Computers in Biology and Medicine 178 (2024) 108705

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Exploring the frontier: Transformer-based models in EEG signal analysis for brain-computer interfaces

Maximilian Achim Pfeffer [a], Steve Sai Ho Ling [a,*], Johnny Kwok Wai Wong [b]

[a] Faculty of Engineering and Information Technology, University of Technology Sydney, CB11 81-113, Broadway, Ultimo, 2007, New South Wales, Australia
[b] Faculty of Design, Architecture and Building, University of Technology Sydney, 15 Broadway, Ultimo, 2007, New South Wales, Australia

ARTICLE INFO

Keywords:
Artificial intelligence
Brain-computer interface
Deep learning
Electroencephalography
Natural language processing
Transformer

ABSTRACT

This review systematically explores the application of transformer-based models in EEG signal processing and brain-computer interface (BCI) development, with a distinct focus on ensuring methodological rigour and adhering to empirical validations within the existing literature. By examining various transformer architectures, such as the Temporal Spatial Transformer Network (TSTN) and EEG Conformer, this review delineates their capabilities in mitigating challenges intrinsic to EEG data, such as noise and artifacts, and their subsequent implications on decoding and classification accuracies across disparate mental tasks. The analytical scope extends to a meticulous examination of attention mechanisms within transformer models, delineating their role in illuminating critical temporal and spatial EEG features and facilitating interpretability in model decision-making processes. The discourse additionally encapsulates emerging works that substantiate the efficacy of transformer models in noise reduction of EEG signals and diversifying applications beyond the conventional motor imagery paradigm. Furthermore, this review elucidates evident gaps and propounds exploratory avenues in the applications of pre-trained transformers in EEG analysis and the potential expansion into real-time and multi-task BCI applications. Collectively, this review distils extant knowledge, navigates through the empirical findings, and puts forward a structured synthesis, thereby serving as a conduit for informed future research endeavours in transformer-enhanced, EEG-based BCI systems.

1. Introduction

Inside the brain, millions of neurons are active at all times [1]. Postsynaptic potentials in neurons of the cerebral cortex perpendicularly to the cortical surface result in what is generally referred to as brain activity [2]. These changes in electrical potentials, or simply electrical activities, can be observed by deploying Electroencephalography (EEG) electrodes on the scalp of a subject. To be precise, the EEG data is captured by placing electrodes on a subject's outside of the scalp, which record the summation of all local potentials (also known as field potentials). Upon collecting the signals with this relatively cost-effective procedure, researchers and developers can utilise the resulting data (with its generally high temporal resolution) to analyse and deploy Brain-Computer Interfaces (BCIs) [3–5].

Recently, the widespread adoption and availability of natural language processing (NLP) has significantly advanced human-computer interaction by enabling machines to interpret and respond to human language inputs. This technology facilitates efficient and intuitive interfaces for diverse applications ranging from simple information retrieval to complex problem-solving and programming tasks. However, NLP is inherently constrained by the modalities of verbal or written communication, which necessitates the explicit externalization of thoughts [6–8].

Emerging EEG-based BCIs represent a critical evolution in this interface dynamic, offering a direct communication pathway between the human brain and computational systems. By capturing and interpreting neural signals, BCIs eliminate the necessity for physical or verbal interaction, thereby providing a more immediate and bandwidth-efficient method of interfacing with technology. This direct neural interaction facilitates a higher throughput of information and a reduction in the latency associated with conventional input methods such as keyboards or speech.

From a scientific standpoint, BCIs extend the capabilities established by NLP by translating neural activity directly into executable commands. This advancement could significantly enhance computational tasks that currently rely on language-based input, such as programming, by allowing for direct brain-to-code generation. Such capabilities not

* Corresponding author.
E-mail addresses: Max.Pfeffer@uts.edu.au (M.A. Pfeffer), Steve.Ling@uts.edu.au (S.S.H. Ling), Johnny.Wong@uts.edu.au (J.K.W. Wong).

https://doi.org/10.1016/j.compbiomed.2024.108705
Received 11 November 2023; Received in revised form 1 May 2024; Accepted 1 June 2024
Available online 8 June 2024
0010-4825/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

List of abbreviations

- AI - Artificial Intelligence
- BCI - Brain-Computer Interface
- BSS - Blind Source Separation
- CSP - Common Spatial Pattern
- dDTF - Directed Transfer Function
- DL - Deep Learning
- ECNN - Ensemble pre-trained Convolutional Neural Network
- EDA - Encoder-Decoder-Attention
- EEG - Electroencephalography
- EOG - Electrooculography
- ERD - Event-Related Desynchronisation
- ERP - Event-Related Potentials
- ERS - Event-Related Synchronisation
- FFN - Feed-Forward Network
- ICA - Independent Component Analysis
- LPL - Linear Projection Layer
- LR - Linear Regression
- LSTM - Long-Short-Term Memory Networks
- MHSA - Multi-Headed Self-Attention layer
- MI - Motor Imagery
- ML - Machine Learning
- MODMA - Multimodal Open Dataset for Mental Disorder Analysis
- NLP - Natural Language Processing
- PCA - Principal Component Analysis
- PTSD - Post-Traumatic Stress Disorder
- RF - Random Forest
- RNN - Recurrent Neural Network
- S3T - Spatial-Temporal Tiny Transformer
- SEED - Shanghai Jiao Tong University Emotion EEG Dataset
- SNR - Signal-to-Noise Ratio
- SVM - Support Vector Machine
- TE - Transformer Encoder
- TD - Transformer Decoder
- TL - Transfer Learning
- t-SNE - t-distributed Stochastic Neighbor Embedding
- TSTN - Temporal Spatial Transformer Network
- XAI - Explainable Artificial Intelligence

only streamline interactions but also enhance the potential for real-time computing applications, where speed of input and processing is critical.

Furthermore, the application of BCIs can profoundly impact cognitive neuroscience research, providing deeper insights into brain function and information processing. This direct interfacing could lead to advancements in understanding neural correlates of cognition and behavior, which are crucial for both medical diagnostics and therapeutic interventions.

In summary, while NLP has facilitated a more natural interaction between humans and machines through language, BCIs promise to transcend this communication barrier by establishing a direct, language-independent neural interface. This represents a significant scientific and technological leap towards creating more integrated, efficient, and intuitive computational systems.

EEG-mediated BCI signal analysis has long been a cornerstone of neuroscientific research, aiming at bridging human cognition with computational interfaces. With the progression of deep learning techniques, the integration of NLP within this domain holds the potential to enhance the real-time classification, signal-to-noise ratio, and multi-class classification accuracy of EEG signals. The crux of this literature review delves into the interplay between EEG-based BCIs and NLP, particularly spotlighting the advent of transformer-based models in the field.

Transformers, despite their relatively recent introduction to the machine learning panorama, have swiftly become the quintessence of innovation in the NLP realm. Their intrinsic capability to handle complex sequential data through innovative network architecture components has rendered them indispensable in contemporary AI applications. Given their accelerated trajectory in the NLP domain, an immediate question emerges: what are the prospects of these models in broad-spectrum data-driven research, especially in contexts where deep learning models have made their mark?

Within the research landscape of EEG-based BCIs, there has been a significant transition from traditional Machine Learning (ML) approaches to the more sophisticated realm of Deep Learning (DL). This literature review anchors itself firmly in the context of this transition, highlighting the pivotal role that DL methodologies are beginning to play in the advancement of EEG-BCI technology. DL's inherent strengths — such as the ability to construct end-to-end learning models, achieve higher classification accuracies, and more effectively map and model complex signal interactions — represent not just incremental advancements but a substantial leap forward in the field of BCIs.

This transition from traditional ML methods to DL-based approaches in EEG-BCI research is reflective of a larger trend within the field of artificial intelligence, where DL has consistently demonstrated its superiority over ML in managing complexity and enhancing predictive accuracy. Considering that DL methods have opened up new capabilities in interpreting and harnessing EEG signals, this review emphasises their ground-breaking impact. The spotlight on transformer-based DL methodologies is particularly warranted as they symbolise the latest evolutionary development, offering solutions to inherent ML challenges, including the dependence on handcrafted features and a limited scope of generalisation. Therefore, a thorough exploration of these cutting-edge DL techniques through this literature review is not just timely but critical for mapping out the future trajectory of EEG-BCI research.

The importance of transformers in BCI is multifaceted. First, their architecture, primarily designed for handling sequential data, aligns seamlessly with EEG data, which is inherently temporal. Additionally, transformers have shown an exceptional capacity to understand and represent long-range dependencies and contextual connections in data, making them particularly effective for EEG data, which is inherently temporal and complex [9–11]. By addressing latency issues, transformers can dramatically bolster the real-time classification of EEG signals, a challenge that has perennially plagued BCIs. Furthermore, by virtue of their depth and attention mechanisms, transformers can potentially ameliorate the signal-to-noise ratio in EEG data, subsequently improving the reliability and classification accuracy of BCI systems. Hence, this alignment of technical strengths with the long-standing challenges of BCIs underscores the exigency of a comprehensive review paper in this domain.

Additionally, the application of transformer models transcends mere classification improvements, extending into the crucial realm of signal denoising. The sophistication of transformer architectures enables them to excel in identifying and isolating relevant patterns within noisy datasets, a common challenge in EEG signal processing. The ability of transformers to handle sequential data makes them particularly suited for temporal signal enhancement, offering a new level of precision in distinguishing between noise and true neural signals. By leveraging these capabilities, transformers hold the promise of significantly elevating the clarity and quality of EEG data, which is paramount for the real-time functionality of BCIs. This advancement is not only expected to improve the reliability and accuracy of EEG interpretations but also to expand the practical applications of BCIs in complex, real-world environments where noise factors are prevalent. This additional focus on denoising underscores the versatility and comprehensive impact of transformer models in EEG-BCI research, warranting the in-depth investigation that this literature review aims to undertake.

This comprehensive review aims to map the burgeoning terrain of transformer-based EEG-BCIs, a field at the cusp of transformative developments. The advent of transformer technology, coupled with its nascent integration into EEG signal analysis, marks a critical juncture for the review of current methodologies and the charting of prospective research directions. Acknowledging the dominance of RNNs, CNNs, and traditional ML methods in EEG classification, this paper identifies a significant gap: the under-utilisation of transformer models in the BCI domain. Current literature on EEG-BCIs has largely overlooked the unparalleled efficiency of transformers in handling sequential data in combination with extracting long-range dependencies and contextual relationships in data, as well as their subsequent potential to enhance the precision and robustness of EEG signal classification and interpretation.

The necessity for this review stems from the observed deficiencies in existing studies, notably the scarcity of exploration into the application of advanced NLP techniques within EEG signal processing. Furthermore, while transformer models revolutionise NLP, their capability to redefine EEG-based BCIs remains largely untapped. By bridging this gap, our review serves as a critical contribution, elucidating the transformative impact of these models on the BCI field. It underscores the urgency for in-depth, targeted research to assimilate transformer technology into EEG-BCI development, paving the way for novel approaches that can tackle the complexities of brainwave data. Thus, this paper not only catalogues the current state of affairs but also serves as a clarion call for the BCI research community to broaden its horizons and embrace the potential of transformers, setting a trajectory for future innovations and advancements.

2. Background

Electroencephalography-based brain-computer interfaces (EEG BCIs) are emerging as a transformative technology in both biomedical applications and socioeconomic realms. In the biomedical sector, EEG BCIs hold significant promise for the development of assistive devices that empower individuals with severe motor disabilities, such as those resulting from stroke or spinal cord injuries, to communicate and control their environment. These interfaces translate neural activity into commands that enable users to operate software or hardware, such as speech-generating devices and robotic limbs, thereby providing a new lease on independence and interaction.

Additionally, EEG BCIs are instrumental in advancing neurorehabilitation techniques. They play a crucial role in facilitating neural plasticity, which is the brain's ability to reorganize itself by forming new neural connections. This capability is essential for recovery of motor functions in patients who have suffered neurological damage. By engaging patients in brain-driven tasks, these technologies help strengthen neural pathways and improve motor outcomes, enhancing the overall rehabilitation process.

Moreover, EEG BCIs are increasingly utilised for seizure detection and management in patients with epilepsy. By continuously monitoring brain waves, these systems can identify the characteristic electrical patterns that precede a seizure as shown in Fig. 1. This advance warning allows for timely intervention, such as the administration of medication or the activation of a neuromodulation device, potentially preventing the seizure or mitigating its severity. This application not only improves patient safety but also contributes to better management of epilepsy, reducing the burden on patients and healthcare systems alike [13]. Socioeconomically, the integration of EEG BCIs can lead to significant cost savings in healthcare by reducing the need for long-term care and rehabilitation services [14]. Moreover, as the technology advances and becomes more accessible, it is expected to play a crucial role in the workplace by augmenting human capabilities and potentially altering traditional job roles [15]. As the landscape of EEG BCIs continues to evolve, the convergence of interdisciplinary research and technological innovation paves the way for groundbreaking applications that challenge our current understanding of neural interaction and computational methodologies.

2.1. EEG and BCI basics

Different stimuli result in the activation of different brain regions, with mostly distinct patterns occurring as a result of different intentions of and reactions of stimuli to the brain [2]. This results in a range of different mental tasks [2,16]. The visual stimulation of subjects, for instance, will trigger different brain regions (i.e., the visual cortex) [17] than the participating subjects' limb movements (the coordination of which is likely to mainly take place in the primary motor cortex and the premotor cortex) [18,19], ultimately forming two distinct categories of mental tasks in BCIs as exemplary shown in Fig. 2.

BCI is an umbrella term for applications and devices that enable a direct communication pathway between the central nervous systems (i.e., the human brain) and an external, digital receptor (i.e., any mobile or desktop application) by translating brain signals into digital commands than can be read by machines [21]. Therefore, BCIs have generally been accepted not only to boast significant potential to improve the quality of life (i.e., for individuals in stroke rehabilitation or suffering from motor-neuron disabilities) but also to revolutionise the way society may interact with technology in the future. In addition, recent advances in Artificial Intelligence (AI), as well as data and signal processing in the past few years have extended the range of possibilities in the research and development field of BCIs, providing new waves of momentum to the research landscape in this domain [22].

Current research in Brain-Computer Interfaces (BCIs) offers a spectrum of approaches, with non-invasive EEG-BCIs, the focus of this review, representing one end, and invasive BCIs representing the other end [23,24].

Non-invasive BCIs establish a one-way information flow from the brain to external devices, allowing users to control external systems through thought commands. However, advancements are being made in invasive BCIs, such as those developed by Neuralink [24]. These devices hold immense promise for individuals with paralysis or neurological conditions by establishing a two-way communication channel between the brain and external devices.

Invasive BCIs directly interface with neural tissue, potentially enabling not only control of external systems but also the much-anticipated addition of sensory feedback. This could revolutionise the lives of paralysed individuals, allowing for a more natural and intuitive user experience.

