

AttentionNet: Monitoring Student Attention Type in Learning with EEG-Based Measurement System

Dhruv Verma^{*†}
University of Toronto
 Toronto, Canada
 dhruvverma@cs.toronto.edu

Sejal Bhalla^{*†}
University of Toronto
 Toronto, Canada
 sejal@cs.toronto.edu

S. V. Sai Santosh^{*†}
NVIDIA Corporation
 Austin, USA
 vsiripurapu@nvidia.com

Saumya Yadav
Human-Machine Interaction Lab
IIT-Delhi
 Delhi, India
 saumya@iiitd.ac.in

Aman Parnami
WEAVE Lab
IIT-Delhi
 Delhi, India
 aman@iiitd.ac.in

Jainendra Shukla
Human-Machine Interaction Lab
IIT-Delhi
 Delhi, India
 jainendra@iiitd.ac.in

Abstract—Student attention is an indispensable input for uncovering their goals, intentions, and interests, which prove to be invaluable for a multitude of research areas, ranging from psychology to interactive systems. However, most existing methods to classify attention fail to model its complex nature. To bridge this gap, we propose AttentionNet, a novel Convolutional Neural Network-based approach that utilizes Electroencephalography (EEG) data to classify attention into five states: Selective, Sustained, Divided, Alternating, and relaxed state. We collected a dataset of 20 subjects through standard neuropsychological tasks to elicit different attentional states. The average across-student accuracy of our proposed model at this configuration is 92.3% (SD=3.04), which is well-suited for end-user applications. Our transfer learning-based approach for personalizing the model to individual subjects effectively addresses the issue of individual variability in EEG signals, resulting in improved performance and adaptability of the model for real-world applications. This represents a significant advancement in the field of EEG-based classification. Experimental results demonstrate that AttentionNet outperforms a popular EEGnet baseline (p-value < 0.05) in both subject-independent and subject-dependent settings, confirming the effectiveness of our proposed approach despite the limitations of our dataset. These results highlight the promising potential of AttentionNet for attention classification using EEG data.

Index Terms—EEG, Attention, Affective Computing, Cognitive Engagement Assessment

I. INTRODUCTION

Education plays a crucial role in the development of individuals and societies, and effective learning is essential for academic success. The field of education is constantly evolving, and with the advent of advanced technologies, there is a growing interest in exploring innovative methods to enhance educational practices and improve learning outcomes. Student engagement is a crucial factor that influences the quality of learning experiences and outcomes. Engaged students are more likely to be actively involved in their learning, motivated to learn, and able to retain and apply knowledge effectively.

* These authors contributed equally.

† This work was conducted during their affiliation with IIT-Delhi, India.

Hence, furnishing teachers with feedback on student engagement holds the utmost significance in the realm of education.

Engagement is a multifaceted construct that encompasses three essential components: behavioral engagement, cognitive engagement, and emotional engagement [1]–[3]. Emotional engagement pertains to the affective reactions exhibited by students, such as expressions of happiness or sadness, while behavioral engagement encompasses outward actions, such as asking questions and showing interest during class, as noted in previous research [4]. On the other hand, cognitive engagement involves internal neuropsychological processes, including attention and investment in learning [5]. The current study investigates the cognitive engagement of students, which refers to the extent of mental investment in learning, including traits such as reflection, persistence, and experiencing a state of flow. Some indications imply that there may be a positive association between cognitive engagement and academic performance in students [6], [7].

Attention is a crucial component of cognitive engagement in students, playing a fundamental role in their ability to focus, process information, and actively participate in learning activities. In this study, attention is considered a latent variable of engagement, and its types are determined using neuroanatomic theories, factor analysis of psychometric tests, cognitive processing, and clinical-based models [8]. The clinical-based model encompasses five distinct types of attention, each with its unique characteristics and relevance to the cognitive processes of students. It emphasizes the multi-dimensional nature of attention, making it highly relevant and appropriate for fine-grained attention modeling. Understanding these different types of attention can be beneficial in shedding light on how students engage and interact with their learning environment:

- **Alternating attention:** It reflects mental flexibility in shifting the focus of attention. It allows students to shift their focus between different tasks or stimuli, helping them to manage multiple information sources and switch

between different cognitive demands effectively. This can be particularly helpful in complex learning situations that require shifting between different activities or subjects.

- **Divided attention:** It represents the ability to respond to multiple tasks simultaneously. It enables students to allocate their cognitive resources to multiple tasks simultaneously. This can be beneficial in situations where students need to multitask or manage multiple information streams, such as when studying with digital resources or taking notes during a lecture.
- **Focused attention:** It describes attention to specific tactile, auditory, or visual stimuli. It allows students to concentrate their mental resources on a specific task or stimulus. This type of attention is crucial for deep learning and critical thinking, as it enables students to thoroughly process information, analyze it, and engage in higher-order cognitive processes.
- **Selective attention:** It depicts the ability to maintain cognitive focus amidst distracting or competing stimuli. This type of attention enables students to filter out irrelevant distractions and focus on relevant information selectively. This can help students to prioritize important information, ignore distractions, and maintain focus on the task at hand, leading to more efficient and effective learning.
- **Sustained attention:** It represents consistent attention during repetitive activities. It allows students to maintain focus and concentration over an extended period of time. This is essential for tasks that require prolonged mental effort, such as reading, problem-solving, or studying. Sustained attention helps students to stay engaged and attentive throughout the learning process, leading to better comprehension and retention of information.

