Capstone Project I Report

Mamba State Space Model: A Comprehensive Evaluation of Its Efficacy in Brain-Computer Interface Using EEG

Rakhat Ospanov, Shaknazar Zhumabekov

A thesis submitted in part fulfilment of the degree of

BS in Robotics Engineering

Supervisor: Berdakh Abibullaev

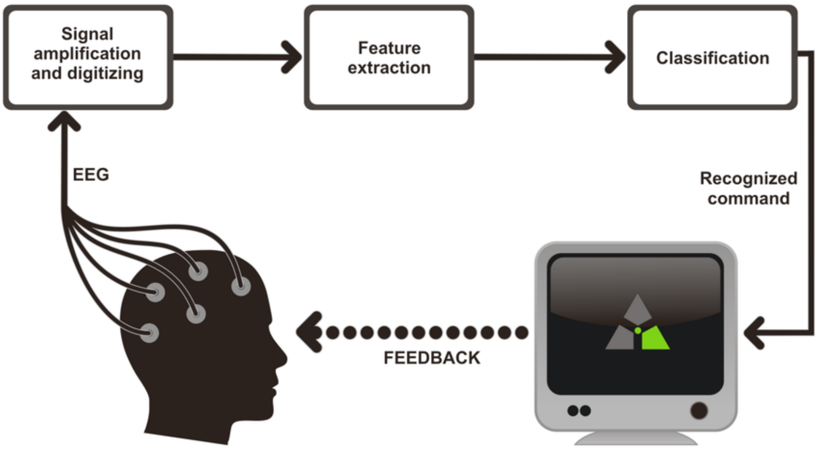
Instructor: Anara Sandygulova

Department of Robotics and Mechatronics
School of Engineering and Digital Sciences
Nazarbayev University
May 5, 2025

Table of Contents

Abstract

The increasing complexity of EEG data analysis in Brain-Computer Interface (BCI) systems calls for robust and scalable machine learning models capable of extracting meaningful spatiotemporal patterns. This project evaluates Mamba SSM, a novel deep learning architecture rooted in State Space Models (SSMs), for EEG signal classification. Building on the limitations of Transformers Mamba incorporates bidirectional Mamba blocks to address challenges such as computational overhead, long-sequence processing, and noise in EEG signals. The project includes an extensive literature review comparing Mamba and Transformer based architectures for EEG analysis, highlighting their strengths and weaknesses. Experimental development involved adapting Mamba for EEG datasets, focusing on architectural enhancements, and hardware efficiency. Initial results demonstrate the model's competitive accuracy in motor imagery classification tasks and its potential for scalability in real-time BCI applications. Challenges such as EEG data variability, dependency management, and computational resource limitations informed iterative development and learning outcomes. These included a deeper understanding of EEG datasets' structures, architecture design for time-series data, model training on resource-constrained systems and dependencies settings. Discussions center on Mamba's balance of efficiency and accuracy, emphasizing its utility for edge-device deployment in assistive technologies. This study investigates the use of Mamba State Space Model in the Domain of BCI Motor Imagery classification tasks. During the investigation, several architectures were proposed either based on pure Mamba blocks or combined with Attention layers. The findings suggest that Mamba SSM holds promise for advancing BCI systems, with future work exploring domain-specific optimizations, self-supervised learning, and broader applications in neuroscience. .



Chapter 1: **Introduction**

The extraction and interpretation of brain activity is of particular value for the diagnosis of various diseases and disorders, such as epilepsy, stroke, or sleep disorders. It is widely recognized that human sensory perception, limb movements, and even skin responses are controlled by constant feedback from the central nervous system, particularly the brain. Medical researchers have long been interested in decoding brain activity, resulting in the development of fMRI, MEG, and EEG formats for recording brain data. While fMRI and MEG are primarily used for diagnostic purposes, EEG has a broader range of applications, including enhancing human interaction with the environment, especially for individuals with movement or limb-control disabilities.

Non-invasive Brain-Computer Interfaces (BCIs) are among the most promising technologies for augmenting human capabilities by enabling interaction with computers, hardware components like robotic manipulators, and even video games. These systems rely on decoding EEG data streams from electrodes to generate output for external devices. However, EEG data analysis and classification remain challenging problems in neuroscience and machine learning. EEG signals are prone to noise and sensor artifacts, and they are highly user-specific due to the unique characteristics of each individual's central nervous system. The challenge lies in developing a generalizable classification model that can independently reduce noise and extract the most relevant features for classifying mental phenomena.

In the machine learning research community, the deep learning paradigm is seen as a promising solution to this problem due to its ability to automatically learn necessary features, requiring minimal pre-processing, such as noise cleaning or filtering.

State Space Models such as Recurrent Neural Networks were used for sequence modelling tasks in EEG domain, prior to popularization of the Transformer architecture among researchers working in Brain-Computer Interface area. However, since State Space Models process sequences of EEG in sequential manner, they suffer from being unable to capture long dependencies among sequences, usually forgetting state details coming earlier to the new sequence. Due to such constraints, RNN-s were not successful in most EEG tasks, shifting researchers' interest towards Transformer based architectures with Attention mechanism for parallel processing of the sequence and modeling inter-relationships between tokens or patches of a sequences, both from the beginning and end of a sequence. Recently, the Mamba deep learning model, based on Selective State Space Models (SSMs) and designed with hardware efficiency in mind, has shown great potential for long-sequence modeling, offering an alternative to popular transformer architectures. Unlike transformers, which rely on self-attention mechanisms to learn and prioritize important features, Mamba introduces bidirectional Mamba blocks that use Selective SSMs to capture temporal dynamics in EEG signals. Selective State Space Models act in a similiar fashion to the Attention mechanism within transformer models by providing a framework to learn and focus on extracting important and meaningful features within long EEG sequence without being overwhelmed in State updates due to long sequences. These blocks also incorporate position embeddings to address the position-sensitivity challenges posed by data, which is crucial for decoding spatial and temporal features in EEG signals.

Although only a few papers have been published on Mamba, this model shows great promise for tasks such as EEG signal processing due to its efficient handling of long-sequence data. Mamba's ability to focus on important features in a similiar way as Attention mechanism within transformers representations and handle long temporal data with minimal computational resources suggests it can effectively reduce noise and extract key features in EEG data, outperforming models like transformers in terms of both efficiency and accuracy for EEG decoding tasks.

This novel architecture marks a significant step forward in the development of BCIs, as Mamba's approach addresses the computation and memory challenges associated with long-sequence data, making it a promising tool for future research and practical applications in EEG-based systems.

Chapter 2: **Literature Review**

The interpretation of brain activity through EEG (electroencephalogram) data has long been a focal point in neuroscience and medical research, particularly in diagnosing neurological disorders like epilepsy, stroke, and sleep disorders [1]. Over the past decades, non-invasive Brain-Computer Interfaces have emerged as a powerful tool, offering new pathways for enhancing human interaction with the environment, especially for individuals with movement or motor control impairments [2]. Despite advancements, EEG signal analysis remains a significant challenge due to its susceptibility to noise, sensor artifacts, and user-specific variability [3]. These complexities have led to an ongoing search for deep learning models capable of decoding brain activity with higher accuracy and generalizability [4]. In recent years, Transformer-based models, especially Vision Transformers, have gained traction in handling time series and image data due to their ability to capture both spatial and temporal features [5]. For instance, Gunter et al. in their work, "SViT: A Spectral Vision Transformer for the Detection of REM Sleep Behavior Disorder" [6], investigated the application of a Spectral Vision Transformer for detecting REM Sleep Behavior Disorder using EEG data. They demonstrated the effectiveness of SViT in classifying RBD patients from healthy controls, highlighting its potential for sleep disorder diagnosis using BCI. Nan Qi et al., in their work "Seizure Prediction Based on Improved Vision Transformer Model for EEG Channel Optimization" [7], utilized an enhanced Vision Transformer (ViT) architecture for seizure prediction using the CHB-MIT dataset. By incorporating a Joint Patchwise Weighting (JPW) branch for robust feature extraction, their model achieved an impressive accuracy of 93.65 per cent, showcasing ViT's ability to efficiently process EEG signals for seizure detection. Additionally, researchers have explored the use of Vision Transformers for emotion recognition from EEG signals [5]. However, transformers come with high computational and memory costs, particularly for long-sequence data, such as EEG [8].

The Mamba deep learning model, based on State Space Models (SSMs), has recently been introduced as a novel alternative to traditional architectures like transformers for BCI applications. Designed with hardware efficiency in mind, Mamba is particularly suited to overcoming the computational and memory limitations that hinder real-time EEG processing on edge devices. According to [9], Mamba's linear scaling with sequence length significantly reduces the computational burden, allowing it to process extended sequences with improved efficiency compared to traditional models like transformers. This makes it especially suitable for EEG applications, where capturing long-range dependencies is crucial for accurate neural signal decoding. Furthermore, Mamba's state space models (SSMs) enable it to selectively attend to relevant contextual information, which is critical for dynamic and unpredictable BCI environments. Its strong performance across various benchmarks, as shown in zero-shot evaluations [9], further underscores its potential to serve as a backbone model for a wide range of sequence modeling tasks, including those in emerging modalities like genomics, audio, and video. This is especially relevant in the context of wearable BCIs, where the computational load must be minimized without sacrificing accuracy. A recent comparative study on neural decoding processes in non-human primates demonstrated Mamba's superiority over other models, including the Gated Recurrent Unit (GRU), RWKV (Receptance Weighted Key Value), and transformers. Mamba and RWKV showed superior inference and calibration speeds, making them ideal candidates for edge deployment where fast processing and scalability are paramount [10].

