Project 3 Report

TOPIC: EEG Classification Model

IE6400 19843

Foundations Data Analytics Engineering SEC 01 Fall 2023

Group Number 7

Aishwariya Alagesan (002818973)

Arya Lokesh Gowda (002249418)

Keerthana Balswamy (002839068)

Deepana Dhakshinamurthy (002834708)

Suja Ganesh Murugan (002839290)

Introduction:

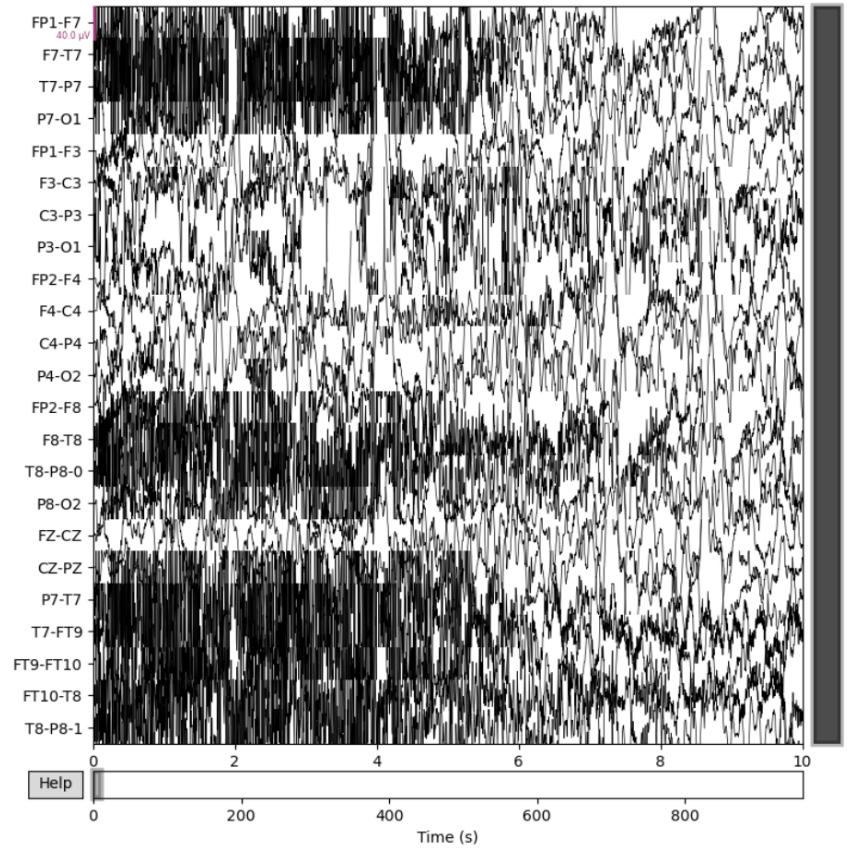
This EEG Classification report presents a comprehensive overview of a dataset containing EEG data that has a diagnosis of epilepsy. In this project, we delve into the intricate task of constructing an EEG classification model, a critical endeavor for deciphering the complexities of brainwave patterns. The application of such models extends to the forefront of medical science, particularly in the diagnosis of epilepsy, where the identification of specific neural signatures is paramount. The report covers various aspects of the dataset, including data preprocessing, Model Selection, Model Training, Model evaluation and performance.

Data Preprocessing:

