IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 19, NO. 2, FEBRUARY 2023 2249

Physics-Informed Attention Temporal Convolutional Network for EEG-Based Motor Imagery Classification

Hamdi Altaheri, *Member, IEEE*, Ghulam Muhammad, *Senior Member, IEEE*, and Mansour Alsulaiman, *Senior Member, IEEE*

***Abstract*—The brain-computer interface (BCI) is a cutting-edge technology that has the potential to change the world. Electroencephalogram (EEG) motor imagery (MI) signal has been used extensively in many BCI applications to assist disabled people, control devices or environments, and even augment human capabilities. However, the limited performance of brain signal decoding is restricting the broad growth of the BCI industry. In this article, we propose an attention-based temporal convolutional network (ATCNet) for EEG-based motor imagery classification. The ATCNet model utilizes multiple techniques to boost the performance of MI classification with a relatively small number of parameters. ATCNet employs scientific machine learning to design a domain-specific deep learning model with interpretable and explainable features, multihead self-attention to highlight the most valuable features in MI-EEG data, temporal convolutional network to extract high-level temporal features, and convolutional-based sliding window to augment the MI-EEG data efficiently. The proposed model outperforms the current state-of-the-art techniques in the BCI Competition IV-2a dataset with an accuracy of 85.38% and 70.97% for the subject-dependent and subject-independent modes, respectively.**

***Index Terms*—Attention, classification, convolution neural network (CNN), deep learning, electroencephalography (EEG), motor imagery, scientific machine learning, temporal convolution networks (TCN).**

I. Introduction

THE brain-computer interface (BCI) is a system that interprets brain activity and converts it into commands to control an external device, such as a wheelchair or a drone. BCI is a cutting-edge technology that has the potential to transform the world and further enhances the quality of life, with a wide range of industrial applications spanning from medical applications to human augmentation [1], [2].

The brain signal can be recorded using various techniques, such as electroencephalography (EEG). EEG is a noninvasive method that records the electrical activities of the brain. The EEG signal is captured on the scalp as a two-dimensional (2-D) matrix of real values (time and channel). EEG is widely used and preferred over other techniques due to its ease of use, low cost, low risk, portability, and high temporal resolution, making it suitable for industrial applications.

Motor imagery (MI) is the activity of thinking about moving a part of the human body without physically moving it. EEG-based MI (MI-EEG) activities have been employed in a variety of medical applications, including stroke rehabilitation, wheelchair control, prostheses control, exoskeleton control, cursor control, speller, and thought-to-text conversion. MI-EEG signals have also been used in nonmedical applications such as vehicle control, drone control, environment control, smart home, security, gaming, and virtual reality [2]. Therefore, MI-EEG signals have great applicability in a variety of medical and nonmedical industry applications. However, the real-world applications are still limited by the decoding performance and generalization ability of the MI-EEG signal.

One of the main challenges for the real-world application of MI-EEG BCI is accurately recognizing human intention from low signal-to-noise ratio and nonstationary brain signals with various sources of artifacts, including biological artifacts (e.g., muscle movements, eye blinking), electronic equipment (e.g., computers and wireless devices), and environmental noise (e.g., light and sound). These artifacts, along with channel correlation, subject dependency, and high dimensionality of EEG signals make the analysis and classification of brain signal a challenging task.

Several conventional machine learning (ML) and deep learning (DL) approaches have been proposed to address the difficulties involved in classifying MI-EEG tasks. Among the conventional ML approaches that rely on manual feature extraction, filter bank common spatial patterns (FBCSP) [3], and its variants achieved the best performance in MI classification. In contrast to conventional methods, DL can learn distinct and latent features from raw EEG data without the requirement for preprocessing or manual feature extraction. DL has been used effectively in a

Manuscript received 27 May 2022; revised 7 July 2022; accepted 26 July 2022. Date of publication 9 August 2022; date of current version 13 December 2022. This work was supported by the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia under Grant DRI-KSU-1354. Paper no. TII-22-2210.*(Corresponding author: Ghulam Muhammad.)*

The authors are with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia (e-mail: haltaheri@ksu.edu.sa; ghulam@ksu.edu.sa; msuliman@ksu.edu.sa).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2022.3197419.

Digital Object Identifier 10.1109/TII.2022.3197419

基于脑电图的运动想象分类的物理信息注意时间卷积网络

Hamdi Altaheri，IEEE 会员，Ghulam Muhammad，IEEE 高级会员，
IEEE 高级会员 Mansour Alsulaiman

摘要：脑机接口 (BCI) 是一项有可能改变世界的尖端技术。脑电图 (EEG) 运动想象 (MI) 信号已广泛应用于许多 BCI 应用中，以帮助残疾人、控制设备或环境，甚至增强人类能力。然而，脑信号解码性能有限，限制了 BCI 行业的广泛发展。在本文中，我们提出了一种基于注意力机制的时间卷积网络 (ATCNet)，用于基于 EEG 的运动想象分类。ATCNet 模型利用多种技术来提高 MI 分类的性能，同时参数数量相对较少。ATCNet 采用科学机器学习来设计具有可解释和可说明特征的领域特定深度学习模型，多头自注意力机制可突出 MI-EEG 数据中最有价值的特征，时间卷积网络可提取高级时间特征，基于卷积的滑动窗口可有效地增强 MI-EEG 数据。所提出的模型在 BCI 竞赛 IV-2a 数据集中的表现优于当前最先进的技术，在主体相关和主体无关模式下的准确率分别为 85.38% 和 70.97%。

索引词——注意力、分类、卷积神经网络 (CNN)、深度学习、脑电图 (EEG)、运动想象、科学机器学习、时间卷积网络 (TCN)。

我

脑机接口 (BCI) 是一种解读大脑活动并将其转化为指令来控制外部设备（例如轮椅或无人机）的系统。BCI 是一项尖端技术，它具有改变世界、进一步提升生活质量的潜力，其应用范围十分广泛。

稿件收到日期：2022 年 5 月 27 日；修订日期：2022 年 7 月 7 日；接受日期：2022 年 7 月 26 日。出版日期：2022 年 8 月 9 日；当前版本日期：2022 年 12 月 13 日。本研究由沙特阿拉伯教育部研究与创新副署资助，资助编号：DRI-KSU-1354。论文编号：TII-22-2210。（通讯作者：Ghulam Muhammad。）作者就职于沙特阿拉伯利雅得 11543 国王沙特大学计算机与信息科学学院计算机工程系（邮箱：haltaheri@ksu.edu.sa；ghulam@ksu.edu.sa；msuliman@ksu.edu.sa）。
本文中一个或多个图片的彩色版本可在 https://doi.org/10.1109/TII.2022.3197419 上找到。

数字对象标识符 10.1109/TII.2022.3197419

工业应用范围涵盖从医疗应用到人类增强 [1]，[2]。

脑信号可以通过多种技术记录，例如脑电图 (EEG)。EEG 是一种记录脑电活动的非侵入式方法。EEG 信号在头皮上被捕获，形成包含实数值（时间和通道）的二维 (2-D) 矩阵。EEG 因其易用性、低成本、低风险、便携性和高时间分辨率而被广泛使用，并比其他技术更受青睐，非常适合工业应用。

运动想象 (MI) 是指在不实际移动身体某一部位的情况下，想象其运动的活动。基于脑电图的 MI (MI-EEG) 活动已广泛应用于各种医疗领域，包括中风康复、轮椅控制、假肢控制、外骨骼控制、光标控制、拼写器以及思维转文本。MI-EEG 信号也已应用于非医疗领域，例如车辆控制、无人机控制、环境控制、智能家居、安防、游戏和虚拟现实 [2]。因此，MI-EEG 信号在各种医疗和非医疗行业应用中都具有广泛的适用性。然而，MI-EEG 信号的解码性能和泛化能力仍然限制了其在实际应用中的应用。

MI-EEG BCI 在实际应用中面临的主要挑战之一是，如何从低信噪比、非平稳的脑信号中准确识别人类意图，这些脑信号包含各种伪影，包括生物伪影（例如肌肉运动、眨眼）、电子设备（例如计算机和无线设备）以及环境噪声（例如光和声）。这些伪影，加上脑电信号的通道相关性、受试者依赖性和高维度性，使得脑信号的分析和分类成为一项极具挑战性的任务。

为了解决 MI-EEG 任务分类难题，已提出了几种传统的机器学习 (ML) 和深度学习 (DL)。在依赖手动特征提取的传统 ML 方法中，滤波器组公共空间模式 (FBCSP) [3] 及其变体在 MI 分类中取得了最佳性能。与传统方法相比，DL 可以从原始 EEG 数据中学习独特且潜在的特征，而无需预处理或手动特征提取。DL 已在以下领域得到有效应用：

variety of applications, including image, video, audio, and text analysis [4], [5], [6], [7]. Recently, motivated by the significant success of DL techniques in other applications, many researchers have employed DL algorithms to classify MI tasks.

In the past five years, the number of studies using DL methods to classify MI tasks has increased rapidly [2]. Different DL architectures was proposed for MI classification including convolutional neural network (CNN) [8], [9], [10], [11], [12], [13], recurrent neural network (RNN) [14], [15], deep belief network (DBN) [16], autoencoder (AE) [17], and hybrid DL models [8], [18]. CNN was the most widely used architecture for MI classification [2]. Standard CNN models with light [12] and deep architectures [19] have been proposed, as well as many other CNN varieties, including inception-based CNN [10], [11], residual-based CNN [20], 3D-CNN [20], multiscale CNN [13], multilayer CNN [18], multibranch CNN [9], [20], and attention-based CNN [8], [9], [10], [11], [13]. Several other DL models have also been suggested by some studies for classifying MI tasks. Xu et al. [16] proposed a DBN model based on restricted Boltzmann machines for feature extraction and a support vector machine (SVM) for classification. Hassanpour et al. [17] proposed a stacked AE (SAE) to classify MI tasks using frequency features. In other studies, researchers have attempted to extract temporal information from the MI-EEG signal using recurrent neural networks. For example, Kumar et al. [14] proposed a long short-term memory (LSTM) model combined with FBCSP features and an SVM classifier. In another study, Luo and Chao [15] adopted FBCSP features and used them as inputs to a gated recurrent unit (GRU) model. The article showed that the GRU model performed better than the LSTM. In general, CNN models have shown better performance in MI task classification than other DL models [2], e.g., RNN, SAE, and DBN. Therefore, many researchers have suggested integrating CNN with other DL models, such as LSTM [8] and SAE [18], and encouraging results have been obtained.

Recently, a new CNN variant called temporal convolutional network (TCN) was specifically designed for time series modeling and classification [21]. TCN outperformed other recurrent networks such as LSTM and GRU in many sequence-related tasks [21]. In contrast to typical CNNs, TCN can exponentially expand the size of the receptive field with a linear increase in the number of parameters, and unlike RNNs it does not suffer from vanishing or exploding gradients. Some recent studies have used TCN architectures to classify MI tasks [22], [23]. Ingolfsson et al. [22] proposed a TCN model named EEG-TCN that combines TCN with the well-known EEGNet architecture [12]. A recent study in [23] attempted to improve the EEG-TCN model using the feature fusion technique. Our research is an ongoing contribution to these works, which utilizes scientific machine learning (SciML) and attention mechanism with TCN architecture.

Scientific machine learning is a new field that combines machine learning and scientific computing to produce domain-aware ML models that are reliable, robust, scalable, and interpretable. SciML aims to derive insights from scientific data to reduce ML model parameters, prevent overfitting, enhance extrapolation, and overcome domain-specific data challenges, including noisy data, high dimensionality, and low signal-to-noise. SciML can produce the next wave of data-driven scientific discovery in the engineering, physical, and medical sciences [24].

The attention mechanism is an effort to emulate human brain behavior of selectively focusing on a few significant elements while ignoring others. Integrating the attention mechanism with DL models helps to focus automatically (by learning) on the most important parts of the input data. The first attention-based model (RNN model) was proposed by Bahdanau et al. [25], known as additive attention. In the same year, Luong et al. [26] proposed an attention layer with a multiplication scoring function, known as multiplicative attention. In 2017, Google researchers proposed a pure attention model with multihead attention, which consists of several self-attention layers [27]. These attention-based models were originally proposed for natural language processing (NLP) and have subsequently been used in other fields. For computer vision, several attention blocks have been proposed, such as squeeze-and-excitation (SE) [28] and convolutional block attention module (CBAM) [29].

Recently, researchers have used attention-based DL models to classify MI-EEG signals [8], [9], [10], [11], [13]. For instance, Zhang et al. [8] employed self-attention with LSTM and graph neural representation to decode MI tasks. Amin et al. [10], [11] combined attention layers with inception-CNN and LSTM. In a more recent study, Altuwaijri et al. [9] proposed an attention-based multibranch CNN model for classifying MI tasks using raw data. The authors used three SE attention blocks as intermediate layers in three CNN branches.

Although the current studies showed promising results in decoding MI-EEG signals, the classification performance is still limited and requires further improvement.

In this article, we propose an attention-based temporal convolutional network, ATCNet, to decode MI-EEG brain signals. This research utilizes SciML to address domain-specific MI-EEG data challenges, which results in a robust, interpretable, and explainable DL model specifically designed for decoding MI-EEG brain signals. The proposed DL model processes the MI-EEG data in three steps: first, encode the MI-EEG signal into a sequence of high-level temporal representations using conventional layers, then, highlight the most valuable information in the temporal sequence using an attention layer, and finally, extract high-level temporal features from the highlighted information using a temporal convolutional layer. The proposed model utilizes a multihead self-attention and convolutional-based sliding window to boost the performance of MI classification. This article highlights the following contributions.

1) We propose a high-performance ATCNet model, which utilizes the powerful of TCN, SciML, attention mechanism, and convolutional-based sliding window.
2) Performing sliding window using convolution helps augment MI data and efficiently enhance accuracy, by parallelizing the process and reducing computations.
3) Self-attention helps the DL model to attend to the most effective MI information in the EEG data, and the multiple heads help to focus on multiple positions, resulting in multiple attention representations.

