Transforming Brainwaves into Language: EEG Microstates Meet Text Embedding Models for Dementia Detection

Quoc-Toan Nguyen[1], Linh Le[1,2], Xuan-The Tran[1],
Dorothy Bai[3], Nghia Duong-Trung[4], Thomas Do[1], Chin-Teng Lin[1]
[1]*University of Technology Sydney, Sydney, Australia*
[2]*Mila - Quebec AI Institute, Montreal, Canada*
[3]*Taipei Medical University, Taipei, Taiwan*
[4]*German Research Center for Artificial Intelligence (DFKI), Berlin, Germany*
✉*Corresponding Authors:{Thomas.Do, Chin-Teng.Lin}@uts.edu.au*

Abstract

This study proposes a novel, scalable, non-invasive and channel-independent approach for early dementia detection, particularly Alzheimer's Disease (AD), by representing Electroencephalography (EEG) microstates as symbolic, language-like sequences. These representations are processed via text embedding and time-series deep learning models for classification. Developed on EEG data from 1001 participants across multiple countries, the proposed method achieves a high accuracy of 94.31% for AD detection. By eliminating the need for fixed EEG configurations and costly/invasive modalities, the introduced approach improves generalisability and enables cost-effective deployment without requiring separate AI models or specific devices. It facilitates scalable and accessible dementia screening, supporting timely interventions and enhancing AD detection in resource-limited communities.

1 Introduction

Dementia is recognised as the seventh leading cause of mortality globally and plays a major role in increasing disability and dependence among older adults (World Health Organization, 2023). Among the various forms of dementia, Alzheimer's Disease (AD) is the most prevalent, accounting for approximately 60% to 80% of all cases (The Alzheimer's Association, 2023; Nguyen, 2024; Nguyen et al., 2024), with a higher incidence observed in individuals aged 65 and above. AD is characterised by progressive cognitive deterioration, memory impairment, and neuronal loss, ultimately resulting in brain atrophy and tissue damage (van der Flier et al., 2023). Because no definitive cure currently exists (The Alzheimer's Association, 2023), detecting the disease at an early stage is critical for decelerating its progression and enhancing individuals' Quality of Life (QoL) through appropriate interventions and supportive care (Dubois et al., 2016; S et al., 2019).

The development of Artificial Intelligence (AI), including Machine Learning (ML) and Deep Learning (DL), has advanced significantly in early AD detection. Nevertheless, many of these techniques rely on costly modalities, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) (Dong et al., 2024; Ou et al., 2024; Altay et al., 2021; Rallabandi and Seetharaman, 2023), which are typically not viable in resource-limited communities. They also depend on invasive biomarkers such as Cerebrospinal Fluid (CSF) (Gogishvili and others., 2023; Nguyen and Duong-Trung, 2025), which can cause pain, reduce willingness to undergo testing, and limit their adoption. Therefore, Electroencephalogram (EEG) presents a non-invasive and more affordable option, making it more suitable for resource-constrained populations (Adebisi et al., 2024; Klepl et al., 2023; Lassi et al., 2023; Sharma et al., 2025; Nguyen, 2025a; Tran et al., 2024). In particular, EEG microstates[1] has emerged as a promising approach for AD detection, demonstrating notable performance over traditional EEG-based features (Smailovic et al., 2019; Mishra et al., 2020; Yang et al., 2024).

However, conventional AI models for EEG-based decision-making systems typically require a fixed number of input channels, necessitating the development of separate models for each EEG channel configuration. This constraint poses a sig-

[1]EEG microstates are quasi-stable periods of electrical topography across the scalp, most commonly derived from clustering EEG signals at peaks in Global Field Power (GFP). These transient states, typically lasting 80–120 milliseconds, represent the building blocks of spontaneous brain activity and provide insight into the temporal organisation of large-scale neural dynamics (Haydock et al., 2025; Nguyen, 2025b).

nificant barrier to the practical and cost-effective deployment of EEG-based AI systems for AD detection, particularly in resource-limited settings. In most clinical environments, EEG devices are expected to function in various medical applications, making it neither practical nor efficient to dedicate a specific system solely to AD detection or develop bespoke AI models for each device across different premises. Developing and maintaining multiple models for varying channel configurations imposes substantial resource demands, increases development and maintenance costs, and undermines the generalisability of these systems in real-world and clinical contexts. Therefore, developing AI models compatible with executing EEG data across varying channel configurations for AD detection is paramount, enhancing scalability, facilitating broader adoption, and improving clinical applicability to better support individuals in need.

Recently, text embedding models[2] have significantly advanced, transforming natural language inputs into semantically informative vector representations. This has enhanced performance across various Natural Language Processing (NLP) tasks, such as text classification and information retrieval (Kalidindi et al., 2024; Darrin et al., 2024; Enevoldsen et al., 2024). Notably, EEG signals also contain semantic representations with patterns that reflect meaningful cognitive states, beyond their electrical nature (Wang et al., 2024a; Mohammadi Foumani et al., 2024a; Feng et al., 2023; Wang and Ji, 2022). Hence, leveraging text embedding models to convert EEG microstates into standardised vector representations offers a promising new way to capture and analyse underlying cognitive patterns, enabling consistent representation across diverse EEG configurations.

This study utilises a dataset of 1001 participants from multiple countries and achieves an accuracy of 0.9431 using an advanced text embedding model (Darrin et al., 2024; Enevoldsen et al., 2024), text-embedding-3-small (Abdullahi et al., 2024), and a deep learning time-series model (Mohammadi Foumani et al., 2024b), Recurrent Neural Network (RNN) (Zucchet and Orvieto, 2024). This approach enables the development of an adaptive, high-performing AI model that generalises across heterogeneous EEG datasets. By removing the dependency on a fixed number of EEG channels, the framework eliminates the need for separate configuration-specific models, reducing financial and computational cost and clinical deployment complexity. In summary, this research addresses the following Research Questions (RQs):

- **RQ1:** Is it feasible to leverage text embedding models to capture meaningful and distinguishable representations from EEG data for AD detection?
- **RQ2:** How can text embedding models be utilised to standardise/generalise EEG microstates across varying channel configurations, allowing for an adaptive AI model applicable to multiple EEG channel setups in AD detection?
- **RQ3:** To what extent do the vector representations of Normal Control (NC) and AD cases reveal meaningful and statistically significant distinctions?

2 Related Work

Many studies have explored AI-based approaches for AD detection using EEG data, incorporating various ML and DL techniques across different channel configurations and sample sizes. This section summarises prominent contributions in the literature. One study proposed LCOWFBs-6 with 16 channels, reaching 0.9860 accuracy using 11 NC and 12 AD participants (Puri et al., 2023). Similarly, another investigation applied a k-NN classifier to 19-channel EEG data, reporting 0.9000 accuracy on a balanced dataset of 20 NC and 20 AD cases (Yifan et al., 2019). A CNN-based model was developed using 128 channels and achieved 0.7945 accuracy with 29 NC and 36 AD participants (Stefanou et al., 2025). The DEL model was presented using 19 channels, obtaining 0.9790 accuracy with 36 NC and 104 AD participants (Nour et al., 2024). Likewise, the DICE-Net approach utilised 19 channels to attain 0.8328 accuracy on 29 NC and 36 AD samples (Miltiadous et al., 2023a). A graph neural network (GNN) method achieved 0.9200 accuracy using 128-channel EEG from 20 NC and 20 AD subjects (Klepl et al., 2022), while a Gaussian Naïve Bayes (GNB) classifier applied to 128-channel EEG reached 0.8100

[2]Text embeddings are numerical representations of language that capture its semantic information (Wang et al., 2024b).

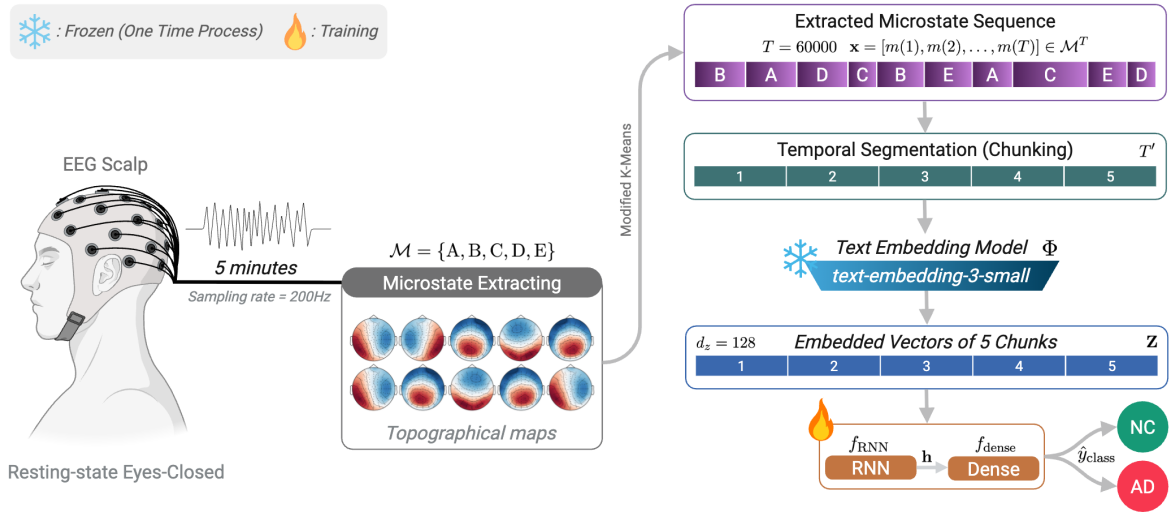


Figure 1: Proposed method of utilising Electroencephalogram (EEG) microstates with text embedding model and time-series deep learning for Alzheimer's Disease (AD) detection. NC: Normal Control, RNN: Recurrent Neural Network.

accuracy with 19 NC and 36 AD participants (Si et al., 2023).

