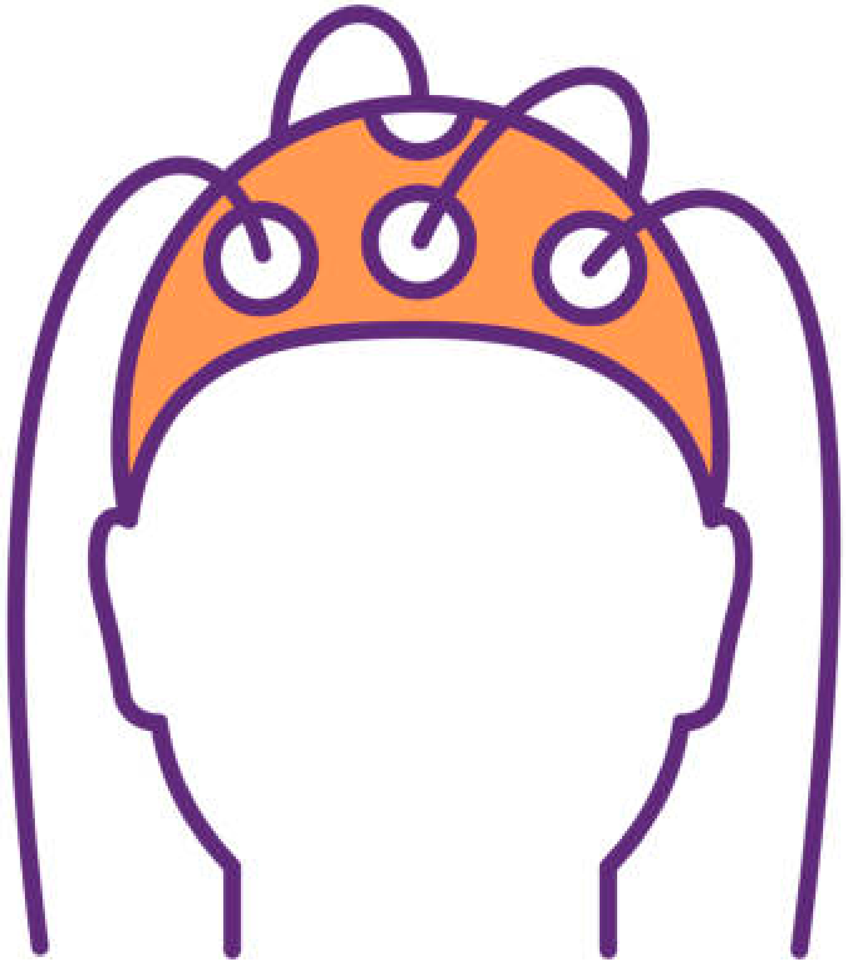


TEAM 3 - EEG DEVICE

Predicting Worker Fatigue using Machine Learning

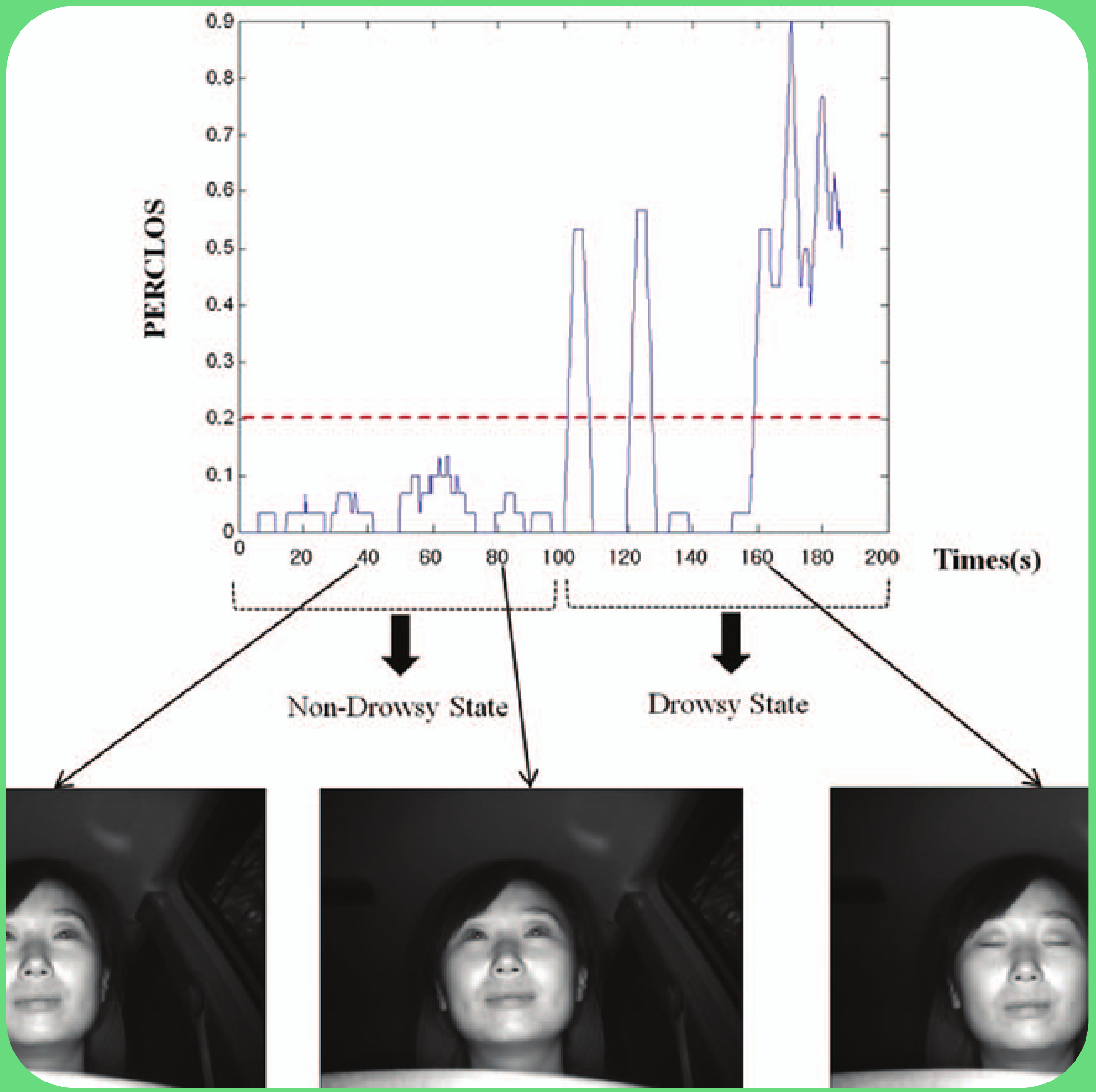
SWAPNIL , AKSHAT
HUSNA , TOMA , MOHAMED

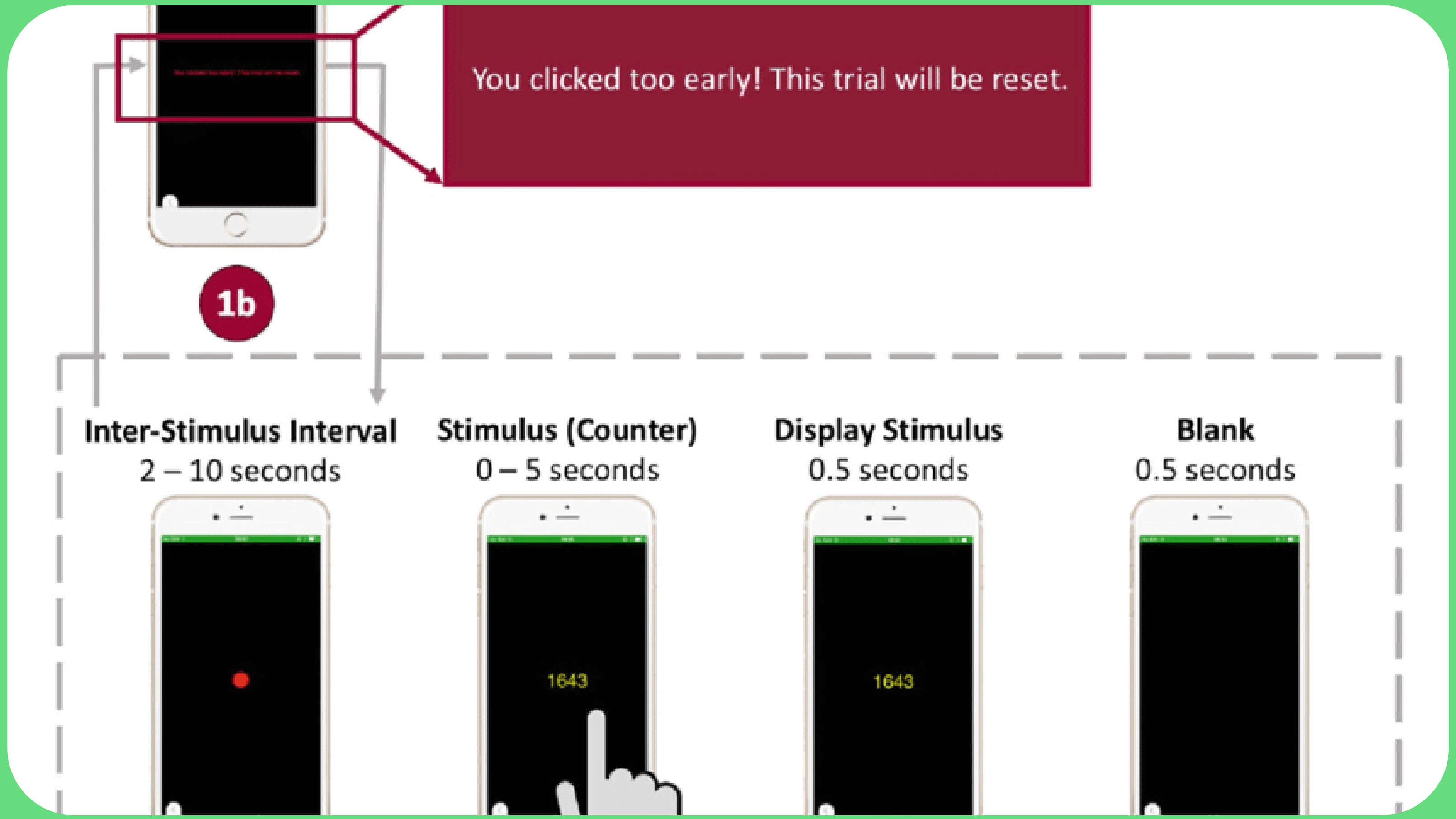
Outline

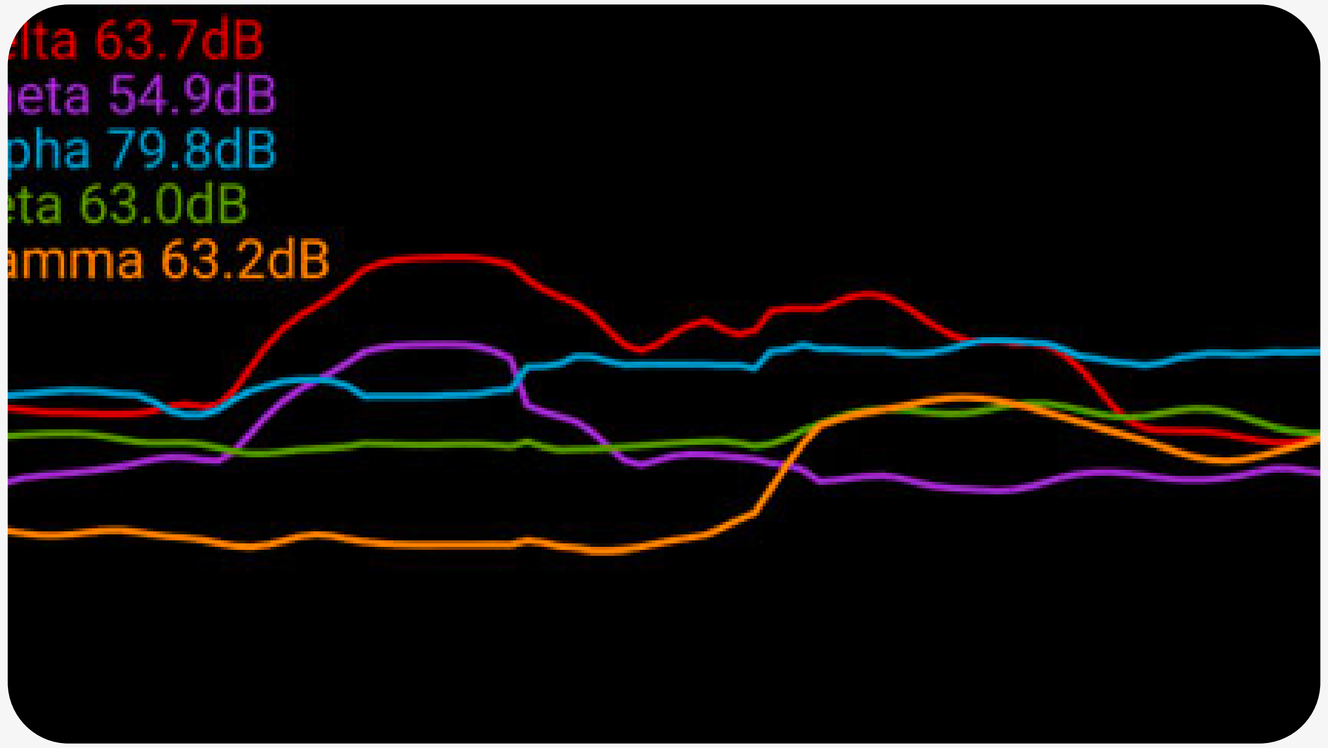
- Problem
- Solution
- Data Collection
- Data Pre-Processing
- ML Algorithms
- Challenges
- Future Work

Problem