深度学习已被广泛应用于图像、视频、音频和文本分析等各种应用领域 [4], [5], [6], [7]。近年来，受深度学习技术在其他应用领域取得巨大成功的推动，许多研究人员开始采用深度学习算法对机器翻译任务进行分类。

在过去五年中，使用深度学习方法对机器翻译任务进行分类的研究数量迅速增加 [2]。针对机器翻译分类，提出了不同的深度学习架构，包括卷积神经网络 (CNN) [8]、[9]、[10]、[11]、[12]、[13]、循环神经网络 (RNN) [14]、[15]、深度信念网络 (DBN) [16]、自编码器 (AE) [17] 和混合深度学习模型 [8]、[18]。CNN 是机器翻译分类中使用最广泛的架构 [2]。已经提出了具有轻量级 [12] 和深度架构 [19] 的标准 CNN 模型，以及许多其他 CNN 类型，包括基于 Inception 的 CNN [10]、[11]、基于残差的 CNN [20]、3D-CNN [20]、多尺度 CNN [13]、多层 CNN [18]、多分支 CNN [9]、[20] 和基于注意力机制的 CNN [8]、[9]、[10]、[11]、[13]。一些研究还提出了几种其他 DL 模型用于对 MI 任务进行分类。Xu 等人 [16] 提出了一种基于受限玻尔兹曼机的 DBN 模型用于特征提取，并使用支持向量机 (SVM) 进行分类。Hassanpour 等人 [17] 提出了一种堆叠 AE (SAE) 模型，使用频率特征对 MI 任务进行分类。在其他研究中，研究人员尝试使用循环神经网络从 MI-EEG 信号中提取时间信息。例如，Kumar 等人 [14] 提出了一种结合 FBCSP 特征和 SVM 分类器的长短期记忆 (LSTM) 模型。在另一项研究中，Luo 和 Chao [15] 采用 FBCSP 特征，并将其作为门控循环单元 (GRU) 模型的输入。该文章表明 GRU 模型的性能优于 LSTM。总体而言，CNN 模型在 MI 任务分类中的表现优于其他深度学习模型 [2]，例如 RNN、SAE 和 DBN。 因此，许多研究人员建议将 CNN 与其他 DL 模型（如 LSTM [8] 和 SAE [18]）相结合，并取得了令人鼓舞的结果。

最近，一种名为时间卷积网络 (TCN) 的新型 CNN 变体被专门设计用于时间序列建模和分类 [21]。在许多与序列相关的任务中，TCN 的表现优于 LSTM 和 GRU 等其他循环网络 [21]。与典型的 CNN 相比，TCN 可以通过参数数量的线性增加以指数方式扩展感受野的大小，并且与 RNN 不同，它不会受到梯度消失或爆炸的影响。一些近期研究已使用 TCN 架构对 MI 任务进行分类 [22]，[23]。Ingolfsson 等人 [22] 提出了一种名为 EEG-TCN 的 TCN 模型，该模型将 TCN 与著名的 EEGNet 架构 [12] 相结合。[23] 中的一项近期研究尝试使用特征融合技术改进 EEG-TCN 模型。我们的研究是对这些工作的持续贡献，它将科学机器学习 (SciML) 和注意力机制与 TCN 架构相结合。

科学机器学习是一个新兴领域，它将机器学习与科学计算相结合，构建可靠、稳健、可扩展且可解释的领域感知型机器学习模型。科学机器学习旨在从科学数据中获取洞察，以减少机器学习模型参数，防止过度拟合，增强外推能力，并克服特定领域的数据挑战。包括噪声数据、高维数据和低信噪比数据。SciML 有望在工程、物理和医学领域掀起下一波数据驱动的科学发现浪潮 [24]。

注意力机制旨在模拟人脑选择性地关注少数重要元素而忽略其他元素的行为。将注意力机制与深度学习模型相结合，有助于通过学习自动聚焦于输入数据中最重要的部分。第一个基于注意力机制的模型（RNN 模型）由 Bahdanau 等人 [25] 提出，称为加性注意力机制。同年，Luong 等人 [26] 提出了一个带有乘法评分函数的注意力层，称为乘性注意力机制。2017 年，谷歌研究人员提出了一种具有多头注意力机制的纯注意力模型，该模型由多个自注意力层组成 [27]。这些基于注意力机制的模型最初是为自然语言处理 (NLP) 提出的，随后被用于其他领域。对于计算机视觉，已经提出了几个注意力模块，例如挤压和激励 (SE) [28] 和卷积块注意力模块 (CBAM) [29]。

最近，研究人员使用基于注意力机制的深度学习模型对 MI-EEG 信号进行分类 [8]、[9]、[10]、[11]、[13]。例如，Zhang 等人 [8] 将自注意力机制与 LSTM 和图神经表征相结合，以解码 MI 任务。Amin 等人 [10]、[11] 将注意力机制与 Inception-CNN 和 LSTM 相结合。在最近的一项研究中，Altuwaijri 等人 [9] 提出了一种基于注意力机制的多分支 CNN 模型，用于使用原始数据对 MI 任务进行分类。作者在三个 CNN 分支中使用了三个 SE 注意力模块作为中间层。

尽管目前的研究在解码 MI-EEG 信号方面取得了有希望的成果，但分类性能仍然有限，需要进一步改进。

在本文中，我们提出了一种基于注意力机制的时间卷积网络 ATCNet，用于解码 MI-EEG 脑信号。本研究利用 SciML 解决特定领域的 MIEEG 数据挑战，最终构建了一个专为解码 MI-EEG 脑信号而设计的稳健、可解释且易于理解的深度学习模型。该深度学习模型分三步处理 MI-EEG 数据：首先，使用常规层将 MI-EEG 信号编码为一系列高级时间表征；然后，使用注意力机制层突出显示时间序列中最有价值的信息；最后，使用时间卷积层从突出显示的信息中提取高级时间特征。该模型利用多头自注意力机制和基于卷积的滑动窗口来提升 MI 分类的性能。本文重点介绍了以下贡献。

1) 我们提出了一个高性能的 ATCNet 模型，该模型充分利用了 TCN、SciML、注意力机制和基于卷积的滑动窗口的强大功能。2) 使用卷积执行滑动窗口有助于增强 MI 数据，并通过并行化处理和减少计算量来有效提高准确率。3) 自注意力机制帮助深度学习模型关注脑电图数据中最有效的 MI 信息，而多个头部结构有助于将注意力集中在多个位置，从而产生多种注意力表征。

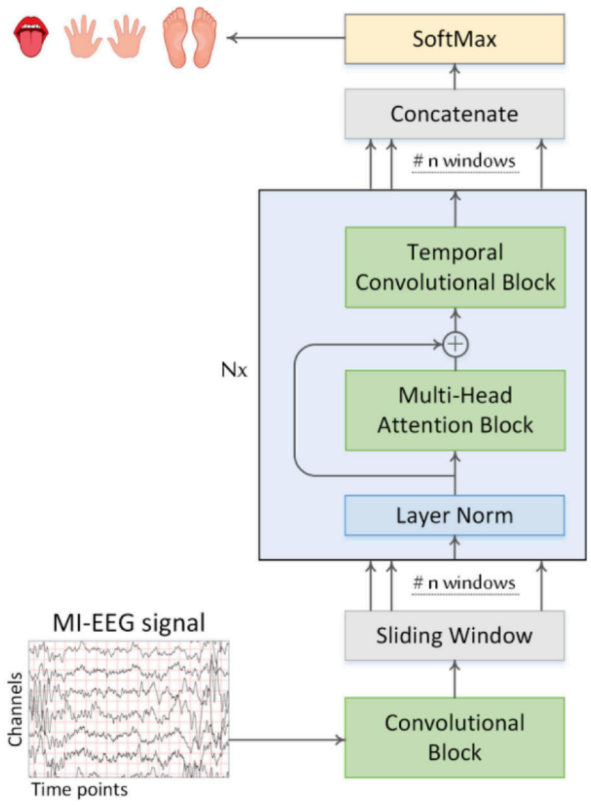


Fig. 1. Components of the proposed ATCNet model.

4) The proposed model achieves outstanding results in the BCI Competition IV-2a (BCI-2a) dataset [30].
5) For reproducibility, the code for this research and the trained models will be released on GitHub.

The rest of this article is organized as follows. Section II describes the proposed ATCNet model. In Section III, we present and discuss the results. Then, Section IV concludes this article.

II. Proposed ATCNet Model

The proposed ATCNet model consists of three main blocks: convolutional (CV) block, attention (AT) block, and temporal convolutional (TC) block, as shown in Fig. 1. CV block encodes low-level spatio-temporal information within the MI-EEG signal through three convolutional layers: temporal, channel depthwise, and spatial convolutions. The output of the CV block is a temporal sequence with a higher level representation. The AT block then highlights the most important information in the temporal sequence using a multihead self-attention (MSA). Finally, the TC block extracts high-level temporal features within the temporal sequence using TCN and feeds them into a fully connected (FC) layer with a SoftMax classifier.

The temporal sequence, output from CV block, can be split into multiple windows and each is fed to AT/TC blacks separately. The output of all windows is then concatenated and fed to a SoftMax classifier. This helps efficiently augment the data and enhances accuracy. The details of ATCNet blocks are described in the following sections.

A. Preprocessing and Input Representation

In this article, we feed raw MI-EEG signals into the proposed model without preprocessing, i.e., the full frequency band, all channels, and without artifact removal.

ATCNet model takes as input a motor imagery trial $X_i \in \mathbb{R}^{C\times T}$ consisting of C channels (EEG electrodes) and T time points. The objective of the ATCNet model is to map the input MI trial X_i to its corresponding class y_i, given a set of m labeled MI trials $S = \{X_i, y_i\}_{i=1}^{m}$, where $y_i \in \{1, \ldots, n\}$ is the corresponding class label for trial X_i and n is the total number of defined classes for set S. For the BCI-2a [30] dataset, $T = 1125$ time points, $C = 22$ EEG channels, $n = 4$ MI classes, and $m = 5184$ MI trials.

B. Convolutional Block

The CV block is similar to the EEGNet architecture proposed in [12]. CV block differs from EEGNet by using 2-D convolution instead of separable convolution, which showed better performance. CV block also uses different parameter values than those used in [12].

CV block consists of three convolutional (conv) layers, as shown in Fig. 2. The first layer performs a temporal convolution using F_1 filters of size $(1, K_C)$, where K_C is the filter length in the time axis. K_C was set to be one-fourth of the sampling rate (64 for BCI-2a). This allows the filters to extract temporal information associated with frequencies above 4 Hz. The output of this layer is F_1 temporal feature maps. The second layer is a depthwise convolution with F_2 filters of size $(C, 1)$, where C is the number of EEG channels. Using depthwise convolution, each filter extracts spatial features (i.e., related to EEG channels) from a single temporal feature map. Therefore, the output of this layer is $F_1 \times D$ feature maps, where D is the number of filters linked to each temporal feature map in the previous layer. D is set empirically to 2. $F_1 \times D$ determine the output dimension of the CV block. The depthwise convolution is followed by an average pooling layer of size $(1, 8)$ to abstract the temporal data by a factor of 8. This reduces the sampling rate of the signal to ~32 Hz. The third convolutional layer consists of F_2 filters of size $(1, K_{C2})$. K_{C2} was set to 16 to decode MI activities within 500 ms (for 32 Hz sampled data). Finally, a second average pooling layer with a size of $(1, P_2)$ is used to reduce the sampling rate to $\sim 32/P_2$ Hz. P_2 is used to control the length of the temporal sequence produced by CV block. The second and third conv layers are followed by batch normalization [31] to speed up network training and then by exponential linear unit (ELU) activation for nonlinearity.

CV block output a sequence $z_i \in \mathbb{R}^{T_c \times d}$ of temporal representation consisting of T_c temporal vectors each with dimension $d = F_2 = F_1 \times D$. We empirically set d *to* 32. The length of the temporal sequence z_i is determined by

$$T_c = T/8P_2 \tag{1}$$

where T refers to the time points of the original EEG signal.

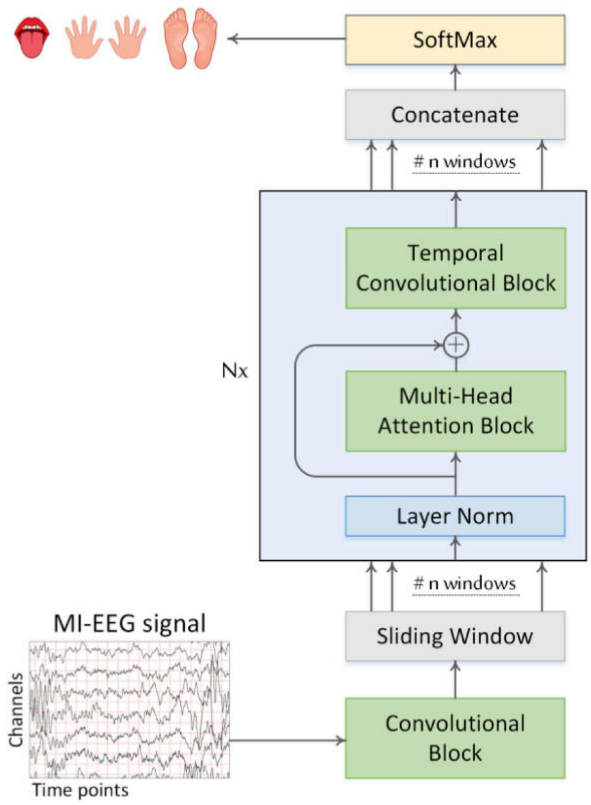


图 1. 所提出的 ATCNet 模型的组成部分。

4）所提出的模型在脑机接口竞赛 IV-2a（BCI-2a）数据集中取得了优异的成绩[30]。5）为了便于复现，本研究的代码和训练好的模型将在 GitHub 上发布。

