VINEURO: A MULTIMODAL EEG-BLOOD FUSION MODEL FOR ALZHEIMER'S RISK PREDICTION

Thanh Trung Nguyen[1,*]

Abstract

Early prediction of Alzheimer's disease risk is crucial for timely intervention but remains challenging in routine clinical practice. Electroencephalography (EEG) is inexpensive and non-invasive, yet EEG alone often lacks sufficient sensitivity and robustness for reliable early-stage risk estimation. In parallel, routine blood tests capture peripheral immune, inflammatory, and metabolic changes associated with cognitive decline, suggesting that combining EEG with blood-based biomarkers could yield more informative risk stratification. In this work, ViNeuro, a multimodal EEG–blood model tailored to Alzheimer's risk prediction, is proposed. A single EEG foundation encoder, termed ViNeuro-EEG, is first pretrained using the dual self-supervised objective of the EEGPT model with the criss-cross backbone and learned positional encoding of the CBraMOD model. Pretraining is conducted on a unified corpus of multi-channel clinical EEG data that includes Vietnamese recordings from 108 Military Central Hospital and international datasets. On top of this encoder, a multimodal extension, ViNeuro-MM, is constructed by projecting routine blood biomarkers into the EEG embedding space and using them as queries in a cross-attention layer over EEG tokens. The proposed framework is evaluated on the PEARL-Neuro cohort for Alzheimer's risk prediction. Compared to its EEG-only counterpart, ViNeuro-MM achieves substantial performance gains, with relative improvements of up to 24.72% in balanced accuracy, demonstrating that fusing routine blood-based biomarkers with EEG foundation representations can markedly enhance early Alzheimer's risk prediction.

Index terms

Clinical EEG; foundation models; Alzheimer's risk prediction; representation learning; self-supervised learning; multimodal learning; blood.

1. Introduction

Alzheimer's disease is a leading cause of dementia worldwide and remains incurable once clinical symptoms are fully manifest. As a result, there is a strong clinical need to identify individuals at elevated risk early, when lifestyle interventions,

[1] Medical Equipment Department, 108 Military Central Hospital

[*] Corresponding author. email: trung.ntc10@benhvien108.vn

DOI: 10.56651/lqdtu.jst.v14.n02.1112.ict

closer monitoring, or enrollment in preventive trials may still alter the course of disease. Current risk assessment often relies on neuropsychological testing, structural or functional imaging, and sometimes cerebrospinal fluid biomarkers. While informative, these modalities can be invasive, expensive, or limited in availability, especially in resource-constrained healthcare systems.

Electroencephalography is an attractive complementary modality for this purpose. It is non-invasive, relatively inexpensive, and widely available, and it directly measures neural activity with high temporal resolution. Numerous studies have reported alterations in EEG rhythms and connectivity patterns in individuals with mild cognitive impairment and early Alzheimer's disease [1]. However, raw EEG is noisy, highly variable across subjects and recording sites, and difficult to interpret reliably. Traditional machine learning models trained on handcrafted features often fail to generalize beyond narrow experimental settings, limiting their impact on routine Alzheimer's risk assessment. In parallel, EEG foundation models have emerged as a promising paradigm for learning robust EEG representations. By training large encoders on diverse EEG corpora using self-supervised objectives, methods such as EEGPT model [2] and CBraMOD model [3] aim to capture generic structure in EEG signals that transfers across tasks and populations. EEGPT model leverages a dual masked autoencoder objective to stabilize representations, while CBraMOD model introduces criss-cross attention and learned 2D positional encodings tailored to the channel-time grid. Despite encouraging results, existing foundation models are largely unimodal, focusing on EEG alone, even though early Alzheimer's risk in practice is assessed in conjunction with other clinical signals.

Among these additional signals, routine blood-based biomarkers are particularly compelling. Clinical and epidemiological studies have shown that individuals at higher risk of dementia exhibit characteristic shifts in peripheral immune, inflammatory, and metabolic profiles, many of which are captured by standard complete blood counts and biochemistry panels [4]. These tests are already part of everyday clinical workflows and provide a low-cost, minimally invasive complement to EEG. From a modeling standpoint, blood biomarkers and EEG offer complementary views: blood reflects systemic and neuroinflammatory processes, whereas EEG reflects real-time brain dynamics. A principled multimodal model that fuses these signals could therefore enhance Alzheimer's risk prediction beyond what either modality can achieve alone.

In this paper, ViNeuro, a multimodal EEG–blood fusion model specifically designed for Alzheimer's disease risk prediction, is introduced. A single EEG foundation encoder, termed ViNeuro-EEG, is first pretrained using the dual self-supervised objective of the EEGPT model, combined with CBraMOD's criss-cross backbone and learned positional encoding. Pretraining is carried out on a unified corpus of multi-channel clinical scalp EEG that includes Vietnamese data from 108 Military Central Hospital and international datasets such as TUEG [5] and NMT-Scalp [6]. The corpus is used exclusively for pretraining purposes, and no dataset contribution is claimed. A multimodal variant, termed ViNeuro-MM, is constructed based on ViNeuro-EEG, in which normalized blood-based biomarkers are projected into the EEG embedding space and utilized as queries in a cross-attention layer over EEG tokens.