It is important to note that invasive BCIs, while exciting, are still in their early stages of development. Surgical implantation is necessary, introducing inherent risks of infection, bleeding, and immune response [25]. Long-term safety and efficacy data are still being gathered. Additionally, the ethical considerations surrounding invasive brain-computer interfaces are substantial. Issues around user privacy, data security, and potential manipulation of brain functions necessitate careful ethical frameworks and ongoing public discourse [26].

Due to the focus on practicality and relative safety within the domain of non-invasive EEG-BCIs, this review will primarily explore advancements in this established technology. However, acknowledging the potential of invasive BCIs provides valuable context for understanding the broader landscape of BCI development. Non-invasive approaches currently offer a more readily available and established option, while invasive techniques hold the promise of future breakthroughs for those willing to accept a higher level of risk.

In contrast, non-invasive BCIs such as Electroencephalography (EEG) based applications can establish a one-way information transmission channel from brain to machine [27]. These BCIs are relatively cost-effective, efficient, and easy to use when developing and

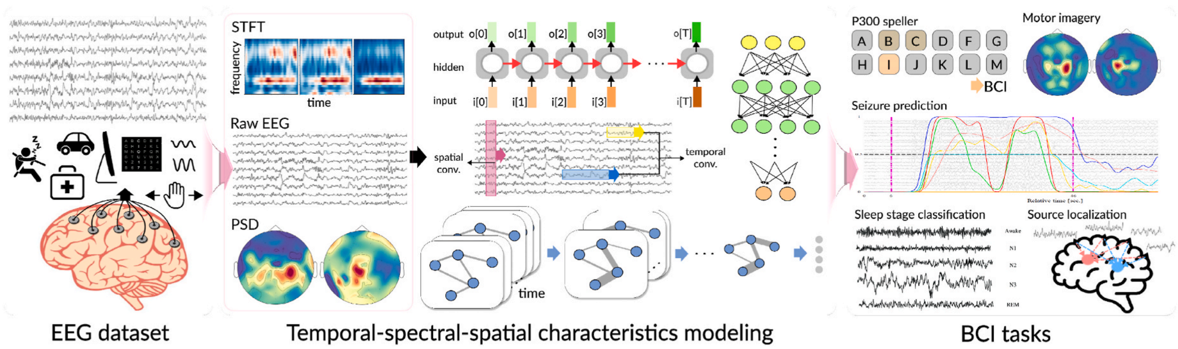

Fig. 1. Signal and data processing pipeline of a Biomedical application of a Brain-Computer Interface for seizure detection [12].

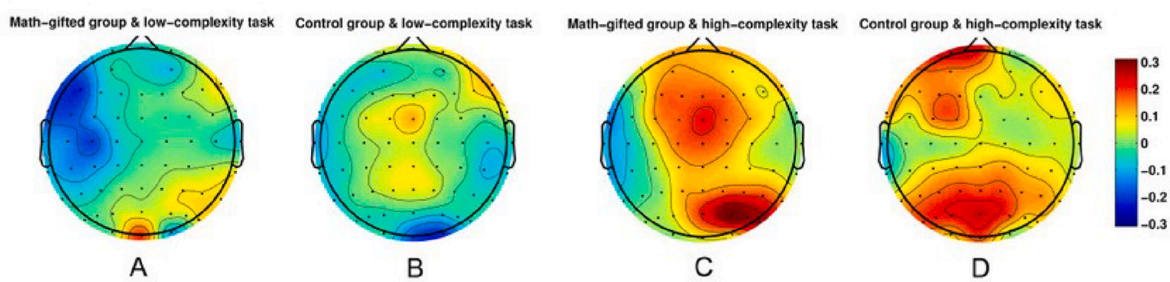

Fig. 2. Visualised neural activity of low and high complexity tasks across different control groups [20].

researching wearable BCI applications [28,29].

To improve the clarity and readability of this review, the abbreviation "BCI" is employed solely to denote non-invasive Brain-Computer Interfaces, unless explicitly stated otherwise.

For the scope of this review, there is a primary focus on the mental task of Motor Imagery (MI) due to its broadly reported intuitiveness for participants, as well as due to the abundance of research and data on this task's domain in the literature [30,31]. Knowledge gaps have been identified in the domain of general EEG signal processing using MI task recordings, however, other tasks, such as inner speech, may ultimately be incorporated in future studies as they hold immense potential for novel BCI applications.

As denoted previously, MI is a common mental task for EEG BCIs. It describes all kinds of tasks in which an individual subject is required to imagine specific motor functions (such as i.e., the closing of a hand, movement of feet, or raising of the tongue) without actually and physically executing the given task, hence the popular term imagined movement [22,32]. In this cognitive process, similar neural networks in the affected brain regions are activated when compared to actually executing the given movement or command, which reportedly makes it one of the most intuitive BCI mental tasks as well as the most suitable for people with motor-neuron disorders [33,34].

Generally, there are multiple EEG patterns/phenomena that are used to classify MI-EEG signals. The most used representations of the power increase and decrease of power are the Event-Related Synchronisation (ERS) and the Event-Related Desynchronisation (ERD) [35,36], which typically manifests within what is referred to as the Mu-band (8–13Hz) and the Beta-band (13 and 30 Hz), respectively [37].

2.2. Pre-processing and noise removal of EEG signals

EEG raw data is generally considered to be notoriously affected by a myriad of artifacts, which are unwanted signals and patterns within the data that arise from unwanted and undesirable events [38,39]. These events include but are not limited to, muscle movements throughout the body, electrode movements, eye movements, and environmental noise and the caused interference thereof [40–44]. Forthgoing in this work, the terms artifact removal, denoising and noise removal are used interchangeably.

These unwanted artifacts pollute the raw data and decrease chances for reliable and accurate classification, and therefore, need to be dealt with before feeding data into an AI of any sort. From basic denoising approaches such as via regression (which erroneously assumes that every channel in the given dataset is part of a total sum of representations of artifacts and source signal representations thereof [40–42]) to various wavelet transformation approaches (i.e. wavelet packet transform), various papers have been published over the past 20 years, with a myriad of numerous applications and assessments of their individual and overarching performances and utilities in EEG signal processing and denoising [41,42,45–47]. Since wavelet transformation and other "manual" EEG signal pre-processing steps are not essential in the proposed NLP-based EEG-BCI, it is covered only briefly in this work. However, its impacts on EEG signal processing and BCI development are significant overall. Generally, wavelet transformation enables pre-processing EEG signals by decomposing the recorded signals into various frequency components [48,49]. The processed signals then enable the detection of transient events (and to determine other changes that occur locally) as the processed data now can be observed in its time and frequency domain at the same time [50–52].

Besides the previously mentioned approaches, in EEG there are other popular pre-processing strategies, perhaps most notably the Principal Component Analysis (PCA) and Independent Component Analysis (ICA) – both of which fall into the domain of Blind Source Separation (BSS) [53]. Summarising, PCA identifies the most significant signal components by estimating high variances in the data, whilst ICA separates original sources independently [54]. This in turn again means that ICA assumes all observed signals to be a linear combination of all source signals, which then allows for the extraction of individual components (i.e. artifacts such as eye movement, muscle movements) [54,55]. Therefore, given its ability to identify statistically independent components, ICA is often the preferred method of proceeding in EEG signal pre-processing [56–58].

Given both the scope of this work and the proposed research's focus on deep learning, however, the following sections will not reference traditional ML benchmarking performances but rather outline a successive development towards Deep Learning-based approaches.

2.3. Data analysis for EEG-based BCIs

2.3.1. Traditional ML

To summarise, traditional Machine Learning (ML) has been utilised in EEG data and signal processing for many years, particularly in the domain of MI task data processing [34,59]. Irrespective of the model, these "traditional" ML approaches do not operate well when faced with raw EEG data. Hence, the feature extraction (and therefore, its accuracy and performance) prior to feeding data into the actual ML classification algorithm is vital to curating an advanced ML BCI solution [32,60].

Generally, the step of the feature extraction itself aims to extract vital information from the raw EEG signal data (herein referred to as raw data), so it can be used more effectively by any ML model. Over the years, a plethora of feasible feature extraction techniques have been developed and scrutinised by researchers [22,61]. Common feature extraction methods for raw data include frequency filtering, spatial filtering, time-domain analysis, and frequency-domain analysis [45,59,62,63]. Depending on raw data, task, subject and the technique deployed, different features may be derived, such as band power ratios or spectral bands. However, many statistical features have been utilised as well, including, but not limited to, signal variance, autoregressive coefficients, or entropy [64–66]. However, it is important to note that these steps basically describe an almost hand-crafted, labour-intensive, and (more or less) manual selection of features, based on few (and often generalised) cognitive assumptions about the prevalent raw data.

Upon retrieving the desired features, one must deploy a traditional ML classification algorithm to train (and ultimately test) the system. Widely used and common ML models such as Naïve Bayes, Linear Regression (LR), Random Forests (RF), and Support Vector Machines (SVMs) have been utilised frequently to categorise EEG signals over the past two decades. Although these have demonstrated good capabilities, as well as indicated a solid generalisation ability, all these ML models ultimately do come with several technical flaws concerning the BCI, with several limitations of which few are briefly discussed below.

Firstly, and as mentioned above, the feature engineering process in itself heavily relies on human and domain expert knowledge and basically represents a manual approach as per today's modern data processing standards. This makes any application prone to error and under-leveraging the given data prior to the AI's training phase already, as subtle yet perhaps relevant features may be completely overlooked by even the most experienced experts [67–69]. This limitation also is one reason for the underdelivering of ML models in terms of capturing complex relationships within EEG data. Moreover, the manual approaches are simply time and labour-intensive, which is not an attractive attribute moving forward.

Secondly, there is a (although debatable) lack of capability to generalise the ML model [70,71], with various researched ML models having reported poor generalisation ability on different tasks, particularly given the manual (and therefore, static, or fixed) nature of the pre-processing and feature extraction pipeline, which usually would need to be adapted from trial to trial, task to task, as well as individual to individual [72].

Lastly, several researchers have reported that ML-based models do not necessarily encapsulate any hierarchical representations of the original data within their final model [73,74]. This means that the final model's capability to accurately classify EEG signals is further limited and usually shallow in nature.

In summary, despite the high-performance metrics achieved in laboratory settings and various limitations inherent to traditional Machine Learning (ML) methodologies, the primary emphasis of this review will be directed towards the exploration and development of innovative Deep Learning (DL) techniques. These DL methods offer several advantages, including superior capabilities for end-to-end learning, enhanced classification accuracy, and more effective capture and modelling of nonlinear relationships within signal data. Furthermore, DL-based approaches demonstrate increased generalization and adaptability in Electroencephalogram-Brain Computer Interface (EEG-BCI) models compared to their ML-based counterparts [59,67,68,73,75–77].

2.3.2. Signal classification using CNNs

With the advances of AI beyond the domain of BCI research and development, various DL approaches have been proposed, most notably using Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), the latter of which seem to dominate the research space in terms of research output using CNNs as methodological preference [67,76–79].

Given the stringent focus on DL-based approaches of this review, as well as the previously mentioned omnipresence of CNNs in this research domain, the subsequent section addresses the basic methodology of CNNs as a classifier, its achievements, and its limitations.

CNNs have become a powerful and guiding component in the development of much of DL-based research and development, both in the academic and the private and industrial sectors. CNNs (and their popularity) originally emerged from the field of digital image processing, pattern recognition and computer vision [80,81]. However, their fundamental characteristics and naturally advantageous attributes have made them a popular choice in other domains over time, EEG signal processing and BCI development being one of which [82]. In fact, several researchers have successfully demonstrated that CNNs are well-positioned to leverage EEG signal data given their inherent competency to facilitate the learning of hierarchical representations in data (which ML models often neglect). This ability, combined with the fact that no manual feature selection is needed due to the CNN's auto-feature nature [83], position them as an extraordinarily attractive methodology for EEG signal classification, which is the reason why they have widely been used for EEG signal classification for years now [84,85].

Although the feature selection is not required to be performed manually with CNNs, there needs to be some sort of pre-processing of the raw data as to that the CNN can retrieve the data in the correct input format, so it can automatically select the features during training. There is no fixed standard procedure on how to conduct pre-processing stage, or as to which methodology and technique for pre-processing is to be deployed. However, there are arguably some industry standards or good practices that help get the most out of the models by utilising some basic signal processing methods such as normalising amplitudes or implementing filters (including, but not limited to band-pass filters) [13,76,79].

In addition to the basic noise removal using i.e., filtering, it is necessary to perform a segmentation of the EEG data, which basically means splitting the usually long recordings (usually multiple minutes) into smaller time segments with a fixed duration, i.e., 4 s per "frame" [86]. This is necessary to ensure that we can feed individual samples into the CNN that all are of equal size, just as one would do when feeding image data into a CNN. Given $n_{channels}$ by $n_{electrodes}$ of the EEG apparatus, the final dimensions of the CNNs input layer DimInput therefore must be equal to $[n_{channels}, n_{timesamples}]$.

CNNs come in different shapes and with different hyperparameter settings, and various layer types and overall architectures. As indicated by their name, their main character is derived from the implementation of convolutional layers, which have the ability to automatically learn local feature representations within data [87]. These layers generally crate feature maps which are often subsequently down samples in their spatial dimensions using pooling layers [87,88], ultimately distilling the most important information of the prior convolution and improving the CNN's overall translational invariance. Through this structure, CNNs enable an automated feature selection, therefore reducing the need for manual intervention during feature extraction.

One such example is delineated by a novel Ensemble pre-trained Convolutional Neural Networks (ECNNs) method for emotion recognition from EEG signals of individuals watching music videos was introduced [89]. utilising scalograms generated through Continuous Wavelet Transform and retraining multiple CNN models, this approach significantly enhances the accuracy of recognising emotions. This methodology was tested on the Database for Emotion Analysis using Physiological signals (DEAP), a widely recognised dataset for analysing physiological responses to emotional stimuli, demonstrating notable success using the proposed methodology [89,90].

Another recent CNN-based approach highlighted the capability of CNNs to interpret complex brain signals for emotion recognition by utilising directed Directed Transfer Function (dDTF) to generate images that represent brain activity, which are then analysed using fine-tuned CNN models like AlexNet, ResNet-50, and more. Tested on the MAHNOB-HCI and DEAP databases, the ResNet-50 model, in particular, showed exemplary performance in capturing emotional states through EEG signal connectivity, achieving high accuracy and F1-scores [91].

Continuing the exploration of CNNs for emotion recognition, this study leverages EEG signals to investigate the relationship between emotions, brain wave patterns, and PTSD. By employing a novel automated CNN-LSTM with ResNet-152 algorithm, the research addresses previous challenges in emotion analysis, achieving a notable accuracy of 98 % [92]. These studies underscore the potential of hybrid deep learning algorithms in accurately identifying emotional states, demonstrating significant advancements over traditional methods.

