Neuroba

Neuroba NCTS Research Series
Published by Neuroba Research | Neuroba.com

Neuroba NCTS Research Series (2026b) • Pages 1–22
DOI: https://doi.org/10.5281/zenodo.20567638

Original Research Article

Transformer-Based Neural Decoding for Real-Time Intent and Emotion Classification from EEG Signals

Neuroba Research[1*]
[1]*Neuroba Research, Neuroba.com*

Abstract: Brain–Computer Interfaces (BCIs) hold transformative potential across diverse applications, from assistive technologies to human-computer interaction. A critical bottleneck in realizing this potential, particularly in real-world scenarios, is the accurate and real-time decoding of complex neural signals, specifically for inferring user intent and emotional states. Traditional machine learning approaches often struggle with the inherent non-stationarity, low signal-to-noise ratio, and high inter-subject variability characteristic of electroencephalography (EEG) signals. Furthermore, the temporal dependencies and long-range correlations within EEG data are often inadequately captured by conventional models, limiting their generalization capabilities. This paper introduces the Neuroba Transformer-Based Neural Decoding Architecture (NTNDA), a novel conceptual framework designed to leverage the power of transformer networks for robust, real-time classification of both user intent and emotion from EEG signals. NTNDA comprises five key modules: EEG Signal Preprocessing Layer, Feature Embedding Layer, Transformer Encoder Layer, Temporal Attention Mechanism, and a Dual Output Head for simultaneous intent and emotion classification. By employing self-attention mechanisms, NTNDA is engineered to effectively model complex temporal dynamics and capture intricate relationships within EEG data, even in the presence of noise. This framework is expected to significantly advance the state-of-the-art in neural decoding, offering improved accuracy, generalization, and real-time performance. While NTNDA presents a comprehensive solution, its practical implementation faces challenges related to computational constraints, dataset limitations, and ethical considerations. Future work will focus on multimodal neural decoding, larger transformer models, and cross-brain generalization to further enhance the capabilities of the Neuroba NCTS Framework.

Keywords: *Brain–Computer Interface (BCI), EEG, Neural Decoding, Transformers, Deep Learning, Emotion Recognition, Intent Classification, Neurotechnology, Human–Computer Interaction*

Corresponding Author: *Neuroba Research*
Neuroba Research, Neuroba.com

I. INTRODUCTION

The burgeoning field of Brain–Computer Interfaces (BCIs) promises to revolutionize how humans interact with their environment, offering unprecedented avenues for communication, control, and cognitive augmentation [1]. At the heart of any effective BCI lies the ability to accurately translate complex neural activity into actionable commands or meaningful insights. While significant progress has been made in signal acquisition, particularly with adaptive multimodal approaches [Neuroba Research (2026a)], the subsequent challenge of decoding these signals, especially for nuanced cognitive states like intent and emotion, remains a formidable barrier to widespread real-world deployment.

A. Evolution of Neural Decoding in BCIs

Early neural decoding efforts in BCIs primarily relied on classical machine learning algorithms such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVMs), and Common Spatial Patterns (CSP) [2], [3]. These methods proved effective for well-defined tasks in controlled laboratory settings, such as motor imagery classification or P300 detection. With the advent of deep learning, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, began to offer improved performance by automatically learning hierarchical features from raw or minimally preprocessed EEG data [4], [5]. These models demonstrated superior capabilities in handling the high-dimensional and complex nature of neural signals, pushing the boundaries of decoding accuracy.

B. Importance of Intent and Emotion Classification

Beyond basic command and control, the ability to accurately classify user intent and emotional states from neural signals is crucial for developing truly intuitive and adaptive BCIs. Intent decoding allows for more natural and proactive interaction, enabling systems to anticipate user needs rather than merely reacting to explicit commands [6]. Emotion recognition, on the other hand, can facilitate personalized neurofeedback, enhance human-AI collaboration, and provide critical insights for mental health monitoring [7]. For instance, a BCI capable of detecting frustration could adapt its interface, while one recognizing focused attention could optimize task delivery. The integration of both intent and emotion decoding moves BCIs closer to understanding the full spectrum of human cognitive and affective states.

C. Limitations of Classical ML Methods

Despite their historical significance, classical machine learning methods and even early deep learning architectures face inherent limitations when confronted with the complexities of real-world EEG data. These include:

- **Inability to Capture Long-Range Dependencies:** EEG signals are time-series data with intricate temporal correlations that span various timescales. Traditional models often struggle to effectively capture these long-range dependencies, which are vital for understanding dynamic cognitive processes [8].
- **Sensitivity to Noise and Artifacts:** While advanced signal acquisition techniques (e.g., NAMSAA [Neuroba Research (2026a)]) can mitigate noise, residual artifacts and the inherent low signal-to-noise ratio of EEG still pose significant challenges for models that lack robust attention mechanisms [9].
- **Inter-Subject Variability:** The unique physiological and anatomical differences between individuals lead to substantial variations in EEG patterns, making it difficult for models to generalize across users without extensive calibration [10].
- **Limited Generalization:** Models trained on specific tasks or datasets often exhibit poor performance when applied to novel scenarios or different user populations, hindering the scalability of BCI solutions [11].
- **Lack of Interpretability:**Many deep learning models operate as "black boxes," making it challenging to understand their decision-making processes, which is critical for clinical and safety-critical BCI applications [12].

D. Challenges in EEG Signal Complexity

EEG signals are inherently complex, characterized by high dimensionality, non-stationarity, and a low signal-to-noise ratio. The subtle neural correlates of intent and emotion are often masked by background brain activity, physiological artifacts (e.g., eye blinks, muscle movements), and environmental noise [13]. Furthermore, the temporal dynamics of these cognitive states can vary significantly, requiring models that can capture both short-term fluctuations and long-term dependencies within the neural data. The spatial distribution of EEG electrodes also provides limited resolution, making it difficult to pinpoint precise brain regions involved in complex cognitive processes [14].

E. Motivation for Transformer-Based Architectures

The advent of transformer architectures, initially popularized in natural language processing (NLP), has revolutionized sequence modeling by effectively capturing long-range dependencies through self-attention mechanisms [15]. Their success in processing sequential data, coupled with their ability to weigh the importance of different parts of an input sequence, makes them highly suitable for analyzing complex EEG time series. Transformers offer a promising solution to overcome the limitations of traditional models by providing a more robust and adaptive framework for neural decoding, particularly for tasks requiring an understanding of dynamic cognitive states like intent and emotion [16].

F. Research Objectives

This paper aims to address the aforementioned challenges by pursuing the following objectives:

1. To critically review the current state-of-the-art in EEG-based emotion recognition, intent decoding, and deep learning approaches, with a particular focus on attention mechanisms and transformer architectures.
2. To identify the key limitations and research gaps in existing neural decoding models for real-time BCI applications.
3. To propose a novel conceptual framework, the Neuroba Transformer-Based Neural Decoding Architecture (NTNDA), designed for robust, real-time classification of user intent and emotional states from EEG signals.
4. To mathematically formulate key components of NTNDA, including the attention mechanism, transformer encoding, and classification loss functions.
5. To outline an implementation roadmap and discuss the potential applications, challenges, and ethical considerations associated with NTNDA.
6. To establish the foundational role of NTNDA within Layer 02 (DECODE) of the broader Neuroba NCTS Framework, elucidating its responsibilities and interface with upstream (SIGNAL) and downstream (TRANSMIT) processing layers.

G. Key Contributions

This paper makes several significant contributions to the field of BCI research and engineering:

- **Novel Conceptual Architecture:** Introduction of NTNDA, a comprehensive transformer-based neural decoding architecture specifically designed for real-time intent and emotion classification from EEG signals.
- **Modular Framework:** Detailed description of NTNDA's five internal modules, outlining their function, inputs, outputs, architecture, benefits, and limitations.
- **Mathematical Formulations:** Provision of mathematical models for critical aspects of transformer-based decoding, including scaled dot-product attention, transformer encoding, and multi-task loss functions.
- **Integration with NCTS:** Elucidation of NTNDA's role as Layer 02 (DECODE) within the Neuroba NCTS Framework, emphasizing its interface with Layer 01 (SIGNAL) [Neuroba Research (2026a)] and Layer 03 (TRANSMIT), and its contribution to semantic interpretation of neural signals.
- **Roadmap for Implementation:** A practical roadmap for the development and deployment of NTNDA, including considerations for real-time inference and optimization strategies.

This paper serves as a foundational document for the Neuroba NCTS Research Series, building upon the robust signal acquisition principles established in Paper 1 [Neuroba Research (2026a)] and laying the groundwork for subsequent papers that will delve into transmission, interpretation, and connection layers of the NCTS framework.

II. LITERATURE REVIEW

The accurate and timely decoding of neural signals is paramount for the efficacy of Brain–Computer Interfaces

(BCIs). This section provides a critical review of existing research on EEG-based emotion recognition, intent decoding, and the evolution of deep learning approaches, culminating in the emergence of transformer architectures for time-series analysis.

