Benchmark and Analysis of Deep Learning Methods to Classify EEG Motor Movement / Imagery Signals

Sathiya Murthi Sankaran
Department of Computer Science
University of Southern California
Los Angeles, California - 90007
sathiyam@usc.edu

Jayanth Jayakumar
Department of Computer Science
University of Southern California
Los Angeles, California - 90007
jjayakum@usc.edu

Jack Praveen Raj Ilango
Department of Computer Science
University of Southern California
Los Angeles, California - 90007
jilango@usc.edu

Pragatheeswaran Ravichandran
Department of Computer Science
University of Southern California
Los Angeles, California - 90007
pragathe@usc.edu

Neranjhana Ramesh
Department of Computer Science
University of Southern California
Los Angeles, California - 90007
neranjha@usc.edu

Sriram Acharya
Department of Computer Science
University of Southern California
Los Angeles, California - 90007
sriramac@usc.edu

Abstract

The study comprehensively evaluates various deep learning strategies to classify Motor Movement / Imagery using Electroencephalogram (EEG) signals, an area pivotal to advancements in brain-computer interfaces (BCIs). This study spans several general Deep Learning methods and previously published works on the task. Focusing on the BCI Competition IV 2A dataset, we explored a range of pre-processing and feature extraction techniques, implemented and evaluated the performance of the benchmarked models, and tested the impact of Data Augmentation. The study also provides an additional perspective by testing the robustness of these models to the addition of noise at different SNR levels to replicate real-world data quality. Our findings provide crucial insights that could facilitate the development of more accurate and robust models.

1 Introduction

Electroencephalography (EEG) captures the brain's electrical activity and is widely utilized in both medical diagnostics and research applications. Accurate classification of EEG data is pivotal for advancing neurotechnology and enhancing brain-computer interface (BCI) systems. These technologies are integral to developments in cognitive load monitoring, medical diagnostics, Bio-metric authentication, and more. The classification of EEG signals is crucial in facilitating effective interactions between humans and machines, aiding in rehabilitation, and improving assistive technologies.

Our paper focuses on Deep Learning and Machine Learning models which are specifically tailored for EEG signal classification. It examines the existing models and potential enhancements to improve the models' effectiveness and reliability are explored. This analysis is centered around the BCI Competition IV 2A dataset instrumental for benchmarking the performances of these models for motor imagery tasks.

1.1 EEG Signals

Electroencephalography (EEG) is a non-invasive technique widely used to monitor brain activity by measuring electrical signals generated by neurons. Electrodes are placed on the scalp which capture these signals, which represent the synchronized firing of neural populations. EEG recordings provide valuable insights into brain function, enabling researchers to study cognitive processes, neurological disorders, and brain-computer interfaces (BCIs). EEG's versatility allows for diverse applications, including diagnosing epilepsy, assessing sleep patterns, and investigating cognitive impairments. Moreover, its real-time monitoring capabilities make EEG instrumental in developing BCIs, which could enable individuals to control devices using their brain activity.

1.2 Dataset

The BCI Competition IV 2A dataset is a collection of EEG signal data from 9 different subjects, each participating in 2 sessions with 48 trials per class. This amounts to a total of 288 trials per session. Data was collected for the following classes using 22 electrodes:

- Left Hand (class 1)
- Right Hand (class 2)
- Both Feet (class 3)
- Tongue (class 4)

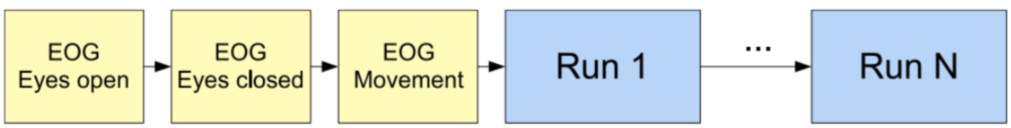

Figure 1: Timing scheme of each session

The dataset is divided into two sets with a total of 5184 samples:

- Train data: 2592 trials (648 trials per class)
- Test data: 2592 trials (648 trials per class)

1.2.1 Dataset Format

The data is stored in the General Data Format (GDF), storing one file per subject and session. Only one session per subject contains class labels for all trials, while the other session was intended to be used for classifier testing and performance evaluation for the competition. Each trial is associated with a target label indicating the intended motor imagery/movement (e.g., left-hand movement or right-hand movement).

1.2.2 Data Collection

In this dataset, the signals were originally sampled at 250 Hz. It was then bandpass-filtered between 0.5 Hz and 100 Hz. Further, an additional 50 Hz notch filter was applied to suppress line noise. The amplifier sensitivity determines the scale of the recorded signals. Here, it was set to 100 μV for EEG channels and 1 mV for EOG channels. To record this dataset, twenty-two Ag/AgCl electrodes were used with the left mastoid serving as the reference and the right mastoid as the ground. This was performed following the international 10-20 system.