Additionally, the DSL-GN hybrid model used 23 EEG channels and reached 0.9400 accuracy on 20 NC and 20 AD participants (Cao et al., 2024). Another work introduced LEADNet with 16 channels, reporting the highest accuracy of 0.9924 on a small dataset of 11 NC and 12 AD (Puri, 2024). An LSTM-based approach using 16-channel EEG achieved 0.9790 accuracy with 15 NC and 20 AD samples (Alessandrini et al., 2022). A comparative study applying k-NN and 19 channels reported 0.9300 accuracy on a dataset of 29 NC and 36 AD (Lal et al., 2024). Lastly, a CNN-based method with 19 channels achieved 0.9860 accuracy with 11 NC and 15 AD participants (Sen et al., 2023).

Despite promising results, three key research limitations exist in the current literature. First, most existing work is trained and validated on a single private dataset with a fixed EEG channel configuration, which restricts their ability to generalise across different EEG devices and clinical settings. Second, the limited sample sizes—often comprising tens of participants per group—undermine the generalisability of the models. Finally, the emphasis on achieving high predictive accuracy often overlooks the importance of thorough error analysis and the interpretation of group-level patterns. These analyses are essential for enhancing the transparency of AI systems, fostering user trust, and enabling more reliable systems.

3 Developed Approach

3.1 Background

3.1.1 Primer of EEG Microstates

The EEG microstate technique models brain signals as a sequence of discrete, non-overlapping topographic maps (Haydock et al., 2025), which are aligned with the original EEG data using spatial correlation methods (Tarailis et al., 2024). These signals are viewed as sequences of topographical patterns (Khanna et al., 2014). EEG microstates have been proven to effectively detect various neurological diseases due to their informative representations, such as AD (Lassi et al., 2023; Smailovic et al., 2019), Parkinson's disease, Mild Cognitive Impairment (MCI) (Chunguang et al., 2022), and epilepsy (SA et al., 2024).

The microstate extraction procedure was performed using the Global Field Power (GFP) method (Thomas et al., 2011). GFP is initially calculated at each time point:

$$GFP(t) = \sqrt{\frac{\sum_{i=1}^{n}(v_i(t) - \bar{v}(t))^2}{n}}, \quad (1)$$

where $v_i(t)$ represents the voltage recorded at electrode i, $\bar{v}(t)$ is the average voltage across all electrodes at time t, and n is the total number of electrodes. EEG scalp maps corresponding to GFP peaks—points of highest signal-to-noise ratio (SNR)—are selected and clustered using a modified k-means algorithm (Pascual-Marqui et al.,

1995). The Global Map Dissimilarity (GMD) (Pascual-Marqui et al., 1995) is used to quantify the similarity between two topographic maps and is computed as:

$$GMD_{u,v} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{u_i}{GFP_u} - \frac{v_i}{GFP_v}\right)^2} \quad (2)$$

As we can see in Figure 1, this study employs four standard microstates—A, B, C, and D—widely recognised in resting-state EEG literature for representing core functional networks: auditory, visual, salience, and attention (Armen et al., 2022). An additional category, microstate E, includes all scalp patterns that do not conform to the above four (Férat et al., 2022).

3.1.2 Text Embedding Models for EEG

Recent advances in pre-trained models originally developed for NLP have opened new avenues for their application to non-text modalities, particularly time-series data (Zhang et al., 2024). For example, the AutoTimes framework was introduced to leverage pre-trained architectures for autoregressive forecasting by encoding time series into a token-based embedding space and generating future values sequentially (Liu et al., 2024). One study explored the use of Large Language Models (LLMs) in mental health domains, focusing on the classification of depression and emotional states (Hu et al., 2024). Another investigation demonstrated the effectiveness of LLMs in handling forecasting tasks involving multivariate time series data (Tan et al., 2024). In a different approach, text embedding models were employed to encode time series data, which were subsequently used as input to classification models across multiple temporal tasks (Kaur et al., 2024). Especially, EEG signals have been shown to contain semantic representations in various tasks (Wang et al., 2024a; Mohammadi Foumani et al., 2024a; Feng et al., 2023; Wang and Ji, 2022).

According to these foundations, leveraging text embedding models to process EEG microstates data for AD detection can be a relevant approach as it aligns naturally with both time-series dynamics and symbolic representations of discrete states. In this paper, we explore using pre-trained text embedding models (Nguyen et al., 2025) to encode sequences of EEG microstates. By translating microstate dynamics into a structured token-like format, our approach facilitates consistent and scalable representation across heterogeneous EEG configurations (Jin et al., 2024), which is utilised as input for a time-series model (Mohammadi Foumani et al., 2024b) to detect AD.

3.2 Proposed Method

As illustrated by Figure 1, let

$$\mathcal{M} = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}\}$$

denote the finite set of EEG microstates. For a subject's EEG recording, the entire microstate sequence is represented as a function

$$m : \{1, 2, \ldots, T\} \to \mathcal{M},$$

where $T = 200 \times 60 \times 5 = 60000$ is the total number of time points for a 5-minute recording sampled at 200 Hz. This yields a symbolic sequence of the form

$$\mathbf{x} = [m(1), m(2), \ldots, m(T)] \in \mathcal{M}^T.$$

Step 1: Temporal Segmentation (Chunking)

Define the segmentation operator

$$\begin{aligned} &\mathcal{S}_N : \mathcal{M}^T \longrightarrow \prod_{i=1}^{N} \mathcal{M}^{T'}, \\ &T' = T/N = 12000, \\ &N = 5. \end{aligned}$$

For each chunk $i \in \{1, \ldots, 5\}$, define the corresponding time interval

$$\mathcal{I}_i = \{(i-1)T' + 1, \ldots, iT'\},$$

and extract the chunk as

$$\mathbf{x}_i = \mathbf{x}|_{\mathcal{I}_i} \in \mathcal{M}^{T'}.$$

Step 2: Text Embedding Transformation

Let `text-embedding-3-small`[3] be a pre-trained language embedding model adapted for EEG microstate sequences. Define the embedding function

$$\Phi_{\text{text-embedding-3-small}} : \mathcal{M}^{T'} \to \mathbb{R}^{d_z}, \quad d_z = 128,$$

which maps each symbolic sequence $\mathbf{x}_i$ (treated as a character string) into a continuous vector space:

$$\mathbf{z}_i = \Phi_{\text{text-embedding-3-small}}(\mathbf{x}_i) \in \mathbb{R}^{128}.$$

[3]The best performing model in this research among others (see Section 5).

All embedded segments are concatenated into a matrix

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_5 \end{pmatrix} \in \mathbb{R}^{5\times 128}.$$

Step 3: RNN-based Classifier

Let the RNN (Zucchet and Orvieto, 2024) be defined as

$$f_{\text{RNN}} : \mathbb{R}^{5\times 128} \to \mathbb{R}^{d_h},$$

which aggregates temporal embeddings into a latent representation:

$$\mathbf{h} = f_{\text{RNN}}(\mathbf{Z}) \in \mathbb{R}^{d_h}.$$

A dense layer f_{dense} maps the RNN output to logits $\mathbf{s} = f_{\text{dense}}(\mathbf{h}) \in \mathbb{R}^2$, from which class probabilities over $\mathcal{Y} = \{\text{NC}, \text{AD}\}$ are computed. The predicted class is

$$\hat{y}_{\text{class}} = \arg\max_{y\in\mathcal{Y}} \hat{\mathbf{y}}(y).$$

4 Experiments

4.1 Datasets

This research includes eyes-closed resting-state wet EEG data from 1001 participants, comprising 715 individuals classified as NC (mean age 58.02±8.91) and 286 as AD (mean age 74.84±8.25). Medical domain professionals clinically assessed and labelled the participants in ten countries. All EEG recordings were acquired by trained technicians following a standardised acquisition protocol, ensuring consistency in resting-state conditions. More information about the included datasets can be found in the Appendix in Table 3.

To maintain consistency and ensure cross-participant compatibility, all EEG data were resampled to 200Hz—a frequency demonstrated to be effective for AD detection in various studies (Rezaee and Zhu, 2025; Gutiérrez-de Pablo et al., 2024; Moguilner et al., 2024). For model training and evaluation, a fixed segment of 5 minutes (300 seconds) was extracted from each participant. EEG preprocessing steps (Haydock et al., 2025) included re-referencing to the average reference, band-pass filtering (1–40Hz), and artefact removal using Independent Component Analysis (ICA). These steps were proven to be essential for microstate analysis in various studies (Haydock et al., 2025).