Overt measures of attention in the classroom, such as tracking device usage or gaze patterns, may only capture external stimuli that grab attention, such as a ringing smartphone. However, these techniques do not consider the attentional state influenced by internal stimuli, such as thoughts. In contrast, physiological signals such as Electroencephalography (EEG), functional Near Infrared Spectroscopy (fNIRS), and Electrodermal Activity (EDA) take into account the unfiltered, involuntary changes in the nervous system, thus accounting for the covert state of mind as well [12]. Among these methods, EEG-based approaches have been extensively studied for detecting dynamic mental states due to their high temporal resolution, lack of clinical risk, affordability, and portability.

To address the challenges, we propose a novel technique leveraging multi-channel dynamics and temporal dependencies of EEG data from neuropsychological tests to comprehensively model student attention. We demonstrate its efficacy in distinguishing between different types of attention based on the Clinical Model. We conducted a user study (N=20) to evaluate the performance of our approach in detecting the nature of student engagement in accordance with the clinical model. Our study focused on a specific context and population, and we acknowledge that the findings may not be generalizable

to all situations. Since EEG data features long-range temporal dependencies, we designed a Convolutional Neural Network (CNN) with self-attention modules, termed as AttentioNet, to classify the different types of attention. CNN was selected for its capability to learn hierarchical representations from complex EEG data, capture spatial patterns, and efficiently learn from limited datasets. To summarize, the key contributions of this paper are as follows:

- Collection of EEG data from 20 subjects and 5 classes through standard neuropsychological tasks to elicit different attentional states.
- We introduce AttentioNet, a novel CNN based approach for classifying various attention types using EEG data. Inclusion of multi-channel information in the EEG data, capturing temporal and spatial dependencies for more accurate attention classification.
- Personalization of the model to individual subjects using a transfer learning-based approach, effectively addressing the issue of individual variability in EEG signals.

II. RELATED WORK

Previous research has predominantly centered on the detection of engagement using EEG [9]–[11]. While recent works have applied various algorithms on EEG data to achieve an accuracy of 89.4% in binary classification between attentive and non-attentive/relaxation states [12] and a maximum of 80.84% in 3-class classification [13]. Engagement, which is a multifaceted concept, plays a crucial role in learning and academic achievement. It's essential to recognize that engagement and attention are distinct cognitive processes, although they are often mistakenly considered the same. In the study [10], [14], [15], attention and engagement are used interchangeably as indicators of engagement. Also, in the past, there have been efforts to identify any single attention type in application-specific scenarios [16]–[18]. However, to our knowledge, just one other work [19] has attempted to classify attention according to the clinical model. It uses thermal imaging and remote eye tracking to infer attentional states but suffers from a lack of reliability since both modalities give overt measures of attention.

The conventional approach to detecting attention using EEG often involves relying on self-report or external annotator as the ground truth, where participants or external annotators are asked to report their own level of attention subjectively. However, these method has several drawbacks. Firstly, self-report can be influenced by various factors, such as social desirability bias or subjective interpretations, which may result in inaccurate or biased measurements of attention. Secondly, self-report can be influenced by subjective biases and may not always accurately reflect the true attentional state of the individual. Similarly, external annotators may introduce errors or biases in their assessments, leading to less reliable results.

These limitations highlight the need for alternative methods that can provide more objective and reliable measures of attention using EEG. To overcome this in the proposed work, standardized neuropsychological assessments are employed to

assess attention, as they comprehensively measure various aspects of attention. These assessments are typically administered by trained psychologists or clinical psychologists and involve task-based evaluations designed to measure attentional processes.

III. STUDY DESIGN

Training data was collected with ethics approval from the Institutional Review Board of Indraprastha Institute of Information Technology, Delhi, India (IIITD/IRB/9/10/2019-1). Participants underwent an experiment, as outlined in Fig. 1 of the paper. Volunteers did not receive any compensation for participating, and all of them provided informed consent before proceeding with the experiment.

A. Participants

A total of 20 healthy participants (10 females, 10 males) within the age range of 18-21 years (mean = 19.58, SD = 0.86) took part in the experiment. Every participant in the study was an undergraduate student. They all self-reported not having color blindness, a history of neurological disease, or hearing disorder, and they had normal or corrected-to-normal vision.

B. Data collection

In this research study, each participant completed a computerized experiment that involved following specific instructions. It took approximately 14 minutes for each participant to complete the test. The experiment comprised four established neuropsychological tasks that stimulate distinct types of attention, including Selective, Divided, Sustained, and Alternating attention. We omitted focused attention from the study as our primary interest lies in examining attentional patterns over extended periods [19]. EEG data and facial video recordings of the participants were recorded during the experiment. Additionally, self-reported measures of distraction (applicable to divided attention tasks only) were collected using a Likert Scale. To minimize any potential disruptions to the experimental flow, these ratings were obtained at the end of the experiment by simultaneously presenting facial and screen recordings of the trial to trigger recall. A total of 20 videos were collected, which were used to construct the AttentioNet dataset.