- Mental fatigue deteriorates the ability to perform daily activities
- Made even more severe when you consider high risk and dangerous jobs
- Current solutions in the market are either too expensive, time consuming or just not practical







Solution

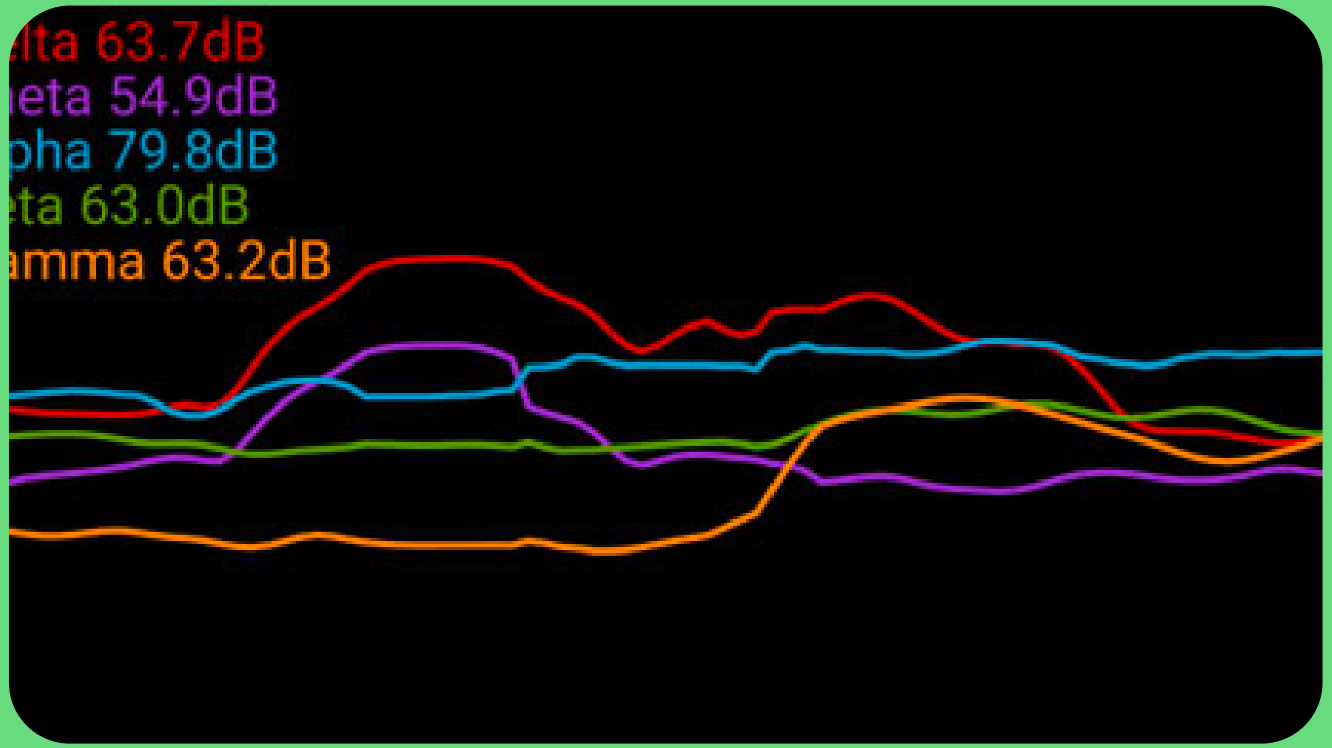
- Machine Learning model to predict mental state using EEG activity
- Trained on both "Normal" and "Fatigued" brain states
- Numerous real-world solutions such as construction workers, pilots, truck drivers
- Commercial applications if companies want to measure the mental health of their employees

Data Collection

- Muse 2 headband and mind monitor app
- Classified fatigue state as right after physical exercise
- Normal state classified as 1-2 hours after waking up.
- 2 people, 4 datasets each (2 fatigue, 2 normal).

