New MCI detection method based on Transformer and EEG Data

1st Siwar Chaabene
MIRACL
University of Sfax
Sfax, Tunisia
siwarchaabene@gmail.com

2nd Brahim Haroun Hassan
MIRACL
University of Sfax
Sfax, Tunisia
harounhassan.brahim@isimsf.u-sfax.tn

3nd Amal Boudaya
MIRACL
University of Sfax
Sfax, Tunisia
amal.boudaya@gmail.com

4nd Lotfi Chaari
Toulouse INP
University of Toulouse
IRIT, France
lotfi.chaari@toulouse-inp.fr

5nd Bassem Bouaziz
MIRACL
University of Sfax
Sfax, Tunisia
bassem.bouaziz@isims.usf.tn

***Abstract*—Preventing health issues is a crucial aspect of the medical field, particularly when it comes to mild cognitive impairment (MCI), which is a risk factor for developing dementia. Early detection of MCI is essential, and there are two types of MCI: amnestic MCI (aMCI) and non-amnestic MCI (naMCI). However, it is challenging to differentiate between individuals with MCI and those who are aging normally. Electroencephalography (EEG) is a promising modality for diagnosing MCI, which provides information about an individual's cognitive state during a clinical examination. This research aims to distinguish between individuals diagnosed with MCI as either aMCI or naMCI, and healthy controls (HC) during a verbal fluency task (VFT). To achieve this, a new MCI detection method based on the transformer architecture was proposed. This method makes use of EEG data and achieves up to 94.78% accuracy.**

***Index Terms*—MCI detection, EEG, verbal fluency task, deep learning, transformer network**

I. Introduction

Mild cognitive impairment (MCI) [1] is a condition that causes a deterioration in cognitive ability that primarily affects the elderly. Besides, MCI is thought to be the crossover between normal aging cognitive changes and early dementia. On the other side, dementia is the ultimate stage of irreversible deterioration in cognitive ability, namely memory and other cognitive processes like language and reasoning. In this regard, Alzheimer's disease (AD) [2] is one of the most well-known and prevalent kinds of dementia. However, the degree of memory impairment distinguishes MCI from dementia. The consequences of MCI have no significant influence on patients' daily activities. Whereas dementia can cause disruptions in daily tasks. However, depending on the severity and the different areas (learning, memory, executive functions, etc.) that it may affect, MCI can be divided into two main subtypes [3]–[5] including *amnestic MCI* (aMCI) and *non-amnestic MCI* (naMCI). However, aMCI occurs when just memory is classed as impaired among the cognitive domains. On the other hand, naMCI refers to impairment in cognitive domains other than memory (language, cognition, judgment, etc.).

The diagnosis of any disease in its early stages is very important. It allows early intervention to treat, delay or reduce the impact and damage of the disease. Degenerative diseases frequently begin early, proceed gradually, and finally significantly impair cognitive functions. In the case of MCI, it should be underlined that early diagnosis should be taken into account to stop its development, especially aMCI to AD. However, MCI can be detected by a clinical neuropsychological assessment such as the Mini-Mental State Examination (MMSE) or the Montreal Cognitive Assessment (MoCA). Nowadays, early diagnosis through advanced neuroimaging techniques seems to be much more practical. These modalities include magnetic resonance imaging (MRI), functional MRI (fMRI), single photon emission computed tomography (SPECT), positron emission tomography (PET), fluorodeoxyglucose (FDG)-PET, magnetoencephalography (MEG), functional near-infrared spectroscopy (fNIRS), electroencephalography (EEG) [6]–[10]. Among these modalities, EEG remains the most used due to its several advantages. It is a non-invasive neuroimaging technique that records and monitors brain activity using electrodes (up to 256) placed on the scalp. As compared to MRI, PET, or CT, this modality collects data from patients for future analysis with a higher temporal resolution and, more crucially, at a cheaper cost. These data will be processed through artificial intelligence (AI) approaches using machine learning (ML) and/or deep learning (DL), allowing for the development of a self-diagnostic concept. However, this modality is commonly used to predict different types of neurological diseases [11]. For instance, in a recent clinical study, Siuly et al. [6] proposed a framework for MCI detection based on EEG during a resting state. They applied ML techniques, namely Extreme Learning Machine (ELM), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), to distinguish MCI from