A. EEG-Based Emotion Recognition Systems

Emotion recognition from EEG signals has garnered significant attention due to its potential in mental health monitoring, adaptive human-computer interaction, and neurofeedback. Early approaches often relied on hand-crafted features extracted from EEG (e.g., power spectral density in alpha, beta, theta bands, asymmetry indices) combined with traditional classifiers like SVM, k-Nearest Neighbors (k-NN), and Random Forests [17], [18]. While these methods showed promise in controlled settings, their performance often degraded in real-world scenarios due to the non-stationary nature of emotions and EEG signals. More recently, deep learning models, particularly CNNs and RNNs, have been employed to automatically learn discriminative features from raw or preprocessed EEG data, achieving improved accuracy and robustness [19], [20]. However, capturing the subtle and dynamic temporal patterns associated with emotional states remains a challenge, often requiring models capable of discerning long-range dependencies.

B. Intent Decoding in BCIs

Intent decoding is a cornerstone of active BCI systems, enabling users to control external devices through their thoughts. Historically, motor imagery (MI) BCIs have dominated this area, utilizing event-related desynchronization/synchronization (ERD/ERS) in sensorimotor rhythms [21]. Decoding algorithms for MI often involve CSP for spatial filtering, followed by LDA or SVM for classification [22]. Other paradigms, such as P300-based spellers and SSVEP, rely on evoked potentials and frequency-tagging, respectively, with decoding focused on detecting specific signal components [23]. While effective, these methods often require extensive calibration and are sensitive to variations in signal quality. The transition to more natural and intuitive control necessitates decoding more abstract forms of intent, moving beyond simple motor commands to higher-level cognitive states [24].

C. Deep Learning Approaches (CNN, RNN, LSTM)

Deep learning has significantly advanced neural decoding by enabling end-to-end learning from raw EEG data. Convolutional Neural Networks (CNNs) excel at extracting spatial features from multi-channel EEG, treating the electrode array as an image-like input [25]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), are well-suited for sequential data like EEG, capable of modeling temporal dependencies [26]. Hybrid models combining CNNs for spatial feature extraction and RNNs/LSTMs for temporal modeling have shown superior performance in various EEG classification tasks, including motor imagery and emotion recognition [27], [28]. These models mitigate some of the limitations of classical methods by automatically learning complex features, but they can still struggle with very long-range dependencies and the inherent noise in EEG signals.

D. Attention Mechanisms in Neural Signal Processing

Attention mechanisms, inspired by human cognitive processes, allow neural networks to selectively focus on the most relevant parts of an input sequence [29]. In EEG analysis, attention mechanisms have been integrated into CNN-RNN architectures to highlight salient time points or electrode channels, thereby improving classification accuracy [30]. For instance, spatial attention can weigh the importance of different electrodes, while temporal attention can emphasize critical moments in the EEG signal. These mechanisms provide a degree of interpretability by indicating which parts of the input contribute most to the model's decision, addressing one of the
key limitations of deep learning models [31].

E. Transformer Architectures in Time-Series Modeling

The introduction of the Transformer architecture, characterized by its self-attention mechanism, marked a paradigm shift in sequence modeling [32]. Unlike RNNs, which process data sequentially, Transformers process the entire sequence simultaneously, enabling them to capture long-range dependencies more effectively and facilitating parallel computation [33]. This capability has led to their widespread adoption in natural language processing and, more recently, in time-series analysis, including EEG decoding [34]. Transformer-based models have demonstrated state-of-the-art performance in various EEG tasks, such as sleep stage classification, emotion recognition, and motor imagery decoding [35], [36]. Their ability to model complex temporal dynamics and weigh the importance of different signal segments makes them highly promising for robust neural decoding.

F. Multimodal EEG Decoding Approaches

Recognizing the limitations of single-modality EEG, researchers have increasingly explored multimodal approaches, combining EEG with other physiological signals (e.g., fNIRS, EMG, EOG) or contextual information (e.g., eye-tracking, facial expressions) [37]. Multimodal decoding aims to leverage the complementary strengths of different modalities to improve accuracy and robustness, particularly in real-world environments where EEG signals are prone to artifacts [38]. While multimodal approaches offer significant advantages, they also introduce challenges related to data synchronization, fusion strategies, and increased computational complexity [39].

G. Limitations of Current Models

Despite the advancements in deep learning and the emergence of transformer architectures, several limitations persist in current neural decoding models:

- **Computational Complexity:** Transformer models, particularly those with deep architectures and large attention heads, can be computationally intensive, posing challenges for real-time inference on resource-constrained edge devices [40].
- **Data Requirements:** Deep learning models, including transformers, typically require large amounts of labeled data for training, which can be difficult and time-consuming to acquire in the BCI domain [41].
- **Generalization Across Subjects:** The high inter-subject variability of EEG signals remains a significant hurdle, with models often struggling to generalize to new users without extensive fine-tuning or calibration [42].
- **Real-Time Performance:** Many state-of-the-art models are evaluated offline, and their performance in real-time, closed-loop BCI systems remains to be fully validated [43].

These limitations underscore the need for novel architectures that balance decoding accuracy with computational efficiency, generalization capabilities, and real-time performance. The proposed Neuroba Transformer-Based Neural Decoding Architecture (NTNDA) aims to address these challenges by providing a robust and adaptive framework for intent and emotion classification.

III. PROBLEM STATEMENT

The realization of robust, real-world Brain–Computer Interfaces (BCIs) is significantly hindered by the complexities inherent in decoding neural signals, particularly for nuanced cognitive states such as intent and emotion. While advanced signal acquisition techniques, as described in Layer 01 (SIGNAL) [Neuroba Research (2026a)], can mitigate noise and artifacts, the fundamental challenges of interpreting the underlying neural dynamics remain. This section delineates the core problems that necessitate a novel, transformer-based approach to neural decoding.

A. Temporal Instability in EEG Signals

EEG signals are highly non-stationary, meaning their statistical properties (e.g., mean, variance, spectral content) change over time [44]. This temporal instability arises from the dynamic nature of brain activity, which is constantly influenced by internal cognitive processes, external stimuli, and physiological changes (e.g., fatigue, attention shifts). Traditional machine learning models, and even some deep learning architectures, often assume stationarity or struggle to adapt to these rapid fluctuations, leading to degraded decoding performance over time [45]. A robust decoding architecture must be capable of modeling and adapting to these complex temporal dynamics.

B. Low Signal-to-Noise Ratio

Despite rigorous preprocessing and artifact removal, EEG signals inherently possess a low signal-to-noise ratio (SNR). The neural correlates of specific intents or emotions are often subtle and embedded within a background of ongoing, task-irrelevant brain activity [46]. This "background noise" can obscure the target signals, making it difficult for decoding algorithms to extract meaningful features. Models must possess sophisticated mechanisms to selectively attend to relevant signal components while suppressing irrelevant noise, a task for which attention mechanisms are particularly well-suited [47].

C. Inter-Subject Variability

One of the most pervasive challenges in BCI research is the high degree of inter-subject variability. The anatomical structure of the brain, the spatial distribution of neural sources, and the individual strategies employed during cognitive tasks vary significantly across users [48]. Consequently, a decoding model trained on data from one group of subjects often performs poorly when applied to a new user. This lack of generalization necessitates extensive, time-consuming calibration sessions for each new user, hindering the practical deployment and scalability of BCI systems [49]. Developing models that can learn subject-invariant features or rapidly adapt to new users is a critical requirement.

D. Lack of Generalization in Models

Beyond inter-subject variability, current decoding models often lack generalization across different tasks, contexts, or environmental conditions. A model trained to recognize emotions in a controlled laboratory setting may fail when deployed in a dynamic real-world environment [50]. This lack of generalization stems from the models' tendency to overfit to specific datasets or experimental paradigms, failing to capture the underlying, invariant neural representations of the target cognitive states. A robust architecture must be capable of learning generalized features that are resilient to contextual variations.

E. Real-Time Inference Constraints

For BCIs to be truly interactive and responsive, neural decoding must occur in real-time, with minimal latency. This requirement imposes strict constraints on the computational complexity of the decoding algorithms [51]. While deep learning models, particularly large transformers, offer superior accuracy, their high computational demands can lead to unacceptable delays in real-time applications, especially when deployed on resource-constrained edge devices (e.g., wearable BCI headsets). Balancing decoding accuracy with computational efficiency and low-latency inference is a critical engineering challenge [52].

These interconnected problems collectively limit the robustness, reliability, and widespread applicability of current BCI decoding systems. Addressing these challenges requires a paradigm shift towards architectures that can effectively model complex temporal dynamics, selectively attend to relevant features, generalize across subjects and contexts, and operate within strict real-time constraints. The Neuroba Transformer-Based Neural Decoding Architecture (NTNDA) is proposed as a comprehensive solution to these critical issues.

IV. PROPOSED FRAMEWORK

To address the multifaceted challenges of temporal instability, low signal-to-noise ratio, inter-subject variability, and real-time inference constraints in neural decoding, we propose the **Neuroba Transformer-Based Neural Decoding Architecture (NTNDA)**. NTNDA is a novel conceptual framework designed to leverage the power of transformer networks and self-attention mechanisms for the robust, real-time classification of user intent and emotional states from EEG signals. This architecture is specifically engineered to operate as Layer 02 (DECODE) within the broader Neuroba NCTS Framework, translating the standardized neural inputs from Layer 01 (SIGNAL) [Neuroba Research (2026a)] into actionable semantic interpretations.