Data Collection

- Automatically converted to csv format
- 4128 rows of data total
- 38 features
- Missing data due to blinking and jaw clenching

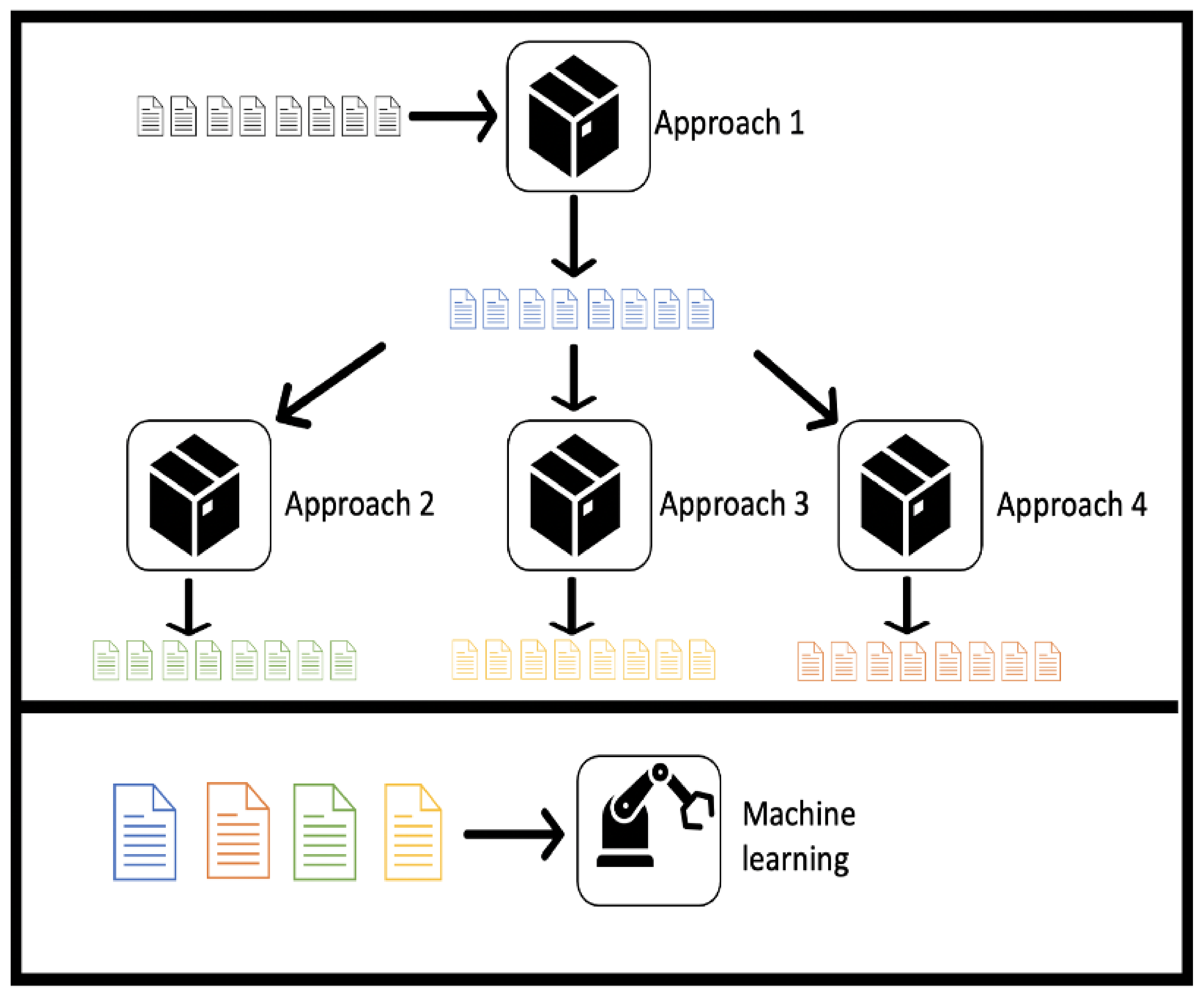


_AF8	Delta_TP10	Theta_TP9	Theta_AF7	Theta_AF8	Theta_TP10	Alpha_TI
.3207145	0.42881608	0.60196716	-0.18698062	-0.13936596	0.35633218	0.82523
27662006	0.3829187	0.6451441	-0.09020477	-0.16846849	0.45644432	0.9690
15880851	0.099144466	0.51457024	-0.18154833	-0.20774347	0.24197535	0.9885
04452983	0.6459588	0.39067212	-0.34338745	0.082143106	0.3756607	0.80666
53867586	0.72533315	0.4206783	-0.4175956	-0.07507378	0.34412867	0.53560
17703442	0.010968491	0.45171264	-0.35931224	-0.1990452	0.16505104	0.8054
.3394353	0.20413157	0.33588305	0.5497157	0.26697856	-0.06340826	0.7988
35774288	0.46079406	0.27564555	0.6178541	0.053712662	0.5290685	0.30752
12645137	0.3178599	0.38405982	0.18402104	7.42e-04	0.39803013	0.39017
7349838	0.03942682	0.40294698	0.4872455	0.16785648	0.2067533	0.592