本文其余部分安排如下。第二部分描述了提出的 ATCNet 模型。第三部分展示并讨论了结果。第四部分对本文进行了总结。

二、PATCNM

所提出的 ATCNet 模型由三个主要模块组成：卷积 (CV) 模块、注意力 (AT) 模块和时间卷积 (TC) 模块，如图 1 所示。CV 模块通过三个卷积层（时间卷积、通道深度卷积和空间卷积）对 MI-EEG 信号中的低级时空信息进行编码。CV 模块的输出是具有更高级别表示的时间序列。然后，AT 模块使用多头自注意力 (MSA) 突出显示时间序列中最重要的信息。最后，TC 模块使用 TCN 提取时间序列中的高级时间特征，并将其输入到带有 SoftMax 分类器的全连接 (FC) 层。

CV 模块输出的时间序列可以拆分成多个窗口，每个窗口分别输入到 AT/TC 黑场。所有窗口的输出随后被连接起来，并输入到 SoftMax 分类器。

这有助于有效地增强数据并提高准确性。ATCNet 模块的详细信息将在以下章节中描述。

A.预处理和输入表示

在本文中，我们将原始 MI-EEG 信号输入到所提出的模型中，无需预处理，即全频带、所有通道，并且不去除伪影。

ATCNet 模型以运动想象试验 $X \in R$ 作为输入，该试验由 C 个通道（EEG 电极）和 T 个时间点组成。ATCNet 模型的目标是将输入的想象试验 X 映射到其对应的类 y。给定一组 m 个带标签的想象试验 $S = \{X, y\}$，其中 $y \in \{1, ..., n\}$ 是试验 X 对应的类标签，n 是集合 S 中已定义类的总数。对于 BCI-2a [30] 数据集，$T = 1125$ 个时间点，$C = 22$ 个 EEG 通道，$n = 4$ 个想象试验类别，以及 $m = 5184$ 个想象试验。

B.卷积块

CV 模块与 [12] 中提出的 EEGNet 架构类似。CV 模块与 EEGNet 的不同之处在于，它使用二维卷积而非可分离卷积，从而展现出更佳的性能。此外，CV 模块使用的参数值也与 [12] 中的不同。

CV 块由三个卷积 (conv) 层组成，如图 2 所示。第一层使用大小为 (1, K) 的 Ffilters 执行时间卷积，其中 K 是时间轴上的滤波器长度。K 设置为采样率的四分之一（对于 BCI-2a 为 64）。这允许滤波器提取与 4 Hz 以上频率相关的时间信息。该层的输出是 Ftemporal 特征图。第二层是具有大小为 (C, 1) 的 Ffilters 的深度卷积，其中 C 是 EEG 通道的数量。使用深度卷积，每个滤波器从单个时间特征图中提取空间特征（即与 EEG 通道相关的特征）。因此，该层的输出是 $F \times D$ 特征图，其中 D 是链接到前一层中每个时间特征图的滤波器数量。D 根据经验设置为 2。$F \times D$ 确定 CV 块的输出维度。深度卷积之后是大小为（1, 8）的平均池化层，将时间数据抽象为原来的 8 倍。这将信号的采样率降低到 ~32 Hz。第三个卷积层由大小为（1, K）的 F 个滤波器组成。将 K 设置为 16，以便在 500 毫秒内解码 MI 活动（对于 32 Hz 采样数据）。最后，使用大小为（1, P）的第二个平均池化层将采样率降低到 ~32/PHz。P 用于控制 CV 块产生的时间序列的长度。第二和第三个卷积层之后是批量归一化 [31] 以加速网络训练，然后是指数线性单元 (ELU) 激活以实现非线性。

CV 块输出一个时间表示序列 $z \in R$，该序列由 T 个时间向量组成，每个时间向量的维度为 $d = F = F \times D$。我们根据经验将 d 设置为 32。时间序列 z 的长度由下式确定

$$T = T/8P \quad (1)$$

其中 T 表示原始 EEG 信号的时间点。

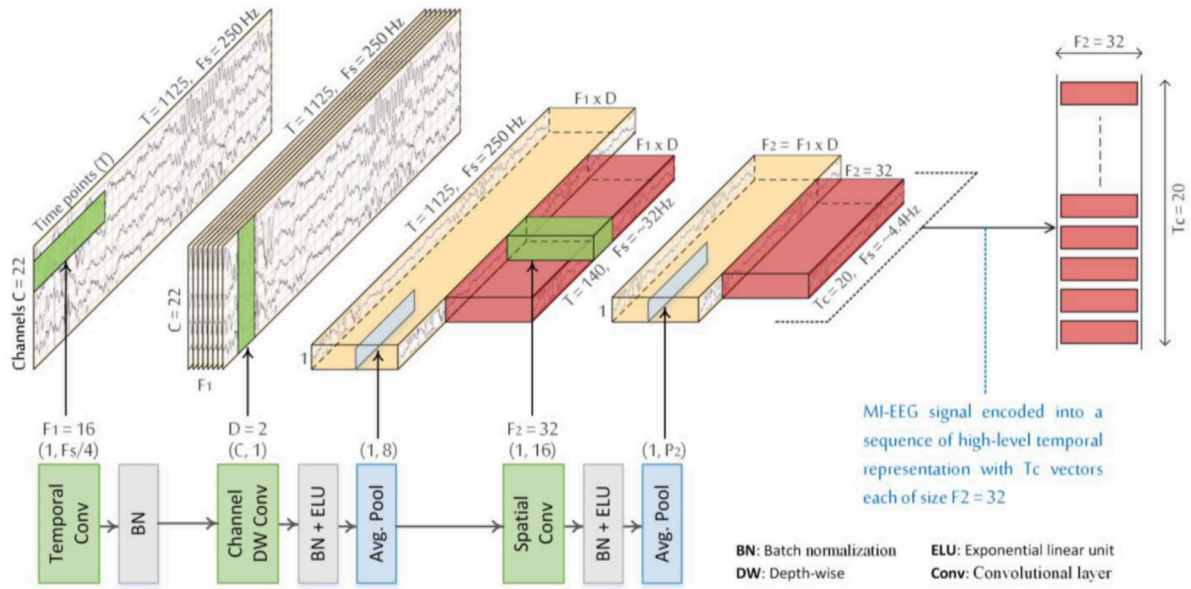


Fig. 2. CV block performs spatio-temporal encoding through three convolutional layers. The CV block receives a raw MI-EEG signal and outputs a temporal sequence with T_C elements. Each element is a vector of size F_2.

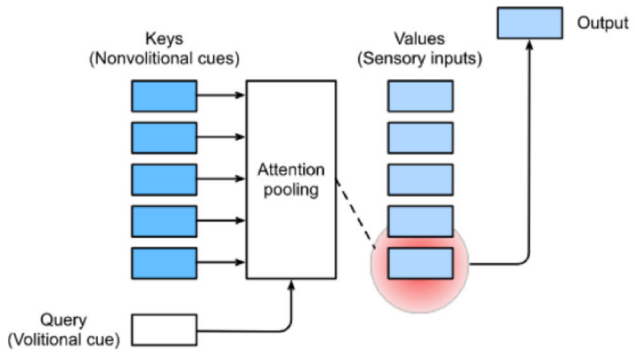


Fig. 3. Interaction of queries and keys creates attention pooling that biases the selection of values.

C. Convolutional-Based Sliding Window (SW)

Instead of entering the whole T_c samples of z_i to the later layers, a sliding window has been used to divide the temporal sequences into multiple windows. This helps to augment the data and enhance the decoding accuracy. However, the sliding window raises the computations, because it requires the input data to be passed through the DL model n times (instead of once), where n stands for the number of windows. As a result, the computations are incremented n times. But in our approach, we used a sliding window as integration with convolutional layers (in the convolutional block). In this approach, convolution computations are performed once for all windows, which reduces training and inference time by parallelizing the process. This technique was originally used in sliding-window-based object detection. The convolutional-based sliding window has been described in detail by Schirrmeister et al. [32].

We used a sliding window of length T_w with one step stride to divide the temporal sequence z_i into multiple windows $z_i^w \in \mathbb{R}^{T_w \times d}$ with $w = 1, \ldots, n$ denoting the window index, and n is the total number of windows. Each window z_i^w is then entered separately to the later AT block and then to the TC block. The window length T_w is determined by

$$T_w = T_c - n + 1,\ T_c > n \geq 1 \tag{2}$$

$$T_w = {}^{T}/_{8P_2} - n + 1. \tag{3}$$

If the CV block performs two temporal pooling of size $P_1 = 8$ and $P_2 = 7$, CV will produce a temporal sequence z_i consisting of $T_c = 20$ vectors [Eq. (1), where $T = 1125$]. Each vector will represent 56 (8×7) time-points in the original MI-EEG signal x_i. Therefore, performing one step sliding in the z_i is equivalent to 56 time-steps sliding in the original signal x_i.

D. Attention Block

In psychology, the cognitive process of selectively focusing on one or a few things while disregarding others is known as attention. In deep neural networks, the attention mechanism is an effort to emulate the human brain behavior of selectively focusing on a few significant elements while ignoring others. In the visual world, subjects use both volitional and nonvolitional cues to selectively focus attention. The former is task-dependent, and the latter is based on the conspicuity and saliency of things in the surroundings. Inspired by the voluntary and involuntary attention cues, the attention mechanism can be emulated using three components: values (sensory inputs), keys (nonvolitional cues), and queries (volitional cues). The interaction of queries and keys creates attention pooling that biases the selection of values, as demonstrated in Fig. 3.

The attention mechanism can be implemented based on attention scores or by different machine learning algorithms such as reinforcement learning. This article adopts an attention scores-based approach, i.e., MSA, due to its large success in various fields, such as NLP and computer vision.

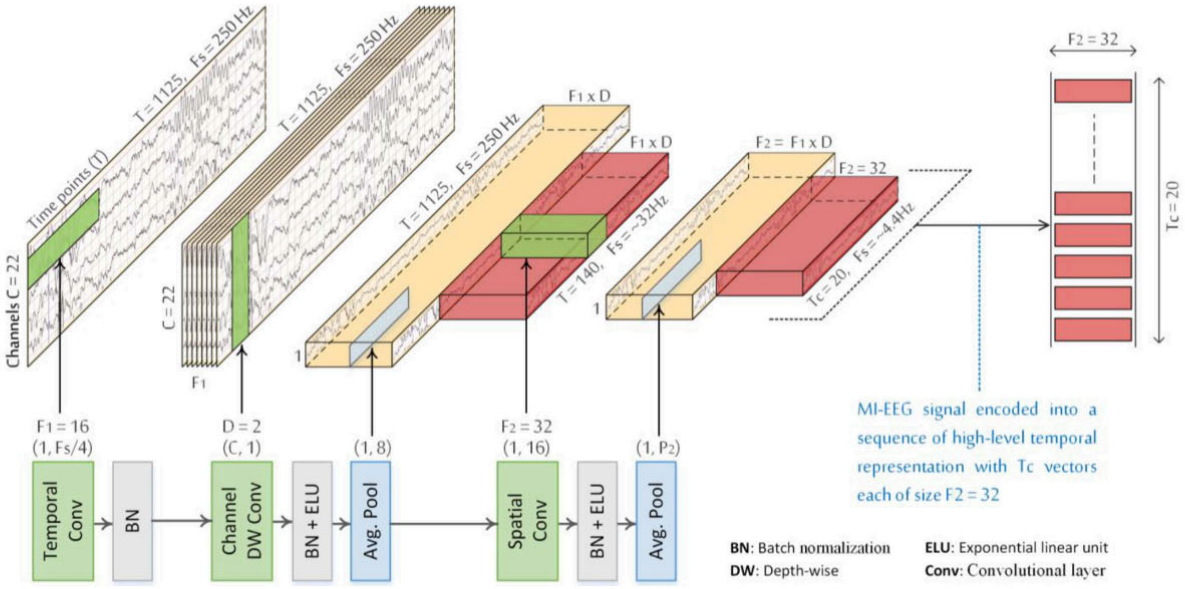


图 2. CV 模块通过三个卷积层执行时空编码。CV 模块接收原始 MI-EEG 信号，并输出包含 Telements 的时间序列。每个元素都是一个大小为 F 的向量。

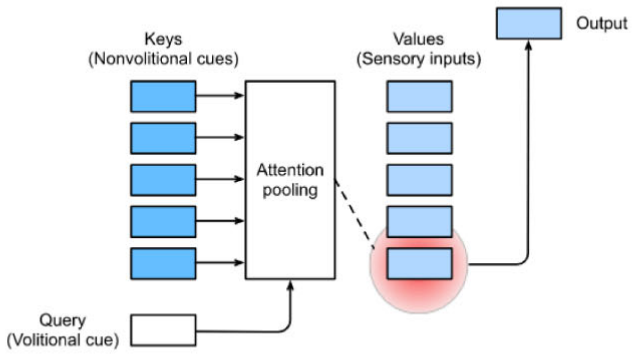


图 3. 查询和键的交互会产生注意力池，从而影响值的选择。

C.基于卷积的滑动窗口（SW）

我们没有将 z 的整个 T 个样本输入到后面的层，而是使用滑动窗口将时间序列分成多个窗口。这有助于扩充数据并提高解码精度。然而，滑动窗口会增加计算量，因为它要求输入数据通过深度学习模型 n 次（而不是一次），其中 n 代表窗口的数量。因此，计算量增加了 n 次。但在我们的方法中，我们使用滑动窗口与卷积层（在卷积块中）集成。在这种方法中，卷积计算对所有窗口执行一次，通过并行化过程减少了训练和推理时间。该技术最初用于基于滑动窗口的目标检测。Schirrmeister 等人 [32] 详细描述了基于卷积的滑动窗口。

我们使用长度为 T 且步长为一步的滑动窗口将时间序列 zin 划分为多个窗口 z∈R，其中 w = 1, . . . , n 表示窗口索引 n 为窗口总数。每个窗口 z 分别进入后面的 AT 块，然后进入 TC 块。窗口长度 T 由以下公式确定

$$T=T-n+1,\ T>n\geqslant 1 \tag{2}$$

$$T=T/\quad 8P-n+1。 \tag{3}$$

如果 CV 模块执行两个大小分别为 P= 8 和 P= 7 的时间池化，CV 将生成一个由 T= 20 个向量组成的时间序列 z[公式 (1)，其中 T = 1125]。每个向量代表原始 MI-EEG 信号 x 中的 56 个（8 × 7）时间点。因此，在 z 中执行一步滑动相当于在原始信号 x 中执行 56 个时间步长滑动。

D.注意力模块

在心理学中，选择性地关注一件或几件事物而忽略其他事物的认知过程被称为注意力。在深度神经网络中，注意力机制旨在模拟人类大脑选择性地关注少数重要元素而忽略其他元素的行为。在视觉世界中，受试者会使用有意识和非有意识的线索来选择性地集中注意力。前者依赖于任务，而后者则基于周围事物的显著性和显著性。受有意识和非有意识注意力线索的启发，注意力机制可以通过三个组成部分来模拟：值（感官输入）、键（非意识线索）和查询（意识线索）。查询和键的相互作用会形成注意力池，从而影响值的选择，如图 3 所示。

注意力机制可以基于注意力分数来实现，也可以通过不同的机器学习算法（例如强化学习）来实现。本文采用基于注意力分数的方法，即 MSA，因为它在自然语言处理（NLP）和计算机视觉等各个领域都取得了巨大的成功。

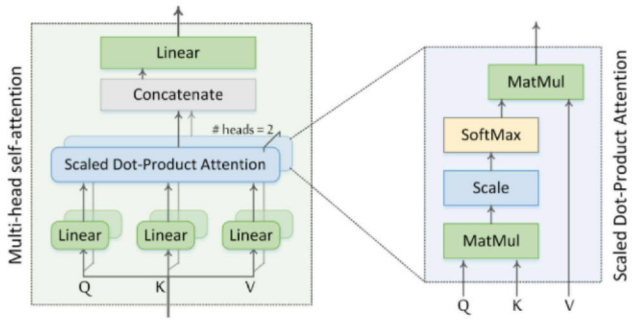


Fig. 4. Multihead self-attention.

