlos N., Born Jan, Sato Takashi R., 2016. Sleep-stage-specific regulation of cortical excitation and inhibition. Curr. Biol 26 (20), 2739–2749. 10.1016/j.cub.2016.08.035. [PubMed: 27693142]

Niethard Niels, Ngo Hong-Viet V., Ehrlich Ingrid, Born Jan, 2018. Cortical circuit activity underlying sleep slow oscillations and spindles. Proc. Natl. Acad. Sci 115 (39), E9220–E9229. 10.1073/pnas.1805517115. [PubMed: 30209214]

Niethard Niels, Brodt Svenja, Born Jan, 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.: Off. J. Soc. Neurosci 41 (19), 4212–4222. 10.1523/JNEUROSCI.1957-20.2021.

Nietz Angela K., Popa Laurentiu S., Streng Martha L., Carter Russell E., Kodandaramaiah Suhasa B., Ebner Timothy J., 2022. Wide-field calcium imaging of neuronal network dynamics in vivo. Biology 11 (11), 1601. 10.3390/biology11111601. [PubMed: 36358302]

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]

Paxinos George, Franklin Keith B.J., 2019. Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates. Academic Press.

Ren Chi, Komiyama Takaki, 2021. Characterizing cortex-wide dynamics with wide-field calcium imaging. J. Neurosci 41 (19), 4160–4168. 10.1523/JNEUROSCI.3003-20.2021. [PubMed: 33893217]

Rosenberg Richard S and Van Hout Steven. n.d. The American Academy of Sleep Medicine Inter-Scorer Reliability Program: Sleep Stage Scoring. Journal of Clinical Sleep Medicine 09 (01): 81–87. 〈10.5664/jcsm.2350〉 .

Saadatnejad Saeed, Oveisi Mohammadhosein, Hashemi Matin, 2020. LSTM-Based ECG classification for continuous monitoring on personal wearable devices. IEEE J. Biomed. Health Inform 24 (2), 515–523. 10.1109/JBHI.2019.2911367. [PubMed: 30990452]

Silasi Gergely, Xiao Dongsheng, Vanni Matthieu P., Chen Andrew C.N., Murphy Timothy H., 2016. Intact skull chronic windows for mesoscopic wide-field imaging in awake mice. J. Neurosci. Methods 267 (July), 141–149. 10.1016/j.jneumeth.2016.04.012. [PubMed: 27102043]

Silber Michael H., Ancoli-Israel Sonia, Bonnet Michael H., Chokroverty Sudhansu, Grigg-Damberger Madeleine M., Hirshkowitz Max, Kapen Sheldon, et al. , 2007. The visual scoring of sleep in adults. J. Clin. Sleep. Med.: JCSM: Off. Publ. Am. Acad. Sleep. Med 3 (2), 121–131.

Tezuka Taro, Kumar Deependra, Singh Sima, Koyanagi Iyo, Naoi Toshie, Sakaguchi Masanori, 2021. Real-time, automatic, open-source sleep stage classification system using single EEG for mice. Sci. Rep 11 (1), 11151 10.1038/s41598-021-90332-1. [PubMed: 34045518]

Turner Kevin L., Gheres Kyle W., Proctor Elizabeth A., Drew Patrick J., 2020. Neurovascular coupling and bilateral connectivity during NREM and REM Sleep. eLife 9 (October), e62071. 10.7554/eLife.62071. [PubMed: 33118932]

Wan Zhijiang, Li Manyu, Liu Shichang, Huang Jiajin, Tan Hai, Duan Wenfeng, 2023. EEGformer: a transformer–based brain activity classification method using EEG signal (https://www.frontiersin.org/articles/). Front. Neurosci 17. 10.3389/fnins.2023.1148855.