Fig. 1 summarizes the AttentioNet experiment. The AttentioNet experiment began with a 1-minute baseline recording with closed eyes, followed by a 1-minute baseline with open eyes for relaxation. The 30-second resting periods and 1-minute intervals for eye-closed and eye-opened states were chosen to ensure reliable data collection and consistent results. Subsequently, the subject performed the neuropsychological tasks, each separated by a 30-second resting period.

- **Stroop Test:** The Stroop test is a well-established tool for evaluating selective attention [20]. It gauges the time taken by individuals to verbally identify the font color of a word that presents an incongruent color. In our study, we utilized a computerized version of the test

where participants reported the font color using arrow keys. In the experiment, a slide deck containing the words 'red', 'green', 'blue', and 'yellow' in random order was presented to the subjects. The participants were then instructed to enter the font color of the word displayed using the corresponding keys 'r', 'g', 'b', or 'y'.

- **Reading Test:** The experiment was a recreation of the study conducted by in [16] to assess the state of divided attention. Divided attention was induced by creating an environment where readers could be distracted by external stimuli, in this case, distracting sounds. A total of five reading sessions, each lasting approximately 30 seconds, were conducted. Afterward, participants were asked to rate their level of distraction during each session using a Likert scale ranging from 1-9, with 9 indicating the highest level of distraction. It should be noted that the purpose of these self-reported ratings is to assess the participants' perception of their own distraction levels, rather than serving as the primary measure of attention in the overall study.
- **Continous Performance Test(CPT):** The current study utilized the Conners' CPT II test [21], which is a task that involves a "No-Go" CPT paradigm. It is a cognitive task that measures sustained attention and divided attention. In this task, participants are presented with a series of stimuli (letters) on a computer screen and are instructed to respond. In the experiment, participants were instructed to press the space bar whenever they see any letter other than "X", indicating a "No-Go" response for "X". The test is typically administered for a duration of approximately 3 minutes.
- **Trail Making Test-B(TMT-B):** TMT, a widely used neuropsychological assessment tool, evaluates cognitive performance and measures various types of attention [22]. TMT comprises two types: TMT-A, involving number sequencing from 1 to 15, and TMT-B, which requires set-shifting, where subjects alternate between numerical and alphabetic sequences (1-A-2-B-3...) [23]. For our study, we exclusively employed the TMT-B task to analyze participants' alternating attention. Participants were instructed to select a sequence of alphanumeric bubbles that turned green upon selection. In this assessment, participants are tasked with connecting 40 circled numbers and letters (20 of each) that are randomly distributed, following alternating numerical and alphabetical sequences, akin to 1-A-2-B-3-C... and so on, providing a challenging task for cognitive evaluation.

IV. METHODOLOGY FOR ATTENTIONNET

A. EEG Data Acquisition and Preprocessing

For our EEG recording, we utilized the Emotiv EPOC+¹ in combination with Emotiv PRO software. The EPOC+ system enabled us to acquire the raw EEG signals wireless via Bluetooth at a high sampling rate of 128

¹<http://www.emotiv.com/epoc>

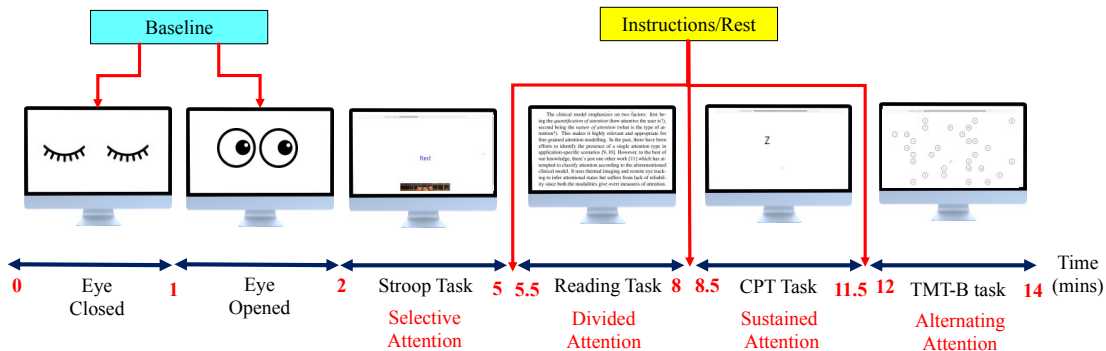


Fig. 1. Experiment timeline for AttentioNet data collection.

Hz from 14 electrodes strategically positioned on the scalp, including AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. To ensure data quality, the collected raw EEG data were bandpass filtered within the frequency range of 0.2 Hz to 45 Hz. Subsequently, the filtered data was segmented and labeled according to the different attention tasks performed by the subjects. To facilitate further analysis, we epoched the segmented data into frames of 1 second, with an overlap of 750 milliseconds, to capture potential transient changes in the EEG signals during the tasks. This meticulous approach to data pre-processing allows for a more accurate and reliable analysis of the EEG data in our study.