HC patients. In addition, another strategy was used by Wen et al. [12] to compare this time MCI patients to healthy ones who all had type II diabetes. They used CNN with multispectral images obtained from EEG signals and achieved 88% accuracy. To distinguish between HC, MCI, and AD, Huggins et al. [13] used topographic images obtained by transforming EEG signals from their time-frequency representation. Concerning classification, they employed an AlexNet DL model. What makes their approach different from others is the transformation of EEG signals into images. Gkenios et al. [14] made a comparison between three recurrent neural networks (RNNs) for diagnosing AD and MCI using EEG. They used a simple long short-term memory (LSTM) model, and two other hybrids combining convolutional neural networks (CNN) with simple LSTM (Conv-LSTM) and CNN with Bidirectional LSTM (Conv-BLSTM) to make the classification between healthy patients, AD and MCI patients. Their results showed that Conv-LSTM performs well in the case of two-class classification (HC vs. MCI) while Conv-BLSTM performs well for all three classes. Recently, Fouladi et al. [15] proposed two methods based on DL architectures to classify between MCI and AD. The first is a modified CNN architecture, and the second is a convolutional autoencoder (Conv-AE) neural network, with which they obtained scores of 92% and 89%, respectively.

The main purpose of this paper is to present a novel approach that can automatically diagnose MCI to highlight the MCI identification issue. This is accomplished by differentiating between individuals with amnesic MCI, non-amnesic MCI, and healthy controls (HC) using EEG data collected during a specific cognitive test known as the verbal fluency task (VFT). This test is the relevant cognitive assessment that has been used to measure individuals' capacity for complex cognitive functioning, such as language and executive functioning [16]. The multiclass classification is an important part of this work and will be done using our proposed method. The remainder of this paper is as follows. In section II, we present the proposed method for MCI detection. Section III presents the findings, while Section IV provides the conclusion.

II. Materials and Methods

This section is reserved for the description of the proposed method for detecting MCI by detailing the experimental data and the different manipulations performed during the procedure. The following pipeline, which is depicted in Figure 1, includes all these processes.

A. Data Description

The EEG data used in this work were collected in a research study conducted by Grässler et al. [17], who provided a detailed description of the data and recording conditions. This data is a part of the data recorded during three successive cognitive tasks, namely Stroop, N-back, and VFT. These cognitive tests are performed on a computer, on which each instruction associated with each test is displayed. 52 individuals provided the data for this study, comprising 26 MCI subjects (13 aMCI and 13 naMCI), as well as 26 HC. Participants are Germans (males and females), educated, and aged between 55 to 80 years. For precision, the recording occurred in the morning for some participants and in the afternoon for others, with an average duration of 9 minutes. The equipment used for recording is the MOVE Brain Vision wireless system, with an EEG/fNIRS cap. The cap is mounted following the standard 10-20 system configuration. In this paper, we focus only on the EEG recordings during the VFT task, which is described as:
EEG: Signals are recorded at a frequency sampling of 1000 Hz using 32 electrodes (AFp1, AFp2, AFF5h, AFF1h, AFF2h, AFF6h, F7, F8, FFC5h, FFC1h, FFC2h, FFC6h, FTT7h, FCC3h, FCC4h, FTT8h, TTP7h, CCP3h, CCP4h, TTP8h, TP9, TP10, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2, PO10).
Verbal fluency task: The VFT is a cognitive assessment tool that consists in evaluating the capacity of the subjects to generate, in a defined time, a maximum of words corresponding to certain criteria of the task [16]. However, the participants have to pronounce as many words (nouns, verbs, adjectives) that start with a specific letter, this is the phonological criterion. Similarly, they are asked to say as many words as possible that belong to a specific category, this is the semantic criterion. Therefore, each assessment condition will be run three times (trials) as a block and will last 30 seconds (s). A resting block of 31-34 s is set between each block.

B. Data Preparation and Pre-processing

We extracted the VFT-related portions of the raw data while leaving out the other test components, taking into consideration the various tests present in the raw data. Subsequently, the data were resampled from 1000.0 Hz to 128.0 Hz to reduce complexity and data size. For filtering, we applied a notch filter to remove the 50 Hz frequency band (according to the European electricity standard) from the data. This filter attenuates the power line disturbance frequency. In addition, we applied a 3rd-order Butterworth bandpass filter with a lower and upper-frequency band of 0.1 Hz and 32 Hz, respectively. This choice is based on the fact that the brain activity of interest for the detection of MCI and AD is set in this frequency range [14]. Finally, a 3-s window with a 1-s overlap was chosen to segment the data into equal-sized epochs.

C. Proposed Model Analysis

1) Used architecture: The proposed method is based on the Transformer architecture [18], which introduces the concept of self-attention, which allows the model to weigh the importance of different parts of the input when making predictions. The key innovation of the transformer is the use of multi-head attention, which allows the model to attend to different parts of the input simultaneously and learn more complex relationships between encoded features. It consists of stacks (up to 6 each containing the same structure) of **Encoders** and **Decoders**. In the proposed model, we built the Encoder part as in the original description, by creating two sub-layers, respectively, a *multi-head attention* module and a *feedforward neural network* with exactly two deeply connected layers (Dense layers).

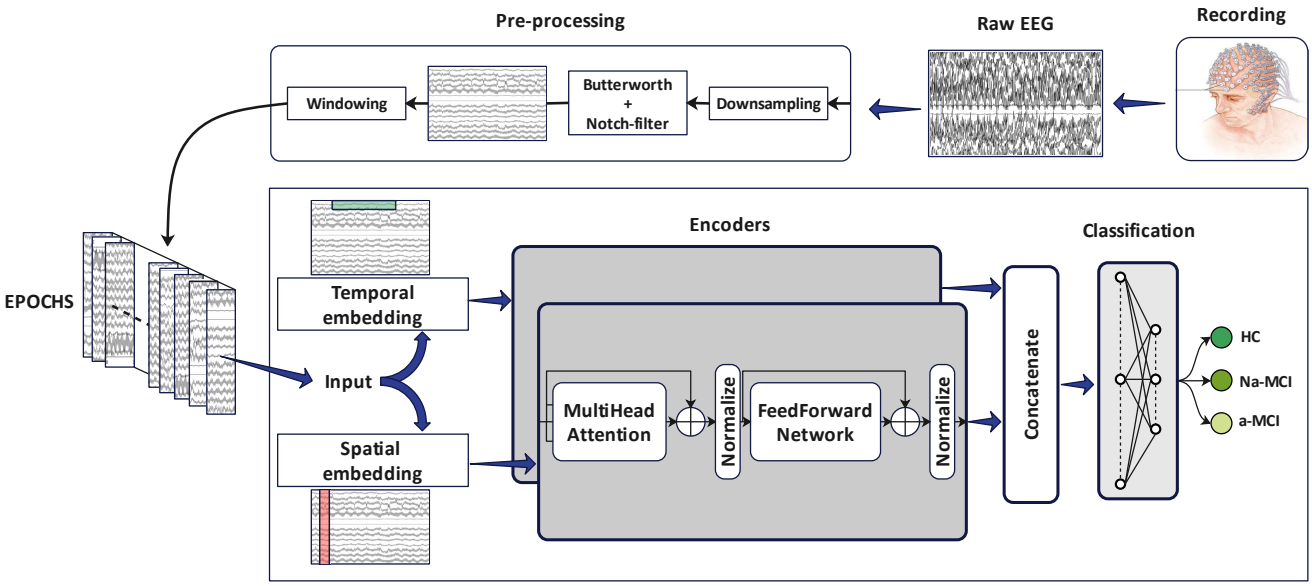

Fig. 1: The pipeline of the proposed method for MCI detection.

The ReLu activation function is applied after the first Dense before passing the output to the second layer. A residual connection, notably an *Add layer* followed by a *Normalization layer* is defined after each of these two blocks. They are used to add the input with the expected output of each block and normalize them in order to serve as input for the next layer. Assuming x, an input of a given sub-layer, it could be represented by $Sublayer(x)$, then its output will be proceeded by the normalization layer as $LayerNorm(x+Sublayer(x))$. Finally, for reasons of conformity with the output size of each sub-layer, and to facilitate subsequent operations, the outputs will be reduced to the same size as that of the layer embedding the initial input. A value corresponding to this dimension will therefore be set by an essential parameter called d_{model}, whose maximum value must not exceed 512 as mentioned in the original paper [18].

2) Contribution: We have added, alongside the existing architecture, the ***temporal embedding*** and ***spatial embedding*** features extractors using different types of convolution operations. These extractors replace the input embedding technique and positional encoding used in the original Encoder. However, convolution operations are well known in the field of computer vision for their exploits, especially in the context of image recognition. They are also used for other tasks, such as signal processing and classification [19]. Their strengths lie in the extraction of features from data using a set of filters. These filters, the higher they are, allow for the detection and extraction of various shapes and patterns from the signals. In this regard, DL architectures based on these concepts are varied and naturally powerful. The best known is CNN, whose main concept is to use both convolution and downsampling operations. This allows for a reduction in the spatial dimension of the data while retaining the important features.

Temporal embedding is defined with the goal of exploring each input data to extract relevant features along the time axis, from the signal produced by each electrode individually. Therefore, we used a regular convolution by applying a *Conv2D* layer with a kernel of $1 \times k$ size instead of $k \times k$ to consider only one signal at a time while sliding the kernel (windows) to reach all electrodes independently. This same process will be repeated F times to produce various feature maps, where F defines the number of filters.

Spatial embedding is performed on the electrode axis to extract features related independently to spatial correlations as well as cross-channel correlations [20] from all electrodes at each time point. For this reason, we applied a *depthwise separable convolutions* [21]. Unlike the standard convolution, which extracts simultaneously the spatial and channel-wise features from a multi-dimension input, a depthwise separable convolution operates separately in two stages: *depthwise convolution* and *pointwise convolution*. However, the depthwise convolution is performed by convolving a $k \times k$ size *depth kernel* with each channel of an input independently. While the *pointwise convolution* performs by iterating point by point over the pooled feature maps outputted by the depthwise convolution to realize a linear combination. It is similar to the standard convolution but rather uses a kernel size of 1×1. Depthwise separable convolutions are less computational-intensive because they significantly reduce the number of parameters resulting from convolution operations [22], [23]. Table I summarizes the different components of these two embeddings and their potential parameters. The outputs of the encoders corresponding respectively to the temporal and spatial features will then be concatenated, then fitted into the classification network. Finally, as in the original Transformers, the number of encoders can vary between 1 and 6 layers. As well as the number of heads of the modules related to Multi-head attention which is between 1 and 8. But it is necessary

to know that for each temporal and spatial embedding, an independent layer of the encoder corresponds exactly. In addition, to have removed the positional encoding unit and the linear embedding of the inputs, we also rejected the notion of masking, used in the native architecture.

TABLE I: Summary of embedding module layers with their parameters.

Layer name	Filter	Value	Output shape
Temporal embedding			
Input	–	–	$C \times T$
Reshape	–	–	$C \times T \times 1$
Conv2D	F_1	$1 \times K_1$	$C \times T \times F_1$
BatchNormalization	–	–	$C \times T \times F_1$
Activation	–	$relu$	$C \times T \times F_1$
AveragePooling2D	–	$1 \times P_1$	$C \times (T/P_1) \times F_1$
Dropout	–	p_1	$C \times (T/P_1) \times F_1$
Reshape	–	–	$F_1 \times (C * (T/P_1))$
Dense	–	d_{model}	$F1 \times d_{model}$
Spatial embedding			
Input	–	–	$C \times T$
Reshape	–	–	$C \times T \times 1$
DepthwiseConv2D	D	$C \times 1$	$1 \times T \times D$
BatchNormalization	–	–	$1 \times T \times D$
Activation	–	$relu$	$1 \times T \times D$
AveragePooling2D	–	$1 \times P_2$	$1 \times (T/P_2) \times D$
Dropout	–	p_2	$1 \times (T/P_2) \times D$
SeparableConv2D	F_2	$1 \times K_2$	$1 \times (T/P_2) \times F_2$
BatchNormalization	–	–	$1 \times (T/P_2) \times F_2$
Activation	–	$relu$	$1 \times (T/P_2) \times F_2$
AveragePooling2D	–	$1 \times P_3$	$1 \times (T/(P_2 * P_3)) \times F_2$
Dropout	–	p_3	$1 \times (T/(P_2 * P_3)) \times F_2$
Reshape	–	–	$F_2 \times (1 * (T/(P_2 * P_3)))$
Dense	–	d_{model}	$F_2 \times d_{model}$

$Channel : C = 32$; $Time : T = 384$; $Filter : F1 = F2 = 64$; $D = 1$, $Kernel : K_1 = K_2 = 32$; $P_1 = 9$; $p_1 = p_2 = p_3 = 0.1$; $P_2 = 4$; $P_3 = 6$; and $d_{model} = 128$.

III. Experimental validation

In this section, we describe the experimental setting and the performance outcomes resulting from our approach.

A. Data Setup

The data is separated into three subsets based on the phases of a classification task: the training set (60%), the validation set (10%), and the testing set (30%), which can be used respectively to train the model, validate the model at each step of the training, and test and validate the performance of the final model.

B. Classification results

All the experiments were carried out on a Hp Pavilion Gaming 15 laptop running on the Windows 11 operating system and equipped with an Intel® Core™ i7-10870H CPU @ 2.20GHz and an NVIDIA GeForce RTX 2060 (Max Q-Design) graphics card supporting CUDA 11.8 (cuDNN 8.1). We trained the model on our dataset, which contained 52 participants in total. Each data input is in the form of the number of electrodes with the number of samples. After splitting the data into a number of segments of the same size. The new data format is defined as (number of segments $\times$ number of electrodes $\times$ number of samples). The segmentation takes into account the segment duration and the overlap, both in seconds. After a number of trials combining different options for segment length and overlap. We selected the 3s-1s combination, which gave a better result.

According to the three subsets, the total number of recordings was (14183, 32, 384), of which (9928, 32, 384) were used for training and (4255, 32, 384) for testing. A total of (9928, 32, 384) recording were made after then, of which (8935, 32, 384) were used for training and (993, 32, 384) for validation. Based on the confusion matrix results, the true positive values of healthy, aMCI, and naMCI are respectively equal to 2082, 979, and 972, (per Figure 2). We evaluated the performance of our model in terms of statistical scores computed from the confusion matrix. This includes the values of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The concerned metrics [24] are accuracy (94.78%), precision (96.39%), recall (i,e,. sensitivity) (93.37%), specificity (93.26%), and F1-score (95.20%).

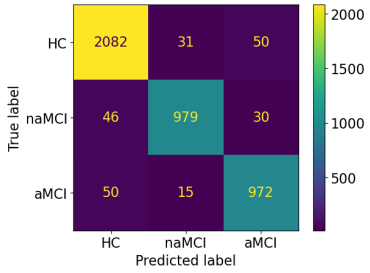

Fig. 2: Confusion Matrix.

The history of the training and validation curves for 100 epochs is shown in Figure 3. After achieving the maximum number of epochs, the accuracy of these curves increases while their losses decrease.

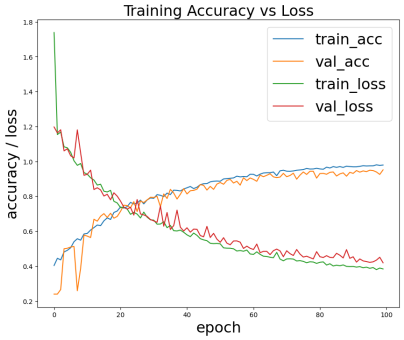

Fig. 3: Training history: training and validation accuracy, and training and validation loss for 100 iterations.

C. Comparison results

Once the best combination was obtained, we compared our method with two others in the literature. In [14], the authors proposed a hybrid model based on CNN and bidirectional LSTM to perform a classification between AD, MCI, and HC. The achieved accuracy test is equal to 94.59%. In [15], their aim was to distinguish people with AD from those with MCI. They reached a score of 89% with the Conv-AE model and 92% with the CNN. In this regard, we have implemented the previously cited models using our dataset. Table II summarizes the obtained results. Our approach achieved a score of 94.59%, surpassing the performance demonstrated by the results of this comparison.

TABLE II: Comparison findings.

Authors	Methods	Testing accuracy
Gkenios et al. [14]	CNN+BLSTM	85%
Fouladi et al. [15]	CNN / Conv-AE	90%
Proposed method	Transformer model	**94.78%**

IV. Conclusion

In this paper, we proposed a new MCI detection method based on the transformer model using EEG signals. The EEG recording was done during a verbal fluency task. Our goal was to distinguish three classes, including HC, aMCI, and naMCI, and we achieved a classification result of 94.78%. Finally, we compared our proposed method with other existing studies in the literature dealing with the same topic. Thanks to the temporal embedding and spatial embedding that we introduced, the results demonstrated the performance and efficiency of our approach based on the Transformer architecture. This approach remains efficient because of the high learning capability due to the parallel processing of the entire input features.

Acknowledgment

This work is a part of the *ICT-rollator* project (Innovative ICT-based Rollator to Promote Active and Healthy Aging in Preventive and Therapeutic Conditions) funded by the German Ministry of Education and Research (BMBF Project No. 01DH21023A) and the Tunisian Ministry of Higher Education and Research.

References

[1] J. Golomb, A. Kluger, and S. H. Ferris, "Mild cognitive impairment: historical development and summary of research," *Dialogues in clinical neuroscience*, 2022.

[2] M. Kashefpoor, H. Rabbani, and M. Barekatain, "Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features," *Journal of medical signals and sensors*, vol. 6, no. 1, p. 25, 2016.

[3] S. Jongsiriyanyong and P. Limpawattana, "Mild cognitive impairment in clinical practice: a review article," *American Journal of Alzheimer's Disease & Other Dementias®*, vol. 33, no. 8, pp. 500–507, 2018.

[4] S. R. Rapp, C. Legault, V. W. Henderson, R. L. Brunner, K. Masaki, B. Jones, J. Absher, and L. Thal, "Subtypes of mild cognitive impairment in older postmenopausal women: the women's health initiative memory study," *Alzheimer disease and associated disorders*, vol. 24, no. 3, p. 248, 2010.

[5] N. D. Anderson, "State of the science on mild cognitive impairment (MCI)," *CNS spectrums*, vol. 24, no. 1, pp. 78–87, 2019.

[6] S. Siuly, Ö. F. Alçin, E. Kabir, A. Şengür, H. Wang, Y. Zhang, and F. Whittaker, "A new framework for automatic detection of patients with mild cognitive impairment using resting-state EEG signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 9, pp. 1966–1976, 2020.

[7] K. Engedal, M. L. Barca, P. Høgh, B. B. Andersen, N. W. Dombernowsky, M. Naik, T. E. Gudmundsson, A.-R. Øksengaard, L.-O. Wahlund, and J. Snaedal, "The power of EEG to predict conversion from mild cognitive impairment and subjective cognitive decline to dementia," *Dementia and Geriatric Cognitive Disorders*, vol. 49, no. 1, pp. 38–47, 2020.