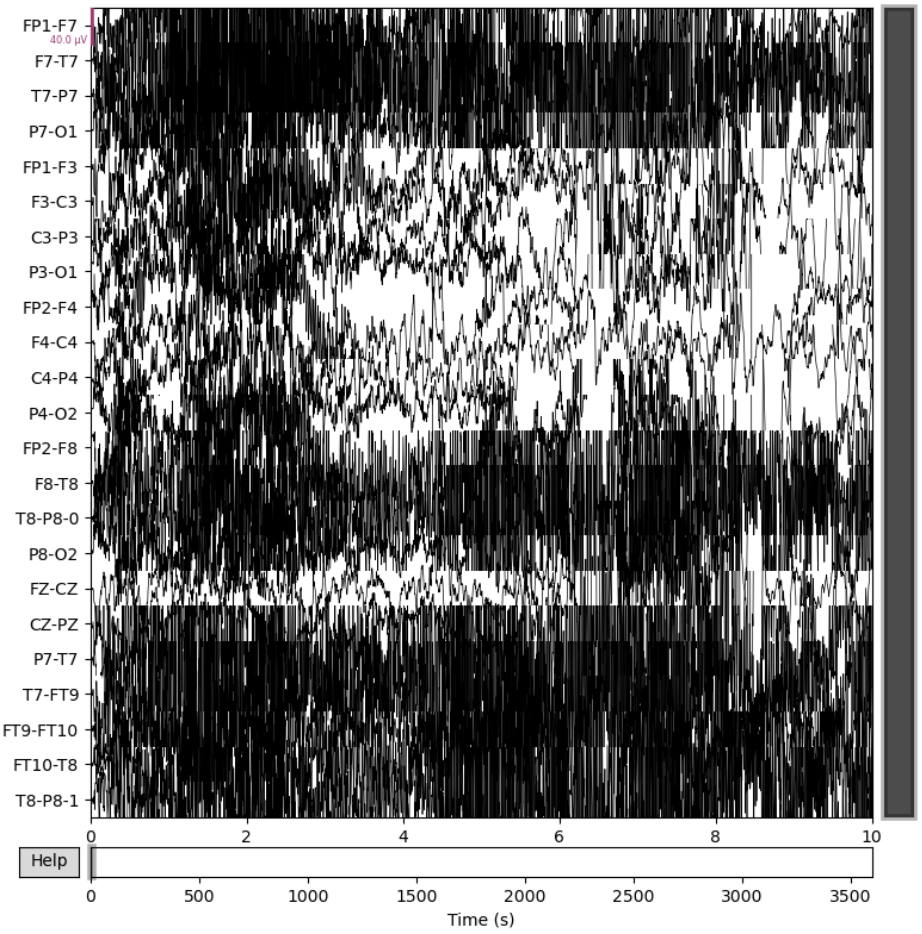
- **Data Preprocessing:**

Upon loading the dataset, we conducted an initial inspection to understand its structure and characteristics. A thorough exploration of the data structure and characteristics precedes the preprocessing of EEG data. This involved addressing missing values, implementing noise reduction techniques, and considering data augmentation as necessary.

```
<Info | 8 non-empty values
 bads: []
 ch_names: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, ...
 chs: 23 EEG
 custom_ref_applied: False
 highpass: 0.0 Hz
 lowpass: 128.0 Hz
 meas_date: 2074-07-25 09:31:46 UTC
 nchan: 23
 projs: []
 sfreq: 256.0 Hz
 subject_info: 1 item (dict)
>
```



```
<Info | 8 non-empty values
 bads: []
 ch_names: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, ...
 chs: 23 EEG
 custom_ref_applied: False
 highpass: 0.0 Hz
 lowpass: 128.0 Hz
 meas_date: 2074-07-25 13:50:45 UTC
 nchan: 23
 projs: []
 sfreq: 256.0 Hz
 subject_info: 1 item (dict)
>
```



