Brain Activity Classification Using Machine Learning

Name: Muhammad Junaid Reg. No. : FA21-EEE-061
Name: Umer Ali Reg. No. : FA21-BCE-008
Name: Ahmed Raheel Reg. No. : FA21-BCE-015

Supervisor: Dr. Shoaib Azmat
Co-Supervisor: Engr. Nauman Khan Tareen

Electrical Domain: Electronics & Computer

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Outline

- Introduction
- Motivation and Need
- Problem Analysis
- Proposed Solution
- Deliverables
- Design and Development
- Results
- Hardware Deployment
- Impact of the Project
- Project Management
- Conclusion
- References

Introduction

- Brain Activity Classification
  - Understanding cognitive functions via EEG signals.
- EEG (electroencephalography)
  - Measure brain's electrical activity using scalp sensors.
- Machine Learning
  - Automates EEG data analysis and classification.
- Project Focus
  - Classify eye (open/close) and mental states (arithmetic/relaxed).

Motivation and Need

- Health and Neurotherapy
  - Mental Health Monitoring
  - Early Detection of Cognitive Decline
- Assistive Technologies
  - Brain-Computer Interface
  - Wearable Attention Tracking Technology
- Neurological Research

Problem Analysis.

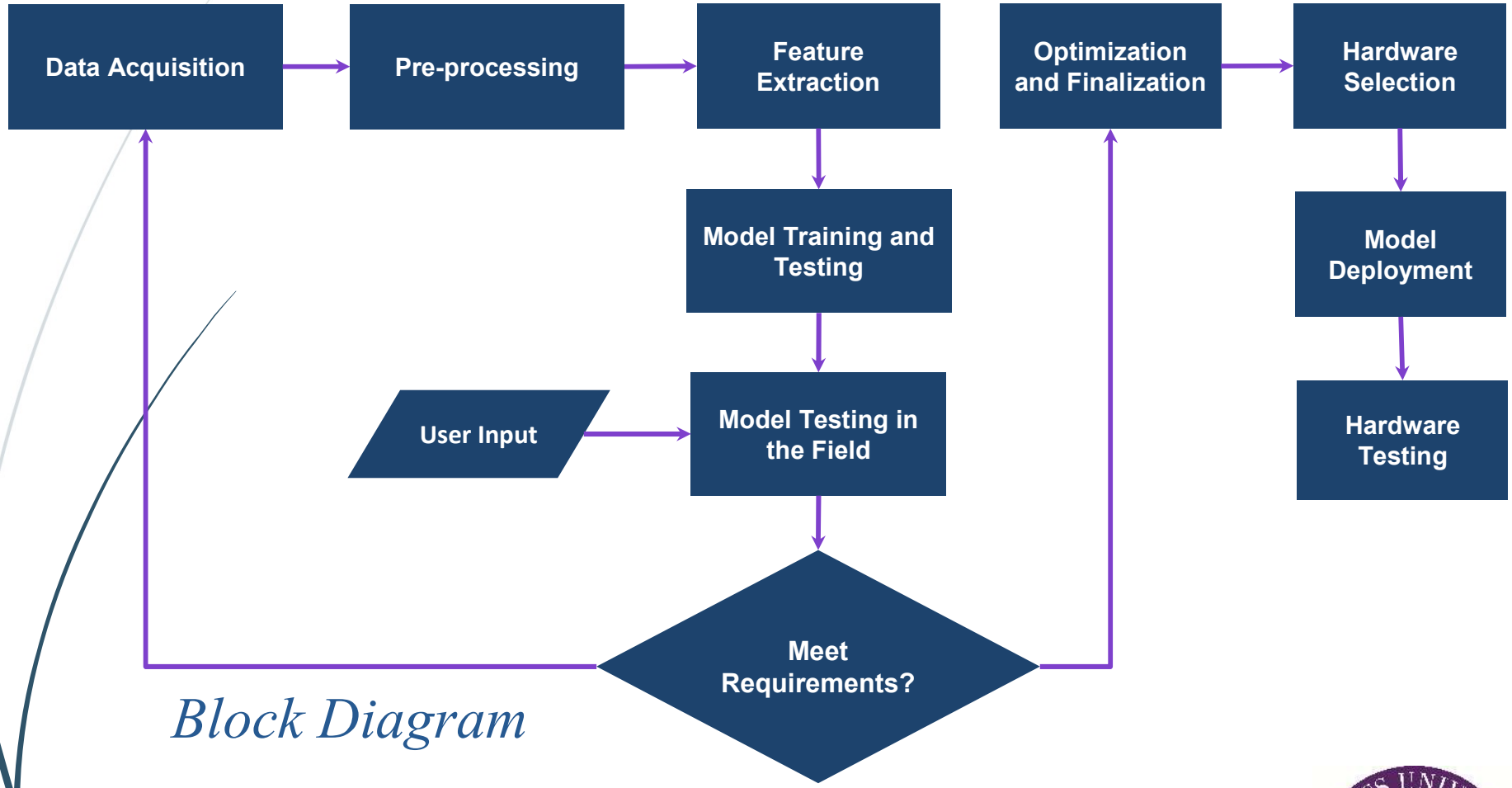
Literature Review

	Title	Model	Accuracy	No. of Electrodes/ Channels
1.	EEG-Based Eye Movement Recognition Using Brain–Computer Interface and Random Forests **[1]**	Random Forest	85.29%	32 channels