[8] T. Nguyen, M. Kim, J. Gwak, J. J. Lee, K. Y. Choi, K. H. Lee, and J. G. Kim, "Investigation of brain functional connectivity in patients with mild cognitive impairment: a functional near-infrared spectroscopy (fNIRS) study," *Journal of Biophotonics*, vol. 12, no. 9, p. e201800298, 2019.

[9] S. Sridhar and V. Manian, "EEG and deep learning based brain cognitive function classification," *Computers*, vol. 9, no. 4, p. 104, 2020.

[10] D. Yang, R. Huang, S.-H. Yoo, M.-J. Shin, J. A. Yoon, Y.-I. Shin, and K.-S. Hong, "Detection of mild cognitive impairment using convolutional neural network: temporal-feature maps of functional near-infrared spectroscopy," *Frontiers in Aging Neuroscience*, vol. 12, p. 141, 2020.

[11] D. Yang, Y.-I. Shin, and K.-S. Hong, "Systemic review on transcranial electrical stimulation parameters and EEG/fNIRS features for brain diseases," *Frontiers in Neuroscience*, vol. 15, p. 629323, 2021.

[12] D. Wen, Y. Zhou, P. Li, P. Zhang, J. Li, Y. Wang, X. Li, Z. Bian, S. Yin, and Y. Xu, "Resting-state EEG signal classification of amnestic mild cognitive impairment with type 2 diabetes mellitus based on multispectral image and convolutional neural network," *Journal of Neural Engineering*, vol. 17, no. 3, p. 036005, 2020.

[13] C. J. Huggins, J. Escudero, M. A. Parra, B. Scally, R. Anghinah, A. V. L. De Araújo, L. F. Basile, and D. Abasolo, "Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046087, 2021.

[14] G. Gkenios, K. Latsiou, K. Diamantaras, I. Chouvarda, and M. Tsolaki, "Diagnosis of Alzheimer's disease and mild cognitive impairment using EEG and recurrent neural networks," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3179–3182, IEEE, 2022.

[15] S. Fouladi, A. A. Safaei, N. Mammone, F. Ghaderi, and M. Ebadi, "Efficient deep neural networks for classification of Alzheimer's disease and mild cognitive impairment from scalp EEG recordings," *Cognitive Computation*, vol. 14, no. 4, pp. 1247–1268, 2022.

[16] L. Olabarrieta-Landa, E. L. Torre, J. C. López-Mugartza, E. Bialystok, and J. C. Arango-Lasprilla, "Verbal fluency tests: Developing a new model of administration and scoring for spanish language," *NeuroRehabilitation*, vol. 41, no. 2, pp. 539–565, 2017.

[17] B. Grässler, F. Herold, M. Dordevic, T. Gujar, S. Darius, I. Böckelmann, N. Mueller, and A. Hökelmann, "Multimodal measurement approach to identify individuals with mild cognitive impairment: Study protocol for a cross-sectional trial," *BMJ Open*, vol. 11, p. 46879, 05 2021.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[19] S. Chaabene, B. Bouaziz, A. Boudaya, A. Hökelmann, A. Ammar, and L. Chaari, "Convolutional neural network for drowsiness detection using EEG signals," *Sensors*, vol. 21, p. 1734, 2021.

[20] R. Shang, J. He, J. Wang, K. Xu, L. Jiao, and R. Stolkin, "Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification," *Knowledge-Based Systems*, vol. 194, p. 105542, 2020.

[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[22] L. Kaiser, A. N. Gomez, and F. Chollet, "Depthwise separable convolutions for neural machine translation," *arXiv preprint arXiv:1706.03059*, 2017.

[23] G. Lu, W. Zhang, and Z. Wang, "Optimizing depthwise separable convolution operations on GPUs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 1, pp. 70–87, 2021.

[24] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.