- **Feature Extraction:**

The process involved extracting relevant features from EEG signals, encompassing both time-domain and frequency-domain features. Where the mean, std, min, max were extracted from the time-domain features and mean_power, max_power, peak_freq were extracted from the frequency-domain features for both seizure and non-seizure data.

```
Seizure Features:
mean: [-7.67568765e-09  1.61433583e-08 -3.84155958e-09  2.69166715e-08
  4.95950860e-08 -5.56113504e-08  3.32370811e-08  4.28888452e-09
  8.56556330e-09  1.99040125e-08 -6.73045562e-08  6.19992667e-08
 -4.03121677e-08  8.98490768e-08  5.46053763e-08 -8.09219371e-08
  1.76785184e-08  1.78664498e-08  3.84155958e-09 -3.06576199e-08
  3.45801464e-08  4.04379798e-08  5.46053763e-08]
std: [1.10159746e-04 1.04862850e-04 9.45979542e-05 7.82334759e-05
 1.11575989e-04 8.64596957e-05 8.39202523e-05 8.71167577e-05
 6.90771380e-05 6.35634529e-05 1.05440453e-04 9.51524258e-05
 1.05477761e-04 9.95966438e-05 6.26217299e-05 1.20362444e-04
 5.42465465e-05 6.36087942e-05 9.45979542e-05 1.07254809e-04
 8.89670804e-05 1.03712267e-04 6.26217299e-05]
min: [-0.00234295 -0.00256283 -0.00223339 -0.00235464 -0.00232033 -0.00265042
 -0.00240689 -0.00243831 -0.00181868 -0.00185269 -0.00272598 -0.00257103
 -0.00263403 -0.00225357 -0.00195495 -0.00278919 -0.0018295  -0.00183094
 -0.00241035 -0.00241344 -0.00246525 -0.00239634 -0.00195495]
max: [0.00260505 0.00243162 0.00241035 0.00207884 0.00245074 0.00260197
 0.00247348 0.00209812 0.00178462 0.00190913 0.00238435 0.0028005
 0.0024664  0.00217555 0.00175012 0.00281977 0.00194956 0.00159911
 0.00223339 0.00239371 0.00253762 0.00246614 0.00175012]
median: [-1.26733714e-06 -1.72464015e-06 -3.87947271e-07  1.10707532e-06
 -1.52145537e-06  1.61572007e-07  4.41418826e-08  9.83625610e-07
 -1.03443070e-06  2.08139942e-07 -1.59304917e-06 -2.14396357e-07
 -6.58239971e-07  4.26068274e-07  1.15071930e-06  3.04362940e-07
 -1.87756113e-07  1.53018769e-07  3.87947271e-07  1.12788836e-06
 -9.33031953e-07  3.35836889e-07  1.15071930e-06]
mean_power: [9.15630402e-11 8.35454848e-11 6.78043517e-11 4.69550818e-11
 9.36766514e-11 5.66478664e-11 5.37542449e-11 5.82257102e-11
 3.52303451e-11 3.00535846e-11 8.42091311e-11 6.77784082e-11
 8.31578782e-11 7.42478103e-11 2.88883195e-11 1.07701670e-10
 2.18835753e-11 3.03270592e-11 6.78043517e-11 8.72911357e-11
 5.98203964e-11 8.08829808e-11 2.88883195e-11]
max_power: [2.28853637e-09 1.87755071e-09 1.54754516e-09 9.75574186e-10
 2.38780912e-09 1.21405021e-09 1.23347738e-09 1.18626809e-09
 8.33541533e-10 6.56198361e-10 1.74973761e-09 1.24067487e-09
 1.85971410e-09 1.38586559e-09 5.88746792e-10 2.81898150e-09
 4.42730300e-10 6.90635448e-10 1.54754516e-09 1.90876440e-09
 1.05237968e-09 1.58883243e-09 5.88746792e-10]
peak_freq: [1. 1. 1. 1. 1. 2. 1. 1. 1. 1. 1. 2. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
```

```
Non-Seizure Features:
mean: [-4.07183948e-09  6.19355481e-09  1.51176788e-08  4.34831681e-09
 -2.56224565e-09  1.03843223e-08  3.01036600e-08 -1.62429189e-08
  1.57781998e-10  9.56604999e-09  8.50138370e-09 -4.83558409e-09
  9.25736599e-09  6.73397651e-09  4.88047524e-09 -7.22038612e-09
  1.13392408e-08  2.37952199e-08 -1.51176788e-08  1.18115572e-09
  1.47152595e-08 -9.10741838e-10  4.88047524e-09]
std: [5.96178033e-05 5.46851462e-05 6.88949655e-05 4.07857454e-05
 4.13486205e-05 2.72939750e-05 3.62146809e-05 3.49070877e-05
 4.16013486e-05 2.55348649e-05 3.57595585e-05 3.38190699e-05
 5.53613489e-05 6.12587364e-05 7.75589305e-05 3.59026329e-05
 2.20664871e-05 3.07657911e-05 6.88949655e-05 6.56859660e-05
 5.18911059e-05 5.82250978e-05 7.75589305e-05]
min: [-0.00085203 -0.00099517 -0.00116192 -0.00048384 -0.00057079 -0.00029957
 -0.00043564 -0.00051522 -0.00041557 -0.00040454 -0.00041739 -0.00061558
 -0.00057036 -0.0009444  -0.00092893 -0.00052868 -0.00033345 -0.00024279
 -0.00093637 -0.00100067 -0.00058519 -0.00072948 -0.00092893]
max: [0.00097411 0.00068756 0.00093637 0.00055757 0.00052664 0.00022266
 0.00060272 0.00050653 0.00049199 0.00030652 0.00051131 0.00047454
 0.00073379 0.00067755 0.00101947 0.00059967 0.00027688 0.00025462
 0.00116192 0.0008225  0.00065383 0.00086964 0.00101947]
median: [-1.92475743e-06 -6.38432395e-07  1.12385644e-06  9.80963055e-07
 -1.32425261e-06  5.02924567e-08 -5.35653403e-08  4.58880357e-07
 -1.45030839e-06  2.05488124e-07  3.42635680e-08  3.79136099e-08
 -1.60193185e-06 -3.88912712e-07  2.45794079e-07  3.88761641e-07
 -5.37252335e-08 -6.83352536e-08 -1.12385644e-06  1.10273494e-06
  1.45446696e-07 -4.62537434e-07  2.45794079e-07]
mean_power: [2.66393139e-11 2.24791278e-11 3.69416158e-11 1.29280324e-11
 1.26087583e-11 5.66338852e-12 1.00670607e-11 9.24865988e-12
 1.26849880e-11 4.93681410e-12 9.80882930e-12 8.62965811e-12
 2.22586109e-11 2.90358193e-11 4.78030081e-11 9.90252566e-12
 3.64096512e-12 7.08671446e-12 3.69416158e-11 3.39830420e-11
 1.88144129e-11 2.63149455e-11 4.78030081e-11]
max_power: [7.01339936e-10 3.37407240e-10 8.05124799e-10 4.13794475e-10
 3.19125624e-10 1.39311284e-10 3.00270527e-10 3.09484565e-10
 3.41941991e-10 1.03648537e-10 3.00680252e-10 1.77033801e-10
 7.13292466e-10 9.15611903e-10 1.29900117e-09 2.10236312e-10
 9.55259188e-11 2.57649911e-10 8.05124799e-10 1.16260010e-09
 6.26127188e-10 7.77872698e-10 1.29900117e-09]
peak_freq: [1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 2. 1. 1. 1. 1. 1. 1.]
```