It can be summarised that CNNs have become an integral part of data-driven analysis and research in the domain of BCI development over the past years and have been applied to a myriad of mental tasks, including, but not limited to MI classification and emotion recognition. By deploying CNNs to classify EEG data, researchers have been able to demonstrate improved robustness, improved generalisability, as well as increased classification accuracies [29,44,79,93,94]. In addition, the use of CNNs has been combined with additional DL models, such as Long-Short-Term Memory Networks (LSTMs) or RNNs [68,79,95,96]. These combinatory approaches have further provided evidence to suggest an improved capturing and handling of complex and noisy EEG data.

In this chapter, we explored and discussed some of the CNN-based BCI applications and unravelled how CNNs enable an automatic handling of non-linear relationships in data, pose an improved generalisability, and easier adaptation to individual features of subjects. Overall, it can be concluded that the overwhelming majority of DL-based EEG-BCIs in published research incorporate some kind of CNN, due to many of the previously described advantages over other architectures and the manual ML approaches, as well as because of their prominence in other domains.

Most recently, however, a relatively new methodology in DL has started to baffle the global tech scene, whilst simultaneously creating inevitable momentum across all research disciplines – Transformer Networks, herein referred to as Transformers, have made a colossal impact on the private sector within weeks of the release of the public beta of OpenAI's ChatGPT 3.5 in November 2022 [97]. In the next section, the methodology of Transformers is explained before outlining and discussing how to potentially utilise them for building EEG-based BCIs in the future.

3. Review methodology

3.1. Search strategy

The methodology for this review was systematically designed to encompass an exhaustive search of the literature pertaining to the application of transformer models in EEG-based BCI systems. The literature search was conducted across electronic databases with a strong focus on Scopus and PubMed [98,99]. The search strategy was tailored to include a combination of key terms and phrases related to EEG, BCI, CNNs, and transformer models using AND operators exclusively across queries.

3.2. Inclusion and exclusion criteria

The criteria for including articles in this review were as follows.

- Published in the period between January 2018 and December 2023 to ensure both recency and relevance of the findings.
- Articles that discuss the development, implementation, or evaluation of transformer models in the context of EEG data analysis and BCI applications.
- Studies that provide empirical evidence on the performance, advantages, or limitations of transformer-based approaches.

The exclusion criteria were defined as follows.

- Articles not published in English.
- Studies that do not explicitly focus on EEG-based BCI systems or do not employ transformer models (or CNNs, see Table 1).
- Conference abstracts, editorials, and non-peer-reviewed literature omitted unless indicated otherwise.

3.3. Search outcome

The initial search for research output including the keywords BCI and EEG yielded 5052 records, of which 2465 entries constitute original research articles. Upon narrowing down the search by including BCI EEG entries using CNNs, a total of 666 entries were identified, of which 367 are research articles. In comparison, the same search for transformer-based research in this field queried a total number of 29 research articles for the period of 5 years only.

These findings have been summarised in Table 1 below, which shows the results of the total counts per query, Fig. 3 shows the annual publication counts of each query, excluding the initial BCI AND EEG query in order to highlight the research trends in this niche research area of Transformers in BCI.

4. Transformers

Relatively recently (on the June 12, 2017), the first transformer networks were proposed by a couple of high-profile Google engineers in a Paper called "Attention Is All You Need" [100], which arguably paved the way for a new era of processing sequential data, particularly for applications in NLP [101,102]. Prior to their introduction, CNNs and RNNs had an almost exclusive position at the forefront of NLP research and applications [103–105]. Despite their advantages as described in the previous sections of this paper, however, RNNs, and particularly CNNs, experience stringent limitations and disadvantages when it comes to capturing dependencies over medium to extended periods of time – which is why the disruptive paper by Vaswani [100] introduced the transformers ability to leverage so-called "self-attention" mechanisms, which allow capturing more dependencies between all given points and positions within a sequence, without any regard to their individual proximity to each other [100,106,107]. This puts transformers in a

Table 1
Scopus Query Results; Total Document Counts for the past 5 years with breakdown by keyword combinations using AND query operator.

Keywords	Total Results	Research Articles
BCI, EEG	5052	2645
BCI, EEG, CNN	666	367
BCI, Transformer	62	32
BCI, Transformer, EEG	51	29

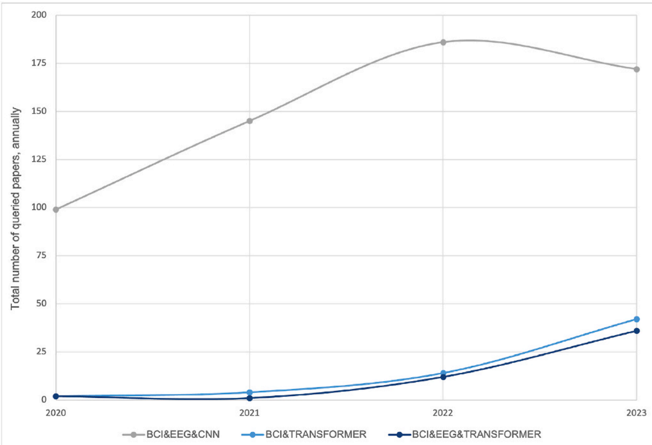

Fig. 3. Comparison of annual research outputs according to the Scopus search results including articles, conferences, books, and more. Transformer-based approaches start to increase towards the end of 2023, while not being a cornerstone of contemporary BCI research.

competitive position compared to previous DL approaches, as this global attention span provides the ability for the model to capture both global and local features, which is an important advancement, particularly in the domain of NLP, where long pieces of texts and entire paragraphs are required to be analysed in various contexts [100,108]. However, since their introduction, transformer models have been shown to provide substantial benefits to a myriad of computational analytical problems, such as in computer vision or speech recognition [109–111].

Whilst CNN-based approaches utilise local convolutional filters to extract and learn spatial and temporal features, transformers generally handle this using self-attention mechanisms [100,112]. Transformers also process input sequences in parallel, which is a more efficient mechanism to capture global relationships within data when compared to CNNs, which mainly proceed hierarchically, and learn features by scaling through multiple successive convolutional procedures [87].

Therefore, in this section, an introduction to the transformer model's architecture and methodology is given, before discussing their potential applications for BCI research studies in the future.

4.1. Architecture

Transformers essentially contain a dualism between an encoder and a decoder, in which input data is processed via the encoder and predictions (outputs) are being produced based on the encoded output of the encoder [40,100,108] as shown in Fig. 4.

4.1.1. Encoder

Positional encoding is crucial in the Transformer Encoder (TE) to compensate for its lack of sequential processing, using sinusoidal functions to establish order within sequences [100] (see Fig. 5). In general, sinusoidal functions ensure that all positional encodings remain differentiable by deploying a geometric progression of frequencies using Sine and Cosine functions [100,113]. This advantage also boasts the ability of the transformer to generalise any sequences of variable lengths and ensures the transformer's capability to make use of various length sequences [111,114]. Overall, however, employing functions other than sinusoidal ones is possible as well and may be subject to research in the near future.

In both NLP and non-NLP tasks, the TE processes input sequences through multiple identical layers, each consisting of sub-layers. These sub-layers include multi-headed self-attention layers followed by feed-forward networks, with residuals and normalisation layers integrated throughout [100,108,114]. Each of these sub-layers within the TE is escorted by residual links, which again adds the previously original

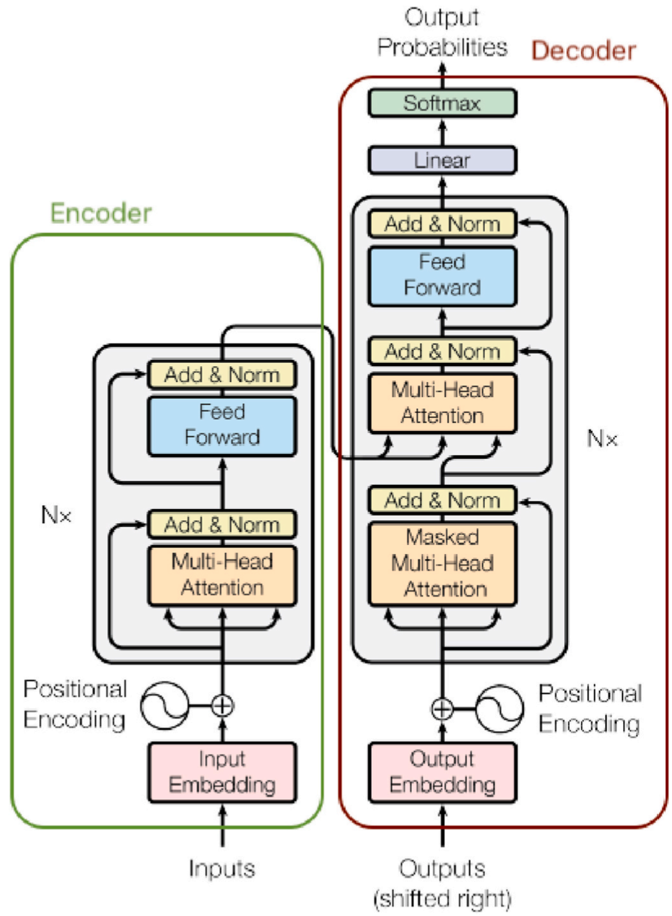

Fig. 4. Original architecture of the first transformer model as introduced by Google Engineers in 2017 [100], modified to highlight both the encoder and the decoder components.

input in addition to the output of the sub-layer, which generally is perceived to be able to prevent effects of a vanishing gradient which may occur just like in the training of CNN or other NN-based DL networks. Lastly, within the sub-layer, the layer normalisation is utilised in subsequence to the previous residual links, ultimately improving the adjustment effectivity, as well as soothing the training progression itself [100,114,115].

Within the TE, the multi-headed self-attention layer (MHSA) provides the model's capability to capture various features and aspects of the sequential input [100,111,114]. Using the MHSA, learnable matricidal representations of data are formed, which are the key (K), the query (Q), as well as the value (V) vectors, all of which can be configured based on the initial input [116–118]. A self-attention score is computed by utilising the dot product, and by utilising a square root of the key vector dimension to scale the result, which measures the degree of relevance between each of the keys and the query [118,119].

Now, for processing the outputs of the MHSAs, position-wise Feed-Forward Networks (FFNs) are deployed as previously mentioned, which enables transformer models to extract and analyse more versatile relationships and patterns of the original input data. For the generally proposed transformer architecture, FFNs within the TE are comprised of two linear layers that encapsulate some kind of activation function [93]. Interestingly, a more efficient parallel computation of the computation of individual sequential elements under disregard of element-wise positioning and positional relationships is enabled by applying the TE's FFNs to all positions of any input sequence independently [93,100].

Again, this combination of self-attention and FFNs within the sub-layers has been reported to enable the TE's capability to extract both

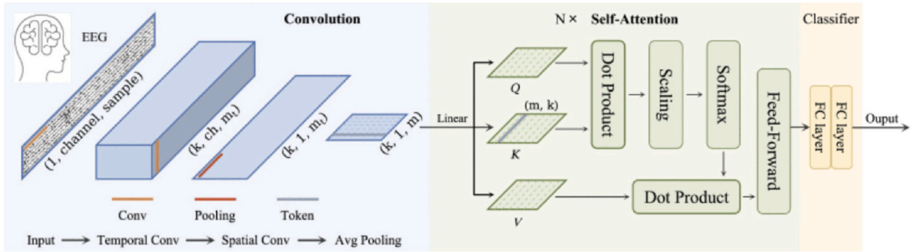

Fig. 5. EEG Conformer Architecture as proposed by the researchers, with both convolutional and transformer components [156].

diverse and rich features from the originally provided input. In terms of NLP, this enables a more nuanced and human-like interaction with the sequential data and allows for more efficient post-processing using i.e., sentiment analysis to perform a wide range of NLP tasks [103,108]. This concept may be a valuable and intrinsically significant component to unlocking improved denoising within EEG data processing when dealing with sequential EEG data. Therefore, this too may be subject to future research and may be considered when designing a transformer-based EEG encoder system. However, one shall consider the risk of overfitting as well, and perhaps aim to experiment with various dropout links and regularisation techniques when designing a novel transformer's building blocks.

Summarising, the design of the first of the two main components of the transformer architecture (the TE) has a significant impact on the transformers' overall ability and decision-making process. Therefore, adaptations and variations of the encoder design may be subjected to future research, as there is a substantiated and ongoing need to refine the TEs' degree of functionality.

4.1.2. Decoder

For each transformer's architecture, the transformer decoder (TD) is the second main component. Like the TE, the TD compromises multiple layers with several sub-layers [100]. More importantly, however, its main responsibility lies within its function to generate the corresponding output sequences based on the encoder's outputs (i.e., continuous representations of the positionally encoded raw data). Generally, as per their initial proposal, the TD layers again are organised as an MHSA, encoder-decoder attention (EDA) mechanism, as well as a position-wise, fully connected FFN, each of which will again be succeeded by residuals with normalisation just as previously described for the encoder module TE. Masked self-attention is generally deployed within the first TD layer, as one of the critical aspects of transformer training is to inhibit the TD from "relying" on future information when proceeding with the generation of sequential outputs [120,121]. Masked self-attention can help circumvent the occurrence of this phenomenon by modifying the MHSA input and setting their attention scores close to negative infinity, effectively making them (or masking them) inaccessible or unfeasible for considerations of inclusion within the generated output sequence [120,121].

In contrast, the EDA is a structure exclusively found within the DT components of the transformer. Its main function has been reported to be attending to the TE's output sequence (i.e., the TD's input sequence) by comparing, and if necessary, correcting, the generation of relevant outputs by the encoder (in NLP, this would be referred to as ensuring that outputs correspond to the correct semantic context). To do so, the EDA applies the same computational efforts as previously discussed within the TE module but with a small yet significant difference: Source values for K and V are derived from the encoder's input, and Q is derived from the output of the preceding layer within the TD module [100,122].

Summarising, the EDA mechanism is like the MHSA, yet uses different source values for its query matrix and the operations thereof. Lastly, the final (i.e., third) TD layer deploys another FNN, just as previously explored when describing the TE architecture.

4.2. Training and loss function

Regarding the training of these models, transformers models allow for the implementation of supervised end-to-end learning, with general similarities to training considerations as i.e., in CNNs, where the most common optimisation objectives are ensured by deploying an analysis of cross-entropy loss during training [123,124]. Just like as observed in various other DL techniques, the training of transformers can quickly result in high computational expenses and the need for extended hardware resources, the need of which generally scales further with increasing model complexity and the dataset's magnitude.