Additionally, Mamba's application extends beyond conventional EEG classification. The DrowzEE-G-Mamba model specifically addresses driver drowsiness detection by utilizing EEG

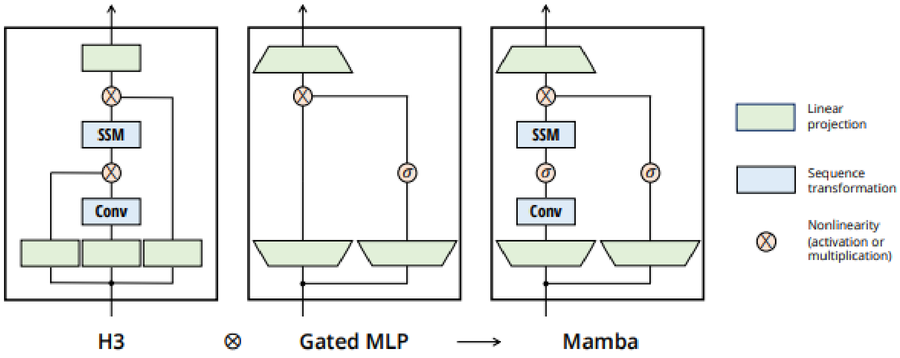


Figure 2.1: Mamba Design Architecture [9]

data to capture state transitions between alertness and drowsiness. This system has shown significant improvements in detection rates, achieving a peak accuracy of 83.24 per cent on the SEED-VIG dataset while maintaining robustness across various complexities, making it suitable for real-time applications with limited resources. The architecture incorporates channel-split, channel-concatenation, and channel-shuffle operations, optimizing information flow from EEG data. Such applications highlight the potential of Mamba-based models to enhance safety in critical scenarios, paving the way for powerful SSM-based AI algorithms in BCIs [11].

Moreover, the introduction of EEGMamba illustrates Mamba's versatility in EEG classification tasks. This universal classification network integrates a Spatio-Temporal-Adaptive (ST-Adaptive) module, Bidirectional Mamba, and Mixture of Experts (MoE) into a unified framework. The ST-Adaptive module addresses variations in signal length and channel count, facilitating unified feature extraction across different EEG datasets. With its ability to balance accuracy and fast inference speed, EEGMamba demonstrates superior performance in various tasks, including seizure detection, emotion recognition, and motor imagery classification. Such advancements affirm Mamba's capability to generalize across multiple tasks, a critical consideration given the variability inherent in EEG data [12].

Mamba's design allows it to handle large datasets and scale efficiently, addressing one of the core challenges in neural decoding: maintaining performance as data volumes increase. Traditional models like GRU, while sufficient in certain tasks, fail to scale effectively, and transformer models, though accurate, suffer from prohibitive computational requirements. This places Mamba in a unique position, especially as BCIs evolve to accommodate more complex tasks and user-specific adaptations in real-time. Additionally, the Mamba architecture, unlike transformers, does not rely on attention mechanisms, making it computationally lighter while still achieving high levels of performance in decoding tasks.

Although Mamba has proven effective in non-human primate experiments, its potential application in human EEG-based BCI remains largely unexplored. Current studies primarily focus on traditional EEG analysis and classification methods, and there is a notable lack of literature directly addressing the integration of Mamba models into BCI frameworks. This gap offers a significant opportunity for future research to assess how the Mamba architecture can be leveraged to improve real-time EEG processing, noise reduction, and classification accuracy in human BCIs. With its demonstrated success in scalability, inference speed, and handling large datasets, Mamba presents a promising avenue for developing more efficient and practical BCI systems, particularly for non-invasive applications. Therefore, this paper aims to investigate the application of Vision Mamba in enhancing the performance and efficiency of BCI systems, addressing the existing research gap and paving the way for future advancements in the field.

Chapter 3: **Methodology**

Proposed Mamba Model Architecture

The Mamba architecture is a novel approach to sequence modeling that combines the strengths of bidirectional sequence processing and state space modeling. This hybrid approach enables the model to effectively capture both data-dependent contextual information and long-range temporal dependencies [9].

The core components of the Mamba model are:

1. **Token Embedding:** The input sequence is first tokenized and converted into dense vector representations. These embeddings serve to transform discrete input tokens into a continuous space, enabling the model to process them effectively.
2. **Positional Encoding:** To incorporate order and temporal structure into the model, positional embeddings are added to the token embeddings. These embeddings allow the model to differentiate between tokens with similar content but occurring at different positions in the sequence.
3. **Mamba Encoder Block:** This block forms the backbone of the model and consists of several key components:
   - **Normalization:** Input tokens are normalized to stabilize training and ensure consistent scaling across features.
   - **Projection Layers:** The normalized embeddings are passed through projection layers to extract higher-level representations of the sequence data.
   - **Forward and Backward Convolution:** Bidirectional convolution operations are applied to capture both forward and backward dependencies, allowing the model to aggregate context from both past and future tokens.
   - **State Space Model Layers:** These layers leverage a state space modeling framework that allows the model to maintain a latent memory of the sequence, capturing both local and global structural information efficiently.
4. **Multilayer Perceptron (MLP) Block:** The output of the Mamba Encoder is passed to an MLP block, which performs the final prediction or classification task. The MLP aggregates the learned representations and outputs a probability distribution or sequence of predictions, depending on the downstream task.

By combining these components, the Mamba model is able to learn complex temporal dynamics and deliver strong performance across a wide range of sequence modeling tasks, including natural language processing, time series forecasting, and beyond.

Bidirectional State Space Models

Bidirectional state space models extend the traditional state space framework by considering both past and future information. This allows the model to capture dependencies that may not be apparent when only considering past information.

Key characteristics of bidirectional state space models:

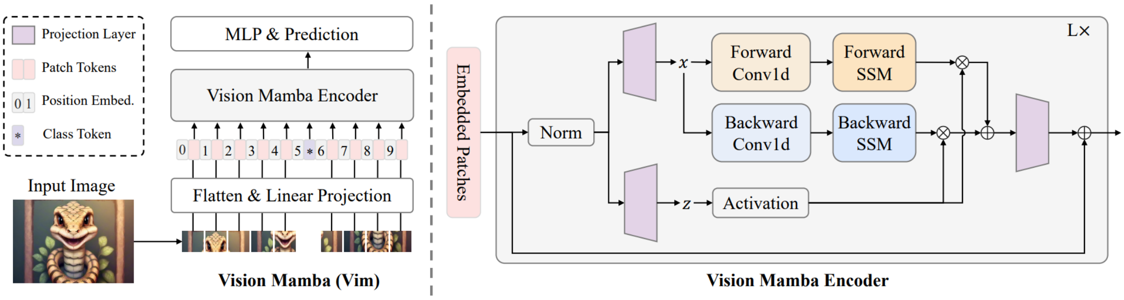


Figure 3.1: Vision Mamba Architecture [13]

- **Two sets of state variables:** One set evolves forward in time and captures the influence of past information, while the other set evolves backward in time and captures the influence of future information.
- **Combined output:** The final output of the model is a combination of the outputs from the forward and backward state variables.

Motor imagery datasets in BCI play a crucial role in evaluating the classification capabilities of deep learning models. These datasets capture brain signals associated with imagined movements, such as hand or foot movements. By training deep learning models on these datasets, researchers can assess the models' ability to accurately distinguish between different types of motor imagery. This evaluation is essential for developing effective brain-computer interfaces (BCIs) that can translate brain signals into control commands for external devices. The complexity and diversity of motor imagery datasets, coupled with the increasing sophistication of deep learning architectures, have contributed to significant advancements in BCI technology.

In our methodology, we prioritize detailed experiments on the BCI 1 and BCI 2 Motor Imagery Datasets. These datasets provide a valuable resource for researchers to develop and evaluate algorithms for decoding brain signals associated with imagined movements. The BCI 1 contains EEG recordings from three subjects performing four different motor imagery tasks: left-hand, right-hand, feet, and tongue. The BCI 2 contains EEG recordings from nine subjects performing two motor imagery tasks: left-hand and right-hand. As another chapter in methodology, we are interested in developing comparison of the performance in terms of classification accuracy as well as on hardware performance among various modifications of the Mamba models with baseline EEG classification models such as EEGNet and various transformer based architectures for EEG. The special emphasis on the Transformer and Mamba comparison is going to be placed, as Transformer and Mamba share a lot of architectural similiarities, and there is a significant evidence within other domains of research showing Mamba outperforming Transformer in parameter size, computational demands and classification accuracy.

Chapter 4: **Ethical Considerations**

Ethical and Legal Issues

Our project "Mamba State Space Model: A Comprehensive Evaluation of Its Efficacy in Brain-Computer Interface Using EEG" raises several ethical and legal concerns, particularly surrounding data privacy, intellectual property (IP), and safety.

First of all, a potential problem can occur with the data privacy. BCI systems collect sensitive brainwave data, which is highly personal and must be handled carefully to protect user confidentiality. In line with "General Data Protection Regulation" guidelines [14], our team is going to take several important steps to ensure compliance. Firstly, we are going to anonymize all the collected EEG data and use encryption techniques for storage, ensuring that no individual could be identified from the dataset. This reflects our commitment to protecting the privacy of participants, even when working with complex neural data.

Additionally, the "Intellectual Property" concerns are considered all throughout the project. As it was found, Vision Mamba is open source, which raised questions about proper usage and licensing. We ensure that all third-party libraries and tools used in the project complied with their licensing agreements. The potential for our work to introduce new, patentable algorithms was carefully considered, with the decision to adopt a research-oriented approach while recognizing the importance of IP laws. Furthermore, in the context of this issue we use [15] and [16].

Informed Judgments and Trade-offs

In the process of developing the BCI system and its potential extension to robotics, we anticipate encountering a specific trade-off that requires careful decision-making: the balance between cost and safety. EEG reading systems must be safe, as they come into close contact with the human scalp, which introduces risks associated with electrical interference or data privacy breaches. Ensuring that the system is both affordable and safe is crucial for making BCI technology accessible while maintaining high ethical standards.

Additionally, if we extend the system to include robotics, particularly in environments where they interact with humans, the same trade-off emerges. We aim to develop affordable systems, but not at the expense of safety. For this reason, we incorporate redundant safety mechanisms, such as system diagnostics and fail-safes, to control any potential malfunctions. These mechanisms are essential for minimizing risks to users and ensuring that the system operates safely under all conditions. This approach demonstrates how we prioritized user safety while carefully managing project costs, balancing innovation and responsible design.

Sustainability and Social Responsibility

From a sustainability and societal perspective, we mainly focused on social impacts of our work. In the process of developing our project and presenting results we are particularly concerned with providing Sustainable and responsible development of the scientific research across Machine Learning community. In order to do that, we stick to the promise of developing comprehensive and comfortable GitHub repository for our project with developed script representing Demo of the project; with development of the Docker containers for easy access to the functions and with adaptable inference code implementations and training loops and data loaders for customised trainings. Thus we are hoping to provide research which is interpretable for scientists based on their own preferences, and custom datasets.

Furthermore, in terms of societal responsibility, there is a global concern about Robotics automation. Automation, especially in medical and industrial fields, could lead to job displacement. Recognizing this, we aimed to focus our robotics work on augmenting human labor rather than replacing it, ensuring that our technology could be integrated in ways that improve job quality and productivity. This aligns with our goal to introduce socially responsible technologies that mitigate the negative effects of automation.

