Colour Prediction using Vision Transformer and Continous Wavelet Transform on EEG signals

Puranjay Mishra[a], Marios Antonakakis[b], IEEE member, Koushlendra Kumar Singh[a], Michalis Zervakis[b], IEEE member

[a.] National Institute of Technology Jamshedpur (NIT JSR), Jamshedpur, Jharkhand, India
puranjaymishra0@gmail.com, koushlendra.cse@nitjsr.ac.in
[b.] School of Electrical and Computer Engineering, Technical University of Crete (TUC), Chania, Crete, Greece
{mantonakakis, mzervakis}@tuc.gr

Abstract— **Electroencephalography (EEG)-based classification of brain disease such as epilepsy or schizophrenia, decoding brain activity during movement and vision have been shown promising results in the last years. Here, we introduce a novel pipeline for the presence of speech information carried on EEG signals. The proposed work includes a new conducted EEG dataset of 15 subjects and a deep learning model to predict the colour information. With a unique experimental set up, the data successfully captures the information about the mental enunciation of the set of used colors. The primary goal is to perform multiclass classification using our custom EEG data which records the brain activity of individuals during mental enunciation and thought about a class of objects, in our case, colours. Continuous Wavelet Transform (CWT) is applied on each of the EEG channels of each participant to obtain time-frequency (TF) based characteristics. A Vision Transformer (ViT) based model is then developed and used to capture information from these TF. The method deals with a 6-class classification problem, for which, the 6 different colors are used as target classes for our model. The proposed model achieves 91.36% cross validation accuracy, 5.48x the random guess accuracy. These results clearly demonstrate the existence of speech information in EEG signals and lay the foundational stone for future research in speech assistive technologies.**

Keywords— EEG, BCI, Colour Prediction, Deep Learning, Vision Transformer

I. EEG IS A METHOD OF RECORDING THE CONTINUOUS/TEMPORAL ELECTRICAL ACTIVITY OF THE BRAIN. THIS IS MEASURED USING ELECTRODES OF THE EEG DEVICE, WHICH CAN BE EITHER INVASIVE (INTRACRANIAL EEG OR IEEG) OR NON-INVASIVE (SCALP EEG). EEG SIGNALS SERVE AS A DIAGNOSTIC METHOD FOR VARIOUS CONDITIONS, INCLUDING EPILEPSY [1-3], BRAIN TUMORS [4-6], BRAIN DAMAGE OR DYSFUNCTION [7-9], STROKE, AND SLEEP DISORDERS [10-12]. IT CAN ALSO HELP DISTINGUISH EPILEPTIC SEIZURES FROM OTHER TYPES OF SPELLS SUCH AS PSYCHOGENIC NON-EPILEPTIC SEIZURES [13] OR DIFFERENTIATE "ORGANIC" ENCEPHALOPATHY OR DELIRIUM FROM PRIMARY PSYCHIATRIC SYNDROMES SUCH AS CATATONIA. HOWEVER, EEG HAS SOME DRAWBACKS, INCLUDING POOR SIGNAL-TO-NOISE RATIO AND LOW SPATIAL RESOLUTION WHEN COMPARED WITH METHODS LIKE FMRI.

EEG signals reflect correlated synaptic activity caused by postsynaptic potentials of brain neurons. They also capture the potential difference of neural backpropagation, the electric impulse that travels in the opposite direction of the action potential in the dendrite, and thus have the potential to capture the thoughts and decision-making processes of the brain. The capabilities of EEG signals are not limited to the human brain, as studies on rats and dogs [14-15] have shown promising results. This inspiration has led to attempts to process and interpret EEG signals for coordination and control, decision-making, and thoughts.

The present study aims to provide concrete evidence of the presence of speech information in EEG signals. It also strives to create a generalized model for color prediction that will eventually lead to an end-to-end thought-to-speech translator. The need for such a translator arises from the necessity of a non-invasive, real-time speech assistance device, as a significant part of the global population suffers from some form of speaking impairment. While technologies like voice amplification systems, fluency assistive devices, communication boards, and artificial larynxes can help patients with such conditions, they either require input from the patient, lack expressibility, or are invasive.

The field of Brain-Computer Interface (BCI) has opened paths for the development of a device that can use brain signal procurement methods like EEG to offer real-time, non-invasive, and expressive speech assistance to patients with disabilities [16-18]. As these devices can capture minute changes in potential difference or polarity of different parts of the brain with high frequency, they are the perfect candidate for the data source of a BCI-based speech assistance device [19].

The use of Machine Learning (ML) and Deep Learning (DL) techniques for BCI has become the method of choice in recent times. BCI systems perform tasks through analyzing Event Related Potentials (ERP). These ERPs are electrophysiological responses arising at a specific period after a specific internal or external event [19] They are observed when exposed to a mental, sensory event or the dereliction of a consistently occurring stimulus. ERPs are employed for determining the subject's brain states in response to distinct stimuli. ML and DL techniques have performed well in capturing information across the continuous varying ERPs; emotion classification [20], motor imagery categorization [21], and limb movement classification [22] are examples of their use.

Our proposed methodology tackles the feature extraction from the ERPs of recorded signals differently. We use the Vision Transformer (ViT), a deep learning architecture that applies the transformer model to image data, enabling impressive performance on various computer vision tasks. Vision Transformers capture long range dependencies and context in the image data due to its architecture, which is explained in further sections of this work. Thus, this paper aims to confirm the existence of information about speech activity in EEG signals, a prerequisite for the fully functional speech assistance device. The proposed method has been validated through limited data obtained by a simple experimental setup and a Vision Transformer-based neural network model.

A. Related work

Speech-based EEG research is currently more popular than ever, with notable works in various fields such as envisioned speech [21,24], image visualization [26], decision and preference classification [26,27]. Our work bears similarities to two specific studies: 'Envisioned speech recognition using EEG sensors' [21] (Fig. 1) and 'A hybrid classifier combination for home automation using EEG signals' [27], although there are a few key differences.

The paper on Envisioned Speech [21] used the following experiment setup: 23 participants were asked to imagine and visualize 10 digits, 10 characters, and 10 animals while EEG data was collected. This EEG data first underwent denoising. Next, four different features were computed on the smoothed signal, namely Standard Deviation (SD), Root Mean Square (RMS), Sum of Values (SUM), and Energy (E) (as defined in the paper). These extracted features were then concatenated to form a new feature vector. Coarse-level (3-class classification among digits, characters, and animals) and fine-level classification (classification of each individual class) were performed on the extracted features using the Random Forest classifier. For EEG signal acquisition, a wireless neuroheadset called Emotiv EPOC+ [28] was utilized.

The work on home automation using EEG signals was set up with the subjects of the experiment giving mental commands to change the state of the light bulb (on to off and vice-versa), and the EEG signal was recorded using the Emotiv EPOC+ headset. The EEG signal underwent denoising through the Savitzky-Golay (S-Golay) filter [29], which convoluted adjacent points and improved signal quality. Parameter-based preprocessing involved time segmentation and down-sampling. Time segmentation divided the total recorded time into smaller sections, determining whether the user intended to turn on or off the bulb. Different time segments (4, 2, 1, 0.5) were tried to balance data samples and performance. Down-sampling reduced the sampling rate of the EEG data, and the model was trained at various rates (128, 64, 32, 16, 8, 4, 2, and 1 Hz). The optimum performance was achieved at 16 Hz and 1 Hz. The key difference between this work and our research lies in the element of mental vocalization of a class by the participants. While the Home Automation paper's decision classification may have elements of such mental vocalization, Envisioned Speech is purely based on the visualization of the image class [30].

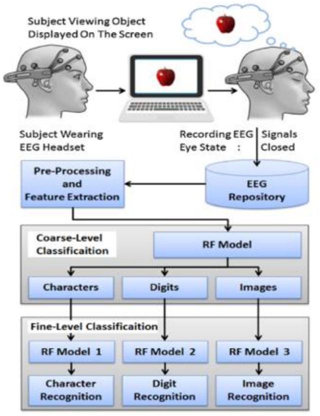


Fig. 1 Work done in the Envisioned Speech paper [21]

Other works [31-32] have attempted to identify mentally vocalized speech. In [31], seven volunteers were studied, and autoregressive coefficients were used with Independent Component Analysis, Artifact removal, and Subspace filtering for imagined syllable classification. The results were inconclusive, requiring further research. In [32], an extensive investigation of BCI applications involved 15 test subjects imagining short words, long words, and vowels. The collected data underwent feature extraction and classification using Relevance Vector Machine (RVM), demonstrating clear suitability for EEG signals in imagined speech for each test subject, achieving notable accuracy figures.

The above-mentioned works suggest that speech prediction as a unified model across all human subjects using EEG signals is a promising approach. However, the absence of the right experimental setup for such a unified model makes proving such a hypothesis challenging. This situation inspired our research, which resulted in the following contributions:

1. Designed and developed a new database of EEG signals that are unique compared to the relevant state-of-the-art techniques.
2. Created a diverse word bank for the model to classify, including 6 colors for mental enunciation: 'red,' 'green,' 'blue,' 'black,' 'white,' and 'yellow.' These words range from single syllables (e.g., 'red' and 'black') to two syllables (e.g., 'blue,' 'white,' and 'yellow'), increasing the complexity of the task.
3. Developed a Vision Transformer-based model that achieved 5.48x higher accuracy than random guessing across all subjects in our experiment.

II. PARTICIPANT AND METHODS

A. *Experimental Setup And Procedure*

In this experiment, we selected 15 volunteers from a diverse age bracket (19-35) to contribute to our dataset. Although the median age of participants lies in the early twenties, our goal was to generate a generalized dataset with no age bias. We used Brain Tech’s 32-channel EEG machine, operating in the frequency band ranges of 0.1-100Hz, with an input range of -8mV to +8mV and electrodes made of Silver-Silver Chloride (Ag-AgCl). The electrode placement followed the 10-20 system of EEG electrode placement, ensuring different inter-electrode spaces, proportional electrode placement to skull size and shape, and coverage of all regions of the brain (Frontal, Temporal, Occipital, Parietal). We cleaned the site of application using cotton buds and placed the EEG electrodes with a Polyoxyethylene-based conductive gel.

All the volunteers were students at National Institute of Technology, Jamshedpur, who were requested to remain calm during the whole process with clear thoughts. Additionally, they were asked not to consume caffeine, alcohol, or smoke prior to the recording process to avoid any undesirable effects of these substances on the nervous system. In a dimly lit room, a slide presentation displaying all the color classes was prepared and shown to the volunteers. Slides were manually changed every 20 seconds. The participants had no prior information about the experiment, and the data recording was done in a single take to avoid temporal correlation, which can lead to inflated performance figures [36]. According to their findings, recording data in random single attempts instead of repeating each word 30 times in a block brings the accuracy down to the chance level.

During the experiment, the participants were asked to mentally enunciate the color displayed on the screen with their utmost concentration, constantly throughout the 20-second block. A fixed mental vocation rate was not set to avoid distracting the participants, and they were allowed to choose their own pace.

Data was collected for a total of 120 seconds (6 classes, each lasting 20 seconds) using the 28 electrodes of the EEG device. The initially recorded matrix from each subject in the dataset has dimensions of (5120,16), where the 0th dimension represents the total number of data points collected in the duration of 20 seconds for the 6 different colors, and the 1st dimension represents the 16 potential differences (PD) of electrodes on the scalp. These 16 PD values were split individually, and treated as separately for the artifact removal. In order to enhance the quality and reliability of the EEG signal time-frequency plots, a series of preprocessing steps were applied. These steps aimed to mitigate the presence of artifacts and unwanted noise, thus providing a clearer representation of the underlying neural activity.

B. EEG preprocessing

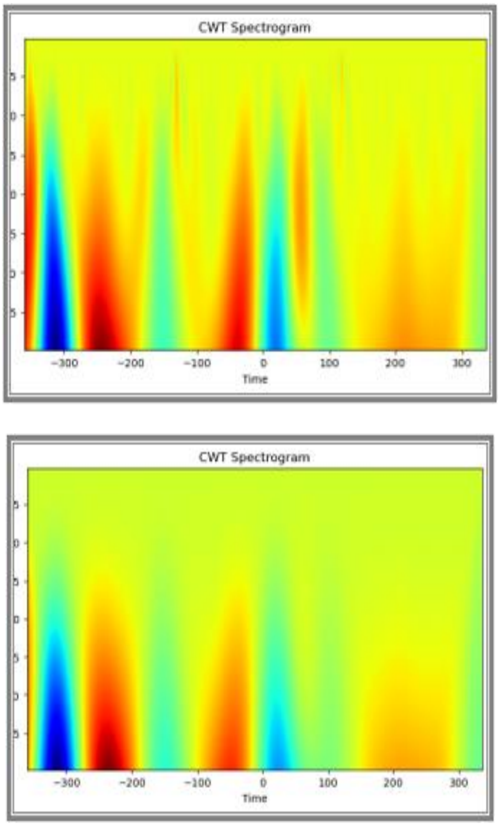
The collected data was manually cleaned and then artifact removal was performed on each data sample.

A baseline shift removal was performed. Baseline shift removal in EEG data refers to the process of eliminating fluctuations in the baseline (average voltage) to ensure it remains centered around zero. This step is crucial as it helps prevent spurious effects and allows researchers to focus on genuine changes in brain activity without interference from baseline variations. This was done using the Asymmetric Least Squaring [33]. The method of Asymmetric least squares is a regression method that assigns different weights to overestimated and underestimated data points, reducing the impact of outliers on the model's fit.

We then performed a Powerline interference correction [34] on the data. Powerline interference correction in EEG data involves eliminating electrical noise at the powerline frequency (typically 50 Hz or 60 Hz) to enhance the signal quality. This correction is crucial as powerline interference can introduce artifacts, distorting the EEG recordings and hindering accurate analysis of brain activity. To address the interference at 60 Hz, the Powerline interference correction was done using the Butterworth bandstop filter using the order of 6, low limit value set to 59 Hz and the higher limit at 61 Hz.

The time-frequency representation was estimated on these 16 potential differences values post artifact removal using Continuous Wavelet Transform (CWT) with the Ricker wavelet. The Ricker wavelet is a symmetric wavelet with a peaked shape resembling a Mexican hat, defined by a single parameter determining its width. The width parameter influences the frequency resolution of the wavelet transform, enabling the analysis of signals in both time and frequency domains simultaneously. This provides insights into how the signal's frequency components change over time, offering a rich representation of the signal's time-varying spectral content.

The spectrogram before and after the preprocessing steps is presented in Fig. 2.



a)

b)

Fig. 2 - CWT plot **a)** without preprocessing and **b)** with preprocessing.

C. The Vision Transformer Model

We used the CWT with the Richter wavelet [35] in conjunction with the pretrained Vision Transformer (version - 'vit-base-patch16-224-in21k'). The output of the model was fed into a set of linear layers to perform classification. In Fig. 3, we can find a simplified illustration of the ViT pipeline.

Our model comprises a total of 16 layers, with 12 of them being part of the pretrained ViT. These layers include Patch Embedding layers, Positional Embedding layer, Multi-Headed Self Attention, Layer Normalization, and a Feed Forward Network. The final classification was performed on the output of the pretrained ViT using 2 layers of fully connected layers. A pictorial illustration of this can be found in the Fig. 4.

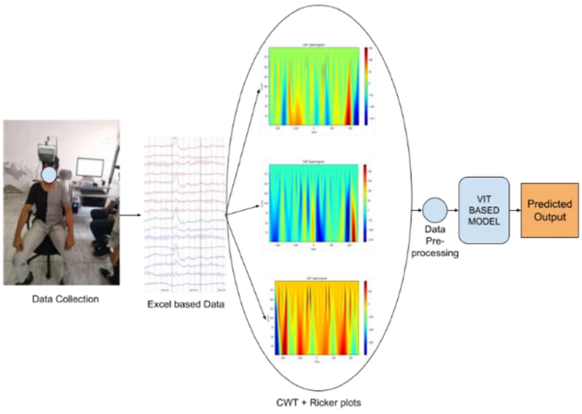


Fig. 3 – Pipeline of the proposed study.

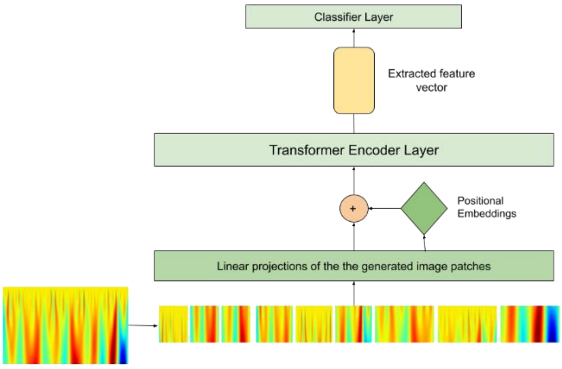


Fig. 4 – Illustration of a simplified view of our ViT model.

The feature extraction process of the ViT model begins with the Patch Embedding layer. The input image is divided into smaller patches, and each patch is linearly embedded into a lower-dimensional representation. This allows the model to extract local information from the image and focus on specific regions, addressing computational challenges associated with processing entire images. The Patch Embedding layer reduces the spatial dimensionality by segmenting the image into patches and generating a series of patch embeddings. These embeddings are utilized by subsequent layers for further processing and feature extraction.

The embedded patches then receive positional information from the Position Embedding layer. As the transformer architecture does not inherently capture spatial relationships between patches, the Position Embedding layer encodes relative or absolute positions of each patch within the image. This incorporation of positional information enables the model to understand dependencies between patches and the spatial organization of the image. Position Embeddings can be pre-existing or learned during training, allowing the ViT model to reason and capture global dependencies by understanding context and relationships between various patches. The Transformer Encoder layer (Fig. 5) is a block of layers that consist of the Multi-Headed Attention block, a Normalize layer and a Feed Forward Neural Network with skip connections, illustrated below.

The Multi-Head Attention mechanism (Fig. 6) enables the model to capture global dependencies and interactions within visual data. It comprises linear projections, scaled dot-product attention, and head concatenation

1. Linear Projections: The input visual features are transformed into queries, keys, and values using linear projections. These projections map the input features into lower-dimensional spaces to facilitate attention computations.
2. Scaled Dot-Product Attention: The queries, keys, and values are used to compute attention weights, which measure the relevance between different patches within the image. This computation involves taking the dot product between the queries and keys, followed by a SoftMax operation to obtain normalized attention weights. The dot product measures the degree of alignment, similarity, or difference between the query and key values. The softmax operation ensures that the attention weights sum up to 1 and represent the importance of each patch for each query.
3. Head Concatenation: Multiple attention heads are employed to capture different types of dependencies and allow the model to learn diverse representations. Each attention head performs separate linear projections, scaled dot-product attention, and subsequent value computations. The outputs of the attention heads are concatenated and linearly transformed to obtain the final output.

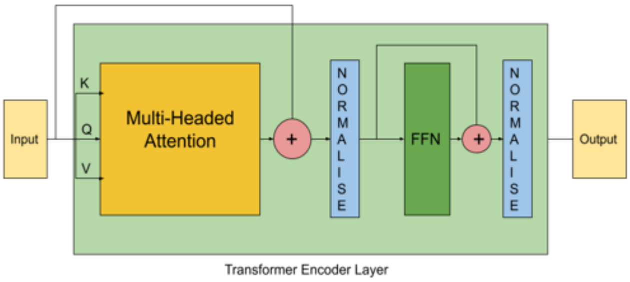


Fig. 5 – Working of the Transformer Encoder Layer.

III. RESULTS

The model described in Section II was developed in PyTorch [37] and used for multiclass classification. Cross Entropy was employed as the loss function, calculating the training loss between the predicted and true labels. Adam was chosen as the optimizer to perform the weight update step for our model.

The model was trained for 20 iterations and each iteration was validated using the 5-fold cross validation split for 25 epochs. The average macro precision and macro recall, confusion matrix and the ROC curve with the AUC score was obtained to observe model performance.

The model achieved an average training accuracy of 99.37 (with standard deviation being 0.97) with the average cross validation accuracy of 91.36 (with standard deviation of 0.70) across the 20 iterations. Furthermore, the model obtained an average macro precision and recall for the cross-validation sets are 0.92 and 0.93 respectively (averaged across the 20 iterations). The graphs for the training loss, training accuracy and accuracy on the 5-fold cross validation set are illustrated below. Figure 12 is the confusion matrix for one of the cross-validation sets and for a random epoch.

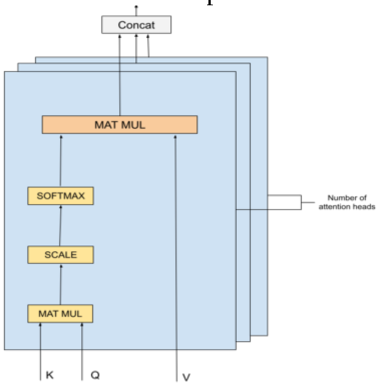


Fig. 6 – Illustration of the Multi-Headed Attention block.

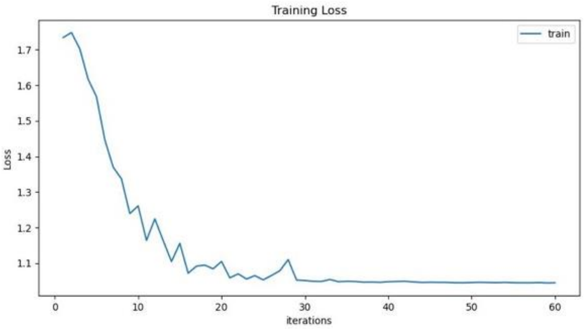


Fig. 8 – Training loss vs epochs

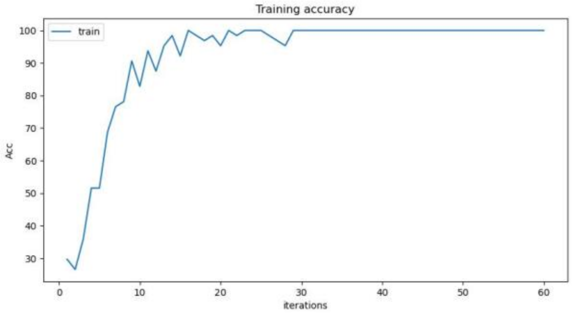


Fig. 9 – Training accuracy vs epochs

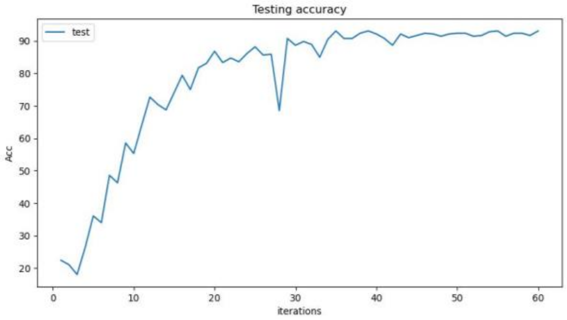
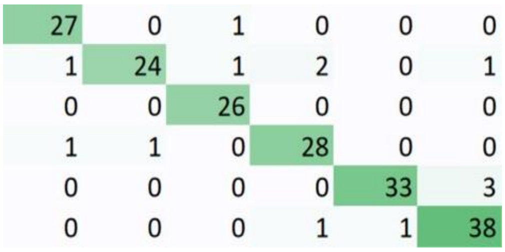


Fig. 10 - Testing accuracy vs epochs



27	0	1	0	0	0
1	24	1	2	0	1
0	0	26	0	0	0
1	1	0	28	0	0
0	0	0	0	33	3
0	0	0	1	1	38

Fig. 12 – Confusion Matrix for the cross validation set

IV. DISCUSSION

This work introduces a novel EEG dataset that aims to prove the direct relationship between mentally vocalisation and EEG signals through the problem of Colour prediction. This dataset underwent artifact removal using Baseline Correction and Powerline Interference Removal. A Continuous Wavelet Transform with the ricker wavelet to obtain the time-frequency plots and then a Vision Transformer based model was used to classify the data into the preset classes. The model was evaluated using 5-fold cross validation, average macro precision and average macro recall across the validation folds. It obtained an accuracy of 91.36% on the 5-fold cross validation and 0.925 and 0.930 on the macro precision and macro recall.

TABLE I. COMPARISON WITH PREVIOUS RESEARCH

No	Research Group	No. Of Classes	Approach	Accuracy (%)
1	Pradeep et. al. [21]	10 characters 10 digits 10 images	classification of envisioned objects using a RF classifier	66.91 68.46 65.72 Avg = 67.03
2	Roy et. al. [25]	3 classes	classification of mental decisions using a LSTM + RF model	68.36
3	Brigham et. al. [29]	2 classes	Subject-wise classification of imagined syllables /ba/ and /ku/ using K Nearest Neighbours	61
4	Nguyen et. al. [30]	3 vowels 3 short words 2 long words	Classification of groups of vowels, short and long words	49 50.1 66.2 Avg = 55.1
5	**Proposed method**	**6 colors**	**Classification of a set of mentally enunciated colors using CWT+ Vision Transformer model**	**91.36**

The model gives state-of-the-art performance amongst its peers that have similar experimental setups and objectives (see Table I). The test accuracy obtained outperforms all the models mentioned in the above table because of the proposed method of detection which uses the Vision Transformer.

Our experimental set up is unique in the sense that it combines the idea of visualization and vocalization of a subject. This resembles the natural process of speech generation, where the brain processes several inputs like visual and auditory along with the words for a conversation.

Our research work has a strong transformative potential as it demonstrates the possibility of using EEG signals as input for an end-to-end Speech Generative model. However, the limited class count and the simplicity of one-word outputs are on its limitations. A complex experimental set-up and the use of sequence models to generate real conversational outputs would be future steps that could overcome the limitation of this work.

V. CONCLUSION

In this paper, we introduced a novel experimental setup with a custom dataset and a ViT-based model for predicting thoughts of speech using EEG signals. By leveraging the CWT with Ricker wavelet and the VIT model, our approach performed competitively with existing works. The model achieved vital results, surpassing chance accuracy by over 5.48 times on our custom dataset and reaching an impressive 91.36% on the validation dataset, highlighting the potential of EEG signals as a source of speech-related information. However, further exploration of the full capabilities of EEG signals is warranted. Future research could explore diverse pre-training strategies to indirectly train the model for speech prediction, investigate alternative inputs in ViT such functional connectivity graphs as well as develop real-time speech prediction language models.

ACKNOWLEDGMENT

This work was supported by the EU co-funded project 'smartHealth' (Type of action: DIGITAL-SIMPLE, grant no. 101083630). Special thanks to Harsh Singh and Aman Pandey for giving their valuable contribution in the data collection phase of the project.

REFERENCES

[1]. U.R. Acharya, et al. "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals." *Computers in biology and medicine* 100 (2018): pp. 270-278, doi:10.1016/j.compbiomed.2017.09.017

[2]. U.R. Acharya, H. Fujita, V.K. Sudarshan, S. Bhat, and Joel EW Koh. "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review." *Knowledge-based systems* 88 (2015): 85-96, doi: 10.1016/j.knosys.2015.08.004

[3]. G. Sdoukopoulou, M. Antonakakis, G. Modde, C. H. Wolters and M. Zervakis, "Interictal Spike Classification in Pharmacoresistant Epilepsy using Combined EEG and MEG," 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE), Kragujevac, Serbia, 2021, pp. 1-6, doi: 10.1109/BIBE52308.2021.963550.

[4]. Murugesan, Muthukumar, and R. Sukanesh. "Automated detection of brain tumor in EEG signals using artificial neural networks." In *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pp. 284-288. IEEE, 2009., doi: 10.1109/ACT.2009.77

[5]. Karameh, Fadi N., and Munther A. Dahleh. "Automated classification of EEG signals in brain tumor diagnostics." In *proceedings of the 2000 American control conference. ACC (IEEE cat. No. 00CH36334)*, vol. 6, pp. 4169-4173. IEEE, 2000, doi: 10.1109/ACC.2000.877006

[6]. Sharanreddy M., and P. K. Kulkarni. "Detection of primary brain tumor present in EEG signal using wavelet transform and neural network." *Int J Biol Med Res* 4.1 (2013): pp. 2855-9

[7]. Albert, Bruno, et al. "Automatic EEG processing for the early diagnosis of traumatic brain injury." *Procedia Computer Science* 96 (2016), pp: 703-712, doi: 10.1016/j.procs.2016.08.253

[8]. Wang, Fang, et al. "An ensemble of Xgboost models for detecting disorders of consciousness in brain injuries through EEG connectivity." *Expert Systems with Applications* 198 (2022), p. 116778, doi: 10.1016/j.eswa.2022.116778

[9]. Goldfine, Andrew M., et al. "Determination of awareness in patients with severe brain injury using EEG power spectral analysis." *Clinical Neurophysiology* 122.11 (2011), pp. 2157-2168, doi: 1016/j.clinph.2011.03.022

[10]. Qureshi, Arooj Ahmed, Canxiu Zhang, Rong Zheng, and Ahmed Elmeligi. "Ischemic stroke detection using EEG signals." In *CASCON*, pp. 301-308. 2018.

[11]. Choi, Yoon-A., et al. "Machine-learning-based elderly stroke monitoring system using electroencephalography vital signals." *Applied Sciences* 11.4 (2021), pp. 1761, doi: 10.3390/app11041761

[12]. Buettner, Ricardo, Annika Grimmeisen, and Anne Gotschlich. "High-performance diagnosis of sleep disorders: a novel, accurate and fast machine learning approach using electroencephalographic data." (2020), url: http://hdl.handle.net/10125/64138

[13]. Chen, David K., Esha Sharma, and W. Curt LaFrance. "Psychogenic non-epileptic seizures." *Current neurology and neuroscience reports* 17 (2017), pp. 1-10, doi: 10.1007/s11910-017-0781-7

[14]. Kujala, Miiamaaria V. et al. "Time-resolved classification of dog brain signals reveals early processing of faces, species and emotion." *Scientific Reports* 10.1 (2020), p. 19846, doi: 10.1038/s41598-020-76806-8

[15]. Li, Bo, et al. "Rat locomotion detection based on brain functional directed connectivity from implanted electroencephalography signals." *Brain Sciences* 11.3 (2021), p. 345, doi: 10.3390/brainsci11030345

[16]. Wolpaw, Jonathan R., et al. "An EEG-based brain-computer interface for cursor control." *Electroencephalography and clinical neurophysiology* 78.3 (1991), pp. 252-259, doi: 10.1016/0013-4694(91)90040-B

[17]. X. Wan et al., "A Review on Electroencephalogram Based Brain Computer Interface for Elderly Disabled," in IEEE Access, vol. 7, pp. 36380-36387, 2019, doi: 10.1109/ACCESS.2019.2903235.

[18]. R.A. Ramadan, S. Refat., M.A. Elshahed, R.A. Ali, (2015). Basics of Brain Computer Interface. In: Hassanien, A., Azar, A. (eds) Brain-Computer Interfaces. Intelligent Systems Reference Library, vol 74. Springer, Cham, doi: 10.1007/978-3-319-10978-7_2

[19]. Värbu, Kaido, Naveed Muhammad, and Yar Muhammad. "Past, present, and future of EEG-based BCI applications." *Sensors* 22.9 (2022), p. 3331, doi: 10.3390/s22093331

[20]. Bajaj, Varun, and Ram Bilas Pachori. "Detection of human emotions using features based on the multiwavelet transform of EEG signals." *Brain-Computer Interfaces: Current Trends and Applications* (2015), pp. 215-240, doi: 10.1007/978-3-319-10978-7_8

[21]. Kumar, Pradeep, et al. "Envisioned speech recognition using EEG sensors." *Personal and Ubiquitous Computing* 22 (2018), pp. 185-199, doi: 10.1007/s00779-017-1083-4

[22]. J. Li, Y. Wang, L. Zhang and T. -P. Jung, "Combining ERPs and EEG spectral features for decoding intended movement direction," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, USA, 2012, pp. 1769-1772, doi: 10.1109/EMBC.2012.6346292.

[23]. Miner, Laurie A., Dennis J. McFarland, and Jonathan R. Wolpaw. "Answering questions with an electroencephalogram-based brain-computer interface." *Archives of physical medicine and rehabilitation* 79.9 (1998): pp. 1029-1033, ISSN 0003-9993, doi: 10.1016/S0003-9993(98)90165-4.

[24]. DaSalla, S. Charles, et al. "Single-trial classification of vowel speech imagery using common spatial patterns." *Neural networks* 22.9 (2009), pp. 1334-1339, doi: 10.1016/j.neunet.2009.05.008

[25]. P. Tirupattur., Y.S. Rawat, C. Spampinato and M. Shah 2018, October. Thoughtviz: Visualizing human thoughts using generative adversarial network. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 950-958), doi: 10.1145/3240508.3240641

[26]. J. Teo, C.L. Hou and J. Mountstephens, 2018. Preference classification using electroencephalography (EEG) and deep learning. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *10*(1-11), pp.87-91, ISSN: 2180-1843, eISSN: 2289-8131

[27]. Partha Pratim Roy, Pradeep Kumar, and Victor Chang. "A hybrid classifier combination for home automation using EEG signals." *Neural Computing and Applications* 32 (2020), pp. 16135-16147, doi: 10.1007/s00521-020-04804-y

[28]. Badcock NA, Mousikou P, Mahajan Y, de Lissa P, Thie J, McArthur G. 2013. Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ* 1:e38 https://doi.org/10.7717/peerj.38

[29]. R. W. Schafer, "What Is a Savitzky-Golay Filter? [Lecture Notes]," in *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111-117, July 2011, doi: 10.1109/MSP.2011.941097.

[30]. Costa, Ernane JX, and Euvaldo F. Cabral Jr. "EEG-based discrimination between imagination of left and right hand movements using adaptive gaussian representation." *Medical engineering & physics* 22.5 (2000): pp. 345-348, doi: 10.1016/S1350-4533(00)00051-5

[31]. K. Brigham and B. V. K. V. Kumar, "Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy," *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, Chengdu, China, 2010, pp. 1-4, doi: 10.1109/ICBBE.2010.5515807.

[32]. Nguyen, Chuong H., George K. Karavas, and Panagiotis Artemiadis. "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features." *Journal of neural engineering* 15, no. 1 (2017): p. 016002. doi: 10.1088/1741-2552/aa8235

[33]. J. Peng, S. Peng, A. Jiang, J. Wei, C. Li and J. Tan, 2010. Asymmetric least squares for multiple spectra baseline correction. *Analytica chimica acta*, *683*(1), pp.63-68, doi: 10.1016/j.aca.2010.08.033

[34]. Reddy, A. Guruva, and Srilatha Narava. "Artifact removal from EEG signals." *International Journal of Computer Applications* 77.13 (2013).

[35]. Tingbo Hou, and Hong Qin. "Continuous and discrete Mexican hat wavelet transforms on manifolds." *Graphical Models* 74.4 (2012): pp. 221-232, doi: 10.1016/j.gmod.2012.04.010

[36]. Porbadnigk, Anne, et al. "EEG-based speech recognition-impact of temporal effects." *International Conference on Bio-Inspired Systems and Signal Processing*. Vol. 1. SciTePress, 2009, doi: 10.1016/j.gmod.2012.04.010

[37]. A Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, Antiga L, Desmaison A. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems. 2019;32, doi: 10.48550/arXiv.1912.01703