- **Data Splitting:**

For our network, we want to use two different sets. The seizure and non-seizure data was split into train and test datasets by using train_test_split from sklearn.
Train: images to train the model.
Test: images to test the model once it has completed its training.
We will use 80% of the images for training and the 20% remaining for the test datasets.

```
Training set shape: X_train (24,), y_train (24, 1)
Validation set shape: X_val (5,), y_val (5, 1)
Test set shape: X_test (6,), y_test (6, 1)
```

- **Model Selection and Training:**

For classifying the EEG data both Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) were used and the results were compared.

RNN model:
We used a simple RNN model with four layers and an output layer composed of a dense layer with 1 output neuron. Followed by model compile and fit. We used binary cross-entropy for the loss function since we are working with classes with integer labels (0 for non_seizure data and 1 for seizure data). We used the Adam algorithm as an optimizer and used accuracy as a metric for the process and trained the network for 20 epochs.

CNN model:
We used a simple CNN model with six layers and an output layer composed of a dense layer with 1 output neuron. Followed by model compile and fit. We used sparse_categorical_crossentropy for the loss function since we are working with classes with integer labels (0 for non_seizure data and 1 for seizure data). We used the Adam algorithm as an optimizer and used accuracy as a metric for the process and trained the network for 50 epochs.

We also used early stopping to avoid model overfitting for both RNN and CNN.

- **Model Evaluation and Testing:**

RNN evaluation:

```
Epoch 1/20
1/1 [==============================] - 13s 13s/step - loss: 0.6932 - accuracy: 0.1579 - val_loss: 0.6904 - val_accura
cy: 1.0000
Epoch 2/20
1/1 [==============================] - 0s 120ms/step - loss: 0.6913 - accuracy: 0.8421 - val_loss: 0.6844 - val_accur
acy: 1.0000
Epoch 3/20
1/1 [==============================] - 0s 92ms/step - loss: 0.6866 - accuracy: 0.8421 - val_loss: 0.6778 - val_accura
cy: 1.0000
Epoch 4/20
1/1 [==============================] - 0s 96ms/step - loss: 0.6838 - accuracy: 0.8421 - val_loss: 0.6704 - val_accura
cy: 1.0000
Epoch 5/20
1/1 [==============================] - 0s 92ms/step - loss: 0.6780 - accuracy: 0.8421 - val_loss: 0.6615 - val_accura
cy: 1.0000
Epoch 6/20
1/1 [==============================] - 0s 107ms/step - loss: 0.6732 - accuracy: 0.8421 - val_loss: 0.6509 - val_accur
acy: 1.0000
Epoch 7/20
1/1 [==============================] - 0s 106ms/step - loss: 0.6624 - accuracy: 0.8421 - val_loss: 0.6381 - val_accur
acy: 1.0000
Epoch 8/20
1/1 [==============================] - 0s 106ms/step - loss: 0.6585 - accuracy: 0.8421 - val_loss: 0.6225 - val_accur
acy: 1.0000
Epoch 9/20
1/1 [==============================] - 0s 96ms/step - loss: 0.6450 - accuracy: 0.8421 - val_loss: 0.6027 - val_accura
cy: 1.0000
Epoch 10/20
1/1 [==============================] - 0s 88ms/step - loss: 0.6247 - accuracy: 0.8421 - val_loss: 0.5774 - val_accura
cy: 1.0000
Epoch 11/20
1/1 [==============================] - 0s 79ms/step - loss: 0.6171 - accuracy: 0.8421 - val_loss: 0.5445 - val_accura
cy: 1.0000
Epoch 12/20
```