B. AttentioNet Architecture

In this section, we describe the architecture of our proposed model, termed as AttentioNet (see Fig. 2). Given a multi-channel frame of raw EEG signals (time steps, channels) which is called an epoch, the goal is to recognize the attention type (from relaxed, selective, sustained, alternating, and divided attention states) represented by this epoch. Recently, [24] demonstrated the use of self-attention in robust subject-independent EEG analysis. Motivated by the same, we design our network to encode the temporal and multi-channel dynamics in EEG using an attention mechanism. Our approach draws inspiration from [25], which proposes to use this technique for efficient affect recognition using EDA. The architecture consists of 4 blocks which are described as follows:

1) *Low-level Feature Extraction*: In our methodology, we begin with a low-level feature extraction (LFE) module, which plays a crucial role in capturing the fundamental characteristics of the acquired EEG data. We extract low-level feature maps X_{LFE} from the multi-channel input X using a 1-D convolutional layer followed by batch normalization. The 1-D convolutional layer operates on the multi-channel input X , which represents a multi-dimensional array with multiple channels. The convolutional layer applies filters to each channel independently, extracting features from each channel separately. This allows the model to capture unique patterns and interactions specific to each channel, while also considering the collective information across channels. Following the convolutional layer, batch normalization is applied

to normalize the output. Batch normalization is a technique that helps stabilize and accelerate the training process of deep neural networks. It normalizes the output across channels, ensuring that features extracted from different channels are on similar scales. The combination of the 1-D convolutional layer and batch normalization allows the proposed network to account for multi-channel dynamics and capture multivariate interactions.

Next, we incorporate a signal attention mechanism on the X_{LFE} to extract temporal and channel-wise features. This attention mechanism enables the model to focus on relevant regions of the EEG signals, capturing important patterns and variations that may be indicative of different cognitive states or tasks. By adaptively attending to different parts of the EEG data, the model can effectively highlight informative features for further analysis.

In the next section, we detail the architecture, including the LFE module and the signal attention mechanism, and how they collaborate to extract meaningful features from the EEG data. This comprehensive approach leverages both low-level and high-level information, resulting in a more robust and discriminative representation for subsequent analysis and interpretation.

2) *Attention Module*: After extracting low-level features, the feature map is processed using a series of sub-modules, including a channel attention module (CA) and a non-local temporal attention module (NTA).

Channel Attention Module It is widely acknowledged that distinct cognitive and affective states elicit specific responses in different regions of the brain, as evidenced by the EEG electrode activity within those regions. Therefore, investigating the inter-channel (region) relationships and identifying the channels that play a pivotal role in discriminating attention types are critical considerations.

The EEG data collected during the experiments comprises 14 channels, each of which may contribute differently to the final prediction. The channel-wise attention mechanism, proven to be portable for various tasks while maintaining excellent performance and robustness to noisy inputs, is utilized to extract the inter-channel relationships. We implement a modified version of the channel attention module inspired

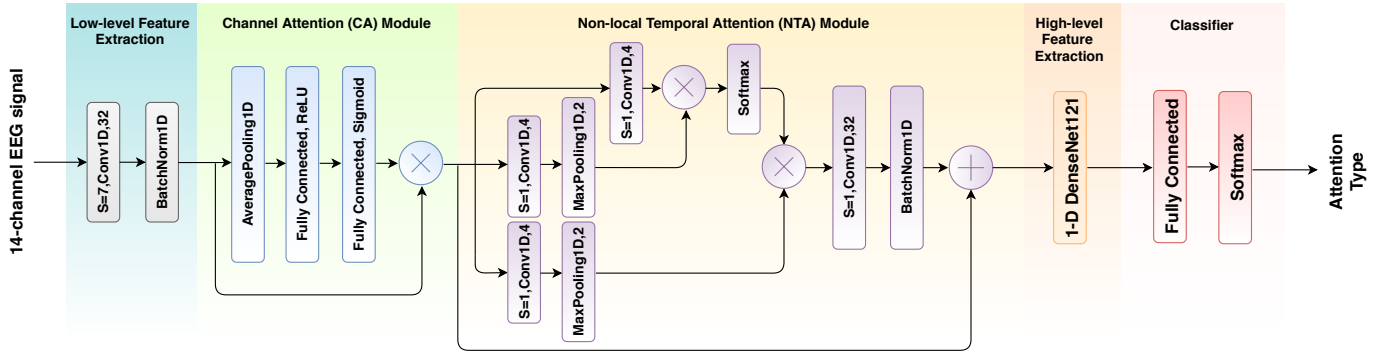


Fig. 2. Overview of the AttentionNet Architecture: A Visual Representation of the Attention-based Neural Network Model

by the Squeeze-and-Excitation Block, as proposed in reference [26], to suit the unique characteristics of EEG signals. This adaptation includes customizing 1D convolutions designed explicitly for EEG data in order to achieve our desired outcome.