NTNDA is structured around five interconnected internal modules, each responsible for a distinct stage of the neural decoding pipeline. These modules work in concert to extract meaningful features, model complex temporal dynamics, and generate accurate classifications.

A. Module 1: EEG Signal Preprocessing Layer

Function: The primary function of this module is to receive the standardized, high-fidelity neural data stream from Layer 01 (SIGNAL) and perform any necessary task-specific preprocessing before feature extraction. While Layer 01 handles general artifact removal and standardization, this layer focuses on preparing the data specifically for the transformer architecture.

Inputs:

- Standardized, artifact-suppressed multimodal neural signals from Layer 01 (SIGNAL) [Neuroba Research (2026a)].

- Associated metadata, including signal quality indicators and artifact flags.

Outputs:

- Segmented and normalized EEG data epochs, ready for feature embedding.

Architecture:

7 **Segmentation (Epoching):** The continuous EEG stream is divided into overlapping or non-overlapping time windows (epochs) suitable for the specific decoding task (e.g., 1-second windows for emotion recognition, shorter windows for rapid intent detection).
8 **Task-Specific Filtering:** Application of bandpass filters to isolate specific frequency bands relevant to the target cognitive states (e.g., alpha and beta bands for emotion, mu and beta bands for motor intent).
9 **Normalization:** Standardization of the data within each epoch (e.g., z-score normalization) to ensure consistent input scaling for the neural network, mitigating the effects of amplitude variations across channels or subjects.

Benefits:

- **Task Optimization:** Tailors the input data to the specific requirements of the decoding task, enhancing the salience of relevant neural features.
- **Improved Stability:** Normalization stabilizes the training process and improves the convergence of the subsequent deep learning layers.

Limitations:

- **Information Loss:** Aggressive filtering or inappropriate epoch lengths can inadvertently discard relevant neural information or disrupt temporal dynamics.
- **Latency:** The epoching process introduces an inherent delay equal to the window length, which must be carefully managed for real-time applications.

B. Module 2: Feature Embedding Layer

Function: This module transforms the preprocessed EEG epochs into a high-dimensional feature representation suitable for processing by the transformer encoder. It maps the raw time-series data into a sequence of continuous embeddings, capturing both spatial and temporal characteristics.

Inputs:

- Segmented and normalized EEG data epochs from Module 1.

Outputs:

- A sequence of high-dimensional feature embeddings, incorporating spatial and temporal information.

Architecture:

10 **Spatial-Temporal Convolution:** A series of 1D or 2D convolutional layers are applied to the EEG epochs to extract local spatial (across channels) and temporal (across time points) features. This step acts as a trainable feature extractor, replacing hand-crafted features.
11 **Linear Projection:** The output of the convolutional layers is flattened and passed through a linear projection layer to map the features into the desired embedding dimension (\$d_{model}\$) required by the transformer.
12 **Positional Encoding:** Since transformers lack an inherent sense of sequence order, positional encodings (e.g., sine and cosine functions of different frequencies) are added to the feature embeddings to inject information about the relative or absolute position of each time step within the epoch.

Benefits:

- **Automated Feature Extraction:** Replaces manual feature engineering with a data-driven approach, allowing the model to learn the most discriminative representations.
- **Sequence Representation:** Transforms the continuous EEG signal into a sequence of discrete tokens, enabling the application of transformer architectures.

Limitations:

- **Computational Overhead:** The convolutional layers add computational complexity, particularly for high-channel-count EEG data.
- **Hyperparameter Sensitivity:** The performance is sensitive to the choice of convolutional kernel sizes, strides, and the embedding dimension.

C. Module 3: Transformer Encoder Layer

Function: The core of NTNDA, this module utilizes a stack of transformer encoder blocks to model the complex, long-range temporal dependencies within the sequence of feature embeddings. It aims to capture the dynamic evolution of neural states associated with intent and emotion.

Inputs:

- Sequence of feature embeddings (with positional encodings) from Module 2.

Outputs:

- A sequence of context-aware, high-level representations of the EEG data.

Architecture:

13 **Multi-Head Self-Attention (MHSA):** This mechanism allows the model to weigh the importance of different time steps within the sequence relative to each other. By employing multiple attention heads, the model can simultaneously attend to information from different representation subspaces at different positions.

14 **Feed-Forward Network (FFN):** A fully connected feed-forward network is applied independently to each position in the sequence, introducing non-linearity and further transforming the representations.

15 **Layer Normalization and Residual Connections:** Layer normalization and residual (skip) connections are employed around each sub-layer (MHSA and FFN) to stabilize training, mitigate the vanishing gradient problem, and facilitate the flow of information through deep networks.

Benefits:

- **Long-Range Dependency Modeling:** Effectively captures complex temporal dynamics and correlations across the entire EEG epoch, overcoming the limitations of RNNs.
- **Parallel Computation:** Unlike RNNs, the self-attention mechanism allows for parallel processing of the sequence, improving computational efficiency during training.

Limitations:

- **Quadratic Complexity:** The computational complexity of self-attention scales quadratically with the sequence length, posing challenges for very long EEG epochs.
- **Data Hunger:** Transformers typically require large amounts of data to train effectively and avoid overfitting.

D. Module 4: Temporal Attention Mechanism

Function: This module acts as a specialized pooling layer, aggregating the sequence of context-aware representations from the Transformer Encoder into a single, fixed-size vector that summarizes the most salient information for the classification tasks.

Inputs:

- Sequence of context-aware representations from Module 3.

Outputs:

- A single, aggregated feature vector representing the entire EEG epoch.

Architecture:

16 **Attention Scoring:** A learned attention mechanism calculates a weight (score) for each time step in the sequence, indicating its relevance to the overall classification task.

17 **Weighted Summation:** The context-aware representations are multiplied by their corresponding attention scores and summed to produce the final aggregated feature vector.

Benefits:

- **Selective Focus:** Allows the model to focus on the most discriminative parts of the EEG epoch, ignoring irrelevant or noisy segments.
- **Interpretability:** The attention scores provide a degree of interpretability, highlighting which time points were most influential in the model's decision.

Limitations:

- **Information Bottleneck:** Aggregating the entire sequence into a single vector can potentially lead to the loss of fine-grained temporal details.

E. Module 5: Dual Output Head (Intent + Emotion Classification)

Function: The final module takes the aggregated feature vector and performs simultaneous classification of user intent and emotional state. This multi-task learning approach leverages shared representations to improve the generalization and robustness of both tasks.

Inputs:

- Aggregated feature vector from Module 4.

Outputs:

- Predicted intent class (e.g., motor command, cognitive state).
- Predicted emotional state (e.g., valence/arousal levels, discrete emotion categories).

Architecture:

18 **Shared Representation:** The aggregated feature vector serves as a shared representation for both classification tasks.

19 **Intent Classification Head:** A fully connected layer followed by a softmax activation function to output the probability distribution over the predefined intent classes.

20 **Emotion Classification Head:** A separate fully connected layer followed by a softmax (for discrete emotions) or linear (for continuous valence/arousal) activation function to output the predicted emotional state.

21 **Multi-Task Loss Function:** The model is trained using a combined loss function that optimizes both intent and emotion classification simultaneously, encouraging the learning of generalized features.

Benefits:

- **Multi-Task Synergy:** Leverages the underlying correlations between intent and emotion to improve the performance of both tasks.
- **Efficiency:** A single shared encoder reduces the overall computational burden compared to training separate models for each task.

Limitations:

- **Task Interference:** If the tasks are not sufficiently related, optimizing for one task might degrade the performance of the other (negative transfer).
- **Loss Weighting:** Balancing the contribution of the intent and emotion loss terms in the combined loss function requires careful tuning.

V. SYSTEM ARCHITECTURE

The Neuroba Transformer-Based Neural Decoding Architecture (NTNDA) is realized through a sophisticated system architecture that integrates the five core modules into a cohesive, real-time inference pipeline. This section details the structural and functional elements of the NTNDA system.

A. EEG Acquisition Pipeline

While NTNDA focuses on decoding, it relies on a robust acquisition pipeline, specifically Layer 01 (SIGNAL) of the NCTS Framework [Neuroba Research (2026a)]. This pipeline provides continuous, synchronized, and artifact-suppressed multimodal neural data (primarily EEG, but potentially including fNIRS or EMG for context). The data is streamed to NTNDA via a high-bandwidth, low-latency interface (e.g., Lab Streaming Layer - LSL).

B. Preprocessing Pipeline

Upon receiving the data stream, the Preprocessing Pipeline (Module 1) segments the continuous data into overlapping epochs (e.g., 1-second windows with a 50% overlap). This overlapping strategy ensures continuous, high-frequency updates for real-time applications. The epochs undergo task-specific bandpass filtering (e.g., 1-50 Hz) and z-score normalization to standardize the input amplitude across channels and time.