Examples from the Project

A hypothetical example would be the societal impact of ”robotic automation in medical diagnostics”. If robotics were to automate some medical tasks, such as diagnostics, it could potentially reduce the need for certain healthcare professionals. However, we emphasized the importance of designing systems that work collaboratively with humans, enhancing healthcare delivery rather than replacing professionals. Furthermore, in regard to the Human Health we should never lie all the responsibility to a robot.

Ethical Summary

In summary, our project involves key ethical, legal, and societal considerations. By addressing issues of data privacy, intellectual property, safety, and societal impact, our team demonstrates strong commitment to ethical responsibility. We take informed steps to navigate trade-off in cost versus safety, while ensuring large social responsibility in our design choices.

Chapter 5: **Literature Progress**

In the initial stages of our project, we explored the use of Transformer architectures for EEG signal analysis in Brain-Computer Interfaces (BCI). We delved deeply into Transformer models, attempting to code a preliminary model and conducting extensive research on their application in the field. However, after further evaluation, we decided to pivot towards using Mamba State Space Models (Mamba SSM), a more novel and promising approach for this task.

We found Mamba SSM stood out due to its efficiency and potential in time-series analysis, which aligns well with EEG signal processing. We conducted a thorough literature review to assess its feasibility and explored the existing code implementations to determine how we could adapt it for BCI EEG data. This shift represents a significant advancement in our methodology, bringing us closer to a more optimized solution.

Name	Dataset	Data Preprocessing	Model Architecture	Results
Benchmarking Neural Decoding Backbones towards Enhanced On-edge iBCI Applications [10]	Signals from Primary Motor and Sensory Cortexes from non-human primates	-	original Mamba [9]	Mamba demonstrate faster inference speeds, lower computational complexity, better scalability compared to GRU and Transformer, emerge as preferred choices for deployment on wearable devices.
DrowzEE-G-Mamba: Leveraging EEG and State Space Models for Driver Drowsiness Detection [11]	SEED-VIG dataset recordings from 23 participants	Convolution Layers	DrowzEE-G-Mamba architecture based on Vmamba	peak accuracy of 83.24 - higher than SVM, EEGNet, TSception, ConvNeXt, LMDA-Net
EEGMAMBA: BIDIRECTIONAL STATE SPACE MODELS WITH MIXTURE OF EXPERTS FOR EEG CLASSIFICATION [12]	Bonn, CHB-MIT, SleepEDF-20, SHHS, DEAP, SEED, Shu, BCI-IV-2a	For the DEAP dataset: downsampled to 128Hz and a bandpass filter with a cutoff frequency of 4-45Hz is applied; for the SEED dataset: downsampled to 200Hz and subjected to a bandpass filter ranging from 0 to 75Hz	EEGMamba, which seamlessly integrates the Spatio-Temporal-Adaptive (ST-Adaptive) module, Bidirectional Mamba, and Mixture of Experts (MoE)	evaluated the model on eight publicly available EEG datasets across four tasks, and experimental results demonstrate the superior performance of our proposed model in multi-task classification scenarios.
MENTALITY: A MAMBA-BASED APPROACH TOWARDS FOUNDATION MODELS FOR EEG [17]	The training data subset contains 579 patients, with 2,138 seizure events	All signals are resampled to 200 Hz, and we remove 60 and 120 Hz interference via a notch filter	CNN + Mamba	AUROC of 0.72
Brain-Mamba: Encoding Brain Activity via Selective State Space Models [18]	The Seizure detection TUH-EEG dataset, MPI-EEG	-	BNMamba, a graph learning method that use special traits of brain networks to efficiently and effectively encode spatiotemporal brain networks. It leverages: 1 MPNNs to encode local dependencies of brain units, 2 a tokenizer and ordering mechanism to order nodes with respect to their functionality in the brain, 2 a selective structured state space model to efficiently select informative and relevant brain regions, and 4 a simple adaptive readout block to learn the brain-level encoding.	achieved the highest accuracy among the other popular models, MBI-EEG - 92.28 per cent, TUH - 92.17 per cent

Table 5.1: Mamba for EEG data analysis review table

As a result of our re-selection towards the Mamba, we performed a comprehensive literature review, comparing the application of Mamba State Space Models (Mamba SSM) 5.1 and Transformers **??** for EEG signal analysis in Brain-Computer Interfaces (BCI). Both approaches have shown potential in time-series analysis, but differ significantly in their architectures, computational efficiency, and ability to handle EEG data characteristics.

Transformers)

Our exploration of Transformers focused on their application to EEG, a complex and non-stationary signal. Through our literature review, we evaluated various parameters that have been employed in previous works:

- **Data preprocessing:** Transformer models often rely on techniques such as short-time Fourier transform (STFT) or continuous wavelet transform (CWT) to convert EEG signals into images or 2D matrices, as demonstrated in several studies like [7], [19],

Name	Dataset	Data Preprocessing	Model Architecture	Results
EEG-Based Emotion Recognition Using a Multi-Head Attention Mechanism with a Novel Loss Function [?]	SEED dataset for emotion recognition.	EEG signals segmented into 1-second windows. Artifact removal using Independent Component Analysis (ICA).	Transformer encoder with multi-head self-attention mechanism. Novel loss function incorporating both classification and distribution learning.	Achieved an average accuracy of 86.06Cross-Subject Emotion Recognition from EEG Signals Using Transferable Spatial-Temporal-Spectral Patterns and Deep Learning [?]
DEAP dataset for emotion recognition.	EEG signals preprocessed with band-pass filtering (4-45 Hz). Short-Time Fourier Transform (STFT) applied to extract spectral features.	Transformer network to model temporal dependencies of spectral features. Spatial information encoded using channel embeddings. Transfer learning strategy employed for cross-subject recognition.	Reported an average accuracy of 78.54A Novel Deep Learning Approach for Motor Imagery EEG Classification Based on Multi-Scale Temporal Convolutional Network and Transformer [?]	BCI Competition IV 2a dataset for motor imagery classification.
EEG signals filtered between 0.5 and 30 Hz. Data augmented using time warping and scaling techniques.	Hybrid architecture combining Multi-Scale Temporal Convolutional Network (MSTCN) for feature extraction and a Transformer encoder for sequence modeling.	Achieved a classification accuracy of 90.32Sleep Stage Classification Using EEG Signals Based on Improved Transformer Network with Attention Mechanism [?]	Sleep-EDF dataset for sleep stage classification.	EEG signals segmented into 30-second epochs. Features extracted using Fast Fourier Transform (FFT).
Improved Transformer network with a novel attention mechanism to capture long-range dependencies in EEG signals.	Obtained an overall accuracy of 87.6Epileptic Seizure Detection Using EEG Signals with a Novel Attention-Based Bidirectional Long Short-Term Memory Network [?]	CHB-MIT dataset for seizure detection.	EEG signals preprocessed with band-pass filtering (0.5-40 Hz). Data segmented into fixed-length windows.	Attention-based Bidirectional Long Short-Term Memory (BiLSTM) network. Attention mechanism used to weigh the importance of different time steps in the EEG sequence.
Reported a seizure detection accuracy of 97.3 Transformer Networks for EEG Signal Classification [?]	Various EEG datasets including motor imagery and emotion recognition datasets.	EEG signals can be directly used as input or transformed into time-frequency representations.	Explored the application of standard Transformer networks for EEG signal classification. Treated EEG signals as sequences and applied Transformer encoder directly.	Demonstrated the feasibility of using Transformer networks for various EEG classification tasks, achieving competitive results compared to existing methods. Highlighted the potential of self-attention for capturing temporal dependencies.
A Frequency Domain Enhanced Transformer for EEG-Based Brain-Computer Interfaces [?]	GigaDB dataset for motor imagery classification.	EEG signals segmented into trials. Power Spectral Density (PSD) features extracted from different frequency bands.	Frequency Domain Enhanced Transformer (FDET) that incorporates frequency domain information into the Transformer architecture. Used multi-head self-attention to learn relationships between different frequency bands and channels.	Achieved an average classification accuracy of 88.43 Utilizing Temporal and Spectral Features with Deep Learning for EEG-Based Emotion Recognition [?]
MAHNOB-HCI dataset for emotion recognition.	EEG signals preprocessed with band-pass filtering. Temporal features extracted using statistical measures. Spectral features extracted using Wavelet Transform.	Deep learning model incorporating both Convolutional Neural Networks (CNNs) for feature extraction and Transformer networks for sequence modeling of temporal and spectral features.	Reported competitive results on the MAHNOB-HCI dataset for emotion recognition. Demonstrated the effectiveness of combining temporal and spectral features with deep learning architectures.	
Graph Transformer Networks for Brain Signal Analysis [?]	Various EEG and MEG datasets for different cognitive tasks.	Brain signals represented as graphs where nodes represent electrodes and edges represent functional connectivity.	Graph Transformer Network (GTN) that extends the Transformer architecture to process graph-structured data. Used graph attention mechanisms to learn relationships between different brain regions.	Showed promising results on various brain signal analysis tasks, including classification of cognitive states and prediction of neurological disorders. Highlighted the potential of graph-based Transformer models for capturing complex brain dynamics.
Learning Long-Range Dependencies in EEG with End-to-End Transformer [?]	A synthetic EEG dataset and a real-world sleep EEG dataset.	Raw EEG signals used as input without extensive preprocessing.	End-to-end Transformer model designed to directly process raw EEG signals. Explored different input representations and attention mechanisms for capturing long-range temporal dependencies.	Demonstrated the ability of the Transformer model to learn meaningful representations from raw EEG data and achieve competitive performance on sleep stage classification. Highlighted the potential for simplifying the EEG analysis pipeline.
Transformer-Based Deep Learning Model for EEG Signal Classification [?]	A custom EEG dataset for a specific cognitive task.	EEG signals segmented and potentially filtered based on the specific task requirements.	Implemented a standard Transformer encoder for classifying EEG signals related to the cognitive task. Explored different configurations of the Transformer model.	Achieved high classification accuracy on the custom EEG dataset, demonstrating the applicability of Transformer models to specific EEG-based applications.
Multi-Modal Emotion Recognition Using EEG and Physiological Signals with Transformer Networks [?]	RECOLA dataset containing EEG and physiological signals for emotion recognition.	EEG and physiological signals preprocessed and synchronized.	Transformer network designed to fuse information from multiple modalities (EEG and physiological signals) for improved emotion recognition. Used cross-modal attention mechanisms.	Showed improved emotion recognition performance by leveraging information from both EEG and physiological signals using a Transformer-based fusion approach.

Table 5.2: Transformer Networks for EEG Data Analysis Review Table

and [20]. This process aids the transformer model in leveraging its powerful attention mechanism for feature extraction.

- **Architecture modifications:** Given that EEG signals are fundamentally different from text sequences or visual data typically fed into the transformers, researchers have adapted transformers by adding convolution layers, attention modules, or even hybrid models (e.g., HViT and LCT) to improve their ability to capture temporal and spatial dependencies in EEG signals.
- **Performance Metrics:** Across different tasks, ViT-based models have achieved promising results in seizure detection and emotion recognition, as evidenced by works such as [21], [22], and [23], achieving high accuracy, sensitivity, and specificity. However, the computational complexity and memory requirements of ViTs were consistently highlighted as concerns, particularly when scaling to large datasets.

Despite the strong performance of ViTs in EEG tasks, we identified limitations in terms of scalability and efficiency, especially when dealing with large datasets or requiring on-edge computation for real-time BCI applications.

Mamba SSM: A More Efficient Alternative

Our shift to Mamba SSM was driven by its potential to better handle time-series data like EEG signals. Mamba SSM offers more efficient processing, reduced computational overhead, and flexibility in adapting to various EEG datasets.

- **Efficiency and Time-Series Alignment:** Mamba SSM's design, rooted in state space models, allows it to efficiently model temporal dependencies. This aligns closely with EEG signals, which are inherently time-dependent. Several studies such as [10], [11], and [18] demonstrated Mamba SSM's superior inference speed and lower complexity compared to other methods like GRU and Transformer models.
- **Model Adaptation and Scalability:** The modular nature of Mamba allows for easy integration of additional components like convolution layers, bidirectional models, and mixture of experts, as seen in works like [12]. This ensures that Mamba SSM can scale across tasks without compromising efficiency or accuracy.
- **Performance Evaluation:** In our literature review, Mamba SSM consistently outperformed traditional approaches in EEG signal analysis tasks, particularly in multi-task classification, emotion recognition, and seizure detection. The reduced computational cost and improved scalability make it a compelling option for real-time BCI applications

Code Implementation To better understand and implement Mamba SSM, we explored existing code repositories, focusing on two key GitHub repositories:

- **VMamba GitHub repository:** This repository provided a foundation for implementing Mamba SSM, with detailed code for time-series analysis, particularly suited to EEG data.
- **Vim GitHub repository:** While initially focused on vision models, this repository offered insights into transformer-based architectures that helped us refine our approach and compare them to Mamba's capabilities.

Both repositories were instrumental in shaping our understanding of the architecture and how it could be adapted to our specific BCI EEG task. The code provided a practical framework for experimentation, allowing us to customize models, test different datasets, and refine preprocessing techniques to suit EEG signals.

Chapter 6: **Initial Experiments and Model Development**

Initial approaches: LSTM experiments

To get initial intuition and development of the State Space Models such as Mamba for EEG, we started with their precursor models such as LSTM, Transformers and LSTM models with Multi-head Attention blocks. LSTMs (Long Short-Term Memory networks) and Transformers serve as foundational models that paved the way for the development of advanced architectures like Mamba. LSTMs introduced a robust mechanism for handling long-range dependencies in sequential data, using gates to regulate the flow of information and mitigate the vanishing gradient problem. This capability made LSTMs a go-to model for tasks like time-series prediction and natural language processing. However, they faced limitations in scalability and parallelization due to their sequential nature. Transformers, with their self-attention mechanism, revolutionized the field by enabling parallel processing and capturing global context in sequences more effectively. This innovation addressed LSTMs' inefficiencies, particularly for tasks involving large-scale data and long sequences. The Mamba model builds on these advancements, combining LSTMs' strength in temporal representation with Transformers' efficiency and scalability, potentially introducing novel mechanisms that integrate both architectures' strengths while addressing their weaknesses.

In first iterations, we developed several LSTM based models for EEG Motor Imagery classification. The first baseline model consisted of single spatial convolution layer, single temporal convolution layer with batch normalization layers followed after every convolutional layer. Then, single LSTM block is used to process temporal features in more detail, Additonally, we apply ReLU activation functions after each convolution to extract non-linear features. To avoid overfitting, we add Dropout after the LSTM layer. Finally, classification layer is used as fully connected layer in which the class probabilities are obtained. The consequent models became more complex, in the second model for example we add 2 more convolutional layers, 1 for temporal and 1 for spatial feature extraction. In third model we add second LSTM model, with Dropout being added to the second LSTM model too. We remain with same 4 layers for temporal and spatial convolution and make sure that each of them is followed by consequent Batch Normalization module. Finally, we add Max Pooling layers after the second spatial and temporal layers to reduce dimensions within the data.

We applied these models to the BCI IV competition part A and B both of which are focused on Motor Imagery. In our case there were two classes present with an imagery of right hand movement or left hand movement. The training was done for 40 epochs with a batch size of 32 EEG data samples per batch. The training algorithm was chosen to be Adam with learning rate of 0.001. Dropout probability was set to be 0.3 across all of the models, as we didn't expect overfit of the model to happen with initially small sized models. LSTM models were easy to train, but their performance was significantly low, within the range of 20-27% accuracy at the beginning, but later 3rd model performed properly to 48%-52% accuracy range. We think these models gave us initial intuition into architecture design of the models for EEG data, but our general conclusion was to gradually remove idea of using LSTM based models in this problem setting.

Initial approaches: Convolutional Mamba approach

The next part of our work was based on the development of Mamba approach. Specifically

our goal was to use CNN along with Mamba SSM and some additional blocks in the case they are required.
In our Mamba CNN pipeline we found a suitable architecture for the initial programming progress [24]. In this github the author provides Transformer based EEG data classification approach and achieves 84.4 percent accuracy on the validation dataset, which is a good result in the regard to EEG Deep Learning module. According to the Figure 6.1 and Figure 6.2 it is clear that the training step of [24] was successful, leading to accurate and competitive results. It must be noted that a significant part of our interest toward this architecture was based on the use .ipynb format by the author, allowing us to research this pipeline in Google Colab. In chapter 7, we are going to explain the reason towards this online tool.

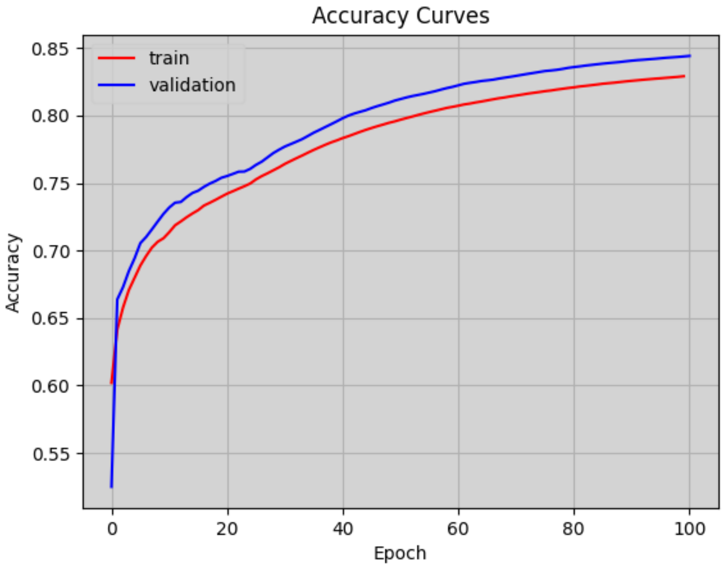


Figure 6.1: Transformer Accuracy curve

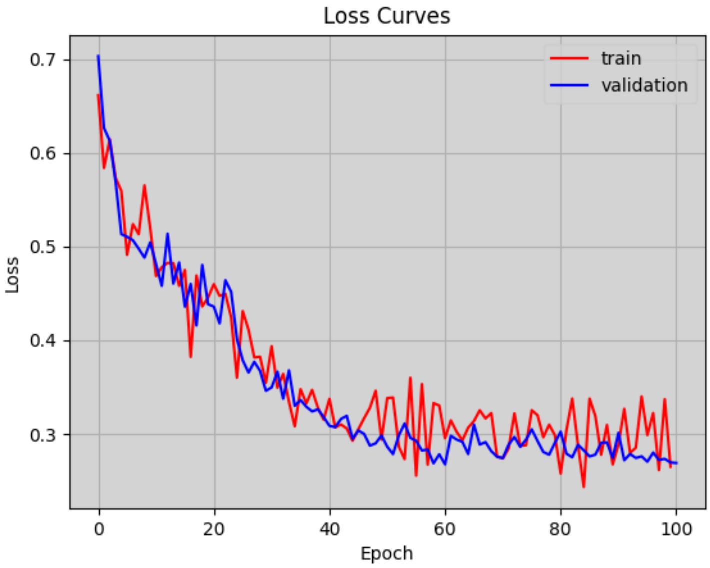


Figure 6.2: Transformer Loss curve

In this paper, we are going to use the dataset used [24] for the justified and direct investiga-

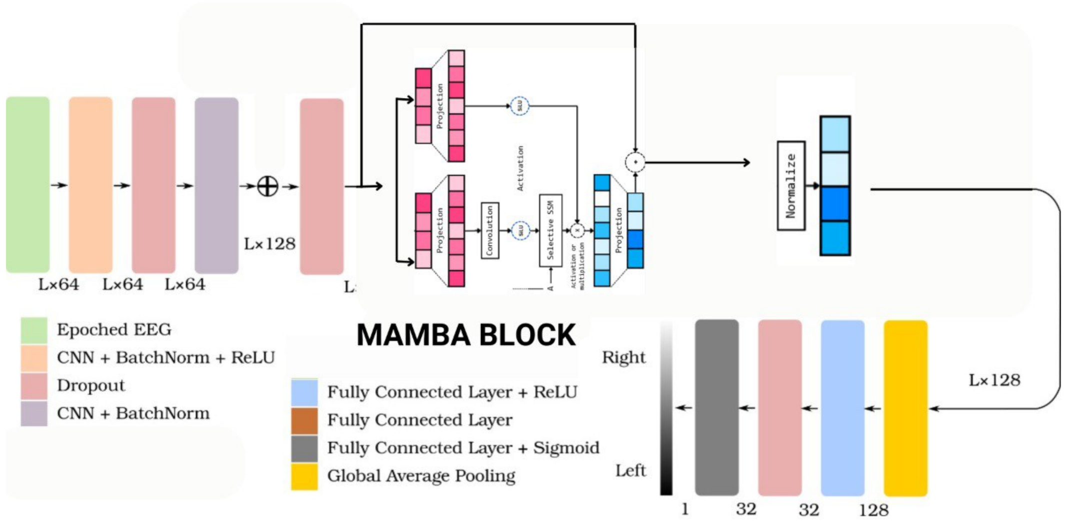


Figure 6.3: Proposed CNN Mamba Architecture

tion of potential differences between the Mamba SSM and Transformers architectures in the regard to Machine learning in Brain Computer Interface.

