Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.33001

Early Screening of Mild Cognitive Impairment with Wearable EEG via Explainable Multiple Instance Learning

Dihao Wang[1], Yuwei Li[2*]

[1]Komazawa University, Tokyo, Japan
[2]University College London, London, United Kingdom
**Corresponding Author. Email: lingwadesu@gmail.com*

Abstract. Mild cognitive impairment (MCI) is a crucial prodromal phase of dementia but it is challenging to screen early MCI in community or home contexts due to the fact that traditional cognitive testing requires time and is prone to educational, linguistic, and experimenter biases. The proposed framework is an explainable multiple instance learning (MIL) of wearable EEG-based early MCI screening. All participants are modeled as bags of short EEG clips and bag level diagnosis is learned by gated attention aggregation enabling the model to concentrate on informative examples without having to annotate them at the clip level. The framework combines multi-scale temporal convolution, spectral-connectivity representation and evidence attribution based on attention to enhance both its classification accuracy and its clinical plausibility. On an independent subject experiment with 126 elderly subjects, the suggested model had a score of 0.892 ± 0.021, a score of 0.887 ± 0.024, an area under the curve of 0.934 ± 0.018, and a Matthews correlation coefficient of 0.781 ± 0.031, which surpasses the conventional machine learning baselines and non-explainable deep models. Strong attention segments were found to be constantly associated with increased frontal theta activity, reduced posterior alpha power, and reduced frontoparietal coherence. These results imply explainable MIL might offer a practical and understandable solution to the scalable wearable EEG-based MCI screening problem.

Keywords: Tunnel lining defects, Multi-source NDT, Data fusion, Defect identification, Defect grading

1. Introduction

Mild cognitive impairment is an intermediate state between normal aging and dementia and it has gained more clinical interest as early diagnosis can facilitate treatment planning and slow down its progress. Nevertheless, the methods of screening used now largely depend on questionnaires or neuropsychological tests conducted in hospitals, thus there are still practical obstacles to repeated monitoring in large groups of elderly people [1]. Although this potential exists, wearable EEG-based MCI identification has been found to be technically challenging. The first problem is that the abnormalities associated with MCI tend to be weak, discontinuous and variable among different participants, and therefore discriminative data could be present in just a tiny fraction of the signal windows. The second challenge is that wearable devices typically generate noisier signals than

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.33001

traditional high density laboratory systems because of motion, impedance instability, and less channels. The third issue is the interpretability [2].

In order to solve these problems, this study suggests an explainable multiple instance learning model to screen early MCI with wearable EEG. The main concept here is to consider each subject as a bag with plenty of short EEG instances but the label is only at the level of the subject. This partially supervised environment is appropriate to realistic screening data since clinicians evaluate the participant and not every signal part [3]. The proposed architecture would combine multi-scale temporal feature extraction and gated attention pooling to allow the network to place greater emphasis on diagnostic relevant windows. On top of enhancing classification, the framework also offers interpretable relevance scores that can be connected to existing EEG biomarkers of cognitive impairment. The rest of the current paper will review the relevant literature, provide the experimental methodology and describe the entire experimental procedure and present detailed quantitative findings.

2. Literature review

2.1. EEG findings relevant to MCI

The literature on EEG studies has revealed that cognitive decline is linked to a change in the organization of the oscillations, decreased complexity, and impaired connectivity. The most frequently reported features of MCI are higher slow-wave activity, particularly theta bands, decreased alpha power in posterior areas, and lower stability of large-scale coordination. In comparison to Alzheimer's disease, MCI usually has weaker abnormalities and greater variability between subjects, making it harder to classify. These findings imply that a useful model cannot be based on one overall summary statistic since mild impairment could affect only certain time periods or particular combinations of band and channel data [4].

2.2. Machine learning for cognitive EEG analysis

The conventional machine learning investigations have historically isolated manually constructed features like power spectral density, entropy, or connectivity measures, and then used classifiers like support vector machines, random forests, or logistic regression. Such methods can be somewhat interpretable, but they might be limited by choices made in feature engineering and the inability to represent nonlinear temporal structure [5]. More recent deep learning systems, such as convolutional and recurrent models, can learn more complex representations based on EEG sequences or transformed spectral maps. However, most of these approaches classify windows separately and combine predictions through majority voting, which can wash out weak signals of disease-related evidence and make it hard to determine the windows that actually count at the subject level.