Data Pre-Processing

4 Different Approaches

4 Large Datasets



Approach 1

Outlier Detection and Removal and Imputation of Missing Values

Three distributions of Gaussian mixture model

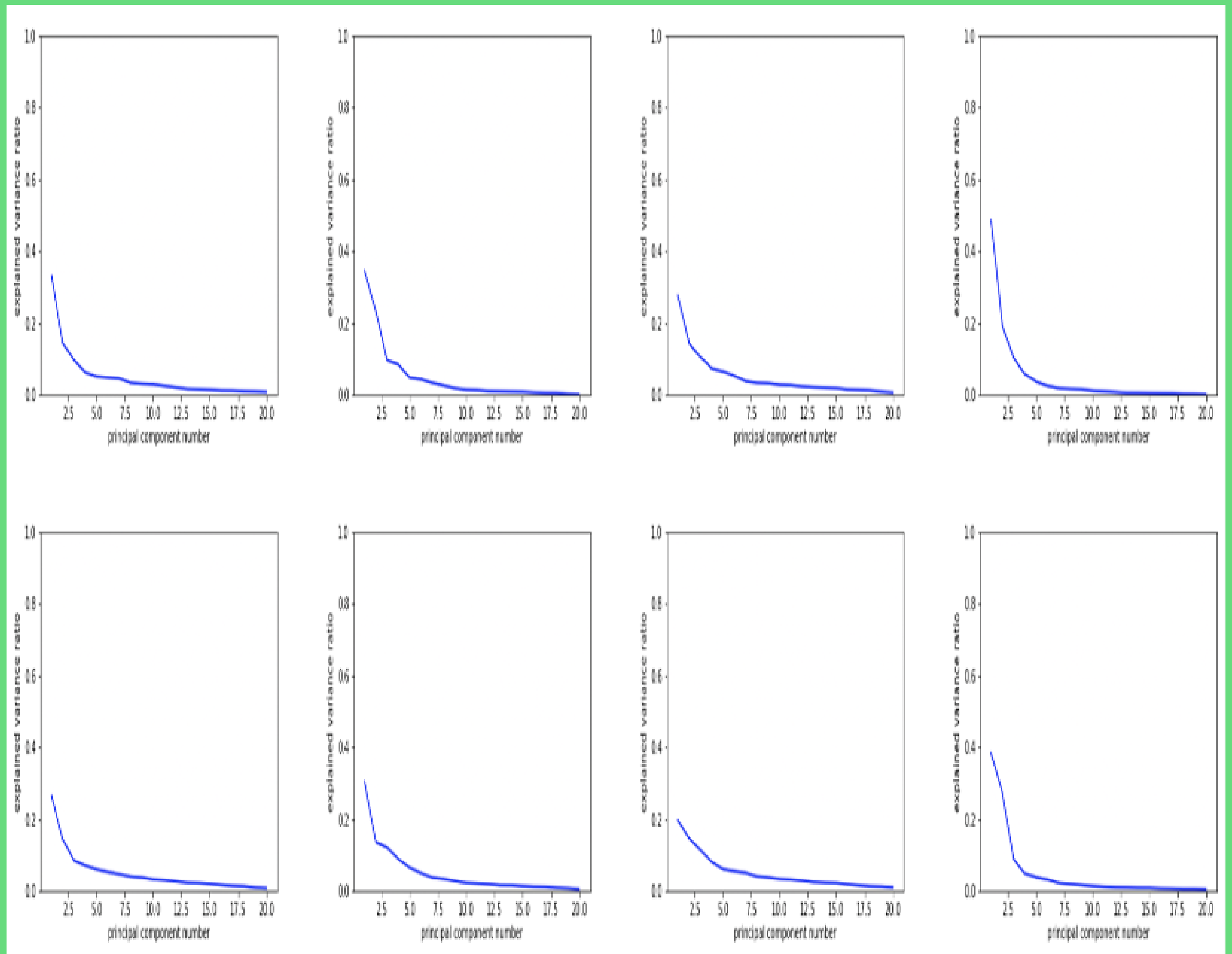
Linear interpolation-based imputer

Threshold< 0.0005.

20 features in total.

Approach 2

Principal Component Analysis (PCA)