2.	Statistical Feature Analysis for EEG Baseline Classification : Eyes Open vs Eyes Closed **[2]**	SVM KNN	73.41% 72.07%	64 channels
3.	Assessment of Model Accuracy in Eyes Open and Closed EEG Data **[3]**	SVM ANN Random Forest KNN	96.9% 97.1% 97.0% 96.1%	24 channels

Problem Analysis. (Continued.)

	Title	Model	Accuracy	No. of Electrodes/ Channels
4.	EEG Based Mental Arithmetic Task Classification Using a Stacked Long Short Term Memory Network for Brain-Computer Interfacing **[4]**	LSTM	91.67%	22 channels
5.	EEG Signal Classification for Mental Stress During Arithmetic Task Using Wavelet Transformation and Statistical Features **[5]**	KNN EBT QSVM	77.3% 68.2% 59.1%	23 channels
6.	EEG mental arithmetic task levels classification using machine learning and deep learning algorithms **[6]**	KNN SVM Desicion Tree ANN LSTM	91% 89% 65% 96.80% 98%	23 channels

Proposed Solution



Block Diagram

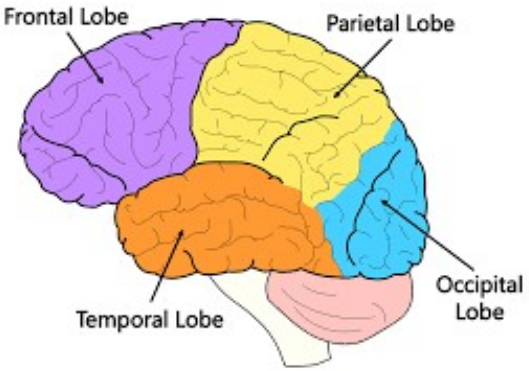
Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Deliverables

- EEG Dataset
  - 10 Subjects
  - 4 Mental Activities
- Trained Machine Learning Models
  - Pre-processing
  - Feature Extraction
  - Model Training and Testing
- Hardware Prototype
  - Model Selection
  - Selected Model Deployed on Edge Device

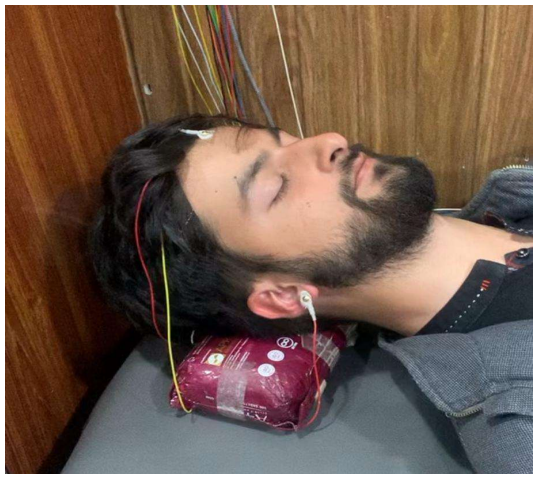
Design Development (Data Acquisition)