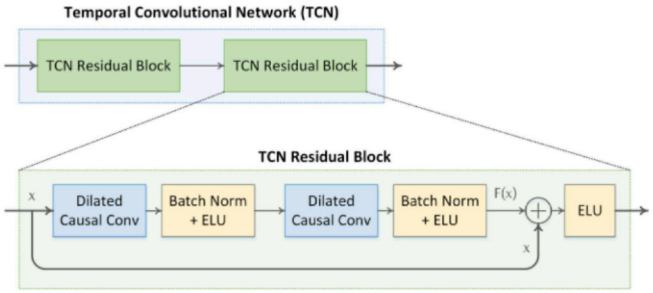


Fig. 5. Architecture of the TCN consisting of two residual blocks.

The attention block consists of an MSA layer, as described in [27]. MSA consists of several self-attention layers (i.e., scaled dot-product attention) called heads, as shown in Fig. 4. Each self-attention layer consists of three main components: query Q, keys K, and values V. Interactions between query and keys produce attention scores that guide selection bias over values. The detailed implementation of this interaction is as follows. Given a window representation z_i^w, encoded by CV block, the query/key/value vectors are calculated for each batch using linear transformation as

$$q_t^h = W_Q^h \; LN\left(z_{i,t}^w\right) \in \mathbb{R}^{d_H}, \; W_Q^h \in \mathbb{R}^{d \times d_H} \tag{4}$$

$$k_t^h = W_K^h \; LN\left(z_{i,t}^w\right) \in \mathbb{R}^{d_H}, \; W_K^h \in \mathbb{R}^{d \times d_H} \tag{5}$$

$$v_t^h = W_V^h \; LN\left(z_{i,t}^w\right) \in \mathbb{R}^{d_H}, \; W_V^h \in \mathbb{R}^{d \times d_H} \tag{6}$$

where LN stands for Layer Normalization [33], $t = 1, \ldots, T_w$ is an index over the temporal vectors in window w and T_w is the length of the window (the total number of temporal vectors), $h = 1, \ldots, H$ is an index over multiple attention heads and H is the total number of heads. The diminution of the attention head is set empirically to $d_H = d/2H$.

Given a query $q_t^h \in \mathbb{R}^{q=d_H}$ and T_w key-value pairs (k_1^h, v_1^h), $\ldots$, $(k_{T_w}^h, v_{T_w}^h)$, where $k_{t'}^h \in \mathbb{R}^{k=d_H}$ and $v_{t'}^h \in \mathbb{R}^{v=d_H}$. The attention pooling f that generates the context vector c_t^h is defined as a weighted sum of the values $v_{t'}^h$

$$c_t^h = f\left(q_t^h, k_{t'}^h, v_{t'}^h\right) = \sum_{t'=1}^{T_w} \alpha_{tt'}^h v_{t'}^h \in \mathbb{R}^{d_H}, \; \sum_{t'=1}^{T_w} \alpha_{tt'}^h = 1. \tag{7}$$

The attention weight (scalar) $\alpha_{tt'}^h$ of the query q_t^h and key $k_{t'}^h$ is calculated by applying the SoftMax function on the corresponding alignment scores $e_{tt'}^h$ as follows:

$$\alpha_{tt'}^h = \text{softmax}\left(e_{tt'}^h\right) = \frac{\exp\left(e_{tt'}^h\right)}{\sum_{k=1}^{T_w} \exp\left(e_{tk}^h\right)} \in \mathbb{R} \cdot \tag{8}$$

The alignment scores $e_{tt'}^h$ are calculated using the attention scoring function a, as in (9). Distinct selections for the attention scoring function a result in different attention pooling behaviors. Two common scoring functions have been proposed: additive attention (Bahdanau attention [25]) and multiplicative attention (Luong attention [26]). In this article, we use multiplicative attention, specifically scaled dot-product attention defined by Vaswani et al. [27], which is more computationally efficient. The dot product operation, however, necessitates that both the query and the key have the same vector length. The scoring function of scaled dot-product attention is defined in 10. The dot product is divided by $\sqrt{d_H}$ to ensure that the variance of the dot product remains constant regardless of vector length

$$e_{tt'}^h = a\left(q_t^h, k_{t'}^h\right) \in \mathbb{R} \tag{9}$$

$$a = \frac{\left(q_t^h\right)^T k_{t'}^h}{\sqrt{d_H}} \in \mathbb{R} \cdot \tag{10}$$

For each head, the context vectors of the scaled dot-product attention for a minibatch with $n = T_w$ queries and $m = T_w$ key-values pairs (global attention) are determined by (11), where keys and queries of length d_H and values of length v (in this article $v = d_H = 8$). Attention context vectors manage and quantify the interdependence either between the input and output components (general attention) or within the input components (self-attention). In this article, we adopt the self-attention mechanism as it helps parallel computing attention to all inputs at the same time

$$C^h = \text{softmax}\left(\frac{Q^h \left(K^h\right)^T}{\sqrt{d_H}}\right) V^h \in \mathbb{R}^{((n=T_w)\times(v=d_H))}$$
$$\text{Where} Q \in \mathbb{R}^{n \times d_H}, K \in \mathbb{R}^{m \times d_H}, \text{ and } V \in \mathbb{R}^{m \times v}. \tag{11}$$

Then, the MSA is computed by projecting the concatenation of the context vectors from all heads and adding the result to the input window z_i^w using a residual connection, as in

$$z_i^w = W_O \; \left[C^1, \ldots, C^H\right] + z_i^w \in \mathbb{R}^{T_w \times d}, \; W_O \in \mathbb{R}^{d_H \times d}. \tag{12}$$

E. Temporal Convolutional Block

The TC block has the same architecture as the TCN described in [22]. TCN consists of a stack of residual blocks. The residual block composes of two dilated causal convolutional layers, each one followed by batch normalization [31] and ELU activation, as shown in Fig. 5.

Causal convolutions are used to prevent any information from traveling from the future to the past, i.e., the output at time t depends only on the inputs from time t and earlier. Dilated convolutions allow the receptive field to be expanded exponentially while increasing the network depth. Therefore, dilated causal convolutions can learn relationships in long sequences. The residual connection performs an elementwise addition of the input and output feature map $F(x) + x$, which is effective

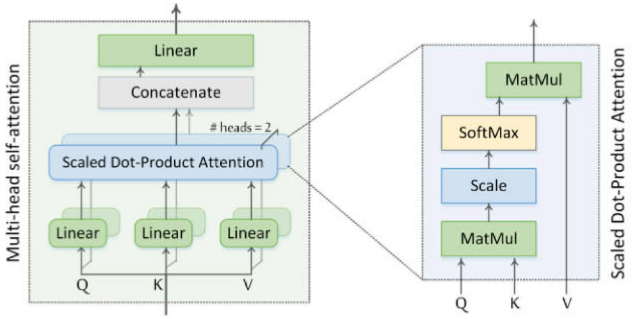


图 4.多头自注意力。

注意力模块由一个自注意力层 (MSA) 组成，如 [27] 中所述。MSA 由多个称为“头”的自注意力层（即缩放点积注意力）组成，如图 4 所示。每个自注意力层由三个主要部分组成：查询 Q、键 K 和值 V。查询和键之间的交互会产生注意力分数，从而引导对值的选择偏差。此交互的具体实现如下。给定一个由 CV 模块编码的窗口表示 z，使用线性变换为每个批次计算查询/键/值向量，如下所示

$$q = W_{LN}(z) \in R, W \in R \quad (4)$$

$$k = W_{LN}(z) \in R, W \in R \quad (5)$$

$$v = W_{LN}(z) \in R, W \in R \quad (6)$$

其中 LN 代表层正则化 [33]，t = 1, … , T 是窗口 w 中时间向量的索引，T 是窗口的长度（时间向量的总数），h = 1, … , H 是多个注意力头的索引，H 是注意力头的总数。注意力头的缩减量根据经验设置为 d= d/2H。

给定一个查询 q∈RandT 个键值对 (k, v),,(k, v)，其中 k∈Randv∈R。生成上下文向量 c 的注意力池 f 被定义为值 v 的加权和

$$c = f(q, k, v) = \sum_{t=1}^{T} \alpha v \in R, \quad \sum_{t=1}^{T} \alpha = 1。$$

（7）查询 q 和键 k 的注意力权重（标量）α 通过将 SoftMax 函数应用于相应的对齐分数来计算，如下所示：

$$\alpha = \text{softmax}(e) = \frac{\text{经验值}(e)}{\sum_{k=1}^{\text{指数}} (e)} \in R \cdot \quad (8)$$

对齐分数使用注意力评分函数 a 计算，如 (9) 所示。不同的注意力评分函数 a 会导致不同的注意力池化行为。目前已提出了两种常见的评分函数：加法注意力（Bahdanau 注意力 [25]）和乘法注意力（Luong 注意力 [26]）。在本文中，我们使用乘法注意力，特别是 Vaswani 等人 [27] 定义的缩放点积注意力，其计算效率更高。

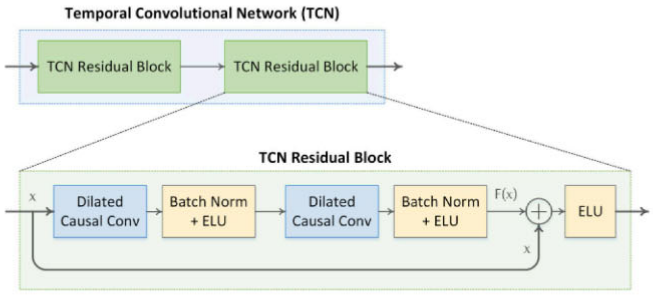


图 5.由两个残差块组成的 TCN 架构。

然而，点积运算要求查询和键具有相同的向量长度。缩放点积注意力机制的评分函数定义在 10 中。点积除以 √ d 确保点积的方差保持不变，无论向量长度如何

$$e = a(q, k) \in R \quad (9)$$

$$a = \frac{(q) \cdot k}{\sqrt{d}} \in R \cdot \quad (10)$$

对于每个头，对于具有 n = T 个查询和 m = T 个键值对（全局注意力）的小批量，缩放点积注意力的上下文向量由（11）确定，其中键和查询的长度为 d，值的长度为 v（在本文中 v = d = 8）。注意力上下文向量管理和量化输入和输出组件之间（通用注意力）或输入组件内部（自注意力）的相互依赖性。在本文中，我们采用自注意力机制，因为它有助于同时并行计算对所有输入的注意力。

$$C = \text{softmax}\left(\frac{Q \in R, K \in R}{\sqrt{d}}\right) \text{其中 } Q \in R, K \in R \text{ 且 } V \in R。 \quad (11)$$

然后，通过投影所有头部的上下文向量的连接，并使用残差连接将结果添加到输入窗口来计算 MSA，如下所示

$$z = W[C, \ldots, C] + z \in R, W \in R。 \quad (12)$$

E.时间卷积块

TC 块的架构与 [22] 中描述的 TCN 相同。TCN 由一堆残差块组成。残差块由两个扩张因果卷积层组成，每个层后接批量归一化 [31] 和 ELU 激活函数，如图 5 所示。

因果卷积用于防止任何信息从未来传递到过去，即时间 t 的输出仅取决于时间 t 及更早的输入。扩张卷积允许感受野呈指数级扩展，同时增加网络深度。因此，扩张因果卷积可以学习长序列中的关系。残差连接对输入和输出特征图 F(x) + x 进行元素级加法，这非常有效

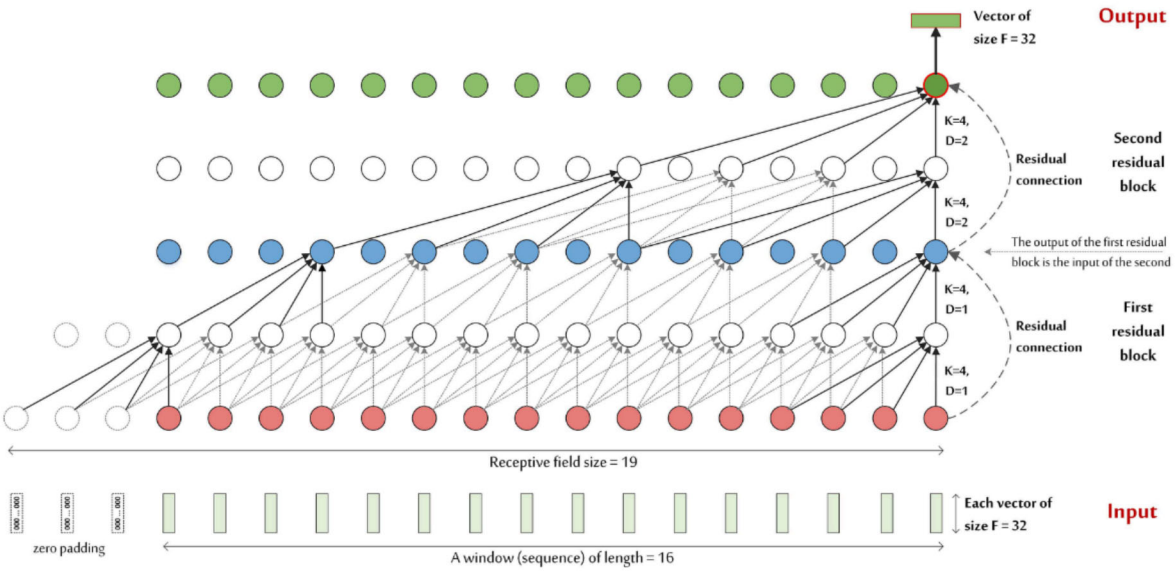


Fig. 6. Visualize a TCN block with a depth of 2, i.e., consisting of two residual blocks, kernel size = 4, and number of filters = 32. The output of the first residual block is the input of the second. The receptive field size for this TCN will be 19, so the input length T_w should be ≤ 19. This figure shows a sequence of 16 temporal elements (T_w= 16) entering the TCN.

in deep networks due to its ability to learn the identity function. In the residual block, we use identity mapping because the input and output dimensions are identical (32), otherwise, a linear transformation, i.e., 1×1 convolution, is used.