Wang Ziyue, Fei Xiang, Liu Xiaotong, Wang Yanjie, Hu Yue, Peng Wanling, Wang Ying-wei, Zhang Siyu, Xu Min, 2022. REM Sleep Is Associated with Distinct Global Cortical Dynamics and Controlled by Occipital Cortex. Nat. Commun 13 (1), 6896. 10.1038/s41467-022-34720-9. [PubMed: 36371399]

West Sarah L., Aronson Justin D., Popa Laurentiu S., Feller Kathryn D., Carter Russell E., Chiesl William M., Gerhart Morgan L., et al. , 2022. Wide-field calcium imaging of dynamic cortical networks during locomotion. Cereb. Cortex 32 (12), 2668–2687. 10.1093/cercor/bhab373. [PubMed: 34689209]

Whitney CW, Gottlieb DJ, Redline S, Norman RG, Dodge RR, Shahar E, Surovec S, Nieto FJ, 1998. Reliability of scoring respiratory disturbance indices and sleep staging. Sleep 21 (7), 749–757. 10.1093/sleep/21.7.749. [PubMed: 11286351]

Wright Patrick W., Brier Lindsey M., Bauer Adam Q., Baxter Grant A., Kraft Andrew W., Reisman Matthew D., Bice Annie R., Snyder Abraham Z., Lee Jin-Moo, Culver Joseph P., 2017. Functional connectivity structure of cortical calcium dynamics in anesthetized and awake mice. edited by daniele marinazzo. PLoS ONE 12 (10), e0185759. 10.1371/journal.pone.0185759. [PubMed: 29049297]

Wu Xin, Yang Juan, Pan Yu, Zhang Xiangmin, Luo Yuxi, 2020. Automatic Sleep-Stage Scoring Based on Photoplethysmographic Signals. Physiol. Meas 41 (6), 065008 10.1088/1361-6579/ab921d. [PubMed: 32392540]

Xue Yihao, Joshi Siddharth, Nguyen Dang, and Mirzasoleiman Baharan. 2024. Understanding the Robustness of Multi-Modal Contrastive Learning to Distribution Shift. arXiv. 〈10.48550/arXiv.2310.04971〉 .

Zhang Xiaohui, Landsness Eric C., Culver Joseph P., Anastasio Mark A., 2022. Identifying Functional Brain Networks from Spatial-Temporal Wide-Field Calcium Imaging Data via a Recurrent Autoencoder. PC11946:PC1194612 Neural Imaging and Sensing 2022. SPIE. 10.1117/12.2626317.

Zhang Xiaohui, Landsness Eric C., Chen Wei, Miao Hanyang, Tang Michelle, Brier Lindsey M., Culver Joseph P., Lee Jin-Moo, Anastasio Mark A., 2022. Automated sleep state classification of wide-field calcium imaging data via multiplex visibility graphs and deep learning. J. Neurosci. Methods 366 (January), 109421. 10.1016/j.jneumeth.2021.109421. [PubMed: 34822945]

Zhang Xiaohui, Gangopadhyay Ahana, Chang Hsi-Ming, Soni Ravi, 2023. Diffusion Model-Based Data Augmentation for Lung Ultrasound Classification with Limited Data. Proceedings of the 3rd machine learning for health symposium. PMLR, pp. 664–676. In: 〈https://proceedings.mlr.press/v225/zhang23a.html〉 .

Zhao Caihong, Li Jinbao, Guo Yahong, 2022. SleepContextNet: a temporal context network for automatic sleep staging based single-channel EEG. Comput. Methods Prog. Biomed 220 (June), 106806 10.1016/j.cmpb.2022.106806.

Zhuang Lan, Dai Minhui, Zhou Yi, Sun Lingyu, 2022. Intelligent automatic sleep staging model based on CNN and LSTM. Front. Public Health 10 (July), 946833. 10.3389/fpubh.2022.946833. [PubMed: 35968483]