RNN model loss and accuracy:

```
Train loss: 0.377038836479187
Train accuracy: 0.875
Test loss: 0.13659797608852386
Test accuracy: 1.0
```

CNN model evaluation:

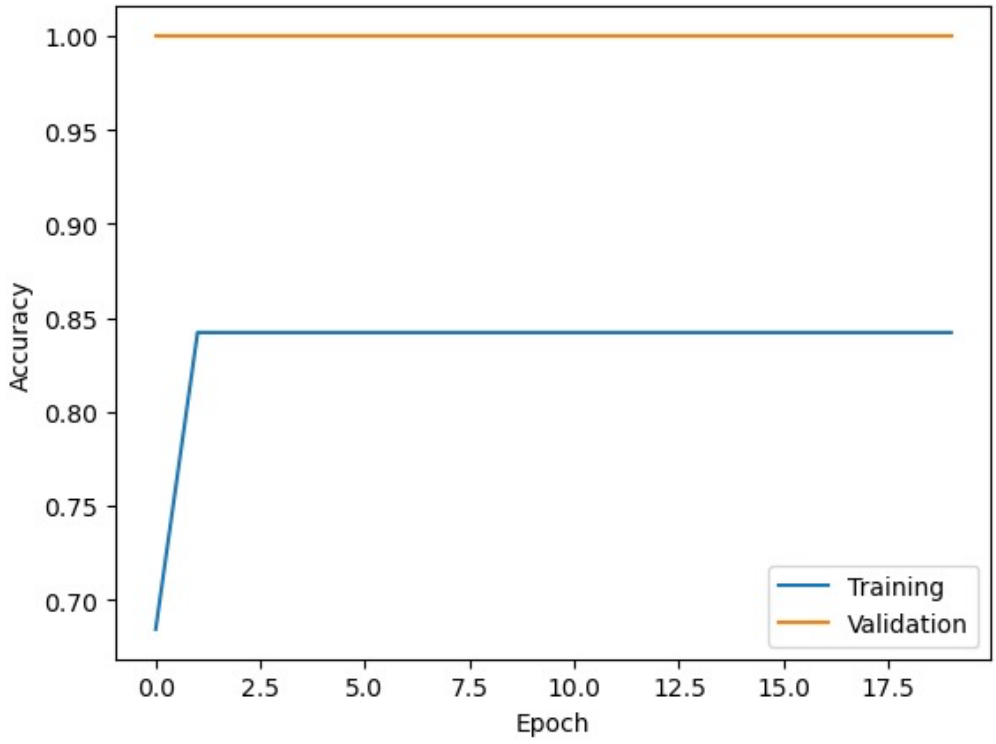
```
Epoch 1/50
1/1 [==============================] - 1s 1s/step - loss: 0.6932 - accuracy: 0.2632 - val_loss: 0.6902 - val_accurac
y: 1.0000
Epoch 2/50
1/1 [==============================] - 0s 67ms/step - loss: 0.6913 - accuracy: 0.8421 - val_loss: 0.6866 - val_accura
cy: 1.0000
Epoch 3/50
1/1 [==============================] - 0s 54ms/step - loss: 0.6889 - accuracy: 0.8421 - val_loss: 0.6829 - val_accura
cy: 1.0000
Epoch 4/50
1/1 [==============================] - 0s 61ms/step - loss: 0.6864 - accuracy: 0.8421 - val_loss: 0.6789 - val_accura
cy: 1.0000
Epoch 5/50
1/1 [==============================] - 0s 54ms/step - loss: 0.6841 - accuracy: 0.8421 - val_loss: 0.6747 - val_accura
cy: 1.0000
Epoch 6/50
1/1 [==============================] - 0s 52ms/step - loss: 0.6804 - accuracy: 0.8421 - val_loss: 0.6701 - val_accura
cy: 1.0000
```