- Contec KT88-2400
- 10 Subjects
- 12 Channels (Frontal, Occipital, Temporal)
- 4 Activities
  - Relaxed with Eyes Opened
  - Relaxed with Eyes Closed
  - Mental Arithmetic with Eyes Opened
  - Mental Arithmetic with Eyes Closed
- EEG Recording Time
  - 5 Minutes per Activity



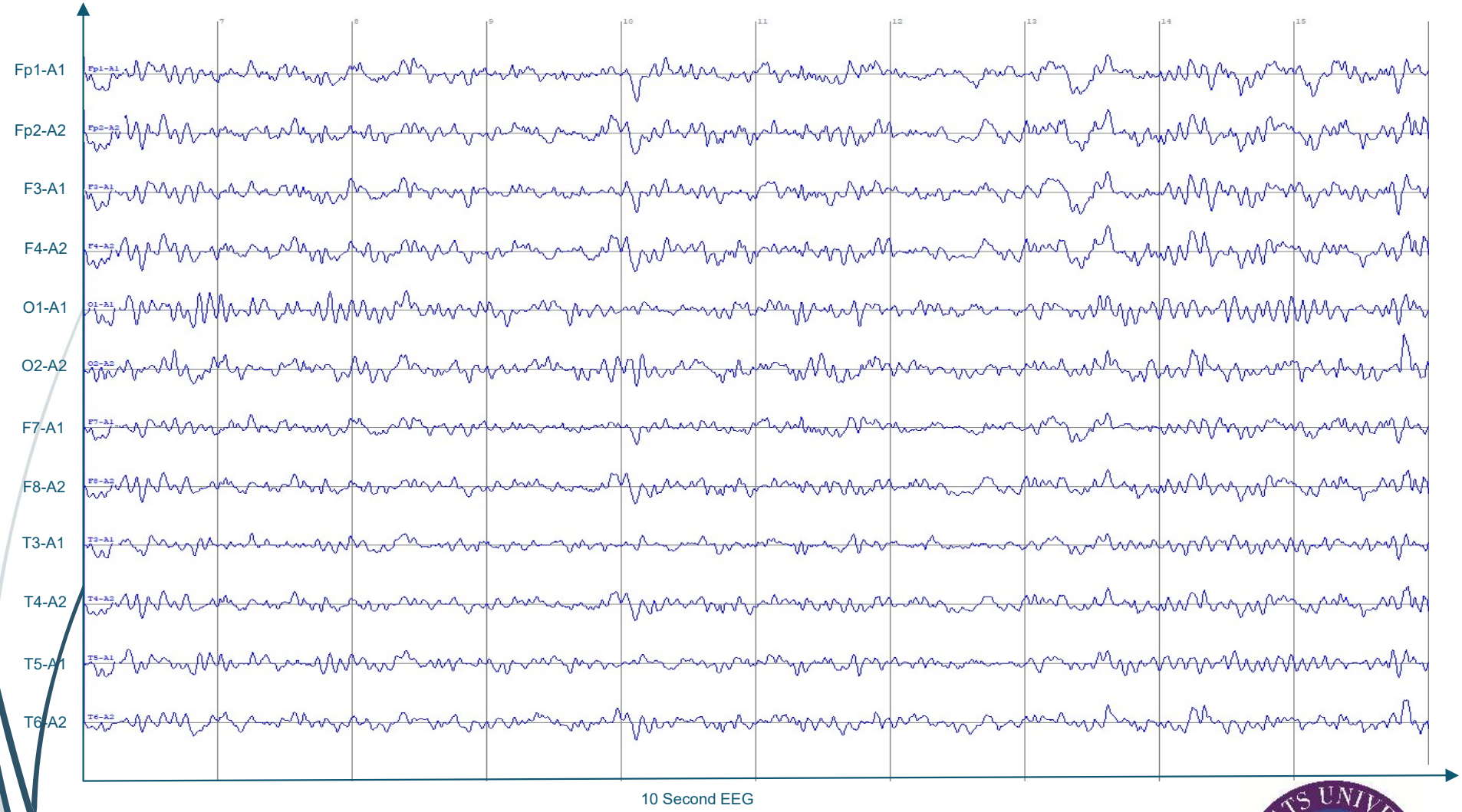
Design Development (Data Acquisition)




Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Design Development (Data Waveforms)



Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Design Development (Data in .csv Format)

Time	Fp1	Fp2	F3	F4	O1	O2	F7	F8	T3	T4	T5	T6
0	-3.25322	0.550546	-0.25025	1.25124	1.451438	7.457389	-2.55253	-2.25223	-0.15015	2.152132	-1.35134	4.254215
0.005	-4.15412	0.650645	-1.65164	1.25124	0.950942	6.256199	-2.85283	-2.25223	-0.55055	0.950942	-1.45144	2.752728
0.01	-4.75471	0.250248	-2.25223	0.350347	0.550546	3.853819	-3.15312	-2.05203	-1.15114	-0.15015	-1.65164	0.350347
0.015	-5.25521	0.05005	-2.65263	-0.25025	-0.25025	0.850843	-3.25322	-2.05203	-1.75174	-1.95193	-2.05203	-1.75174
0.02	-5.35531	-0.25025	-2.25223	-0.45045	-0.75074	-0.75074	-3.15312	-1.65164	-2.05203	-2.65263	-2.05203	-3.15312
0.025	-5.25521	0.05005	-1.45144	0.05005	-0.75074	-1.65164	-2.65263	-1.35134	-2.05203	-2.55253	-2.05203	-3.55352
0.03	-4.45441	0.350347	0.05005	2.052033	0.05005	-1.45144	-1.95193	-0.85084	-1.75174	-1.45144	-1.65164	-2.85283
0.035	-3.55352	1.25124	0.650645	4.154116	0.850843	-1.65164	-1.35134	-0.55055	-1.35134	-0.25025	-1.45144	-1.95193
0.04	-2.85283	2.052033	1.751736	5.755703	0.950942	-2.25223	-1.15114	-0.45045	-1.05104	0.650645	-1.45144	-1.35134
0.045	-2.55253	2.152132	1.551537	6.856794	0.05005	-3.25322	-1.35134	-0.45045	-1.05104	1.551537	-1.75174	-1.05104
0.05	-2.65263	1.851835	0.950942	7.157092	-1.05104	-3.75372	-1.65164	-0.45045	-1.05104	2.352331	-1.95193	-0.55055
0.055	-3.15312	0.850843	0.05005	6.956893	-1.65164	-3.45342	-2.05203	-0.75074	-1.05104	2.752728	-1.75174	0.05005
0.06	-3.25322	0.05005	-0.75074	6.356298	-0.25025	-1.75174	-2.05203	-0.75074	-0.75074	3.353323	-0.85084	0.650645

Design Development (Pre-processing)

Pre-processing

- ICA
  - To remove EMG and EOG artifacts from data.
- Data Segmentation
  - 5 second window
  - 80% overlap (4 second)
  - Dataset: 11,840 samples
    - 80% for training (9,472 samples)
    - 20% for testing (2,368 samples)

Design Development (Feature Extraction)

Feature Extraction

$$\text{Variance}(\sigma^2) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\textit{Mean Absolute Deviation} = \frac{1}{N}\sum_{i=1}^{N} | x_i - \mu |$$

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}\right)^3}$$

$$\text{Standard Deviation}(\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2\right)^2}$$

$$P_{\text{band}} = \sum_{f=f_1}^{f_2} |X(f)|^2$$

Delta Band: 0.5 Hz$\leq f <$4 Hz

Theta Band: 4 Hz$\leq f <$8 Hz

Alpha Band: 8 Hz$\leq f <$13 Hz

Beta Band: 13 Hz$\leq f <$30 Hz

Gamma Band: 30 Hz$\leq f \leq$100 Hz

Design Development (Feature Extraction)