We start by performing average temporal pooling on the low-level feature map X_{LFE} to extract channel-wise statistics. These statistics are then passed through a multi-layer perceptron (MLP) with 64 hidden units for further processing. ReLU activation is employed to introduce non-linearity, and Sigmoid activation is applied at the end to generate normalized channel weights ranging from 0 to 1. Finally, the original low-level feature map X_{LFE} is re-weighted by element-wise multiplication with the channel attention weights, resulting in the channel attention-enhanced feature map denoted as X_{CA} . Finally, the output of the X_{CA} module can be shown as follows:

$$X_{CA} = \text{Sigmoid}(\text{MLP}(\text{AvgPool}(X_{LFE}))) \cdot X_{LFE} \quad (1)$$

Non-local Temporal Attention Module (NTA) The acquired EEG data in our experiments is sequential, which implies that long-range samples may impact the final outcome differently at different positions. Therefore, it is crucial to investigate these long-range dependencies in the signal to gain a comprehensive understanding of the data.

In this context, non-local operations have been proven effective in various applications such as video classification, object detection, instance segmentation, and keypoint detection [27]. The generalizability and success of non-local operations in these domains have inspired us to apply them to leverage the long-range relationships in neuropsychological signals for attention-type recognition.

A generic non-local operation in a deep neural network, as shown in [27], [28]. The equation defining \hat{x}_i is as follows:

$$\hat{x}_i = \frac{1}{c(x)} \sum_{\forall_j} f(x_i, x_j)g(x_j), \quad (2)$$

In this equation, x represents the input feature map, denoted as X_{CA} in our case. The index i represents the target position in time for the output, while the index j refers to the set of all possible positions that contribute to the response x_i . The

function $f(\cdot)$ extracts the relationship between index i and all possible positions j , while $g(\cdot)$ re-weights the input features. The function $c(\cdot)$ represents a normalization function applied before obtaining the final output.

In this work, we use the convolution operation coupled with a max-pooling operation for the linear operation $g(\cdot)$ as:

$$g(x_j) = \text{MaxPool}(W_g x_j), \quad (3)$$

where W_g represents the convolution kernel. For the pairwise operation $f(\cdot)$, we apply the commonly used Embedded Gaussian formulation [29] (a widely used measure for computing similarity in an embedding space) to define $f(\cdot)$ as:

$$f(x_i, x_j) = e^{\theta x_i^T \phi x_j} \quad (4)$$

where $\theta(\cdot)$ and $\phi(\cdot)$ represent convolution operations. For the normalization operator $c(\cdot)$, we adopt the adaptation from reference [27], where it is set as $\sum_{\forall_j} f(x_i, x_j)$, and the final pairwise operation is performed using the softmax operation $\frac{1}{c(x)} f(x_i, x_j)$. The overall operation of the NTA block can be expressed as:

$$X_{NL} = \text{Softmax}(X_{CA}^T W_\theta^T \text{MaxPool}(W_\phi X_{CA})) \text{MaxPool}(W_g X_{CA}) \quad (5)$$

where W_θ and W_ϕ are convolution kernels and finally, the output of the NTA module can be shown as:

$$X_{NTA} = W_w F_{NL} + F_{CA} \quad (6)$$

Where W_w denotes the 1D convolution operation followed by batch normalization. The output X_{NTA} refers to the re-weighted result obtained from the Attention Module.

3) *High-level Feature Extraction*: We employ DenseNet-121 [30] as the backbone for extracting high-level feature maps, as it offers improved gradient flow and tractable optimization. It is worth noting that other state-of-the-art feature extraction backbones such as VGG, ResNet, or GoogLeNet exist, but a comparative analysis of these networks is not the primary focus of this work. To make DenseNet-121

²A detailed diagram for the modified DenseNet-121 is skipped due to space constraints.

compatible with the input signal features, we modify its original architecture by replacing the 2D convolutions with 1D convolutions. The backbone takes the input X_{NTA} and produces a feature vector denoted as X_{HFE} .

4) *Classifier*: Finally, the feature vector X_{HFE} is fed into a Fully Connected Layer with Softmax activation, consisting of 5 units, for the purpose of classifying attention into their respective types.

C. Model Training

After pre-processing the data (as described in Section 3.C), we obtained a total of 62,530 windows, each with a length of 128 samples, across 20 subjects and five classes (as detailed in Section 3.D). To train our neural network, a leave-one-subject-out scheme was adopted, where the data of one subject was reserved for validation while the remaining subjects' data was utilized for training. The Adam optimizer [31] was employed, with a learning rate of 0.001, and was decayed by a factor of 10 if the validation loss plateaued after an epoch. The Categorical Cross-entropy function served as the loss function. Mini-batches of size 32 were used for training, with model checkpoints saved at the epoch with the highest validation accuracy.

V. EXPERIMENT RESULTS

To assess the performance of our approach in detecting student attention type, we conducted experiments in two different settings: subject-independent and subject-dependent (adaptive) evaluations. Additionally, since this is the first attempt to categorize attention using EEG data, we compare our model's performance to EEGNet [32].