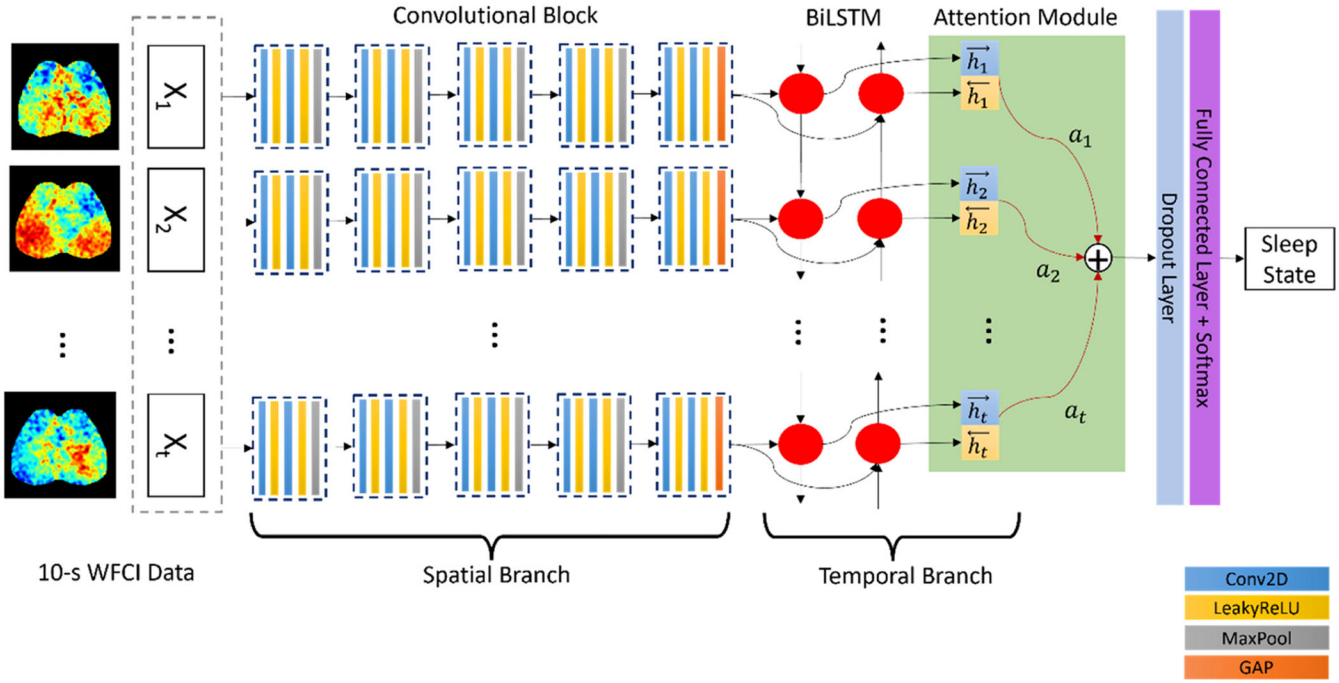

Fig. 1.

Illustration of the attention-based CNN-BiLSTM model architecture used in the study. The CNN-BiLSTM is a hybrid model consisting of a spatial branch to extract spatial features of WFCI image frames and a temporal branch to capture the temporal dependency among time steps in the WFCI recordings. Five convolutional blocks (within black dashed line) were employed in the CNN, in which two convolutional layers (blue) activated by Leaky ReLU (yellow) and max pooling layer (gray) were included. A global average pooling layer (orange) was placed after the final convolutional layer. The following bidirectional LSTM architecture receives the extracted spatial features from CNN as inputs. The BiLSTM layer (red) is composed of 64 units in each direction. A temporal attention mechanism was applied to the BiLSTM outputs before they are combined by use of a fully connected later with softmax activation to perform the final classification.