4.3. Performance of transformer-based models

Transformer architecture offers several advantages over traditional CNNs and RNNs for sequence modelling tasks. As per the following pages, its inherent parallelism, self-attention mechanism, well-suited encoder-decoder structure, and flexibility make it a powerful tool for various applications [9,100,125]. However, researchers are still exploring ways to address the computational cost and interpretability challenges associated with Transformers. In this section, their general advantages and disadvantages are delineated before delving into their impact in the BCI domain.

4.3.1. Advantages

Unlike RNNs, Transformers achieve faster training times due to inherent parallelism in the self-attention mechanism [100]. This is evident in a study by Ref. [126], where their Transformer-XL model achieved state-of-the-art performance on various language modelling benchmarks while requiring significantly less training time compared to RNN-based models. Research from other domains such as speech recognition confirm these findings, such as a study by Ref. [127], in which the authors demonstrate significant performance benefits of Transformer models over RNNs in various benchmarks, including a notable superiority in 13 out of 15 automatic speech recognition tasks [127].

Moreover, the Transformer's encoder-decoder architecture with attention is particularly well-suited for machine translation. Vaswani et al. (2017) demonstrated that their Transformer model achieved significantly better translation quality on English-to-French and English-to-German tasks compared to traditional RNN-based encoder-decoder models [100]. This improvement was attributed to the attention mechanism allowing the decoder to focus on relevant parts of the encoded source sentence during translation, leading to more accurate and coherent translations [100].

Another advantage of Transformers may be their significant flexibility, allowing for adaptation to various tasks, such as recent advancements like the Vision Transformer by Dosovitskiy et al. (2020) further demonstrate this flexibility by achieving competitive results on image classification tasks traditionally dominated by CNNs [128].

However, this versatility of transformer-based models is underpinned in various other research domains, too, including protein structure prediction applications in which the performance of the transformer-based models outperformed state-of-the-art unsupervised structure learning methods by a wide margin, with far greater parameter efficiency than prior state-of-the-art protein language models [129]. Other examples of applications in which transformers outperformed previous state-of-the-art models (such as RNNs and CNNs) were recently given including, but not limited to long-sequence time-series forecasting [130], traffic-flow prediction [131], molecular dynamics-based drug discovery [132], and more.

4.3.2. Disadvantages

While parallelisable, transformers can still be computationally expensive to train compared to simpler models [100]. This may limit their application in resource-constrained environments. In addition, the interpretability of how these models make decisions can be challenging due to the complex nature of the self-attention mechanism, which currently hinders interpretability compared to more simple models [133,134].

In conclusion, the Transformer architecture offers several advantages over traditional CNNs and RNNs for sequence modelling tasks. Its inherent parallelism, self-attention mechanism, well-suited encoder-decoder structure, and flexibility make it a powerful tool for various applications. However, researchers are still exploring ways to address the computational cost and interpretability challenges associated with Transformers. Whilst powerful, they also face significant challenges such as high memory requirements and susceptibility to overfitting. The extensive memory demand arises from the self-attention mechanism, which computes interactions across all elements in the input sequence, making them less viable for tasks with very long sequences or limited computational resources. Additionally, their large parameter space increases the risk of overfitting, especially when trained on smaller or less diverse datasets, where they may fail to generalise well to new, unseen data [100,135,136].

4.4. Transformers in BCI research

4.4.1. Overview

As discussed in the previous section, transformers have been found to be a powerful deep learning architecture traditionally used in NLP, which however can be highly valuable for EEG data processing and MI classification tasks. However, since transformers are a relatively recent development, with an even more recent introduction to EEG processing, only a few applications have been explored so far - only 29 original research articles were found via *Scopus* query as shown in Table 1.

EEG signals generally comprise non-linear, temporal relationships which cause even RNNs to face problems capturing the patterns properly, usually under the interference of the vanishing gradient problem [137,138]. Transformers may be able to capture these long-ranging dependencies more effectively than RRNs (and CNNs) due to their self-attention mechanisms as described in the previous section, ultimately leading to an improved understanding of temporal features in i.e., MI tasks.

Another promising consideration in the design of new BCI applications using transformer DL models is the fact that they can be deployed using parallel processing, despite them being sequential models. RNNs are not able to do this, which may enable transformers to a more efficient, cost-effective, and more timely analysis in various. This also makes them more suitable for applications outside of the lab, where models with swift processing and response times are required in real-life settings, which currently is another main hurdle yet to be taken using the DL-based BCI applications that are currently being researched [107,139].

As transformer models can attend to both temporal information and spatial information simultaneously, the case for a transformer-based EEG analysis further solidifies as this implies that transformer-based BCIs could carry a superior discriminative capability for the various kinds of BCI mental tasks when compared to either CNN or RNN-based approaches [100,140].

In addition, not only do transformers support end-to-end supervised learning, but they do also incredibly well in applications where adaptability and Transfer Learning (TL) are required. In fact, this may be one of the strongest advantages advocating for the case of future transformer-based BCIs, given the fact that all currently popular and commercial language models (i.e., Google's Bard or OpenAI's ChatGPT) are pre-trained on training data of enormous magnitudes, and improve over time using transferable knowledge [102,108,141]. This in turn concludes that it would be possible to overcome some of the current limitations of BCI development (i.e., inter-subject variability in brain activity) using effective transfer-learning on more generalised and well-informed transformer models, similar to TL approaches using CNNs and RNNs that researchers in this domain have undertaken previously [78,79,142,143]. Here, the exact methodology of the novel, pre-trained BCI transformer is to be addressed in future research before deploying it on collected data, i.e., by experimenting with approaches using the open-source EEG-BCI baseline datasets such as the *physioNet EEG Motor Movement/Imagery Dataset*, which contains 109 patients and is a commonly used in EEG BCI research [144].

Next, transformer-based EEG signal analysis may pose additional advantages in certain settings (i.e., when deployed as medical devices), given their potential evolving around interpretability, ultimately paving the way for potentially developing an Explainable Artificial Intelligence (XAI) [145]. Such XAI-BCI would be useful, particularly in clinical environments, where practitioners and clinical experts are often faced with liabilities and responsibility when basing any diagnostic decisions of medical AI systems, which more often than not occur to the human decision makers as black boxes [146]. However, a transformer-based signal analysis can easily visualise its learned attention weights [145], which ultimately would enable the decision makers (as well as researchers that previously trained and tested the models) to make more informed decisions by understanding the time points and regions within the EEG recording that have been most apparent for the decision of the classification model.

Lastly, transformer-based EEG-BCIS may be less susceptible to noisy data. EEG signals are infamously contaminated with various artifacts and other noise from various sources [42,44,147], which is often considered one of the main hurdles yet to take in the development of reliable and effective BCIs. Applications have shown that transformed-based applications could potentially be less susceptible to noise, and more effective in removing unwanted signals without depending on human interference [40,107,148]. One such application is a recent piece of work in which researchers have proposed an EEG decoding method using the Spatial-Temporal Tiny Transformer (S3T), which utilises attention mechanisms to transform EEG data into a more characterised representation [140]. By considering EEG spatial and temporal information, the proposed S3T captures global dependencies and determines the importance of different feature channels for classification [140]. According to the researchers, this method is more cost-effective and consists of pre-processing, spatial and temporal transformation, classification, and additional merging steps. The S3T demonstrated promising potential for EEG decoding, enhancing classification accuracy and efficiency. However, decoding is not the only aspect that may be able to be improved by the implementation of attention mechanisms by transformer models.

4.4.2. Contemporary research and preliminary findings

In one of the few recent publications from January 2023, researchers have adopted the previously mentioned transformer-based deep learning network approach for MI classification tasks on EEG data [149]. The researchers utilised a Temporal Spatial Transformer Network (TSTN), which incorporates three kinds of key processes, which was basically implemented by adopting the architecture as proposed by Ref. [140].

Firstly, TSTN employs a Common Spatial Pattern (CSP) approach to construct discriminative relationships by employing several spatial filters. Different brain functions are attributed to distinct brain areas, and therefore, the implementation of CSP allows for the creation of particular spatial filters to extract task-induced neuronal activities effectively [150].

Secondly, TSTN performs using self-attention to improve the data as obtained through the CSP. When juxtaposing different prevalent methodologies (i.e., RNN or CNN), where the choice of kernel size is crucial for classification, TSTNs do not require explicit kernel selection [151–153]. Instead, self-attention is applied to the feature-channel data, allowing for the weighting of relevant channels pertaining to an individual's performance of the given mental task. In addition, CNNs generally disregard temporal dependencies of time-series information [149,152].

Lastly, the Temporal Segment Transformer Network (TSTN) as proposed by the researchers segments the improved feature-channel data into information batches and utilises transformation techniques to identify dependencies along the temporal dimension. In comparison to RNNs or long short-term memory networks (LSTMs) as utilised in prior literature, both RNNs and LSTMs suffer from the issue of vanishing gradients [137,149]. Therefore, RNN or LSTM-based solutions may encounter difficulties when faced with analysing MI-task-based EEG data with prolonged time frames in which participants perform the corresponding actions for the tasks. For instance, one example the authors of this article referenced is stroke patients, who generally are expected to process the information only slowly when moving the corresponding limb, requiring an extended time for executing the given MI tasks [149,154].

Amongst other findings of the study, it was found that in various scenarios, the acquisition of EEG signal data points is closely related to factors including, but not limited to, a subject's emotional state or cognitive abilities [149]. Artificial neural network-based approaches, on the other hand, demonstrate better flexibility in adapting to sequential data learning [155]. However, one notable drawback of the transformer-based network, such as TSTN, is its substantial model size. The implementation of TSTN on wearable devices might be challenging due to the significant training resources needed and the prerequisite of elevated computing power to result in high accuracies.

Nonetheless, these research findings have highlighted the advantages of the transformer-based approach, which effectively incorporates spatial filtering, self-attention, and multi-head transforming to address the challenges associated with EEG data processing in MI classification.

Another recent publication by the same researchers in December 2022 proposed an *"EEG Conformer"*, which resembles a convolutional transformer for the computation of EEG decoding as shown in Fig. 5 [156].

The architecture as proposed by the researchers encompasses not only transformer modules, but in fact three key elements: a convolutional module, a self-attention module, and a fully connected classifier [156]. For the convolutional component, the input comprises the raw two-dimensional EEG trials. By applying convolutional modules on both temporal dimension and electrode channel dimensions, the researchers propose to attend temporal and spatial convolutions. As reported by their work, this would allow for the model to assess both temporal and spatial patterns that are inherent in the EEG data. To enhance generalization and suppress noise interference, subsequent pooling is performed [156].

Subsequent to the convolutional layer, the output as spatial-temporal representation is processed by the self-attention module, where long-term temporal dependencies are extracted [156]. According to the researchers, attending to these dependencies provides the model's capability to capture essential relationships and patterns across time points.

The pre-processing steps for this application included the application of a Chebyshev filter, which is a type of digital filter commonly used in signal processing, including EEG signal pre-processing. It is designed based on Chebyshev polynomials, which provide the optimal trade-off between the filter's ripple in the passband and its roll-off rate in the stopband [157,158]. Generally (as well as in EEG signal pre-processing), Chebyshev filters are used to attenuate or eliminate unwanted frequency components, such as noise or artifacts, while preserving the desired frequency components related to brain activity [159,160]. They can be particularly useful for filtering EEG signals due to their ability to efficiently suppress noise and provide well-defined frequency response, since some of their advancements as compared to other filter types include a Steep Roll-off, an adjustable ripple, as well as a generally compact filter design: Chebyshev filters can achieve a sharp roll-off in the stopband, meaning they can quickly attenuate frequencies outside the desired passband [157,160,161]. This characteristic allows for the effective removal of unwanted noise or interference from the EEG signal. In addition, they offer control over the trade-off between the passband ripple and the stopband attenuation. This is possible because Chebyshev filters are designed based on Chebyshev polynomials, which optimize the ripple in the passband and the roll-off rate in the stopband. By adjusting the filter parameters, the user can customize the level of passband ripple and stopband attenuation according to specific requirements [157,161]. This flexibility allows for fine-tuning the filter's performance to balance the preservation of desired frequency components and the suppression of unwanted noise or interference in EEG signal processing [162]. Lastly, as indicated before, Chebyshev filters can furthermore achieve a given level of performance using fewer filter coefficients compared to other filter designs. This compactness is advantageous in terms of computational efficiency and memory requirements.

However, there are some drawbacks to selecting this kind of filter when compared to other pre-processing filter designs: Chebyshev filters with steep roll-off can introduce some signal distortion, particularly in the transition band between the passband and the stopband [162,163]. This distortion may affect the shape and timing of EEG signal components, potentially altering the characteristics of interest. In addition, like many other filters, Chebyshev filters can introduce phase distortion, leading to a time delay or phase shift in the filtered EEG signal. This phase distortion may affect the synchronisation of EEG signals with other analyses that are time-dependent, such as the study of event-related potentials (ERP).

Another drawback (although non-technical) is that designing Chebyshev filters requires determining parameters such as the ripple tolerance, stopband attenuation, and cut-off frequency. Selecting appropriate values for these parameters can be challenging and may require expertise in signal processing.

However, other filter types may be deployed first in the proposed architecture to evaluate their actual impacts on final classification outcomes.

In the proposed application by the researchers, the overall results of the EEG conformer slightly outperform binary classification results as given per state-of-the-art CNNs on three separate datasets, with the researchers demonstrating how well the introduced transformer module can improve overall performance by enhancing the segregation ability of the model as shown via their implementation of t-distributed stochastic neighbor embedding (t-SNE) [156]. As per Fig. 6, the transformer enables easier discrimination between data point clusters.