C. Embedding Strategy

The normalized epochs are fed into the Feature Embedding Layer (Module 2). Here, a spatial-temporal convolutional block extracts local features. For instance, a 1D convolution along the time axis captures temporal frequency patterns, followed by a spatial convolution across channels to capture spatial distributions. The output is flattened and linearly projected into a sequence of vectors of dimension d_{model} (e.g., 256 or 512). Crucially, sinusoidal positional encodings are added to these vectors to retain temporal order information, creating the final sequence of embeddings.

D. Transformer Stack

The sequence of embeddings enters the Transformer Encoder Layer (Module 3), which consists of a stack of N identical encoder blocks (e.g., $N=4$ or 6). Each block contains a Multi-Head Self-Attention (MHSA) mechanism and a position-wise Feed-Forward Network (FFN). The MHSA allows each embedding to attend to all other embeddings in the sequence, capturing complex, long-range temporal dependencies. Layer normalization and residual connections ensure stable gradient flow during training. The output of the final encoder block is a sequence of context-rich representations.

E. Attention Heads

Within the MHSA, multiple attention heads (e.g., $h=8$) operate in parallel. Each head projects the input sequence into different query, key, and value subspaces, allowing the model to attend to different aspects of the temporal dynamics simultaneously. The outputs of all heads are concatenated and linearly transformed back to the d_{model} dimension.

Following the transformer stack, the Temporal Attention Mechanism (Module 4) applies a final attention layer over the sequence of context-rich representations. This layer learns to assign higher weights to the most discriminative time steps, aggregating the sequence into a single, fixed-size feature vector that summarizes the entire epoch.

F. Classification Layers

The aggregated feature vector is passed to the Dual Output Head (Module 5). This consists of two parallel fully connected (dense) networks. The Intent Classification Head maps the feature vector to the predefined intent classes (e.g., left hand, right hand, rest) using a softmax activation. Simultaneously, the Emotion Classification Head maps the same feature vector to emotional states (e.g., positive, negative, neutral) using its own softmax activation.

G. Real-Time Inference Design

To achieve real-time performance, NTNDA is designed with several optimization strategies:

- **Streaming Processing:** The system processes data in a continuous stream using overlapping epochs, providing frequent classification updates (e.g., every 500 ms).
- **Model Optimization:** Techniques such as model quantization (reducing the precision of weights) and pruning (removing redundant connections) can be applied to reduce the computational footprint of the transformer model without significantly sacrificing accuracy.
- **Edge Deployment:** For wearable BCI applications, the optimized NTNDA model can be deployed on edge devices (e.g., specialized AI accelerators or mobile processors) to minimize latency and ensure privacy by processing data locally.

VI. MATHEMATICAL FORMULATION

This section provides the mathematical underpinnings for key processes within the Neuroba Transformer-Based Neural Decoding Architecture (NTNDA), focusing on the attention mechanism, transformer encoding, and the multi-task classification loss functions.

A. Attention Mechanism (Scaled Dot-Product Attention)

The core of the transformer architecture is the scaled dot-product attention mechanism. Given a sequence of input embeddings, the mechanism computes a weighted sum of values, where the weight assigned to each value is determined by the compatibility of a query with the corresponding key.

Let Q (Queries), K (Keys), and V (Values) be matrices representing the input sequence, projected into different subspaces. The attention output is computed as:

$$ \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V $$

where d_k is the dimension of the keys (and queries). The term $\sqrt{d_k}$ is a scaling factor that prevents the dot products from growing too large, which could push the softmax function into regions with extremely small gradients.

In Multi-Head Self-Attention (MHSA), this process is performed in parallel by h different heads:

$$ \text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O $$

where $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$, and W_i^Q, W_i^K, W_i^V, W^O are learned parameter matrices.

B. Transformer Encoding

A single transformer encoder block consists of the MHSA mechanism followed by a position-wise Feed-Forward Network (FFN), with residual connections and layer normalization (LayerNorm) applied around each sub-layer.

Let X be the input sequence to the encoder block. The output of the MHSA sub-layer is:

$$ Z = \text{LayerNorm}(X + \text{MHSA}(X, X, X)) $$

The FFN consists of two linear transformations with a ReLU activation in between:

$$ \text{FFN}(Z) = \max(0, ZW_1 + b_1)W_2 + b_2 $$

The final output of the encoder block is:

$$ Y = \text{LayerNorm}(Z + \text{FFN}(Z)) $$

This process is repeated for N stacked encoder blocks.

C. Temporal Embedding Representation

Before entering the transformer, the raw EEG epoch $E \in \mathbb{R}^{C \times T}$ (where C is the number of channels and T is the number of time points) is transformed into a sequence of embeddings $X \in \mathbb{R}^{L \times d_{model}}$ (where L is the sequence length).

This is achieved through a spatial-temporal convolution block f_{conv} and a linear projection W_{proj}:

$$ X_{conv} = f_{conv}(E) $$
$$ X_{proj} = X_{conv}W_{proj} $$

To retain sequence order, positional encodings PE are added:

$$ X = X_{proj} + PE $$

where $PE_{(pos, 2i)} = \sin(pos / 10000^{2i/d_{model}})$ and $PE_{(pos, 2i+1)} = \cos(pos / 10000^{2i/d_{model}})$.

D. Signal Feature Mapping (Temporal Attention)

The output of the final transformer encoder block is a sequence of context-aware representations $Y \in \mathbb{R}^{L \times d_{model}}$. The Temporal Attention Mechanism (Module 4) aggregates this sequence into a single vector $v \in \mathbb{R}^{d_{model}}$.

A learned attention vector w_a computes a score α_i for each time step i:

$$ \alpha_i = \frac{\exp(Y_i \cdot w_a)}{\sum_{j=1}^{L} \exp(Y_j \cdot w_a)} $$

The aggregated feature vector v is the weighted sum of the representations:

$$ v = \sum_{i=1}^{L} \alpha_i Y_i $$

E. Classification Loss Functions

NTNDA employs a multi-task learning approach, simultaneously optimizing for intent and emotion classification.

Let y_{intent} and $\hat{y}_{intent}$ be the true and predicted intent labels, respectively. The intent classification loss is typically the categorical cross-entropy:

$$ \mathcal{L}_{intent} = -\sum_{c=1}^{C_{intent}} y_{intent, c} \log(\hat{y}_{intent, c}) $$

Similarly, let $y_{emotion}$ and $\hat{y}_{emotion}$ be the true and predicted emotion labels. The emotion classification loss is:

$$ \mathcal{L}_{emotion} = -\sum_{c=1}^{C_{emotion}} y_{emotion, c} \log(\hat{y}_{emotion, c}) $$

The total multi-task loss $\mathcal{L}_{total}$ is a weighted sum of the individual losses:

$$ \mathcal{L}_{total} = \lambda_{intent} \mathcal{L}_{intent} + \lambda_{emotion} \mathcal{L}_{emotion} $$

where λ_{intent} and $\lambda_{emotion}$ are hyperparameters that balance the contribution of each task during training.

VII. INTENT & EMOTION DECODING MODEL

The NTNDA framework utilizes a sophisticated multi-task learning model to simultaneously decode user intent and emotional state. This approach offers several advantages over training separate models for each task.

A. Multi-Task Learning Approach

Multi-task learning (MTL) aims to improve the generalization performance of a model by leveraging domain-specific information contained in the training signals of related tasks [53]. In the context of NTNDA, intent and emotion are considered related cognitive processes that often co-occur and influence each other. By training a single model to predict both simultaneously, the model is forced to learn a more robust and generalized representation of the underlying neural dynamics that is beneficial for both tasks.

B. Shared Encoder vs. Dual Heads

The NTNDA model architecture employs a shared encoder (Modules 1-4) and dual output heads (Module 5). The shared encoder, comprising the preprocessing, embedding, and transformer layers, is responsible for extracting high-level, context-aware features from the raw EEG data. These features represent a generalized understanding of the user's cognitive state.

The dual output heads then take this shared representation and specialize it for their respective tasks. The Intent Head maps the features to specific commands or actions, while the Emotion Head maps them to affective states. This architecture balances the benefits of shared feature learning with the need for task-specific specialization, reducing overall computational complexity compared to two independent models.

C. Cross-Subject Generalization Strategy

A major challenge in BCI decoding is inter-subject variability. NTNDA addresses this through several strategies:

22 **Robust Feature Extraction:** The transformer architecture, with its self-attention mechanism, is inherently better at capturing invariant temporal patterns across subjects compared to traditional methods.
23 **Data Augmentation:** Techniques such as adding noise, time-shifting, or channel-mixing can be applied during training to artificially increase the diversity of the dataset and improve the model's robustness to subject-specific variations.
24 **Domain Adaptation:** Advanced techniques like adversarial training can be employed to align the feature distributions of different subjects in the shared representation space, encouraging the model to learn subject-invariant features [54].

D. Personalization Layer

While cross-subject generalization is crucial for out-of-the-box usability, optimal performance often requires some degree of personalization. NTNDA can incorporate a personalization layer, typically implemented by fine-tuning the final classification heads (or the last few layers of the encoder) using a small amount of calibration data from the specific user. This allows the model to adapt to the unique neural signatures of the individual while retaining the generalized knowledge learned from the broader population.

VIII. REAL-TIME DECODING FRAMEWORK

Deploying NTNDA in real-world BCI applications requires a framework optimized for real-time inference, balancing decoding accuracy with strict latency constraints.

A. Latency Constraints

Real-time BCI systems demand low latency to ensure a seamless and responsive user experience. For applications like neuroprosthetic control or interactive gaming, the delay between the user's intent and the system's response must be minimal (typically < 100-300 ms) to avoid frustration and maintain a sense of agency [55]. NTNDA's architecture must be optimized to process incoming EEG epochs and generate classifications within these tight timeframes.

B. Streaming EEG Processing

To achieve low latency, NTNDA employs a streaming processing approach. Instead of waiting for a long, discrete trial to complete, the system continuously buffers incoming EEG data and processes it in overlapping sliding windows (epochs). For example, a 1-second window might slide forward by 100 ms at a time, providing a new classification output every 100 ms. This continuous stream of predictions allows for rapid detection of changes in intent or emotion.

C. Edge vs. Cloud Inference

The deployment of NTNDA can be tailored to the specific application requirements:

- **Edge Inference:** For mobile or wearable BCIs, deploying the optimized NTNDA model directly on the edge device (e.g., a smartphone or a dedicated AI chip on the headset) is preferred. This minimizes latency by eliminating the need for data transmission to a remote server and enhances user privacy by keeping sensitive neural data local.
- **Cloud Inference:** For applications requiring the highest possible accuracy or involving very large, complex models that exceed edge device capabilities, cloud inference can be utilized. However, this approach introduces potential latency issues due to network transmission and raises privacy concerns regarding the transfer of neural data.

D. Optimization Strategies

To facilitate edge deployment and meet real-time constraints, several optimization strategies can be applied to the NTNDA model:

- **Model Quantization:** Reducing the precision of the model's weights and activations (e.g., from 32-bit floating-point to 8-bit integer) significantly reduces memory footprint and accelerates computation with minimal loss in accuracy [56].
- **Knowledge Distillation:** Training a smaller, faster "student" model to mimic the behavior of a larger, more complex "teacher" model (the full NTNDA).
- **Efficient Attention Mechanisms:** Exploring variants of the standard self-attention mechanism (e.g., sparse attention, linear attention) that reduce the quadratic computational complexity, making the transformer more efficient for long sequences [57].

IX. APPLICATIONS

The robust, real-time decoding capabilities of the Neuroba Transformer-Based Neural Decoding Architecture (NTNDA) enable a wide range of transformative BCI applications.

A. Assistive Communication

For individuals with severe motor impairments (e.g., ALS, locked-in syndrome), NTNDA can power advanced communication interfaces. By accurately decoding intent (e.g., selecting letters or words) and simultaneously monitoring emotional state (e.g., detecting frustration or fatigue), the system can dynamically adjust its interface, pacing, or predictive text algorithms to optimize the user's communication experience and reduce cognitive load.

B. Neuroprosthetics

In neuroprosthetics, precise intent decoding is critical for controlling robotic limbs or exoskeletons. NTNDA's ability to model complex temporal dynamics can improve the fluidity and naturalness of prosthetic control. Furthermore, integrating emotion recognition can provide valuable feedback; for instance, detecting pain or discomfort could automatically trigger a safety halt or adjust the prosthetic's grip force.

C. Mental Health Monitoring

Continuous, real-time emotion classification using NTNDA offers a powerful tool for mental health monitoring and intervention. Wearable BCIs could track affective states throughout the day, providing objective data for diagnosing and managing conditions like depression, anxiety, or bipolar disorder. This continuous monitoring could also trigger timely interventions, such as personalized neurofeedback or alerts to healthcare providers.

D. Human-AI Interaction Systems

As AI systems become more integrated into daily life, NTNDA can facilitate more intuitive and empathetic human-AI interaction. An AI assistant equipped with NTNDA could understand not only the user's explicit commands (intent) but also their underlying emotional state, allowing it to respond more appropriately, offer support when the user is stressed, or adjust its communication style based on the user's mood.

E. Adaptive User Interfaces

In computing and gaming, NTNDA can drive adaptive user interfaces that respond dynamically to the user's cognitive and affective state. A learning environment could adjust the difficulty of tasks based on the student's level of focus and frustration. A video game could alter its narrative or gameplay mechanics in real-time to maximize player engagement and emotional impact.

F. Neuroadaptive Systems

More broadly, NTNDA serves as a core component for neuroadaptive systems environments or technologies that automatically adapt to the user's neural state. This could range from smart home environments that adjust lighting and music based on the user's mood to high-stakes professional settings (e.g., air traffic control, surgery) where the system monitors operator fatigue and cognitive workload to prevent errors and optimize performance.

X. CHALLENGES AND LIMITATIONS

While NTNDA represents a significant advancement in neural decoding, its development and deployment face several challenges and limitations that must be addressed.

A. EEG Noise and Artifacts

Despite the robust preprocessing provided by Layer 01 (SIGNAL), residual noise and artifacts remain a significant challenge. The low signal-to-noise ratio of EEG means that subtle neural patterns can still be obscured, particularly in dynamic real-world environments. The transformer model must be highly resilient to these perturbations, and continuous improvements in artifact suppression techniques are necessary to maximize decoding accuracy.

B. Computational Constraints

Transformer architectures are inherently computationally intensive, particularly regarding memory usage and processing power. Deploying large, multi-layer transformers for real-time inference on low-power edge devices (e.g., wearable BCI headsets) is a major engineering hurdle. Balancing model complexity with the strict latency and power constraints of real-world applications requires ongoing research in model optimization and efficient hardware design.

C. Dataset Limitations

Deep learning models, especially transformers, require massive amounts of labeled data to train effectively and generalize well. In the BCI domain, acquiring large, diverse, and high-quality datasets is notoriously difficult, time-consuming, and expensive. The lack of standardized, large-scale datasets for intent and emotion classification limits the ability to fully train and validate complex models like NTNDA, often leading to overfitting on smaller, specific datasets.

D. Ethical Concerns

The ability to accurately decode user intent and emotional states raises profound ethical questions. Issues of cognitive liberty, mental privacy, and the potential for manipulation must be carefully considered. The deployment of systems capable of inferring intimate cognitive states necessitates robust ethical frameworks, transparent user consent mechanisms, and clear guidelines on how such data can be used and shared.

E. Privacy of Neural Data

Neural data is highly sensitive and uniquely identifiable. The continuous collection and processing of this data by NTNDA pose significant privacy risks. Ensuring the secure storage, transmission, and processing of neural data is paramount. Techniques such as federated learning, where the model is trained locally on the user's device without transmitting raw data, and differential privacy must be explored to protect user confidentiality.

F. Real-Time Deployment Issues

Transitioning from offline validation to real-time, closed-loop deployment introduces numerous practical challenges. System latency, hardware reliability, user comfort, and the stability of the electrode-skin interface over extended periods all impact the real-world performance of the BCI. Ensuring that NTNDA operates reliably and consistently outside the laboratory environment requires rigorous field testing and iterative refinement.

XI. RELATIONSHIP TO NEUROBA NCTS FRAMEWORK

The Neuroba Transformer-Based Neural Decoding Architecture (NTNDA) is the defining component of **Layer 02: DECODE** within the Neuroba Neural Communication and Translation System (NCTS) Framework. The NCTS is a conceptual, layered architecture designed to facilitate seamless brain-to-device communication. NTNDA's role is to bridge the gap between raw physiological signals and actionable semantic meaning.

A. Semantic Interpretation of Neural Signals

As Layer 02, NTNDA is responsible for the semantic interpretation of the neural data provided by Layer 01. It transforms the standardized, high-fidelity signal stream into a structured understanding of the user's cognitive state. This involves moving beyond simple feature extraction to a higher-level comprehension of what the neural patterns represent in terms of human intent and emotion.

B. Mapping Signals to Intent/Emotion

The core function of NTNDA within the NCTS Framework is to map the complex, multi-dimensional EEG signals to specific, predefined categories of intent and emotional states. By employing the multi-task transformer architecture, NTNDA provides a robust and generalized mapping that can adapt to different users and contexts, forming the basis for subsequent system actions.

C. Interface with Layer 01 (SIGNAL)

NTNDA (Layer 02) relies entirely on the output of Layer 01 (SIGNAL) [Neuroba Research (2026a)]. Layer 01 provides the clean, standardized, and synchronized multimodal data stream that NTNDA requires for accurate decoding. The interface between these layers is defined by strict data formatting protocols and quality metrics, ensuring that NTNDA only processes reliable neural information.

D. Interface with Layer 03 (TRANSMIT)

The output of NTNDA the classified intent and emotional state, along with associated confidence scores serves as the direct input to Layer 03 (TRANSMIT). Layer 03 is responsible for formatting these semantic interpretations into standardized communication protocols and transmitting them to external devices or applications. NTNDA ensures that the information passed to Layer 03 is accurate, timely, and semantically meaningful.

E. Frame NCTS as a Conceptual Architecture

It is important to emphasize that the Neuroba NCTS Framework, including NTNDA, is currently a conceptual architecture and a roadmap for future research and development. While the individual components and algorithms (e.g., transformers, adaptive filtering) are based on established scientific principles, the fully integrated, real-time, real-world system described by the NCTS Framework represents an aspirational goal that requires significant ongoing research and engineering effort to fully realize.

XII. FUTURE WORK

The development of NTNDA opens several exciting avenues for future research within the Neuroba NCTS Framework.

A. Multimodal Neural Decoding (EEG + fNIRS)

Future iterations of NTNDA will focus on true multimodal decoding, integrating EEG with other modalities like functional Near-Infrared Spectroscopy (fNIRS). While EEG provides high temporal resolution, fNIRS offers superior spatial resolution and robustness to certain artifacts. Developing transformer architectures capable of effectively fusing and decoding these disparate data streams will significantly enhance the accuracy and reliability of intent and emotion classification.

B. Larger Transformer Models for Brain Signals

As computational resources and dataset sizes grow, exploring larger, more complex transformer models specifically tailored for brain signals is a key priority. These models could potentially capture even deeper, more abstract representations of cognitive states, leading to unprecedented decoding performance. Research will focus on optimizing these large models for efficient training and inference in the BCI domain.

C. Foundation Models for Brain Data

A major goal is the development of "foundation models" for brain data large-scale transformer models pre-trained on vast, diverse datasets of neural signals across various tasks and subjects. These foundation models could then be fine-tuned with minimal data for specific BCI applications (e.g., a specific user's intent decoding), drastically reducing calibration time and improving cross-subject generalization.

D. Real-World Deployment Systems

Transitioning NTNDA from conceptual design and offline validation to robust, real-world deployment systems is a critical next step. This involves developing optimized edge-computing hardware, refining real-time streaming protocols, and conducting extensive longitudinal studies in naturalistic environments to validate the system's performance, usability, and long-term stability.

E. Cross-Brain Generalization

Achieving true cross-brain generalization where a model trained on a population can seamlessly decode the intent and emotion of a completely new user with zero calibration remains the holy grail of BCI research. Future work will explore advanced domain adaptation, meta-learning, and subject-invariant feature learning techniques within the NTNDA framework to move closer to this goal.

XIII. CONCLUSION

The accurate and real-time decoding of complex neural signals is a critical prerequisite for the widespread adoption of Brain–Computer Interfaces. This paper has introduced the **Neuroba Transformer-Based Neural Decoding Architecture (NTNDA)**, a novel conceptual framework designed to address the limitations of traditional decoding methods. By leveraging the power of transformer networks and self-attention mechanisms, NTNDA offers a robust solution for the simultaneous classification of user intent and emotional states from EEG signals.

The proposed architecture, encompassing preprocessing, feature embedding, transformer encoding, temporal attention, and a dual output head, provides a comprehensive pipeline for modeling the intricate temporal dynamics and long-range dependencies inherent in neural data. The multi-task learning approach enhances generalization, while the focus on real-time inference optimization ensures applicability in dynamic, real-world scenarios.

As Layer 02 (DECODE) of the Neuroba NCTS Framework, NTNDA plays a pivotal role in translating the high-fidelity signals acquired by Layer 01 [Neuroba Research (2026a)] into actionable semantic meaning. While challenges regarding computational constraints, dataset limitations, and ethical considerations remain, NTNDA represents a significant theoretical advancement. Future research focusing on multimodal integration, foundation models, and cross-brain generalization will further solidify NTNDA's position as a cornerstone technology for next-generation, adaptive, and intuitive brain-computer ecosystems.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767-791, 2002.

[2] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update,"*Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[3] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis,"*IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41-56, 2007.

[4] 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.

[5] V. J. Lawhern, A. J. Huber, J. N. Wu, D. Rule, and K. Robbins, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces,"*Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[6] M. A. Lebedev and J. R. Wolpaw, "Brain-machine interfaces: past, present and future,"*Trends in Neurosciences*, vol. 29, no. 9, pp. 536-546, 2006.

[7] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey,"*IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374-393, 2017.

[8] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review,"*Journal of Neural Engineering*, vol. 16, no. 3, p. 031001, 2019.

[9] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: a review,"*Sensors*, vol. 19, no. 5, p. 987, 2019.

[10] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain-computer interface: a brief review,"*Journal of Neuroscience Methods*, vol. 12, no. 6, p. 062001, 2015.

[11] H. Yang, S. Fazli, and K.-R. Müller, "Deep learning for EEG-based brain-computer interfaces: A review,"*IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 1-1, 2019.

[12] P. Bashivan, I. Rish, M. Yeagle, and B. Riedner, "Learning representations from EEG with deep recurrent-convolutional neural networks,"*arXiv preprint arXiv:1511.06448*, 2015.

[13] S. Sanei and J. A. Chambers,*EEG Signal Processing*. John Wiley & Sons, 2013.

[14] P. L. Nunez and R. Srinivasan,*Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in*Advances in Neural Information Processing Systems*, 2017, pp. 5998-6008.

[16] E. Vafaei and M. Hosseini, "Transformers in EEG analysis: a review of architectures and applications in motor imagery, seizure, and emotion classification,"*Sensors*, vol. 25, no. 5, p. 1293, 2025.

[17] 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. 6, no. 3, pp. 162-175, 2015.

[18] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening,"*IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798-1806, 2010.

[19] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks,"*IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532-541, 2018.

[20] Y. Li, W. Zheng, Z. Cui, T. Zhang, and Y. Zong, "A novel neural network model based on cerebral hemisphere asymmetry for EEG emotion recognition," in*Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 1561-1567.

[21] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles,"*Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842-1857, 1999.

[22] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement,"*IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441-446, 2000.

[23] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,"*Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510-523, 1988.

[24] J. d. R. Millán, R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb, et al., "Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges,"*Frontiers in Neuroscience*, vol. 4, p. 161, 2010.

[25] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals,"*Journal of Neural Engineering*, vol. 14, no. 1, p. 016003, 2016.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory,"*Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[27] D. Zhang, L. Yao, X. Chen, S. Wang, J. Chang, and Y. Fang, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition,"*IEEE Transactions on Cybernetics*, vol. 49, no. 8, pp. 2877-2889, 2019.

[28] C. Wang, B. L. Lu, and J. R. Duann, "EEG-based emotion recognition using deep learning," in*Neural Information Processing*, Springer, 2015, pp. 1-8.

[29] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate,"*arXiv preprint arXiv:1409.0473*, 2014.

[30] J. Wang, "Attention mechanisms in brain–computer interfaces,"*arXiv preprint arXiv:2502.19281*, 2025.

[31] Y. Zhang, C. S. Nam, A. Zhou, J. Yin, D. Macwan, and A. R. Paiva, "Attention-based deep learning for EEG-based brain-computer interfaces,"*IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1-10, 2021.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in*Advances in Neural Information Processing Systems*, 2017, pp. 5998-6008.

[33] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Qi, J. Wang, and L. Sun, "Transformers in time series: A survey,"*arXiv preprint arXiv:2202.07125*, 2022.

[34] H. Zhang and H. Li, "Transformer-based EEG decoding: A survey,"*arXiv preprint arXiv:2507.02320*, 2025.

[35] Y. Ding, "EmT: A Novel Transformer for Generalized Cross-Subject EEG Emotion Classification,"*IEEE Transactions on Affective Computing*, vol. 16, pp. 1-10, 2025.

[36] W. Liao, "Advancing BCI with a transformer-based model for motor imagery decoding,"*Journal of Neural Engineering*, vol. 22, p. 016003, 2025.

[37] G. Pfurtscheller, B. Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, and N. Birbaumer, "The hybrid BCI,"*Frontiers in Neuroscience*, vol. 4, p. 3, 2010.

[38] S. Fazli, M. Danóczy, J. Scheer, M. O. von Bünau, K.-R. Müller, and B. Blankertz, "Enhanced performance by a hybrid NIRS-EEG brain computer interface,"*NeuroImage*, vol. 59, no. 1, pp. 519-529, 2012.

[39] R. J. Deligani, M. S. Al-Quraishi, I. Elamvazuthi, T. B. Tang, M. Al-Qurishi, and S. A. Al-Quraishi, "Multimodal fusion of EEG-fNIRS: a mutual information-based approach,"*IEEE Access*, vol. 9, pp.

1-10, 2021.

[40] W. U. R. Qamar, "Multi-scale EEG feature decoding with Swin Transformers for real-time BCI,"*Scientific Reports*, vol. 16, p. 12820037, 2026.

[41] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review,"*Journal of Neural Engineering*, vol. 16, no. 3, p. 031001, 2019.

[42] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain-computer interface: a brief review,"*Journal of Neuroscience Methods*, vol. 12, no. 6, p. 062001, 2015.

[43] Y. Ding, "Neural decoding for EEG-BCI: from conventional machine learning to deep learning,"*Artificial Intelligence in Medicine*, vol. 150, p. 102838, 2026.

[44] S. Sanei and J. A. Chambers,*EEG Signal Processing*. John Wiley & Sons, 2013.