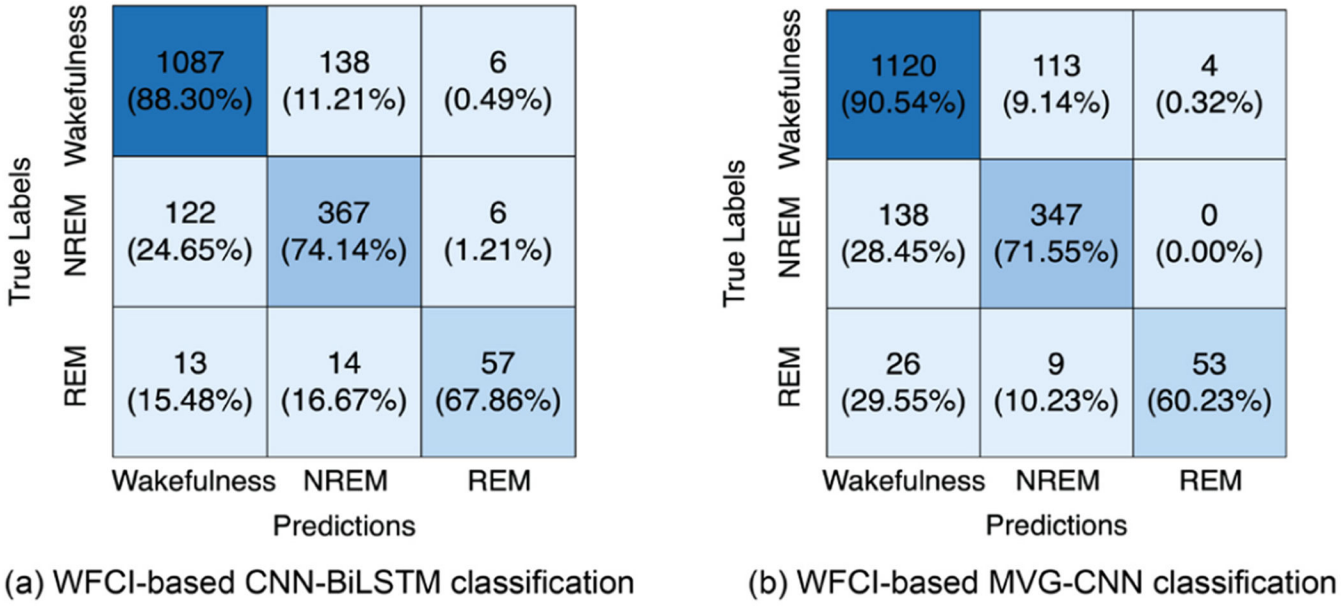

Fig. 2.

Confusion matrices for the CNN-BiLSTM and the MVG-CNN models, respectively, for classification of wakefulness, NREM, and REM in test set (n=1810 epochs). Manual EEG/EMG-based scoring is on the x-axis and MVG-CNN predictions are on y-axis. The diagonal cells 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.

