HIGH SCHOOL EDITION
Journal of Student Research

Volume 14 Issue 1 (2025)

Machine Learning Approaches for Electroencephalography-Based Biomarker Discovery in Autism Spectrum Disorder

Sooeun Ban[1] and Young ui Min[#]

[1]Hankuk academy of Foreign Studies, Republic of Korea
[#]Advisor

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, and repetitive behaviors. Electroencephalography (EEG), with its high temporal resolution, offers valuable insights into the neural dynamics associated with ASD. This paper proposes a novel convolutional neural network architecture designed to enhance the accuracy of ASD screening by separately extracting spatio-temporal features from EEG signals. The network is structured into three main modules: preprocessing, feature extraction, and ASD screening. Initially, EEG signals are transformed into topological maps to serve as input for the CNN. The feature extraction module then processes these maps to independently capture spatial and temporal features, which are subsequently aggregated and used by the ASD screening network. Additionally, the model incorporates demographic information as auxiliary input to improve screening accuracy. We also conducted brain anatomy-driven experiments, filtering out specific brain regions from the input EEG signals to determine which parts are most highly correlated with ASD screening. The ablation study demonstrated the effectiveness of the proposed spatio-temporal approach, and additional t-SNE evaluations further validated its robustness. The proposed network achieved state-of-the-art performance, with an accuracy of 97.09% on a public dataset.

Introduction

ASD (Autism Spectrum Disorder) is a neurological and developmental disorder that affects how people interact with others, communicate, learn, and behave. Autism has become significantly prevalent in recent years. While 6.7 in 1000 children were diagnosed with ASD in 2000, the number had risen to 27.6 by 2020. Early diagnosis of autism is crucial for early intervention, as it has been shown that early diagnosis allows for enhancement of developmental outcomes and improved adaptive skills. Additionally, early intervention makes it possible for caregivers to provide specialized interventions tailored for specific needs.

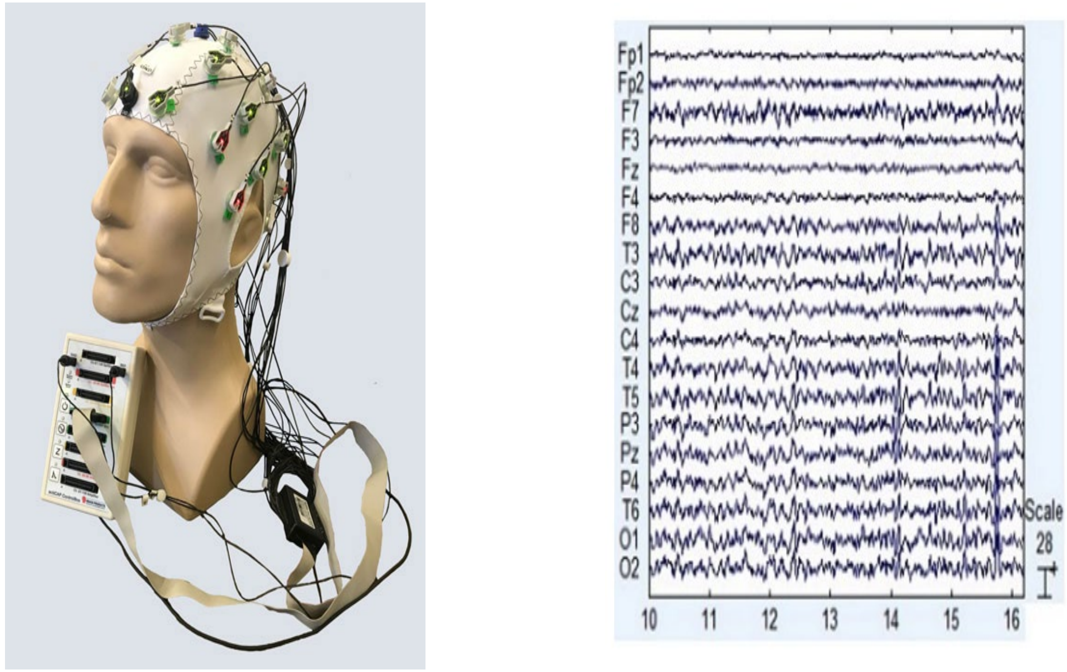
However, while characteristics that hint autism may be detected in early childhood, autism is often not diagnosed until much earlier. Since autism is highly heterogeneous and its symptoms vary from one patient to another, there are many difficulties in finding a biomarker for autism. Autism also has high comorbidity rates, making it harder to identify autism at an early age. Identifying a biomarker that can diagnose ASD at an early stage with certainty has therefore become a crucial task for dealing with ASD. Although there have been several approaches to identify a biomarker using EEG and machine learning, there were a lot of limitations because the environment of filming EEGs often varied from case to case. There also exists a problem that there is a lack of data and that data is non stationary.

