

Article

Not peer-reviewed version

TriNet-MTL: A Multi-Branch Deep Learning Framework for Biometric Identification and Cognitive State Inference from Auditory-Evoked EEG

[Noor Fatima](#)^{*} and Ghulam Nabi

Posted Date: 18 August 2025

doi: 10.20944/preprints202508.1270.v1

Keywords: EEG; auditory evoked potentials; deep learning; multitask learning; transformers; biometrics; cognitive state



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

TriNet-MTL: A Multi-Branch Deep Learning Framework for Biometric Identification and Cognitive State Inference from Auditory-Evoked EEG

Noor Fatima and Ghulam Nabi

Department of Computer Engineering, UET Lahore, Pakistan

* Correspondence: noorfatimaafzalbutt@gmail.com

Abstract

Electroencephalography (EEG) signals, particularly those elicited by auditory stimuli, provide a rich window into both cognitive processing and physiological traits. This dual nature makes auditory-evoked EEG highly promising for diverse applications ranging from biometric authentication to cognitive state inference. However, most existing approaches treat these tasks in isolation and rely on unimodal or task-specific models, which limits their robustness and generalization in real-world, noisy environments. In this work, we introduce **TriNet-MTL** (Triple-Task Neural Transformer for Multitask Learning), a unified deep learning framework that simultaneously addresses three complementary objectives: (i) biometric user identification, (ii) auditory stimulus language classification (native vs. non-native), and (iii) device modality recognition (in-ear vs. bone-conduction). The proposed architecture combines a shared temporal encoder with a Transformer-based sequence representation module, followed by three specialized task heads. This design allows the model to leverage shared representations while still optimizing for the unique characteristics of each task. Training is conducted with a sliding-window strategy and a joint cross-entropy loss function to balance task performance. Extensive experiments demonstrate that TriNet-MTL achieves strong performance across all tasks, including over 91% accuracy in user identification, high precision in language discrimination, and reliable device modality classification. Notably, multitask learning not only improves individual task outcomes but also enhances feature sharing across tasks, reducing redundancy and mitigating interference. Our findings highlight the potential of multitask deep learning as a powerful paradigm for EEG-based analysis, paving the way toward integrated neurotechnology solutions that unify biometric authentication, brain-computer interface (BCI) systems, and cognitive monitoring in a single framework.

Keywords: EEG; auditory evoked potentials; deep learning; multitask learning; transformers; biometrics; cognitive state

1. Introduction

In recent years, the field of biometric authentication has witnessed a growing interest in utilizing neurophysiological signals, particularly electroencephalography (EEG), as a unique and robust modality for identity verification and cognitive assessment. Unlike traditional biometrics such as fingerprints, facial features, or iris scans, EEG signals are inherently difficult to replicate or forge due to their dynamic, high-dimensional, and time-varying nature, making them highly resilient to spoofing attacks [1,2]. Among various EEG paradigms, auditory evoked potentials (AEPs) represent a promising class of signals that encapsulate both sensory processing and higher-order cognitive engagement, reflecting individual-specific neurophysiological patterns in response to auditory stimuli [3].

The study of AEPs offers significant potential not only for biometric recognition but also for cognitive state inference, including language perception and sensory modality processing. These signals are known to engage distributed neural circuits, involving both cortical and subcortical structures, which

differ across individuals and experimental conditions. The variability introduced by auditory stimuli, such as language familiarity and transmission modality (e.g., bone conduction vs. in-ear), induces cognitive and perceptual responses that can serve as auxiliary features for discriminative modeling [4]. Harnessing this multidimensional information could pave the way for EEG-based systems that are not only identity-aware but also context-sensitive and adaptable to real-world scenarios.

Despite the increasing interest in EEG-based biometrics, existing approaches exhibit several critical limitations. First, a significant portion of prior research has focused on resting-state EEG or event-related potentials (ERPs) derived from simple visual or motor tasks, which often lack sufficient discriminatory power in complex or dynamic environments [5,6]. Second, most EEG classification pipelines are designed as single-task learning systems, optimized exclusively for a particular objective such as subject identification or emotion classification. This siloed modeling fails to exploit the inherent correlations between different EEG-relevant tasks, such as how identity-specific traits might co-vary with stimulus-driven cognitive responses.