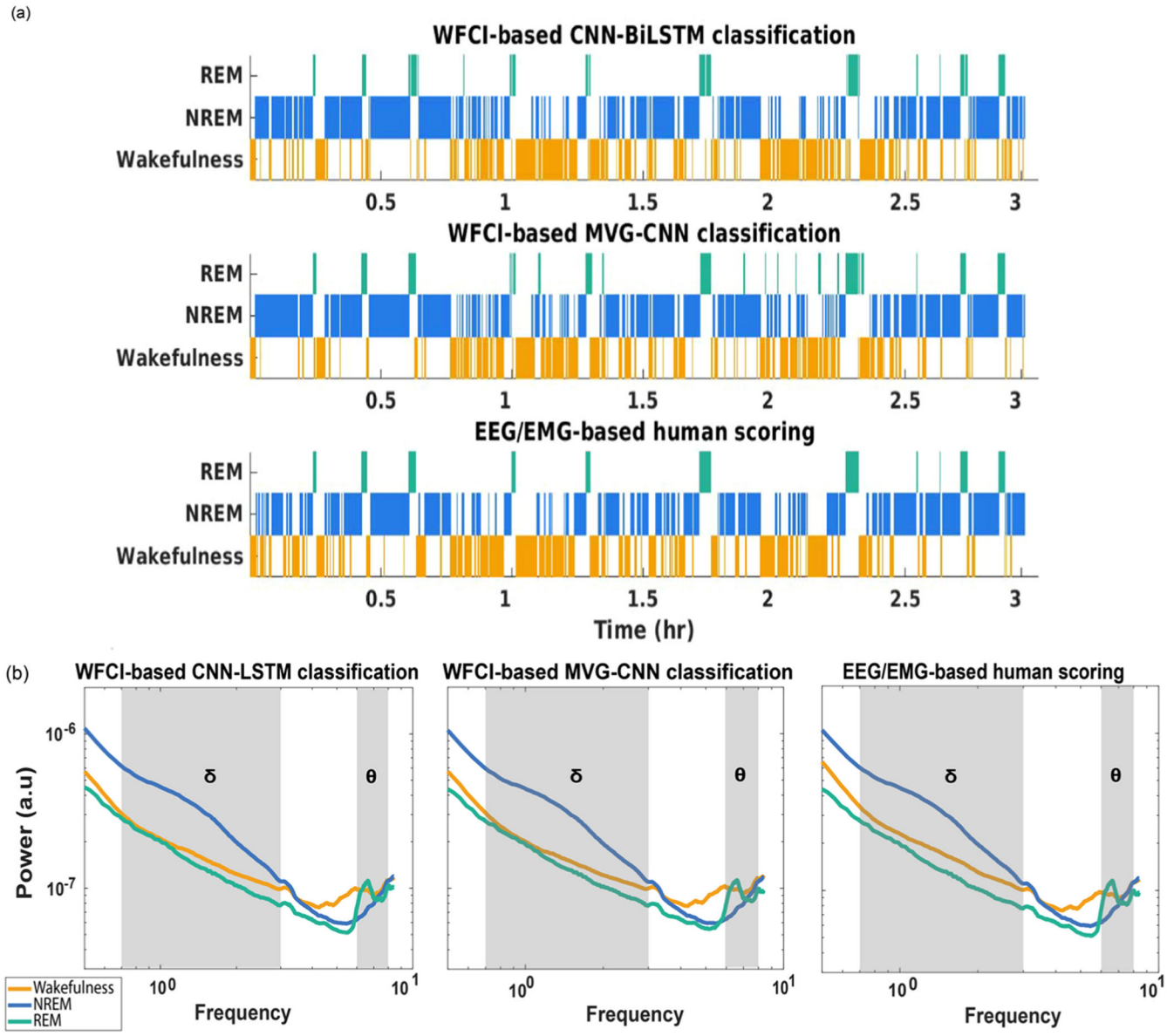

Fig. 3.

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

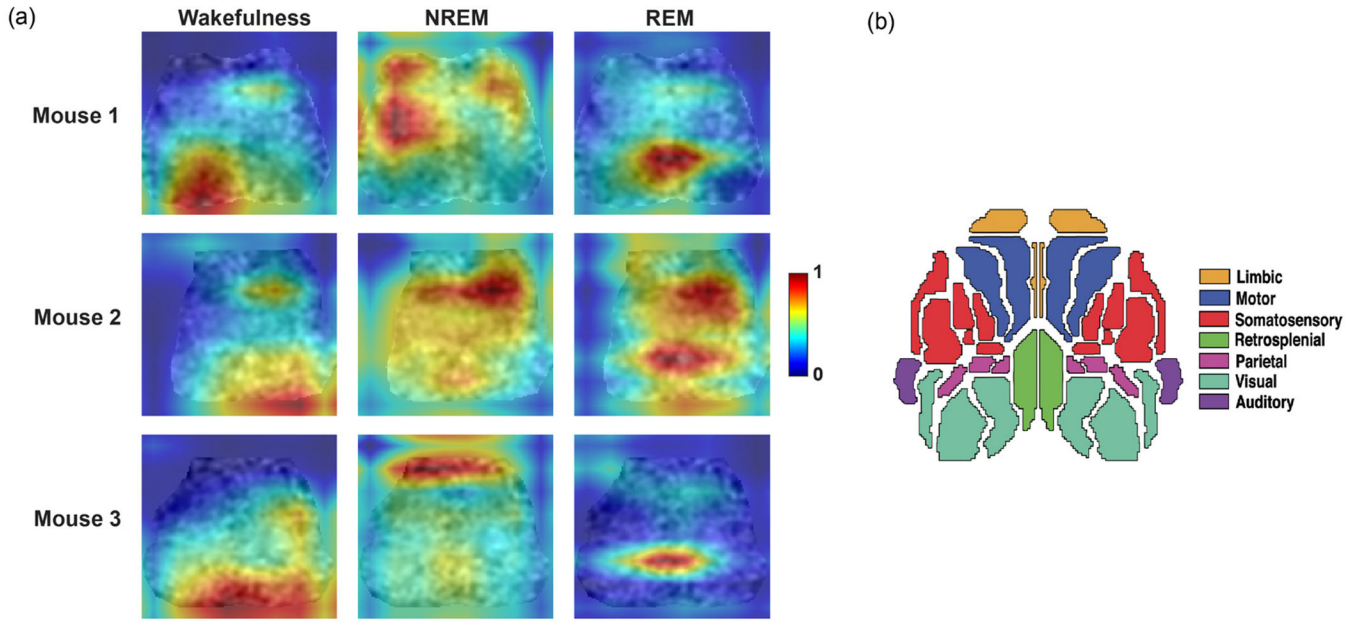

Fig. 4.