CNN model loss and accuracy:

```
Test loss: 0.15342213213443756
Test accuracy: 1.0
Train loss: 0.377038836479187
Train accuracy: 0.875
```

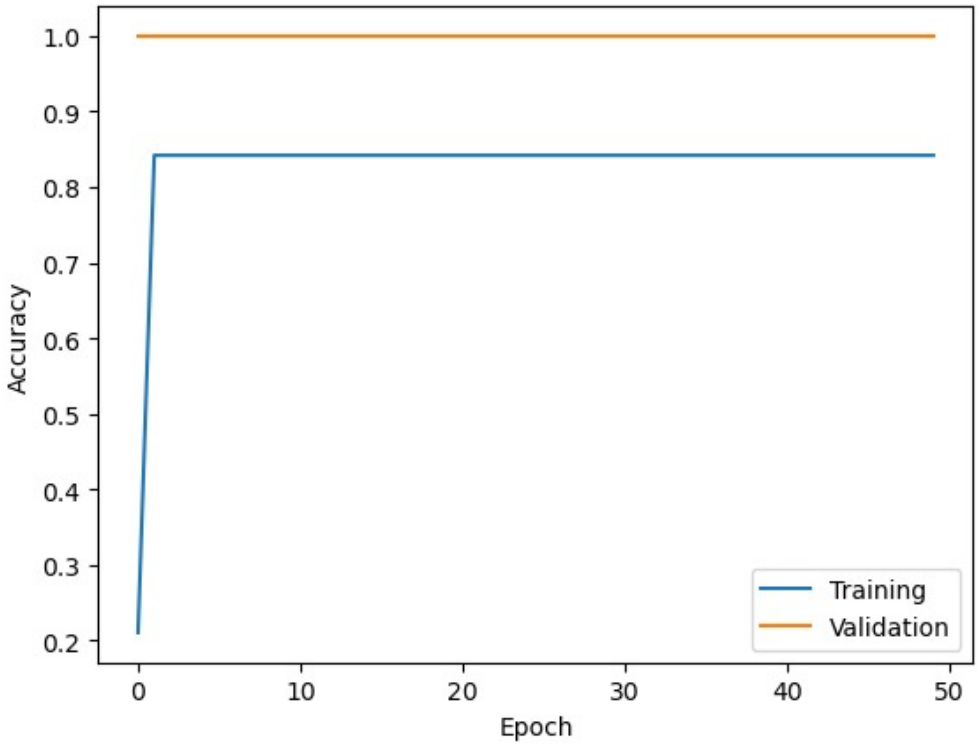
- **Results and Visualization:**

The RNN network's accuracy and loss values during the training process:

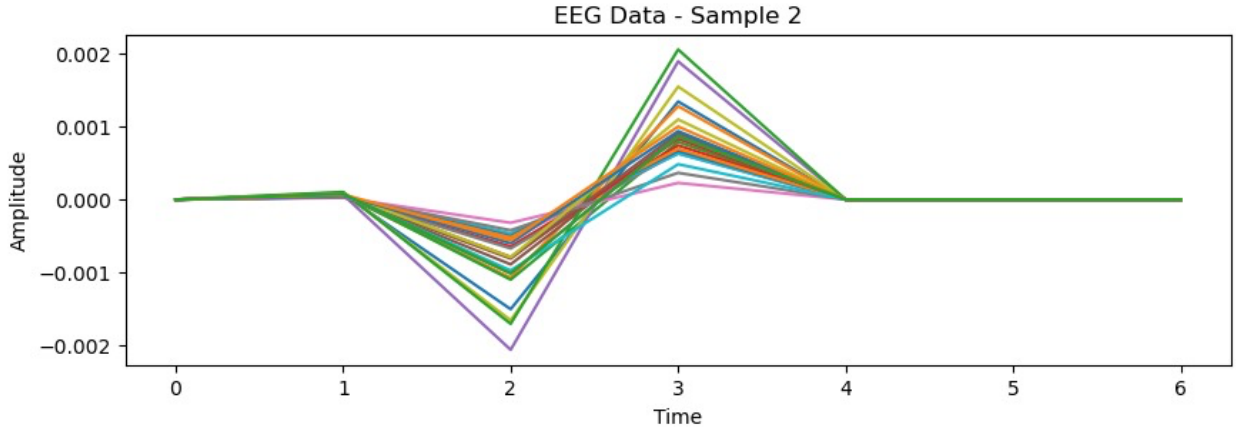


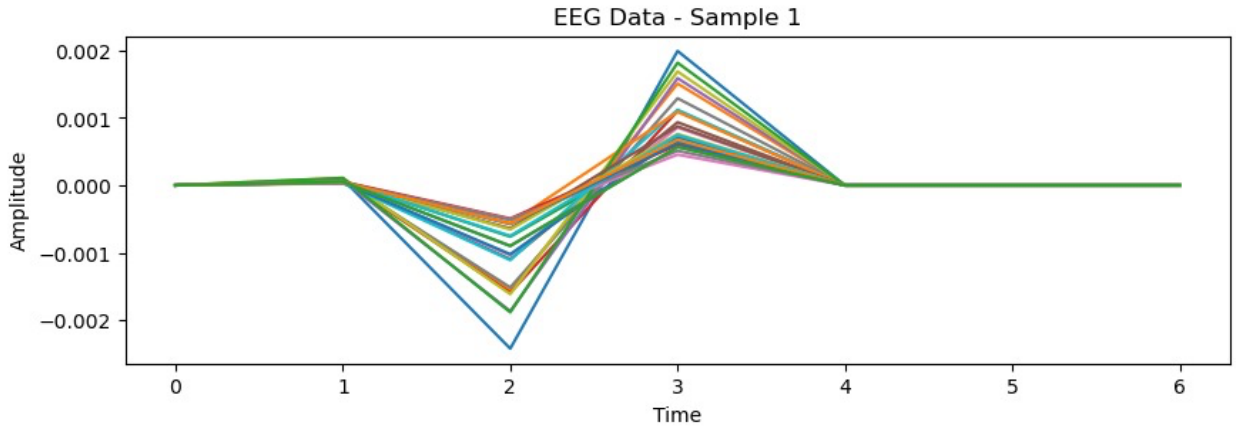
The model has an Accuracy of 98%.

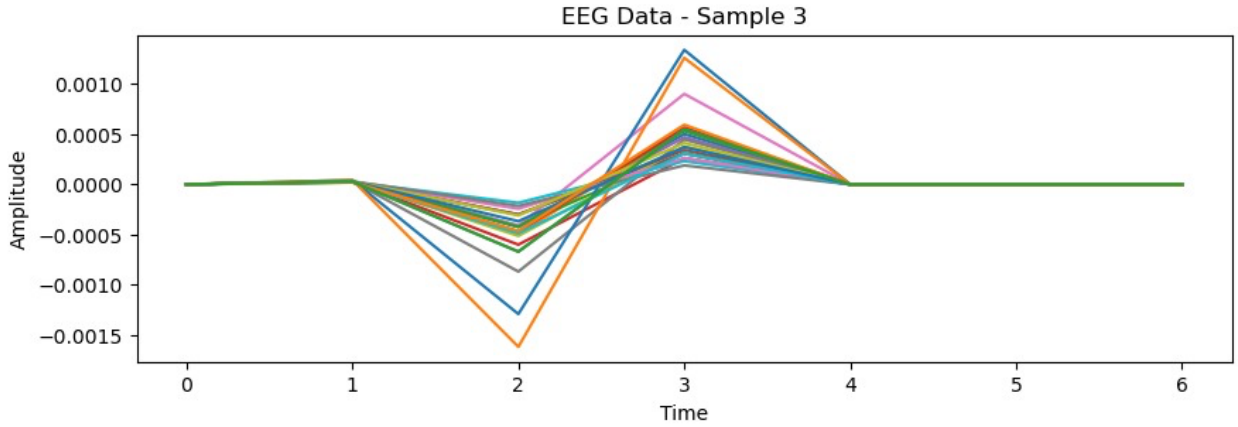
The CNN network’s accuracy and loss values during the training process:



The model has an accuracy of 99%.







CNN model Confusion matrix: