



Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Efficacy of transformer networks for classification of EEG data

Gourav Siddhad^{a,*}, Anmol Gupta^a, Debi Prosad Dogra^b, Partha Pratim Roy^a^a Department of Computer Science and Engineering, Indian Institute of Technology, Roorkee, 247667, Uttarakhand, India^b School of Electrical Sciences, Indian Institute of Technology, Bhubaneswar, 752050, Odisha, India

ARTICLE INFO

Keywords:

Age and Gender
Classification
Deep learning
EEG
Mental workload
Transformer network

ABSTRACT

With the unprecedented success of transformer networks in natural language processing (NLP), recently, they have been successfully adapted to areas like computer vision, generative adversarial networks (GAN), and reinforcement learning. Classifying electroencephalogram (EEG) data has been challenging and researchers have been overly dependent on pre-processing and hand-crafted feature extraction. Despite having achieved automated feature extraction in several other domains, deep learning has not yet been accomplished for EEG. In this paper, the efficacy of the transformer network for the classification of raw EEG data (cleaned and pre-processed) is explored. The performance of transformer networks was evaluated on a local (age and gender data) and a public dataset (STEW). First, a classifier using a transformer network is built to classify the age and gender of a person with raw resting-state EEG data. Second, the classifier is tuned for mental workload classification with open access raw multi-tasking mental workload EEG data (STEW). The network achieves an accuracy comparable to state-of-the-art accuracy on both the local (Age and Gender dataset; 94.53% (gender) and 87.79% (age)) and the public (STEW dataset; 95.28% (two workload levels) and 88.72% (three workload levels)) dataset. The accuracy values have been achieved using raw EEG data without feature extraction. Results indicate that the transformer-based deep learning models can successfully abate the need for heavy feature-extraction of EEG data for successful classification.

1. Introduction

The field of natural language processing (NLP) witnessed unprecedented success with the introduction of transformer network by Vaswani et al. [1]. Radford et al. [2] soon scaled the original transformer network to generative pre-trained (GPT) language models with billions of parameters. They achieved incredible success across different NLP tasks, including question-answering, translation, and many more. In the latest iteration, GPT-3 [3], the authors scaled the parameters to 175 billion and achieved state-of-the-art results. Recently, Fedus et al. [4] introduced switch-transformers with a trillion parameters and observed substantial gains over the state-of-the-art in over 100 languages. Transformer network has not only proved to be a resounding success in NLP, but it has also been successfully adapted to the field of reinforcement learning [5], computer vision [6], and even in GANs [7].

This paper explored the use of transformer networks to classify raw electroencephalography (EEG) data. EEG has long been used for different purposes like disease diagnosis [8], BCI [9], neuro-feedback [10], workload estimation [11], motor imagery classification [12], and even brain-to-brain communication [13]. EEG records the electrical activity at the human scalp produced by the ionic currents resulting from

simultaneous activation of multiple neurons in the brain [14]. There are several advantages of using EEG to measure brain activity, including but not limited to its low cost, portability, and being non-invasive with a very high temporal resolution [15].

However, EEG has several disadvantages as well. The most important of which are noisy raw data [16], high inter-subject variability [17], and low spatial resolution [18]. Artifacts from many non-physiological sources such as power line, bad electrode contact, broken electrodes, etc. and physiological sources such as cardiac pulse, muscle activity, sweating, movement, etc., contaminate raw EEG data [19]. Due to these limitations, pre-processing raw data in EEG plays a vital role and it usually serves as a bottleneck for any classifier. The input fed to the classifier is usually cleaned, filtered, and pre-processed data. Even then, the classifiers could not extract the relevant features as the pre-processing usually results in loss of a significant amount of information and artifact correction cannot recover the EEG signal from the noise in a satisfactory way [20].

With the recent advances in deep learning [21], one expects that the classifier should learn the required features automatically and not rely on hand-engineered features. However, because of the limitations of EEG, most studies have used hand-crafted features from

* Corresponding author.

E-mail addresses: g_siddhad@cs.iitr.ac.in (G. Siddhad), agupta@cs.iitr.ac.in (A. Gupta), dpdogra@iitbbs.ac.in (D.P. Dogra), partha@cs.iitr.ac.in (P.P. Roy).<https://doi.org/10.1016/j.bspc.2023.105488>

Received 10 February 2022; Received in revised form 28 August 2023; Accepted 12 September 2023

Available online 20 September 2023

1746-8094/© 2023 Elsevier Ltd. All rights reserved.

EEG data, even when classifying with complex neural network architectures like deep recurrent convolutions, deep-stacked autoencoders, frequency-dependent CNN, spatial-frequency CNN, etc. [22].

To build an end-to-end classification system for EEG, we chose to explore the highly effective transformer networks. To better evaluate the efficacy of transformer networks in classifying the raw EEG data, we used one local (Age and Gender classification) [23] and one public dataset (STEW; workload classification) [24].

The contributions of this paper are:

- Transformer network is used for raw EEG data (cleaned and pre-processed, no feature extraction) classification. To the best of our knowledge, transformer networks have not been used previously for raw EEG data classification.
- The proposed framework achieved a state-of-the-art accuracy of 94.53% (gender) and 87.79% (age) on a locally collected dataset of age and gender. It also achieved an accuracy of 95.28% (two workload levels) and 88.72% (three workload levels) on a public dataset of workload classification.
- The classification results (accuracy, precision, recall, and F1-score) are also compared with the state-of-the-art attention networks. It has been shown that the transformer networks are a better choice, especially without needing any hand-crafted feature extraction.

The rest of the paper is organized as follows. Section 2 presents a review of the latest articles that have classified age, gender, and mental workload with EEG data. Section 3 describes the methodology used for the whole experiment, including the task, pre-processing, and the transformer network's details. Section 4 reports the results and the conclusion is given in Section 5.