EEGNet is a architecture specifically designed for analyzing EEG data. It addresses the unique challenges posed by EEG signals, such as non-stationarity and temporal dependencies, by incorporating depthwise and separable convolutions. These convolutions efficiently capture spatial and temporal patterns in the EEG data. Moreover, EEGNet employs a parallel filter-bank structure to process different frequency bands separately, enabling the model to extract multi-scale information present in EEG signals effectively.

Due to its effectiveness in handling EEG data, EEGNet is widely used in EEG-based classification tasks, including motor imagery classification [33], emotion recognition [34], and various cognitive tasks [32]. Researchers often use EEGNet as a benchmark model to evaluate and compare the performance of their proposed approaches in EEG data analysis.

A. Subject Independent Evaluation

To evaluate the model's performance without subject-specific data, we employed a leave-one-subject-out (LOSO) cross-validation approach. The dataset was split based on individual subjects, where data from 19 participants were used for training, and the data of one participant was held out for validation. This process was repeated for each subject, ensuring a comprehensive assessment of the model's performance across different subjects.

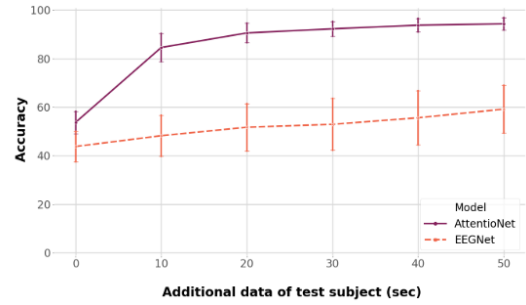


Fig. 3. Performance of AttentionNet and EEGNet tuned with different amounts of data from the test subject. Error bar indicates the standard error.

The average LOSO accuracy for AttentionNet in the 5-class attention classification task was 53.75% (SD=4.58, chance=20%), while for EEGNet, it was 43.8% (SD=6.20). Our approach significantly outperformed EEGNet, and we confirmed this with a Students' T-test ($p : 1.5 \times 10^{-6}$).

However, we recognized that considering the inherent variability of EEG signals across participants could lead to further improvements in performance. As a result, we decided to adopt a subject-dependent (adaptive) approach, enabling us to personalize the model for each individual. This adaptability takes into account the unique characteristics of each subject's EEG data, which has been demonstrated to lead to significant enhancements in the model's performance [35].

B. Subject Dependent Evaluation

To account for individual variability, we employed a transfer learning-based approach using the LOSO cross-validation method. For each left-out test subject, we fine-tuned a base model pre-trained on the data from the remaining subjects. This fine-tuning process involved using a subset of the left-out subject's data. To determine the optimal size of this subset, indicating the number of samples needed for effective personalization, we incrementally added 10 seconds of the subject's data per class. After each step, we fine-tuned the model and evaluated the performance of AttentionNet in comparison to EEGNet.

As depicted in Fig. 3, the transfer learning approach led to a significant improvement in performance for both networks, with AttentionNet exhibiting a notably higher increase. Notably, using 30 seconds of data per class was found to be sufficient to personalize the model for any given subject. At this configuration, the average across-subject accuracy of AttentionNet was 92.3% (SD=3.04), making it well-suited for end-user applications. Any further increase in the tuning data beyond this point resulted in an insignificant gain in accuracy, suggesting that additional data can be traded off to minimize the data required for fine-tuning the base model.

Fig. 4 summarizes the performance of our approach and EEGNet, each tuned according to the aforementioned optimal configuration, in a 5-class classification task for each subject. This analysis highlights the effectiveness of our subject-

dependent adaptive approach in achieving highly accurate attention classification results across various individuals.

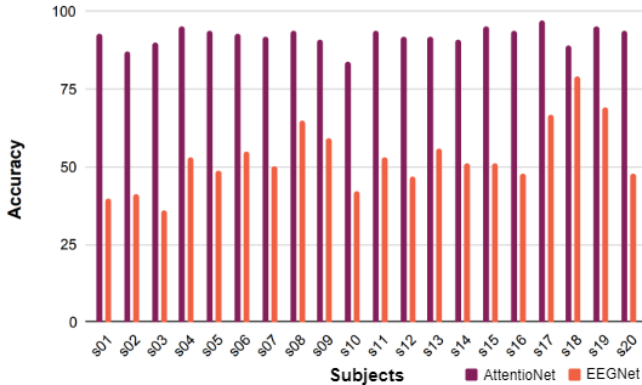


Fig. 4. Comparison of EEGNet and AttentioNet, tuned with 30 seconds of subject data per class, for each subject.

VI. CONCLUSION

In this work, we demonstrate the effectiveness of a novel attention classification mechanism that utilizes deep neural networks to analyze EEG data and differentiate between different attention states, as specified by the clinical model of attention. Unlike previous research that treated attention as a unidimensional variable, our approach acknowledges the multidimensional nature of human attention. To evaluate the performance of our approach, we collected data from 20 students who performed a range of controlled neuropsychological tasks, specifically designed to elicit distinct attention states. For classifying attention types, we introduced AttentioNet, a CNN-based architecture that incorporates non-local temporal and channel-wise attention mechanisms. This integration enables robust classification of attention types using EEG data. Comparing our model’s performance to EEGNet, a well-established EEG baseline, we observed that AttentioNet outperformed EEGNet with average accuracies of 53.75% and 92.3% in subject-independent and subject-dependent settings, respectively (the latter achieved through personalization with 30 seconds of EEG data). Furthermore, our subject-dependent analysis highlighted the importance of personalization in EEG-based attention monitoring, underscoring its potential for enhancing performance in real-world attention-based applications.