Dataset

The data was provided by Physionet. The dataset was collected from 109 participants leading to the overall number of 1500. The samples are represented as one and two minute EEG recordings. In the beginning of each sample collection, volunteers were asked to open and close their eyes. Consequently, it was required for the participants to firstly, open one of their fists followed by imaging the same action. Next, participants had to either open and close both of their fists or feet and similarly imagine the same movement. Regarding the annotations, it should be mentioned that authors included three classes: T0 - rest, T1 - left fist and both fists, T2 - right fist and booth feet. The data was recorded via 64 channels by BCI2000 system. Figure 6.4 provides a deeper look into the hardware system of acquiring the data in terms of Physionet EEGmmidb dataset. Figure 6.5 presents a sample signal from the training data.

Mamba SSM architecture[1]

For the implementation of Mamba SSM block, we replaced corresponding parts of the transformer with our desired state space model architecture. It should be noted, that positional embedding part was also reduced in the context of our research. Figure 6.3 represents our proposed CNN + Mamba + MultiLayer Perceptron architecture.

The data entering CNN is in shape (3377, 64, 497) where 3377 is the quantity of epochs, 64 is the number of channels and 497 is the amount of time points in each epoch. It should be mentioned that epochs in regard to BCI is divided original time series of sample by trials. The first block of the CNN is Conv1D layer with the same number of output neurons as the input followed by ReLU activation function, Batch Normalization and Dropout. The second block is Conv1D with twice the number of channels in the output shaping the data to (10, 128, 497) with the consequent Batch Normalization. 3377 epochs have been divided into batches with 10 samples. Afterwards, via the permutation input of size (10, 497, 128) (batches, time points, 2 * channels) is considered as the input to Mamba. Next, global pooling find the mean of the time points, decreasing the shape to (10, 128) (batches, 2 * channels). Finally, the MultiLayer Perceptron maps the output into two classes.

Training Results

We chose the number of epochs to be equal to 100. As a result, the test accuracy is 71 per-

---

[1] `https://github.com/RakhaOsp/capstone_github_BCI/blob/main/worked_try_mamba.ipynb`

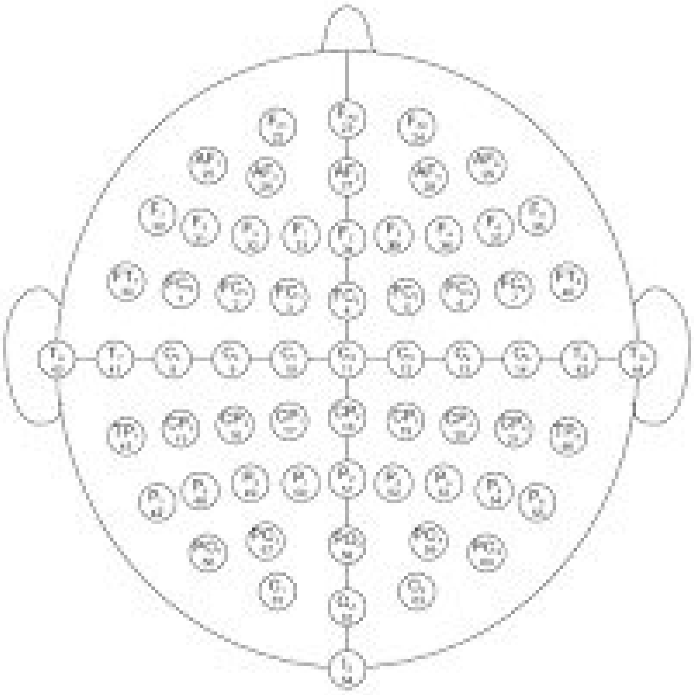


Figure 6.4: Channel structure of Physionet EEGmmidb

cent, while the validation and train accuracies are 78 and 79 percent, respectively. Figures 6.6 and 6.7 vizualize the results.

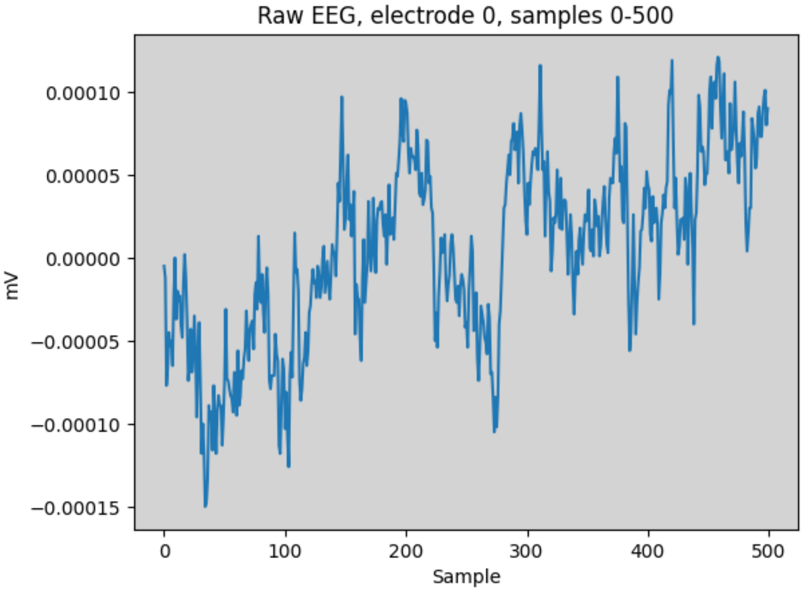


Figure 6.5: Sample of Physionet EEGmmidb

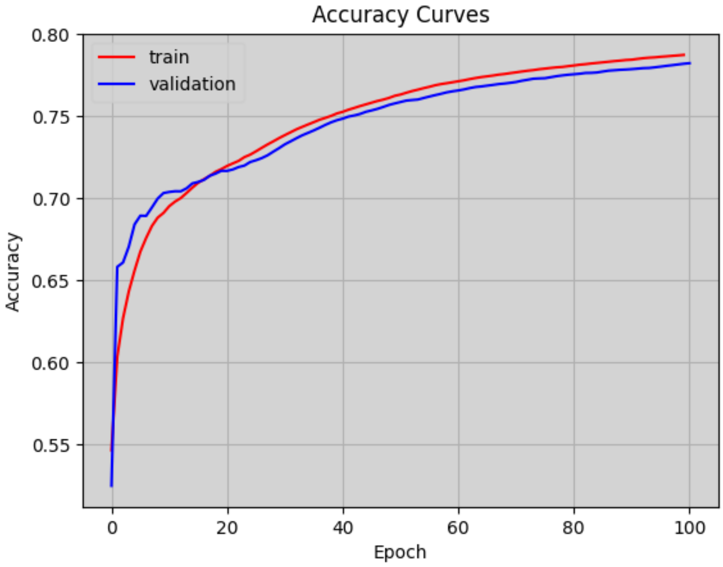


Figure 6.6: Proposed Mamba Accuracy

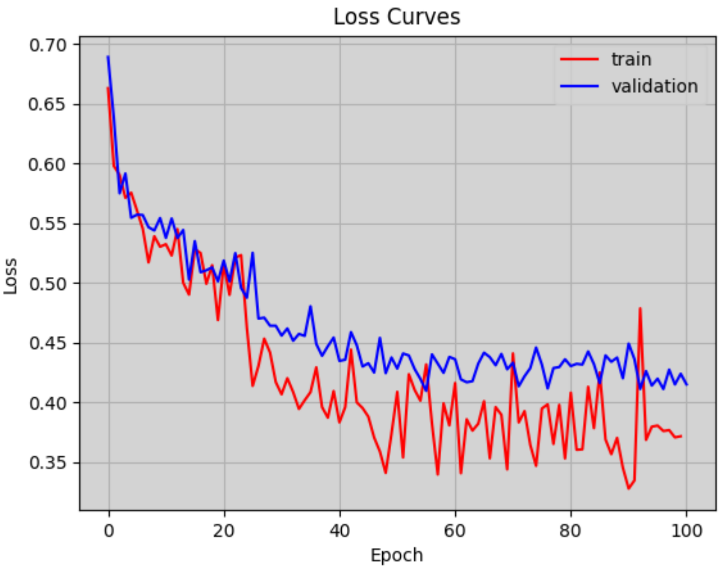


Figure 6.7: Proposed Mamba Loss

Chapter 7: **Final experiments and proposed architectures**

The final part of our work was based on the investigation of Mamba SSM capabilities in the context of BCI EEG motor imagery classification. Based on our previous experience from chapter 6 and literature review, we tried to develop new models achieving high resutls, sufficient GPU memory utilization and high inference speed in comparison to transformers based models. In this chapter we propose our best model STCMA and present several other models built on Vision Mamba and Hybrid transformers+Mamba SSM architectures. The following models developments were based on BCI Competition IV 2a dataset.

Proposed Spatial Temporal Convolutional Mamba Architecture (STCMA)

The development of the Spatial Temporal Convolutional Mamba Architecture (STCMA) went through multiple iterations (v1, v2) before arriving at the final optimized design. In this section, we describe the evolution of the model architecture from the early versions to the final version, highlighting the rationale behind each change and the corresponding performance outcomes.