The receptive field size (RFS) of the TCN increases exponentially with the number of stacked residual blocks L due to the exponential increase in dilation D with each succeeding block. The RFS is controlled by two parameters: the number of residual blocks L and the kennel size K_T, as defined in

$$\text{RFS} = 1 + 2\,(K_T - 1)\left(2^L - 1\right). \tag{13}$$

In the proposed ATCNet, the TC block consists of a TCN with residual blocks and 32 filters of size $K_T = 4$ for all convolutional layers. With this TCN, the RFS is 19, i.e., the TCN can process up to 19 elements in a sequence, as shown in Fig. 6. Therefore, the temporal sequence entered in the TC block should be less than or equal to 19 to allow TCN to process all temporal information without loss. For sequences that are longer than RFS, they can be split into multiple windows each with a length less than RFS. Each window is then entered separately into the TC block. In this article, we fixed RFS to 19 and changed the length of windows entering the TC block (T_w), as defined in (3).

Fig. 6 shows a sequence of 16 temporal elements (T_w= 16) entering the TCN. Each element is a vector of size F_2 (#filters in CV block). The output of the TCN is the last element in the sequence, which is a vector of size F_T (# filters in TCN). In this article, $F_2 = F_T = 32$. The outputs of the TC block from all windows are concatenated and then fed to an FC layer with four neurons, as the number of MI classes, followed by a SoftMax classifier, as shown in Fig. 1.

Unless otherwise noted, hyperparameters used for all experiments in this article are shown in Table I. These parameters were set empirically based on several experiments and were fixed for all subjects.

TABLE I
HYPERPARAMETER SETTING THAT USED FOR ALL SUBJECTS

Attention (AT) block		Convolutional (CV) block	
# of attention heads (H)	2	# Temporal filters (F_1)	16
Head size (d_H)	8	Kernel size (K_c)	64
Dropout rate (p_a)	0.5	Depth multiplier (D)	2
		2nd pooling size (P_2)	7
		Dropout rate (p_c)	0.3
Temporal Convolutional (TC) block		# of windows (n)	5
# of residual blocks (L)	2		
Kernel size (K_T)	4		
# Filters (F_T)	32		
Dropout rate (p_t)	0.3		

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Selected Dataset and Evaluation Approaches

BCI Competition IV-2a (BCI-2a) dataset [30] is used to train and evaluate the proposed model. BCI-2a is a well-known public MI-EEG dataset created by Graz University of Technology in 2008. BCI-2a has been widely used in the research community and is thus considered a benchmark dataset in MI-EEG decoding. It contains a limited number of samples captured under uncontrolled conditions with a considerable amount of artifacts, which makes decoding MI tasks using this dataset a challenging process.

BCI-2a dataset consists of 5184 trials (samples) of MI-EEG data recorded using 22 EEG electrodes from nine subjects (576 trials per subject). MI trials last 4 s and were sampled at 250 Hz and filtered between 0.5 and 100 Hz. Each trial belongs to one of four MI tasks: imagining of movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). For each subject, two sessions were recorded on different days. Each session consists of 288 trials per subject. One of these sessions is used to train the model and the other for evaluation.

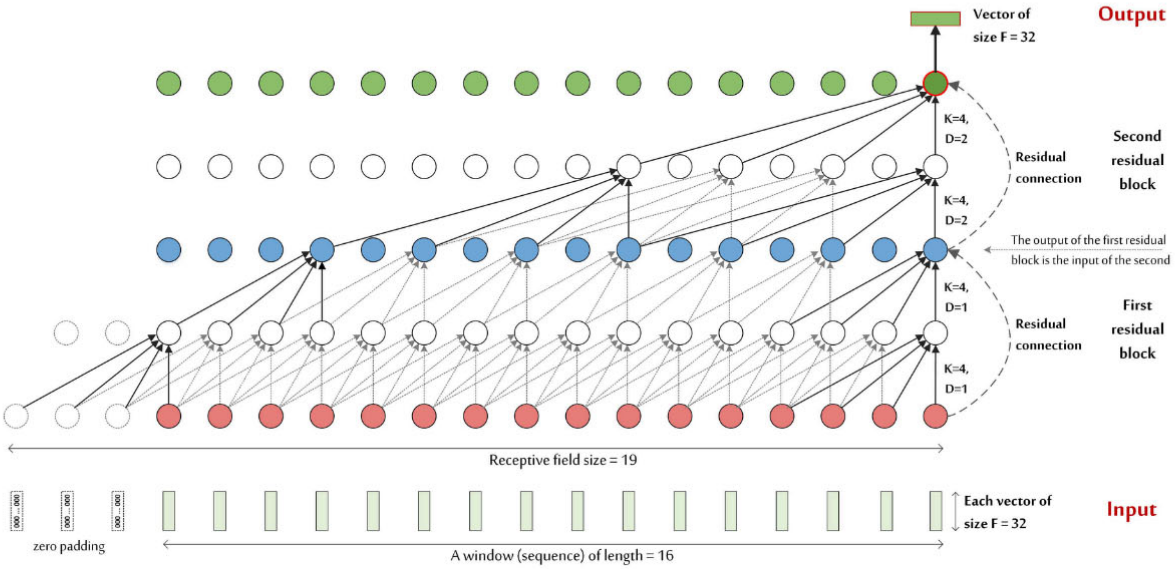


Fig. 6. Visualize a TCN block with a depth of 2, i.e., consisting of two residual blocks, kernel size = 4, and number of filters = 32. The output of the first residual block is the input of the second. The receptive field size for this TCN will be 19, so the input length Tshould be ⩽ 19. This figure shows a sequence of 16 temporal elements (T= 16) entering the TCN.

在深度网络中，由于其学习恒等函数的能力，它被广泛使用。在残差块中，由于输入和输出维度相同（32），我们使用恒等映射；否则，我们使用线性变换，即 1×1 卷积。

TCN 的感受野大小 (RFS) 随着堆叠残差块数量 L 的增加而呈指数增长，这是由于每个后续块的扩张量 D 呈指数增长。RFS 由两个参数控制：残差块数量 L 和卷积核大小 K，定义见

$$\text{RFS} = 1 + 2\,(K-1)\left(2^{L}-1\right). \tag{13}$$

在所提出的 ATCNet 中，TC 块由一个带有残差块的 TCN 和所有卷积层大小为 K= 4 的 32 个滤波器组成。使用此 TCN，RFS 为 19，即 TCN 可以处理序列中最多 19 个元素，如图 6 所示。因此，输入 TC 块的时间序列应小于或等于 19，以允许 TCN 处理所有时间信息而不会造成丢失。对于长度大于 RFS 的序列，可以将其拆分成多个长度小于 RFS 的窗口。然后将每个窗口分别输入到 TC 块中。在本文中，我们将 RFS 固定为 19，并更改了进入 TC 块的窗口长度 (T)，如 (3) 中所定义。

图 6 显示了进入 TCN 的 16 个时间元素序列 (T= 16)。每个元素都是一个大小为 F（CV 块中的滤波器数量）的向量。TCN 的输出是序列中的最后一个元素，它是一个大小为 F（TCN 中的滤波器数量）的向量。在本文中，F= F= 32。所有窗口的 TC 块输出被连接起来，然后馈送到具有四个神经元的 FC 层，数量与 MI 类的数量相同，然后是 SoftMax 分类器，如图 1 所示。

除非另有说明，本文所有实验所使用的超参数均如表一所示。这些参数是根据多次实验经验设定的，并且对于所有受试者都是固定的。

TABLE I

HS TUA S

Attention (AT) block		Convolutional (CV) block	
# of attention heads (H)	2	# Temporal filters (F_1)	16
Head size (d_H)	8	Kernel size (K_c)	64
Dropout rate (p_a)	0.5	Depth multiplier (D)	2
		2^{nd} pooling size (P_2)	7
		Dropout rate (p_c)	0.3
Temporal Convolutional (TC) block		# of windows (n)	5
# of residual blocks (L)	2		
Kernel size (K_T)	4		
# Filters (F_T)	32		
Dropout rate (p_t)	0.3		

III. ERD

A. Selected Dataset and Evaluation Approaches

BCI 竞赛 IV-2a (BCI-2a) 数据集 [30] 用于训练和评估所提出的模型。BCI-2a 是一个著名的公共 MI-EEG 数据集，由格拉茨技术大学于 2008 年创建。BCI-2a 在研究界得到了广泛的应用，因此被认为是 MI-EEG 解码的基准数据集。该数据集包含有限数量的样本，且是在不受控制的条件下捕获的，并且存在大量的伪影，这使得使用该数据集解码 MI 任务变得极具挑战性。

BCI-2a 数据集包含 5184 个 MI-EEG 数据试验（样本），这些数据使用来自 9 名受试者的 22 个 EEG 电极记录（每位受试者 576 个试验）。MI 试验持续 4 秒，采样率为 250 Hz，滤波范围为 0.5 至 100 Hz。每个试验属于四个 MI 任务之一：想象左手（第 1 类）、右手（第 2 类）、双脚（第 3 类）和舌头（第 4 类）的运动。每位受试者在不同日期记录两个会话。每个会话包含每位受试者 288 个试验。其中一个会话用于训练模型，另一个会话用于评估。

The proposed model is evaluated using subject-dependent (subject-specific) and subject-independent approaches. For subject-dependent, we used the same training and testing data as the original competition, i.e., 288×9 trials in session 1 for training, and 288×9 trials in session 2 for testing. For subject independent, we used cross-subject evaluation, known as "leaving one subject out" (LOSO). In LOSO, the model is trained and evaluated by several folds, equal to the number of subjects, and for each fold, one subject is used for evaluation and the others for training. The LOSO evaluation technique ensures that separate subjects (not visible in the training data) are used to evaluate the model.

B. Performance Metrics

The proposed models in this article are evaluated using accuracy (14), and Kappa score (15)

$$\text{ACC} = \frac{\sum_{i=1}^{n} \text{TP}_i / I_i}{n} \tag{14}$$

where TP_i is the true positive, i.e., the number of correctly predicted samples in class i, l_i is the number of samples in class i, and n indicates the number of classes

$$\kappa\_\text{score} = \frac{1}{n} \sum_{a=1}^{n} \frac{P_a - P_e}{1 - P_e} \tag{15}$$

where n is the number of classes, P_a is the actual percentage of agreement, and P_e is the expected percentage chance of agreement.

C. Training Procedure

The models were trained and tested by a single GPU, Nvidia GTX 2070 8 GB, using the TensorFlow framework. For all experiments, we used the following training configurations. Glorot uniform initializer is used to initialize the weights. The models are trained using the Adam optimizer with a learning rate of 0.0009, batch size of 64, and a categorical cross-entropy loss over 1000 epochs with a patience of 300. These hyperparameters were determined through several experiments to help the models generalize well.

The proposed ATCNet model achieves an overall accuracy of 85.38% and a κ-score of 0.81, which is better than the state-of-the-art results.

D. Contributions of ATCNet Blocks

In this section, we perform an ablation analysis to measure the effectiveness of each block in the ATCNet model. Table II presents the impact of removing one or more blocks in the ATCNet model on the performance of MI classification using the BCI-2a dataset. Blocks were removed before training and validation operations. The results showed that the AT block increased the overall accuracy by 1.54% and SW by 2.28%. The addition of the TC block also increased accuracy by 1.04% compared to using the CV block only. The results showed that each block adds its contribution regardless of the other blocks except for the attention block. Attention block improves accuracy if followed by TC block. If the TC block is removed, the accuracy drops to 79.44%, which is lower than the accuracy after removing both AT and TC blocks, 82.60%, and even removing all blocks, 81.71%. This means that placing the attention layer at the end of the model harms the performance unless followed by an additional classification layer.