To address this challenge, I propose a novel convolutional neural network architecture that separately extracts spatio-temporal features to enhance the accuracy of ASD screening. The proposed network comprises three modules: preprocessing, feature extraction, and ASD screening. First, the EEG signals are converted into topological maps to be input into the convolutional neural network. The feature extraction module processes these topological maps to extract spatio-temporal features independently. These features are then aggregated and passed to the subsequent ASD screening network. Additionally, I introduce an approach that incorporates demographic information as auxiliary input to the ASD screening network which aims to further improve the screening accuracy.

The structure of the following chapters in this paper is as follows: Chapter 2 provides an overview of EEG and CNNs to enhance understanding. Chapter 3 details the proposed system, including network architectures and the loss function. Chapter 4 presents the experimental results, and Chapter 5 summarizes the findings of the paper.

Background Knowledge

Electroencephalography



(a) (b)

Figure 1. (a): EEG cap (A1 Medical Integration 2024) and (b): EEG signal (Saccá et al. 2018)

An electroencephalography (EEG) device is a non-invasive tool used to measure electrical activity in the brain. As shown in Figure 1 (a), It consists of multiple electrodes placed on the scalp, which detect voltage fluctuations resulting from ionic current flows within neurons. These electrical signals are recorded as waveforms, representing various brain wave frequencies, such as delta, theta, alpha, beta, and gamma waves (Übeyli 2009). The recorded EEG data provide a high temporal resolution view of brain activity, making it valuable for examining the dynamic processes of the brain in real-time (Figure 1 (b)). The analysis of EEG signals involves preprocessing to remove noise and artifacts, followed by feature extraction to identify meaningful patterns related to brain function.

In the context of Autism Spectrum Disorder (ASD) screening, EEG signals can serve as effective biomarkers for identifying ASD. ASD is associated with atypical neural connectivity and brain activity, which can be reflected in EEG patterns. For instance, individuals with ASD may exhibit differences in the power and coherence of specific frequency bands, altered event-related potentials, and variations in resting-state brain activity. By analyzing these

EEG features using machine learning algorithms, researchers can develop models to distinguish between individuals with ASD and typically developing individuals (Wadhera 2021). The ability to non-invasively monitor brain activity through EEG offers a promising avenue for early diagnosis and intervention in ASD, potentially leading to better outcomes through timely support and treatment.

Convolutioanl Neural Network

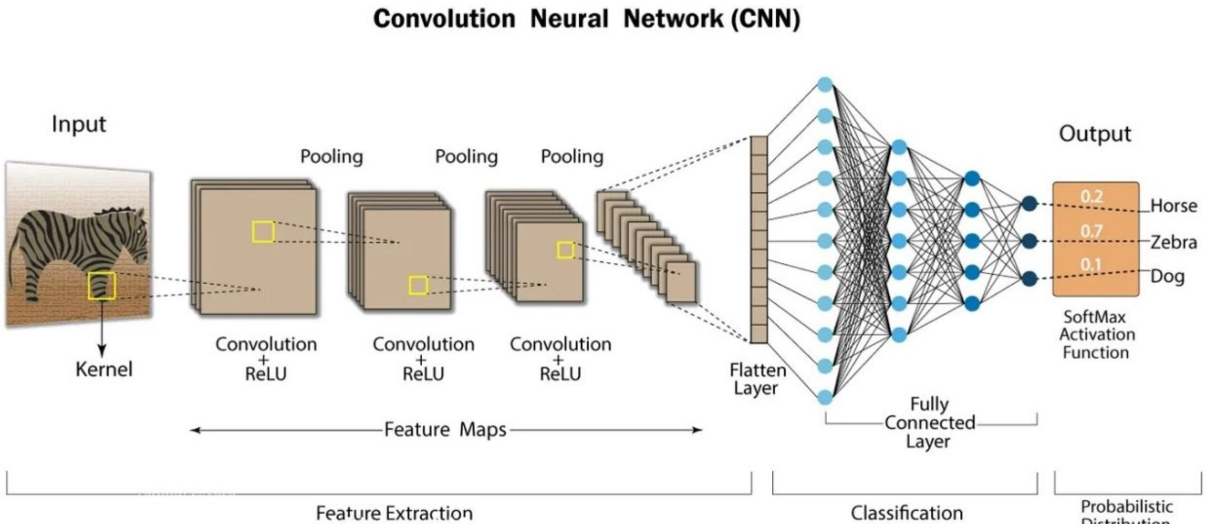


Figure 2. Convolutional neural network architecture pattern (Haque 2023)