4.2 Experimental Settings

The microstates are extracted using the Pycrostate library (Férat et al., 2022). RNN was configured with 32 units, followed by a dense output layer with softmax activation for binary classification (NC vs. AD). The model was trained using the Adam optimiser ($\alpha = 0.001$) and categorical cross-entropy loss, for up to 300 epochs with early stopping (patience $=$ 30) and a batch size of 32. We utilised OpenAI's text-embedding-3-small API[4] to generate fixed-dimensional embeddings from symbolic EEG microstate sequences, enabling consistent input representations. A 5-fold cross-validation was employed to comprehensively evaluate the model's performance across different data subsets. Evaluation metrics included accuracy, F1-score (Rainio et al., 2024), and the Brier score (Ovadia et al., 2019), providing a thorough assessment of both classification effectiveness and confidence calibration—key indicators of reliability in clinical AI applications.

5 Results

5.1 Model Results

Across all evaluated configurations, *text-embedding-3-small* emerged as the best-performing model, particularly when using an embedding size of 32 and a chunk size of 12000. Under this configuration, it achieved an accuracy of *0.9431±0.0288*, F1-score of *0.9023±0.0379*, and a Brier score of *0.0464±0.0192*, marking the highest accurate classification and calibration among all tested setups. These results indicate that *text-embedding-3-small* is not only highly effective in capturing discriminative patterns from EEG microstate sequences but also benefits substantially from longer input chunks while maintaining compact embedding dimensionality. Its stable and superior performance across both evaluation settings makes it a strong candidate for EEG-based AD detection tasks.

With embedding size fixed at 32 (see Table 1, Figure 7a in the Appendix), increasing the chunk size led to notable performance improvements for *text-embedding-3-small*, rising from *0.8701±0.0483* accuracy at 3000 to *0.9431±0.0288* at 12000. This trend was not universally observed across all models. While some models like *Solon-embeddings-*

[4] https://platform.openai.com

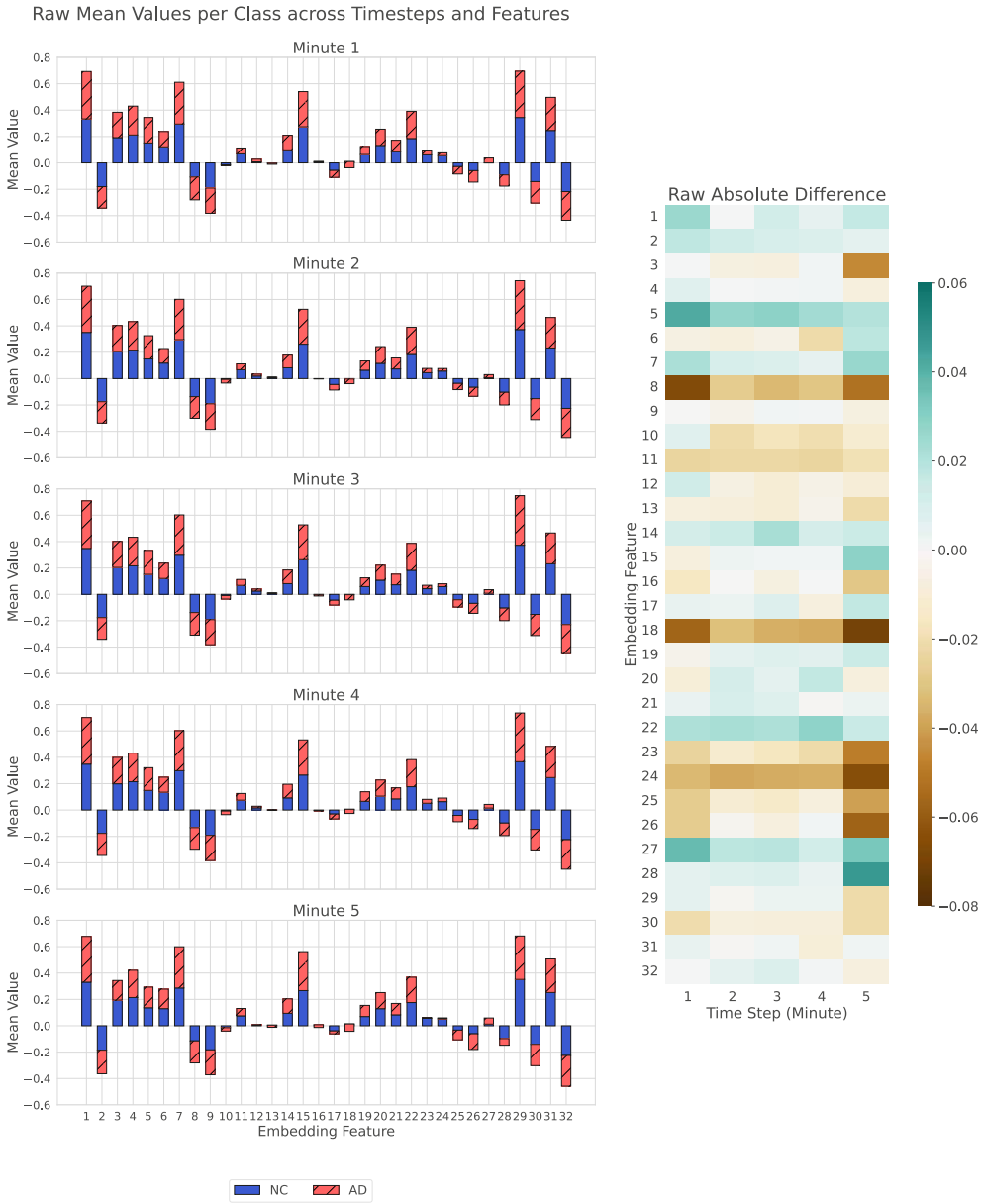


Figure 2: Feature distribution of Normal Control (NC) and Alzheimer's Disease (AD) with raw absolute difference.

large-0.1 maintained relatively stable performance across chunk sizes, others like *granite-embedding-278m-multilingual* and *bge-m3* experienced declining accuracy and F1 scores with longer chunks. For instance, *granite-embedding-278m-multilingual* dropped in accuracy from *0.7832*±*0.0198* to *0.7343*±*0.0376* as chunk size increased. This highlights that while longer sequence contexts can enrich temporal patterns for classification, model-specific architectural design dictates the extent to which such information can be effectively utilised.

At a fixed chunk size of 12000 (see Table 2, Figure 7b in the Appendix), smaller embedding sizes generally resulted in better performance across models. *text-embedding-3-small* again led with an accuracy of *0.9431*±*0.0288* at embedding size 32, while its performance gradually decreased at 64 and 128 dimensions. For other models, the performance drop was more noticeable; for example, `Solon-embeddings-large-0.1` saw a decrease in F1-score from *0.5721*±*0.0807* at size 32 to just *0.2879*±*0.3288* at size 128. These findings suggest that lower-dimensional embeddings may more effectively retain task-relevant signal representations, potentially mitigating the risk of overfitting and reducing the propagation of irrelevant noise often associated with high-dimensional latent spaces, particularly in EEG microstates.

Compared to prior studies (see Table 4 in the

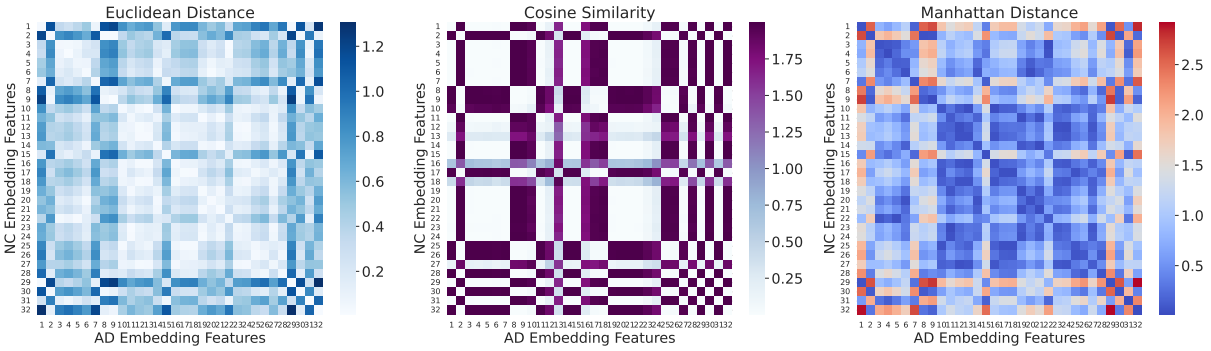


Figure 3: Feature-wise distance between embedded vectors of Normal Control (NC) and Alzheimer's Disease (AD) groups.