TABLE II
CONTRIBUTION OF EACH BLOCK IN THE ATCNET MODEL TO THE PERFORMANCE OF MI CLASSIFICATION USING THE BCI-2A DATASET. AT: ATTENTION, SW: SLIDING WINDOW, TC: TEMPORAL CONVOLUTION

Removed block	Accuracy %	κ-score
None (ATCNet)	85.38	0.805
AT	83.84	0.784
SW	83.10	0.775
SW + AT	82.75	0.770
TC	79.44	0.726
SW + TC	80.48	0.740
AT + TC	82.60	0.768
SW + AT + TC	81.71	0.756

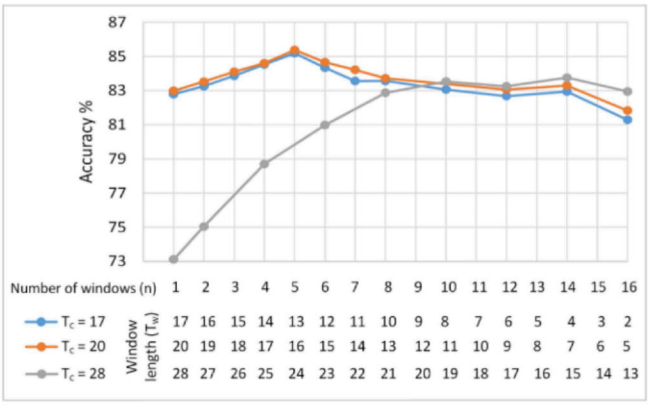


Fig. 7. Accuracy on BCI-2a as a function of the number of windows using three temporal sequences of length 17, 20, and 28. These sequences were studied while varying the number of windows from one window, that is, the whole sequence, to 16 windows. Each window has a different length (T_w) depending on the length of the original sequence (T_c). The 20-sample sequence showed better performance than the others and the best performance was using five windows.

E. Varying the Temporal Sequence Length

In the following experiments, we investigate the effect of changing the length of the temporal sequence (T_c) produced by the CV block as well as the number of windows n. The sequence length is controlled by the size of the second pooling layer (P_2) in the CV block, as defined in (1). Fig. 7 shows the accuracy of MI classification using three temporal sequences of lengths 17, 20, and 28, which encode the original MI-EEG signal with a resolution of 64 samples (256 ms), 56 samples (225 ms), and 40 samples (160 ms), respectively. Each sequence is studied while increasing the number of windows from 1 to 16. ATCNet works best when the window length is less than RFS (19). Using a longer window than RFS significantly reduces accuracy due to information loss, as shown in the 28-sample

所提出的模型使用受试者相关（受试者特定）和受试者无关的方法进行评估。对于受试者相关方法，我们使用与原始竞赛相同的训练和测试数据，即在第 1 阶段进行 288×9 次试验用于训练，在第 2 阶段进行 288×9 次试验用于测试。对于受试者无关方法，我们使用跨受试者评估，也称为“留一受试者”（LOSO）。在 LOSO 中，模型通过与受试者数量相等的多次迭代进行训练和评估，每次迭代使用一个受试者进行评估，其他受试者用于训练。LOSO 评估技术确保使用单独的受试者（在训练数据中不可见）来评估模型。

B. Performance Metrics

The proposed models in this article are evaluated using accuracy (14), and Kappa score (15)

$$ACC = \frac{\sum_{i=1} TP/I}{n} \quad (14)$$

其中 TP 为真阳性，即第 i 类中正确预测的样本数，l 为第 i 类的样本数，n 为类数

$$\kappa\_score = \frac{1}{n} \sum_{a=1} \frac{P - P}{1 - P} \quad (15)$$

其中 n 是类别数，P 是实际一致百分比，P 是预期一致百分比概率。

C. Training Procedure

这些模型使用 TensorFlow 框架，在单个 GPU（Nvidia GTX 2070 8 GB）上进行训练和测试。所有实验均采用以下训练配置。使用 Glorot 均匀初始化器初始化权重。模型使用 Adam 优化器进行训练，学习率为 0.0009，批次大小为 64，分类交叉熵损失函数超过 1000 个 epoch，耐心值为 300。这些超参数是通过多次实验确定的，旨在帮助模型实现良好的泛化。
所提出的 ATCNet 模型实现了 85.38% 的总体准确率和 0.81 的 κ 分数，优于最先进的结果。

D. Contributions of ATCNet Blocks

在本节中，我们进行消融分析以衡量 ATCNet 模型中每个块的有效性。表 II 显示了在 ATCNet 模型中删除一个或多个块对使用 BCI-2a 数据集进行 MI 分类性能的影响。在训练和验证操作之前删除了块。结果表明，AT 块使整体准确率提高了 1.54%，SW 提高了 2.28%。与仅使用 CV 块相比，添加 TC 块也使准确率提高了 1.04%。结果表明，除了注意力块之外，每个块都会增加其贡献，而不管其他块如何。注意力块提高了

TABLE II

CE B ATCNM
P MI CUBCI-2D. AT:
A, SW: S W , TC: TC

Removed block	Accuracy %	κ-score
None (ATCNet)	85.38	0.805
AT	83.84	0.784
SW	83.10	0.775
SW + AT	82.75	0.770
TC	79.44	0.726
SW + TC	80.48	0.740
AT + TC	82.60	0.768
SW + AT + TC	81.71	0.756

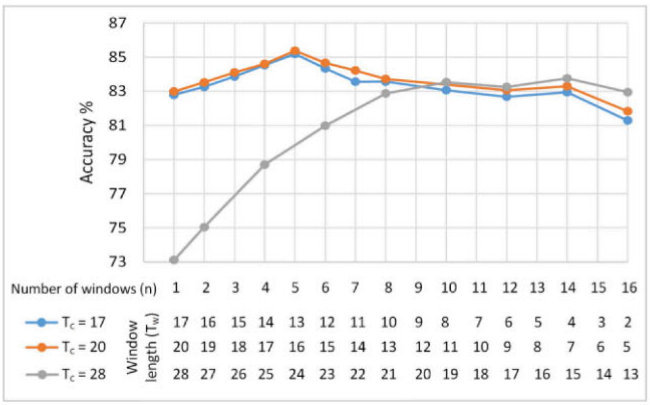


图 7. 使用长度分别为 17、20 和 28 的三个时间序列，BCI-2a 上的准确率与窗口数量的关系。研究这些序列时，窗口数量从一个窗口（即整个序列）变化到 16 个窗口。每个窗口的长度 (T) 取决于原始序列的长度 (T)。20 个样本的序列表现出优于其他序列的性能，其中使用五个窗口时性能最佳。

如果后接 TC 块，准确率会有所提升。如果移除 TC 块，准确率会下降到 79.44%，低于同时移除 AT 和 TC 块后的准确率 82.60%，甚至低于移除所有块后的准确率 81.71%。这意味着将注意力层放在模型的末尾会损害性能，除非在其后附加一个分类层。

E. Varying the Temporal Sequence Length

在接下来的实验中，我们研究了改变 CV 块产生的时间序列 (T) 的长度以及窗口数量 n 的影响。序列长度由 CV 块中第二个池化层 (P) 的大小控制，如 (1) 中定义。图 7 显示了使用长度分别为 17、20 和 28 的三个时间序列进行 MI 分类的准确率，这三个时间序列分别以 64 个样本（256 毫秒）、56 个样本（225 毫秒）和 40 个样本（160 毫秒）的分辨率对原始 MI-EEG 信号进行编码。在将窗口数量从 1 增加到 16 的同时对每个序列进行研究。当窗口长度小于 RFS (19) 时，ATCNet 效果最佳。使用比 RFS 更长的窗口会因信息丢失而显著降低准确率，如 28 个样本的

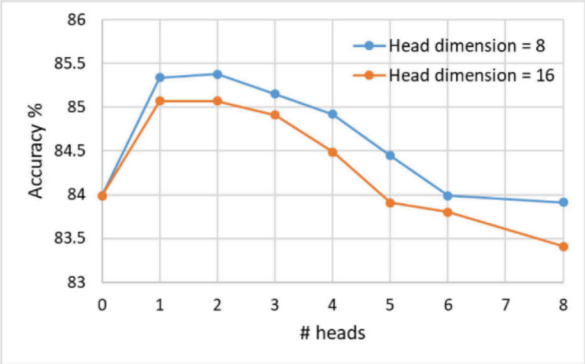


Fig. 8. Accuracy on BCI-a2 as a function of the number of attention heads using head sizes of 8 and 16. Reducing head size as well as the number of heads showed better performance due to the small size of the dataset and its light representation. The best performance was using two 8-size heads.

TABLE III
ATCNET MODEL PERFORMANCE USING DIFFERENT ATTENTION SCHEMES: MSA WITH 8 AND 16 HEAD SIZE, SE, AND CBAM

Attention mechanism	Accuracy %	κ-score
No Attention	83.84	0.784
MSA-8	**85.38**	**0.805**
MSA-16	85.07	0.801
SE	84.07	0.788
CBAM	84.30	0.791

The bold font highlights the best results among the different methods.

sequence (while $T_W = 20$–28). In general, increasing the number of windows improved classification accuracy as this helps to augment the data and helps the model learn the changing MI information from different time positions. However, this increase in accuracy reaches a point where the window contains a narrow signal that may not contain enough MI information to train the model. For example, by dividing the 20-sample sequence into 12 windows with a stride of one, each window will have six samples corresponding to 384 samples (∼1.5 s) in the original signal with a stride of 64. This justifies the decrease in accuracy for sequences of lengths 17 and 20 starting in six windows. The 20-sample sequence performed better than the other sequences in many windows and the best performance was achieved using five windows (each window of length 16). This indicates that encoding the original MI-EEG signal at a resolution of 56 samples (225 ms) provides a good representation compared to a lower (e.g., 160 ms) or higher (e.g., 256 ms) resolution.

F. Comparing Different Attention Schemes

Fig. 8 compares the performance of the MSA block with a different number of heads using dimension sizes of 8 and 16, i.e., the size of each attention head for query/key/value vectors. The results showed that using two heads each of size 8 gave the best results. This is because the MI-EEG dataset has a limited number of samples, which requires a light MSA layer to converge well. In addition, the temporal data entered into the MSA layer has a light representation, i.e., sequence length = 16 and embedding size = 32, which requires few parameters to train.

In Table III, we compare the performance of the proposed model using three attention mechanisms: MSA [27], SE [28], and CBAM [29]. The number of MSA heads was set to 2 and the head size was set to 8 and 16. The reduction ratio for both SE and CBAM was experimentally set to 8. The results showed that all attention mechanisms improved the performance of the ATCNet model while the best performance was achieved by MSA, indicating that MSA is more suitable for a 2-D EEG representation.

TABLE IV
PERFORMANCE (ACCURACY (%) AND κ-SCORE) COMPARISON OF SUBJECT-SPECIFIC CLASSIFICATION USING BCI-2A DATASET FOR THE PROPOSED MODEL WITH OTHER REPRODUCED MODELS

	Proposed (ATCNet)		EEGNet [12]		EEG-TCNet [22]		TCNet Fusion [23]	
Sub.	%	κ	%	κ	%	κ	%	κ
1	88.5	0.85	88.5	0.85	84.0	0.79	86.1	0.81
2	70.5	0.61	66.0	0.55	66.3	0.55	66.0	0.55
3	97.6	0.97	95.1	0.94	94.1	0.92	93.4	0.91
4	81.0	0.75	73.6	0.65	72.6	0.63	72.6	0.63
5	83.0	0.77	75.4	0.67	76.0	0.68	79.9	0.73
6	73.6	0.65	64.2	0.52	62.9	0.50	66.7	0.56
7	93.1	0.91	90.3	0.87	89.9	0.87	90.3	0.87
8	90.3	0.87	85.8	0.81	84.7	0.80	85.8	0.81
9	91.0	0.88	86.5	0.82	85.4	0.81	85.4	0.81
Mean	**85.4**	**0.81**	80.6	0.74	79.6	0.73	80.7	0.74
St.D.	**9.1**	**0.12**	11.1	0.15	10.7	0.14	10.1	0.13

The bold font highlights the best results among the different methods.

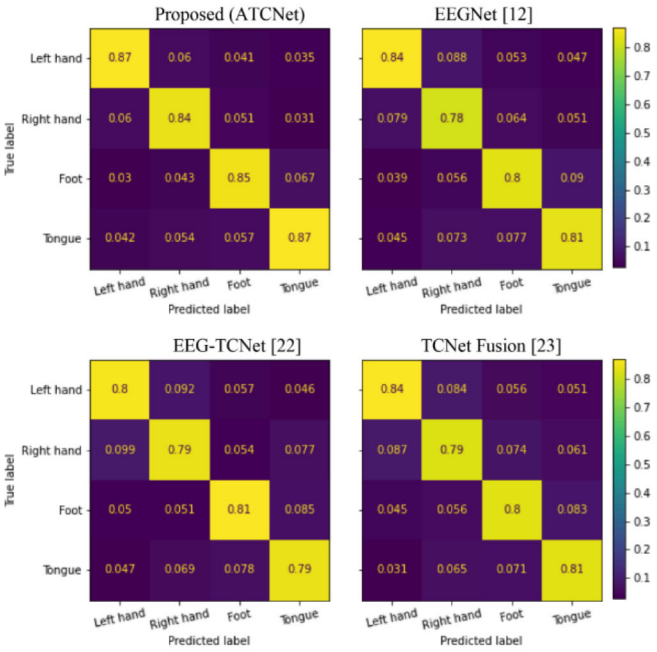


Fig. 9. Average confusion matrices of the proposed ATCNet and the reproduced EEGNet, EEG-TCNet, and TCNet_Fusion models. The results showed that ATCNet improved MI decoding for all MI tasks compared to equivalent models.

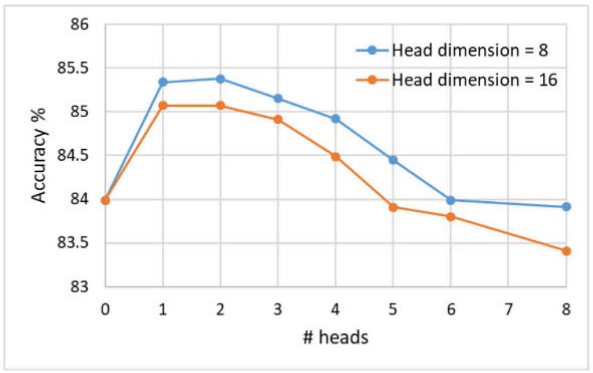


图 8. 使用 8 和 16 尺寸的头部，BCI-a2 上的准确率与注意力头数量的关系。由于数据集规模较小且表示形式简洁，减少头部尺寸和头部数量可获得更好的性能。使用两个 8 尺寸的头部时，性能最佳。

TABLE III
ATCNM P UDAS :
MSA W 8 16 HS , SE, CBAM

Attention mechanism	Accuracy %	κ-score
No Attention	83.84	0.784
MSA-8	**85.38**	**0.805**
MSA-16	85.07	0.801
SE	84.07	0.788
CBAM	84.30	0.791

The bold font highlights the best results among the different methods.