1.2.3 Technical Specifications

- 25 electrodes are used, first 22 are EEG, last 3 are EOG
- Sampling frequency (f_s) is 250Hz
- 9 subjects

Figure 1 shows the timing scheme of each trial. Each trial lasts for 7-8 seconds, sampled at 250Hz.

- 1 trial = 7-8s
- 1 run = 48 trials = 336-384s
- 1 session = 6 runs = 288 trials = 2016-2304s

1.2.4 Additional EOG Channels Data

In addition to the EEG channels, three monopolar EOG channels were recorded. Before the start of each session, EOG influence was sampled with the use of a 5-minute recording. This was further divided into three blocks:

- Eyes Open
- Eyes Closed
- Eye Movements

2 Related Work

- **[1] Vernon J Lawhern et al.** presents EEGNet, a compact Convolutional Neural Network designed for EEG-based BCI classification and interpretation. The model uses depth-wise and separable convolutions which help enhance EEG feature extraction while also reducing trainable parameters. The model was tested across four BCI paradigms - P300, ERN, MRCP, and SMR - demonstrating improved classification performance with limited training data and effectively generalized across all paradigms. EEGNet efficiently achieves comparable performance to more specialized models with significantly fewer parameters. In addition to this, EEGNet facilitates extraction of neuro-physiologically interpretable features which are crucial for neuroscience applications where understanding CNN-derived phenomena is vital. The design of the model is validated across multiple BCI paradigms ensuring that the performance is not affected by noise.
- **[2] Brenda E. Olivas-Padilla et al.** propose a DeepConvNet model architecture is a monolithic deep convolutional neural network model which is optimized for classifying multiple motor imagery tasks using EEG signals. This model uses a complex architecture consisting of several convolutional layers, which extract features from EEG data processed through Discriminative Filter Bank Common Spatial Patterns (DFBCSP). The optimization of the network's hyperparameters is performed using Bayesian optimization to maximize the classification accuracy. The final model achieves an impressive classification accuracy, with the monolithic network attaining an average accuracy of 78.41% across the dataset used, demonstrating a robust capability in handling the multi-class motor imagery classification challenge with a relatively high degree of precision, especially given the complex nature of the EEG data and the multiple motor imagery tasks involved
- **[3] Hermosilla et al.** propose a ShallowConvNet model architecture to address the unique challenges of EEG signal classification in brain-computer interfaces (BCIs) i.e., the high dimensionality but relatively low complexity of EEG data in comparison to data types like images. Other deep learning models such as deep Convolutional Neural Networks (CNNs) can be overly complex and prone to overfitting due to the small datasets sizes typically available for BCI. The ShallowConvNet architecture was designed with fewer layers and filters to efficiently capture the spatial and temporal features necessary for MI task classification, while being computationally less demanding and more robust against overfitting. In the study, this model shows greater performance when compared to more complex systems such as DeepConvNet (which achieved an accuracy of 70.9%), achieving a promising classification results of 73.7% on the same dataset.
- **[4] Abbas Salami et al.** present a deep learning architecture - EEGITNet - with fewer trainable parameters, optimized for converting outputs into intuitive visualization structures like topographic maps, alongside a methodology for achieving this transformation. It reports significant improvements in signal classification on a motor imagery BCI dataset across various scenarios, outperforming other end-to-end architectures. The model enhances its efficiency and performance by replacing standard convolutions with inception modules

and depth-wise causal convolutions with dilation, achieving superior results with reduced complexity compared to similar models.

- **[5] Yonghao Song et al.** introduce a compact Convolutional Transformer named EEG Conformer, designed to encapsulate both local and global features within a unified EEG classification framework. The convolution module of the EEG Conformer is tasked with learning low-level local features through one-dimensional temporal and spatial convolution layers. It is complemented by a self-attention module, which extracts global correlations from these local temporal features. A simple classifier module consisting of fully-connected layers then follows to predict categories for EEG signals. To make it interpretability, a visualization strategy is also developed to project class activation mapping onto brain topography. Extensive experiments have been conducted to evaluate the effectiveness of this method across three public datasets, focusing on EEG-based motor imagery and emotion recognition paradigms.
- **[6]Hamdi Altaheri et al.** employs a deep learning model with interpretable and explainable features, ATCNet, multihead self-attention to highlight the most valuable features in MI-EEG data, temporal convolutional network to extract high-level temporal features, and convolutional-based sliding window to augment the MI-EEG data efficiently.
- **[7] Danilo Avola et al.** present a benchmark study on Machine and Deep Learning for EEG signal classification. They used four widely known models: Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), to determine which model could be a good starting point for developing EEG classification models. The study highlights the importance of choosing the right model for specific problems and data types in EEG signal analysis.