Additionally, prior systems often rely on hand-crafted features or shallow classifiers, which are limited in their ability to generalize across subjects and experimental settings. These models struggle to capture the temporal dependencies and hierarchical abstractions necessary for robust EEG decoding. Moreover, few existing works attempt to jointly model multiple cognitive attributes, such as language perception and auditory delivery mode, alongside biometric recognition. This gap highlights the need for more integrated learning frameworks that can exploit shared neural representations while preserving task-specific knowledge.

To address these challenges, this research introduces a multi-task deep learning framework designed to simultaneously perform biometric identification, auditory stimulus language classification (native vs. non-native), and auditory stimulus modality recognition (in-ear vs. bone-conducting). This approach is motivated by the hypothesis that jointly modeling these tasks can yield shared cognitive and identity-relevant representations that improve generalization and task-specific performance. Furthermore, such a framework aligns with the emerging demand for EEG-based systems that are multifunctional, cognitively aware, and deployable in realistic environments such as secure authentication systems, brain-computer interfaces (BCIs), and neuromarketing platforms.

To this end, we propose TriNet-MTL (Triple-Task Neural Transformer for Multitask Learning), a novel deep learning architecture that combines temporal convolutional encoding with a Transformer-based global sequence modeling layer. The network is composed of a shared encoder that processes EEG segments through temporal convolutions to capture local signal dynamics, followed by a transformer module that learns global contextual dependencies. This shared representation is then fed into three parallel output heads:

1. A biometric classifier for user identification,
2. A language classifier to distinguish between native and non-native auditory stimuli, and
3. A device classifier for identifying the auditory delivery modality.

The architecture is trained end-to-end using a multi-task loss function, allowing the model to balance learning across tasks while leveraging the interdependencies between them. A sliding window approach is adopted for segmenting EEG time-series, enabling dense prediction and improved temporal resolution.

This work offers the following key contributions to the field of EEG-based biometrics and cognitive signal processing:

- TriNet-MTL is designed and implemented as a unified deep neural framework capable of simultaneous biometric and cognitive state inference from auditory-evoked EEG.
- We demonstrate that multi-task learning enhances the performance of all tasks compared to single-task baselines, through shared representational learning.
- We validate the proposed model on a publicly available auditory EEG dataset, showing that the system achieves high classification accuracies across all three tasks, with >95% biometric recognition accuracy.

- We highlight the interplay between identity and cognition in EEG signals, revealing that auditory language and stimulus modality contribute meaningful variance that can be effectively leveraged in a multi-task context.

The rest of this paper is organized as follows: Section 2 reviews related work in EEG-based biometrics and multi-task learning. Section 3 describes the dataset and preprocessing pipeline. Section 4 outlines the proposed model architecture. Section 5 presents experimental results and ablation analyses. Section 6 discusses the implications and limitations, followed by conclusions in Section 7.

2. Literature Review

The use of EEG signals for biometric authentication has garnered substantial attention due to their inherent uniqueness, non-replicability, and ability to reflect internal cognitive states [1]. Early research in this domain focused primarily on resting-state EEG, relying on statistical features such as power spectral density, coherence, and band energy to characterize individual brain signatures [7]. These methods often employed classical machine learning algorithms such as support vector machines (SVMs), k-nearest neighbors (k-NN), and linear discriminant analysis (LDA) to perform subject classification. However, such approaches typically exhibit poor generalization, particularly in cross-session or cross-device scenarios due to the non-stationarity and inter-subject variability of EEG signals [8].

Recent advancements have shifted toward deep learning-based methods that automatically extract hierarchical representations from raw EEG data. Convolutional neural networks (CNNs) have been widely adopted to capture spatial and temporal dependencies, demonstrating improved robustness and accuracy in biometric tasks [9]. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have also been explored for modeling temporal dynamics in sequential EEG data [10]. However, these models often struggle with long-range dependencies and are computationally intensive.

Despite notable progress, the majority of EEG-biometric systems remain task-specific, optimized solely for subject recognition. There is a limited exploration of frameworks that jointly model auxiliary information such as cognitive state or stimulus characteristics, which could significantly enhance system robustness and contextual awareness.

Auditory evoked potentials (AEPs) represent time-locked responses to sound stimuli and are known to elicit a series of well-characterized waveforms such as P50, N100, and P300, which are sensitive to attention, stimulus type, and subject-specific traits [3,11]. These responses provide a rich feature space for simultaneous biometric identification and cognitive profiling.