(a) 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 when making corresponding decisions. (b) Brain regions defined by the Paxinos atlas (Paxinos and Franklin, 2019) within the field of view.

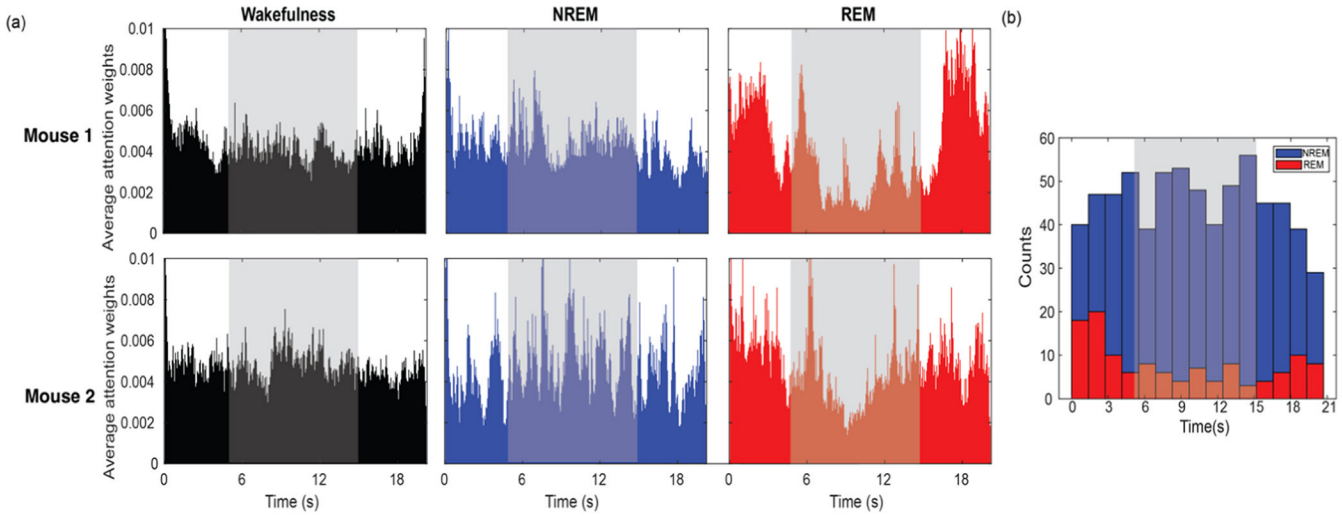

Fig. 5.

(a) Mean values of attention weights of two individual mouse for 20-s data epoch of wakefulness (black), NREM (blue) and REM (red). The shaded gray areas represent the central 10-s of data within the epoch. (b) Histogram showing the time step of highest attention weights of all individual 20-s epochs for NREM and REM for the two mice in (a).