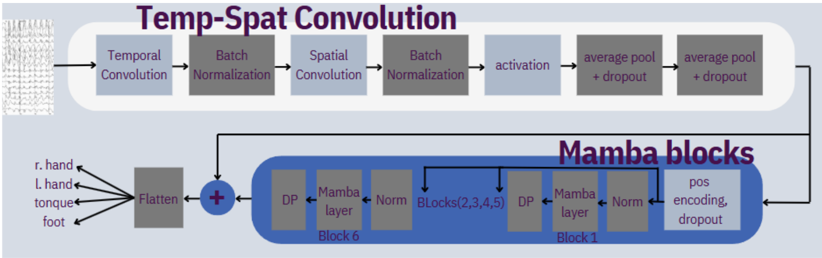


Figure 7.1: STCMA v1.0

STCMA v1: Initial Architecture and Limitations

STCMA v1 was the initial prototype combining convolutional neural networks with a state-space model (Mamba) for EEG motor-imagery classification. As shown in Figure 6.8, the v1 architecture consisted of a simple pipeline: an initial convolutional layer for feature extraction, followed by a single Mamba SSM (Selective State-Space Model) encoder block, and finally a classification layer. In essence, STCMA v1 included one spatial convolution and one temporal convolution (with appropriate batch normalization and activation after each), feeding into a single Mamba sequence modeling layer, and then a fully connected output layer. This straightforward design was inspired by earlier LSTM-based baselines (replacing the LSTM with a Mamba block) to capture temporal dynamics. However, the performance of STCMA v1 was found to be suboptimal. The model struggled to effectively learn the complex spatial patterns in multi-channel EEG data, likely because the single-pass convolutional feature extractor was too shallow to capture relevant EEG features. Additionally, the integration of the Mamba encoder in this initial form did not yield a significant accuracy boost – possibly due to insufficient preprocessing of the signals or inadequate capacity of the single Mamba block. As a result, STCMA v1 achieved relatively low classification accuracy and was ineffective compared to existing approaches. These shortcomings highlighted the need for a more sophisticated architecture to fully leverage the Mamba's potential.

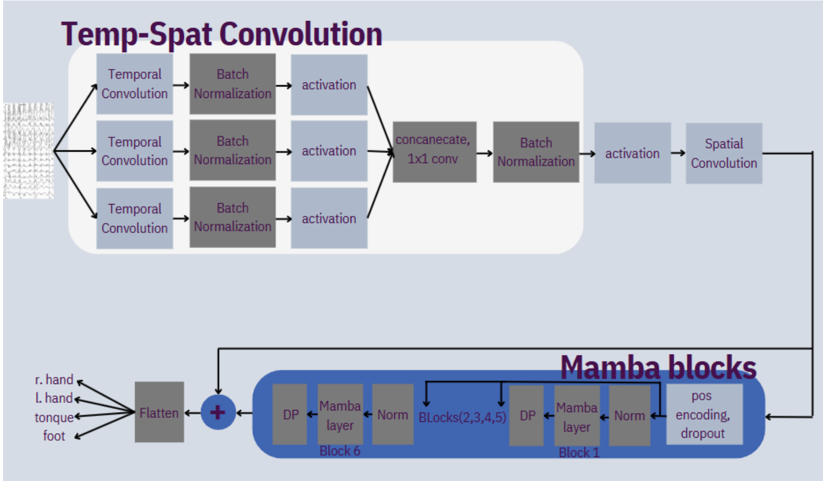


Figure 7.2: STCMA v2.0

STCMA v2: Architectural Refinements with Added Depth

To address the limitations of v1, a revised architecture was developed as STCMA v2 (illustrated in Figure 6.9). STCMA v2 introduced several modifications aimed at improving feature extraction and sequence modeling. First, the convolutional front-end was made deeper: additional temporal and spatial convolution layers were added (each still followed by batch normalization and nonlinear activation) to allow the network to capture a richer hierarchy of EEG features. For example, whereas v1 had only one layer for each of temporal and spatial filtering, v2 included two more second temporal layers and a second spatial convolution layer, with an intermediate pooling step to reduce dimensionality. This expanded CNN block was intended to better isolate important frequency-band information and inter-channel relationships in the EEG signals. These architectural refinements did lead to some performance improvement over v1, indicating that the model was learning more nuanced features. Furthermore, during the

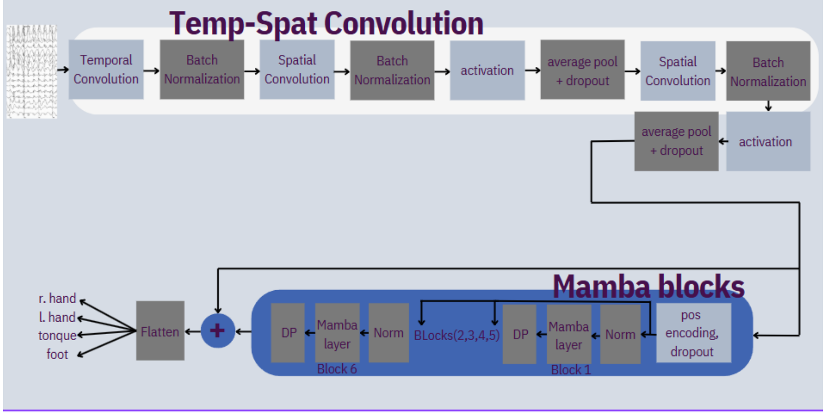
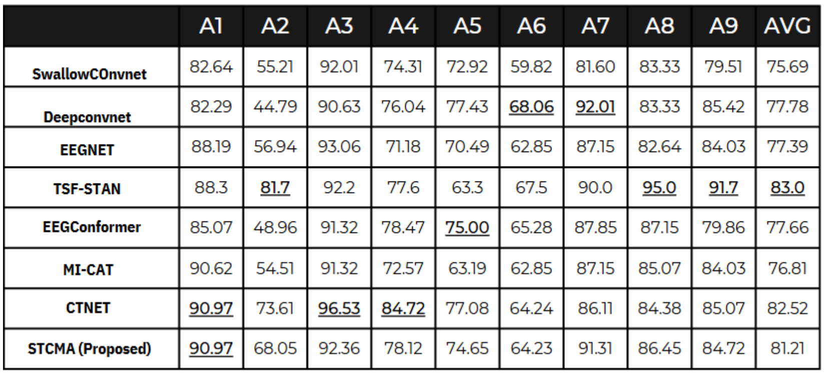


Figure 7.3: STCMA final

Final STCMA: Optimized Design with Temp-Spat Block and Mamba Encoder

The final STCMA architecture represents a significant redesign that synergizes CNN-based feature extraction with the Mamba state-space encoder, resulting in improved performance and faster inference. Figure 6.10 depicts the final architecture, which is composed of three main components: (1) a CNN-based Temporal-Spatial feature extraction block, (2) a Mamba encoder, and (3) a classification head. In this design, the Temp-Spat block is a dedicated convolutional module that sequentially (or in some configurations, concurrently) handles temporal and spatial filtering of the EEG signals. Concretely, the block first applies temporal convolutions to each EEG channel – essentially learning frequency-specific filters that capture temporal patterns (e.g. rhythms in the EEG associated with motor imagery). This is followed by spatial convolution (such as depthwise or pointwise convolutions across channels) to learn spatial filters that combine information from multiple electrodes. This separation of temporal and spatial feature extraction is known to be effective in EEG signal processing, as it mirrors the idea of bandpass filtering (for temporal frequencies) followed by spatial filtering (as in common spatial patterns). The Temp-Spat CNN block in the final STCMA is deeper and more optimized than in v1, yet more streamlined than the overbuilt v2; it provides a rich, low-dimensional representation of the input EEG by the time it reaches the next stage.

Following the CNN block, the Mamba encoder takes the extracted features as input. The Mamba encoder is a Selective State-Space Model layer that excels at modeling sequential data with long-range dependencies. In the final STCMA, we use six Mamba SSM blocks as the temporal sequence modeling component. This Mamba layer maintains the ability to capture temporal dynamics over the entire trial length of EEG, effectively serving a role similar to a recurrent neural network or Transformer encoder, but with far greater computational efficiency. In fact, the Mamba architecture operates with linear-time complexity in sequence length, as opposed to the quadratic complexity of self-attention in Transformers, which makes it highly efficient for long signals. By integrating this Mamba encoder, the model can selectively propagate and update latent temporal states, capturing the evolution of motor imagery patterns over time without blowing up computation or memory. Finally, the classification head of the network takes the encoded feature sequence (often after a global pooling or flattening of the Mamba output) and produces the class probabilities. This head is typically a small fully-connected (dense) layer or series of layers culminating in a softmax output for the four motor imagery classes of the BCI Competition IV 2a dataset. In our implementation, we kept the classification head lightweight to preserve fast inference. The overall architecture is thus a cascaded design: CNN feature extractor → Mamba sequence modeler → classifier. Importantly, this final STCMA architecture is not just more accurate but also more computationally efficient than the earlier versions. By using a single powerful Mamba block (instead of multiple sequential layers) and a well-designed CNN front-end, the model achieves a good balance between complexity and performance. This efficient design yields faster inference times, an important advantage for real-time BCI applications where low latency is critical. In summary, the final STCMA leverages the CNN to handle local feature extraction and the Mamba SSM to handle long-range temporal modeling, a combination that plays to the strengths of each component.



	A1	A2	A3	A4	A5	A6	A7	A8	A9	AVG
SwallowCOnvnet	82.64	55.21	92.01	74.31	72.92	59.82	81.60	83.33	79.51	75.69
Deepconvnet	82.29	44.79	90.63	76.04	77.43	68.06	92.01	83.33	85.42	77.78
EEGNET	88.19	56.94	93.06	71.18	70.49	62.85	87.15	82.64	84.03	77.39
TSF-STAN	88.3	81.7	92.2	77.6	63.3	67.5	90.0	95.0	91.7	83.0
EEGConformer	85.07	48.96	91.32	78.47	75.00	65.28	87.85	87.15	79.86	77.66
MI-CAT	90.62	54.51	91.32	72.57	63.19	62.85	87.15	85.07	84.03	76.81
CTNET	90.97	73.61	96.53	84.72	77.08	64.24	86.11	84.38	85.07	82.52
STCMA (Proposed)	90.97	68.05	92.36	78.12	74.65	64.23	91.31	86.45	84.72	81.21

Figure 7.4: Comparison of the results

Performance and Generalization The proposed final STCMA architecture delivered markedly improved results compared to its predecessors (v1 and v2) and other competing models. As summarized in Figure 6.11, in subject-dependent evaluations on the BCI

Competition IV 2a dataset (within-subject training and testing), STCMA achieved an average classification accuracy of 81.21%, which is a substantial improvement over the earlier versions. This accuracy not only surpasses the v1 and v2 models by a wide margin, but also exceeds the performance of other state-of-the-art models reported on the same dataset. For instance, an advanced deep learning framework combining filter bank common spatial patterns with a CNN classifier achieved about 78.22% average accuracy on BCI 2a – by comparison, STCMA's 81.21% represents a new high in this domain. In addition to its strong accuracy, the final model maintained efficient inference, owing to Mamba's linear scalability and high throughput. This means that STCMA not only provides better accuracy but can also operate with lower latency, an important consideration for practical BCI systems.