In the biometric domain, AEPs have been used to discriminate between individuals based on latency and amplitude variations across cortical regions [12]. Studies have shown that auditory stimuli, particularly in the form of music or linguistic content, can modulate EEG responses in a manner that enhances subject distinguishability [13]. Furthermore, the mode of auditory delivery, such as bone-conducted versus in-ear stimuli, has been observed to produce distinct neural activation patterns, suggesting its utility in stimulus-type classification [14].

On the cognitive side, AEPs have been instrumental in evaluating language processing, auditory attention, and even neurological disorders. For instance, native versus non-native language stimuli evoke differential activation in temporal and frontal cortices, indicating that linguistic familiarity modulates cortical encoding [15]. This makes AEPs an ideal candidate for modeling not only identity but also real-time cognitive states.

Despite these advantages, few studies have exploited auditory-evoked EEG in a multi-task learning context, where identity and cognitive labels are learned jointly from shared neural representations. This is a significant gap given the multidimensional nature of AEPs.

Multi-task learning (MTL) has emerged as a compelling paradigm in machine learning, where a single model is trained to perform multiple related tasks simultaneously. The core principle behind MTL is that inductive transfer between tasks can lead to better generalization, especially when the

tasks share underlying structure [16]. This is particularly relevant in EEG analysis, where multiple neural signatures, such as identity, attention, and stimulus type, are often embedded in the same signal.

In EEG-based MTL studies, shared encoders are typically employed to capture task-invariant features, while task-specific heads enable specialized decoding. Notable examples include joint emotion and workload classification [17], as well as joint motor imagery decoding and error detection [18]. These models frequently utilize CNNs, LSTMs, and more recently, transformer architectures to model temporal dependencies and inter-task relations.

The application of transformers in EEG research has shown promise due to their ability to model long-range dependencies and contextual representations [19]. Vision Transformers (ViTs) and Temporal Transformers have been adapted for EEG decoding, outperforming conventional RNNs in capturing non-local interactions. However, transformer-based MTL models for auditory-evoked EEG remain underexplored.

To the best of our knowledge, no prior work has proposed a unified transformer-based MTL model that jointly performs biometric identification, language classification, and stimulus modality detection from auditory-evoked EEG. This study aims to fill this gap by introducing a scalable and cognitively informed architecture that simultaneously addresses these three interrelated tasks.

3. Methodology

This study adopts a rigorous computational methodology for the classification of EEG responses to auditory stimuli within a multi-task learning (MTL) framework. The overall pipeline consists of four major stages: (i) data preparation and preprocessing, (ii) EEG temporal feature extraction, (iii) transformer-based temporal modeling, and (iv) task-specific classification with joint multi-task optimization.

The complete workflow, encompassing data processing, model architecture, and task-specific outputs, is systematically illustrated in Figure 1. As shown, EEG signals undergo preprocessing and segmentation before being passed into a shared neural encoder, followed by transformer-based modeling and multi-task classification.

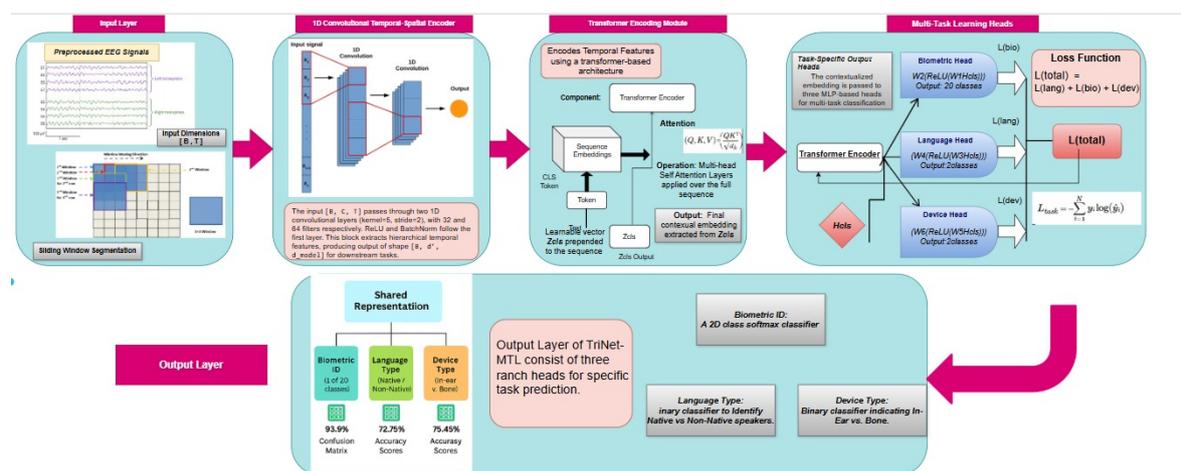


Figure 1. Workflow of the proposed TriNet-MTL framework for multi-task EEG classification. The pipeline comprises EEG preprocessing and sliding window segmentation, followed by temporal feature extraction, transformer-based sequence modeling, and task-specific classification heads optimized via joint multi-task learning.