序列（T= 20–28 时）。一般来说，增加窗口数量可以提高分类准确率，因为这有助于扩充数据并帮助模型学习来自不同时间位置的变化的 MI 信息。然而，这种准确率的提高会达到一个程度，即窗口包含的信号很窄，可能不包含足够的 MI 信息来训练模型。例如，通过将 20 个样本序列分成 12 个窗口，步幅为 1，每个窗口将有 6 个样本，对应于原始信号中的 384 个样本（约 1.5 秒），步幅为 64。这证明了从六个窗口开始长度为 17 和 20 的序列准确率的下降是合理的。20 个样本序列在许多窗口中的表现都优于其他序列，使用五个窗口（每个窗口长度为 16）可获得最佳性能。这表明，以 56 个样本（225 毫秒）的分辨率对原始 MI-EEG 信号进行编码，与较低（例如，160 毫秒）或较高（例如，256 毫秒）的分辨率相比，可以提供良好的表示。

F. Comparing Different Attention Schemes

图 8 比较了不同注意头数量（维度大小分别为 8 和 16，即每个注意头用于查询/键/值向量的大小）的 MSA 块的性能。结果表明，使用两个大小均为 8 的注意头可获得最佳结果。这是因为 MI-EEG 数据集的样本数量有限，因此需要轻量级的 MSA 层才能很好地收敛。此外，输入 MSA 层的时间数据具有

TABLE IV
P (A(%) κ-S) C
S -S CUBCI-2D
P M W ORM

	Proposed (ATCNet)		EEGNet [12]		EEG-TCNet [22]		TCNet Fusion [23]	
Sub.	%	κ	%	κ	%	κ	%	κ
1	88.5	0.85	88.5	0.85	84.0	0.79	86.1	0.81
2	70.5	0.61	66.0	0.55	66.3	0.55	66.0	0.55
3	97.6	0.97	95.1	0.94	94.1	0.92	93.4	0.91
4	81.0	0.75	73.6	0.65	72.6	0.63	72.6	0.63
5	83.0	0.77	75.4	0.67	76.0	0.68	79.9	0.73
6	73.6	0.65	64.2	0.52	62.9	0.50	66.7	0.56
7	93.1	0.91	90.3	0.87	89.9	0.87	90.3	0.87
8	90.3	0.87	85.8	0.81	84.7	0.80	85.8	0.81
9	91.0	0.88	86.5	0.82	85.4	0.81	85.4	0.81
Mean	**85.4**	**0.81**	80.6	0.74	79.6	0.73	80.7	0.74
St.D.	**9.1**	**0.12**	11.1	0.15	10.7	0.14	10.1	0.13

The bold font highlights the best results among the different methods.

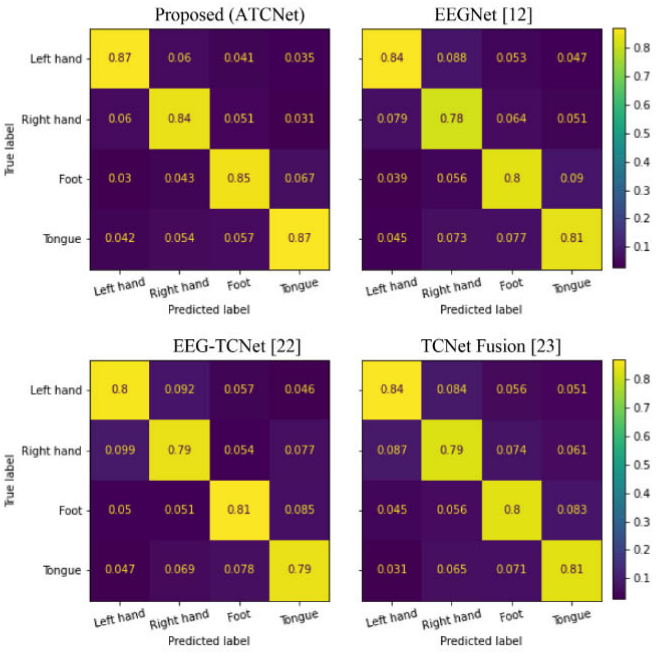


Fig. 9. Average confusion matrices of the proposed ATCNet and the reproduced EEGNet, EEG-TCNet, and TCNet_Fusion models. The results showed that ATCNet improved MI decoding for all MI tasks compared to equivalent models.

轻量级表示，即序列长度 = 16 和嵌入大小 = 32，需要很少的参数来训练。

在表三中，我们比较了使用三种注意力机制（MSA[27]、SE[28]和 CBAM[29]）对所提模型的性能的影响。MSA 的头数量设置为 2，头大小分别设置为 8 和 16。SE 和 CBAM 的缩减比例均通过实验设置为 8。结果表明，所有注意力机制均能提升 ATCNet 模型的性能，但 MSA 的性能最佳，这表明 MSA 更适合二维脑电图表示。

TABLE V
SUBJECT-SPECIFIC PERFORMANCE ON THE BCI-2A DATASET USING THE SAME ORIGINAL COMPETITION DIVISION (HOLD-OUT APPROACH: 50% TRAINING TRIALS AND 50% TEST TRIALS). ACCURACY (%) AND κ-SCORE ARE THE AVERAGES FOR ALL SUBJECTS

Method	Accuracy	κ-score
Shallow CNN [32]	74.31	0.66
EEGNet: CNN [12]*	80.59	0.74
DBN-AE [17]	71.0	_
Multi-layer-CNN and MLP [18]	75.0	_
EEG-TCNet: CNN and TCN [22]*	79.55	0.73
Attention multi-scale CNN [13]	79.9	_
TCNet_Fusion: multi-layer CNN + TCN [23]*	80.67	0.74
Attention-inception CNN & LSTM [10]	82.84	_
Attention multi-branch CNN [9]	82.87	0.772
ATCNet: Attention-CNN and TCN (Proposed)	**85.38**	**0.805**

* *Reproduced.*
The bold font highlights the best results among the different methods.

TABLE VI
SUBJECT-INDEPENDENT PERFORMANCE ON THE BCI-2A DATASET USING LOSO CROSS VALIDATION. ACCURACY (%) AND κ-SCORE ARE THE AVERAGES FOR ALL SUBJECTS

Method	Accuracy	κ-score
Attention graph convolutional network [8]	60.1	-
Multi-layer-CNN and AE [18]	55.3	-
EEGNet: CNN [12]*	68.79	0.584
Attention multi-branch CNN [9]	69.10	-
EEG-TCNet: CNN and TCN [22]*	69.52	0.594
TCNet_Fusion: multi-layer CNN + TCN [23]*	70.58	0.608
ATCNet: Attention-CNN and TCN (Proposed)	**70.97**	**0.613**

* *Reproduced.*
The bold font highlights the best results among the different methods.

G. Comparison to Recent Studies

Table IV summarizes the accuracy and κ-score of the proposed ATCNet model using the BCI-2a dataset and its comparison with the reproduced EEGNet [12], EEG-TCNet [22], and TCNet_Fusion [23], as these models have some similarities with the proposed model. The results of the reproduced models are based on the hyperparameters identified in the original articles, while preprocessing, training, and evaluation followed the same procedure defined in this article. Table IV shows that ATCNet performed better than EEGNet, EEG-TCNet, and TCNet_Fusion for all subjects with an average accuracy of 85.38% and a κ-score of 0.81. This represents a 4.71% increase in accuracy over these models. In addition, the proposed model achieved the best standard deviation among subjects with a value of 9.08%, indicating that the accuracy is more robust over all subjects. The average confusion matrices of ATCNet and the reproduced models are shown in Fig. 9. ATCNet demonstrated an improvement in MI decoding for all MI classes compared to the other models.

Table V presents the reported overall accuracy and κ-score of recent studies in the subject-specific MI-EEG classification using the BCI-2a dataset. The proposed ATCNet model performs better than the recent studies using raw EEG data and without pre-processing. In addition to the subject-specific (subject-dependent) results, we evaluated the performance of the proposed model in subject-independent classification, which is a measure of the model's generalization ability. The proposed model achieved the best subject-independent performance on the BCI-2a dataset, as shown in Table VI.

IV. CONCLUSION

This article proposed a novel ATCNet for EEG-based motor imagery classification. ATCNet consists of three main blocks: the CV block, to encode the raw MI-EEG signal into a compact temporal sequence, the multihead self-AT block, to highlight the most effective information in the temporal sequence, and the TC block, to extract high-level temporal features from the temporal sequence. This article also implemented a convolutional-based SW combined with CV block to improve the performance of MI classification efficiently by parallelizing the process. The ablation analysis showed that each block in the ATCNet model made a significant contribution to the performance of the ATCNet model. The AT block increased overall accuracy by 1.54%, the SW by 2.28%, and the TC by 1.04% compared to using the CV block only. The proposed ATCNet model outperformed recent techniques in MI-EEG classification using the BCI-2a dataset with an accuracy of 85.38% and 70.97% for the subject-dependent and subject-independent modes, respectively. These high results came with a relatively small number of parameters (115.2K), which makes ATCNet applicable to industrial devices with limited resources. The proposed model demonstrated a powerful ability to extract MI features from a raw EEG signal without artifact removal and with minimal preprocessing using a limited-size and challenging dataset. ATCNet showed an overall improvement in EEG decoding for all MI classes and all subjects in the BCI-2a dataset proving that ATCNet can learn to find generic EEG representations across classes and subjects.

For future work, the proposed model can be further improved by using attention mechanisms in several domains. Effective MI information occurs in the EEG data at specific time intervals, channel locations, and frequency bands; Thus, developing a DL model that automatically attends to the most important information in these domains is a promising direction for improving the performance of MI-EEG classification. The proposed model can also be refined using preprocessing methods to remove artifacts and deep generative models to increase the size of the dataset.

REFERENCES

[1] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where," *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5031–5042, Aug. 2022.
[2] H. Altaheri et al., "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, pp. 1–42, 2021.
[3] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Front. Neurosci.*, vol. 6, 2012, Art. no. 56013.
[4] M. S. Hossain, M. Al-Hammadi, and G. Muhammad, "Automatic fruit classification using deep learning for industrial applications," *IEEE Trans. Ind. Inform.*, vol. 15, no. 2, pp. 1027–1034, Feb. 2019.
[5] H. Altaheri, M. Alsulaiman, and G. Muhammad, "Date fruit classification for robotic harvesting in a natural environment using deep learning," *IEEE Access*, vol. 7, pp. 117115–117133, 2019.

TABLE V
S -S P ERFORMANCE ON THE BCI-2DU
S O CD(H-O A : 50% TT50% TT). A(%) κ-S
ARE THE AA S

Method	Accuracy	κ-score
Shallow CNN [32]	74.31	0.66
EEGNet: CNN [12]*	80.59	0.74
DBN-AE [17]	71.0	_
Multi-layer-CNN and MLP [18]	75.0	_
EEG-TCNet: CNN and TCN [22]*	79.55	0.73
Attention multi-scale CNN [13]	79.9	_
TCNet_Fusion: multi-layer CNN + TCN [23]*	80.67	0.74
Attention-inception CNN & LSTM [10]	82.84	_
Attention multi-branch CNN [9]	82.87	0.772
ATCNet: Attention-CNN and TCN (Proposed)	**85.38**	**0.805**

* *Reproduced.*
The bold font highlights the best results among the different methods.

TABLE VI
S -IP ERFORMANCE ON THE BCI-2DU
LOSO CV. A(%) κ-S
AA S

Method	Accuracy	κ-score
Attention graph convolutional network [8]	60.1	-
Multi-layer-CNN and AE [18]	55.3	-
EEGNet: CNN [12]*	68.79	0.584
Attention multi-branch CNN [9]	69.10	-
EEG-TCNet: CNN and TCN [22]*	69.52	0.594
TCNet_Fusion: multi-layer CNN + TCN [23]*	70.58	0.608
ATCNet: Attention-CNN and TCN (Proposed)	**70.97**	**0.613**

* *Reproduced.*
The bold font highlights the best results among the different methods.

G. Comparison to Recent Studies

表 IV 总结了使用 BCI-2a 数据集的所提出的 ATCNet 模型的准确度和 κ 分数，并将其与复现的 EEGNet [12]、EEG-TCNet [22] 和 TCNet_Fusion [23] 进行了比较，因为这些模型与所提出的模型有一些相似之处。复现模型的结果基于原始文章中确定的超参数，而预处理、训练和评估遵循与本文定义的相同程序。表 IV 显示，对于所有受试者，ATCNet 的表现都优于 EEGNet、EEG-TCNet 和 TCNet_Fusion，平均准确度为 85.38%，κ 分数为 0.81。这意味着准确度比这些模型提高了 4.71%。此外，所提出的模型在受试者中取得了最佳标准差，值为 9.08%，表明准确度对所有受试者都更为稳健。 ATCNet 和重现模型的平均混淆矩阵如图 9 所示。与其他模型相比，ATCNet 在所有 MI 类的 MI 解码方面都得到了改进。

表五展示了近期研究使用 BCI-2a 数据集进行特定受试者 MI-EEG 分类时报告的总体准确率和 κ 值。提出的 ATCNet 模型的表现优于近期使用原始 EEG 数据且未经预处理的研究。

除了特定于受试者（与受试者相关）的结果外，我们还评估了所提模型在与受试者无关的分类中的表现，这是衡量模型泛化能力的指标。所提模型在 BCI-2a 数据集上取得了最佳的与受试者无关的性能，如表 VI 所示。

IV. C

本文提出了一种用于基于脑电图 (EEG) 运动想象分类的新型 ATCNet 网络。ATCNet 由三个主要模块组成：CV 模块，用于将原始 MI-EEG 信号编码为紧凑的时间序列；多头自 AT 模块，用于突出时间序列中最有效的信息；以及 TC 模块，用于从时间序列中提取高级时间特征。本文还实现了基于卷积的 SW 与 CV 模块相结合，通过并行化流程有效提升 MI 分类的性能。消融分析表明，ATCNet 模型中的每个模块都对 ATCNet 模型的性能做出了显著贡献。与仅使用 CV 模块相比，AT 模块使总体准确率提高了 1.54%，SW 提高了 2.28%，TC 提高了 1.04%。所提出的 ATCNet 模型在使用 BCI-2a 数据集的 MI-EEG 分类中优于近期技术，在受试者相关和非受试者模式下的准确率分别为 85.38% 和 70.97%。这些优异的成果源于相对较少的参数数量（115.2K），这使得 ATCNet 适用于资源有限的工业设备。所提出的模型展示了强大的能力，能够在不去除伪影的情况下，以极少的预处理，使用规模有限且具有挑战性的数据集，从原始脑电图信号中提取 MI 特征。ATCNet 在 BCI-2a 数据集中对所有 MI 类别和所有受试者的脑电图解码能力均有整体提升，这证明了 ATCNet 能够学习找到跨类别和受试者的通用脑电图表征。

未来，可以通过在多个领域使用注意力机制来进一步改进所提出的模型。有效的 MI 信息出现在 EEG 数据中的特定时间间隔、通道位置和频段；因此，开发一个能够自动关注这些领域中最重要的信息的深度学习模型，是提升 MI-EEG 分类性能的一个有前景的方向。此外，还可以使用预处理方法去除伪影，并使用深度生成模型来改进所提出的模型，以增加数据集的大小。

R

[1] I. Ahmed、G. Jeon 和 F. Piccialli，“从人工智能到工业 4.0 中的可解释人工智能：关于什么、如何和在哪里的调查”，IEEE Trans. Ind. Inform.，第 18 卷，第 8 期，第 5031-5042 页，2022 年 8 月。[2] H. Altaheri 等人，“用于脑电图 (EEG) 运动想象 (MI) 信号分类的深度学习技术：综述”，Neural Comput. Appl.，第 1-42 页，2021 年。[3] KK Ang、ZY Chin、C. Wang、C. Guan 和 H. Zhang，“BCI 竞赛 IV 数据集 2a 和 2b 上的滤波器组常见空间模式算法”，Front. Neurosci.，第 6 卷，2012 年，Art. no. 56013。[4] MS Hossain、M. Al-Hammadi 和 G. Muhammad，“使用深度学习进行工业应用的水果自动分类”，IEEE Trans.

Ind. Inform., vol. 15, no. 2, pp. 1027–1034, Feb. 2019.
[5] H. Altaheri, M. Alsulaiman, and G. Muhammad, “Date fruit classification for robotic harvesting in a natural environment using deep learning,” IEEE Access, vol. 7, pp. 117115–117133, 2019.

[6] I. Ahmed, S. Din, G. Jeon, and F. Piccialli, "Exploring deep learning models for overhead view multiple object detection," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5737–5744, Jul. 2020.

[7] M. Qamhan, H. Altaheri, A. H. Meftah, G. Muhammad, and Y. A. Alotaibi, "Digital audio forensics: Microphone and environment classification using deep learning," *IEEE Access*, vol. 9, pp. 62719–62733, 2021.

[8] D. Zhang, K. Chen, D. Jian, and L. Yao, "Motor imagery classification via temporal attention cues of graph embedded EEG signals," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 9, pp. 2570–2579, Sep. 2020.

[9] G. A. Altuwaijri, G. Muhammad, H. Altaheri, and M. Alsulaiman, "A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for EEG-based motor imagery signals classification," *Diagnostics*, vol. 12, no. 4, 2022, Art. no. 995.

[10] S. U. Amin, H. Altaheri, G. Muhammad, W. Abdul, and M. Alsulaiman, "Attention-inception and long short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation," *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5412–5421, Aug. 2022.

[11] S. U. Amin, H. Altaheri, G. Muhammad, M. Alsulaiman, and W. Abdul, "Attention based inception model for robust EEG motor imagery classification," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2021, pp. 1–6.

[12] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 56013.

[13] D. Li, J. Xu, J. Wang, X. Fang, and Y. Ji, "A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2615–2626, Dec. 2020.

[14] S. Kumar, R. Sharma, and A. Sharma, "OPTICAL+: A frequency-based deep learning scheme for recognizing brain wave signals," *Peer J. Comput. Sci.*, vol. 7, 2021, Art. no. e375.

[15] T. Luo and F. Chao, "Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network," *BMC Bioinf.*, vol. 19, no. 1, 2018, Art. no. 344.

[16] J. Xu, H. Zheng, J. Wang, D. Li, and X. Fang, "Recognition of EEG signal motor imagery intention based on deep multi-view feature learning," *Sensors*, vol. 20, no. 12, 2020, Art. no. 3496.

[17] A. Hassanpour, M. Moradikia, H. Adeli, S. R. Khayami, and P. Shamsinejadbabaki, "A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals," *Expert Syst.*, vol. 36, no. 6, 2019, Art. no. e12494.

[18] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Gener. Comput. Syst.*, vol. 101, pp. 542–554, 2019.

[19] M.-A. Li, J.-F. Han, and L.-J. Duan, "A novel MI-EEG imaging with the location information of electrodes," *IEEE Access*, vol. 8, pp. 3197–3211, 2020.

[20] T. Liu and D. Yang, "A densely connected multi-branch 3D convolutional neural network for motor imagery EEG decoding," *Brain Sci.*, vol. 11, no. 2, 2021, Art. no. 197.

[21] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv1803.01271*.

[22] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces," 2020, *arXiv2006.00622*.

[23] Y. K. Musallam et al., "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," *Biomed. Signal Process. Control*, vol. 69, 2021, Art. no. 102826.

[24] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific machine learning through physics-informed neural networks: Where we are and what's next," *J. Sci. Comput.*, vol. 92, 2022, Art. no. 88.

[25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv1409.0473*.

[26] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Lisbon, Portugal, Sep. 2015, pp. 1412–1421.

[27] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998–6008, 2017.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[29] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[30] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," *Inst. Knowl. Discov. Graz Univ. Technol.*, vol. 16, pp. 1–6, 2008.

[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[32] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[33] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv1607.06450*.

Hamdi Altaheri (Member, IEEE) received the master's degree in computer engineering in 2019 from King Saud University, Riyadh, Saudi Arabia, where he is currently working toward the Ph.D. degree in decoding biomedical signals with the Department of Computer Engineering, College of Computer and Information Sciences.

His research interests include computer vision, bioengineering, machine learning, and deep learning.

Ghulam Muhammad (Senior Member, IEEE) received the B.S. degree in computer science and engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 1997, and the M.S. degree in knowledge-based information engineering in 2003, and the Ph.D. degree in electrical and computer engineering from Toyohashi University and Technology, Toyohashi, Japan, in 2006.

He is a Professor with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He holds two U.S. patents. He has authored and coauthored more than 300 publications including IEEE/ACM/Springer/Elsevier journals, and flagship conference papers. His research interests include AI, machine learning, image and speech processing, and smart healthcare.

Prof. Muhammad was a recipient of the Japan Society for Promotion and Science fellowship from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Mansour Alsulaiman (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Iowa State University, Ames, IA, USA, in 1987.

He is currently a Professor with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He was the Editor-in-Chief with the King Saud University Journal Computer and Information Systems. He is the Director of Center of Smart Robotics Research, King Saud University, Riyadh, Saudi Arabia. His research interests include automatic speech/speaker recognition, automatic voice pathology assessment systems, computer-aided pronunciation training system, and robotics.

[6] I. Ahmed、S. Din、G. Jeon 和 F. Piccialli，“探索用于俯视图多物体检测的深度学习模型”，IEEE Internet Things J.，第 7 卷，第 7 期，第 5737-5744 页，2020 年 7 月。[7] M. Qamhan、H. Altaheri、AH Meftah、G. Muhammad 和 YA Alotaibi，“数字音频取证：使用深度学习进行麦克风和环境分类”，IEEE Access，第 9 卷，第 62719-62733 页，2021 年。[8] D. Zhang、K. Chen、D. Jian 和 L. Yao，“通过图形嵌入 EEG 信号的时间注意线索进行运动意象分类”，IEEE J. Biomed.

Health Inform., vol. 24, no. 9, pp. 2570–2579, Sep. 2020.

[9] G. A. Altuwaijri, G. Muhammad, H. Altaheri, and M. Alsulaiman, “A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for EEG-based motor imagery signals classification,” Diagnostics, vol. 12, no. 4, 2022, Art. no. 995.

[10] S. U. Amin, H. Altaheri, G. Muhammad, W. Abdul, and M. Alsulaiman, “基于注意力起始和长短期记忆的脑电图分类在康复运动想象任务中的应用”，IEEE Trans. Ind. Inform.，第 18 卷，第 8 期，第 5412-5421 页，2022 年 8 月。

[11] S. U. Amin, H. Altaheri, G. Muhammad, M. Alsulaiman, and W. Abdul, “Attention based inception model for robust EEG motor imagery classification,” in Proc. IEEE Int. Instrum. Meas. Technol. Conf., 2021, pp. 1–6.

[12] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces,” J. Neural Eng., vol. 15, no. 5, 2018, Art. no. 56013.

[13] D. Li, J. Xu, J. Wang, X. Fang, and Y. Ji, “A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 28, no. 12, pp. 2615–2626, Dec. 2020.

[14] S. Kumar, R. Sharma, and A. Sharma, “OPTICAL+: A frequency-based deep learning scheme for recognizing brain wave signals,” Peer J. Comput. Sci., vol. 7, 2021, Art. no. e375.

[15] T. Luo and F. Chao, “Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network,” BMC Bioinf., vol. 19, no. 1, 2018, Art. no. 344.

[16] J. Xu, H. Zheng, J. Wang, D. Li, and X. Fang, “Recognition of EEG signal motor imagery intention based on deep multi-view feature learning,” Sensors, vol. 20, no. 12, 2020, Art. no. 3496.

[17] A. Hassanpour, M. Moradikia, H. Adeli, S. R. Khayami, and P. Shamsinejadbabaki, “A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals,” Expert Syst., vol. 36, no. 6, 2019, Art. no. e12494.

[18] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, “Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion,” Future Gener. Comput. Syst., vol. 101, pp. 542–554, 2019.

[19] M.-A. Li, J.-F. Han, and L.-J. Duan, “A novel MI-EEG imaging with the location information of electrodes,” IEEE Access, vol. 8, pp. 3197–3211, 2020.

[20] T. Liu and D. Yang, “A densely connected multi-branch 3D convolutional neural network for motor imagery EEG decoding,” Brain Sci., vol. 11, no. 2, 2021, Art. no. 197.

[21] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” 2018, arXiv1803.01271.

[22] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, “EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces,” 2020, arXiv2006.00622.

[23] Y. K. Musallam et al., “Electroencephalography-based motor imagery classification using temporal convolutional network fusion,” Biomed. Signal Process. Control, vol. 69, 2021, Art. no. 102826.

[24] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, “Scientific machine learning through physics-informed neural networks: Where we are and what’s next,” J. Sci. Comput., vol. 92, 2022, Art. no. 88.

[25] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2014, arXiv1409.0473. [26] M.-T. Luong, H. Pham, and C. D. Manning, “Effective approaches to attention-based neural machine translation,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Lisbon, Portugal, Sep. 2015, pp. 1412–1421.

[27] A. Vaswani et al., “Attention is all you need,” Adv. Neural Inf. Process. Syst., vol. 30, pp. 5998–6008, 2017.

[28] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 7132–7141.

[29] S. Woo、J. Park、J.-Y. Lee 和 IS Kweon，“Cbam：卷积块注意模块”，载于 Proc. Eur. Conf. Comput. Vis.，2018 年，第 3-19 页。[30] C. Brunner、R. Leeb、G. Müller-Putz、A. Schlögl 和 G. Pfurtscheller，“2008 年 BCI 竞赛——格拉茨数据集 A”，Inst. Knowl. Discov. Graz Univ. Technol., vol. 16, pp. 1–6, 2008.

[31] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in Proc. Int. Conf. MachLearn., 2015, pp. 448–456.

[32] R. T. Schirrmeister et al., “Deep learning with convolutional neural networks for EEG decoding and visualization,” Hum. Brain Mapping, vol. 38, no. 11, pp. 5391–5420, 2017.

[33] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” 2016, arXiv1607.06450.

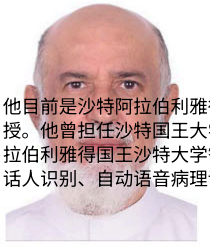
Hamdi Altaheri（IEEE 会员）于 2019 年获得沙特阿拉伯利雅得国王沙特大学计算机工程硕士学位，目前正在该大学计算机与信息科学学院计算机工程系攻读解码生物医学信号博士学位。

His research interests include computer vision, bioengineering, machine learning, and deep learning.

Ghulam Muhammad（IEEE 高级会员）于 1997 年获得孟加拉国达卡工程技术大学计算机科学与工程学士学位，2003 年获得知识型信息工程硕士学位，2006 年获得日本丰桥大学电气与计算机工程博士学位。

他是沙特阿拉伯利雅得国王沙特大学计算机与信息科学学院计算机工程系的教授。他拥有两项美国专利。他撰写或合作发表了 300 多篇出版物，包括 IEEE/ACM/Springer/Elsevier 期刊和旗舰会议论文。他的研究方向包括人工智能、机器学习、图像和语音处理以及智能医疗。

Prof. Muhammad was a recipient of the Japan Society for Promotion and Science fellowship from the Ministry of Education, Culture, Sports, Science and Technology, Japan.



Mansour Alsulaiman (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Iowa State University, Ames, IA, USA, in 1987.

他目前是沙特阿拉伯利雅得国王沙特大学计算机与信息科学学院计算机工程系教授。他曾担任沙特国王大学《计算机与信息系统》期刊的主编。他目前担任沙特阿拉伯利雅得国王沙特大学智能机器人研究中心主任。他的研究方向包括自动语音/说话人识别、自动语音病理评估系统、计算机辅助发音训练系统以及机器人技术。