Appendix), the proposed method offers greater generalisability and reliability by supporting diverse EEG channel configurations (19/64/128 channels) and a significantly larger participant cohort, making it especially suitable for real-world clinical applications.

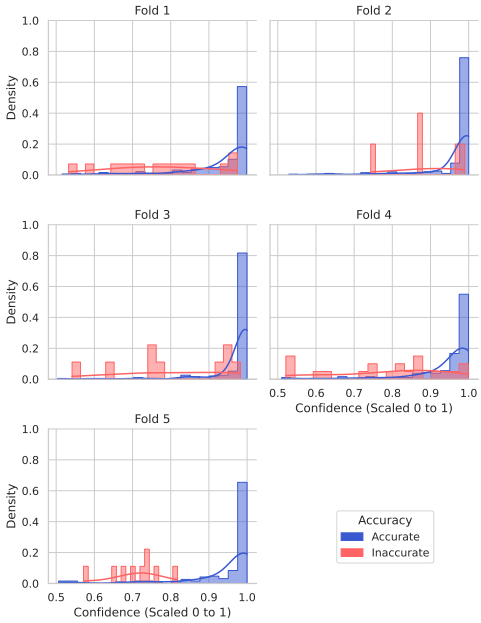


Figure 4: Confidence histograms of five folds with accurately and inaccurately classified sample distribution.

5.2 Error Analysis

This section details the error analysis of the best-performing model (text-embedding-3-small) as presented in the previous section. The model demonstrates consistent performance in classifying AD and NC cases across all validation folds (see Figure 5). True positive counts for AD range from 45 to 56, while true negatives for NC remain high at 124 to 146, indicating strong sensitivity and specificity. Misclassifications are infrequent,

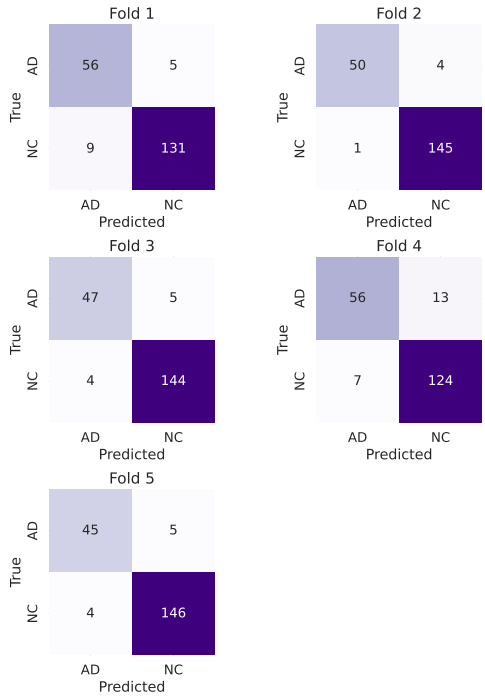


Figure 5: Confusion matrices across all five folds.

with false positives ranging from 1 to 7 and false negatives between 4 and 13, reflecting balanced model behaviour. Notably, even in Fold 4—where AD misclassification was highest—the model preserved a strong detection rate.

This stability is further proven by the model's confidence scores (see Figure 4 and Table 5 in the Appendix), which are a vital component of a reliable AI model. Correctly classified NC cases consistently exhibit high confidence (0.953–0.977), and AD cases follow closely (0.882–0.955), though the latter suggests potential for improvement. Importantly, across all folds, confidence scores for correctly predicted samples are significantly higher than those for incorrect

Table 1: Results of text embedding models with embedding size 32 and different chunk sizes.

Text Embedding Model	Chunk	Accuracy ↑	F1 ↑	Brier ↓
Solon-embeddings-large-0.1	3000	0.8042±0.0354	0.6264±0.1132	0.1299±0.0182
Solon-embeddings-large-0.1	6000	0.8002±0.0238	0.6248±0.0166	0.1308±0.0145
Solon-embeddings-large-0.1	12000	0.7912±0.0271	0.5721±0.0807	0.1409±0.0125
bge-m3	3000	0.8052±0.0128	0.6379±0.0627	0.1422±0.0058
bge-m3	6000	0.7782±0.0356	0.5052±0.1040	0.1550±0.0255
bge-m3	12000	0.7752±0.0171	0.5038±0.0863	0.1598±0.0106
granite-embedding-278m-multilingual	3000	0.7832±0.0198	0.5753±0.0619	0.1478±0.0104
granite-embedding-278m-multilingual	6000	0.7612±0.0129	0.4463±0.1123	0.1632±0.0087
granite-embedding-278m-multilingual	12000	0.7343±0.0376	0.4122±0.1191	0.1685±0.0209
gte-multilingual-base	3000	0.8172±0.0344	0.6409±0.0631	0.1325±0.0241
gte-multilingual-base	6000	0.7972±0.0406	0.5448±0.1002	0.1397±0.0188
gte-multilingual-base	12000	0.7702±0.0339	0.5445±0.0602	0.1547±0.0131
multilingual-e5-large-instruct	3000	0.7673±0.0410	0.5532±0.0620	0.1505±0.0179
multilingual-e5-large-instruct	6000	0.7882±0.0268	0.5805±0.0558	0.1422±0.0201
multilingual-e5-large-instruct	12000	0.7772±0.0199	0.5350±0.0842	0.1455±0.0080
snowflake-arctic-embed-l-v2.0	3000	0.8382±0.0390	0.7048±0.0325	0.1122±0.0225
snowflake-arctic-embed-l-v2.0	6000	0.8002±0.0277	0.6100±0.0517	0.1410±0.0166
snowflake-arctic-embed-l-v2.0	12000	0.7602±0.0310	0.3688±0.1957	0.1599±0.0134
text-embedding-3-small	3000	0.8701±0.0483	0.7735±0.0432	0.0922±0.0249
text-embedding-3-small	6000	0.9141±0.0224	0.8490±0.0450	0.0595±0.0149
text-embedding-3-small	12000	0.9431±0.0288	0.9023±0.0379	0.0464±0.0192

predictions ($p < 0.001$), with most misclassified samples exhibiting scores below 0.80, allowing the model to effectively signal its uncertainty and support clinical decision-making. However, occasional overconfidence in misclassified AD samples (e.g., 0.925 in Fold 2) and limited statistical significance in error trends (only Fold 4 with $p < 0.05$) suggest the need for further improvement. These issues likely stem from the class imbalance—smaller AD sample sizes (50–69 per fold) compared to NC (131–150), which may hinder learning and affect confidence calibration. While the imbalance between NC and AD samples, particularly the limited representation of AD cases, likely contributes to variability in confidence calibration, addressing this issue remains challenging due to the time-intensive nature of collecting clinically validated datasets. Nonetheless, the model's current performance demonstrates strong potential, and the observed trends highlight an important area for future refinement through more balanced data collection efforts.

5.3 Pattern Analysis

To investigate group-wise distinctions in embedded representations generated by text-embedding-3-small (see Figures 6 and 2), we conducted Mann–Whitney U tests across 32 embedding features, segmented by five minutes and across different distance metrics. The statistical analysis revealed that a substantial number of embedding dimensions demonstrated significant distributional differences between the NC and AD groups.

Across five one-minute segments (see Figure 2 and Table 6 in the Appendix), features such as 2, 3, 5–8, 10–11, 13–14, and 18–25 consistently yielded $p < 0.001$, underscoring that these are feasible to capture group-level divergence over time. Features such as 1, 4, and 9 exhibited inconsistent statistical significance across time windows and distance metrics, suggesting that their discriminative power can be highly dependent on transient, non-systematic variations in the data, such as inter-individual variability or momentary signal fluctuations unrelated to disease status.

Distance-based comparisons using Euclidean, Cosine, and Manhattan metrics further validated the discriminative capacity of the embedding space (see Figure 3 and Table 7). Of the 32 embedding features, over two-thirds (22 features) demonstrated statistically significant differences (at least $p < 0.05$) between NC and AD groups under two/three distance measures. A subset of features (approximately 20% remained consistently significant ($p < 0.001$) across all three metrics, underscoring their ability as class-discriminative markers in the latent space.

Further, Kruskal–Wallis tests conducted independently within the NC and AD groups (see Table 8) revealed that more than one-third of the embedding features exhibited significant intra-group distributional differences ($p < 0.01$). This observation suggests that these features not only capture between-group separability but also reflect internal heterogeneity within each clinical cohort, po-

Table 2: Results of text embedding models with chunk size 12000 and different embedding sizes.