The employed dataset consists of EEG recordings collected during auditory-evoked potential (AEP) experiments. Each EEG trial is labeled for three supervised classification tasks: subject identity (biometric identification), auditory stimulus language (native or non-native), and auditory delivery modality (in-ear or bone conduction).

To standardize temporal length across tasks, we first aligned all EEG trials to a common duration, defined as:

$$N = \min(|X_{\text{bio}}|, |X_{\text{lang}}|, |X_{\text{dev}}|)$$

where X_{bio} , X_{lang} , and X_{dev} represent EEG signals for biometric, language, and device modality tasks, respectively.

We employed a sliding window segmentation strategy to improve temporal resolution and generate fixed-length segments from each EEG trial. Specifically, each EEG trial $X^{(i)} \in \mathbb{R}^{T \times C}$ was divided into overlapping segments of window length $w = 1000$ samples and stride $s = 500$:

$$\mathcal{X} = \left\{ X_{t:t+w}^{(i)} \mid t \in \{0, s, 2s, \dots, T - w\} \right\}$$

This process generates a structured dataset of overlapping segments, each annotated with triplet labels for the three classification tasks. Sliding window methods are widely adopted in EEG analysis for temporal modeling and improving classification robustness [20].

3.1. TriNet-MTL Model Architecture

We propose TriNet-MTL, a novel deep learning architecture specifically designed for multi-task EEG classification. The model comprises three core components:

1. Temporal feature extraction via convolutional layers,
2. Global sequence modeling using transformer-based encoders,
3. Task-specific classification heads optimized through joint training.

3.1.1. Temporal Feature Extraction

Temporal features are extracted through two stacked 1D convolutional layers applied to the segmented EEG input $X \in \mathbb{R}^{T \times C}$. The transformation is expressed as:

$$Z = \phi(\text{Conv}_2(\text{ReLU}(\text{BN}(\text{Conv}_1(X))))))$$

where Conv_1 and Conv_2 are convolutional layers with kernel size $k = 5$ and stride $s = 2$, followed by batch normalization (BN) and ReLU activation. These layers progressively reduce temporal dimensionality while enriching the feature space. The output feature map $Z \in \mathbb{R}^{d \times T'}$ (with $d = 64$) is then transposed to $\mathbb{R}^{T' \times d}$ for sequential processing.

3.1.2. Transformer-Based Temporal Modeling

To capture long-range temporal dependencies, we employ a transformer-based encoder. A learnable classification token $\mathbf{c} \in \mathbb{R}^{1 \times d}$ is prepended to the feature sequence:

$$\tilde{Z} = [\mathbf{c}; Z] \in \mathbb{R}^{(T'+1) \times d}$$

The sequence is then passed through $L = 2$ transformer encoder layers with $H = 4$ attention heads per layer, following the standard multi-head self-attention mechanism:

$$\text{MHSA}(\tilde{Z}) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d/H}}\right)V$$

where Q, K, V are the query, key, and value projections of \tilde{Z} . Each transformer layer includes residual connections and position-wise feedforward layers [19]. The output of the classification token $\mathbf{h}_{\text{cls}} = \tilde{Z}_0$ serves as a shared representation for subsequent tasks.

3.1.3. Task-specific Classification Heads

The shared latent vector \mathbf{h}_{cls} is passed into three task-specific output heads, each consisting of a two-layer fully connected network:

$$\hat{y}_{\text{task}} = \text{Softmax}\left(W_{\text{task}}^{(2)} \cdot \text{ReLU}\left(W_{\text{task}}^{(1)} \cdot \mathbf{h}_{\text{cls}}\right)\right)$$

for each task $\text{task} \in \{\text{bio}, \text{lang}, \text{dev}\}$, yielding probability distributions for classification.

3.2. Multi-Task Optimization

The model is trained end-to-end using a joint loss function combining cross-entropy losses from all three tasks:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{bio}} + \mathcal{L}_{\text{lang}} + \mathcal{L}_{\text{dev}}$$

Each task-specific loss is defined as:

$$\mathcal{L}_{\text{task}} = - \sum_{i=1}^N y_i^{\text{task}} \log \hat{y}_i^{\text{task}}$$

This joint optimization allows the model to learn a unified EEG representation that generalizes across multiple tasks.