3 Methodology

3.1 Pre-processing

3.1.1 Band Pass Filtering

Bandpass filtering can be applied to retain only relevant frequency components associated with motor imagery. Motor imagery tasks usually happen in the alpha and beta range of EEG signals.

The raw EEG data from the corpus is filtered between a low frequency of 4 Hz and a high frequency of 38 Hz. The band range includes all the relevant signal information for motor movement tasks. This pre-processing step helps to eliminate non-essential information in the data preserving the efficiency during deep learning approaches.

3.1.2 Exponential Moving Standardization

Exponential Moving Standardization is a technique used to normalize time-series data by applying an Exponentially Weighted Moving Average (EWMA) to the data. Initially, a block of data points is used to compute the mean and standard deviation. As new data points arrive, they are standardized using an exponentially weighted moving average of the mean and standard deviation. The weighting given to each data point decreases exponentially as it moves further away from the present. Each data point is then standardized using the exponentially weighted mean and standard deviation. The purpose of exponential moving standardization is to normalize the data while also adapting to changes in the data distribution over time. It's particularly useful for time-series data like EEG where the statistical properties may vary over different time periods.

3.2 Feature Extraction

3.2.1 Common Spatial Pattern (CSP)

The Common Spatial Patterns (CSP) algorithm finds spatial filters that are useful in discriminating different classes of Electroencephalogram (EEG) signals such as those corresponding to different types of motor activities. CSP involves separating a multivariate signal into additive subcomponents that have maximum differences in variance between two windows. Effective motor imagery decoding relies on temporal, spatial, and frequency features. Here, motor imagery of single limbs is reflected in the μ rhythm (8–13 Hz) and β rhythm (13–30 Hz). However, significant temporal features may not manifest throughout the entire motor imagery process. This distinction between classes using CSP feature extraction can play a cardinal part in increasing the efficiency of classification tasks. Eight frequency band coefficients were chosen between 4 to 32 Hz and the signal was decomposed using Wavelet Packet Decomposition. These feature bands were standardized to generate CSP features and were used to train the general Deep Learning models [MLP, CNN, Conv-LSTM].

3.3 Dataset Loading

3.3.1 Window Data Split

Before the actual train test split of the dataset, we create window samples of the dataset where each window represents a particular subject and target class. Table 2 shows the specification of the sample data split into windows. Following the original dataset specifications, there are 9 subjects and the EEG data is sampled at 250Hz. Here each window represents a trial. Each window created from the samples is 4.5 seconds which corresponds to 1125 time steps per window [4.5 seconds * 250 samples per second]. A particular window corresponds to one of the subject's sessions and has a target class associated with it.

Table 1: Metadata of window data split

i_window_in_trial	**i_start_in_trial**	**i_stop_in_trial**	**target**	**subject**	**session**	**run**
0	0	125	3	1	0train	0
1	0	2128	3	3253	0train	0
2	0	4046	2	5171	0train	0
3	0	5998	1	7123	0train	0
...	...	...	...	...	...	...
43	0	86625	0	87750	1test	5
44	0	88531	3	89656	1test	5
45	0	90459	0	91584	1test	5
46	0	92573	1	93698	1test	5
47	0	94632	0	95757	1test	5

BCI Competition IV 2a is pre-split as 50% of train data and 50% of test data with 9 subjects each in train and test. '0train' represents the windows of train data and '1test' represents the windows of test data. There are 288 trials for each subject which gives 5184 trials [2592 windows in train and test].

3.3.2 Train Test Split

For better training, the pre-mandated 50-50 data split is changed to a subject-wise split. The training session of all 9 subjects and the test session of 6 subjects are taken as the training set. This approach aligns with real-world test data for EEG classification applications. Based on the above specified window data split and subject-wise split specification there are a total of 4320 training samples and 864 testing samples. Table 2 portrays the custom train and test split used for deep learning models.