This mechanism produces a fused representation in which the contribution of EEG patterns is explicitly conditioned on the patient's peripheral blood profile. The proposed framework is evaluated on the PEARL-Neuro [4] cohort, which provides 64-channel EEG recordings, routine blood test measurements, and dementia risk labels for middle-aged individuals. Following prior work, three experimental paradigms are considered: the Multi-Source Interference Task (MSIT), the Sternberg Memory Task (SMT), and resting-state EEG. Across all three paradigms, the unimodal ViNeuro-EEG is compared with the multimodal ViNeuro-MM, and ViNeuro-MM is further benchmarked against multimodal baselines derived from the CBraMOD model and EEGPT models. Experimental results indicate that incorporating blood-based biomarkers through ViNeuro-MM leads to substantial and consistent improvements in balanced accuracy, area under the precision–recall curve (AUPR), and area under the receiver operating characteristic curve (AUROC), with the largest relative gains observed in the resting-state condition, where EEG abnormalities are often subtle.

In summary, the main contributions of this work are as follows:

- *An EEG foundation encoder pretrained on clinical multi-channel EEG.* A single encoder, ViNeuro-EEG, is trained on a unified corpus of Vietnamese and international clinical EEG data, including routine multi-channel recordings from 108 Military Central Hospital, using a dual self-supervised objective that combines the EEGPT training framework with the CBraMOD model criss-cross architecture and learned positional encoding.
- *A lightweight EEG–blood fusion model for Alzheimer's risk prediction.* A multimodal model, ViNeuro-MM, is proposed to integrate routine blood-based biomarkers with EEG representations via a cross-attention mechanism, enabling adaptive emphasis on EEG patterns most relevant to a patient's peripheral blood profile.
- *Extensive evaluation on the PEARL-Neuro cohort.* Experiments conducted on PEARL-Neuro demonstrate that ViNeuro-MM consistently outperforms its EEG-only counterpart and is competitive with or superior to strong multimodal baselines derived from CBraMOD model and EEGPT model across the MSIT, SMT, and resting-state paradigms, highlighting the effectiveness of multimodal EEG–blood modeling for early Alzheimer's disease risk prediction.

2. Preliminaries and related work

EEG foundation models are large neural networks trained on massive EEG datasets to learn general, reusable patterns of brain activity, which are then adapted to specific tasks. Rather than building a new model from scratch for each application, a foundation model is first exposed to diverse EEG signals, often without requiring manual labels, so that it automatically discovers features capturing common rhythms, artifacts, and disease-related signatures. For clinical users, these models can be regarded as pretrained EEG decision-support systems that have been exposed to a broad spectrum of recordings, enabling efficient adaptation to specific diagnostic tasks with reduced annotation requirements and improved robustness.

Among current open-source efforts, EEGPT model [2] and CBraMOD model [3] stand out as two of the most mature and influential EEG foundation models. Both are trained at scale, released with reproducible implementations, and explicitly positioned as general backbones for downstream EEG applications. In this work, the EEGPT and CBraMOD models are employed as backbone EEG encoders within a multimodal EEG–blood framework for Alzheimer's risk prediction.

2.1. Masked autoencoders for EEG foundation models

Let $x \in \mathbb{R}^{C \times T}$ denote an EEG segment with C channels and T time points. First, divide x into non-overlapping time windows of length τ, yielding a tensor in Eq. 1.

$$X \in \mathbb{R}^{C \times N \times \tau}, \quad N = \left\lfloor \frac{T}{\tau} \right\rfloor, \tag{1}$$

where, $X = \{x_{i,j} \mid i \in \{1, \dots, C\}, j \in \{1, \dots, N\}\}$ and each $x_{i,j} \in \mathbb{R}^{\tau}$ is a local patch.

A binary mask $M = \{m_{i,j}\} \in \{0,1\}^{C \times N}$ is sampled (typically from a Bernoulli distribution with a fixed mask ratio). The masked input $\tilde{X} = \{\tilde{x}_{i,j}\}$ is obtained by replacing selected patches with a learnable (or fixed) mask token $x_M \in \mathbb{R}^{\tau}$ in Eq. 2

$$\tilde{x}_{i,j} = \begin{cases} x_{i,j}, & m_{i,j} = 0, \\ x_M, & m_{i,j} = 1. \end{cases} \tag{2}$$

In a generic masked autoencoder (MAE), an encoder f_θ processes the masked input and a decoder d_ϕ reconstructs the original signal from the latent representation z. Following the standard formulation, the MAE objective can be written as Eq. 3.

$$\min_{\theta,\phi} \mathbb{E}_{x \sim D} \Big[\mathcal{H}\big(d_\phi(z),\, x \odot (1 - M)\big) \Big], \qquad z = f_\theta(x \odot M), \tag{3}$$

where, $\odot$ is the element-wise product, $\mathcal{H}(\cdot, \cdot)$ is a reconstruction loss (e.g., ℓ_2 or ℓ_1), and D denotes the pre-training distribution. This mask reconstruction paradigm underlies most recent EEG foundation models, including EEGPT model and CBraMOD model, provides the basic building block for EEG encoders used in our multimodal architecture.