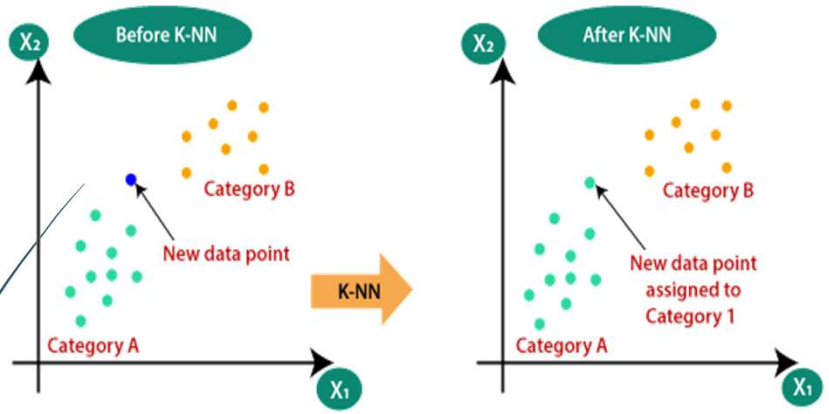
Selected Features

- Statistical Features
  - $Standard\ Deviation(\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$
  - $Mean\ Absolute\ Deviation = \frac{1}{N}\sum_{i=1}^{N} | x_i - \mu |$

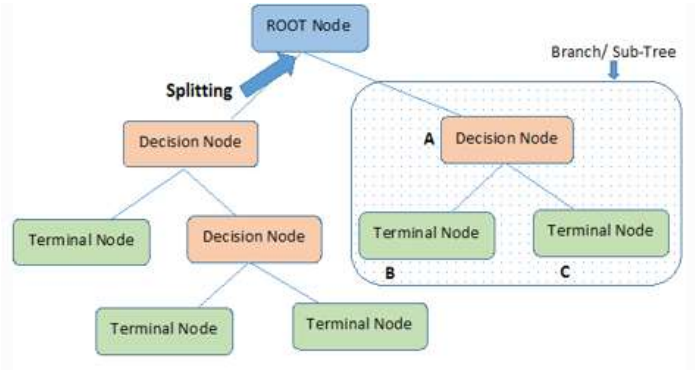
Design Development (Classification)

Used Classifiers

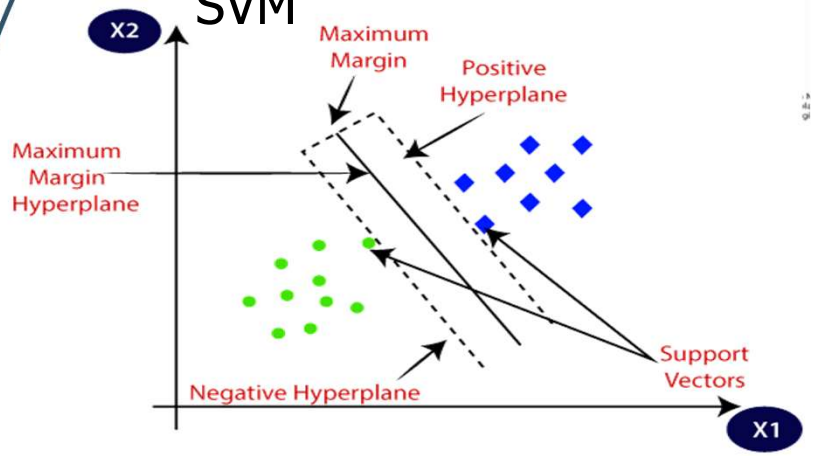
KNN

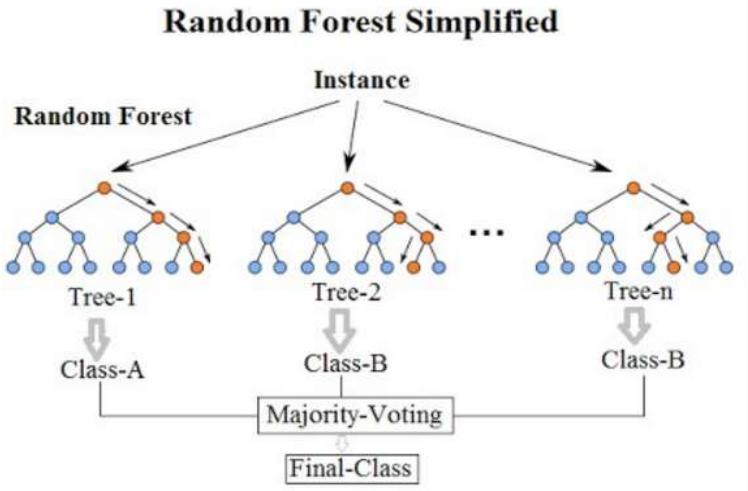


Decission Tree



SVM





Design Development (Classification)