CNN (Convolutioanl Neural Network) is a machine learning architecture used for image analysis. A main component of a CNN network is convolve operation. When given an image as the input, the model converts the image into a 3-dimensional vector(3xHxW) and convolves the filter with the image, which means to start from the upper left part of the vector, slide over the image spatially and perform pixel-wise multiplications with a filter(a 3-dimensional vector with height and depth set as hyperparameters;to prevent feature maps from shrinking, a padding is used which adds zeroes around the input)

CNNs are widely applied in various fields due to their exceptional ability to automatically and hierarchically extract features from raw data. In medical imaging, for instance, CNNs are used to detect and classify abnormalities in X-rays, MRIs, and CT scans with high accuracy. In natural language processing, CNNs can capture local dependencies in text for tasks such as sentiment analysis and text classification. In the domain of EEG signal analysis, CNNs are particularly effective as they can learn spatial and temporal patterns associated with different brain states which makes them valuable for applications such as diagnosing neurological disorders.

Electroencephalography-Based Autism Spectrum Disorder Biomarker

Preprocessing

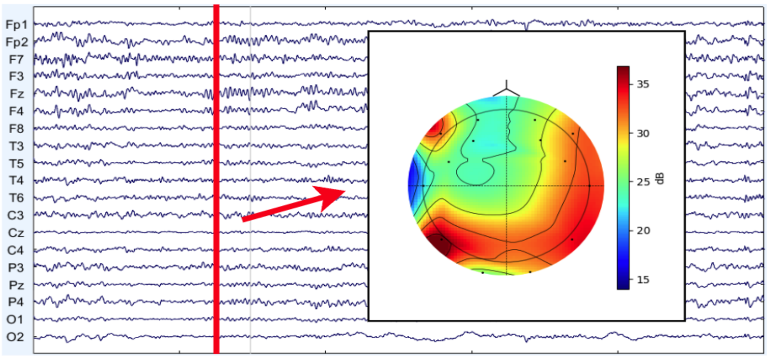


Figure 3. Illustration of preprcoessing (topological map)

Figure 3 shows how EEG data is converted into an image and passed onto the Autiscope. To conduct an EEG, electrodes are pasted onto one's scalp. The waves show how much signal each electrode receives. EEG then converts these signals to an image of a topological map. In each timestep, EEG performs this conversion, and the resulting images are passed onto the feature extraction step.

Feature Extraction

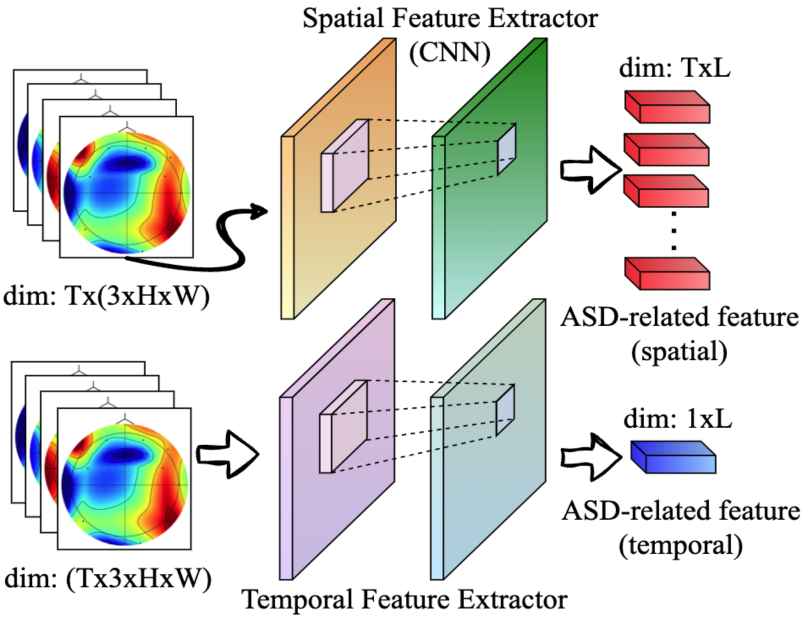


Figure 4. Architecture of the feature extraction module

With images that depict how much signal each part of the brain produces in each timestep, we can now extract features from the images. We will divide features into spatial and temporal features and extract them with a CNN.