Text Embedding Model	Embedding Size	Accuracy ↑	F1 ↑	Brier ↓
Solon-embeddings-large-0.1	32	0.7912±0.0271	0.5721±0.0807	0.1409±0.0125
Solon-embeddings-large-0.1	64	0.7752±0.0367	0.5155±0.1429	0.1549±0.0202
Solon-embeddings-large-0.1	128	0.7552±0.0375	0.2879±0.3288	0.1710±0.0249
bge-m3	32	0.7752±0.0171	0.5038±0.0863	0.1598±0.0106
bge-m3	64	0.7422±0.0528	0.3500±0.2228	0.1781±0.0299
bge-m3	128	0.7233±0.0388	0.0917±0.2050	0.1940±0.0186
granite-embedding-278m-multilingual	32	0.7343±0.0376	0.4122±0.1191	0.1685±0.0209
granite-embedding-278m-multilingual	64	0.7153±0.0383	0.0182±0.0407	0.1946±0.0172
granite-embedding-278m-multilingual	128	0.7143±0.0388	0.0000±0.0000	0.2004±0.0158
gte-multilingual-base	32	0.7702±0.0339	0.5445±0.0602	0.1547±0.0131
gte-multilingual-base	64	0.7832±0.0361	0.4694±0.2695	0.1445±0.0225
gte-multilingual-base	128	0.7903±0.0691	0.4739±0.2813	0.1522±0.0500
multilingual-e5-large-instruct	32	0.7772±0.0199	0.5350±0.0842	0.1455±0.0080
multilingual-e5-large-instruct	64	0.7393±0.0115	0.1812±0.2227	0.1793±0.0129
multilingual-e5-large-instruct	128	0.7392±0.0591	0.1410±0.3152	0.1816±0.0336
snowflake-arctic-embed-l-v2.0	32	0.7602±0.0310	0.3688±0.1957	0.1599±0.0134
snowflake-arctic-embed-l-v2.0	64	0.7992±0.0391	0.5507±0.1551	0.1337±0.0235
snowflake-arctic-embed-l-v2.0	128	0.7352±0.0525	0.1322±0.2956	0.1815±0.0273
text-embedding-3-small	32	0.9431±0.0288	0.9023±0.0379	0.0464±0.0192
text-embedding-3-small	64	0.9291±0.0135	0.8701±0.0340	0.0558±0.0129
text-embedding-3-small	128	0.8761±0.0751	0.7127±0.2493	0.0899±0.0520

tentially encoding subtle variations in cognitive-linguistic patterns or disease stage progression.

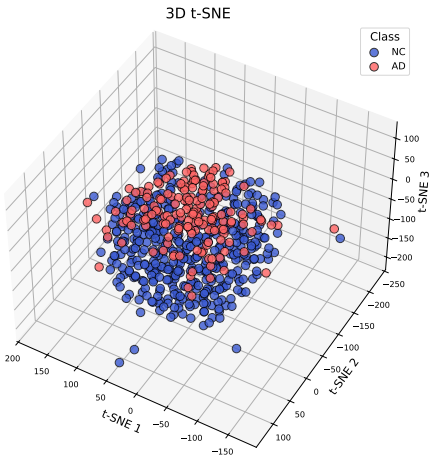


Figure 6: t-SNE of embedded vectors of Normal Control (NC) and Alzheimer's Disease (AD).

6 Conclusion and Discussion

This study presents a high-performing and scalable approach for AD detection using EEG data. Leveraging a large-scale dataset of 1001 participants, the proposed method achieves an accuracy of *0.9431* and a well-calibrated Brier score of *0.0464*. The method is beneficial for broader community use, as it leverages the affordability of EEG and adapts to varying channel configurations, enabling scalable and cost-effective deployment in resource-limited settings for early AD detection.

For **RQ1**, we demonstrate that text embedding models can effectively extract meaningful and discriminative representations from EEG data. The proposed method utilises EEG microstate sequences as text-like symbolic inputs and applies a deep learning architecture with the *text-embedding-3-small* model and RNN as key components. Furthermore, in response to **RQ2**, this approach enables standardisation across varying EEG channel configurations by transforming heterogeneous microstate sequences into a unified embedding space. This allows for the development of an adaptive AI model having high performance across different EEG setups, enhancing its generalisability and clinical applicability.

For **RQ3**, statistical analyses revealed that over two-thirds of the embedding features exhibited significant differences ($p < 0.05$) between NC and AD groups across multiple time segments and distance metrics. Notably, a consistent subset of features remained highly significant ($p < 0.001$), indicating that the vector representations derived from EEG microstates effectively capture meaningful and discriminative patterns associated with AD.

Future work will focus on addressing current limitations by expanding evaluation across larger and more diverse populations, assessing fairness across demographic groups, improving model explainability, and optimising performance for shorter EEG recordings to support real-world use. Additionally, efforts will be made to reduce dependency on third-party APIs to enhance transparency, reproducibility, and facilitate local deployment.

References

Tassallah Abdullahi, Ritambhara Singh, and Carsten Eickhoff. 2024. Retrieval augmented zero-shot text classification. In *Proceedings of the 2024 ACM SIGIR international conference on theory of information retrieval*, pages 195–203.

Abdulyekeen T Adebisi, Ho-Won Lee, and Kalyana C Veluvolu. 2024. EEG-based brain functional network analysis for differential identification of dementia-related disorders and their onset. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Michele Alessandrini, Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, Simona Luzzi, and Claudio Turchetti. 2022. EEG-based alzheimer's disease recognition using robust-pca and lstm recurrent neural network. *Sensors*, 22(10):3696.

Fatih Altay, Guillermo Ramón Sánchez, Yanli James, Stephen V Faraone, Senem Velipasalar, and Asif Salekin. 2021. Preclinical stage alzheimer's disease detection using magnetic resonance image scans. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15088–15097.

Bagdasarov Armen and 1 others. 2022. Spatiotemporal dynamics of EEG microstates in four-to eight-year-old children: Age-and sex-related effects. *Developmental cognitive neuroscience*, 57:101134.

Jun Cao, Lichao Yang, Ptolemaios Georgios Sarrigiannis, Daniel Blackburn, and Yifan Zhao. 2024. Dementia classification using a graph neural network on imaging of effective brain connectivity. *Computers in Biology and Medicine*, 168:107701.

Chu Chunguang and 1 others. 2022. An enhanced EEG microstate recognition framework based on deep neural networks: an application to parkinson's disease. *IEEE Journal of Biomedical and Health Informatics*, 27(3):1307–1318.

Maxime Darrin, Philippe Formont, Ismail Ayed, Jackie CK Cheung, and Pablo Piantanida. 2024. When is an embedding model more promising than another? *Advances in Neural Information Processing Systems (NeurIPS)*, 37:68330–68379.

Zijian Dong, Ruilin Li, Yilei Wu, Thuan Tinh Nguyen, Joanna Chong, Fang Ji, Nathanael Tong, Christopher Chen, and Juan Helen Zhou. 2024. Brain-jepa: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:86048–86073.

Bruno Dubois, Alessandro Padovani, Philip Scheltens, Andrea Rossi, and Grazia Dell'Agnello. 2016. Timely diagnosis for alzheimer's disease: a literature review on benefits and challenges. *Journal of Alzheimer's disease*, 49(3):617–631.

Patrycja Dzianok and Ewa Kublik. 2024. Pearl-neuro database: EEG, fmri, health and lifestyle data of middle-aged people at risk of dementia. *Scientific Data*, 11(1):276.

Kenneth Enevoldsen, Márton Kardos, Niklas Muennighoff, and Kristoffer L Nielbo. 2024. The scandinavian embedding benchmarks: Comprehensive assessment of multilingual and monolingual text embedding. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:40336–40358.

Xiachong Feng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2023. Aligning semantic in brain and language: A curriculum contrastive method for electroencephalography-to-text generation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:3874–3883.

Victor Férat, Mathieu Scheltienne, Denis Brunet, Tomas Ros, and Christoph Michel. 2022. Pycrostates: a python library to study EEG microstates. *Journal of Open Source Software*, 7(78):4564.

Gogishvili and others. 2023. Discovery of novel csf biomarkers to predict progression in dementia using machine learning. *Scientific reports*, 13(1):6531.

Víctor Gutiérrez-de Pablo, Jesús Poza, Aarón Maturana-Candelas, Víctor Rodríguez-González, Miguel Ángel Tola-Arribas, Mónica Cano, Hideyuki Hoshi, Yoshihito Shigihara, Roberto Hornero, and Carlos Gómez. 2024. Exploring the disruptions of the neurophysiological organization in alzheimer's disease: An integrative approach. *Computer Methods and Programs in Biomedicine*, 250:108197.

Masahiro Hata, Yusuke Watanabe, Takumi Tanaka, Noriyuki Awata, and 1 others. 2023. Precise discrimination for multiple etiologies of dementia cases based on deep learning with electroencephalography. *Neuropsychobiology*, 82(2):81–90.

David Haydock, Shabnam Kadir, Robert Leech, Chrystopher L Nehaniv, and Elena Antonova. 2025. EEG microstate syntax analysis: A review of methodological challenges and advances. *NeuroImage*, page 121090.

Yongquan Hu, Shuning Zhang, Ting Dang, Hong Jia, Flora D Salim, Wen Hu, and Aaron J Quigley. 2024. Exploring large-scale language models to evaluate EEG-based multimodal data for mental health. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP)*, pages 412–417.

Yuya Ikegawa, Ryohei Fukuma, Hidenori Sugano, Satoru Oshino, Naoki Tani, Kentaro Tamura, Yasushi Iimura, Hiroharu Suzuki, Shota Yamamoto, Yuya Fujita, and 1 others. 2024. Text and image generation from intracranial electroencephalography using an embedding space for text and images. *Journal of Neural Engineering*, 21(3):036019.

Ming Jin, Yifan Zhang, Wei Chen, Kexin Zhang, Yuxuan Liang, Bin Yang, Jindong Wang, Shirui Pan, and Qingsong Wen. 2024. Position: What can large language models tell us about time series analysis. In *Forty-first International Conference on Machine Learning (ICML)*.

Sai Sushanth Varma Kalidindi, Hadi Banee, Hans Karlsson, and Amy Loutfi. 2024. Adaptive context embedding for temperature prediction in residential buildings. In *ECAI 2024*, pages 4727–4733. IOS Press.

Rachneet Kaur, Zhen Zeng, Tucker Balch, and Manuela Veloso. 2024. Lets-c: Leveraging language embedding for time series classification. *arXiv preprint arXiv:2407.06533*.

Arjun Khanna, Alvaro Pascual-Leone, and Faranak Farzan. 2014. Reliability of resting-state microstate features in electroencephalography. *PloS one*, 9(12):e114163.

Ann-Kathrin Kiessner, Robin T Schirrmeister, Lukas AW Gemein, Joschka Boedecker, and Tonio Ball. 2023. An extended clinical EEG dataset with 15,300 automatically labelled recordings for pathology decoding. *NeuroImage: Clinical*, 39:103482.

Min-jae Kim, Young Chul Youn, and Joonki Paik. 2023. Deep learning-based EEG analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset. *NeuroImage*, 272:120054.

Dominik Klepl, Fei He, Min Wu, Daniel J Blackburn, and Ptolemaios Sarrigiannis. 2022. EEG-based graph neural network classification of alzheimer's disease: An empirical evaluation of functional connectivity methods. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2651–2660.

Dominik Klepl, Fei He, Min Wu, Daniel J Blackburn, and Ptolemaios Sarrigiannis. 2023. Adaptive gated graph convolutional network for explainable diagnosis of alzheimer's disease using EEG data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Utkarsh Lal, Arjun Vinayak Chikkankod, and Luca Longo. 2024. A comparative study on feature extraction techniques for the discrimination of frontotemporal dementia and alzheimer's disease with electroencephalography in resting-state adults. *Brain Sciences*, 14(4):335.

Michael Lassi, Carlo Fabbiani, Salvatore Mazzeo, Rachele Burali, Alberto Arturo Vergani, Giulia Giacomucci, Valentina Moschini, Carmen Morinelli, Filippo Emiliani, Maenia Scarpino, and 1 others. 2023. Degradation of EEG microstates patterns in subjective cognitive decline and mild cognitive impairment: Early biomarkers along the alzheimer's disease continuum? *NeuroImage: Clinical*, 38:103407.

Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. 2024. Autotimes: Autoregressive time series forecasters via large language models. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:122154–122184.

Andreas Miltiadous, Emmanouil Gionanidis, Katerina D Tzimourta, Nikolaos Giannakeas, and Alexandros T Tzallas. 2023a. Dice-net: a novel convolution-transformer architecture for alzheimer detection in EEG signals. *IEEE Access*, 11:71840–71858.

Andreas Miltiadous, Katerina D Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G Tsalikakis, Pantelis Angelidis, Markos G Tsipouras, Euripidis Glavas, Nikolaos Giannakeas, and 1 others. 2023b. A dataset of scalp EEG recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine eeg. *Data*, 8(6):95.

Ashutosh Mishra, Bernhard Englitz, and Michael X Cohen. 2020. EEG microstates as a continuous phenomenon. *Neuroimage*, 208:116454.

Sebastian G Moguilner, Courtney Berezuk, Alex C Bender, Kyle R Pellerin, Stephen N Gomperts, Sydney S Cash, Rani A Sarkis, and Alice D Lam. 2024. Sleep functional connectivity, hyperexcitability, and cognition in alzheimer's disease. *Alzheimer's & Dementia*, 20(6):4234–4249.

Navid Mohammadi Foumani, Geoffrey Mackellar, Soheila Ghane, Saad Irtza, Nam Nguyen, and Mahsa Salehi. 2024a. Eeg2rep: enhancing self-supervised EEG representation through informative masked inputs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5544–5555.

Navid Mohammadi Foumani, Lynn Miller, Chang Wei Tan, Geoffrey I Webb, Germain Forestier, and Mahsa Salehi. 2024b. Deep learning for time series classification and extrinsic regression: A current survey. *ACM Computing Surveys*, 56(9):1–45.

Quoc-Toan Nguyen. 2024. Advancing Early Alzheimer's Disease Detection in Underdeveloped Areas with Fair Explainable AI Methods. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 47–49.

Quoc-Toan Nguyen. 2025a. Echo-GRU: Emotion Recognition Using Wearable EEG Supporting Early Alzheimer's Disease Detection. In *International Conference on Pattern Recognition*, pages 3–17. Springer.

Quoc-Toan Nguyen. 2025b. Standardising Number of EEG Sensors for AI-Driven Dementia Detection. *IEEE Sensors Letters*.

Quoc-Toan Nguyen and Nghia Duong-Trung. 2025. Predicting progression from mild cognitive impairment to alzheimer's using an ai-based multimodal

approach. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 216–228. Springer.

Quoc-Toan Nguyen, Zheng Huiru, Tahia Tazin, Linh Le, Tuan L Vo, Nhu-Tri Tran, David Williams-King, and Benjamin Tag. 2025. Emotion Recognition Using Text Embedding Models: Wearable and Wireless EEG Without Fixed EEG Channel Configurations. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*, pages 476–488.

Quoc-Toan Nguyen, Linh Le, Xuan-The Tran, Thomas Do, and Chin-Teng Lin. 2024. Fairad-xai: Evaluation framework for explainable ai methods in alzheimer's disease detection with fairness-in-the-loop. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 870–876.

Majid Nour, Umit Senturk, and Kemal Polat. 2024. A novel hybrid model in the diagnosis and classification of alzheimer's disease using EEG signals: Deep ensemble learning (del) approach. *Biomedical Signal Processing and Control*, 89:105751.

Zaixin Ou, Caiwen Jiang, Yongsheng Pan, Yuanwang Zhang, Zhiming Cui, and Dinggang Shen. 2024. A prior-information-guided residual diffusion model for multi-modal pet synthesis from mri. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4769–4777.

Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua Dillon, Balaji Lakshminarayanan, and Jasper Snoek. 2019. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *Advances in neural information processing systems (NeurIPS)*, 32.

Roberto D Pascual-Marqui, Christoph M Michel, and Dietrich Lehmann. 1995. Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Transactions on Biomedical Engineering*, 42(7):658–665.

Prado Pavel and 1 others. 2023. The brainlat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds. *Scientific Data*, 10(1):889.

Digambar V Puri. 2024. Leadnet: detection of alzheimer's disease using spatiotemporal EEG analysis and low-complexity cnn. *IEEE Access*.

Digambar V Puri, Sanjay L Nalbalwar, Anil B Nandgaonkar, Jayanand P Gawande, and Abhay Wagh. 2023. Automatic detection of alzheimer's disease from EEG signals using low-complexity orthogonal wavelet filter banks. *Biomedical Signal Processing and Control*, 81:104439.

Oona Rainio, Jarmo Teuho, and Riku Klén. 2024. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1):6086.

VP Subramanyam Rallabandi and Krishnamoorthy Seetharaman. 2023. Deep learning-based classification of healthy aging controls, mild cognitive impairment and alzheimer's disease using fusion of mri-pet imaging. *Biomedical Signal Processing and Control*, 80:104312.

Khosro Rezaee and Min Zhu. 2025. Diagnose alzheimer's disease and mild cognitive impairment using deep cascadenet and handcrafted features from EEG signals. *Biomedical Signal Processing and Control*, 99:106895.

Eikelboom Willem S and 1 others. 2019. Early recognition and treatment of neuropsychiatric symptoms to improve quality of life in early alzheimer's disease: Protocol of the beat-it study. *Alzheimer's research & therapy*, 11:1–12.

Asha SA, Subodh PS, Arya ML, Devika Kumar, Sanjeev V Thomas, Ramshekhar N Menon, and 1 others. 2024. Resting state EEG microstate profiling and a machine-learning based classifier model in epilepsy. *Cognitive Neurodynamics*, pages 1–14.

Sena Yagmur Sen, Ozlem Karabiber Cura, and Aydin Akan. 2023. Classification of dementia EEG signals by using time-frequency images for deep learning. In *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–6. IEEE.

Gulshan Sharma, Surbhi Madan, Maneesh Bilalpur, Abhinav Dhall, and Ramanathan Subramanian. 2025. EEG-based cognitive load estimation of acoustic parameters for data sonification. *IEEE Transactions on Cognitive and Developmental Systems*.

Yajing Si, Runyang He, Lin Jiang, Dezhong Yao, Hongxing Zhang, Peng Xu, Xuntai Ma, Liang Yu, and Fali Li. 2023. Differentiating between alzheimer's disease and frontotemporal dementia based on the resting-state multilayer EEG network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:4521–4527.