We also compared the performance of our model in comparison to other well known architectures in terms of Memory consumption and inference speed. It was found that, the final version of STCMA achieved of at least 1.5 times faster inference speed and 4x less GPU Memory consumtion (1.5 GB). The comparison was based on our STCMA vs the other transformer based models from Figure 7.4.

We also assessed the model's generalization capability in a subject-independent context using a leave-one-subject-out (LOSO) evaluation. In this challenging setup, the model is trained on data from all but one subject and tested on the held-out subject, to gauge how well it can generalize to unseen individuals. The STCMA attained an average accuracy of 71.18% in the LOSO scenario, with a corresponding Cohen's kappa score of 0.6157. This kappa value (0.62) indicates a moderate to substantial agreement between the model's predictions and the ground truth, after accounting for chance agreement. The subject-independent accuracy, as expected, is lower than the subject-specific training accuracy (71.18% vs 81.21%), reflecting the well-known variability in EEG patterns across different people. Nonetheless, an accuracy above 70% in a 4-class classification task with no subject-specific calibration is an encouraging result, on par with or better than many existing approaches under similar cross-subject conditions. It demonstrates that the STCMA is learning features that are not purely subject-specific but have some general relevance to the motor imagery tasks. At the same time, the drop in performance from within-subject to LOSO highlights that there is still significant room for improving the model's generalization. Factors like individual differences in EEG signatures and the limited amount of training data per subject pose challenges that the current model partially mitigates but does not fully overcome.

In conclusion, the proposed STCMA, through its iterative development, has evolved into a robust architecture that effectively captures the spatial-temporal characteristics of EEG signals. The final design (CNN Temp-Spat block + Mamba encoder + classifier) addresses the shortcomings of the initial versions and achieves state-of-the-art performance on the BCI Competition IV 2a dataset. It outperforms other contemporary models in accuracy while also offering faster inference (at least 1.5 times in comparison with other models from the table), fulfilling both accuracy and efficiency criteria. The subject-dependent results underline the model's efficacy when trained on individual-specific data, and the subject-independent results provide insight into its generalization capabilities. This progression of results validates the architectural choices made in STCMA and suggests that combining convolutional feature extraction with a powerful state-space model is a promising direction for advanced BCI signal classifiers. The insights gained from STCMA's development will inform future work, where further enhancements (such as integrating more explicit spatial structure or pre-training strategies) could narrow the generalization gap and push performance even higher.

Hybrid Mamba-Transformer architectures
For the next part, we tried to test the combination of both Mamba State Space Models' GPU optimizations and Transformers architectures. For this part we introduce several our custom hybrid models and the a model a based on the use of Nvidia Hymba [25]. The experiments were based similarly on BCI IV 2a dataset. As for the architecture we used the same CNN

processing layers from STCMA final version.

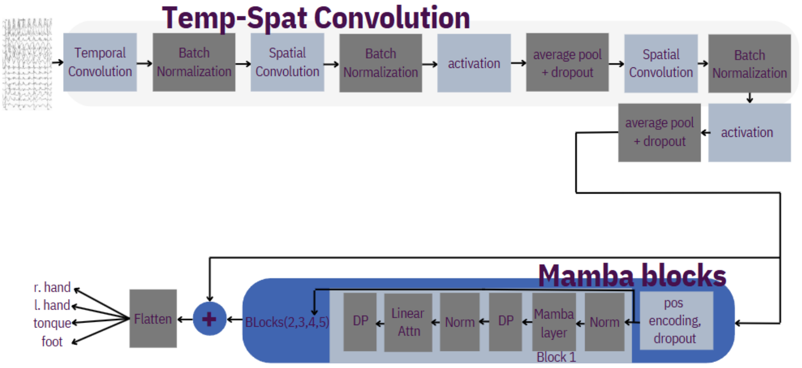


Figure 7.5: Hybrid Mamba, Linear Attention

Figure 7.5 represents our hybrid model based on the Mamba SSM block followed by a linear attention. The Temp-Stat Convolution Layers is totally the same as in the proposed STCMA. Throughout, the evaluation of this model we tried to experiment with a different number of Mamba+linear attn layers. Nevertheless, it was found that this parameter does not significantly affect the results. Overall, the average accuracy was lower in comparison to our raw Mamba SSM based model (STCMA final version), as we got average accuracy of 78.3%.

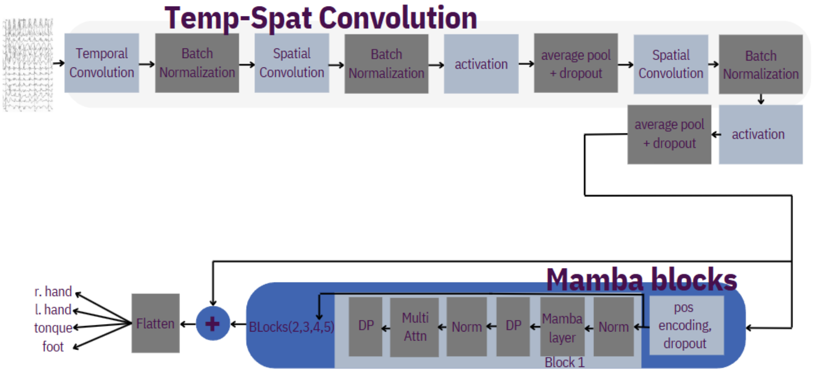


Figure 7.6: Hybrid Mamba, Multihead Attention

Figure 7.6 represents similar hybrid model, but this time based on MultiHead Attention block. Observing the same results for the depth of the Mamba+attention layers, we in addition found that the classification results for subject-dependent classification were similar to the previous Hybrid model, with slightly higher average accuracy of 78.6%.

These results, have left us with the idea that probably, the combination of Mamba SSM with the Transformers' attention mechanisms may lead to incompatibility in regard to the BCI domain. Therefore, next, we investigated the Hymba model [25] proposed by the Nvidia. Hymba is a family of small language models developed by NVIDIA that combines transformer attention mechanisms with state space models (SSMs) in a hybrid-head architecture. This design enables efficient processing by using attention heads for high-resolution recall and SSM heads for context summarization, leading to improved performance and reduced computational overhead [25].

Figure 7.7 demonstrates our architecture based on Hymba blocks. As can be seen, the

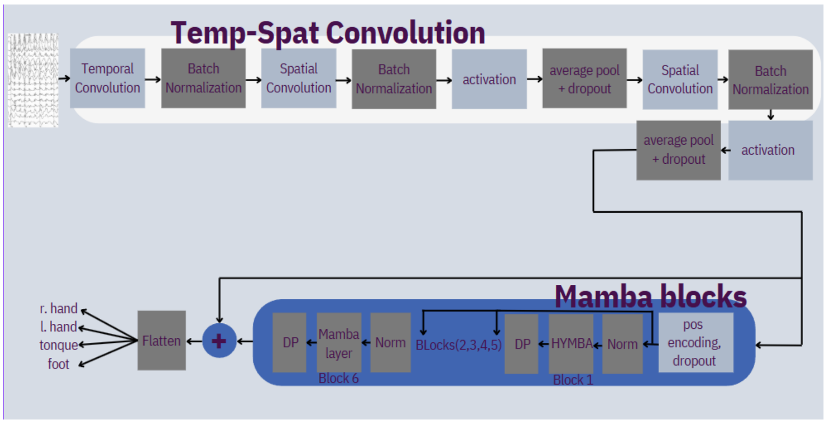


Figure 7.7: Hybrid Mamba based on Hymba block

architecture is similar to Figure 7.3, but with the difference in the Mamba block being replaced by the Hymba block. Regarding the results in accuracy for motor-imagery classification, we achieved 78.9% in average accuracy for the subject-dependent classification.

Therefore, the results achieved by the use of Hymba proved our theory of possible incompatibility of Mamba SSM with Transformers attention mechanisms, as the results achieved by pure Mamba SSM blocks (STCMA v.final) provide higher accuracy.

Chapter 8: **Challenges and Solutions**

The application of the Mamba model to EEG data presents several challenges. One significant hurdle is the **lack of large-scale, well-annotated EEG datasets**. The scarcity of such data limits the model's ability to learn robust representations and generalize to unseen data. Additionally, **overfitting** is a common issue due to the high-dimensional nature of EEG signals and the potential for the model to memorize the training data rather than learning generalizable patterns. Another challenge arises from the **noisy nature of EEG data**, which can be contaminated by artifacts such as muscle activity, eye movements, and external interference. Furthermore, the **high variability of EEG signals** across individuals and experimental conditions can make it difficult to train a model that performs consistently well on different subjects and tasks. Finally, **intersubject variability** poses a significant challenge, as EEG signals can vary greatly between individuals due to differences in brain anatomy, physiology, and recording conditions.

Dependency Management Challenges

The installation of the Mamba SSM was another obstacle in our way. Specifically, Mamba SSM is widely popular due to its inference speed. This advantage is based on some difficult Hardware trick. As far as we understood, a core library for the implementation of this "trick" is called Triton. Triton is an open-source programming language resealed by OpenAI, which enables its users to write efficient GPU based solutions. The specific problem for our team was the fact this Triton is purely Linux based library.

As a result, initially we attempted to work with the Triton and Mamba in Uubntu Dualbot. Nevetheless, this methodology ended in CUDA installation troubles that crashed the Ubuntu and all of the other Recovery cores. As it was found, these issues arised because of some incompatibilities between our Nvidia GPU 1660 TI and Ubuntu 20 Dualbot.To be specific, the challenge was associated with constant changes in the version of CUDA toolkit in the automatic recommendations by our system followed by the emergence of a large amount of errors in other drivers.

Therefore, next we tried to compile Mamba SSM with the Windows operating system. Several custom wheels have been found during the research in the browser across different forums. Specifically, we found the wheel for Mamba under the Windows as OS which required python 3.11. But even if it was our usual working version for python, as the majority of our local environments are based in py 3.11, another Windows wheel for some other requirement was 3.10 py. Therefore, we created totally new local environment with 3.10 python version and observed that neither in 3.10 and 3.11 python versions these two wheels cannot compile together. Further research in the internet resulted in validation by the majority of commentaries that for now it looks impossible to work with Mamba SSM in Windows OS.