Whilst the authors of this novel research demonstrated the potential utility of transformer networks in the binary classification of EEG signals in MI, no other mental tasks have been assessed. In addition, this is highly conceptual research only which does not necessarily improve any real-time BCI performance given the already high performances in binary classification MI tasks. Consequently, the magnitude of the influence exerted by transformer-embedded systems and attention-based components remains an unexplored area of study. In fact, the researchers themselves further mentioned how "the role of multi-heads in the self-attention module remains unclear" [156], which boasts further space for research in this domain. Hence, the impacts of transformer networks on multi-class MI classification tasks may be subjected to

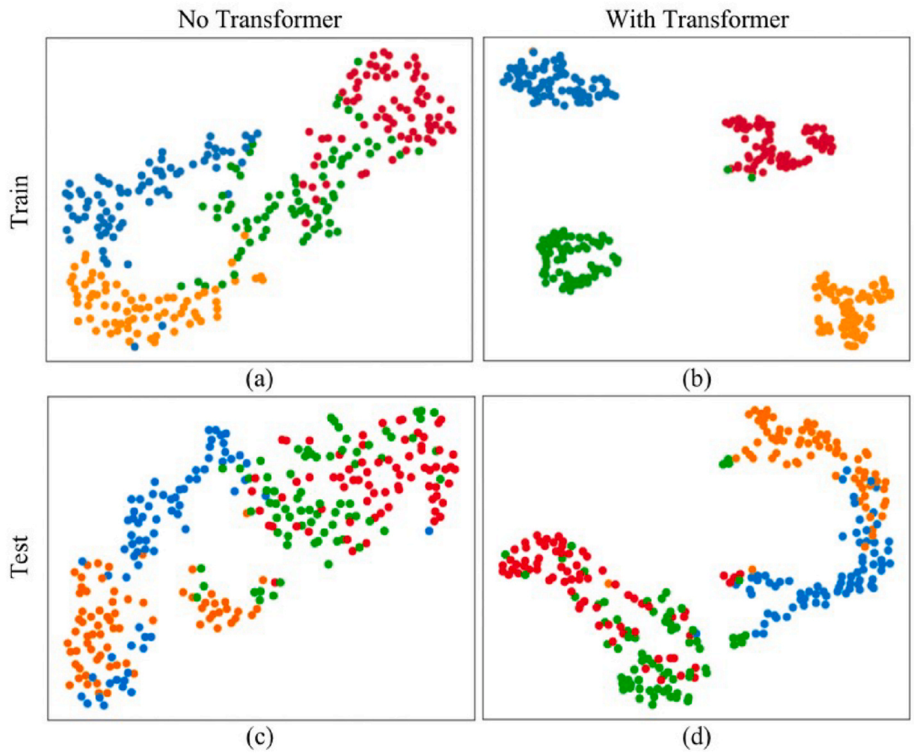

Fig. 6. Feature learning results in categorical discrimination in t-SNE visualisation by the researchers, with and without transformer module.

future research as it may pose the most impactful advantages of transformer modules in EEG signal processing. Currently, there is not sufficient research in the domain of real-time and multi-class classification tasks using transformer-based BCIs.

For instance, as of March 2024, the SCOPUS search string of TITLE-ABS-KEY ("EEG" AND "transformer" AND "brain-computer interface" AND "multi-class") OR TITLE-ABS-KEY ("electroencephalography" AND "transformer" AND "BCI" AND "Multi-class classification") returns only one single article, in which the authors report an outstanding classification accuracy of 99.7 % for binary classification, and 84 % for four-class classification suing transformers in MI. This is in stark contrast to 96 articles when compared to 93 articles when searching using the key "binary" [164].

Another aspect of this developing field, which has not been assessed as of May 2023, is the incorporation of pre-trained transformers. As described in the earlier parts of this literature review, pre-trained transformers come from the domain of NLP, where their model architecture and their performance have received an enormous amount of research attention and media coverage over the past few months. However, these incredibly useful NLP models have been pre-trained on huge datasets of billions of text samples, before eventually being deployed for generative purposes [101,102,165]. In the currently limited literature of transformer-based EEG signal processing, however, investigations of pre-trained models remain absent and therefore, boast a major knowledge gap that urges to be addressed in the near future.

Given that pre-trained weights may reduce the required (and sometimes extensive) sample acquisition and training time, this is an important aspect to investigate. Furthermore, pre-trained EEG transformers may be able to benefit from improved complex pattern-capturing capabilities. In addition, pre-trained models may be a viable potential solution for the remaining issue of cross-subject variability in EEG signal classification for BCIs. Pre-training with additional data, hence potentially improving the generalisability of the end-user's model itself is to be investigated in future research. This generalisation further may boast increased flexibility of the model, which then in turn would support more extensive fine-tuning and adaption to more specific tasks.

Another field for potential improvements using Transformer-based EEG-BCIs is the denoising of EEG signals, given their previously mentioned low signal-to-noise ratios (SNR). In a recent paper by Ref. [40], the researchers introduce the inclusion of a transformer for denoising EEG signals, as many other DL architectures have already been shown to leverage their denoising capabilities more effectively than traditional denoising methods [?]. In this particular paper, the researchers propose a 1-D EEG denoising network with a 2-D transformer (EEGDnet) [40]. In this recent paper, Electrooculogram (EOG; artifacts that generally arise from natural eye movements) and Electromyogram (EMG; artifacts that naturally arise from muscle activity such as on the scalp, mouth, or even other parts of the body) artifacts were removed (in separate experiments) using a transformer-based denoising module, achieving 18 % and 11 % improvement of correlation coefficients, respectively [40]. This may be interpreted as an important first step towards denoising EEG signals for real-time applications outside of the lab. However, the research leaves a couple of issues remain unaddressed.

Firstly, EOG and EMG artifacts were removed separately, and it is inconclusive whether the proposed transformer-based denoising approach is capable of correctly removing separately induced noise artifacts at the same time, i.e., EOG + EMG removal simultaneously.

Secondly, it is unclear how the proposed EEGDnet would perform under uncontrolled conditions, such as in real-life settings in noisy environments outside of the lab. In addition, no attempts have been made to evaluate the classification utility and performance of the proposed network when online, which is an essential knowledge gap to bridge in future research, given the EEG-BCI application's extensive records of underperforming real-time classification scenarios outside of the laboratory [82,166].

Transformers have also been subjected to preliminary research in EEG tasks other than MI. For instance, they have shown outstanding performances in tri-class emotion recognition [167], another popular mental task category for BCI applications. Whilst achieving high performances of between 91.9 % and 98.7 % on the SEED-IV and SEED dataset [168,169], respectively, the authors simultaneously state that the model's ability to generalise results is yet to be improved in the future, as per with previous DL architectures described in the previous sections, too [167]. In addition, as of March 2023, researchers have for the first time demonstrated how the introduction of transformer models can significantly improve the speech signal and EEG-based detection of clinical depression for diagnostic purposes, with outstanding performances and improved detection accuracies of 97.3 % in the Multimodal Open Dataset for Mental Disorder Analysis (MODMA) [170,171].

In addition, a recent paper proposed how the transformer's self-attention mechanisms can improve the EEG decoding for inner speech EEG tasks. Preliminary results here demonstrated the feasibility of this approach and also indicated their efficacy as transformer applications with a single channel only were utilised to decode the imagined speech [172]. However, as the very few transformer-EEG papers are not all accessible to the public (and hidden behind paywalls that were not circumventable by utilising university or third-party resources) it is difficult to provide an in-depth analysis at this point in time (which applies to this particular paper, as well as unfortunately, other papers that could not be included in this literature review). In the near future, more insights may be derived here as the body of literature in this domain only started to form recently.

In summary, it can be hypothesised that transformers may offer an entire repertoire of advantageous traits that may advance the current state of EEG-based BCI benchmarking performances for a range of mental tasks. However, the concept of utilising NLP-derived models to interfere with and develop novel EEG signal analysis applications is too new to accurately determine the developments and results in the future, with only a handful of papers currently being dedicated to exploring the domain of the aforementioned applications.

5. Discussion

Through rigorous exploration and analysis of the existing body of research, it is discerned that transformer-based models are beginning to redefine the landscape of EEG signal processing, particularly within MI classification tasks. The advent of innovative architectures, such as the 'EEG Conformer', exemplifies this shift, exhibiting superior binary classification prowess when benchmarked against traditional CNNs. However, these advancements are, at present, largely confined to theoretical constructs and have not yet been fully harnessed in real-time BCI applications. Moreover, the focus has remained relatively narrow, with a predominance of studies dedicated to binary classification, while the spectrum of multi-class MI tasks and other mental tasks await more comprehensive exploration.

Overall, a conspicuous lacuna in the research was identified — the sparse application of pre-trained transformers within the EEG-BCI domain. This contrasts sharply with the NLP domain, where pre-trained models have catalysed significant breakthroughs. This nascent stage of application in EEG signal processing signals an untapped reservoir of potential, notably in diminishing the extensive demands for data acquisition and training, enhancing signal pattern recognition, and surmounting the perennial challenge of cross-subject variability.

Advancements in DL for EEG-BCIs must address several key technological challenges to enhance efficacy and user applicability. Foremost among these is the trade-off between the portability of wireless DL-EEG-BCIs and their operational stability. Wireless systems offer significant benefits for real-world applications by enabling mobile monitoring; however, they are susceptible to fluctuations in signal quality, exacerbated by poor SNR in dynamic environments. This issue is critical in online classification tasks where real-time data processing demands high SNR to maintain accuracy and reliability. Future developments should focus on robust signal processing techniques and adaptive DL algorithms capable of compensating for environmental noise.

The debate between the need for advanced neuro-sensors versus refined AI and computing tools further complicates technological progress in BCIs. Whilst the development of ultrasensitive, non-invasive sensors could dramatically improve the quality of neural data acquisition, the complexities of neural signals demand equally sophisticated AI models for effective interpretation and utilisation. The challenge lies in enhancing sensor technology to provide high-quality data while simultaneously advancing AI to decode intricate neural patterns efficiently.

Additionally, the dichotomy between personalised BCIs and standardised models presents significant implications for scalability and performance. Personalised systems, whilst potentially more effective for individual users, require extensive customisation that may limit widespread deployment. Conversely, standardisation facilitates broader application but may not address individual neurophysiological variability. Innovative solutions might include developing modular AI frameworks that support customisation through user-centric adaptive algorithms, offering a compromise between personalisation and standardisation.

Additionally, the noise attenuation capabilities intrinsic to transformer architectures, such as those demonstrated in EEGDnet, represent a nascent yet promising pathway. Challenges, however, remain unabated, particularly concerning the concurrent elimination of multiple artifact types and the implementation of these methods in the cluttered and unpredictable environments of real-world settings. The extant body of literature is bereft of substantial evidence supporting the enhancement of real-time, online classification performance — a keystone for the operational deployment of BCIs in practical scenarios.

In the succeeding paragraphs, the ultimate key findings of the literature review have been delineated, which will be succeeded by a précis of the prospective directions this burgeoning field may take.

- Pre-trained transformers, while transformative in natural language processing (NLP), have not been extensively adapted or evaluated within EEG signal processing, indicating a significant gap in the current literature.
- The efficacy of transformer modules, such as those incorporating multi-head self-attention mechanisms, has been preliminarily validated; however, the role and optimisation of these modules warrant further investigation.
- Transformer-based models may have the potential for a superior performance in binary motor imagery (MI) classification tasks across several classification domains, which can be attributed primarily to their enhanced feature segregation capabilities.
- There is a critical need to extend the evaluation of transformer networks beyond binary classification tasks to encompass a wider range of mental tasks within EEG-based BCIs, which may reveal more comprehensive benefits.
- Investigating the application of pre-trained transformer models to EEG signal processing could offer insights into their generalization capabilities, potentially reducing the need for extensive data acquisition and model training.
- Future research should explore the denoising capabilities of transformer networks in EEG signal processing, particularly in real-world, noisy environments, and their effectiveness in online, real-time classification scenarios.
- Addressing cross-subject variability through pre-training may improve model generalisability and flexibility, opening pathways for fine-tuned, task-specific BCIs.
- Preliminary research into the application of transformers in EEG tasks beyond MI, such as emotion recognition and clinical diagnostics, has been promising, suggesting a broader applicability of these models that should be further probed.

Within the context of this literature review, several unaddressed questions and knowledge gaps emerge, warranting further investigation in future studies. These questions reflect areas of potential inquiry that could significantly contribute to the advancement of the field.

1. Can transformer-based EEG signal encoders effectively identify and eliminate multiple types of artifacts simultaneously, thus improving BCI performance?
2. Are alternative positional encoding methods, beyond the conventional sinusoidal approach, capable of enhancing the denoising, overall performance, and robustness of transformer-based models?
3. Is it possible to evolve and approximate optimised transformer-based network architectures to create end-to-end processing pipelines that ensure reliable performance tailored to specific target populations?
4. How reliable are transformer-based models in supporting multi-class classification tasks, particularly in more complex mental processes?

These questions underscore the need for further research and experimentation to fill existing knowledge gaps and expand our understanding of the potential benefits and limitations of transformer-based EEG signal processing within the domain of brain-computer interfaces.

5.1. Ethical implications and Governance

Amidst the excitement and the recent advances in AI, ethical considerations regarding the unfettered use of AI in EEG-BCIs necessitate careful examination. Recent regulatory frameworks, such as the European Union (EU) AI Act [173], aim to address these concerns and paves the way for future regulations [174]. Hence, understanding these frameworks is paramount for advancing EEG-BCI democratisation.

The EU AI Act categorizes AI systems based on potential risk. EEG-BCI applications intended for medical diagnosis or critical infrastructure control (e.g., autonomous vehicles) would likely fall under the high-risk category. This necessitates stricter regulations to ensure user safety and trustworthy operation. Transparency and explainability of the AI's decision-making processes are paramount for EEG-BCI applications. Users need to understand how the AI interprets their brain signals, fostering trust and enabling informed decision-making [175].

Furthermore, the Act prohibits discriminatory AI systems. Biases within EEG-BCI training data can lead to unfair outcomes. Mitigating bias through balanced data collection and responsible algorithm design is essential to ensure inclusivity and equitable access to this technology. The EU AI Act serves as a springboard for anticipated global regulations. Specific neurotechnology regulations are likely to emerge, echoing the principles outlined in the EU Act. Existing data privacy regulations, such as the EU's General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA), are expected to encompass EEG-BCI data. These regulations will ensure user control over their brain data and prioritize data security.

In addition to these overarching concerns, regulations may target specific high-risk applications of EEG-BCIs. Medical diagnosis, for instance, demands rigorous validation and testing to ensure diagnostic accuracy and patient safety. Similarly, brain-computer interfaces for controlling critical infrastructure will necessitate robust security measures to prevent unauthorised access or manipulation [176].

6. Conclusion and outlook

The development of EEG-based BCIs has made substantial progress over recent years, particularly via advances in DL-driven denoising and signal-processing techniques. In particular, CNNs, RNNs, and LSTMs were at the forefront of this development, and still remain in the spotlight. However, with the recent introduction of transformer-based DL models, a new shift in data-driven EEG analysis may be before us, which is new territory to biomedical research and has just begun to be examined over the past few months. Overall, we discussed and outlined how transformers may be able to boast certain advantages over their DL-predecessor methodologies in EEG analytics, yet further research is necessary to make precise declarations in this regard. For future research in this domain, their ability to capture both spatial and temporal features may be examined, as well as their ability to auto-denoise live streaming data during online classification applications for wearable BCI applications. In addition, one may focus on optimizing the architecture for the specific needs of EEG data, exploring hybrid models that combine the strengths of both transformers and CNNs and investigating methods for reducing the computational complexity of transformer models to make their applications more practicable for large-scale and real-time EEG analysis.

In this paper, a critical examination of the burgeoning domain of transformer-based EEG analysis within the realm of BCIs has been conducted. This review has cast a spotlight on the transformative impact that transformer models have begun to wield in the field of EEG signal processing. As a result, this paper traces the significant strides made in this domain, whilst concurrently delineating the extant challenges and fertile grounds for research that persist. The revolutionary impact of transformer technology is unequivocally acknowledged in the sphere of natural language processing (NLP); yet, its foray into EEG-BCI is comparatively embryonic.