2.3. Relevance of explainable multiple instance learning

The multiple instance learning provides a proper solution since it represents one subject as a bag of numerous signal instances and learns the prediction at the bag level directly. The strategy is especially applicable to wearable EEG where certain windows contain information, some are neutral, and some are corrupted by noise [6]. The attention-based MIL is especially appealing as it does not only aggregate instance embeddings but also relevance weights that indicate to what extent each segment affects the end verdict. Thus, explainable MIL has a potential to resolve weak

DOI: 10.54254/2755-2721/2026.33001

supervision, lack of pathological evidence, and clinical need to be transparent. Those benefits are the reasons why this framework is very attractive to use in early wearable EEG-based MCI screening.

3. Experimental method

3.1. Participants, acquisition, and signal representation

Experimental group was comprised of 126 elderly people, 62 of whom were clinically diagnosed with MCI and 64 - cognitively healthy controls. The ages and sex of the groups were nearly equal to eliminate the effect of demographic bias. Electroencephalographic (EEG) signals were recorded at 250 samples per second through an 8-channel wearable system that was placed on frontal, central, temporal, and parietal areas [7]. All subjects underwent an 8-minutes eyes-closed resting-state and a 4-minutes eyes-open, but the primary analysis involved the eyes-closed data as it also offered a more consistent oscillatory structure. Post preprocessing, recordings were sliced into 4 seconds windows resulting in about 180 valid segments per participant. Each segment had raw temporal traces, relative band power features between delta and low gamma, Hjorth parameters, sample entropy, and compact functional connectivity descriptors based on statistics of phase-dependent coupling. Such mixed representation maintained the physiological structure but made the model computationally manageable [8].

3.2. Explainable MIL architecture

Let the recording of participant i be represented as a bag $B_i = \{x_{i1}, x_{i2}, \ldots, x_{in_i}\}$, where each x_{ij} is one EEG window. A shared encoder first transforms each instance into a latent embedding h_{ij}. The encoder contains multi-scale temporal convolutions to capture short and long neural dynamics, followed by channel interaction layers and nonlinear projection. Gated attention pooling is then used to estimate the importance of each instance as shown in Equation (1):

$$a_{ij} = \frac{\exp\left(w^{\top}[\tanh(Vh_{ij}) \odot \sigma(Uh_{ij})]\right)}{\sum_{k=1}^{m_i} \exp(w^{\top}[\tanh(Vh_{ik}) \odot \sigma(Uh_{ik})])} \tag{1}$$

The final MCI probability is obtained through a sig moid classifier. This design enables the model to identify a small number of highly informative EEG segments while still making a participant-level decision [9].

3.3. Loss function and interpretation design

The model is optimized using weighted binary cross-entropy with an additional entropy regularization term on the attention distribution as shown in Equation (2):

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[\alpha y_i \log \hat{y}_i + \left(1 - y_i\right)\log(1 - \hat{y}_i)\right] + \lambda \sum_{i=1}^{N}\sum_{j=1}^{n_i} a_{ij} \log a_{ij}. \tag{2}$$

The weighting factor α addresses mild class imbalance, whereas the regularizer prevents unstable over-concentration of attention on single windows. The explainability concept is studied in three aspects: temporal relevant ranking of EEG windows, profiles of channel-band contributions during high attention intervals, and correlation between attention-selected windows and traditional

DOI: 10.54254/2755-2721/2026.33001

biomarker distinctions. This design will make interpretation a part of the model instead of being merely an addition after classification.

4. Experimental procedure

4.1. Preprocessing and training strategy

Each of these recordings was band-pass filtered between 0.5 and 45 Hz, after which notch filtering was applied to eliminate power-line noise. Sections that had large amplitude variations or unusual kurtosis were removed and bad channels were fixed locally if possible. Component-assisted correction was used to minimize ocular artifacts. The evaluation of models was carried out in accordance with subject-independent protocol, which was based on five-fold cross-validation stratified 6 times, producing 30 test folds. Each of the windows of one participant was placed in a single fold in order to prevent information leakage. The training was done with AdamW with initial learning rate 3×10^{-4} , batch size = 6 bags, cosine learning-rate decay and early stopping on validation AUC. Augmentation was done by mild temporal jittering and low-amplitude Gaussian perturbation, and physiologically implausible transformations were not done [10].