2.2. EEGPT: A transformer-based EEG foundation model with dual self-supervision

EEGPT model is a transformer-based foundation model for EEG that operates on patch-level tokens derived from $x \in \mathbb{R}^{C \times T}$. The pre-processing first converts x into patches $X \in \mathbb{R}^{C \times N \times \tau}$ as in Eq. 1, then applies a local spatio-temporal embedding to map each patch $x_{i,j}$ to a token $e_{i,j} \in \mathbb{R}^d$.

EEGPT model extends the MAE formulation by introducing a dual self-supervised objective with an explicit representation branch. Concretely, a masked encoder f_θ produces a latent representation

$$z = f_\theta(x \odot M), \tag{4}$$

while a decoder d_ϕ reconstructs the unmasked portions of the signal and an additional branch aligns z with the features produced by applying the encoder to the unmasked input. This leads to the dual objective:

$$\min_{\theta,\phi} \mathbb{E}_{x\sim D}\Big[\mathcal{H}\big(d_\phi(z),\, x \odot (1-M)\big) + \mathcal{H}\big(z,\, f_\theta(x)\big)\Big], \qquad z = f_\theta(x \odot M), \tag{5}$$

The first term is a conventional reconstruction loss over the unmasked parts, while the second term encourages the masked representation z to be close to the "full-context" representation $f_\theta(x)$. Intuitively, this alignment makes z a more informative and stable summary of the entire EEG segment, improving robustness and downstream transfer.

2.3. CBraMOD: A criss-cross brain foundation model

Most prior EEG transformer models rely on generic full self-attention and fixed, hand-designed positional encodings defined over the channel–time grid, which often limit their flexibility in adapting to heterogeneous montages and diverse acquisition protocols. These design choices implicitly assume a uniform electrode layout and stable recording conditions, which may not hold across real-world EEG datasets. To our knowledge, CBraMOD is the first EEG foundation model to explicitly redesign both the attention pattern and positional encoding to better capture EEG spatial–temporal characteristics and accommodate heterogeneous montages

Given the masked patch tensor $\tilde{X} = \{\tilde{x}_{i,j}\} \in \mathbb{R}^{C\times N\times\tau}$ in Eq. 2, CBraMOD model first applies a local patch encoder g_ψ to obtain patch embeddings

$$e_{i,j} = g_\psi(\tilde{x}_{i,j}) \in \mathbb{R}^d, \qquad E = \{e_{i,j}\} \in \mathbb{R}^{C\times N\times d}. \tag{6}$$

To encode spatial–temporal locations, CBraMOD model replaces fixed sinusoidal or index-based schemes with an asymmetric conditional positional encoder h_α, implemented as a 2D convolution with kernel (k_θ, k_τ) on the channel-time grid:

$$E^p = h_\alpha(E) = \{e^p_{i,j}\} \in \mathbb{R}^{C\times N\times d}, \qquad E^o = \{e^o_{i,j}\} = \{e_{i,j} + e^p_{i,j}\}, \tag{7}$$

where, typically $k_\theta > k_\tau$ so that h_α can capture longer-range spatial and shorter-range temporal dependencies. This data-driven positional encoding is learned from EEG itself and is not hard-coded to a particular montage or sampling scheme, making the backbone more adaptable to diverse montage settings.

Let $E^{(b)} \in \mathbb{R}^{C\times N\times d}$ denote the input to the b-th criss-cross Transformer block. After layer normalization, a criss-cross attention operator CC_β factorizes self-attention into spatial and temporal stripes that explicitly follow the EEG topology. Concretely, for each head among K heads, spatial attention operates along channels at fixed time index j, while temporal attention operates along time at fixed channel index i:

$$F^{(s,k)}_{:,j} = \mathrm{Attn}\left((E^{(b)}_{:,j})W^{Q,k}_{S},\ (E^{(b)}_{:,j})W^{K,k}_{S},\ (E^{(b)}_{:,j})W^{V,k}_{S}\right), \tag{8}$$

$$F^{(t,k)}_{i,:} = \mathrm{Attn}\left((E^{(b)}_{i,:})W^{Q,k}_{T},\ (E^{(b)}_{i,:})W^{K,k}_{T},\ (E^{(b)}_{i,:})W^{V,k}_{T}\right), \tag{9}$$

with spatial heads $k \leq K/2$ and temporal heads $k > K/2$. The criss-cross attention output is then:

$$\mathrm{CC}_{\beta}(E^{(b)}) = \mathrm{Concat}\Big(\{F^{(S,k)}\}_{k<=K/2},\ \{F^{(T,k)}\}_{k>K/2}\Big), \tag{10}$$