The exploration of transformer-based models has started to venture beyond the confines of MI tasks, with early ventures into domains such as emotion recognition and clinical depression detection showing encouraging results. Nonetheless, the robustness and generalisability of these novel models necessitate further empirical corroboration. Moreover, the application of transformer models in the decoding of inner speech through single-channel EEG showcases the versatility and adaptability of this approach but also accentuates the limited accessibility and inclusivity of the research conducted to date.

In synthesis, our review submits that while transformers exhibit substantial promise for the evolution of EEG-based BCI systems, their full potential remains unrealised. This paper has elucidated these extant research gaps, thereby charting a course for forthcoming inquiry. It can be hypothesised that, by leveraging advancements from NLP, transformer models are well-poised to significantly advance benchmark performances across an array of mental tasks within BCIs. However, given the nascent nature of this interdisciplinary venture, definitive outcomes and trajectories are still emergent and speculative. One can expect a significant increase in output in the domain of transformer-based BCI research over the coming years, and this review paper intends to ignite further exploration and application of transformer technology within the EEG-BCI sphere, potentially ushering in a novel paradigm in BCI research and application. As the field of research in this domain continues to evolve further, additional research is expected to profoundly impact and enhance the capabilities and utility of future BCIs, ultimately driving innovation in the field of interactions between the human brain and machines.

CRediT authorship contribution statement

Maximilian Achim Pfeffer: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Steve Sai Ho Ling:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Johnny Kwok Wai Wong:** Writing – review & editing, Supervision, Investigation.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

References

[1] Brain Basics: The Life and Death of a Neuron | National Institute of Neurological Disorders and Stroke." [Online]. Available: https://www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-life-and-death-neuron.
[2] T. Kirschstein, R. Köhling, What is the source of the EEG? Clin. EEG Neurosci. 40 (3) (2009) 146–149 [Online]. Available, https://pubmed.ncbi.nlm.nih.gov/19715175/.

[3] M. Rashid, N. Sulaiman, A.P.P. Abdul Majeed, R.M. Musa, A.F. Ahmad, B.S. Bari, S. Khatun, Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review, Front. Neurorob. 14 (jun 2020).
[4] I. Lazarou, S. Nikolopoulos, P.C. Petrantonakis, I. Kompatsiaris, M. Tsolaki, EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century, Front. Hum. Neurosci. 12 (jan 2018) 14.
[5] D.J. McFarland, J.R. Wolpaw, EEG-based brain–computer interfaces, Current Opinion Biomed. Eng. 4 (dec 2017) 194–200 [Online]. Available:/pmc/articles/PMC5839510//pmc/articles/PMC5839510/?report=abstract, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5839510/.
[6] M. Abdullah, A. Madain, Y. Jararweh, Chatgpt: fundamentals, applications and social impacts, in: 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2022, pp. 1–8, plus 0.5em minus 0.4emIeee.
[7] C.K. Lo, What is the impact of chatgpt on education? a rapid review of the literature, Educ. Sci. 13 (4) (2023) 410.
[8] M.T. Baldassarre, D. Caivano, B. Fernandez Nieto, D. Gigante, A. Ragone, The social impact of generative ai: an analysis on chatgpt, in: Proceedings of the 2023 ACM Conference on Information Technology for Social Good, 2023, pp. 363–373.
[9] A. Zeng, M. Chen, L. Zhang, Q. Xu, Are transformers effective for time series forecasting? Proc. AAAI Conf. Artif. Intell. 37 (9) (2023) 121–211.
[10] L. Li, X. Gao, J. Deng, Y. Tu, Z.-J. Zha, Q. Huang, Long short-term relation transformer with global gating for video captioning, IEEE Trans. Image Process. 31 (2022) 2726–2738.
[11] J. Grigsby, Z. Wang, N. Nguyen, Y. Qi, Long-range transformers for dynamic spatiotemporal forecasting, arXiv preprint arXiv:2109.12218 (2021).
[12] K. U. Machine Intelligence Laboratory, Designing deep neural networks capturing temporal-spectral-spatial characteristics of eeg, Tech. Rep. (2024). https://milab.korea.ac.kr/research/brain-computer-interface/.
[13] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, Bci2000: a general-purpose brain-computer interface (bci) system, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1034–1043.
[14] M.A. Lebedev, M.A. Nicolelis, Brain–machine interfaces: past, present and future, Trends Neurosci. 29 (9) (2006) 536–546.
[15] R.P. Rao, *Brain-computer Interfacing: an Introduction.* Plus 0.5em Minus 0.4emCambridge University Press, 2013.
[16] J.D.R. Millán, J. Mouriño, M. Franzé, F. Cincotti, M. Varsta, J. Heikkonen, F. Babiloni, A local neural classifier for the recognition of EEG patterns associated to mental tasks, IEEE Trans. Neural Network. 13 (3) (may 2002) 678–686.
[17] V. Müller, W. Lutzenberger, H. Preißl, F. Pulvermüller, N. Birbaumer, Complexity of visual stimuli and non-linear EEG dynamics in humans, Cognit. Brain Res. 16 (1) (mar 2003) 104–110.
[18] L. Deecke, Bereitschaftspotential as an indicator of movement preparation in supplementary motor area and motor cortex, Ciba Foundat. Sympos. 132 (1987) 231–250 [Online]. Available, https://onlinelibrary.wiley.com/doi/full/10.1002/9780470513545.ch14.
[19] T. Mima, J. Steger, A.E. Schulman, C. Gerloff, M. Hallett, Electroencephalographic measurement of motor cortex control of muscle activity in humans, Clin. Neurophysiol. : Off. J. Int. Federat. Clinical Neurophysiol. 111 (2) (feb 2000) 326–337 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/10680569/.
[20] L. Zhang, H. Wang, J.Q. Gan, Eeg-based cortical localization of neural efficiency related to mathematical giftedness, in: Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part I 20, 2013, pp. 25–32, plus 0.5em minus 0.4emSpringer.
[21] J.J. Shih, D.J. Krusienski, J.R. Wolpaw, Brain-computer interfaces in medicine, Mayo Clin. Proc. 87 (3) (2012) 268 [Online]. Available:/pmc/articles/PMC3497935//pmc/articles/PMC3497935/?report=abstract, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3497935.
[22] A. Saibene, M. Caglioni, S. Corchs, F. Gasparini, EEG-based BCIs on motor imagery paradigm using wearable technologies: a systematic review, Sensors 23 (5) (mar 2023) 2798. https://www.mdpi.com/1424-8220/23/5/2798/htm. mdpi.com/1424-8220/23/5/2798.
[23] N. Birbaumer, Breaking the silence: brain–computer interfaces (BCI) for communication and motor control, Psychophysiology 43 (6) (nov 2006) 517–532 [Online]. Available, https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8986.2006.00456.x.
[24] E. Musk, An integrated brain-machine interface platform with thousands of channels, J. Med. Internet Res. 21 (10) (oct 2019) e16194.
[25] M. Shaima, N. Nabi, M.N.U. Rana, M.T. Islam, E. Ahmed, M.I. Tusher, M. H. Mukti, Q. Saad-Ul-Mosaher, Elon musk's neuralink brain chip: a review on 'brain-reading'device, J. Comp. Sci. Technol. Stud. 6 (1) (2024) 200–203.
[26] L. Drew, Neuralink brain chip: advance sparks safety and secrecy concerns, Nature 627 (8002) (2024) 19, 19.
[27] F. Babiloni, F. Cincotti, M. Mattiocco, A. Timperi, S. Salinari, M.G. Marciani, M. Donatella, Brain computer interface: estimation of cortical activity from non invasive high resolution EEG recordings, Ann. Int. Conf. IEEE Eng. Med. Biol. Proc. 26 VI (2004) 4375–4376.
[28] P.L. Lee, C.H. Wu, J.C. Hsieh, Y.T. Wu, Visual evoked potential actuated brain computer interface: a brain-actuated cursor system, Electron. Lett. 41 (15) (jul 2005) 832–834.
[29] S.A. Yohanandan, I. Kiral-Kornek, J. Tang, B.S. Mshford, U. Asif, S. Harrer, A robust low-cost EEG motor imagery-based brain-computer interface, Annual International Conference, vol. 2018, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE Engineering in Medicine and Biology Society, oct 2018, pp. 5089–5092 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30441485/.
[30] S. Park, J. Ha, D.H. Kim, L. Kim, Improving motor imagery-based brain-computer interface performance based on sensory stimulation training: an approach focused on poorly performing users, Front. Neurosci. 15 (nov 2021) 1526.
[31] A. Singh, A.A. Hussain, S. Lal, H.W. Guesgen, A. Singh, A.A. Hussain, S. Lal, H.W. A. Guesgen, A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface, Sensors 21 (2021) 2173, vol. 21, no. 6, p. 2173, mar 2021. [Online]. Available, https://www.mdpi.com/1424-8220/21/6/2173/htm. mdpi.com/1424-8220/21/6/2173.
[32] K.P. Pa Aung, K.H. Nwe, Regions of Interest (Roi) Analysis for Upper Limbs Eeg Neuroimaging Schemes, 2020, pp. 53–58.
[33] M. Lotze, U. Halsband, Motor imagery, J. Physiol. Paris 99 (4–6) (jun 2006) 386–395.
[34] K.K. Ang, C. Guan, EEG-based strategies to detect motor imagery for control and rehabilitation, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (4) (apr 2017) 392–401.
[35] C.S. Nam, Y. Jeon, Y.J. Kim, I. Lee, K. Park, Movement imagery-related lateralization of event-related (de)synchronization (ERD/ERS): Motor-imagery duration effects, Clin. Neurophysiol. 122 (3) (mar 2011) 567–577.
[36] Y. Jeon, C.S. Nam, Y.J. Kim, M.C. Whang, Event-related (De)synchronization (ERD/ERS) during motor imagery tasks: implications for brain–computer interfaces, Int. J. Ind. Ergon. 41 (5) (sep 2011) 428–436.
[37] E. Thomas, J. Fruitet, M. Clerc, Investigating brief motor imagery for an ERD/ERS based BCI, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2012, pp. 2929–2932.
[38] J. Kevric, A. Subasi, Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system, Biomed. Signal Process Control 31 (jan 2017) 398–406.
[39] Y. Shin, S. Lee, M. Ahn, H. Cho, S.C. Jun, H.N. Lee, Noise robustness analysis of sparse representation based classification method for non-stationary EEG signal classification, Biomed. Signal Process Control 21 (aug 2015) 8–18.
[40] X. Pu, P. Yi, K. Chen, Z. Ma, D. Zhao, Y. Ren, EEGDnet: fusing non-local and local self-similarity for EEG signal denoising with transformer, Comput. Biol. Med. 151 (dec 2022) 106248.
[41] C. Kaur, P. Singh, S. Sahni, EEG artifact removal system for depression using a hybrid denoising approach, Basic Clin. Neurosci. 12 (4) (jun 2021) 465 [Online]. Available:/pmc/articles/PMC8817173//pmc/articles/PMC8817173/?report=abstract, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8817173/.
[42] M. Grobbelaar, S. Phadikar, E. Ghaderpour, A.F. Struck, N. Sinha, R. Ghosh, M. Zaved, I. Ahmed, A survey on denoising techniques of electroencephalogram signals using wavelet transform, vol. 3, no. 3, pp. 577–586, Signals 2022 3 (aug 2022) 577–586 [Online]. Available, https://www.mdpi.com/2624-6120/3/3/35/htm. mdpi.com/2624-6120/3/3/35.
[43] S. Vaid, P. Singh, C. Kaur, EEG signal analysis for BCI interface: a review, in: International Conference on Advanced Computing and Communication Technologies, ACCT 2015-April, apr 2015, pp. 143–147.
[44] Z. Khademi, F. Ebrahimi, H.M. Kordy, A review of critical challenges in MI-BCI: from conventional to deep learning methods, J. Neurosci. Methods 383 (jan 2023) 109736.
[45] P. He, M. Kahle, G. Wilson, C. Russell, Removal of ocular artifacts from EEG: a comparison of adaptive filtering method and regression method using simulated data, Ann. Int. Conf. IEEE Eng. Med. Biol. Proc. 7 (2005) 1110–1113. VOLS.
[46] Y. Hou, L. Zhou, S. Jia, X. Lun, A novel approach of decoding eeg four-class motor imagery tasks via scout esi and cnn, J. Neural. Eng. 17 (1) (2020) 016048.
[47] S. Saminu, G. Xu, Z. Shuai, I.A. El Kader, A.H. Jabire, Y.K. Ahmed, I.A. Karaye, I. S. Ahmad, A recent investigation on detection and classification of epileptic seizure techniques using EEG signal, 2021, Vol. 11, Page 668, Brain Sci. 11 (5) (may 2021) 668 [Online]. Available, https://www.mdpi.com/2076-3425/11/5/668/htm. mdpi.com/2076-3425/11/5/668.
[48] S. Blanco, C.E. D'Attellis, S.I. Isaacson, O.A. Rosso, R.O. Sirne, Time-frequency analysis of electroencephalogram series. II. Gabor and wavelet transforms, Phys. Rev. 54 (6) (dec 1996) 6661 [Online]. Available: https://journals.aps.org/pre/abstract/10.1103/PhysRevE.54.6661.
[49] W. Zhou, Y. Li, EEG multiresolution analysis using wavelet transform, Ann. Int. Conf. IEEE Eng. Med. Biol. Proc. 2 (2001) 1854–1856.
[50] A.S. Al-Fahoum, A.A. Al-Fraihat, A. Grant, J.A. Hinojosa, M.S. Oliveira, Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains [Online]. Available: https://doi.org/10.1155/2014/730218, 2014.
[51] Y. Wang, C. Song, T. Zhang, Z. Yao, Z. Chang, D. Wang, Feature extraction of motor imagery EEG via discrete wavelet transform and generalized maximum fuzzy membership difference entropy: a comparative study, 2023, Vol. 12, Page 2207, Electronics 12 (10) (may 2023) 2207 [Online]. Available, https://www.mdpi.com/2079-9292/12/10/2207/htm. mdpi.com/2079-9292/12/10/2207.
[52] S. Blanco, R.Q. Quiroga, O.A. Rosso, S. Kochen, Time-frequency analysis of electroencephalogram series, Phys. Rev. 51 (3) (mar 1995) 2624 [Online]. Available: https://journals.aps.org/pre/abstract/10.1103/PhysRevE.51.2624.
[53] M. Kaczorowska, M. Plechawska-Wojcik, M. Tokovarov, R. Dmytruk, Comparison of the ICA and PCA methods in correction of EEG signal artefacts, in: 2017 10th International Symposium on Advanced Topics in Electrical Engineering, ATEE 2017, apr 2017, pp. 262–267 [Online]. Available: https://www.researchgate.net/publication/316446617_Comparison_of_the_ICA_and_PCA_methods_in_correction_of_EEG_signal_artefacts.
[54] A. Subasi, M.I. Gursoy, EEG signal classification using PCA, ICA, LDA and support vector machines, Expert Syst. Appl. 37 (12) (dec 2010) 8659–8666.