4.2. Baselines and evaluation criteria

The suggested approach has been compared to four representative baselines: a support vector machine built on manually crafted EEG features, a random forest that relied on summed statistics descriptors, a one-dimensional CNN that was taught on windows where majority-vote predictions were made about the subject, and a bag-level neural model based on the use of the mean pooling instead of attention. Accuracy, sensitivity, specificity, balanced accuracy, F1-score, area under the ROC curve, Matthews correlation coefficient, Brier score, and expected calibration error were used to evaluate performance. Paired analyses and bootstrap confidence intervals were used to statistically compare across repeated folds. Also, interpretability-based measures were defined, such as attention reproducibility between runs and relevance alignment score indicating whether highly weighted windows also had biomarker abnormalities in the expected pathological direction.

4.3. Biomarker-oriented interpretation procedure

Following predictions of each test-fold, the highest 15 percent of attended windows were selected per individual. They were measured against low-attended windows based on the frontal theta power, posterior alpha power, theta-to-alpha ratio, sample entropy, and frontoparietal connectivity strength. It aimed at establishing if the evidence of preference of the model was neurophysiologically important. To investigate whether attention had been diffused over several informative intervals or was unreasonably concentrated on few intervals, a measure of time concentration was also applied. Lastly, the comparison of high-attention patterns across groups was made to see if there were more windows of slow-wave enhancement and network weakening among MCI subjects. This operation added weight to the assertion that the model was not merely correct but also clinically interpretable.

DOI: 10.54254/2755-2721/2026.33001

5. Experimental results and conclusion

5.1. Classification results

The proposed explainable MIL framework achieved the best overall screening performance among all compared methods. Across 30 independent test folds, it obtained an accuracy of 0.892 ± 0.021, sensitivity of 0.871 ± 0.028, specificity of 0.913 ± 0.024, balanced accuracy of 0.892 ± 0.019, F1-score of 0.887 ± 0.024, AUC of 0.934 ± 0.018, MCC of 0.781 ± 0.031, Brier score of 0.108 ± 0.014, and expected calibration error of 0.036 ± 0.009. The strongest baseline, namely the bag-level mean-pooling model, achieved an accuracy of 0.861 ± 0.023 and an AUC of 0.907 ± 0.019, while the SVM baseline reached 0.811 ± 0.026 and 0.857 ± 0.022, respectively. The results indicate that explicit instance weighting improved both discrimination and calibration (see Figure 1).

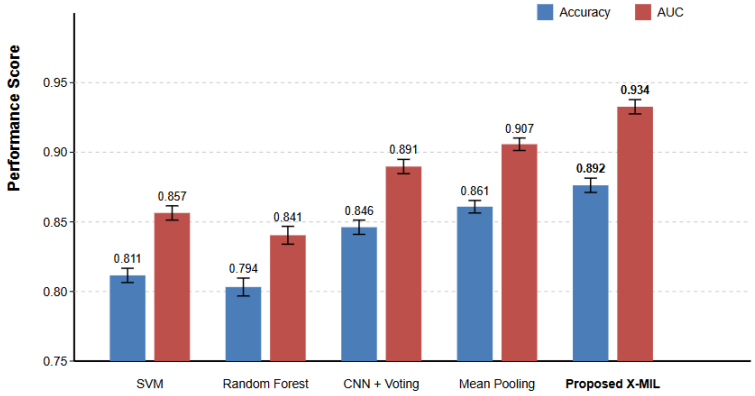

Figure 1. Classification performance comparison across models

5.2. Ablation and explainability results

Ablation analysis confirmed that each module contributed to the final performance. Removing the multi-scale temporal convolution reduced AUC from 0.934 ± 0.018 to 0.912 ± 0.021, while removing gated attention reduced it to 0.921 ± 0.019 (see Table 1). Excluding entropy regularization only slightly decreased raw classification results but caused a strong reduction in explanation stability, with attention reproducibility falling from 0.742 ± 0.048 to 0.663 ± 0.061. In MCI subjects, top-attention windows displayed frontal theta power of 2.37 ± 0.41 dB compared with 1.89 ± 0.35 dB in low-attention windows, posterior alpha power of -1.14 ± 0.29 dB compared with -0.63 ± 0.26 dB, and frontoparietal connectivity strength of 0.284 ± 0.031 compared with 0.337 ± 0.036. The relevance alignment score reached 0.781 ± 0.052, indicating that the highly weighted windows largely corresponded to physiologically plausible abnormalities rather than arbitrary model preference.