Table 2: Train and Test Data Splits

Class	Train Samples	Test Samples
0	1080	216
1	1080	216
2	1080	216
3	1080	216

3.4 Training Methodology

Utilizing the custom split train and test data, we implemented and benchmarked various deep learning models. Initially, the general deep learning models like MLP, CNN and ConvLSTM are used to evaluate the classification task. Due to the simplicity of the network in these models, it did not train well on the raw data. EEG corresponds to time series data, the correlation between each time step needs to be captured. To improve the accuracy and make the model data well-trainable we used CSP as the feature extraction technique. The feature-extracted data is then fed into these models providing a more acceptable model that is comparable with the published models. Further, we explored several contemporary models like ShallowConvNet, DeepConvNet, EEGNet(v1,v4), EEGConformer, ATCNet, EEG-ITNet, Deep4Net, and Shallow-FPCSP. Each of these models portrays various deep learning techniques that analyze various aspects of the EEG data and try to capture the relationship to the motor imagery task. The model's hyperparameters were tuned to ensure that they converged sufficiently. Model training curves and evaluation metrics like Accuracy, F1-score, Precision, and Recall were used to compare and benchmark the models.

3.5 Models tested

3.5.1 EEGNet

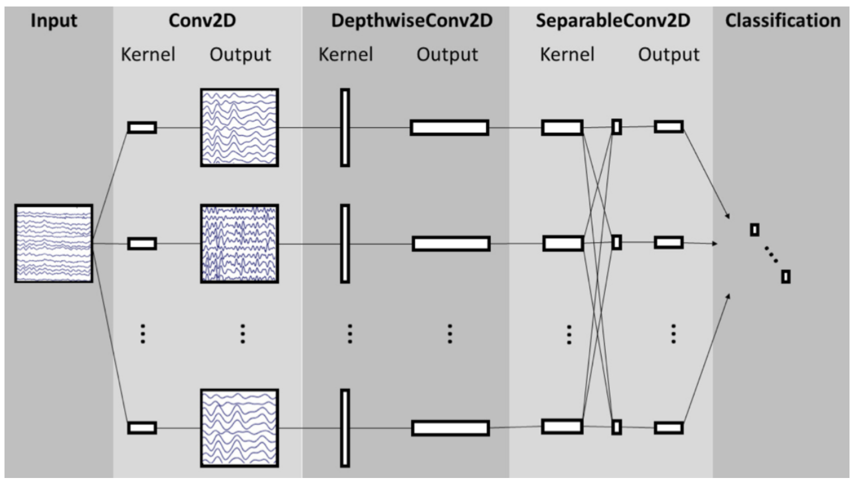

Figure 2: EEGNet architecture

The architecture shown in Figure 2 begins with a temporal convolution layer to learn frequency filters, followed by a depth wise convolution that applies to each feature map individually to extract frequency-specific spatial filters. Next, a separable convolution, which combines a depth-wise convolution (learning a temporal summary for each feature map) followed by a point-wise convolution (optimizing the combination of feature maps), is used. This structure allows EEGNet to effectively process and interpret EEG signals by optimizing the connectivity between inputs and outputs across different convolutional stages.

3.5.2 ShallowConvNet

The architecture of the proposed model as shown in Figure 3 is primarily based on the Shallow Convolutional Network (ShallowConvNet), with several key modifications aimed at enhancing its performance. Notably, the input size has been increased from 500x22x1 to 1000x22x1. The Temporal Convolution Layer in the proposed architecture uses a 45x1 filter with 40 channels, which is wider than the original, producing a 478x22x40 tensor. This increase is achieved by downsampling from a 250 Hz sampling rate to 125 Hz using a stride of 2, allowing the kernel to cover approximately 200 ms and capture a broader range of temporal features. Additionally, the Spatial Convolution Layer consists of 40 channels using a 1x22 filter to process spatial features across all EEG channels. More finely-grained stride settings in both the convolution and pooling layers have been introduced to enhance the resolution and detail of feature extraction, improving the model's ability to capture and classify nuanced temporal and spatial patterns present in EEG signals.