[55] I. Rejer, P. Gorski, Benefits of ICA in the case of a few channel EEG, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2015, IEEE Engineering in Medicine and Biology Society. Annual International Conference, nov 2015, pp. 7434–7437 [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26738010/.
[56] W. Yang, A. Pilozzi, X. Huang, An overview of ICA/BSS-based application to alzheimer's brain signal processing, 2021, Vol. 9, Page 386, Biomedicines 9 (4) (apr 2021) 386 [Online]. Available, https://www.mdpi.com/2227-9059/9/4/386/htm. mdpi.com/2227-9059/9/4/386.
[57] J. Metsomaa, J. Sarvas, R.J. Ilmoniemi, Blind source separation of event-related EEG/MEG, IEEE Trans. Biomed. Eng. 64 (9) (sep 2017) 2054–2064 [Online]. Available, https://pubmed.ncbi.nlm.nih.gov/28113203/.
[58] A. Cichocki, S.L. Shishkin, T. Musha, Z. Leonowicz, T. Asada, T. Kurachi, EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease, Clin. Neurophysiol. : Off. J. Int. Federat. Clinical Neurophysiol. 116 (3) (2005) 729–737 [Online]. Available, https://pubmed.ncbi.nlm.nih.gov/15721088/.
[59] S. Aggarwal, N. Chugh, Signal processing techniques for motor imagery brain computer interface: a review, Array 1–2 (jan 2019) 100003.
[60] W. Mumtaz, S. Rasheed, A. Irfan, Review of challenges associated with the EEG artifact removal methods, Biomed. Signal Process Control 68 (jul 2021) 102741.
[61] M.K. Islam, A. Rastegarnia, Z. Yang, Methods for artifact detection and removal from scalp EEG: a review, Neurophysiol. Clinique/Clinical Neurophysiol. 46 (4–5) (nov 2016) 287–305.
[62] F. Lotte, M. Congedo, EEG feature extraction, Brain-Computer Interf.: Foundat. Methods (aug 2016) 127–143. https://onlinelibrary.wiley.com/doi/full/10.1002/9781119144977.ch7.
[63] P. Boonyakitanont, A. Lek-uthai, K. Chomtho, J. Songsiri, A review of feature extraction and performance evaluation in epileptic seizure detection using EEG, Biomed. Signal Process Control 57 (mar 2020) 101702.
[64] A. Zhang, B. Yang, L. Huang, Feature extraction of EEG signals using power spectral entropy, BioMed. Eng. Informat.: New Develop. Future - Proc. Int. Conf. BioMed. Eng. Informat. BMEI 2 (2008) 435–439.
[65] E. Derya Übeyli, Statistics over features: EEG signals analysis, Comput. Biol. Med. 39 (8) (aug 2009) 733–741.
[66] Z. Liu, J.A. de Zwart, P. van Gelderen, L.W. Kuo, J.H. Duyn, Statistical feature extraction for artifact removal from concurrent fMRI-EEG recordings, Neuroimage 59 (3) (feb 2012) 2073–2087.
[67] A. Al-Saegh, S.A. Dawwd, J.M. Abdul-Jabbar, Deep learning for motor imagery EEG-based classification: a review, Biomed. Signal Process Control 63 (jan 2021) 102172.
[68] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, J. Neural. Eng. 16 (3) (apr 2019) 031001 [Online]. Available, https://iopscience.iop.org/article/10.1088/1741-2552/ab0ab5. iopscience.iop.org/article/10.1088/1741-2552/ab0ab5/meta.
[69] E. Polytechnique, F. De Lausanne, J.B. Ochoa, EEG Signal Classification for Brain Computer Interface Applications, 2002.
[70] N. Wagh, J. Wei, S. Rawal, B.M. Berry, Y. Varatharajah, Evaluating latent space robustness and uncertainty of EEG-ML models under realistic distribution shifts, Adv. Neural Inf. Process. Syst. 35 (dec 2022) 142–221.
[71] A. Apicella, P. Arpaia, G. D'Errico, D. Marocco, G. Mastrati, N. Moccaldi, R. Prevete, Machine learning strategies to improve generalization in EEG-based emotion assessment: a systematic review [Online]. Available: https://arxiv.org/abs/2212.08744v1, dec 2022.
[72] S. Saha, M. Baumert, Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: a review, Front. Comput. Neurosci. 13 (jan 2020) 87.
[73] L. Bozhkov, P. Georgieva, Overview of deep learning architectures for EEG-based brain imaging, in: Proceedings of the International Joint Conference on Neural Networks, 2018-July, oct 2018.
[74] E. Lashgari, J. Ott, A. Connelly, P. Baldi, U. Maoz, An end-to-end CNN with attentional mechanism applied to raw EEG in a BCI classification task, J. Neural. Eng. 18 (4) (aug 2021) 460e3 [Online]. Available, https://iopscience.iop.org/article/10.1088/1741-2552/ac1ade. iopscience.iop.org/article/10.1088/1741-2552/ac1ade/meta.
[75] G. Avina-Cervantes, W. Alsumari, M. Hussain, L. Alshehri, H.A. Aboalsamh, EEG-based person identification and authentication using deep convolutional neural network [Online]. Available: https://doi.org/10.3390/axioms12010074, 2023.
[76] H. Altaheri, G. Muhammad, M. Alsulaiman, S.U. Amin, G.A. Altuwaijri, W. Abdul, M.A. Bencherif, M. Faisal, Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: a review [Online]. Available, Neural Comput. Appl. 2021 (aug 2021) 1–42, https://link.springer.com/article/10.1007/s00521-021-06352-5.
[77] S. Aggarwal, N. Chugh, Review of machine learning techniques for EEG based brain computer interface, Arch. Comput. Methods Eng. 29 (5) (aug 2022) 3001–3020 [Online]. Available, https://link.springer.com/article/10.1007/s11831-021-09684-6.
[78] X. Gu, Z. Cao, A. Jolfaei, P. Xu, D. Wu, T.P. Jung, C.T. Lin, EEG-based brain-computer interfaces (BCIs): a survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications, IEEE ACM Trans. Comput. Biol. Bioinf 18 (5) (2021) 1645–1666.
[79] J. Li, S. Qiu, Y.Y. Shen, C.L. Liu, H. He, Multisource transfer learning for cross-subject EEG emotion recognition, IEEE Trans. Cybern. 50 (7) (jul 2020) 3281–3293.
[80] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, H. Ghayvat, CNN variants for computer vision: history, architecture, application, challenges and future scope, 2021, Vol. 10, Page 2470, Electronics 10 (20) (oct 2021) 2470 [Online]. Available, https://www.mdpi.com/2079-9292/10/20/2470/htm. mdpi.com/2079-9292/10/20/2470.
[81] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, Deep learning for computer vision: a brief review, Comput. Intell. Neurosci. 2018 (2018).
[82] K. Douibi, S. Le Bars, A. Lemontey, L. Nag, R. Balp, G. Breda, Toward EEG-based BCI applications for industry 4.0: challenges and possible applications, Front. Hum. Neurosci. 15 (aug 2021) 456.
[83] M. JoginMohana, M.S. Madhulika, G.D. Divya, R.K. Meghana, S. Apoorva, Feature extraction using convolution neural networks (CNN) and deep learning, in: 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2018 - Proceedings, may 2018, pp. 2319–2323.
[84] P. Lahane, J. Jagtap, A. Inamdar, N. Karne, R. Dev, A review of recent trends in EEG based Brain-Computer Interface, in: ICCIDS 2019 - 2nd International Conference on Computational Intelligence in Data Science, Proceedings, feb 2019.
[85] J. Minguillon, M.A. Lopez-Gordo, F. Pelayo, Trends in EEG-BCI for daily-life: requirements for artifact removal, Biomed. Signal Process Control 31 (jan 2017) 407–418.
[86] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (sep 2018) 270–278.
[87] K. O'Shea, R. Nash, An introduction to convolutional neural networks, Int. J. Res. Appl. Sci. Eng. Technol. 10 (12) (nov 2015) 943–947 [Online]. Available, https://arxiv.org/abs/1511.08458v2.
[88] A. Ajit, K. Acharya, A. Samanta, A review of convolutional neural networks, in: International Conference on Emerging Trends in Information Technology and Engineering, Ic-ETITE 2020, feb 2020.
[89] S. Bagherzadeh, K. Maghooli, A. Shalbaf, A. Maghsoudi, Emotion recognition using continuous wavelet transform and ensemble of convolutional neural networks through transfer learning from electroencephalogram signal, Front. Biomed. Technol. 10 (1) (2023) 47–56.
[90] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: a database for emotion analysis; using physiological signals, IEEE Transact. Affect. Comput. 3 (1) (2011) 18–31.
[91] S. Bagherzadeh, K. Maghooli, A. Shalbaf, A. Maghsoudi, Emotion recognition using effective connectivity and pre-trained convolutional neural networks in eeg signals, Cognitive Neurodynamics 16 (5) (2022) 1087–1106.
[92] B. Chakravarthi, S.-C. Ng, M. Ezilarasan, M.-F. Leung, Eeg-based emotion recognition using hybrid cnn and lstm classification, Front. Comput. Neurosci. 16 (2022) 1019776.
[93] Y. Liu, J. Chen, Y. Chang, S. He, Z. Zhou, A novel integration framework for degradation-state prediction via transformer model with autonomous optimizing mechanism, J. Manuf. Syst. 64 (jul 2022) 288–302.
[94] A. Nagarajan, N. Robinson, C. Guan, Investigation on robustness of EEG-based brain-computer interfaces, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2021-January, EMBS, 2021, pp. 6334–6340.
[95] P. Wang, A. Jiang, X. Liu, J. Shang, L. Zhang, LSTM-based EEG classification in motor imagery tasks, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (11) (nov 2018) 2086–2095.
[96] S. Tortora, S. Ghidoni, C. Chisari, S. Micera, F. Artoni, Deep learning-based BCI for gait decoding from EEG with LSTM recurrent neural network, J. Neural. Eng. 17 (4) (jul 2020) 046011 [Online]. Available, https://iopscience.iop.org/article/10.1088/1741-2552/ab9842. iopscience.iop.org/article/10.1088/1741-2552/ab9842/meta.
[97] N. specified, What is ai chatbot phenomenon chatgpt and could it replace humans? [Online; accessed 11-May-2023]. [Online]. Available, https://web.archive.org/web/20230116100346, 2022, Dec. theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans.
[98] A.N. Guz, J.J. Rushchitsky, Scopus: a system for the evaluation of scientific journals, Int. Appl. Mech. 45 (2009) 351–362.
[99] K. Canese, S. Weis, Pubmed: the bibliographic database, NCBI Handbook 2 (1) (2013).
[100] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[101] B.D. Lund, T. Wang, Chatting about ChatGPT: how may AI and GPT impact academia and libraries? Libr. Hi Tech News 40 (3) (2023) 26–29.
[102] I.L. Alberts, L. Mercolli, T. Pyka, G. Prenosil, K. Shi, A. Rominger, A. Afshar-Oromieh, Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? Eur. J. Nucl. Med. Mol. Imag. 50 (6) (may 2023) 1549–1552 [Online]. Available: https://link.springer.com/article/10.1007/s00259-023-06172-w.
[103] A. Reshamwala, D. Mishra, P. Pawar, REVIEW on natural LANGUAGE PROCESSING image retrieval view project detection and classification of fabric defects for textile industry view project alpa reshamwala narsee monjee institute of management studies review on natural LANGUAGE PROCESSING, Int. J. (ESTIJ 3 (1) (2013) 2250–3498 [Online]. Available: https://www.researchgate.net/publication/235788362.
[104] E. Cambria, B. White, Jumping NLP curves: a review of natural language processing research, IEEE Comput. Intell. Mag. 9 (2) (2014) 48–57.
[105] P.M. Nadkarni, L. Ohno-Machado, W.W. Chapman, Natural language processing: an introduction, J. Am. Med. Inf. Assoc. 18 (5) (sep 2011) 544–551 [Online]. Available: https://academic.oup.com/jamia/article/18/5/544/829676.