Each block combines criss-cross attention with a feed-forward network FF_{γ} and residual connections:

$$\tilde{E}^{(b)} = E^{(b)} + \mathrm{CC}_{\beta}\big(\mathrm{LN}(E^{(b)})\big), \tag{11}$$

$$E^{(b+1)} = \tilde{E}^{(b)} + \mathrm{FF}_{\gamma}\big(\mathrm{LN}(\tilde{E}^{(b)})\big). \tag{12}$$

After $E^{(\beta)}$ such blocks, the resulting representation $E^{(\beta)}$ is reshaped into a latent tensor z and used as the encoder output $f_{\theta}(x \odot M)$ in the masked reconstruction objective of Eq. 12 and Eq. 11.

3. Methodology

In the following, 108MCH-EEG is incorporated into the pretraining corpus to derive a Vietnamese EEG foundation model for neurological disorder prediction, termed ViNeuro, which also serves as the backbone for its multimodal EEG–blood extensions. This design choice allows the proposed framework to leverage large-scale routine clinical EEG while maintaining compatibility with multimodal clinical settings, thereby enabling a systematic evaluation of the added value of integrating blood-based biomarkers with EEG representations for clinically relevant Alzheimer's disease risk prediction, as illustrated in Fig. 1. The overall framework comprises three stages: (1) pretraining data preprocessing, where routine clinical EEG from 108MCH-EEG is combined with selected public datasets and processed through a harmonized pipeline including montage standardization, filtering, resampling, and segmentation to ensure cross-dataset consistency; (2) EEG encoder pretraining (ViNeuro-EEG), in which masked EEG patches with asymmetric conditional positional encodings (ACPE) are processed by a criss-cross transformer encoder under a dual self-supervised objective incorporating both alignment and reconstruction losses to learn robust EEG representations; and (3) multimodal EEG–blood fusion

(ViNeuro-MM), where the pretrained ViNeuro-EEG is used to extract EEG tokens, routine blood-based biomarkers are projected into the same embedding space, and a cross-attention fusion module followed by a prediction head is applied to output the probability of Alzheimer's disease risk.

3.1. Pre-training data pre-processing

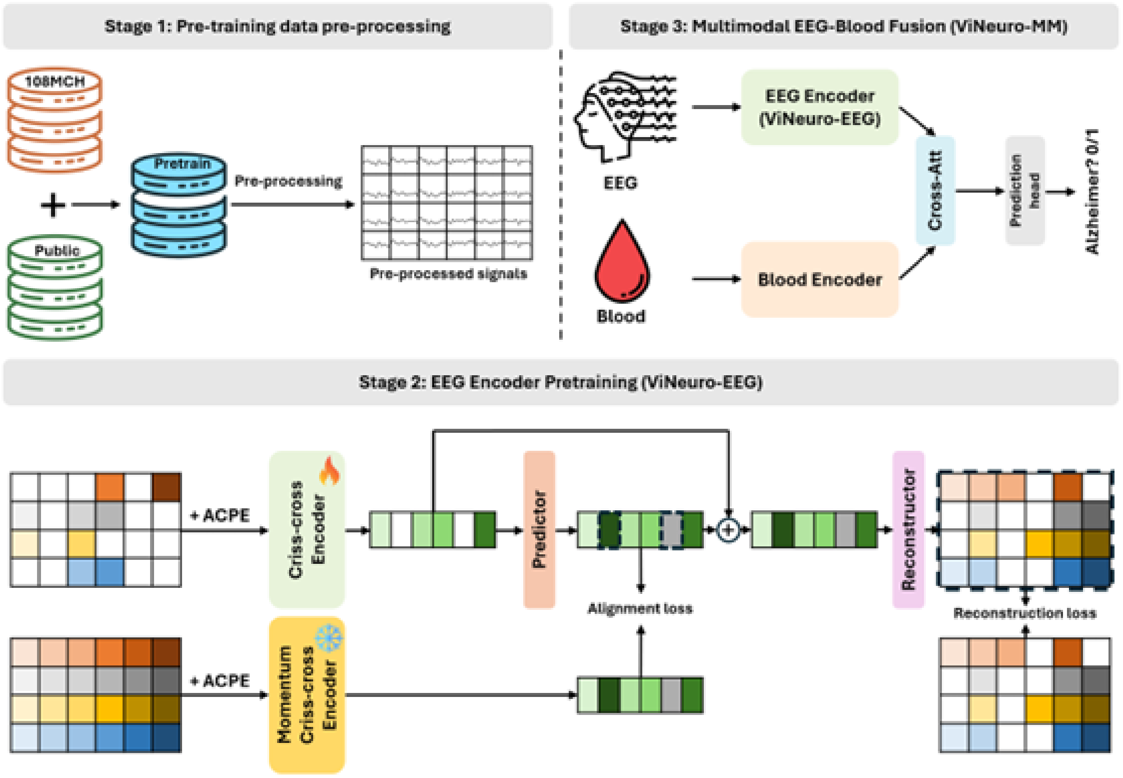

Fig. 1. Overview of the proposed ViNeuro framework.