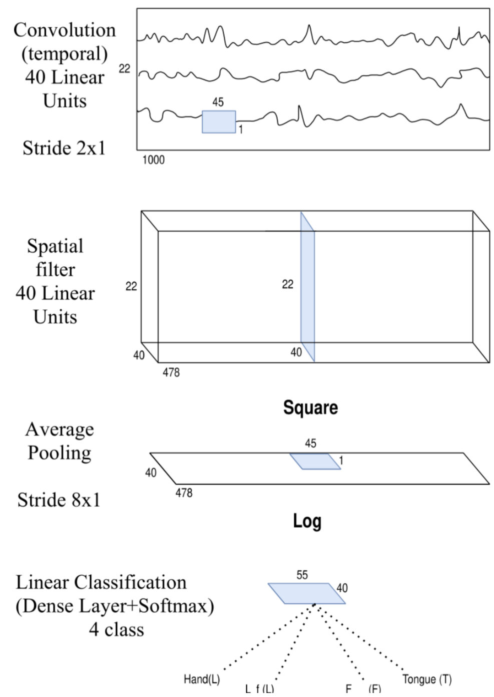

Figure 3: ShallowConvNet architecture

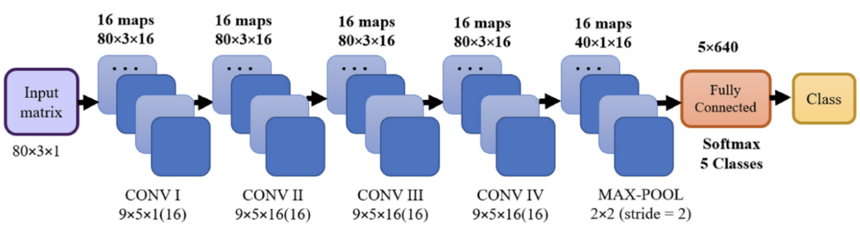

Figure 4: DeepConvNet architecture

3.5.3 DeepConvNet

ShallowConvNet is a simplified CNN that has one feature layer, one convolutional layer, and one average pooling layer. The convolutional layer has filtering and pooling layers to reduce the offset of the estimated value of the parameter error. The ShallowConvNet Excel architecture as shown in Figure 4, consists of multiple convolution layers followed by max-pooling layers to extract hierarchical features from the input EEG data. The output of the convolution layers is flattened and fed into one or more fully connected layers, which perform the final classification based on the learned features.

3.5.4 EEG-ITNet

The architecture comprises four main blocks: the Inception Block, Temporal Convolution (TC) block, Dimension Reduction (DR) block, and Classification block. The inception block uses parallel convolutional layers for frequency and spatial filtering. The TC block utilizes a temporal convolution network architecture with depth-wise causal convolutions and batch normalization to extract key temporal features, ensuring robust performance. The DR block combines these temporal features using a 1×1 convolution and reduces dimensions with pooling to manage feature complexity and prevent overfitting. Finally, the classification block, which can be adapted based on the problem set, employs a softmax activation function in a fully connected layer to perform the classification task, showcasing the network's ability to process and classify EEG signals efficiently.

3.5.5 EEG Conformer

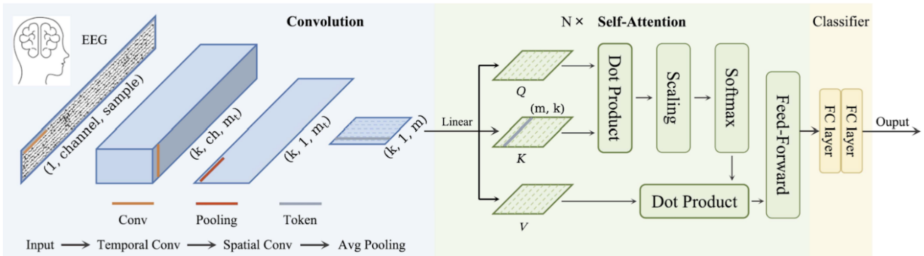

Figure 5: EEG Conformer architecture

The architecture shown in Figure 5, described comprises three main components: a convolution module, a self-attention module, and a fully-connected classifier. The convolution module processes raw two-dimensional EEG trials, applying temporal and spatial convolution layers along the time and electrode channel dimensions, respectively, followed by an average pooling layer to suppress noise and improve generalization. This processed output is then fed into the self-attention module, which extracts long-term temporal features by analyzing the global correlations among different time positions within the feature maps. The architecture concludes with a classifier segment that consists of several fully connected layers, designed to produce the decoding results.

3.5.6 ATCNet

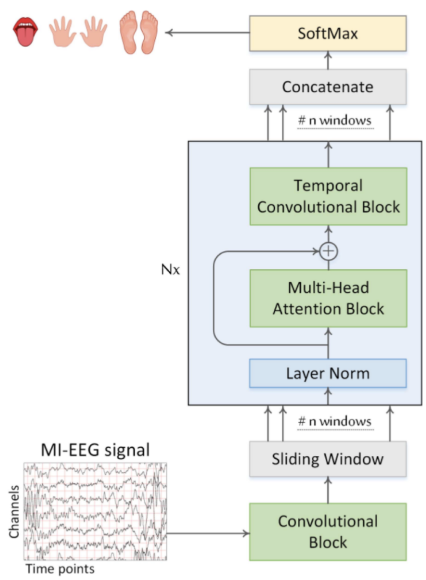

Figure 6: ATCNet architecture

The proposed ATCNet model as shown in Figure 6, consists of three main blocks: Convolutional (CV) block, Attention (AT) block, and Temporal Convolutional (TC) block. CV block encodes low-level spatio-temporal information within the MI-EEG signal through three convolutional layers: temporal, channel depthwise, and spatial convolutions. The output of the CV block is a temporal sequence with a higher level representation. The AT block then highlights the most important information in the temporal sequence using a multihead self-attention (MSA). The TC block extracts high-level temporal features within the temporal sequence using TCN and feeds them into a fully connected (FC) layer with a SoftMax classifier. The temporal sequence, output from CV block, can be split into multiple windows and each is fed to AT/TC blacks separately. The output of all windows is then concatenated and fed to a SoftMax classifier which helps to augment the data and increase accuracy.