[45] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain-computer interfaces,"*Proceedings of the IEEE*, vol. 100, no. 10, pp. 1567-1584, 2012.

[46] P. L. Nunez and R. Srinivasan,*Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.

[47] J. Wang, "Attention mechanisms in brain–computer interfaces,"*arXiv preprint arXiv:2502.19281*, 2025.

[48] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain-computer interface: a brief review,"*Journal of Neuroscience Methods*, vol. 12, no. 6, p. 062001, 2015.

[49] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update,"*Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[50] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey,"*IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374-393, 2017.

[51] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control,"*Clinical Neurophysiology*, vol. 113, no. 6, pp. 767-791, 2002.

[52] W. U. R. Qamar, "Multi-scale EEG feature decoding with Swin Transformers for real-time BCI,"*Scientific Reports*, vol. 16, p. 12820037, 2026.

[53] R. Caruana, "Multitask learning,"*Machine Learning*, vol. 28, no. 1, pp. 41-75, 1997.

[54] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks,"*The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096-2030, 2016.

[55] M. A. Lebedev and J. R. Wolpaw, "Brain-machine interfaces: past, present and future,"*Trends in Neurosciences*, vol. 29, no. 9, pp. 536-546, 2006.

[56] W. U. R. Qamar, "Multi-scale EEG feature decoding with Swin Transformers for real-time BCI,"*Scientific Reports*, vol. 16, p. 12820037, 2026.

[57] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Qi, J. Wang, and L. Sun, "Transformers in time series: A survey,"*arXiv preprint arXiv:2202.07125*, 2022.

[58] Neuroba Research, "Adaptive Multimodal EEG Signal Acquisition for Robust Real-World Brain–Computer Interfaces,"*Neuroba NCTS Research Series*, 2026

APPENDIX: FIGURES

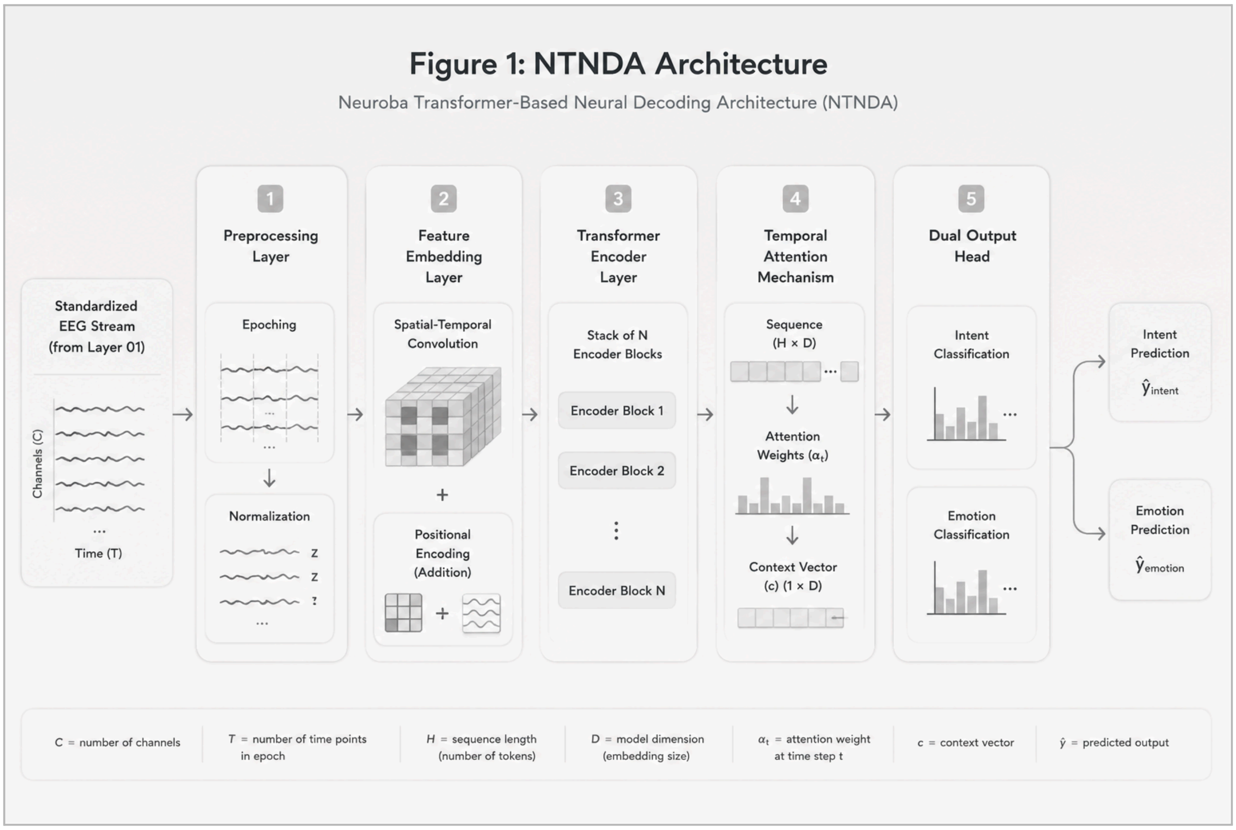


Figure 1: NTNDA Architecture

Neuroba Transformer-Based Neural Decoding Architecture (NTNDA)

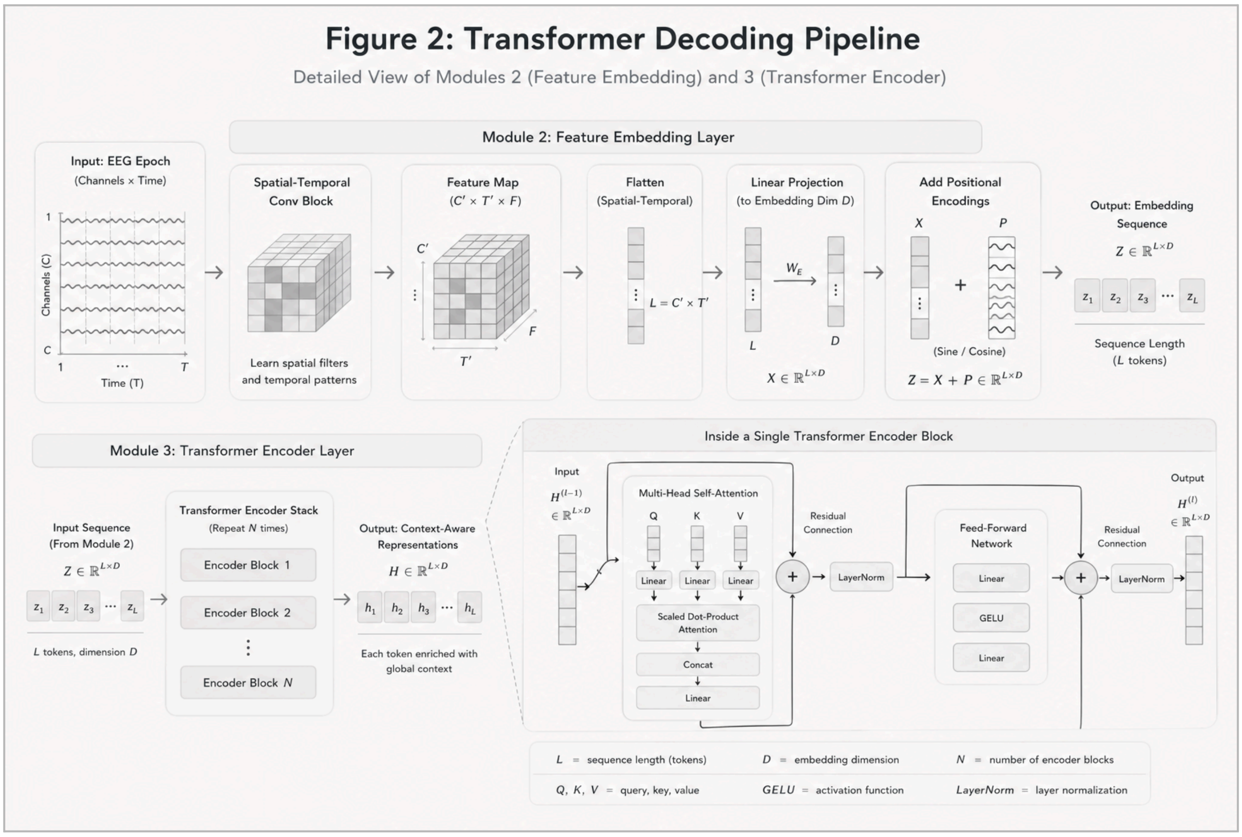


Figure 2: Transformer Decoding Pipeline

Detailed View of Modules 2 (Feature Embedding) and 3 (Transformer Encoder)

Figure 3: Attention Mechanism over EEG Sequences

How Self-Attention (within Module 3) and Temporal Attention Mechanism (Module 4) operate on EEG data

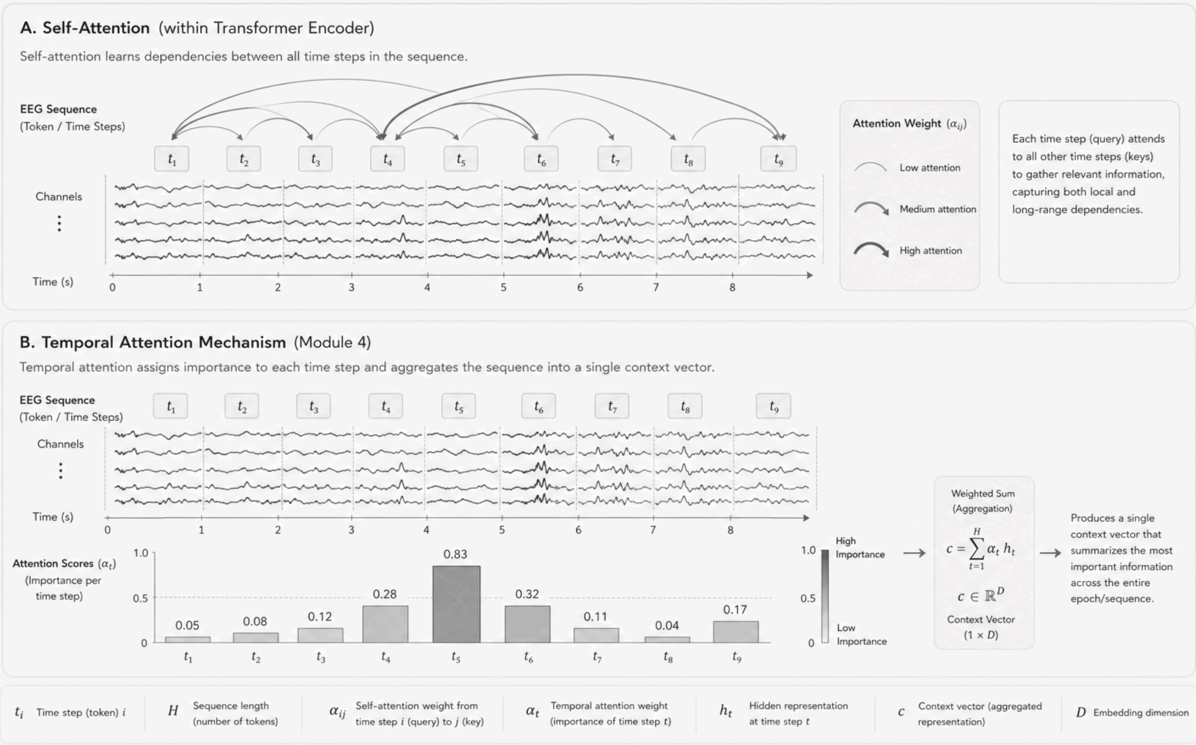
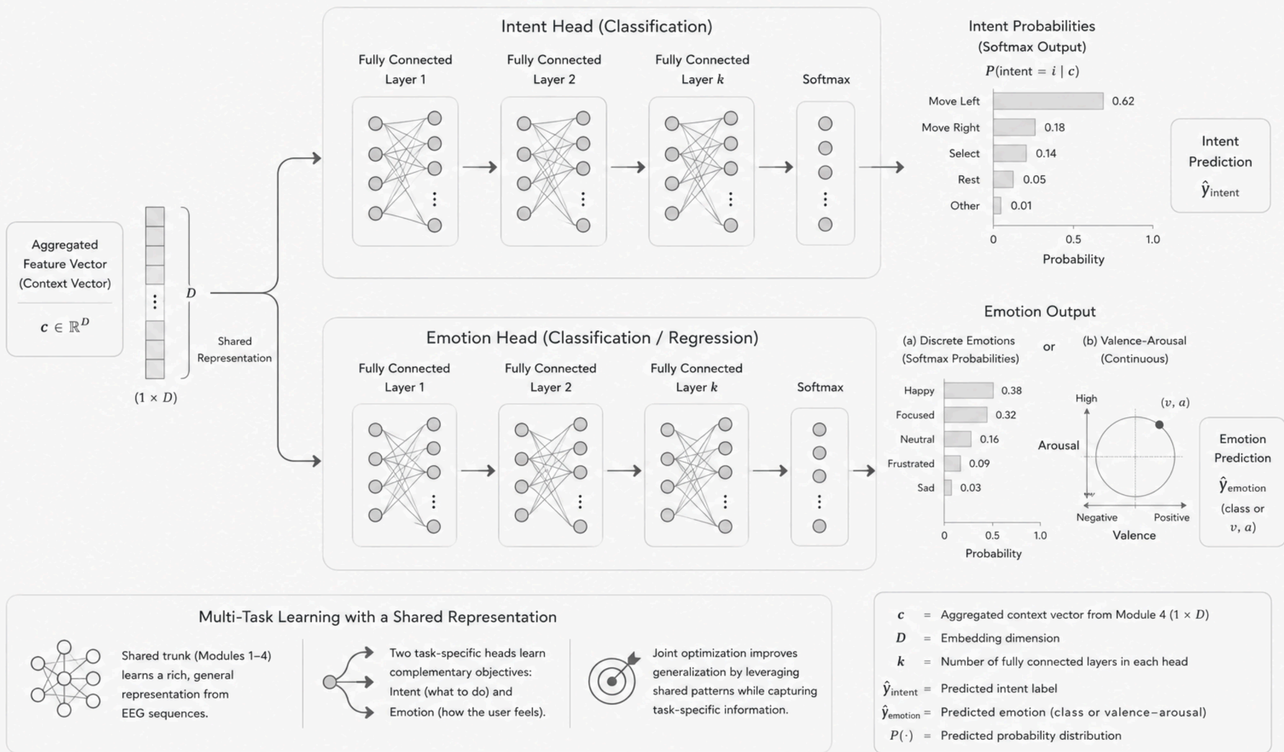


Figure 4: Dual-Head Output System (Intent + Emotion)

Module 5: Dual Output Head



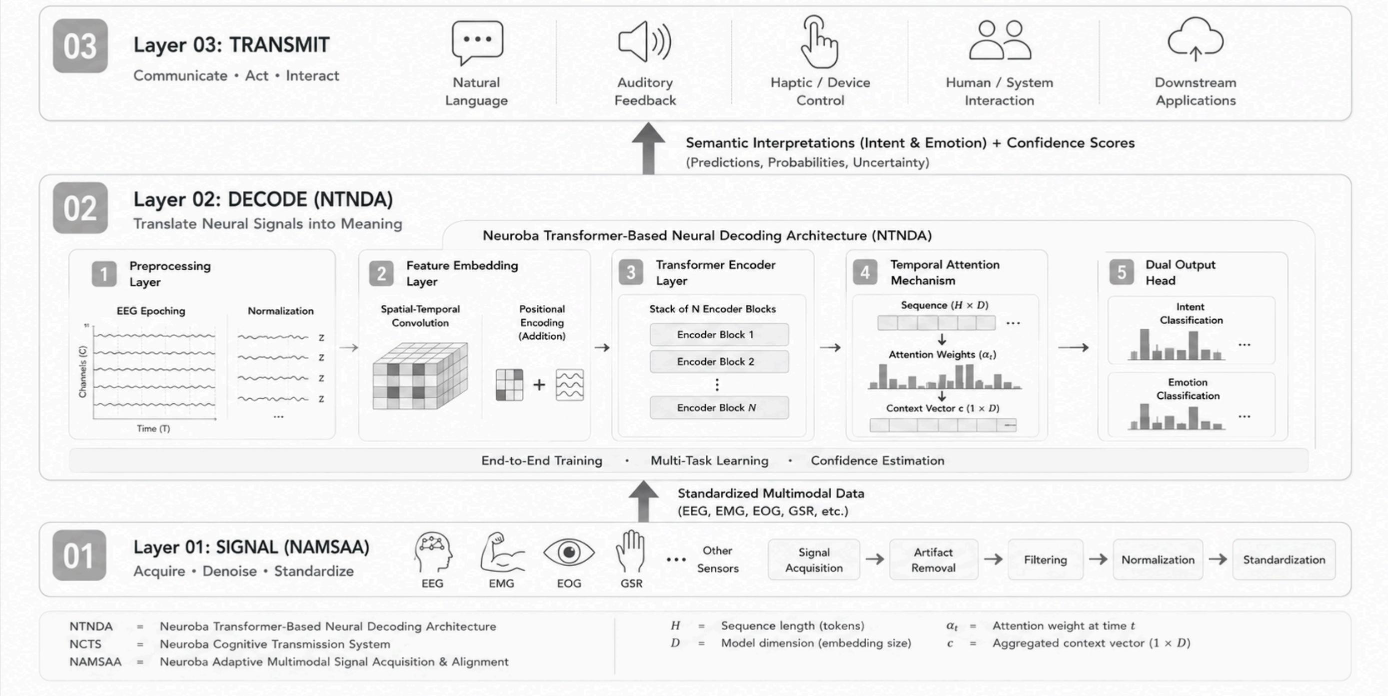


Figure 5: Integration with NCTS Layer 02

Position of NTNDA within the Neuroba NCTS Framework