The pretraining EEG data are curated and preprocessed in two stages. Section 3.1.1 describes the selection of clinically grounded, multi-channel scalp EEG corpora suitable for large-scale representation learning, while Section 3.1.2 details the preprocessing pipeline used to construct the final unified pretraining dataset, including channel selection, filtering, resampling, segmentation, and artifact removal.

3.1.1. Pre-training data curation

The 108MCH-EEG dataset, comprising approximately 800 hours of clinical EEG recordings from 5,134 patients at 108 Military Central Hospital in Vietnam, taken as the core of the pretraining corpus and is augmented with carefully selected international datasets to increase geographical and clinical diversity. Attention is restricted to corpora whose recording protocols, channel configurations, and clinical context are suitable for large-scale representation learning with multi-channel scalp

EEG. In particular, patient-based clinical recordings are prioritized over narrowly defined experimental paradigms (e.g. [7]–[12]), and sufficient spatial coverage is required to support meaningful modeling of spatiotemporal dynamics.

Within this design space, task-oriented datasets focusing on specific cognitive or motor paradigms are excluded, as they may bias learned representations toward predefined objectives rather than broadly useful clinical structure. Although the Sleep Heart Health Study dataset [13] offers an attractive sample size, it provides only two EEG channels in a sleep-focused setting, limiting its suitability for general-purpose pretraining. The Harvard EEG Database (HEEDB) [14] is likewise deferred due to its massive scale and ongoing integration effort, and is reserved for future work once harmonization is complete. As a result, the final pretraining corpus consists of three complementary sources: the Temple University EEG dataset [5], the NMT Scalp EEG dataset from Pakistan [6], and 108MCH-EEG from Vietnam. Collectively, these datasets span multiple hospitals, geographic regions, and acquisition protocols, yielding a diverse yet clinically grounded training set for ViNeuro-EEG.

3.1.2. Data preprocessing

The preprocessing procedure largely follows the CBraMOD model [3] and is conducted in a stepwise manner to suppress noise. For the Temple University EEG dataset, the first and last minute of each recording are discarded, and 19 common channels based on the international 10–20 system are retained. A 0.3–75 Hz band-pass filter is then applied together with a 60 Hz notch filter. All signals are subsequently resampled to 200 Hz, segmented into non-overlapping 30 s windows, and any window containing samples with amplitudes exceeding $100\,\mu\mathrm{V}$ is removed. The remaining windows are normalized to the range $[-1, 1]\,\mu\mathrm{V}$. For the NMT Scalp EEG dataset [6] and 108MCH-EEG, the same preprocessing pipeline is adopted, with the 60 Hz notch filter replaced by a 50 Hz notch filter to account for local power-line frequency. In addition, Independent Component Analysis is performed to further suppress ocular and muscular artifacts.

3.2. EEG encoder pretraining (ViNeuro-EEG)

A single EEG foundation encoder, termed ViNeuro-EEG, is pretrained using the dual self-supervised objective of the EEGPT model (Section 2.2), instantiated with a CBraMOD-style backbone (Section 2.3). Given an EEG segment, patching and masking are performed following the masked autoencoder framework described in Section 2.1. As in CBraMOD, masked patches are projected into embedding vectors and augmented with asymmetric conditional positional encodings on the channel–time grid. The resulting tokens are then processed by stacked criss-cross attention blocks that factorize attention into spatial and temporal components. The final token sequence is used as the learned EEG representation.

Following the EEGPT model, two views are constructed for each EEG window: a masked view and a full-context view. The encoder produces representations Z_{mask} and Z_{full} for the two views, and a lightweight decoder reconstructs the unmasked signal from

Z_{mask}. The dual pretraining loss extends the masked autoencoder objective in Eq. (3) with a representation alignment term between Z_{mask} and Z_{full}, encouraging ViNeuro-EEG to learn context-aware representations that are robust to masking and transferable across datasets. After pretraining on combined corpus, decoder is discarded, and the ViNeuro-EEG encoder is used as an EEG feature extractor in all multimodal experiments.

3.3. Multimodal EEG-blood fusion (ViNeuro-MM)

The integration of multimodal data streams has emerged as a promising approach for enhancing the early diagnosis of cognitive impairments such as Alzheimer's disease. Extensive medical research has shown that these neurological conditions are often associated with measurable shifts in peripheral blood profiles, including changes in the absolute counts and relative proportions of circulating immune cells [4]. Such alterations serve as indirect indicators of underlying neuroinflammatory processes and systemic immune dysregulation, which are hallmarks of progressive cognitive decline. A key advantage of blood-based biomarkers lies in their accessibility, offering a cost-effective and non-invasive alternative to more resource-intensive imaging modalities or cerebrospinal fluid analyses, and thereby enabling routine screening in clinical settings. Building on this foundation, a novel multimodal framework is introduced that synergistically combines blood test–derived biomarkers with EEG signals. By fusing these complementary data sources, physiological markers from blood and high-temporal-resolution neural activity from EEG. This approach aims to reveal subtle patterns of cognitive deterioration that may remain undetected in unimodal analyses.