For now, it was decided to temporaly work in Google Colab [1] as it is Linux based online system. No problems was observed during this method, except for the fact that it is necessary to reinstall mamba ssm and causal-conv1d libraries. It should be highlighted that while no errors occured, there were severe limitations in free subscription. We had a choice of using CPU, GPU T4 or TPU v2.8 there. And the only operating variant was GPU T4 due to Mamba SSM inner incompatibility with the CPU based systems. The free subscription allowed us only to work with GPU for 1-1.5 hours per day, where preparatory code blocks took in at least half of this time. It also required us to do everything in almost first try, while we had to also consider that we cannot verify the code with the CPU choice, as no Mamba SSM blocks can be run this way.

---

[1] `https://github.com/RakhaOsp/capstone_github_BCI/blob/main/worked_try_mamba.ipynb`

Chapter 9: **Discussions and Conclusions**

Currently, the scope of our findings has been focused on exploring State Space Models, such as LSTM and Mamba, for EEG Motor Imagery classification tasks. We began by using LSTM as a simple and well-established state space model for sequential data. Subsequently, we explored Mamba, a more advanced model, which offered promising results. However, we now aim to use Mamba, leveraging its vision component to capture the spatial relationships inherent in EEG data, which could enhance the model's performance. This part of the research is still to be explored in future work.

Moreover, we noticed that on classification tasks transformer models still perform better than Mamba models, as in classification tasks the epochs of EEG data are sufficiently small to be easily processed by attention mechanisms within transformers. Mamba blocks allow to capture more long sequence data without being overwhelmed, as the mamba selective state space mechanism is linear in progression of a sequence compared to quadratic increase in computational complexity with attention. Therefore, we are interested in showing that Mamba is more reliable and capable of capturing useful information in long sequence data compared to transformer architecture.

At this stage, we observed that relying on single State Space Model blocks—either Mamba or LSTM—combined with several CNN layers for temporal and spatial modeling does not yield classification accuracy surpassing that of existing transformer-based models. Furthermore, the combination of Transformers attention mechanisms with the Mamba blocks presented lower results in comparison to pure Mamba SSM based architectures. Therefore, we anticipate that significant architectural changes will be necessary to improve performance and achieve better accuracy than transformers.

However, we believe the scope of our research should extend beyond a simple comparison of classification accuracy. The initial advantages of Mamba models, particularly their adaptability for hardware implementations and faster inference compared to transformer models, should be emphasized. Mamba models possess linear complexity, whereas transformers have quadratic processing complexity, which makes them more computationally efficient, especially in real-time applications requiring fast inference and low-latency processing.

Thus, we intend to conduct further experiments focusing on the inference speed and hardware performance of Mamba models. This will provide a more comprehensive understanding of their potential applications in EEG data processing, particularly in resource-constrained or real-time environments. Moreover, we recognize the need for broader validation of Mamba models across a wider range of datasets. Since current evaluations are limited to the Physionet and BCI IV 2a datasets, testing Mamba models on additional datasets and EEG-related tasks will be essential to assess their generalizability.

In conclusion, while we have made significant progress in exploring the use of State Space Models like LSTM and Mamba for EEG Motor Imagery classification, future work will focus on integrating Vision Mamba to capture spatial relationships in EEG data. Additionally, further experimentation is needed to explore architectural innovations and validate the models across diverse datasets, as well as to evaluate their efficiency in real-time EEG processing. If successful, these models could become valuable tools for next-generation brain-computer interfaces (BCIs) and neurotechnologies.

Chapter 10: **Future Work**

---

1. Designing an Efficient Vision Mamba for EEG Data:

Given the unique characteristics of EEG data, such as its high temporal resolution and the presence of artifacts, it is essential to adapt the Mamba model for optimal performance. Future work should focus on:

- **Temporal Convolutional Networks (TCNs):** Incorporating TCNs to capture the temporal dependencies in EEG signals more effectively.
- **Artifact Removal:** Developing techniques to automatically remove common artifacts, such as eye blinks and muscle movements, to improve data quality.
- **Feature Engineering:** Exploring new feature engineering techniques tailored for EEG data, such as wavelet transforms and frequency band analysis.

2. Self-Supervised Pretraining for Vision Mamba:

Self-supervised learning has shown promising results in various domains, including computer vision. Applying this technique to the Vision Mamba model could enhance its performance on EEG data. Future work should investigate:

- **Masked Autoencoders:** Adapting the Masked Autoencoders principle to EEG data by masking portions of the EEG signals and training the model to reconstruct them.
- **Contrastive Learning:** Exploring contrastive learning approaches to learn meaningful representations from EEG data without explicit supervision.
- **Pre-trained Models:** Leveraging pre-trained models from other domains, such as natural language processing or computer vision, to initialize the Vision Mamba model and improve its performance.

3. Additional Considerations:

- **Domain Adaptation:** Addressing the issue of domain shift between different EEG datasets by exploring domain adaptation techniques.
- **Interpretability:** Developing methods to interpret the learned representations of the Vision Mamba model to gain insights into the underlying brain processes.
- **Real-World Applications:** Investigating the potential applications of the enhanced Vision Mamba model in real-world scenarios, such as brain-computer interfaces for individuals with disabilities.

By addressing these areas, future research can significantly advance the capabilities of the Vision Mamba model and contribute to the development of more effective and reliable EEG-based applications.

Bibliography

---

[1] Y. P. Penev, A. Beneke, K. T. Root, E. Meisel, S. Kwak, M. Diaz, J. Root, M. R. Hosseini, and B. Lucke-Wold, "Therapeutic effectiveness of brain computer interfaces in stroke patients: a systematic review," *Journal of experimental neurology*, vol. 4, no. 3, pp. 87–93, 2023.

[2] N. Wannawas and A. A. Faisal, "Towards ai-controlled fes-restoration of arm movements: Controlling for progressive muscular fatigue with gaussian state-space models," in *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 1–4, IEEE, 2023.

[3] S. Saha, K. A. Mamun, K. Ahmed, R. Mostafa, G. R. Naik, S. Darvishi, A. H. Khandoker, and M. Baumert, "Progress in brain computer interface: Challenges and opportunities," *Frontiers in Systems Neuroscience*, vol. 15, 2021.

[4] T. Hosman, M. Vilela, D. Milstein, J. N. Kelemen, D. M. Brandman, L. R. Hochberg, and J. D. Simeral, "Bci decoder performance comparison of an lstm recurrent neural network and a kalman filter in retrospective simulation," in *2019 9th International IEEE/EMBS conference on neural engineering (NER)*, pp. 1066–1071, IEEE, 2019.

[5] A. Arjun, A. S. Rajpoot, and M. Raveendranatha Panicker, "Introducing attention mechanism for eeg signals: Emotion recognition with vision transformers," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 5723–5726, 2021.

[6] K. M. Gunter, A. Brink-Kjær, E. Mignot, H. B. Sørensen, E. During, and P. Jennum, "Svit: A spectral vision transformer for the detection of rem sleep behavior disorder," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[7] N. Qi, Y. Piao, H. Zhang, Q. Wang, and Y. Wang, "Seizure prediction based on improved vision transformer model for eeg channel optimization," *Computer methods in biomechanics and biomedical engineering*, pp. 1–12, 03 2024.

[8] J. Liu, E. Dong, J. Tong, S. Yang, and S. Du, "Classification of eeg signals based on cnn-transformer model," in *2023 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 2095–2099, 2023.

[9] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," 2024.

[10] Z. Zhou, G. He, Z. Zhang, L. Leng, Q. Guo, J. Liao, X. Song, and R. Cheng, "Benchmarking neural decoding backbones towards enhanced on-edge ibci applications," 2024.

[11] G. Siddhad, S. Dey, and P. P. Roy, "Drowzee-g-mamba: Leveraging eeg and state space models for driver drowsiness detection," 2024.

[12] Y. Gui, M. Chen, Y. Su, G. Luo, and Y. Yang, "Eegmamba: Bidirectional state space models with mixture of experts for eeg classification," 2024.

[13] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," 2024.

[14] and European Union Agency for Fundamental Rights, *The General Data Protection Regulation – One year on – Civil society: awareness, opportunities and challenges.* Publications Office, 2019.

[15] M. of Justice of the Republic of Kazakhstan, "State registers of inventions, utility models, industrial designs, trademarks, appellations of origin of goods, and selection achievements of the republic of kazakhstan.."

[16] World Intellectual Property Organization, "Access to specialized patent information (aspi)." `https://www.wipo.int/aspi/ru/index.html`, 2024. Accessed: 2024-10-13.

[17] S. Panchavati and W. Speier, "Mentality," in *ICLR 2024 Workshop on Learning from Time Series For Health*, 2024.

[18] A. Behrouz and F. Hashemi, "Brain-mamba: Encoding brain activity via selective state space models," in *Conference on Health, Inference, and Learning*, pp. 233–250, PMLR, 2024.

[19] S. Rukhsar and A. Tiwari, "Lightweight convolution transformer for cross-patient seizure detection in multi-channel eeg signals," 05 2023.

[20] Z. Deng, C. Li, R. Song, X. Liu, R. Qian, and X. Chen, "Eeg-based seizure prediction via hybrid vision transformer and data uncertainty learning," *Engineering Applications of Artificial Intelligence*, vol. 123, 05 2023.

[21] A. Arjun, A. Singh, and M. Raveendranatha Panicker, "Introducing attention mechanism for eeg signals: Emotion recognition with vision transformers," vol. 2021, pp. 5723–5726, 11 2021.

[22] I. E. Tibermacine, A. Tibermacine, W. Guettala, C. NAPOLI, and S. Russo, "Enhancing sentiment analysis on seed-iv dataset with vision transformers: A comparative study," in *Proceedings of the 2023 11th International Conference on Information Technology: IoT and Smart City*, ICIT '23, (New York, NY, USA), p. 238–246, Association for Computing Machinery, 2024.

[23] Y. Zhou and J. Lian, "Identification of emotions evoked by music via spatial-temporal transformer in multi-channel eeg signals," *Frontiers in Neuroscience*, vol. 17, 07 2023.

[24] R. D. H. Al-Fahsi, "Eeg motor imagery classification using cnn, transformer, and mlp."

[25] X. Dong, Y. Fu, S. Diao, W. Byeon, Z. Chen, A. S. Mahabaleshwarkar, S.-Y. Liu, M. V. Keirsbilck, M.-H. Chen, Y. Suhara, Y. Lin, J. Kautz, and P. Molchanov, "Hymba: A hybrid-head architecture for small language models," 2024.