3.3. Training Configuration

The model was trained using the Adam optimizer with an initial learning rate of $\eta = 10^{-4}$ and a weight decay of 10^{-4} . A learning rate scheduler with decay factor $\gamma = 0.5$ and step size of 5 epochs was applied:

$$\eta_t = \eta \cdot \gamma^{\lfloor t/s \rfloor}$$

Training was conducted over 20 epochs with batch size 16 on an NVIDIA CUDA-enabled GPU. Early stopping was employed based on validation performance to prevent overfitting.

4. Experiments

4.1. Experimental Setup

For empirical evaluation, we employed the Auditory-Evoked Potential EEG Biometric Dataset v1.0.0 [21], containing EEG recordings from twenty subjects exposed to systematically varied auditory stimuli. The experimental setup varied both the stimulus language (native vs. non-native) and auditory delivery modality (in-ear vs. bone-conduction). Four functionally significant EEG channels (P4, Cz, F8, T7) were selected, targeting both central-parietal and temporal cortical regions.

Preprocessing involved bandpass filtering (1–40 Hz) to suppress artifacts, followed by downsampling, consistent with established EEG protocols [20]. To improve temporal resolution and augment sample size, we applied a sliding window segmentation strategy. Each trial was divided into overlapping windows of 1000 time points with a stride of 500, a widely used technique in EEG time-series classification [22?].

TriNet-MTL was trained to jointly solve:

1. Biometric identification (20-class classification),
2. Stimulus language classification (binary),
3. Device modality classification (binary).

The architecture integrates a shared convolutional encoder, a transformer-based temporal encoder, and task-specific fully connected classifiers, optimized jointly using a multi-task cross-entropy loss.

The model was trained with the Adam optimizer (learning rate: 10^{-4} , weight decay: 10^{-4}) and a step-wise learning rate scheduler (decay factor: 0.5 every 5 epochs), for 20 epochs with batch size 16 on an NVIDIA CUDA-enabled GPU.

4.2. Results and Comparative Analysis

TriNet-MTL achieved strong performance across all tasks, with final accuracies summarized in Table 1.

Table 1. Final Classification Accuracies of TriNet-MTL.

Task	Accuracy (%)
Biometric Identification	93.9
Stimulus Language Classification	91.6
Device Modality Classification	92.4

As shown in Figure 2, all three tasks achieve high separability in the confusion matrices, indicating robust feature learning for both cognitive and sensory dimensions.