Formally, let $r \in \mathbb{R}^m$ denote the normalized vector of blood-based biomarkers. Applying a lightweight projection network $\mathrm{MLP}(\cdot)$ that maps r into the EEG token embedding space:

$$q = \mathrm{MLP}(r) \in \mathbb{R}^d, \tag{13}$$

Given EEG embedded token $Z = \mathcal{E}_\theta(X) \in \mathbb{R}^{L\times d}$, where, $\mathcal{E}_\theta$ is the encoder, implement late fusion by treating q as a query attending to the EEG tokens:

$$\alpha = \mathrm{softmax}\left(\frac{(qW_q)(ZW_K)^\top}{\sqrt{d_k}}\right), \qquad h = \alpha(ZW_V)W_o \in \mathbb{R}^d. \tag{14}$$

The resulting cross-modal representation h serves as input to a prediction head for downstream tasks. At a high level, adopting cross-attention since it enables adaptive alignment between biomarker information and EEG dynamics: the biomarker query can selectively attend to the most informative EEG patterns rather than relying on a static combination. This flexibility is particularly important when the contribution of blood-based signals varies across patients or disease stages. When instantiated on top of the pretrained ViNeuro-EEG encoder, the fusion module yields a multimodal variant, ViNeuro-MM, which augments EEG representations with an explicit EEG-blood integration pathway.

Table 1. Performance of ViNeuro and multimodal baselines on PEARL for Alzheimer's risk prediction across MSIT, SMT, and RST tasks. ViNeuro-EEG uses EEG only (w/o BBB), while ViNeuro-MM, CBraMOD model, and EEGPT model use EEG plus blood-based biomarkers (w/ BBB). "Gain" indicates the relative improvement (%) of ViNeuro-MM over ViNeuro-EEG.

Tasks	**Architectures**	**Modality**	**Balanced Accuracy**		**AUPR**		**AUROC**	
		w/BBB	*Perf.*	*Gain*	*Perf.*	*Gain*	*Perf.*	*Gain*
MSIT	CBraMOD	Yes	0.6373		0.6863		0.7235	
	EEGPT	Yes	0.5560		0.6056		0.5023	
	ViNeuro-EEG	No	0.5422		0.5790		0.5877	
	ViNeuro-MM	Yes	0.6558	+20.95%	0.7225	+24.78%	0.7507	+27.74%
SMT	CBraMOD	Yes	0.6213		0.6043		0.6554	
	EEGPT	Yes	0.5226		0.5745		0.5285	
	ViNeuro-EEG	No	0.5296		0.4692		0.5040	
	ViNeuro-MM	Yes	0.6288	+18.73%	0.6774	+44.37%	0.7156	+41.98%
RST	CBraMOD	Yes	0.5793		0.6666		0.6416	
	EEGPT	Yes	0.5722		0.4856		0.4310	
	ViNeuro-EEG	No	0.5113		0.4445		0.4580	
	ViNeuro-MM	Yes	0.6377	+24.72%	0.7219	+62.41%	0.7100	+55.02%

4. Results and discussion

In this section, ViNeuro is evaluated for Alzheimer's disease risk prediction using EEG and blood-based biomarkers from the PEARL cohort. The unimodal EEG encoder (ViNeuro-EEG) is compared with its multimodal extension (ViNeuro-MM) to quantify the contribution of blood-based biomarkers, and ViNeuro-MM is further benchmarked against two strong multimodal baselines, namely CBraMOD and EEGPT model.

4.1. Experimental settings

The PEARL-Neuro dataset [4] is used, providing 64-channel EEG, routine blood tests, and clinical risk labels for middle-aged individuals at risk of dementia. Following prior work, three paradigms are considered: the Multi-SIT, the SMT, and RST. EEG is preprocessed and windowed as described in Section 3, and each segment is paired with the corresponding same-day blood panel. The prediction target is a binary Alzheimer's risk label.

For ViNeuro-EEG, fine-tuning is performed using EEG data only. For ViNeuro-MM, the same encoder is fine-tuned jointly with blood-based biomarkers through the fusion module described in Section 3.3. The CBraMOD model and the EEGPT model are trained as multimodal baselines that take both EEG and blood features as input. All models are evaluated at the subject level, and performance is reported using balanced accuracy, AUPR, and AUROC.