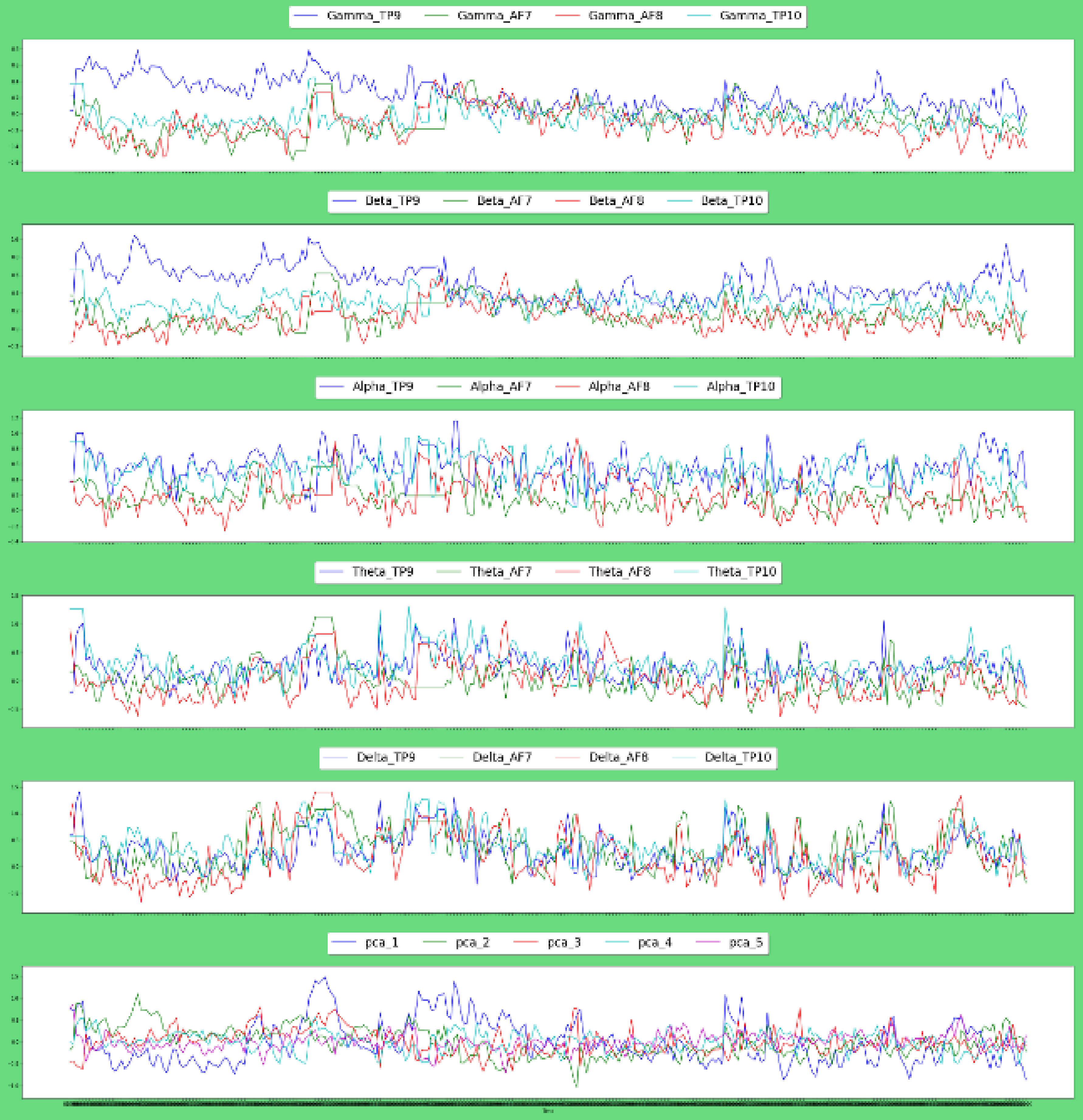
PCA as a feature engineering tool

25 features in total

Elbow method

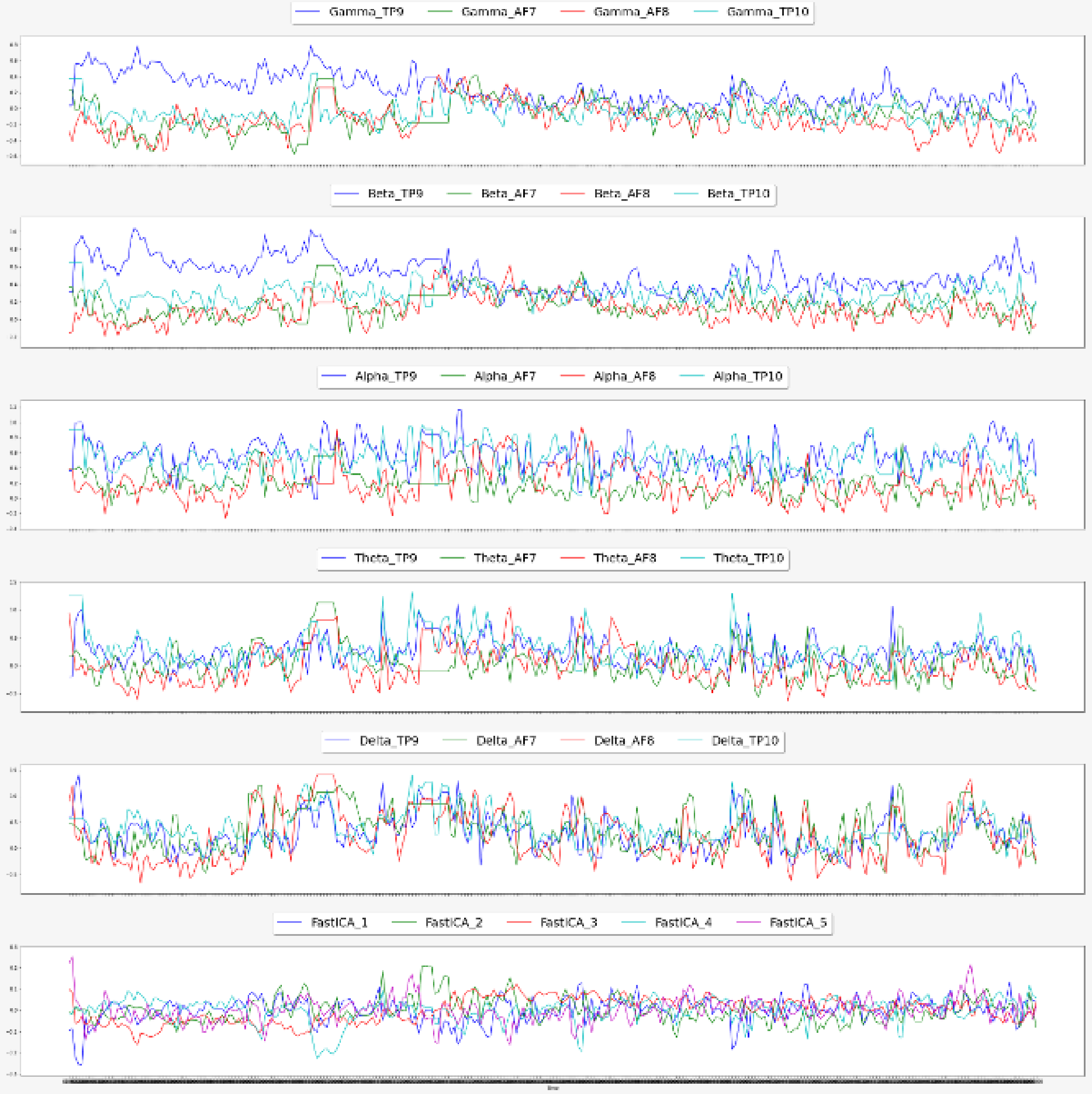
Approach 2 (Continued)

Principal Component Analysis (PCA)



Approach 3

Fast Independent Component Analysis (Fast ICA)



Fast ICA as a feature engineering tool

25 features in total

Approach 4

Independent Component Analysis (ICA) and Time and Frequency Domain Feature Extraction

Feature	Domain	Description
Mean	Time	Mean value among the datapoint values in a dataset
median	Time	Median value among the datapoint values in a dataset
Minimum	Time	Minimum data point value in a window
Maximum	Time	Maximum data point value in a window
Standard deviation	Time	Standard deviation based on the datapoints in a window
Slope	Time	Line slope when fitting a linear regression to the data points in a window
Maximum frequency	Frequency	The frequency with the highest amplitude. This metric provides an indication of the most important frequency in the windows under consideration.
Frequency weighted signal average	Frequency	This metric provides information on the average frequency observed in the window (given the amplitudes) and might shed a light on the entire spectrum of frequencies. It is calculated by multiplying the frequencies with their amplitude and normalizing them by the total sum of the amplitudes.
Power spectral entropy	Frequency	This metric represents how much information is contained within the signal. In other words, the power spectral entropy determines whether there are one or a few discrete frequencies standing out of all others. To calculate the metric, the power spectral density is first calculated (squaring the amplitude and normalizing by the number of frequencies), normalize the values to a total sum of 1 such that we can view it as a probability density function and compute the entropy via the standard entropy calculation.