Spatial features represent the spatial structures of an image in a certain timestep. To extract spatial features, we will perform convolve operations on each of the images to make a TxL ASD-related feature. To extract temporal features that recognize changes according to time, we will perform another convolve operation with a Tx3xHxW dimension input this time, which gives as a 1 x L ASD-related feature. These features are then passed onto the model for disorder classification.

Autism Spectrum Disorder Classification

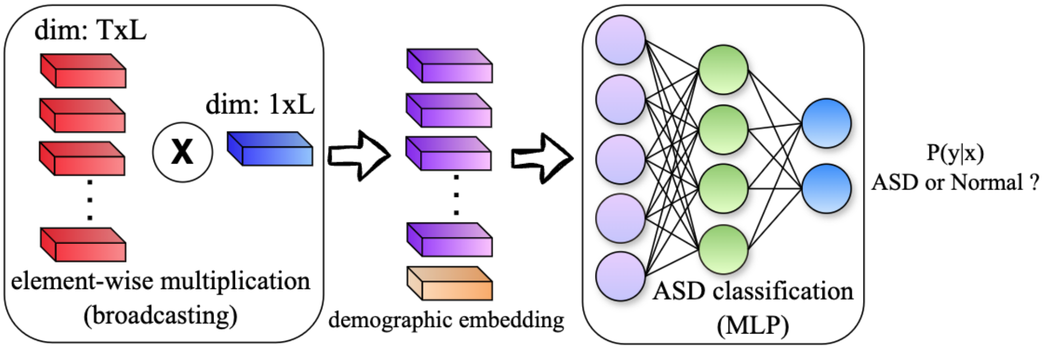


Figure 5. Architecture of the autism spectrum disorder classification module

The spatial and temporal features that are extracted from the preprocessed images should then be merged for them to be adequate for CNN. To do this, we perform an element wise multiplication between spatial and temporal features.

To address the problem that EEG results often differ by person, we can add a demographic embedding that distinguishes demographics such as age and sex.

The result of the element-wise multiplication of the spatial and temporal features, along with the demographic embeddings, are passed onto a neural network as inputs. The neural network then gives the probability that one does not have ASD as an output.

Equation 1: Feature aggregation

$$z(t,l) = z_S(t,l) \times z_T(l) \ \forall \ \ 1 \leq t \leq T \ and \ 1 \leq l \leq L$$

Here, Zs is the spatial features and Zt is the temporal features. The product of a certain input in a spatial feature(1 ≤ l ≤ L) in a timestep with a corresponding input in a time feature is iteratively measured to aggregate spatial and temporal features.

Equation 2: Binary cross entropy loss function

$$J(w,b) = -[\ y \times log_e\big(Autiscope(x;w,b)\big) + (1-y) \times log_e\big(1 - Autiscope(x;w,b)\big)\ ]$$

Here, J(w, b) is set as the binary cross-entropy loss function of the ground-truth and the prediction made by the autiscope. As the prediction is made in a form of binary, either y or (1-y) becomes 0 and erases one part of the formula. The predictions made by an autiscope with parameters set as w and b put into a log function is multiplied by a ground-truth.

Experimental Results

Dataset

The model used a dataset collected from Severance Hospital, Seoul (AI Hub 2024). It contained 250 samples from the control group and 250 samples from patients with ASD. Among the total patients, children aged 0 to 7 accounted for 40.7% of the data set, with those aged 7 to 12 and 13 to 18 accounting for 32.93% and 26.80%, respectively.

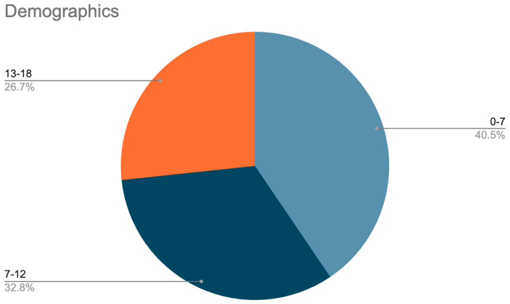


Figure 6. Demographic information of the dataset

Evaluation Metric

Evaluation metrics used in the research are accuracy, precision, recall, and F1-score. Accuracy measures proportion of correct predictions to the total number of predictions. Precision is a metric used to determine how many of the positive predictions made by the model are true positives. Recall is used to measure how many of the positive ground-truths the model successfully detects as positive. F1-score is a harmonic mean of precision and recall scores.

Equation 3: Accuracy

$$Accuracy = \frac{(correct\ predictions)}{(total\ number\ of\ predictions)}$$

Equation 4: Precision

$$Precision = \frac{(True\ positive)}{(True\ positive + False\ positive)}$$

Equation 5: Recall

$$Recall = \frac{(True\ positive)}{(True\ positive + False\ negative)}$$

Equation 6: F1-Score

$$F1 - score = \frac{2 \cdot (Recall \cdot Precision)}{(Recall + Precision)}$$

Performance Comparison

In this chapter, we evaluate the performance of the proposed network by training four different models with varying setups. First, we trained the ResNet-152 architecture (He et al. 2016) without any modifications, which we refer to as the baseline. For the second model, we included demographic information during training and testing, referring to this setup as Experiment-1. The third model utilizes the proposed spatio-temporal feature extraction method explained in Chapter 3, referred to as Experiment-3. Lastly, we trained and evaluated the full model, which incorporates both spatio-temporal feature extraction and demographic information, referring to this setup as Experiment-4.

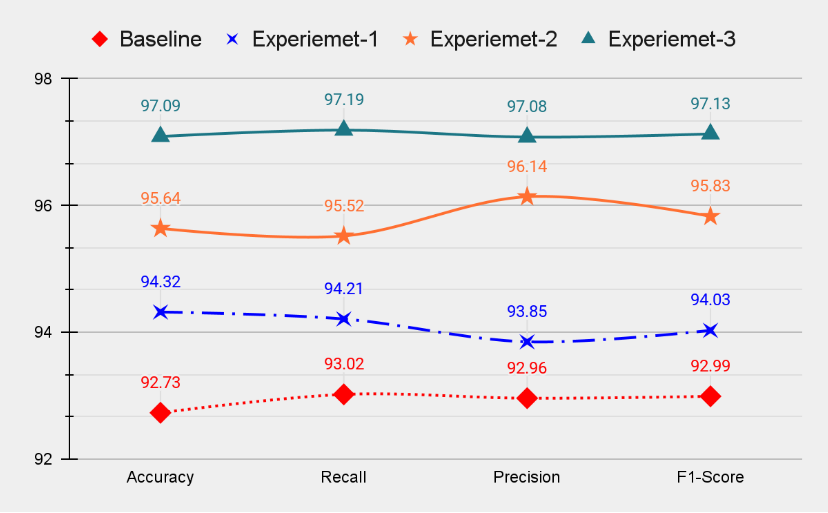


Figure 7. Performance comparison (ablation study)

Figure 7 illustrates the performance comparison of the four different experimental models. The Experiment-1 model achieved an F1-score of 94.03, which is 1.05 points higher than the baseline, demonstrating that incorporating demographic information enhances the accuracy of the trained model. The Experiment-3 model achieved an F1-score 2.84 points higher than the baseline which indicates that the proposed spatio-temporal feature extraction significantly improves accuracy. Finally, Experiment-4, which utilizes both approaches, achieved the highest accuracy with an F1-score of 97.13 which shows the effectiveness of combining spatio-temporal feature extraction and demographic information.