[106] P. Shaw, J. Uszkoreit, A. Vaswani, Self-attention with relative position representations, NAACL HLT 2018 - 2018 Conf. North Am. Chapt. Assoc. Computat. Linguist.: Human Lang. Technol. Proc. Conf. 2 (mar 2018) 464–468 [Online]. Available: https://arxiv.org/abs/1803.02155v2.
[107] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, Y. Zhan, A transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2126–2136.
[108] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T.L. Scao, S. Gugger, M. Drame, Q. Lhoest, A.M. Rush, Transformers: State-Of-The-Art Natural Language Processing, nov 2020, pp. 38–45 [Online]. Available, https://aclanthology.org/2020.emnlp-demos.6.
[109] A. Parvaiz, M.A. Khalid, R. Zafar, H. Ameer, M. Ali, M.M. Fraz, Vision Transformers in medical computer vision—a contemplative retrospection, Eng. Appl. Artif. Intell. 122 (jun 2023) 106126 [Online]. Available, http://arxiv.org/abs/2203.15269.
[110] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, D. Tao, A survey on vision transformer, IEEE Trans. Pattern Anal. Mach. Intell. 45 (1) (jan 2023) 87–110.
[111] K. Lu, A. Grover, P. Abbeel, I. Mordatch, Pretrained transformers as universal computation engines, arXiv preprint arXiv:2103.05247 1 (2021).
[112] Y. Hao, L. Dong, F. Wei, K. Xu, Self-attention attribution: interpreting information interactions inside transformer, Proc. AAAI Conf. Artif. Intell. 35 (14) (may 2021) 963–1012 [Online]. Available, https://ojs.aaai.org/index.php/AAAI/article/view/17533.
[113] X. Liu, H.-F. Yu, I.S. Dhillon, C.-J. Hsieh, Learning to Encode Position for Transformer with Continuous Dynamical Model, nov 2020, pp. 6327–6335 [Online]. Available: https://proceedings.mlr.press/v119/liu20n.html.
[114] M. Dehghani, S. Gouws, O. Vinyals, J. Uszkoreit, Ł. Kaiser, Universal transformers, in: 7th International Conference on Learning Representations, ICLR 2019, jul 2018 [Online]. Available: https://arxiv.org/abs/1807.03819v3.
[115] R. Xiong, Y. Yang, D. He, K. Zheng, S. Zheng, C. Xing, H. Zhang, Y. Lan, L. Wang, T.-Y. Liu, On Layer Normalization in the Transformer Architecture, nov 2020, pp. 524–610 [Online]. Available: https://proceedings.mlr.press/v119/xiong20b.html.
[116] A. Raganato, J. Tiedemann, An Analysis of Encoder Representations in Transformer-Based Machine Translation, 2018, pp. 287–297 [Online]. Available, https://helda.helsinki.fi/handle/10138/263704.
[117] K. Wu, H. Peng, M. Chen, J. Fu, H. Chao, Rethinking and improving relative position encoding for vision transformer, 10 033–10 041, https://github.com/microsoft/Cream/tree/main/iRPE, 2021.
[118] J. Vig, A multiscale visualization of attention in the transformer model. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of System Demonstrations, jun 2019, pp. 37–42 [Online]. Available: https://arxiv.org/abs/1906.05714v1.
[119] A. Katharopoulos, A. Vyas, N. Pappas, F. Fleuret, Transformers Are RNNs: Fast Autoregressive Transformers with Linear Attention, nov 2020, pp. 5156–5165 [Online]. Available: https://proceedings.mlr.press/v119/katharopoulos20a.html.
[120] H. Deguchi, A. Tamura, T. Ninomiya, Dependency-based self-attention for transformer NMT, in: International Conference Recent Advances in Natural Language Processing, RANLP, 2019-September, 2019, pp. 239–246 [Online]. Available: https://aclanthology.org/R19-1028.
[121] C. Chelba, M. Chen, A. Bapna, N. Shazeer, Faster transformer decoding: N-gram masked self-attention [Online]. Available, https://arxiv.org/abs/2001.04589v1, jan 2020.
[122] Y. Li, Y. Lin, T. Xiao, J. Zhu, An efficient transformer decoder with compressed sub-layers, Proc. AAAI Conf. Artif. Intell. 35 (15) (may 2021) 315–413 [Online]. Available, https://ojs.aaai.org/index.php/AAAI/article/view/17572.
[123] Z. Dan, Y. Zhao, X. Bi, L. Wu, Q. Ji, Multi-task transformer with adaptive cross-entropy loss for multi-dialect speech recognition, 2022, Vol. 24, Page 1429, Entropy 24 (10) (oct 2022) 1429 [Online]. Available, https://www.mdpi.com/1099-4300/24/10/1429/htm. mdpi.com/1099-4300/24/10/1429.
[124] R. Castro, I. Pineda, W. Lim, M.E. Morocho-Cayamcela, Deep learning approaches based on transformer architectures for image captioning tasks, IEEE Access 10 (2022) 679–733.
[125] Q. Pu, Z. Xi, S. Yin, Z. Zhao, L. Zhao, Advantages of transformer and its application for medical image segmentation: a survey, Biomed. Eng. Online 23 (1) (2024) 14.
[126] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q.V. Le, R. Salakhutdinov, Transformer-xl: attentive language models beyond a fixed-length context, arXiv preprint arXiv:1901.02860 (2019).
[127] S. Karita, N. Chen, T. Hayashi, T. Hori, H. Inaguma, Z. Jiang, M. Someki, N.E. Y. Soplin, R. Yamamoto, X. Wang, et al., A comparative study on transformer vs rnn in speech applications, in: 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2019, pp. 449–456, plus 0.5em minus 0.4emIEEE.
[128] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2020 *arXiv preprint arXiv:2010.11929*.
[129] R.M. Rao, J. Liu, R. Verkuil, J. Meier, J. Canny, P. Abbeel, T. Sercu, A. Rives, Msa transformer, in: International Conference on Machine Learning, 2021, pp. 8844–8856, plus 0.5em minus 0.4emPMLR.
[130] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: beyond efficient transformer for long sequence time-series forecasting, Proc. AAAI Conf. Artif. Intell. 35 (12) (2021) 11 106–11 115.
[131] L. Cai, K. Janowicz, G. Mai, B. Yan, R. Zhu, Traffic transformer: capturing the continuity and periodicity of time series for traffic forecasting, Trans. GIS 24 (3) (2020) 736–755.
[132] F. Wu, Q. Zhang, D. Radev, J. Cui, W. Zhang, H. Xing, N. Zhang, H. Chen, 3d-transformer: Molecular Representation with Transformer in 3d Space, 2021.
[133] Y. Li, D. Li, S. Prakash, P. Wang, Toward interpretability of dual-encoder models for dialogue response suggestions, arXiv preprint arXiv:2003.04998 (2020).
[134] Z. Zhang, H. Zhang, L. Zhao, T. Chen, S.Ö. Arik, T. Pfister, Nested hierarchical transformer: towards accurate, data-efficient and interpretable visual understanding, Proc. AAAI Conf. Artif. Intell. 36 (3) (2022) 3417–3425.
[135] Z. Li, E. Wallace, S. Shen, K. Lin, K. Keutzer, D. Klein, J. Gonzalez, Train big, then compress: rethinking model size for efficient training and inference of transformers, in: International Conference on Machine Learning, 2020, pp. 5958–5968, plus 0.5em minus 0.4emPMLR.
[136] C. Sanford, D.J. Hsu, M. Telgarsky, Representational strengths and limitations of transformers, Adv. Neural Inf. Process. Syst. 36 (2024).
[137] A.H. Ribeiro, K. Tiels, L.A. Aguirre, T. Schön, Beyond Exploding and Vanishing Gradients: Analysing RNN Training Using Attractors and Smoothness, jun 2020, pp. 2370–2380 [Online]. Available, https://proceedings.mlr.press/v108/ribeiro20a.html.
[138] M.K. Chowdary, J. Anitha, D.J. Hemanth, Emotion recognition from EEG signals using recurrent neural networks, 2022, Vol. 11, Page 2387, Electronics 11 (15) (jul 2022) 2387 [Online]. Available, https://www.mdpi.com/2079-9292/11/15/2387/htm. mdpi.com/2079-9292/11/15/2387.
[139] J. Sun, J. Xie, H. Zhou, EEG classification with transformer-based models. LifeTech 2021 - 2021 IEEE 3rd Global Conference on Life Sciences and Technologies, mar 2021, pp. 92–93.
[140] Y. Song, X. Jia, L. Yang, L. Xie, Transformer-based spatial-temporal feature learning for EEG decoding [Online]. Available: https://github.com/anranknight/EEG-Transformer, 2021.
[141] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, B. Ge, Summary of ChatGPT/GPT-4 research and perspective towards the future of large language models [Online]. Available: https://arxiv.org/abs/2304.01852v3, apr 2023.
[142] D. Wu, Y. Xu, B.L. Lu, Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016, IEEE Transactions on Cognitive and DevelopmentalIEEE Transact. Cognit. Develop. Syst.l Systems 14 (1) (mar 2022) 4–19.
[143] Z. Wan, R. Yang, M. Huang, N. Zeng, X. Liu, A review on transfer learning in EEG signal analysis, Neurocomputing 421 (jan 2021) 1–14.
[144] H. Goldberger A, L. Amaral, L. Glass, J. Hausdorff, P.C. Ivanov, R. Mark, J. E. Mietus, G.B. Moody, C.K. Peng, Stanley, EEG motor movement/imagery dataset v1.0.0 [Online]. Available: https://physionet.org/content/eegmmidb/1.0.0/, 2009.
[145] A. Ali, T. Schnake, O. Eberle, G. Montavon, K.-R. Müller, L. Wolf, XAI for Transformers: Better Explanations through Conservative Propagation, jun 2022, pp. 435–451 [Online]. Available: https://proceedings.mlr.press/v162/ali22a.html.
[146] A.M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B.A. Becker, C. Mooney, Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review, 2021, Vol. 11, Page 5088, Appl. Sci. 11 (11) (may 2021) 5088 [Online]. Available, https://www.mdpi.com/2076-3417/11/11/5088/htm.
[147] C.Q. Lai, H. Ibrahim, M.Z. Abdullah, J.M. Abdullah, S.A. Suandi, A. Azman, Artifacts and noise removal for electroencephalogram (EEG): a literature review, in: ISCAIE 2018 - 2018 IEEE Symposium on Computer Applications and Industrial Electronics, jul 2018, pp. 326–332.
[148] G. Siddhad, A. Gupta, D.P. Dogra, P.P. Roy, Efficacy of transformer networks for classification of raw EEG data. https://arxiv.org/abs/2202.05170v1, feb 2022.
[149] P.L. Lee, S.H. Chen, T.C. Chang, W.K. Lee, H.T. Hsu, H.H. Chang, Continual learning of a transformer-based deep learning classifier using an initial model from action observation EEG data to online motor imagery classification, Bioengineering 10 (2) (feb 2023) 186, 2023, Vol. 10, Page 186, https://www.mdpi.com/2306-5354/10/2/186/htm.
[150] J. Kim, J. Yeon, J. Ryu, J.Y. Park, S.C. Chung, S.P. Kim, Neural activity patterns in the human brain reflect tactile stickiness perception, Front. Hum. Neurosci. 11 (sep 2017) 445.
[151] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (11) (nov 2017) 5391–5420 [Online]. Available, https://onlinelibrary.wiley.com/doi/full/10.1002/hbm.23730.
[152] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces, J. Neural. Eng. 15 (5) (jul 2018) 056013 [Online]. Available, https://iopscience.iop.org/article/10.1088/1741-2552/aace8c. iopscience.iop.org/article/10.1088/1741-2552/aace8c/meta.
[153] M. Hassan, M. Shamas, M. Khalil, W. El Falou, F. Wendling, Eegnet: an open source tool for analyzing and visualizing m/eeg connectome, PLoS One 10 (9) (2015) e0138297.
[154] I. Winkens, C.M. Van Heugten, D.T. Wade, E.J. Habets, L. Fasotti, Efficacy of time pressure management in stroke patients with slowed information processing: a

randomized controlled trial, Arch. Phys. Med. Rehabil. 90 (10) (oct 2009) 1672–1679.
[155] C. Janiesch, P. Zschech, K. Heinrich, Machine learning and deep learning, Electron. Mark. 31 (3) (sep 2021) 685–695 [Online]. Available: https://link.springer.com/article/10.1007/s12525-021-00475-2.
[156] Y. Song, Q. Zheng, B. Liu, X. Gao, EEG conformer: convolutional transformer for EEG decoding and visualization, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2023) 710–719.
[157] J. Neirynck, L. Milić, Equal ripple tolerance characteristics, Int. J. Circ. Theor. Appl. 4 (1) (jan 1976) 99–104 [Online]. Available, https://onlinelibrary.wiley.com/doi/full/10.1002/cta.4490040110.
[158] J.D. Rhodes, S.A. Alseyab, The generalized Chebyshev low-pass prototype filter, Int. J. Circ. Theor. Appl. 8 (2) (apr 1980) 113–125 [Online]. Available, https://onlinelibrary.wiley.com/doi/full/10.1002/cta.4490080205.
[159] Y. Li, M.L. Luo, K. Li, A multiwavelet-based time-varying model identification approach for time–frequency analysis of EEG signals, Neurocomputing 193 (jun 2016) 106–114.
[160] I. Ahmad, A. Shufian, M.D. Rahman Barno, S. Datta, A novel approach to remove ocular artifact from EEG signal, in: 2019 IEEE 5th International Conference for Convergence in Technology, I2CT 2019, mar 2019.
[161] P. Podder, M.M. Hasan, M.R. Islam, M. Sayeed, Design and implementation of butterworth, Chebyshev-I and elliptic filter for speech signal analysis, Int. J. Comput. Appl. 98 (7) (feb 2020) 12–18 [Online]. Available, http://arxiv.org/abs/2002.03130, 10.5120/17195-7390.
[162] M. Sandhu, S. Kaur, A study on design and implementation of butterworth, Chebyshev and elliptic filter with MatLab, Int. J. Emerg. Trends Eng. Res. 4 (6) (2016) [Online]. Available: www.ijeter.everscience.org.
[163] A. Carini, G.L. Sicuranza, A study about Chebyshev nonlinear filters, Signal Process. 122 (may 2016) 24–32.
[164] N. Sharma, A. Upadhyay, M. Sharma, A. Singhal, Deep temporal networks for eeg-based motor imagery recognition, Sci. Rep. 13 (1) (2023) 18813.
[165] G. Yenduri, C. G. Selvi, G. Srivastava, P. Kumar Reddy Maddikunta, D. G. Raj, R. H. Jhaveri, W. Wang, A. V. Vasilakos, T. Reddy Gadekallu, and D. G. Raj, "GPT (Generative Pre-trained Transformer)-A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions.".
[166] N. Tibrewal, N. Leeuwis, M. Alimardani, Classification of motor imagery EEG using deep learning increases performance in inefficient BCI users, PLoS One 17 (7 July) (jul 2022) e0268880 [Online]. Available, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0268880.
[167] L. Gong, M. Li, T. Zhang, W. Chen, EEG emotion recognition using attention-based convolutional transformer neural network, Biomed. Signal Process Control 84 (jul 2023) 104835.
[168] W.L. Zheng, W. Liu, Y. Lu, B.L. Lu, A. Cichocki, EmotionMeter: a multimodal framework for recognizing human emotions, IEEE Trans. Cybern. 49 (3) (mar 2019) 1110–1122.
[169] Shanghai Jiao Tong University, "SEED dataset." [Online]. Available: https://bcmi.sjtu.edu.cn/~seed/seed-iv.html.
[170] A. Qayyum, I. Razzak, M. Tanveer, M. Mazher, B. Alhaqbani, High-density electroencephalography and speech signal based deep framework for clinical depression diagnosis, IEEE ACM Trans. Comput. Biol. Bioinf (mar 2023) 1–11.
[171] H. Cai, Z. Yuan, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, Q. Zhao, Z. Liu, Z. Yao, M. Yang, H. Peng, J. Zhu, X. Zhang, G. Gao, F. Zheng, R. Li, Z. Guo, R. Ma, J. Yang, L. Zhang, X. Hu, Y. Li, B. Hu, A multi-modal open dataset for mental-disorder analysis, 2022 9:1, Sci. Data 9 (1) (apr 2022) 1–10 [Online]. Available: https://www.nature.com/articles/s41597-022-01211-x.
[172] Y.E. Lee, S.H. Lee, EEG-transformer: self-attention from transformer architecture for decoding EEG of imagined speech, in: International Winter Conference on Brain-Computer Interface, BCI, 2022-February, 2022.
[173] Future of Life Institute, The eu artificial intelligence act, 2024-04-15. [Online]. Available: https://artificialintelligenceact.eu/, 2024.
[174] A. Karanasiou, On being transhuman: commercial bcis and the quest for autonomy. The Cambridge Handbook of the Law of Algorithms, CUP 2020, 2020.
[175] A. Dato, T. Crepax, Brain Computer Interface: a Data Protection Perspective, Tilburg Law School, 2018.
[176] M. Veale, F. Zuiderveen Borgesius, Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach, Computer Law Rev. Int. 22 (4) (2021) 97–112.