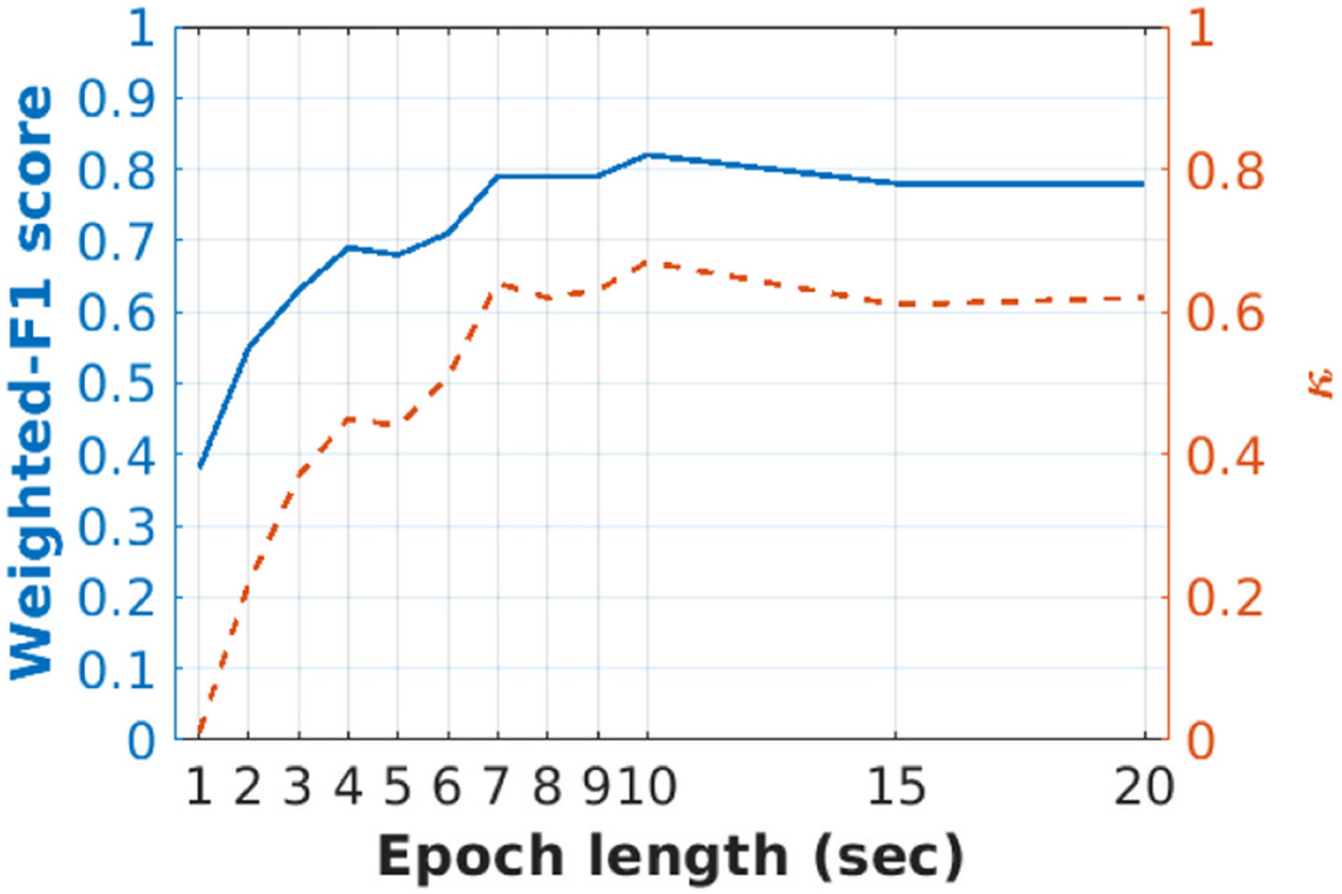

Fig. 6.

The sleep state classification performance with respect to the epoch durations used in the CNN-BiLSTM. As epoch duration was varied, weighted F1-score (left y-axis, solid blue line) and the Cohen's Kappa statistic (right y-axis, dotted red line) were compared.

Table 1

Metrics to evaluate the sleep stage classification performance on test data (n=1810 epochs). Both WFCI-based CNN-BiLSTM model and MVG-CNN model achieved substantial agreement of $\kappa = 0.64$ compared to manual EEG/EMG-based scoring. Prec., precision; Rec., recall; Acc., accuracy; κ, Cohen's Kappa statistic.

Scoring method	Wakefulness		NREM		REM		Acc.	F1-score	κ
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.			
CNN-BiLSTM	0.89	0.88	0.71	0.74	0.83	0.68	0.83	0.84	0.64
MVG-CNN	0.87	0.91	0.74	0.72	0.93	0.60	0.84	0.84	0.64

J Neurosci Methods. Author manuscript; available in PMC 2025 May 26.

Table 2

Comparison of sleep fragmentation among WFCI-based sleep state classification by use of the proposed CNN-BiLSTM model, the MVG-CNN method and the EEG/EMG-based human scoring. The state transitions refer to change from any state to any other state.

Scoring method	Average sleep state length (sec)			Number of state transitions
	Wakefulness	NREM	REM	
CNN-BiLSTM	35	57	37	238
MVG-CNN	29	61	35	245
Human annotator	47	66	81	182

Table 3

Mean and standard deviation of peak activation in all Grad-CAM maps for individual mouse. Lower values are indicative of frontal peak activation and higher values of posterior peak activation.

Sleep state	Mouse 1	Mouse 2	Mouse 3
Wakefulness	107.71±18.10	98.95±28.76	106.88±18.82
NREM	59.76±24.45	53.34±21.54	48.49±36.63
REM	88.40±10.14	71.44±27.81	87.42±0.51