3.6 Data Augmentation

In some of the models, it was observed that the training accuracy was relatively higher compared to validation and test accuracies. The observed discrepancy between high training accuracy and lower validation and test accuracies in our model suggests that it may be overfitting to the training data. To mitigate this, data augmentation was employed as a strategy to enhance the generalizability of the model. Data Augmentation artificially expands the training dataset by creating modified versions of training samples, thereby helping the model learn more robust features and reducing overfitting.

Two specific augmentation transformations are utilized: Frequency Shift and Sign Flip. The FrequencyShift transformation is configured with a 50% probability to alter the EEG signal frequencies, adjusting them randomly within a range of -2 to +2 Hz. This shift is intended to simulate natural variations in EEG signal frequencies that might occur due to different physiological or environmental factors, helping the model learn to handle slight frequency variations in input data. The SignFlip transformation, applied with a 10% probability, flips the sign of the EEG signals, essentially mirroring the signal waveform about the time axis. This can be particularly useful for augmenting data in scenarios where the EEG data may include symmetric patterns. Together, these transformations diversify the training dataset, aiming to prevent overfitting and enhance the model's ability to generalize across varied EEG signal conditions not seen during training.

3.7 Robustness to Noise

The robustness of the models to noise was systematically evaluated by introducing Additive White Gaussian Noise (AWGN) to the test datasets. This method of testing is crucial for understanding how the models perform under less-than-ideal conditions that might mimic real-world environmental noise or electronic interference when collecting the signals. AWGN was chosen because it represents a common type of noise in electronic systems and signal processing, characterized by a constant power spectrum across various frequencies, thus serving as a standard benchmark for noise resilience.

To conduct this analysis, AWGN was added to the EEG test data at two distinct signal-to-noise ratio (SNR) levels: 8 dB and 15 dB. These levels were selected to simulate moderate and mild noise conditions, respectively. By testing at different SNRs, the impact of varying noise intensities on the model's performance could be assessed, providing insights into the model's sensitivity and stability across different noise environments.

The performance metrics, such as accuracy, precision, and recall, were recorded before and after the noise addition. The comparative results, encapsulated in Table 5, showcase how each model's performance metrics declined as a result of noise interference.

4 Experiment Results and Discussions

Figure 7 shows a few of the sample training curves for — MLP, EEGNet, Conv-LSTM, and DeepConvNet. Among these, it can be seen that EEGNet demonstrates a better generalization capability as evidenced by its closely aligned training and validation loss curves and a minimal discrepancy between training and validation accuracies. This contrasts with the overfitting observed in both the MLP and Conv-LSTM models that we trained on CSP features, where substantial gaps between training and validation accuracies were noted.

Table 3 shows the performance metrics for various EEG signal classification models. ATCNet outperforms other models with the highest accuracy of 67.82% and an F1 score of 0.6751, indicating superior performance in accuracy, precision and recall. EEGNet-v4 also shows a robust performance with an accuracy of 65.16% and F1 score of 0.6439. The EEGConformer and CNN models exhibit the lowest accuracies on the BCI data.

Table 4 presents a comprehensive comparison of model performances enhanced by data augmentation techniques and highlights the improvements in performances. Among the models, ATCNet achieves highest increase in F1 score of 4.59% which indicates its adaptability. EEGNet-v4, EEGConformer, ShallowFBCSP also show improvements in performances. However, the transformations we have used for augmentation did not seem to help the performance of EEGNet-v1, Deep4Net, and EEGITNet models.