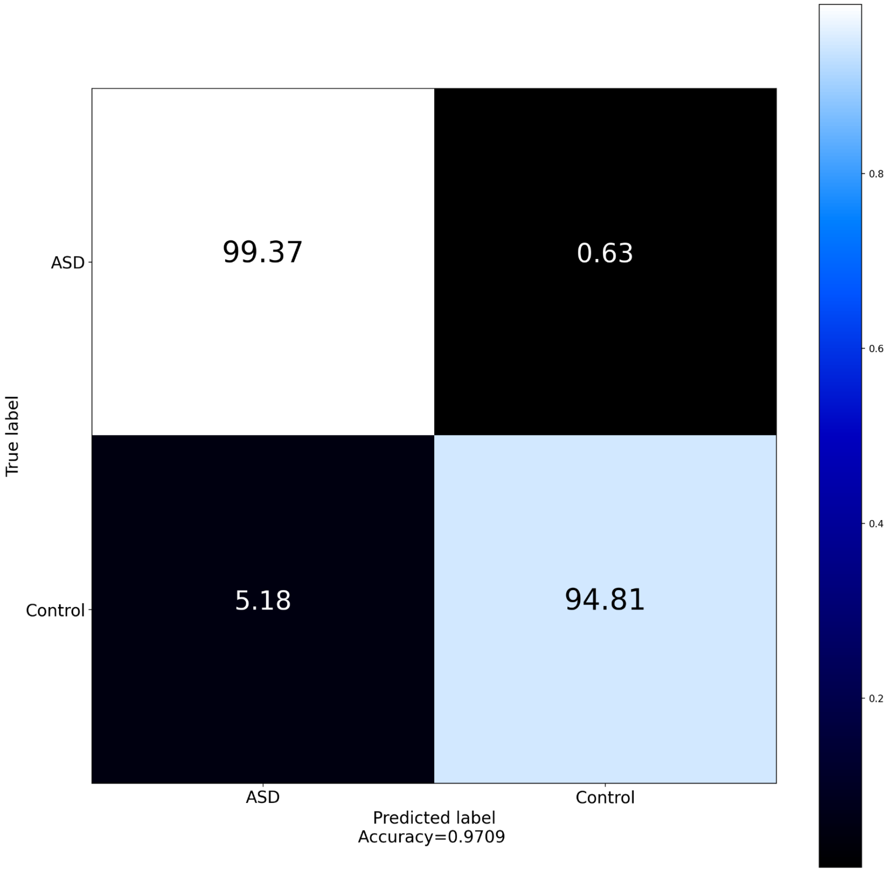


Figure 8. Confusion matrix

Figure 8 demonstrates the confusion matrix evaluation. The proposed network achieved a true positive ratio of 99.37% and a true negative ratio of 94.81%. The false positive ratio is 5.18% which indicates that the model tends to over-predict samples as ASD. This is acceptable in the medical domain, where it is generally preferred to err on the side of over-predicting positive cases, as this allows for further investigation and reduces the risk of missed diagnoses.

Brain Anatomy-Driven Experiment

In addition, we conducted brain anatomy-driven experiments to examine which brain regions are most highly correlated with ASD. We trained and evaluated four models with different experimental setups, where specific channels were set to zero for particular electrodes. For example, the model labeled "w/o F" in Figure 9 represents the model trained with EEG samples excluding the frontal lobe. Similarly, "T" refers to the temporal lobe, while "P" and "O" represent the parietal and occipital lobes, respectively.

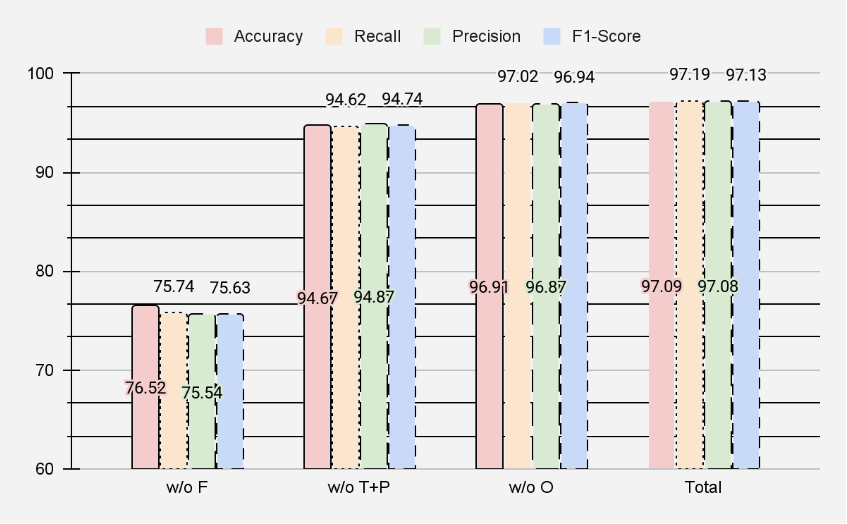
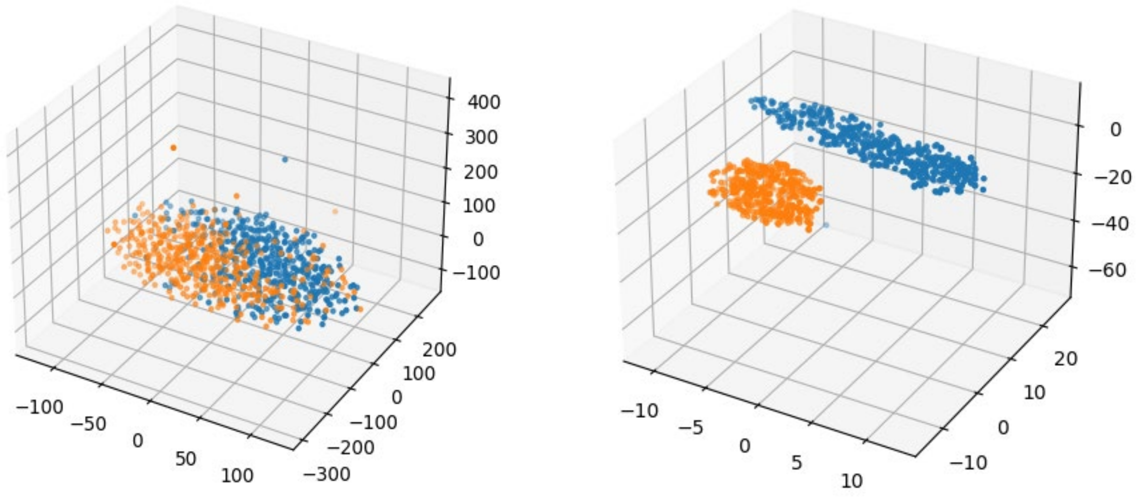