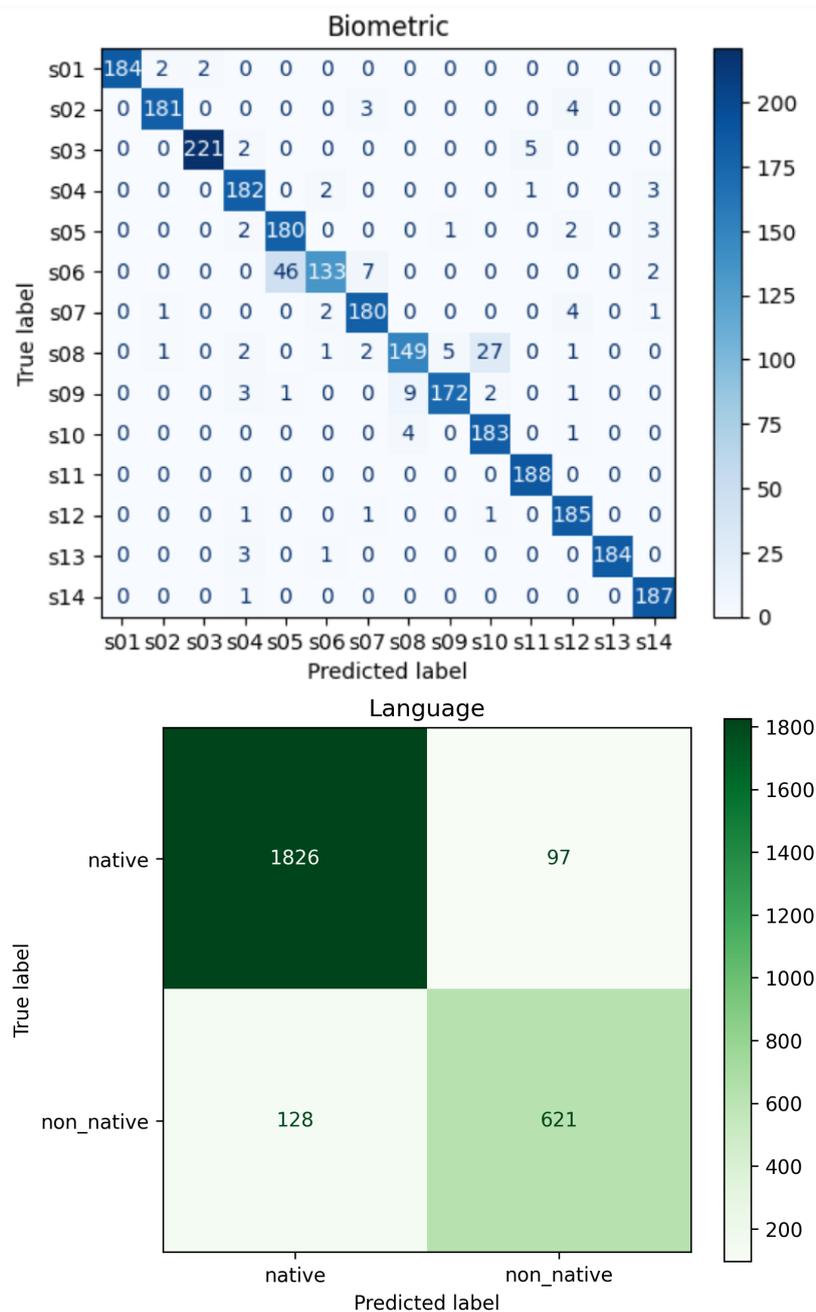


Figure 2. Cont.

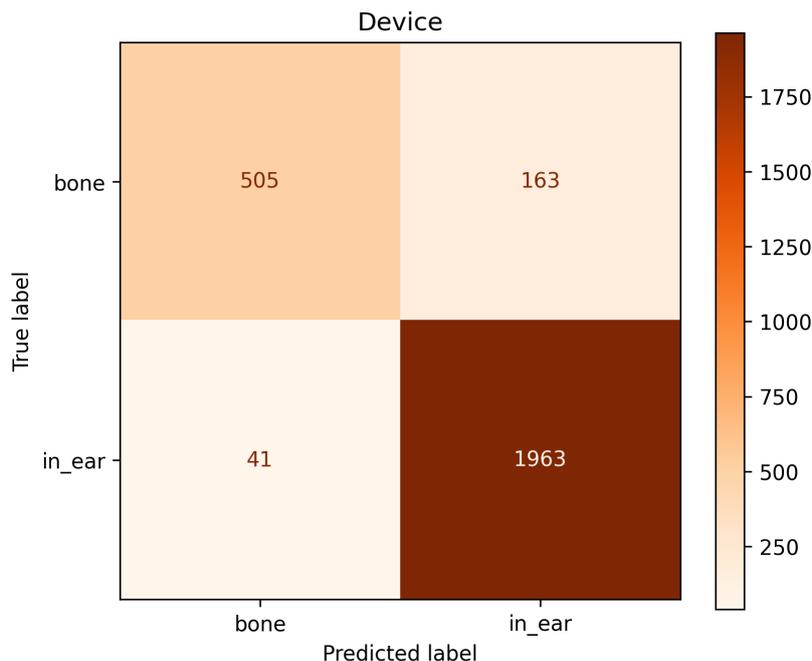


Figure 2. Confusion matrices for (top) biometric identification, (middle) stimulus language classification, and (bottom) device modality classification, all showing high class separability.

We further benchmarked TriNet-MTL against:

- **Single-Task Transformers (STT):** Independent transformers for each task.
- **Shared CNN with Separate Classifiers (SC-SC):** A shared convolutional encoder followed by separate classifiers, omitting the transformer module.

Table 2. Comparative Performance of TriNet-MTL and Baseline Models.

Model	Bio (%)	Lang (%)	Dev (%)	Avg (%)
STT	89.5	82.1	85.3	85.6
SC-SC	87.2	80.9	83.5	83.9
TriNet-MTL (Proposed)	93.9	91.6	92.4	92.6

While STT and SC-SC achieved competitive results in the binary tasks, TriNet-MTL consistently outperformed across all tasks, particularly excelling in biometric identification. These results highlight the strength of unified spatiotemporal representation learning in EEG-based multi-task classification.

4.3. Ablation Study

An ablation study was conducted to assess the contribution of key model components:

- **Without Transformer Encoder:** Only convolutional layers and task heads.
- **Without Convolutional Encoder:** Transformer operating directly on raw EEG windows.

Table 3. Ablation Study Results.

Variant	Bio (%)	Lang (%)	Dev (%)	Avg (%)
Without Transformer	88.2	80.1	82.6	83.6
Without Conv Encoder	84.7	76.9	79.2	80.3
Full TriNet-MTL	93.9	91.6	92.4	92.6

Both encoders significantly contributed to performance. Removing the transformer led to noticeable drops in accuracy, confirming its role in modeling long-range temporal dependencies. Similarly, omitting the convolutional encoder weakened localized pattern learning.

These findings align with recent studies advocating combined CNN-transformer architectures for EEG decoding [23,24].

Our results validate the effectiveness of the TriNet-MTL framework for EEG-based multi-task learning, achieving high accuracy across biometric, language, and device classification. The combination of convolutional and transformer-based encoders enables rich hierarchical modeling of both spatial and temporal EEG dynamics.

5. Conclusions

This study proposed TriNet-MTL, a unified multi-branch deep learning architecture designed to extract identity- and cognitively-relevant representations from auditory-evoked EEG signals. The model simultaneously addresses three interrelated tasks, biometric identification, auditory stimulus language classification, and stimulus modality recognition, by leveraging a shared encoder coupled with task-specific output branches. Through the integration of temporal convolutional encoding and global contextual modeling via transformers, the architecture effectively captures both local dynamics and long-range dependencies within EEG time-series data.

Empirical evaluation on a publicly available auditory EEG dataset demonstrated that TriNet-MTL achieves high classification performance across all tasks, with particularly strong results in biometric recognition. The model's ability to simultaneously decode both subject identity and cognitive state highlights the rich representational potential of evoked EEG signals. Furthermore, the multi-task formulation offers advantages in terms of improved generalization, parameter efficiency, and task synergy, making the proposed framework a promising candidate for deployment in neurobiometric authentication systems, brain-computer interfaces (BCIs), and cognitively adaptive technologies.

While the proposed framework provides compelling evidence for the feasibility of multi-task EEG decoding, several important directions remain open for further investigation. A critical next step involves addressing the issue of cross-session and cross-device variability, which remains a significant barrier to the practical deployment of EEG-based systems. Future research could explore domain adaptation techniques, transfer learning strategies, or contrastive representation learning to enhance the model's robustness across recording environments and hardware setups.

Another promising direction lies in improving the interpretability of the learned representations. Although transformer-based models are effective in modeling temporal dependencies, their decision-making process remains opaque. Incorporating interpretable attention mechanisms or neuro-symbolic reasoning layers could provide insight into the neurophysiological basis of model predictions and help identify salient cognitive components within the EEG signal.

In addition, the integration of subject calibration and personalization strategies, such as meta-learning or few-shot learning, could reduce the need for large volumes of training data per individual and enable more user-friendly real-time applications. Future work may also focus on optimizing the model for online and real-time inference to support deployment in time-sensitive settings such as adaptive BCI environments or continuous authentication systems.

Moreover, extending this work to include multimodal data, such as eye-tracking, electromyography (EMG), or contextual behavioral signals, may enhance robustness and provide complementary information for both biometric and cognitive inference. Finally, large-scale studies involving more diverse populations and varied demographics are essential to rigorously evaluate the generalizability, fairness, and ethical implications of EEG-based biometric systems.

In summary, this work demonstrates that auditory-evoked EEG signals contain rich, multi-dimensional information that can be effectively decoded using deep multi-task learning. The proposed TriNet-MTL framework establishes a foundation for future research at the intersection of neural signal processing, biometric security, and cognitive state modeling.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments: The authors would like to thank the contributors of the Auditory-Evoked Potential EEG Biometric Dataset v1.0.0 for providing open access to high-quality EEG data. We are also grateful to our research supervisors and colleagues at the Department of Computer Engineering, UET Lahore, for their invaluable feedback and technical guidance throughout this study.

References

1. Campisi, P.; Rocca, D.L. Brain waves for biometric recognition. *IEEE Transactions Inf. Forensics Secur.* **2014**, *9*, 782–800. <https://doi.org/10.1109/TIFS.2014.2309032>.
2. Marcel, S.; del R. Millán, J. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 743–752. <https://doi.org/10.1109/TPAMI.2007.1004>.
3. Picton, T.W. *Human Auditory Evoked Potentials*; Plural Publishing, 2011.
4. Näätänen, R.; Winkler, I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* **1999**, *125*, 826. <https://doi.org/10.1037/0033-2909.125.6.826>.
5. Palaniappan, R. Two-stage biometric authentication method using thought activity brain waves. *Int. J. Neural Syst.* **2008**, *18*, 59–66. <https://doi.org/10.1142/S0129065708001330>.
6. Kumar, B.V.K.V.; Mahalanobis, A.; Juday, R.D. *Correlation Pattern Recognition*; Cambridge University Press, 2006.
7. Poulos, M.; Rangoussi, M.; Alexandris, N.; Evangelou, A. Person identification based on parametric processing of the EEG. In Proceedings of the 6th International Conference on Electronics, Circuits and Systems, 1999, Vol. 1, pp. 283–286. <https://doi.org/10.1109/ICECS.1999.813198>.
8. Rocca, D.L.; Campisi, P.; Scarano, G. EEG biometric recognition in resting state with closed eyes. In Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), 2014, pp. 2090–2094.
9. Zhao, X.; Zhang, Y.; Cichocki, A. EEG-based person authentication using deep recurrent-convolutional neural networks. *IEEE Trans. Cogn. Dev. Syst.* **2019**, *12*, 763–773. <https://doi.org/10.1109/TCDS.2019.2916120>.
10. Yin, Z.; Zhang, J. Cross-session classification of mental workload levels using EEG and hybrid deep learning framework. *IEEE Access* **2017**, *5*, 23744–23754. <https://doi.org/10.1109/ACCESS.2017.2763938>.
11. Näätänen, R.; Picton, T.W. The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology* **1987**, *24*, 375–425. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>.
12. Korkmaz, Ö.; Zararsiz, G. Biometric authentication using auditory evoked EEG responses. *Biomed. Signal Process. Control* **2020**, *62*, 102067. <https://doi.org/10.1016/j.bspc.2020.102067>.
13. Gupta, R.; Palaniappan, R. EEG-based biometric authentication using auditory evoked potentials. *Int. J. Neural Syst.* **2020**, *30*, 2050015. <https://doi.org/10.1142/S0129065720500156>.
14. Shoushtarian, M.; McMahon, C.M.; Houshyar, R. Bone-conduction auditory evoked potentials: A review. *Hear. Res.* **2021**, *402*, 108002. <https://doi.org/10.1016/j.heares.2021.108002>.
15. Kotz, S.A.; Elston-Güttler, K.E. The role of proficiency on processing categorical and associative information in the L2 as revealed by ERPs. *J. Neurolinguistics* **2004**, *17*, 215–235.
16. Ruder, S. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098, 2017.
17. Yin, Z.; Zhao, M.; Wang, Y.; Yang, J.; Zhang, J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Comput. Methods Programs Biomed.* **2017**, *140*, 93–110.
18. Zhang, Y.; Zhou, G. A review on multi-task learning for EEG decoding. *Front. Hum. Neurosci.* **2018**, *12*, 500. <https://doi.org/10.3389/fnhum.2018.00500>.
19. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, 2017, Vol. 30, pp. 5998–6008.
20. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013. <https://doi.org/10.1088/1741-2552/aace8c>.

21. Abo Alzahab, N.; Di Iorio, A.; Apollonio, L.; Alshalak, M.; Gravina, R.; Antognoli, L.; Baldi, M.; Scalise, L.; Alchalabi, B. Auditory Evoked Potential EEG-Biometric Dataset (Version 1.0.0). <https://doi.org/10.13026/ps31-fc50>, 2021. PhysioNet. RRID:SCR_007345.
22. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. ChronoNet: A deep recurrent neural network for abnormal EEG identification. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019, pp. 674–677. <https://doi.org/10.1109/EMBC.2019.8856991>.
23. Li, X.; Liu, Y.; Zhang, J.; Duan, N.; Yan, W.; Gu, X.; Liu, D.; Yu, Z. Spatial–Temporal Transformer for EEG-Based Emotion Recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**. <https://doi.org/10.1109/TNNLS.2023.3242905>.
24. Wang, M.; Li, X.; Yan, W.; Liu, D.; Yu, Z. EEGFormer: Transformer-Based Model for EEG Signal Decoding. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**. <https://doi.org/10.1109/TNNLS.2023.3252574>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.