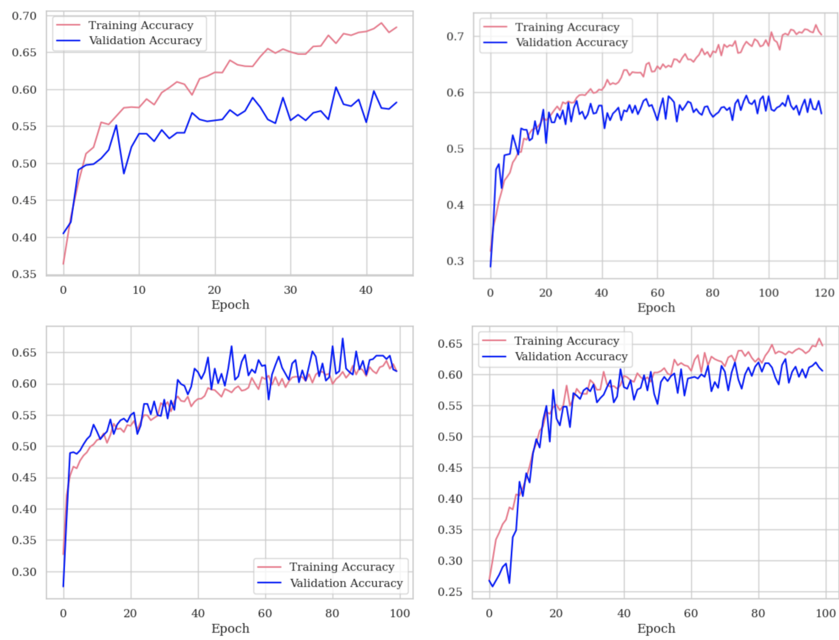

Figure 7: Training curves-Accuracy. Top left: MLP. Top right: Conv LSTM. Bottom left: EEGNet. Bottom right: DeepConvNet.

The data in Table 5 highlights the robustness of various EEG classification models under different signal-to-noise ratio (SNR) conditions. It's evident that the performance of all models deteriorates with the addition of noise to the test data. When the SNR is changed from 15 to 8, the extent of the performance drop varies significantly across models. Models such as ShallowFBCSP and EEGConformer demonstrate remarkable resilience to noise with minimal F1 score reductions of 0.48% and 1.96% respectively. In stark contrast, generic models like MLP, CNN, and Conv-LSTM experience drastic performance declines, exceeding 70% at an SNR of 8, reflecting their vulnerability to noise. This substantial variation underscores the efficiency of the models designed specifically for EEG Classification.

Of the tested models, ATCNet model has the highest accuracy while also showing improvements to our data augmentation techniques.

Table 3: Comparison of Model Performances

Model	**Feature Extraction**	**Accuracy**	**F1**
EEGNet-v4	Raw-filtered	0.6516	0.6439
ShallowConvNet	Raw-filtered	0.6145	0.6160
DeepConvNet	Raw-filtered	0.6099	0.5893
ATCNet	Raw-filtered	0.6782	0.6751
EEGConformer	Raw-filtered	0.5555	0.5505
ShallowFBCSPNet	FBCSP	0.6157	0.6145
EEGNet-v1	Raw-filtered	0.6064	0.6025
Deep4Net	Raw-filtered	0.5763	0.5855
EEGITNet	Raw-filtered	0.5833	0.5894
MLP	CSP	0.5637	0.5621
CNN	CSP	0.5208	0.5202
Conv-LSTM	CSP	0.5868	0.5864

Table 4: Comparison of Model Performances with Augmentation

Model	Feature Extraction	Accuracy	F1	% increase in F1
EEGNet-v4	Raw-filtered	0.6678	0.6679	3.727%
ATCNet	Raw-filtered	0.7072	0.7061	4.591%
EEGConformer	Raw-filtered	0.5637	0.5588	1.507%
ShallowFBCSPNet	FBCSP	0.6296	0.6296	2.457%
EEGNet-v1	Raw-filtered	0.5648	0.5490	-8.879%
Deep4Net	Raw-filtered	0.5277	0.4964	-15.217%
EEGITNet	Raw-filtered	0.5197	0.4839	-17.899%

Table 5: Comparison of Model Performances with Different SNR Levels

Model	Feature Extraction	F1 (SNR = 15)	% drop in F1	F1 (SNR = 8)	% drop in F1
EEGNet-v4	Raw-filtered	0.6131	4.78	0.4876	24.27
ShallowConvNet	Raw-filtered	0.5622	8.73	0.3848	37.53
DeepConvNet	Raw-filtered	0.5777	1.96	0.5479	7.02
ATCNet	Raw-filtered	0.6064	10.18	0.4573	32.18
EEGConformer	Raw-filtered	0.5386	2.15	0.5143	6.55
ShallowFBCSPNet	FBCSP	0.6115	0.48	0.5581	9.18
EEGNet-v1	Raw-filtered	0.5723	5.02	0.4530	24.83
Deep4Net	Raw-filtered	0.6039	3.15	0.5769	1.47
EEGITNet	Raw-filtered	0.5537	6.07	0.5381	8.69
MLP	CSP	0.2287	59.37	0.1097	80.45
CNN	CSP	0.2216	57.24	0.1049	79.84
Conv-LSTM	CSP	0.2557	56.55	0.1654	71.78

5 Limitations

One of the major challenges that we anticipated was the reproduction of results of large models. While we were able to implement and test most of the models that we intended to, in our attempt to benchmark FusionNet and FusionNetv2 for EEG data classification, we encountered significant compute limitations that hindered our ability to evaluate its performance.