Figure 9. Brain anatomy-driven experiment results

The model trained without the frontal lobe achieved the lowest accuracy, with an F1-score of 75.63. This noticeable performance gap highlights the importance of the frontal lobe in screening for ASD. The model trained without the temporal and parietal lobes achieved an F1-score of 94.74, which is 2.49 points lower than the model trained with full EEG samples. Interestingly, the model trained without the occipital lobe showed no performance drop, indicating that the occipital lobe is not highly correlated with ASD screening.



(a)

(b)

Figure 10. t-SNE evaluation
(a): baseline and (b): proposed approach

Finally, we conducted a t-SNE (Van der Maaten and Hinton, 2008) evaluation to visually assess the effectiveness of the proposed approach. Figure 10(a) shows the feature map visualization of the baseline, as explained in Section 4.3. Each dot represents an EEG sample, with different colors indicating positive and negative labels. The features of each label are entangled, making it difficult to clearly separate them in the feature space. In contrast, Figure 10(b) displays the feature map visualization of the proposed approach, where the features of each label are more distinctly separated compared to the baseline. This result clearly demonstrates the effectiveness of the proposed approach in enhancing feature separation.

Conclusion

In this research, I proposed using a CNN model for analyzing the results of an EEG and identifying the biomarkers crucial for early diagnosis of ASD. More specifically, the contributions were that types of features were divided into spatial and temporal features to reflect the changes in brain waves and that demographic embedding were added for medical preciseness. After going through multiple evaluations like ablation analysis, confusion matrix, and t-sne, it was proved that using both demographic embeddings and the division of two feature types were beneficial to the quality of the model. External validations from actual implementations of this model in the real world would positively affect our progress on ASD diagnosis.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

A1 Medical Integration. (2024, Nov 18). “*EEG Cap with LED Wires*”: A1 Medical Integration. https://a1props.com/product/eeg-cap-with-leds/

AI Hub. (2024, Nov 18). “Pediatric electroencephalography data”: AI Hub. https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=71356

Haque, N. (2023, Apr 3). “*What is Convolutional Neural Network — CNN (Deep Learning)*”: LinkedIn. https://www.linkedin.com/pulse/what-convolutional-neural-network-cnn-deep-learning-nafiz-shahriar/

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Saccá, V., Campolo, M., Mirarchi, D., Gambardella, A., Veltri, P., & Morabito, F. C. (2018). On the classification of EEG signal by using an SVM based algorithm. Multidisciplinary approaches to neural computing, 271-278.

Übeyli, E. D. (2009). Statistics over features: EEG signals analysis. Computers in Biology and Medicine, 39(8), 733-741.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

Wadhera, T. (2021). Brain network topology unraveling epilepsy and ASD Association: Automated EEG-based diagnostic model. Expert Systems with Applications, 186, 115762.