Una Smailovic, Thomas Koenig, Erika J Laukka, Grégoria Kalpouzos, Thomas Andersson, Bengt Winblad, and Vesna Jelic. 2019. EEG time signature in alzheimer s disease: functional brain networks falling apart. *NeuroImage: Clinical*, 24:102046.

Konstantinos Stefanou, Katerina D Tzimourta, Christos Bellos, Georgios Stergios, Konstantinos Markoglou, Emmanouil Gionanidis, Markos G Tsipouras, Nikolaos Giannakeas, Alexandros T Tzallas, and Andreas Miltiadous. 2025. A novel cnn-based framework for alzheimer's disease detection using EEG spectrogram representations. *Journal of Personalized Medicine*, 15(1):27.

Mingtian Tan, Mike Merrill, Vinayak Gupta, Tim Althoff, and Tom Hartvigsen. 2024. Are language models actually useful for time series forecasting? *Advances in Neural Information Processing Systems (NeurIPS)*, 37:60162–60191.

Povilas Tarailis, Thomas Koenig, Christoph M Michel, and Inga Griškova-Bulanova. 2024. The functional aspects of resting EEG microstates: a systematic review. *Brain topography*, 37(2):181–217.

The Alzheimer's Association. 2023. 2023 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 19(4):1598–1695.

Koenig Thomas and 1 others. 2011. Ragu: a free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational intelligence and neuroscience*, 2011:1–14.

Xuan-The Tran, Quoc-Toan Nguyen, Linh Le, Thomas Do, and Chin-Teng Lin. 2024. EEG-Based Contrastive Learning Models For Object Perception Using Multisensory Image-Audio Stimuli. In *Proceedings of the 1st International Workshop on Brain-Computer Interfaces (BCI) for Multimedia Understanding*, pages 39–47.

Pedro A Valdes-Sosa. 2021. The cuban human brain mapping project, a young and middle age population-based EEG, mri, and cognition dataset. *Scientific data*, 8(1):45.

Wiesje M van der Flier, Marjolein E de Vugt, Ellen MA Smets, Marco Blom, and Charlotte E Teunissen. 2023. Towards a future where alzheimer's disease pathology is stopped before the onset of dementia. *Nature aging*, 3(5):494–505.

Jiaqi Wang, Zhenxi Song, Zhengyu Ma, Xipeng Qiu, Min Zhang, and Zhiguo Zhang. 2024a. Enhancing EEG-to-text decoding through transferable representations from pre-trained contrastive EEG-text masked autoencoder. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 7278–7292.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024b. Improving text embeddings with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 11897–11916.

Zhenhailong Wang and Heng Ji. 2022. Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5350–5358.

World Health Organization. 2023. Dementia. *World Health Organization*. Accessed: 2024-08-19. Available at: https://www.who.int/news-room/fact-sheets/detail/dementia.

Xiaoli Yang, Zhipeng Fan, Zhenwei Li, and Jiayi Zhou. 2024. Resting-state EEG microstate features for alzheimer's disease classification. *PloS one*, 19(12):e0311958.

Zhao Yifan and 1 others. 2019. Imaging of nonlinear and dynamic functional brain connectivity based on EEG recordings with the application on the diagnosis of alzheimer's disease. *IEEE transactions on medical imaging*, 39(5):1571–1581.

Xiyuan Zhang, Ranak Roy Chowdhury, Rajesh K Gupta, and Jingbo Shang. 2024. Large language models for time series: a survey. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 8335–8343.

Nicolas Zucchet and Antonio Orvieto. 2024. Recurrent neural networks: vanishing and exploding gradients are not the end of the story. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:139402–139443.

A Additional Dataset Information

Table 3 summarises the datasets used in this study, comprising a total of 1001 individuals collected from multiple countries: Republic of Korea (Kim et al., 2023), Poland (Dzianok and Kublik, 2024), Greece (Miltiadous et al., 2023b), Cuba (Valdes-Sosa, 2021), Argentina, Chile, Colombia, Mexico, and Peru (Pavel et al., 2023), and the USA (Kiessner et al., 2023). All included datasets in this paper are publicly available, and ethical approvals were obtained by the respective original data providers following proper regulations and institutional review boards. All data were fully anonymised before public release, ensuring no personally identifiable information was accessible. The reuse of these datasets complies with open science policies and legal data-sharing frameworks. Furthermore, no sensitive information was transmitted through external APIs used for model inference, as only preprocessed, anonymised features were utilised.

Table 3: Summary of included EEG datasets. NC: Number of Normal Control individuals. AD: Number of Alzheimer's Disease individuals.

Dataset	Channels	NC	AD
CAUEEG (Kim et al., 2023)	19	0	230
PEARL-Neuro (Dzianok and Kublik, 2024)	128	69	0
DS004504 (Miltiadous et al., 2023b)	19	29	29
CHBMP (Valdes-Sosa, 2021)	64	19	0
BrainLat (Pavel et al., 2023)	128	30	27
TUAB (Kiessner et al., 2023)	23	568	0

B Additional Model Results

Figure 7 illustrates the performance of the proposed method across various text embedding models, embedding sizes, and chunk sizes. Table 4 summarises prominent existing studies on AI-based EEG approaches for AD detection.

C Details of Pattern Analysis

As the best-performing model was achieved using embeddings from text-embedding-3-small, the corresponding data with an embedding size of 32 was selected for all subsequent analyses. Figure 2 illustrates the feature distribution of NC and AD groups based on raw absolute differences, while Figure 3 presents the feature-wise distances between their embedded vector representations.

To evaluate the statistical significance of feature differences between NC and AD groups, we employed two non-parametric tests (Ikegawa et al., 2024): the Mann–Whitney U test and the Kruskal–Wallis test. These tests were selected because they do not assume normal distribution of the data, an important consideration given the complex and potentially non-Gaussian nature of EEG-derived features. The Mann–Whitney U test assesses whether the distributions of a single feature differ significantly between two independent groups (NC vs. AD) without assuming normality. It was applied across each embedding feature and time segment, as well as across different distance metrics, to detect fine-grained inter-group differences (see Tables 6 and 7). In parallel, the Kruskal–Wallis test, a generalisation of the Mann–Whitney test for comparing more than two groups, was used to examine intra-group variability across the five one-minute EEG segments within each class (NC and AD) (see Table 8). These tests enabled robust identification of embedding features that consistently exhibit statistically significant discriminative power, both across groups and within temporal dynamics.

D Acknowledgements

This work was supported in part by the Australian Research Council (ARC) under Grants DP220100803 and DP250103612; in part by the Industrial Transformation Research Hub (ITRH) under Grant IH240100016; in part by the Australian National Health and Medical Research Council (NHMRC) Ideas Grant APP2021183; in part by the UTS Human-Centric AI Centre through GrapheneX (2023–2031); in part by the Australian Defence Innovation Hub under Grant P18-650825; and in part by the Australian Defence Science and Technology Group (DSTG) under Grant 12549.

E Author Contributions

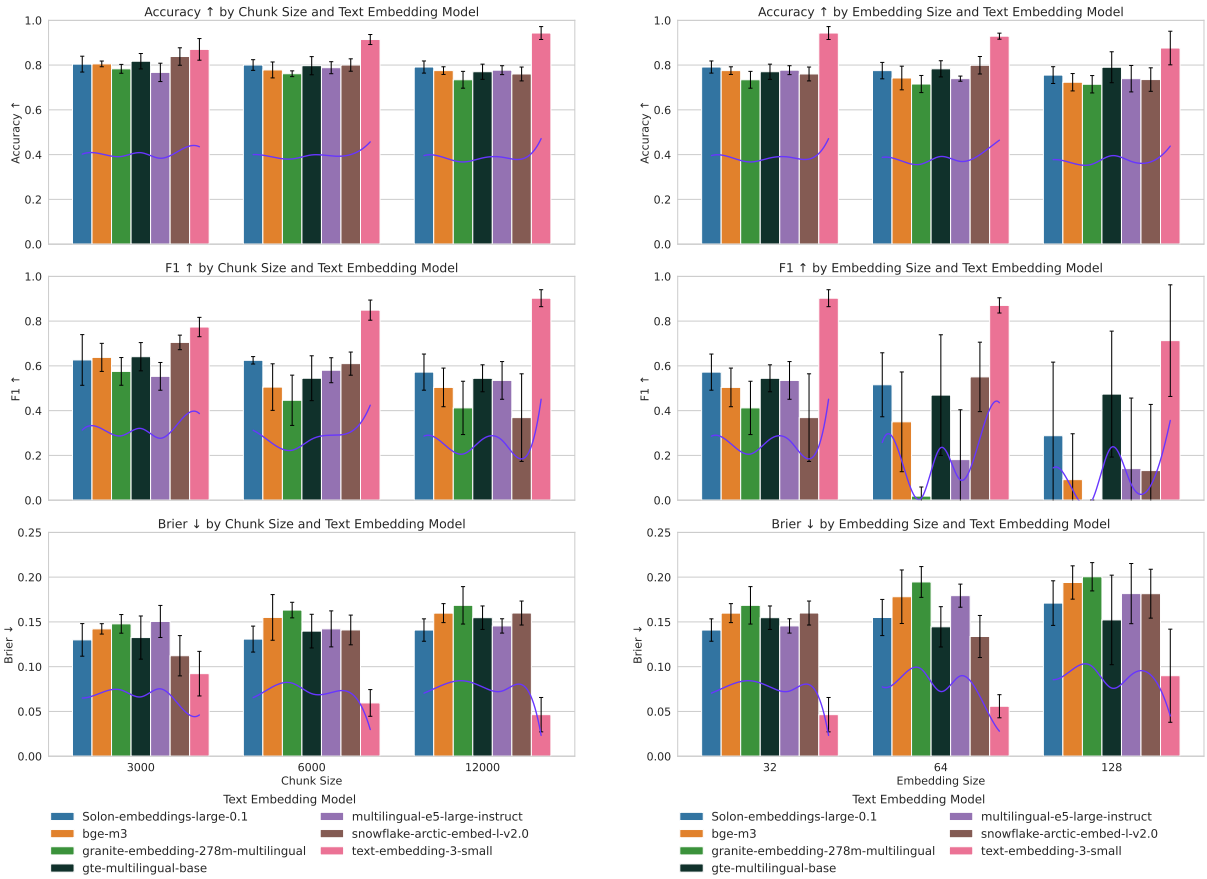
Q.T.N. proposed the methods, developed the code, analysed the data, visualised the figures, and wrote the main parts of the manuscript. X.T.T. contributed to manuscript writing. D.B. provided supervision on the medical aspects of the proposed methods. L.L. and N.D.T. contributed to ensuring the validation of the AI components. T.D. and C.T.L., as project investigators, supervised the overall project.