VII. LIMITATIONS AND FUTURE WORK

As with any study, our research also comes with certain limitations that warrant discussion and consideration.

Firstly, our dataset comprises a relatively small number of subjects, primarily consisting of students. To develop a more widely applicable and generalizable system, future work should focus on expanding the dataset to include a more diverse range of subjects.

Secondly, the controlled settings used to evoke attentional states may not fully replicate the complexities of real-world

situations. Addressing this limitation involves conducting experiments in more naturalistic settings to capture user responses in authentic, everyday scenarios.

Thirdly, our data collection was confined to a single session. To achieve a more comprehensive evaluation of personalized algorithms, future research should incorporate data from multiple sessions to assess performance more holistically.

Despite these limitations, our findings illustrate the promising potential of EEG-based systems for attention monitoring and classification. Looking ahead, we envision integrating our approach with portable EEG devices like AttentivU glasses [11], facilitating real-time attention measurement. Additionally, by incorporating advanced sensing technologies, our work has the potential to provide enhanced insights into teaching, thereby improving learning outcomes in educational settings.

VIII. ACKNOWLEDGEMENTS.

This research was supported by the Centre for Design and New Media (a TCS Foundation Initiative supported by Tata Consultancy Services) and the Infosys Centre for Artificial Intelligence at IIIT-Delhi, India. Gratitude is also extended to all the participants who contributed to this study.

IX. ETHICAL IMPACT STATEMENT

Our research delves into the use of deep neural networks to analyze human cognitive states, particularly attention, with a focus on young populations like students. Consequently, our work extends beyond technological advancements and raises various ethical implications that demand careful consideration.

Firstly, the collection of EEG data from human subjects for research purposes necessitates strict adherence to ethical guidelines. This includes obtaining informed consent from participants, safeguarding their confidentiality, and ensuring minimal discomfort or harm during data acquisition. To ensure ethical compliance, our study was approved by the IRB at our institution.

Secondly, we recognize the potential impact on privacy and data security. EEG data is sensitive and can reveal personal information about an individual’s cognitive processes and attentional states. To safeguard privacy, measures were implemented, such as data anonymization, secure storage, and restricted access to the data.

Thirdly, the limitations of the dataset used for training and evaluation must be acknowledged, as it may not fully represent the diversity of the general population in terms of age, gender, ethnicity, and other demographic factors. Efforts were made to carefully interpret the findings and avoid overgeneralization.

Moreover, the potential for bias in the proposed model has been carefully evaluated. Convolutional Neural Networks (CNNs) are known to learn patterns from data, and if the training data is biased, it can lead to biased predictions. Thus, the model’s performance across different demographic groups was thoroughly assessed and ensured that it does not perpetuate existing biases or discriminate against specific populations.

Furthermore, the intended use and application of AttentionNet have been considered. If AttentionNet is used in real-world settings for monitoring and classifying attention states, it may raise concerns about surveillance, autonomy, and fairness. Ethical considerations will be taken into account when deploying AttentionNet in practical applications.

In conclusion, while AttentionNet represents a significant advancement in EEG-based attention classification, it is imperative to adhere to ethical principles throughout the research process. Transparent reporting of limitations, potential biases, and ethical considerations is crucial to ensuring the responsible and ethical use of this technology. By upholding ethical standards, we strive to maximize the positive impact of our research while minimizing potential risks or unintended consequences.