4.2. Alzheimer's risk prediction results

Table 1 summarizes Alzheimer's risk prediction performance on PEARL across the three paradigms. Across all tasks, augmenting ViNeuro-EEG with blood-based biomarkers (ViNeuro-MM) yields consistent and substantial gains. For example, in

MSIT, ViNeuro-MM improves over ViNeuro-EEG by +20.95% BA, +24.78% AUPR, and +27.74% AUROC. The gains are even more pronounced in the resting-state setting, where ViNeuro-MM achieves +24.72% BA, +62.41% AUPR, and +55.02% AUROC relative to its EEG-only counterpart. These improvements are obtained on top of baselines that already exceed chance-level balanced accuracy, underscoring the added value of incorporating routine blood-based biomarkers. When compared to multimodal CBraMOD model and EEGPT model, ViNeuro-MM is competitive or superior across most metrics and paradigms, indicating that pretraining ViNeuro-EEG on a unified clinical corpus and extending it with an explicit EEG-blood fusion pathway provides an effective and robust approach for Alzheimer's risk prediction.

5. Limitations

Although ViNeuro-EEG is pretrained on a unified corpus that includes Vietnamese and international clinical EEG, downstream evaluation is restricted to a single external cohort (PEARL-Neuro) and a single endpoint (Alzheimer's risk), so broader validation across sites, populations, and clinical labels is still needed. In addition, only two routinely available modalities, EEG and a limited panel of blood tests, are modeled, while other potentially informative data, such as imaging, medications, comorbidities, and longitudinal outcomes, are not considered. Finally, our multimodal design relies on a relatively simple fine-tuning scheme with a lightweight fusion head; exploring fully end-to-end training, calibration, and interpretability remains an important direction for future work.

6. Conclusion

In this work, ViNeuro is presented as a multimodal EEG–blood model for Alzheimer's disease risk prediction. ViNeuro comprises an EEG foundation encoder, termed ViNeuro-EEG, which is pretrained using an EEGPT model-style dual self-supervised objective and a CBraMOD-inspired criss-cross backbone on clinical EEG data from 108 Military Central Hospital and international datasets. Built upon this encoder, ViNeuro-MM is introduced as a cross-attention fusion module that integrates routine blood-based biomarkers with EEG representations. On the PEARL cohort, ViNeuro-MM consistently outperforms its EEG-only counterpart as well as strong multimodal baselines, achieving substantial relative gains in balanced accuracy, AUPR, and AUROC. These results demonstrate that fusing routine blood tests with EEG foundation representations constitutes a practical and effective strategy for enhancing early Alzheimer's disease risk prediction.

References

[1] P. M. Rodrigues, J. P. Teixeira, C. Garrett, D. Alves, and D. Freitas, "Alzheimer's early prediction with electroencephalogram," *Procedia Computer Science*, Vol. 100, pp. 865–871, 2016. DOI: 10.1016/j.procs.2016.09.236

[2] G. Wang, W. Liu, Y. He, C. Xu, L. Ma, and H. Li, “EEGPT: Pretrained transformer for universal and reliable representation of EEG signals,” *Advances in Neural Information Processing Systems*, Vol. 37, pp. 39 249–39 280, 2024. DOI: 10.52202/079017-1239

[3] J. Wang, S. Zhao, Z. Luo, Y. Zhou, H. Jiang, S. Li, T. Li, and G. Pan, “CBraMod: A criss-cross brain foundation model for EEG decoding,” *arXiv preprint arXiv:2412.07236*, 2024. DOI: 10.48550/arXiv.2412.07236

[4] P. Dzianok and E. Kublik, “PEARL-Neuro Database: EEG, fMRI, health and lifestyle data of middle-aged people at risk of dementia,” *Scientific Data*, Vol. 11, No. 1, p. 276, 2024. DOI: 10.1038/s41597-024-03106-5

[5] I. Obeid and J. Picone, “The temple university hospital EEG data corpus,” *Frontiers in Neuroscience*, Vol. 10, p. 196, 2016. DOI: 10.3389/fnins.2016.00196

[6] H. A. Khan, R. Ul Ain, A. M. Kamboh, H. T. Butt, S. Shafait, W. Alamgir, D. Stricker, and F. Shafait, “The NMT scalp EEG dataset: An open-source annotated dataset of healthy and pathological EEG recordings for predictive modeling,” *Frontiers in Neuroscience*, Vol. 15, 2022. DOI: 10.3389/fnins.2021.755817

[7] W.-L. Zheng and B.-L. Lu, “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks,” *IEEE Transactions on Autonomous Mental Development*, Vol. 7, No. 3, pp. 162–175, 2015. DOI: 10.1109/TAMD.2015.2431497

[8] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” *Circulation*, Vol. 101, No. 23, pp. e215–e220, 2000. DOI: 10.1161/01.CIR.101.23.e215

[9] G. Huang, Z. Hu, W. Chen, S. Zhang, Z. Liang, L. Li, L. Zhang, and Z. Zhang, “M3CV: A multi-subject, multi-session, and multi-task database for EEG-based biometrics challenge,” *NeuroImage*, Vol. 264, 2022. DOI: 10.1016/j.neuroimage.2022.119666

[10] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, “Deep learning with convolutional neural networks for EEG decoding and visualization,” *Human Brain Mapping*, Vol. 38, No. 11, pp. 5391–5420, 2017. DOI: 10.1002/hbm.23730