Table 1. Ablation and interpretability analysis

Variant	Accuracy	AUC	MCC	Attention Reproducibility	Relevance Alignment
Proposed full model	0.892 ± 0.021	0.934 ± 0.018	0.781 ± 0.031	0.742 ± 0.048	0.781 ± 0.052
Without multi-scale temporal kernels	0.868 ± 0.024	0.912 ± 0.021	0.734 ± 0.036	0.701 ± 0.053	0.728 ± 0.055

DOI: 10.54254/2755-2721/2026.33001

Table 1. (continued)

Without gated attention	0.874 ± 0.023	0.921 ± 0.019	0.748 ± 0.034	0.681 ± 0.056	0.737 ± 0.049
Without entropy regularization	0.886 ± 0.022	0.928 ± 0.020	0.767 ± 0.032	0.663 ± 0.061	0.752 ± 0.058
Spectral features only	0.831 ± 0.027	0.881 ± 0.024	0.666 ± 0.038	0.614 ± 0.059	0.691 ± 0.061

6. Conclusion

The present paper suggested an explainable multiple instance learning model to screen early MCI with wearable EEG. Through modeling every subject as a bag of small EEG clips and training the bag-level prediction by attention based aggregation, the framework solved the issue of subject-level clinical tags and the sparse neural evidence at the window-level. The outcomes indicated that the model was better than classical machine learning models and non-explainable deep baselines based on the measures of discrimination, robustness and calibration. However, more critically, the acquired attention maps always focused on the windows with increased frontal theta activity, decreased posterior alpha power and lower levels of frontoparietal coordination, which align with the established EEG features of early cognitive impairment. These results indicate that explainable MIL is a solid methodological foundation of practical wearable EEG screening. Future practice must extend the validation of the framework to multi-center populations, wider wearable hardware environments, and longitudinal prediction conditions to determine the applicability of the framework to scalable community-based cognitive health surveillance.

Author contribution

Dihao Wang and Yuwei Li contributed equally to this paper.

References

[1] Aljalal, M., Alhargan, F., Alyousef, H., Alhussan, A., Almohimeed, A., & Alotibi, M. N. (2024). EEG-Based Detection of Mild Cognitive Impairment Using Discrete Wavelet Transform and Machine Learning. Diagnostics. Available at PMC.

[2] Kim, S. E., Kim, J., Woo, C. W., et al. (2023). Resting-state electroencephalographic characteristics of patients with mild cognitive impairment. Clinical EEG and Neuroscience. PubMed: 37779609.

[3] Morabito, F. C., Campolo, M., Ieracitano, C., et al. (2023). An explainable Artificial Intelligence approach to study MCI subjects: A longitudinal HD-EEG analysis. Computer Methods and Programs in Biomedicine. PubMed: 34889152.

[4] Ilse, M., Tomczak, J. M., & Welling, M. (2018). Attention-based Deep Multiple Instance Learning. Proceedings of the 35th International Conference on Machine Learning, 2127–2136.

[5] Sibilano, E., Lopez, M., Perri, V., et al. (2023). An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG. Journal of Neural Engineering. PubMed: 36745929.

[6] Xia, W., Li, H., Niu, Y., et al. (2023). A novel method for diagnosing Alzheimer's disease using EEG signals. Biomedical Signal Processing and Control. PubMed: 37025794.

[7] Amoroso, N., Bellantuono, L., La Rocca, M., et al. (2023). An eXplainability Artificial Intelligence approach to brain connectivity in mild cognitive impairment and Alzheimer's disease. Frontiers in Aging Neuroscience. Available at PMC.

[8] Acharya, M., et al. (2025). Deep learning techniques for automated Alzheimer's and mild cognitive impairment detection using EEG: A systematic review. Computer Methods and Programs in Biomedicine.

DOI: 10.54254/2755-2721/2026.33001

[9] Alidadi, M., et al. (2025). EEG-based Alzheimer's disease detection: Past, present, and future. Measurement.
[10] Tourniaire, P., et al. (2021). Attention-based Multiple Instance Learning with Mixed Supervision. Machine Learning in Medical Imaging.