FusionNet's design incorporates multiple branches tailored to accommodate different EEG configurations, each comprising distinct convolution blocks optimized for the characteristics of the input data. These branches process EEG data independently, allowing FusionNet to capture and extract relevant features from diverse EEG sources efficiently. While implementing this, we faced memory issues due to the increased computational demands of running multiple branches simultaneously.

Another challenge arises from the limitation of the dataset which contains samples from only 9 subjects. A small sample size may not fully capture the variability and complexity of real-world scenarios, leading to concerns about the generalization and robustness of the trained models. But this doesn't affect our benchmark, as all the models are tested on the same dataset, ensuring consistency and providing a fair basis for comparison between different architectures despite the limited sample size. Testing these models on synthetic data could provide some valuable insights.

6 Future Work

- Extending the Red-team attacks to other distortions that could potentially affect real-world data collection could help improve the robustness of EEG classification models by identifying vulnerabilities and testing the model's resilience.
- Embedding techniques such as Graph Embedding and Auto Encoder could be used to transform raw EEG data into a more meaningful and compact representation, enhancing model performance and interpretability. These techniques can be particularly useful in dealing with the high-dimensional nature of EEG data.

- Access to better and higher computational resources can be crucial to test and evaluate the available models.
- Extending the benchmark across other datasets to check cross-dataset generalization is another direction that can be pursued.

7 Conclusion

This study has conducted a comprehensive analysis and benchmark of various Deep Learning models for classifying EEG signals associated with motor imagery. Our findings indicate that while some models demonstrate better performance, challenges such as model complexity, computational demand, and vulnerability to noise persist. The ATCNet and EEGNet-v4 have shown particularly promising results in terms of accuracy and robustness against noise, making them suitable candidates for further development and potential clinical application. However, the limitations observed due to the small dataset size and model complexity highlight the need for ongoing research to enhance the generalization and efficiency of these models. Overall, the insights gained from this research pave the way for future developments in EEG-based BCI technologies, with the ultimate goal of improving the interface between humans and machines for various practical applications.

References

[1] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," Journal of Neural Engineering, vol. 15, no. 5, pp. 056013, Jul. 2018. [Online]. Available: http://dx.doi.org/10.1088/1741-2552/aace8c

[2] B. E. Olivas-Padilla and M. I. Chacon-Murguia, "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters," *Applied Soft Computing*, vol. 75, pp. 461-472, 2019. doi: 10.1016/j.asoc.2018.11.031

[3] D. Milanés Hermosilla, R. Trujillo Codorniú, R. López Baracaldo, R. Sagaró Zamora, D. Delisle Rodriguez, Y. Llosas Albuerne, and J. R. Núñez Álvarez, "Shallow Convolutional Network Excel for Classifying Motor Imagery EEG in BCI Applications," *IEEE Access*, vol. 9, pp. 98275-98286, 2021. doi: 10.1109/ACCESS.2021.3091399

[4] Salami, A., Andreu-Perez, J. & Gillmeister, H. (2023) EEG-ITNet: An Explainable Inception Temporal Convolutional Network for Motor Imagery Classification. In *Proceedings of the Neural Information Processing Systems Conference (NeurIPS)*, Vancouver, Canada.

[5] Song, Y., Zheng, Q., Liu, B. & Gao, X. (Year) EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. In *Proceedings of the IEEE Conference on Neural Networks*

[6] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-Informed Attention Temporal Convolutional Network for EEG-Based Motor Imagery Classification," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2249-2258, 2023. doi: 10.1109/TII.2022.3197419

[7] D. Avola, M. Cascio, L. Cinque, A. Fagioli, G. L. Foresti, M. R. Marini, and D. Pannone, "Analyzing EEG Data with Machine and Deep Learning: A Benchmark," arXiv:2203.10009 [cs.LG], Mar. 2022.

[8] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008 – Graz Data Set A," in Institute for Knowledge Discovery, Graz University of Technology, Austria, and Institute for Human-Computer Interfaces, Graz University of Technology, Austria, 2008.

[9] R. R. Chowdhury, Y. Muhammad, and U. Adeel, "Enhancing Cross-Subject Motor Imagery Classification in EEG-Based Brain-Computer Interfaces by Using Multi-Branch CNN," *Sensors*, vol. 23, no. 18, pp. 7908, Sep. 15, 2023. doi: 10.3390/s23187908. PMID: 37765965; PMCID: PMC10536894.