Table 4: Performance comparison between the proposed method and prominent research on AI-based EEG approaches for Alzheimer's Disease (AD) detection. NC: Normal Control.

Method	Channel	Participant (NC/AD)	Accuracy
Ours	19, 23, 64, 128	715 / 286	0.9431
MNet (Hata et al., 2023)	19	55 / 101	0.8170
LCOWFBs-6 (Puri et al., 2023)	16	11 / 12	0.9860
k-NN (Yifan et al., 2019)	19	20 / 20	0.9000
CNN (Stefanou et al., 2025)	128	29 / 36	0.7945
DEL (Nour et al., 2024)	19	36 / 104	0.9790
DICE-Net (Miltiadous et al., 2023a)	19	29 / 36	0.8328
GNN (Klepl et al., 2022)	128	20 / 20	0.9200
GNB (Si et al., 2023)	128	19 / 36	0.8100
DSL-GN (Cao et al., 2024)	23	20 / 20	0.9400
LEADNet (Puri, 2024)	16	11 / 12	0.9924
LSTM (Alessandrini et al., 2022)	16	15 / 20	0.9790
k-NN (Lal et al., 2024)	19	29 / 36	0.9300
CNN (Sen et al., 2023)	19	11 / 15	0.9860

Table 5: Confidence summary by folds between Normal Control (NC) and Alzheimer's Disease (AD) groups with p-values of the Mann-Whitney U test. ✓: Accurately classified, ✗: Inaccurately classified.

Fold	Total Sample		✓ Sample		Confidence Score (✓)			Confidence Score (✗)		
	NC	AD	NC	AD	NC	AD	p-value	NC	AD	p-value
1	140	61	131	56	0.957 ± 0.082	0.882 ± 0.122	<0.001	0.759 ± 0.119	0.788 ± 0.178	0.699
2	146	54	145	50	0.971 ± 0.080	0.938 ± 0.095	<0.001	0.743 ± 0.115	0.925 ± 0.063	0.400
3	148	52	144	47	0.977 ± 0.069	0.955 ± 0.080	<0.001	0.796 ± 0.196	0.814 ± 0.136	0.904
4	131	69	124	56	0.953 ± 0.095	0.916 ± 0.089	<0.001	0.707 ± 0.121	0.832 ± 0.158	<0.05
5	150	50	146	45	0.958 ± 0.096	0.898 ± 0.115	<0.001	0.699 ± 0.086	0.718 ± 0.063	0.904



(a) Result of different chunk sizes.

(b) Result of different embedding sizes.

Figure 7: Visualisation of results of text embedding models for Alzheimer's Disease (AD) detection using EEG microstates.

Table 6: Results with p-values of Mann–Whitney U test by raw feature values and time step between Normal Control (NC) and Alzheimer's Disease (AD).

Embedding Feature	Minute 1	Minute 2	Minute 3	Minute 4	Minute 5	All
1	<0.05	0.233	<0.01	0.162	0.713	<0.001
2	<0.001	<0.001	<0.001	<0.001	0.809	<0.001
3	<0.01	<0.01	<0.001	<0.05	<0.001	<0.001
4	0.051	0.665	0.096	<0.05	<0.001	0.15
5	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
6	<0.05	<0.001	<0.001	<0.001	<0.001	<0.001
7	<0.001	<0.01	<0.01	<0.05	<0.001	<0.001
8	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
9	0.450	0.732	0.063	0.374	0.213	0.97
10	<0.01	<0.001	<0.001	<0.001	<0.001	<0.001
11	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
12	0.279	0.119	<0.05	0.150	<0.001	<0.01
13	<0.05	<0.001	<0.001	<0.01	<0.001	<0.001
14	<0.001	<0.01	<0.001	<0.01	<0.001	<0.001
15	<0.01	0.297	0.484	0.165	<0.001	<0.001
16	0.068	0.648	<0.05	0.654	<0.001	<0.001
17	0.140	0.649	0.058	0.765	<0.001	<0.001
18	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
19	0.740	0.108	<0.05	<0.01	<0.001	<0.001
20	<0.01	<0.01	<0.01	<0.001	<0.01	<0.001
21	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001
22	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
23	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001
24	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
25	<0.001	<0.05	<0.01	0.098	<0.001	<0.001
26	<0.001	0.506	0.080	0.691	<0.001	<0.001
27	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
28	0.192	0.104	<0.05	0.237	<0.001	<0.001
29	<0.01	<0.05	0.300	0.566	<0.001	<0.001
30	<0.001	<0.01	<0.001	<0.001	<0.001	<0.001
31	<0.05	0.849	0.995	0.078	<0.001	<0.05
32	0.225	<0.01	<0.001	0.051	<0.01	<0.01

Table 7: Results with p-values of Mann–Whitney U test by raw feature values with types of distance for each feature between Normal Control (NC) and Alzheimer's Disease (AD).

Embedding Feature	Euclidean	Cosine	Manhattan
1	<0.001	0.39	<0.001
2	<0.001	0.71	<0.001
3	<0.001	<0.001	<0.001
4	<0.05	0.83	<0.01
5	0.38	<0.001	0.31
6	<0.001	<0.01	<0.001
7	<0.01	<0.05	<0.05
8	<0.001	<0.001	<0.001
9	<0.001	0.43	<0.001
10	<0.001	<0.001	<0.001
11	<0.001	0.05	<0.001
12	<0.001	0.87	<0.001
13	<0.001	<0.001	<0.001
14	0.76	<0.01	0.99
15	0.64	0.82	0.24
16	<0.001	<0.001	<0.001
17	0.62	0.07	0.80
18	<0.05	<0.001	<0.01
19	<0.001	<0.001	<0.001
20	<0.001	0.12	<0.001
21	<0.001	<0.001	<0.001
22	<0.001	<0.001	<0.001
23	<0.001	<0.01	<0.001
24	<0.001	<0.01	<0.001
25	<0.001	<0.001	<0.001
26	0.25	<0.001	0.15
27	<0.05	<0.01	<0.01
28	<0.001	0.62	<0.001
29	<0.001	<0.01	<0.001
30	0.43	<0.05	0.36
31	<0.001	<0.001	<0.001
32	<0.001	0.29	<0.001

Table 8: Results with p-values of Kruskal–Wallis by raw feature values across all five minutes between Normal Control (NC) and Alzheimer's Disease (AD).

Embedding Feature	NC	AD
1	0.190	0.961
2	<0.001	0.251
3	<0.001	0.150
4	<0.001	0.284
5	<0.001	<0.01
6	<0.001	<0.01
7	<0.001	<0.01
8	0.265	<0.001
9	0.332	0.051
10	<0.001	0.796
11	<0.05	<0.01
12	<0.001	<0.001
13	<0.001	<0.001
14	0.117	<0.001
15	<0.001	<0.001
16	<0.001	0.266
17	<0.001	<0.001
18	<0.001	<0.001
19	<0.01	0.248
20	0.587	<0.05
21	<0.05	<0.001
22	<0.001	<0.05
23	<0.001	<0.001
24	<0.001	0.079
25	<0.001	0.319
26	<0.001	0.314
27	<0.001	<0.01
28	<0.001	<0.01
29	<0.001	<0.001
30	0.188	<0.001
31	<0.001	<0.001
32	<0.001	<0.001