REFERENCES

- [1] S. D’Mello, E. Dieterle, and A. Duckworth, “Advanced, analytic, automated (aaa) measurement of engagement during learning,” *Educational psychologist*, vol. 52, no. 2, pp. 104–123, 2017.
- [2] R. Azevedo, “Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues,” *Educational psychologist*, vol. 50, no. 1, pp. 84–94, 2015.
- [3] K. Xie, B. C. Heddy, and B. A. Greene, “Affordances of using mobile technology to support experience-sampling method in examining college students’ engagement,” *Computers & Education*, vol. 128, pp. 183–198, 2019.
- [4] K. A. Renninger and J. E. Bachrach, “Studying triggers for interest and engagement using observational methods,” *Educational Psychologist*, vol. 50, no. 1, pp. 58–69, 2015.
- [5] S. Lackmann, P.-M. Léger, P. Charland, C. Aubé, and J. Talbot, “The influence of video format on engagement and performance in online learning,” *Brain Sciences*, vol. 11, no. 2, p. 128, 2021.
- [6] B. W. Miller, “Using reading times and eye-movements to measure cognitive engagement,” *Educational psychologist*, vol. 50, no. 1, pp. 31–42, 2015.
- [7] B. A. Greene, “Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research,” *Educational Psychologist*, vol. 50, no. 1, pp. 14–30, 2015.
- [8] M. M. Sohlberg and C. A. Mateer, “Effectiveness of an attention-training program,” *Journal of clinical and experimental neuropsychology*, vol. 9, no. 2, pp. 117–130, 1987.
- [9] M. Hassib, S. Schneegass, P. Eiglsperger, N. Henze, A. Schmidt, and F. Alt, “Engagemeter: A system for implicit audience engagement sensing using electroencephalography,” in *Proceedings of the 2017 Chi conference on human factors in computing systems*, 2017, pp. 5114–5119.
- [10] Ç. İ. Açı, M. Kaya, and Y. Mishchenko, “Distinguishing mental attention states of humans via an eeg-based passive bci using machine learning methods,” *Expert Systems with Applications*, vol. 134, pp. 153–166, 2019.
- [11] N. Kosmyna, C. Morris, U. Sarawgi, and P. Maes, “Attentivu: a biofeedback system for real-time monitoring and improvement of engagement,” in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–2.
- [12] B. Hamadicharef, H. Zhang, C. Guan, C. Wang, K. S. Phua, K. P. Tee, and K. K. Ang, “Learning eeg-based spectral-spatial patterns for attention level measurement,” in *2009 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2009, pp. 1465–1468.
- [13] B. Hu, X. Li, S. Sun, and M. Ratcliffe, “Attention recognition in eeg-based affective learning research using cfs+ knn algorithm,” *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 1, pp. 38–45, 2016.
- [14] A. Al-Nafjan and M. Aldayel, “Predict students’ attention in online learning using eeg data,” *Sustainability*, vol. 14, no. 11, p. 6553, 2022.
- [15] D. Szafir and B. Mutlu, “Pay attention! designing adaptive agents that monitor and improve user engagement,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 11–20.
- [16] M. Rodrigue, J. Son, B. Giesbrecht, M. Turk, and T. Höllerer, “Spatio-temporal detection of divided attention in reading applications using eeg and eye tracking,” in *Proceedings of the 20th international conference on intelligent user interfaces*, 2015, pp. 121–125.
- [17] Z. Mohamed, M. El Halaby, T. Said, D. Shawky, and A. Badawi, “Characterizing focused attention and working memory using eeg,” *Sensors*, vol. 18, no. 11, p. 3743, 2018.
- [18] L.-W. Ko, O. Komarov, W. D. Hairston, T.-P. Jung, and C.-T. Lin, “Sustained attention in real classroom settings: An eeg study,” *Frontiers in human neuroscience*, vol. 11, p. 388, 2017.
- [19] Y. Abdelrahman, A. A. Khan, J. Newn, E. Velloso, S. A. Safwat, J. Bailey, A. Bulling, F. Vetere, and A. Schmidt, “Classifying attention types with thermal imaging and eye tracking,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–27, 2019.
- [20] B. C. Wright, “What stroop tasks can tell us about selective attention from childhood to adulthood,” *British Journal of Psychology*, vol. 108, no. 3, pp. 583–607, 2017.
- [21] F. Ghassemi, M. H. Moradi, M. T. Doust, and V. Abootelebi, “Classification of sustained attention level based on morphological features of eeg’s independent components,” in *2009 ICME International Conference on Complex Medical Engineering*. IEEE, 2009, pp. 1–6.
- [22] M. M. Sohlberg and C. A. Mateer, “Improving attention and managing attentional problems: Adapting rehabilitation techniques to adults with add,” *Annals of the New York Academy of Sciences*, vol. 931, no. 1, pp. 359–375, 2001.
- [23] A. Prange, M. Barz, A. Heimann-Steinert, and D. Sonntag, “Explainable automatic evaluation of the trail making test for dementia screening,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–9.
- [24] D. Zhang, L. Yao, K. Chen, and J. Monaghan, “A convolutional recurrent attention model for subject-independent eeg signal analysis,” *IEEE Signal Processing Letters*, vol. 26, no. 5, pp. 715–719, 2019.
- [25] G. Yin, S. Sun, D. Yu, D. Li, and K. Zhang, “A efficient multimodal framework for large scale emotion recognition by fusing music and electrodermal activity signals,” *arXiv preprint arXiv:2008.09743*, 2020.
- [26] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [27] X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.
- [28] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, vol. 2. Ieee, 2005, pp. 60–65.
- [29] 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.
- [30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [31] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [32] 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,” *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.
- [33] A. C. Hernandez-Ruiz, D. Enériz, N. Medrano, and B. Calvo, “Motor-imagery eegnet-based processing on a low-spec soc hardware,” in *2021 IEEE Sensors*, 2021, pp. 1–4.
- [34] M. Tunnell, H. Chung, and Y. Chang, “A novel convolutional neural network for emotion recognition using neurophysiological signals,” in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 792–797.
- [35] D. Nath, Anubhav, M. Singh, D. Sethia, D. Kalra, and S. Indu, “A comparative study of subject-dependent and subject-independent strategies for eeg-based emotion recognition using lstm network,” in *Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis*, ser. ICCDA 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 142–147. [Online]. Available: <https://doi.org/10.1145/3388142.3388167>