Approach 4 (Continued)

Independent Component Analysis (ICA) and Time and Frequency Domain Feature Extraction

The process that was used to obtain the time and frequency domain features is summarized as follows:

- 2 second-sized windows
- 50% window overlap
- Several time and frequency domain metrics

5 Fast ICA feature components

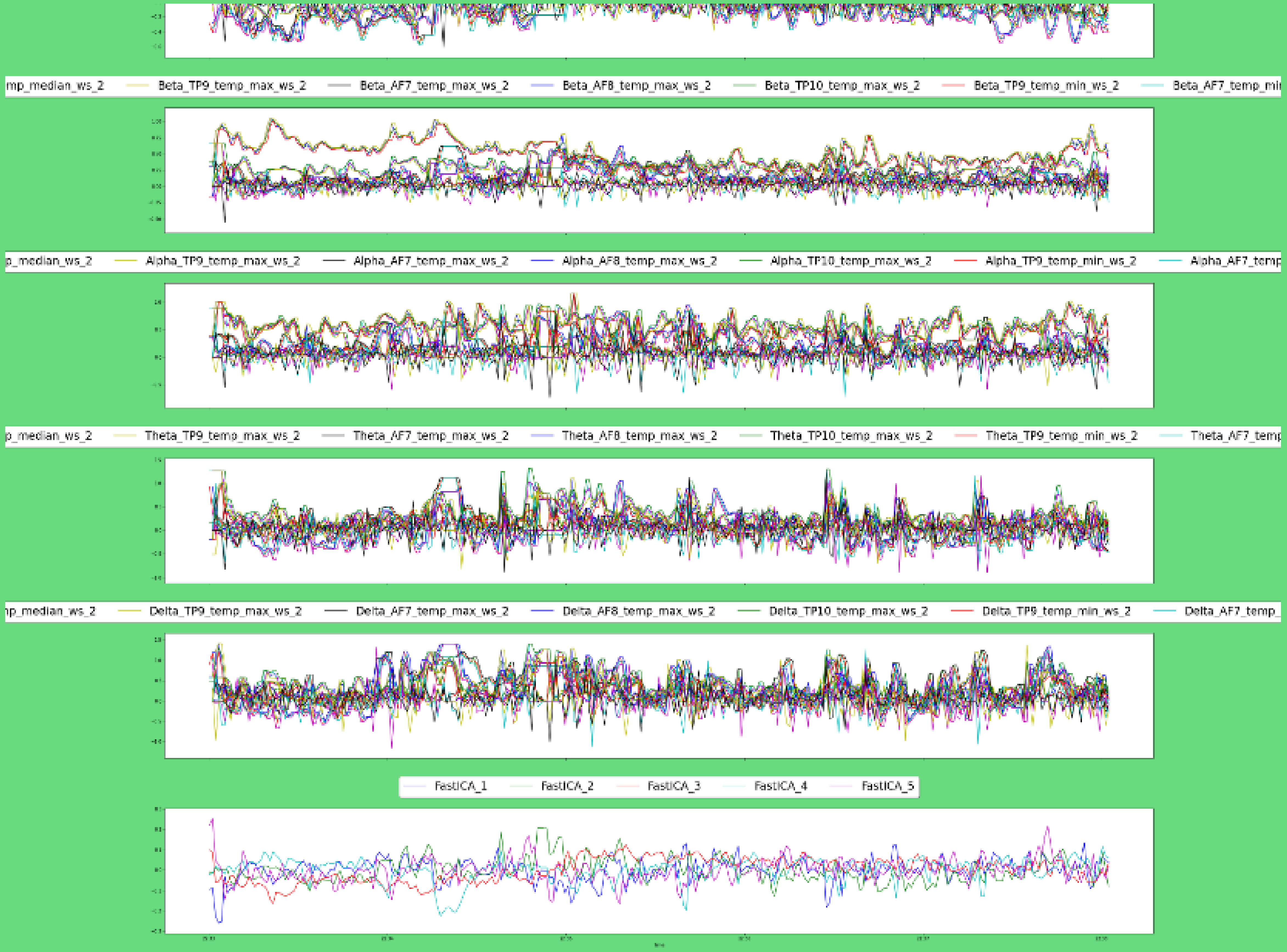
120 temporal-based features

100 frequency-based features

245 features in total

Approach 4 (Continued)

Independent Component Analysis (ICA) and Time and Frequency Domain Feature Extraction



Prediction Algorithm: AutoML

ML models used:

- Baseline
- Linear
- Decision Tree
- Random Forest
- XGBoost
- Neural Network

AutoML Evaluations

Standard

Best Model: XGBoost

Accuracy: 94.76% in 4.12s

PCA

Best Model: Neural Network

Accuracy: 99.12% in 1.16s

FastICA

Best Model: XGBoost

Accuracy: 96.07% in 4.43s

Fast ICA and Temporal Frequency Extraction

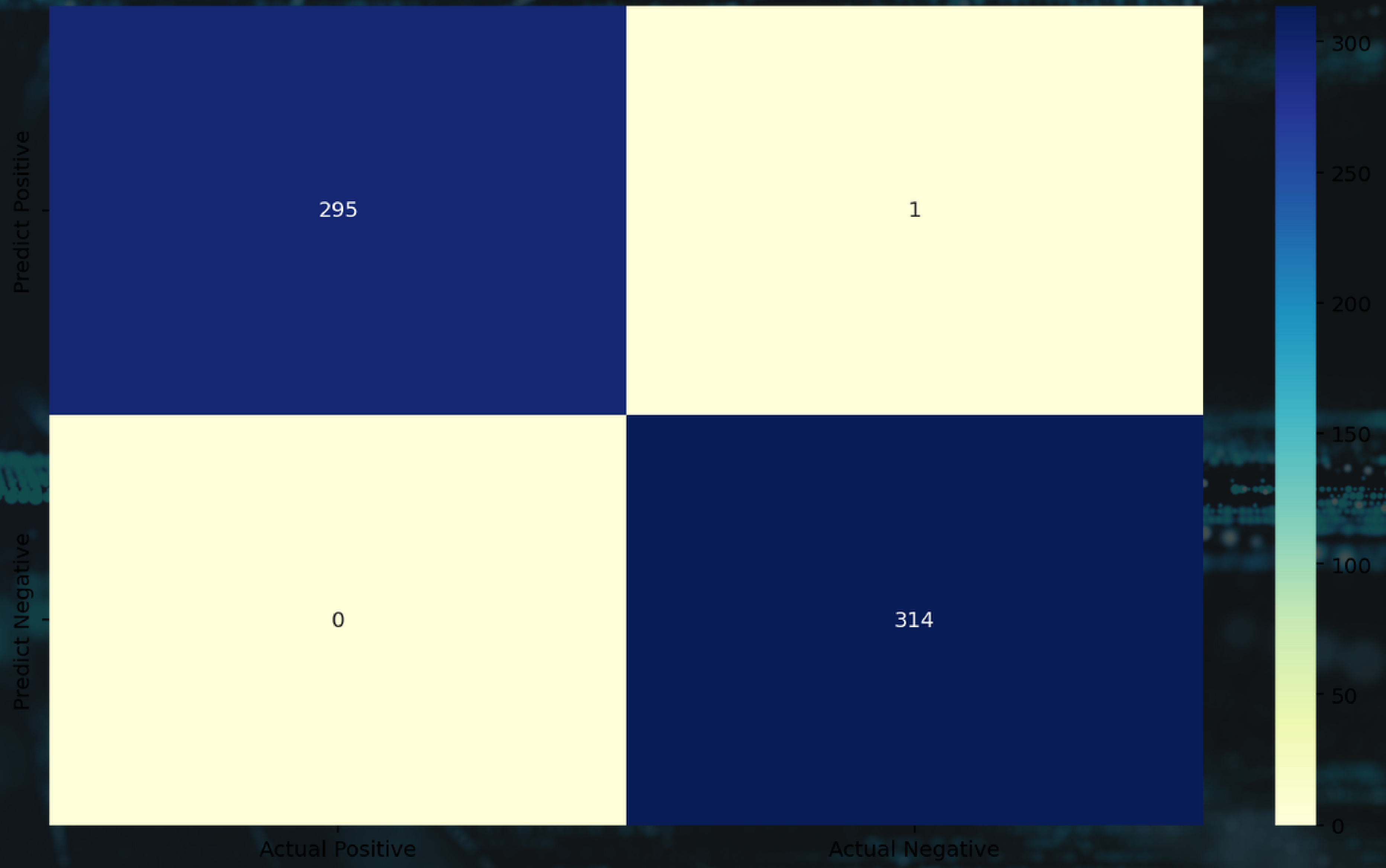
Best Model: XGBoost

Accuracy: 95.29% in 27.76s

ANN Classifier

Trained with PCA Dataset

99.84% Accuracy



Density of input layer 1: 25 - Rectified Linear Unit (ReLU)

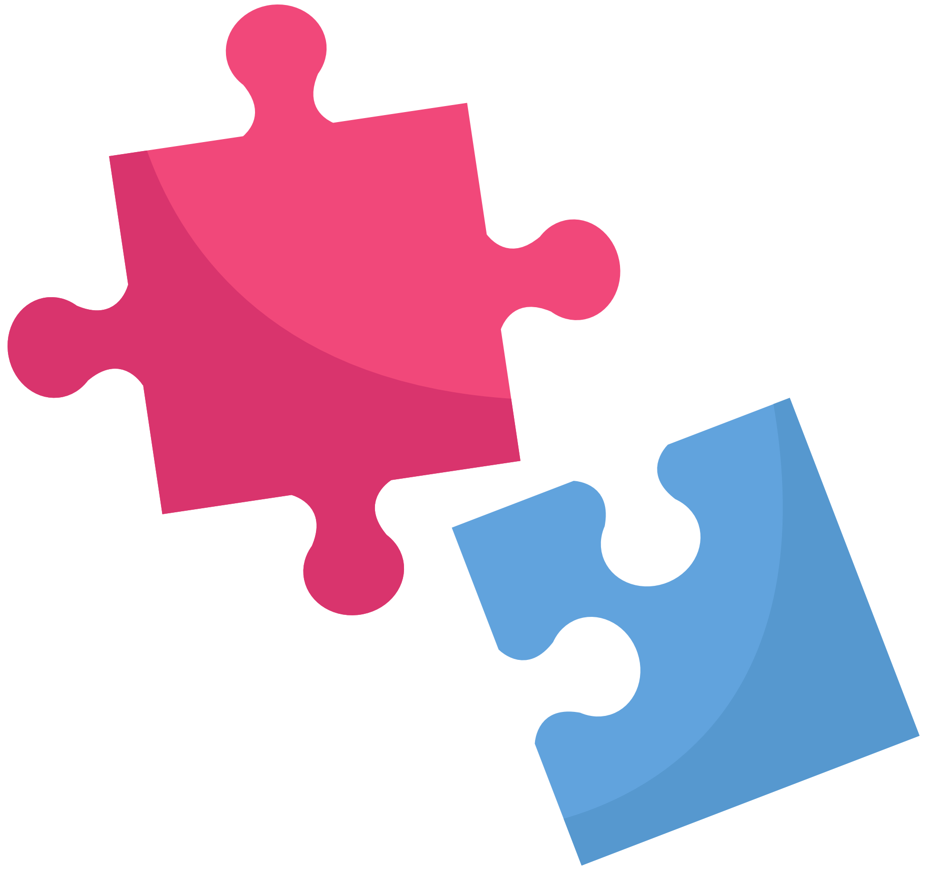
Density of hidden layer 2: 16 - Rectified Linear Unit (ReLU)

Density of output layer 3: 1 - Sigmoid

Compiled with: Binary Crossentropy

Learning rate: 0.05

Challenges



1 Data fetching issues with involuntary activities like blinking of eyes etc. The device also records when you move, fidget, or open your eyes.

2 It is difficult to differentiate between fatigue and normal data :)

3 Alignment issues of Muse device(headband) leading to null values.

4 The brain sensing is not 100 percent error free.

Next Steps

Collect more data

Real Time Predictions

Diverse Conditions

Conclusion

We are open to any questions