2. Background

For testing the transformer network, we first selected our local Age and Gender EEG dataset collected by Kaur et al. [23]. The dataset consists of resting-state EEG data of 60 individuals. To further validate the performance of the proposed methodology, it was tuned and applied on the open-access mental workload dataset known as STEW [24]. In STEW, data of 48 subjects were collected while doing a SIMKAP task (details in the Methodology section). We chose STEW dataset since a similar 14-channel EEG headset was used for the collection and the classification paradigm included both binary and multi-class classification identical to the local dataset. In the following subsections, the relevant literature for both datasets is discussed.

2.1. Age and gender classification

There is limited available literature where EEG data has been used to identify age and gender, especially for age. If the discussion is restricted to deep learning with EEG data, then the field is still very much unexplored. Putten et al. [25] applied CNN on EEG data to classify gender and achieved an accuracy of 80%. This might be the most extensive study with EEG data for gender recognition as they collected data of 1308 subjects using 26-channel EEG with 47% male subjects and 53% female subjects. Zhang et al. [26] used the phase locking values (PLV) as a feature for gender classification on the DEAP dataset using extreme learning machines (ELM) and achieved an accuracy of 80.0%.

In our previous work, Kaushik et al. [27] with the same dataset (Age and Gender), used DWT to decompose the EEG signals into five frequency bands. Deep neural networks, namely long short-term memory (LSTM), bi-directional LSTM (BLSTM), and BLSTM-LSTM, were presented, out of which BLSTM-LSTM gave the best accuracy 97.52% and 93.70% on beta frequency band for gender and age classification, respectively. As evident, the field of identifying age and gender using

EEG and deep learning is yet to be explored. Kaur et al. [23] pre-processed EEG data using Savitsky–Golay (SG) filter, smoothing is done, and discrete wavelet transform (DWT) decomposition is performed for extracting five frequency bands. Statistical features are extracted from the frequency bands. Then random forest classifier is used, which gave an accuracy of 96.66% and 88.33% for gender and age classification, respectively.

2.2. Workload classification

EEG can help us provide an objective, absolute, and temporally sensitive workload measure. Therefore, it has been used time and again for workload classification. Qu et al. [28] used support vector machine (SVM) for classifying EEG data collected from 13 participants while doing the NASA MATB-II task [29]. They used energy from the independent components extracted from independent component analysis (ICA) as features and achieved an accuracy of 79.8% for the three workload levels (low, medium, and high). Different machine learning algorithms such as linear discriminant analysis [30], random forest [31], KNN, and SVM [32] have been utilized for workload classification from EEG data.

Different neural network architectures have already been shown to classify mental workload successfully. Hefron et al. [33] used stacked LSTMs for classifying data of six participants using the MATB-II. They achieved an accuracy of 93%. Sun et al. [34] proposed WLNet specifically for classifying mental workload and compared it to temporally constrained sparse group spatial patterns (TCSGSP, a deep version of the CSP method proposed by Zhang et al. [35]) and EEGNet [36] (proposed for BCI applications, consisting of depth-wise and separable convolution layers). WLNet was a shallow neural network consisting of one-dimensional convolution layers, thereby requiring fewer parameters. They did binary workload classification on the dataset provided by Fabien Lotte research group at INRIA [17] and found that WLNet out-performed TCSGSP and EEGNet. Saha et al. [37] used a stacked denoising autoencoder to classify three workload levels and achieved 89.51% accuracy.

Specifically for the STEW dataset, different neural networks have been applied and compared. Chakladar et al. [38] compared the performances of stacked LSTM, BLSTM, CNN-LSTM, stacked autoencoders with LSTM/BLSTM, and their proposed model of hybrid BLSTM-LSTM network. Their proposed model achieved state-of-the-art accuracy of 86.33% and 82.57% for binary and tertiary classification on the STEW dataset. Lim et al. [24] made use of artifact subspace reconstruction (ASR) and features were extracted using power spectral density (PSD) and neighborhood component analysis (NCA). Using the support vector regression (SVR) model, the accuracy of 69.20% was achieved for the “SIMKAP multi-task” classification. Zhu et al. [39] constructed oblique visibility graphs (OVG) and frequency domain features (mean degree, clustering coefficient, and degree distribution) were extracted. Decision tree (DT) and SVM were used for classification, where SVM performed better and yielded 89.60% and 79.50% for binary and tertiary classification, respectively.

Attention-based networks have also been used for EEG classification. Zhu et al. [40] proposed a neural network based on CNN and attention mechanism to perform automatic sleep staging. Zheng and Chen [41] proposed an end-to-end attention-based Bi-LSTM method, named Bi-LSTM-AttGW for visual cognition task. The attention weighting method is applied to Bi-LSTM output.

None of the above-cited literature for both datasets worked with raw EEG data. The input fed to the deep neural network has varied from simple band segregated data (alpha, beta, gamma, etc.) to simple features like mean, variance, etc. of the band power to much more complex features like PLV and PLI. Thus, we can conclude that EEG poses a unique challenge of complex pre-processing and feature selection mechanisms.

In this paper, we explored the possibility of using raw EEG data for doing classification. Given the available literature, it was evident

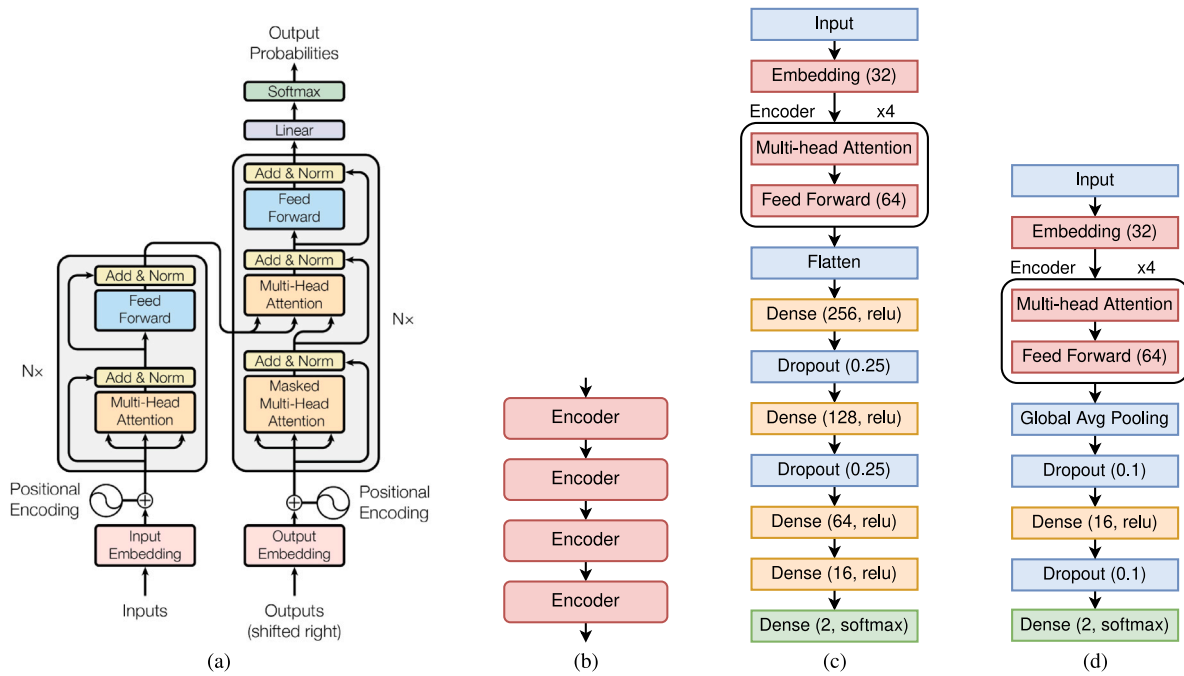


Fig. 1. Network architecture for (a) Original transformer network [1] (left is encoder and decoder is on the right); (b) Arrangement of encoders in transformer block (stack of 4 encoders are used in the proposed method); Proposed architecture for (c) Age and Gender, and (d) STEW dataset.

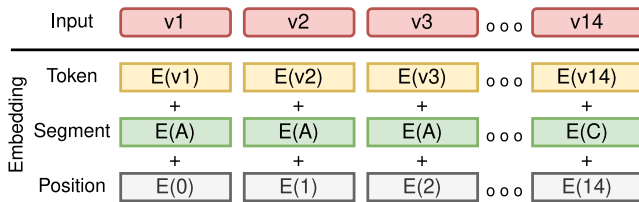


Fig. 2. This figure shows the working of embedding and positional encoding. E with numbers represents the position embedding, E with different alphabets represents different segments, and E with various v represents the token vectors for the input. These are sum together to produce the input for the transformer network.

that we needed to use a proficient neural network architecture that can automatically extract the relevant features. Recently, transformer networks have seen a phenomenal rise and are used in state-of-the-art accomplishments in almost every domain [6,7]. In the context of EEG, transformer networks have been rarely used but have been shown to generate state-of-the-art results for sleep stage classification [42] and speech recognition [43]. Therefore, we decided to use a transformer network for the classification.

3. Methodology

This section discusses the proposed transformer network for the classification of age and gender and mental workload estimation. After acquisition and pre-processing is done on EEG data, epochs are created with a window size of 0.25 s. The transformer network is then trained for classification on this raw EEG data.

3.1. Experimental data

3.1.1. Age and gender dataset

This dataset [23] consists of raw EEG signals from 60 users (25 female and 35 male) with an age range between 6 and 55 years. The data was collected with Emotiv Epoch Plus (14 channels) with a sampling rate of 128 Hz. The data was acquired at a resting state for 10 sessions of 10 s each.

3.1.2. STEW dataset

The simultaneous task EEG workload (STEW) dataset [24] consists of raw EEG data of 48 subjects who participated in a multitasking workload experiment that utilized the simultaneous capacity (SIMKAP) multitasking test. The signals were captured using the Emotiv EPOC EEG headset, with 16-bit A/D resolution, 128 Hz sampling frequency, and 14 channels, namely AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 according to the 10–20 international system with two reference channels (CMS, DRL). There are two parts to the experiment. First, the data was acquired for 2.5 min with subjects at rest or “No task”. Next, subjects performed the SIMKAP test with EEG being recorded and the final 2.5 min were used as the workload condition. Subjects rated their perceived mental workload after each segment of the experiment on a rating scale of 1–9.

3.2. EEG and pre-processing

EEG signals in their raw form (captured from a device) contain noise and artifacts and must be cleaned before use. EEG data is imported and bandpass filtering in the range [0.5 Hz, 70 Hz] is done to remove environmental/muscle noise in scalp EEG. After creating epochs with a window size of 0.25 s and removing bad epochs from the data, independent component analysis (ICA) is applied, and bad channels are manually removed. The dataset is used with a sampling rate of 128 Hz, the same as during acquisition.

3.3. Transformers

The transformer network [1] is a neural network based on a self-attention mechanism. It has proved higher quality and lower computation requirements for language translation tasks than recurrent and convolution models. It applies a self-attention mechanism that directly models relationships between all input parts, regardless of their position. The result of these comparisons is an attention score for every other part in the input, which determines their contributions. The transformer network (shown in Fig. 1(a)) consists of encoders and decoders. The encoders are stacked in the transformer model as

Table 1
Classification analysis for age and gender and STEW dataset.

Model	Features	Classifier	Accuracy	
Age and Gender dataset			Gender (2 class)	Age (6 class)
Kaur et al. [23]	Statistical	Random forest	96.66	88.33
Kaushik et al. [27]	Frequency	BLSTM-LSTM	97.50	93.70
Proposed work	Not used	Transformer	94.53	87.79
STEW dataset			No vs. SIMKAP task (2 class)	SIMKAP multi-task (3 class)
Lim et al. [24]	PSD, NCA	SVR	–	69.20
Zhu et al. [39]	Frequency	DT, SVM	89.60	79.50
Chakladar et al. [38]	Frequency, Time	BLSTM-LSTM	86.33	82.57
Proposed work	Not used	Transformer	95.28	88.72

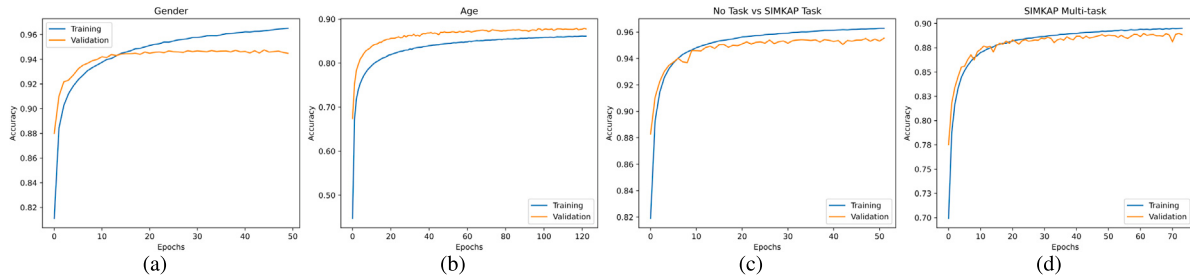


Fig. 3. Train and validation accuracy for Age and Gender (a) Gender, (b) Age; and STEW dataset (c) No vs. SIMKAP task, (d) SIMKAP multi-task.

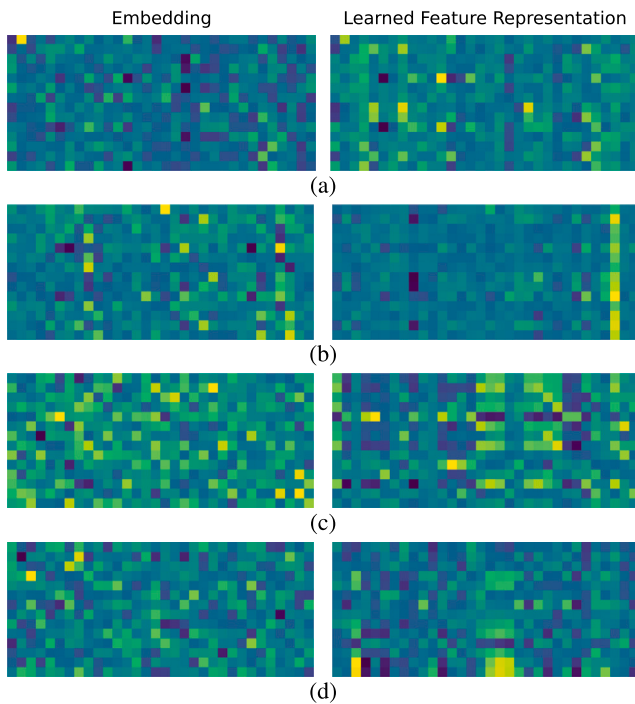


Fig. 4. Visualization of embedding (left) and learned feature representation (right) for Age and Gender dataset (a) Gender, (b) Age; and STEW dataset (c) No vs. SIMKAP task, (d) SIMKAP multi-task.

shown in Fig. 1(b). Each encoder has two sub-layers, a multi-head self-attention mechanism, and a position-wise fully connected feed-forward network. There exists a residual connection around both sub-layers, followed by a normalization layer. The decoder is not used here.

Embedding and positional encoding are done before sending the input to the encoders. Encoders operate with vectors and therefore, the input is converted into embedding vectors using a fixed representation. Unlike recurrent neural networks (RNN) and LSTM, transformer networks do not have a way to capture relative positions of the input. To

provide this contextual information, positional encoding is used with each input vector. Positional encoding is not part of the architecture of the model but the pre-processing. The positional encoding vector is generated to be the same size as the embedding for each input. The positional encoding vector is added to the embedding vector after calculation. The ‘injected’ pattern into the embedding vector allows the algorithm to learn this spatial information. This mechanism is shown in Fig. 2.

3.4. Classification

For Age and Gender classification, the architecture for gender (2 class) classification is shown in Fig. 1(c). The last layer is modified as *Dense(6)* for age (6 class) classification network. The number of attention heads is also increased to 8. The rest of the parameters are kept the same. The dataset is evaluated for two classification scenarios. First, for gender classification, the dataset has been categorized based on the gender of the subject, male and female. This network (2 class) is trained for 49 epochs with early stopping (Fig. 3(a)). The criterion for early stopping is that the training stops when there is no improvement in validation loss for 25 epochs. Second, for age classification, the dataset has been categorized into six age groups, 6–11, 12–16, 18–24, 25–30, 33–39, and 42–56. These age groups are then labeled as 0–5. This network (6 class) is trained for 122 epochs with early stopping (Fig. 3(b)).

For classification of the STEW dataset, the network architecture for “No vs. SIMKAP task” is shown in Fig. 1(d). The last layer is modified as *Dense(3)* for the “SIMKAP multi-task” network. For embedding, 32 dimensions are used with 4 attention heads. The hidden layer size of the feed-forward network in the transformer is 64. Adam is used as an optimizer with default parameters and sparse categorical cross-entropy is used for loss. Both datasets are used as a 70-15-15 percent split for train, validation, and test sets during the experimentation. STEW dataset is evaluated for two classification scenarios. First, for “No vs. SIMKAP task”, the dataset has been categorized into two workload levels, at rest as low and performing SIMKAP tasks as high. This network (2 class) is trained for 51 epochs with early stopping (Fig. 3(c)). Second, for the “SIMKAP multi-task”, the dataset is categorized into three workload levels, using the ratings provided by each subject. One can perceive a rating of 1–3 as low, 4–6 as moderate, and 7–9 as high workload. This network (3 class) is trained for 73 epochs with early stopping (Fig. 3(d)).

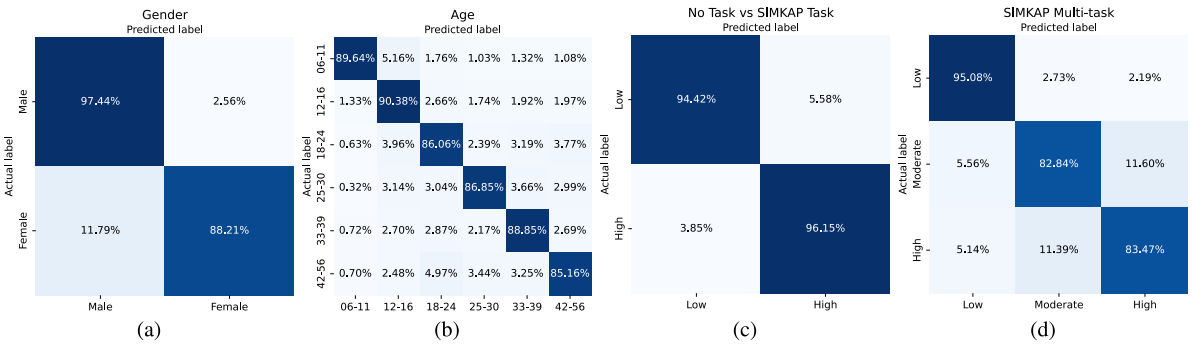


Fig. 5. Confusion matrix for Age and Gender dataset (a) Gender, (b) Age; and STEW dataset (c) No vs. SIMKAP task, (d) SIMKAP multi-task.

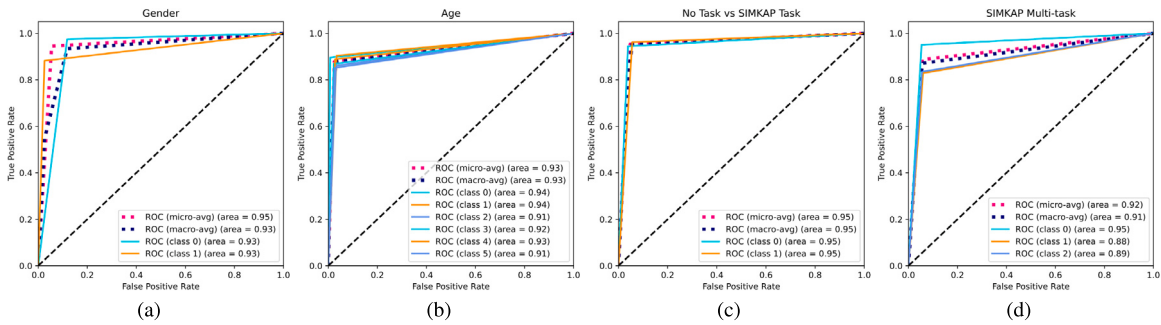


Fig. 6. ROC for Age and Gender dataset (a) Gender, (b) Age; and STEW dataset (c) No vs. SIMKAP task, (d) SIMKAP multi-task.

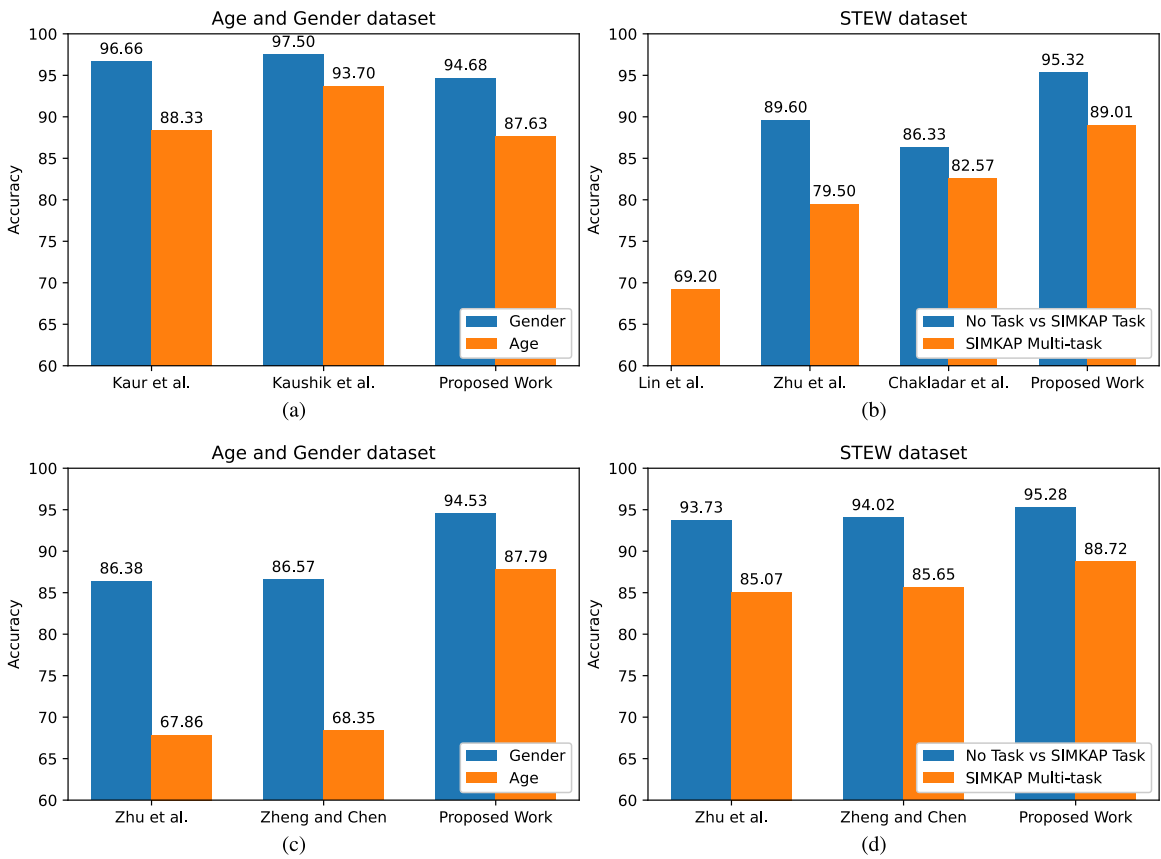


Fig. 7. Comparison of methods for (a) Age and Gender and (b) STEW dataset; and Ablation Study for (c) Age and Gender and (d) STEW dataset.

Table 2
Ablation study.

Model	Method	Accuracy		Precision		Recall		F1-score	
		2 class	6 class	2 class	6 class	2 class	6 class	2 class	6 class
Age and gender									
Zhu et al. [40]	CNN	86.38	67.86	87.72	68.88	80.21	68.41	82.64	67.72
Zheng and Chen [41]	Bi-LSTM-AttGW	86.57	68.35	85.74	68.46	82.25	68.25	83.66	68.24
Proposed work	Transformer	94.53	87.79	94.39	88.19	92.82	87.82	93.55	87.99
STEW									
Zhu et al. [40]	CNN	93.73	85.07	93.74	83.51	93.73	82.60	93.73	82.99
Zheng and Chen [41]	Bi-LSTM-AttGW	94.02	85.65	94.02	83.94	94.01	83.66	94.02	83.78
Proposed work	Transformer	95.28	88.72	95.29	87.31	95.28	87.12	95.28	87.21

4. Results and discussion

This section discusses the comparative analysis between the proposed model and other approaches, including machine learning and popular deep neural networks. The results are summarized in Table 1, confusion matrices are shown in Fig. 5, receiver operating characteristics (ROC) are shown in Fig. 6, comparative analysis is shown in Fig. 7, and ablation study is summarized in Table 2. Also, the features generated from the network corresponding to their input embedding can be seen in Fig. 4 for both classes of both the datasets.

4.1. Age and gender dataset

The proposed work make use of transformer network (Fig. 1(c)) on raw EEG data and achieved an accuracy of 94.68% and 87.63% for gender and age classification, respectively (Fig. 7(a)). The corresponding confusion matrices are shown in Figs. 5(a) and 5(b) respectively and ROC are shown in Figs. 6(a) and 6(b) respectively. The accuracy achieved with the transformer network trained on pre-processed data is not far behind the state-of-the-art accuracy. The reason could be the positional encoding, which can be done in different ways. The positional encoding used in transformer networks for text processing is not tailor-made for EEG data. Also, the features are not used, which are more effective in some cases (Figs. 4(a) and 4(b)).

4.2. STEW dataset

The proposed work make use of transformer network (Fig. 1(d)) for classification and achieved test accuracy of 95.32% and 89.01% for “No vs. SIMKAP task” and “SIMKAP multi-task” classification on raw EEG data. The corresponding confusion matrices are shown in Figs. 5(c) and 5(d) respectively and ROC are shown in Figs. 6(c) and 6(d) respectively. Transformer network trained with pre-processed EEG data clearly out-performs the existing methods with feature extraction for both classification scenarios on the STEW dataset (Fig. 7(b)). One of the reasons could be the effective learning of embedding and positional encoding used which provides contextual information.

4.3. Ablation study

For the ablation study, highly successful networks proposed by Zhu et al. [40] and Zheng and Chen [41] are used. These are trained on both datasets used in this work. The accuracies achieved by the proposed model achieved better results than both the methods for both datasets. The improved accuracy might result from multi-head attention, which learns better and faster than a single attention layer used in these networks, resulting in high performance. The comparison of results with these attention-based networks is summarized in Table 2. The comparison is also shown in Figs. 7(c) and 7(d) for Age and Gender and STEW dataset respectively. It can be seen that the proposed transformer network for EEG outperforms existing deep learning networks on raw EEG data.

5. Conclusion

This study explored the use of transformer networks for the classification of raw (cleaned) EEG data. To evaluate such models’ efficacy, proposed network’s performance was tested on a local (Age and Gender) and a public dataset (STEW). The robustness of the classifier was also tested as both datasets had different classification paradigms. Accuracies of 94.68% (gender), 87.63% (age), 95.59% (two workload levels), and 89.40% (three workload levels) were achieved. The proposed method gave state-of-the-art accuracy for STEW and comparable for Age-and-gender. This proves the effectiveness of the transformers for processing raw EEG data and can reduce the need for feature extraction in EEG as deep learning has done in other domains such as image processing.

However, the study needs to be further replicated and validated with different datasets and more comparisons. The positional encoding used in transformers is not tailor-made for EEG data and features are not used, which are more effective in some cases. Both of these may be the reason for lower performance on the Age and Gender dataset. It can be further improved by generating robust EEG embedding using multiple datasets similar to the embedding used in NLP. Similarly, new positional encoding methods specifically to EEG can be explored in the future. In conclusion, transformer networks seem to be effective in classifying EEG data as well. They might also reduce the over-reliance on the feature extraction still used in the EEG domain.

CRedit authorship contribution statement

Gourav Siddhad: Methodology, Writing – original draft, Visualization. **Anmol Gupta:** Validation, Writing – original draft. **Debi Prosad Dogra:** Writing – review & editing. **Partha Pratim Roy:** Writing – review & editing, Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is available online.

References

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, et al., Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.
- [2] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, et al., Improving language understanding by generative pre-training, 2018, p. 12.
- [3] T. Brown, B. Mann, N. Ryder, M. Subbiah, J.D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Adv. Neural Inf. Process. Syst. 33 (2020) 1877–1901.
- [4] W. Fedus, B. Zoph, N. Shazeer, Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity, J. Mach. Learn. Res. 23 (1) (2022) 5232–5270.

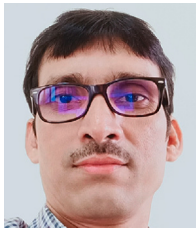
- [5] O. Vinyals, I. Babuschkin, W.M. Czarnecki, M. Mathieu, A. Dudzik, et al., Grandmaster level in StarCraft II using multi-agent reinforcement learning, *Nature* (ISSN: 1476-4687) 575 (7782) (2019) 350–354, <http://dx.doi.org/10.1038/s41586-019-1724-z>.
- [6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929).
- [7] Y. Jiang, S. Chang, Z. Wang, Transgan: Two transformers can make one strong gan, 2021, arXiv preprint [arXiv:2102.07074](https://arxiv.org/abs/2102.07074).
- [8] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, et al., DeprNet: A deep convolution neural network framework for detecting depression using EEG, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–13.
- [9] X. Chen, B. Zhao, Y. Wang, X. Gao, Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm, *J. Neural Eng.* 16 (2) (2019) 026012.
- [10] C. Dousset, H. Kajosch, A. Ingels, E. Schröder, C. Kornreich, et al., Preventing relapse in alcohol disorder with EEG-neurofeedback as a neuromodulation technique: A review and new insights regarding its application, *Addict. Behav.* 106 (2020) 106391.
- [11] W.-L. Zheng, B.-L. Lu, A multimodal approach to estimating vigilance using EEG and forehead EOG, *J. Neural Eng.* 14 (2) (2017) 026017.
- [12] M.T. Sadiq, X. Yu, Z. Yuan, M.Z. Aziz, S. Siuly, W. Ding, Toward the development of versatile brain–computer interfaces, *IEEE Trans. Artif. Intell.* 2 (4) (2021) 314–328.
- [13] A. Pérez, M. Carreiras, J.A. Duñabeitia, Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening, *Sci. Rep.* 7 (1) (2017) 1–12.
- [14] J.C. Henry, Electroencephalography: Basic principles, clinical applications, and related fields, *Neurology* 67 (11) (2006) 2092.
- [15] M. Hämäläinen, R. Hari, R.J. Ilmoniemi, J. Knuutila, O.V. Lounasmaa, Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain, *Rev. Mod. Phys.* 65 (2) (1993) 413.
- [16] A. Puce, M.S. Hämäläinen, A review of issues related to data acquisition and analysis in EEG/MEG studies, *Brain Sci.* (ISSN: 2076-3425) 7 (6) (2017) <http://dx.doi.org/10.3390/brainsci7060058>.
- [17] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, et al., A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update, *J. Neural Eng.* 15 (3) (2018) 031005.
- [18] E.D. Kondylis, T.A. Wozny, W.J. Lipski, A. Popescu, V.J. DeStefino, et al., Detection of high-frequency oscillations by hybrid depth electrodes in standard clinical intracranial EEG recordings, *Front. Neurol.* 5 (2014) 149.
- [19] C.M. Michel, D. Brunet, EEG source imaging: A practical review of the analysis steps, *Front. Neurol.* (ISSN: 1664-2295) 10 (2019) <http://dx.doi.org/10.3389/fneur.2019.00325>.
- [20] A. Pedroni, A. Bahreini, N. Langer, Automagic: Standardized preprocessing of big EEG data, *NeuroImage* (ISSN: 1053-8119) 200 (2019) 460–473, <http://dx.doi.org/10.1016/j.neuroimage.2019.06.046>.
- [21] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [22] Z. Gao, W. Dang, X. Wang, X. Hong, L. Hou, et al., Complex networks and deep learning for EEG signal analysis, *Cogn. Neurodyn.* (2020) <http://dx.doi.org/10.1007/s11571-020-09626-1>.
- [23] B. Kaur, D. Singh, P.P. Roy, Age and gender classification using brain–computer interface, *Neural Comput. Appl.* 31 (10) (2019) 5887–5900, <http://dx.doi.org/10.1007/s00521-018-3397-1>.
- [24] W.L. Lim, O. Sourina, L.P. Wang, STEW: Simultaneous task EEG workload data set, *IEEE Trans. Neural Syst. Rehabil. Eng.* (ISSN: 1558-0210) 26 (11) (2018) 2106–2114, <http://dx.doi.org/10.1109/TNSRE.2018.2872924>.
- [25] M.J. Van Putten, S. Olbrich, M. Arns, Predicting sex from brain rhythms with deep learning, *Sci. Rep.* 8 (1) (2018) 1–7.
- [26] J. Zhang, H. Xu, L. Zhu, W. Kong, Z. Ma, Gender recognition in emotion perception using EEG features, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 2883–2887, <http://dx.doi.org/10.1109/BIBM47256.2019.8983332>.
- [27] P. Kaushik, A. Gupta, P.P. Roy, D.P. Dogra, EEG-based age and gender prediction using deep BLSTM-LSTM network model, *IEEE Sens. J.* 19 (7) (2019) 2634–2641, <http://dx.doi.org/10.1109/JSEN.2018.2885582>.
- [28] H. Qu, Y. Shan, Y. Liu, L. Pang, Z. Fan, et al., Mental workload classification method based on EEG independent component features, *Appl. Sci.* 10 (9) (2020) 3036, <http://dx.doi.org/10.3390/app10093036>.
- [29] Y. Santiago-Espada, R.R. Myer, K.A. Latorella, J.R. Comstock Jr., *The Multi-Attribute Task Battery Ii (Matb-Ii) Software for Human Performance and Workload Research: A User's Guide*, NASA, 2011.
- [30] M. Bagheri, S.D. Power, EEG-based detection of mental workload level and stress: The effect of variation in each state on classification of the other, *J. Neural Eng.* (ISSN: 1741-2552) 17 (5) (2020) 056015, <http://dx.doi.org/10.1088/1741-2552/abbc27>.
- [31] Z. Pei, H. Wang, A. Bezerianos, J. Li, EEG-based multiclass workload identification using feature fusion and selection, *IEEE Trans. Instrum. Meas.* (ISSN: 1557-9662) 70 (2021) 1–8, <http://dx.doi.org/10.1109/TIM.2020.3019849>.
- [32] S.S. Gupta, R.R. Manthalkar, Classification of visual cognitive workload using analytic wavelet transform, *Biomed. Signal Process. Control* (ISSN: 1746-8094) 61 (2020) 101961, <http://dx.doi.org/10.1016/j.bspc.2020.101961>.
- [33] R.G. Hefron, B.J. Borghetti, J.C. Christensen, C.M.S. Kabbani, Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation, *Pattern Recognit. Lett.* (ISSN: 01678655) 94 (2017) 96–104, <http://dx.doi.org/10.1016/j.patrec.2017.05.020>.
- [34] Z. Sun, B. Li, F. Duan, H. Jia, S. Wang, et al., WLnet: Towards an approach for robust workload estimation based on shallow neural networks, *IEEE Access* (ISSN: 2169-3536) 9 (2021) 3165–3173, <http://dx.doi.org/10.1109/ACCESS.2020.3044732>.
- [35] Y. Zhang, C.S. Nam, G. Zhou, J. Jin, X. Wang, et al., Temporally constrained sparse group spatial patterns for motor imagery BCI, *IEEE Trans. Cybern.* 49 (9) (2018) 3322–3332.
- [36] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, et al., EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces, *J. Neural Eng.* 15 (5) (2018) 056013.
- [37] A. Saha, V. Minz, S. Bonela, S.R. Sreeja, R. Chowdhury, et al., Classification of EEG signals for cognitive load estimation using deep learning architectures, in: U.S. Tiwary (Ed.), *Intelligent Human Computer Interaction*, in: *Lecture Notes in Computer Science*, Springer International Publishing, Cham, ISBN: 978-3-030-04021-5, 2018, pp. 59–68, http://dx.doi.org/10.1007/978-3-030-04021-5_6.
- [38] D. Das Chakladar, S. Dey, P.P. Roy, D.P. Dogra, EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm, *Biomed. Signal Process. Control* (ISSN: 17468094) 60 (2020) 101989, <http://dx.doi.org/10.1016/j.bspc.2020.101989>.
- [39] G. Zhu, F. Zong, H. Zhang, B. Wei, F. Liu, Cognitive load during multitasking can be accurately assessed based on single channel electroencephalography using graph methods, *IEEE Access* (ISSN: 2169-3536) (2021) 1, <http://dx.doi.org/10.1109/ACCESS.2021.3058271>.
- [40] T. Zhu, W. Luo, F. Yu, Convolution- and attention-based neural network for automated sleep stage classification, *Int. J. Environ. Res. Public Health* (ISSN: 1660-4601) 17 (11) (2020) <http://dx.doi.org/10.3390/ijerph17114152>.
- [41] X. Zheng, W. Chen, An attention-based bi-LSTM method for visual object classification via EEG, *Biomed. Signal Process. Control* 63 (2021) 102174.
- [42] D. Kostas, S. Aroca-Ouellette, F. Rudzicz, BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data, *Front. Hum. Neurosci.* (ISSN: 1662-5161) 15 (2021) 253, <http://dx.doi.org/10.3389/fnhum.2021.653659>.
- [43] G. Krishna, C. Tran, M. Carnahan, A.H. Tewfik, EEG based continuous speech recognition using transformers, 2020, [arXiv:2001.00501](https://arxiv.org/abs/2001.00501) [cs, eess, stat].



Gourav Siddhad is currently pursuing a Ph.D. degree with the Department of Computer Science and Engineering, Indian Institute of Technology, Roorkee, India, working under the guidance of Dr. P. P. Roy. He received the M.Tech degree from the PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India, in 2020 and the B.Tech degree from the Hansraj College, University of Delhi, New Delhi, India, in 2017. He is mainly interested in the areas of brain–computer interfaces, machine learning, image processing, and cancelable biometrics.



Anmol Gupta is currently pursuing a Ph.D. degree with the Department of Computer Science and Engineering, Indian Institute of Technology, Roorkee, India, working under the guidance of Dr. P. P. Roy. He received the M.Tech. degree from the National Institute of Technology, Hamirpur, India, in 2016 and the B.Tech. degree from the National Institute of Technology, Uttarakhand, India, in 2014. He is mainly interested in the areas of machine learning and brain–computer interfaces.

G. Siddhad et al.

Debi P. Dogra received the Ph.D. degree from IIT Kharagpur, Kharagpur, India, in 2012, M.Tech. degree from IIT Kanpur, Kanpur, India, in 2003, and B.Tech. degree from the Haldia Institute of Technology, India, in 2001.

He is currently an Assistant Professor with the School of Electrical Sciences, IIT Bhubaneswar, Bhubaneswar, India. He was with Samsung Research, India, from 2011 to 2013. He worked with ETRI, South Korea, from 2006 to 2007. He worked as the Faculty of Computer Science & Engineering at the Haldia Institute of Technology (HIT) from 2003 to 2006. He has published more than 100 research papers in international journals and conferences. He has obtained three U.S. patents. His research interests include visual surveillance, augmented reality, and human-computer interface.



Partha P. Roy received the M.S. and Ph.D. degrees from the Autonomous University of Barcelona, Spain, in 2010 and the B.Tech. degree from the University of Kalyani, India, in 2002.

He is currently an Associate Professor with IIT Roorkee, Roorkee, India. He was with Samsung Research Institute, India, from 2013 to 2014. He worked as a Postdoctoral Research Fellow with Synchromedia Lab, Canada, and with RFAI Lab, France, from 2010 to 2013. He has worked as an Assistant System Engineer with TCS, India, from 2003 to 2005. He has published more than 225 research papers in international journals and conferences. His research interests are pattern recognition, computer vision, bio-signal analysis, EEG-based pattern analysis, and multilingual text recognition.

Dr. Roy is an Associate Editor of IET Image Processing, IET Biometrics, IEICE Transactions on Information and Systems, and Nature Computer Science (Springer).

Biomedical Signal Processing and Control 87 (2024) 105488