[11] C. Ma Thi *et al.*, “UET175: EEG dataset of motor imagery tasks in Vietnamese stroke patients,” *Frontiers in Neuroscience*, Vol. 19, 2025. DOI: 10.3389/fnins.2025.1580931

[12] T. D. Ngo, H. D. Kieu, M. H. Nguyen, T. H.-A. Nguyen, V. M. Can, B. H. Nguyen, and T. H. Le, “An EEG & eye-tracking dataset of ALS patients & healthy people during eye-tracking-based spelling system usage,” *Scientific Data*, Vol. 11, No. 1, p. 664, 2024. DOI: 10.1038/s41597-024-03501-y

[13] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, “The National Sleep Research Resource: towards a sleep data commons,” *Journal of the American Medical Informatics Association*, Vol. 25, No. 10, pp. 1351–1358, 2018. DOI: 10.1145/3233547.3233725

[14] C. Sun *et al.*, “Harvard electroencephalography database: A comprehensive clinical electroencephalographic resource from four Boston hospitals,” *Epilepsia*, Vol. 66, No. 9, pp. 3411–3425, 2025. DOI: 10.1111/epi.18487

Manuscript received 06-11-2025; Accepted 19-12-2025.

■

Thanh Trung Nguyen received Bachelor’s degree in Biomedical Engineering from Le Quy Don Technical University, Vietnam, in 2010, Master’s degree in Biomedical Engineering in 2015, and the PhD. degree in Electronic Engineering from Hanoi University of Science and Technology in 2020. He is currently the Vice Head of the Department of Equipment at the 108 Military Central Hospital and a lecturer at the Department of Diagnostic Imaging, Institute of Clinical Medical and Pharmaceutical Sciences 108. His research interests include biomedical instrumentation, medical imaging, and electronic systems in healthcare.
Email: trung.ntc10@benhvien108.vn

VINEURO: MÔ HÌNH HỢP NHẤT ĐA PHƯƠNG THỨC EEG-MÁU CHO DỰ ĐOÁN NGUY CƠ ALZHEIMER

Nguyễn Thành Trung

Tóm tắt

Dự đoán sớm nguy cơ mắc bệnh Alzheimer có ý nghĩa then chốt để can thiệp kịp thời, nhưng vẫn còn nhiều thách thức trong thực hành lâm sàng thường quy. Điện não đồ *(EEG)* là phương tiện rẻ tiền và không xâm lấn, tuy nhiên chỉ riêng EEG thường không đủ độ nhạy và độ ổn định để ước lượng nguy cơ ở giai đoạn sớm một cách đáng tin cậy. Song song đó, các xét nghiệm máu thường quy phản ánh những thay đổi về miễn dịch ngoại vi, tình trạng viêm và chuyển hoá liên quan tới suy giảm nhận thức, gợi ý rằng việc kết hợp EEG với các chỉ dấu sinh học từ máu có thể giúp phân tầng nguy cơ giàu thông tin hơn. Trong công trình này, tác giả đề xuất ViNeuro, một mô hình đa mô thức EEG-máu được thiết kế chuyên biệt cho bài toán dự đoán nguy cơ Alzheimer. Trước hết là bước tiền huấn luyện một bộ mã hoá nền tảng EEG duy nhất, gọi là ViNeuro-EEG, sử dụng mục tiêu tự giám sát kép *(dual self-supervised objective)* của mô hình EEGPT kết hợp với backbone criss-cross và mã hoá vị trí học được *(learned positional encoding)* của mô hình CBraMOD. Quá trình tiền huấn luyện được thực hiện trên một corpora EEG lâm sàng đa kênh đã được chuẩn hoá, bao gồm cả các bản ghi EEG từ Bệnh viện Trung ương Quân đội 108 ở Việt Nam và các bộ dữ liệu quốc tế. Trên nền bộ mã hoá này xây dựng ViNeuro-MM, mô hình đa phương thức chiếu các chỉ dấu sinh học từ xét nghiệm máu vào không gian embedding của EEG và sử dụng chúng như các truy vấn trong một lớp cross-attention trên các token EEG. Nghiên cứu đánh giá ViNeuro trên cohort PEARL-Neuro cho bài toán dự đoán nguy cơ Alzheimer. So với phiên bản chỉ dùng EEG, ViNeuro-MM đạt mức cải thiện hiệu năng đáng kể, với mức tăng tương đối lên tới 24,72% về độ chính xác cân bằng *(balanced accuracy)*, cho thấy rằng việc kết hợp các chỉ dấu sinh học từ xét nghiệm máu thường quy với các biểu diễn nền tảng từ EEG có thể nâng cao rõ rệt khả năng dự đoán sớm nguy cơ Alzheimer.

Từ khóa

EEG lâm sàng; mô hình nền tảng; dự đoán nguy cơ Alzheimer; biểu diễn đặc trưng, học tự giám sát; học đa mô thức; máu.