Classes

- Binary Classification
  - Visual vs No Visual (2 Classes)
  - Arithmetic vs No Arithmetic (2 Classes)
- Multiclass Classification (4 Classes)
  - Relaxed with Eyes Opened
  - Relaxed with Eyes Closed
  - Arithmetic with Eyes Opened
  - Arithmetic with Eyes Closed

Design Development (Classification)

Data Size

- Binary Classification
  - Training Data: 10,900 samples (80%)
  - Test Data: 1160 samples (20%)
- Multiclass Classification (4 Classes)
  - Training Data: 10,900 samples (80%)
  - Test Data: 1160 samples (20%)

Design Development (Data Segmentation)

Data Segmentation Techniques

- Two Classes
  - Visual
    - Combined data of "Relaxed with Eyes Opened" and "Arithmetic with Eyes Opened"
  - No Visual
    - Combined data of "Relaxed with Eyes Closed" and "Arithmetic with Eyes Closed"

Design Development (Data Segmentation)

Model Test Accuracies

Models	1 Second Windows
Logistic Regression	74.8
Decision Tree	72.1
Random Forest	83.3
SVM (linear)	75.6
SVM (rbf)	82.3
KNN	78.2

Design Development (Data Segmentation)

Model Test Accuracies

Models	2 Second Windows	
	No Overlap	50% Overlap
Logistic Regression	79.5	77.7
Decision Tree	78.2	79.4
Random Forest	85.9	89.2
SVM (linear)	80.2	79.1
SVM (rbf)	85.0	87.4
KNN	84.1	87.6

Design Development (Data Segmentation)

Model Test Accuracies

Models	3 Second Windows		
	No Overlap	33% Overlap	66% Overlap
Logistic Regression	84.4	81.3	81.7
Decision Tree	83.8	82.6	80.3
Random Forest	89.5	91.0	91.9
SVM (linear)	85.1	82.2	82.6
SVM (rbf)	89.0	89.1	90.4
KNN	88.6	90.6	92.3

Design Development (Data Segmentation)

Model Test Accuracies

Models	4 Second Windows		
	No Overlap	25% Overlap	75% Overlap
Logistic Regression	83.3	83.1	83.8
Decision Tree	84.8	81.8	83.9
Random Forest	90.0	90.9	95.4
SVM (linear)	83.0	83.2	84.6
SVM (rbf)	87.8	88.8	92.2
KNN	90.0	89.4	94.5

Design Development (Data Segmentation)

Model Test Accuracies

Models	5 Second Windows		
	No Overlap	20% Overlap	80% Overlap
Logistic Regression	83.3	84.0	86.9
Decision Tree	85.4	82.4	88.9
Random Forest	91.0	89.9	97.0
SVM (linear)	84.6	84.0	87.3
SVM (rbf)	87.9	88.5	94.2
KNN	90.2	91.9	97.2

Design Development (Data Segmentation)

Model Test Accuracies

Models	6 Second Windows		
	No Overlap	33% Overlap	66% Overlap
Logistic Regression	84.8	87.0	87.1
Decision Tree	82.3	86.1	90.0
Random Forest	91.0	93.1	95.7
SVM (linear)	84.6	86.5	87.4
SVM (rbf)	88.7	89.5	94.8
KNN	92.0	92.4	95.8

Design Development (Classification Results)

Model Test Accuracies (Best Technique)

Models	5 Second Window (80% Overlap)		
	Visual	Mental	Multiclass
Logistic Regression	86.9	65.8	61.1
Decision Tree	88.9	65.2	57.1
Random Forest	97.0	94.9	94.1
SVM (linear)	87.3	66.4	63.8
SVM (rbf)	94.2	85.7	83.0
KNN	97.2	93.9	92.3

Design Development (Continued.)

Result Comparison

- Online Dataset **[7]**
  - 10 Subjects
  - Two Classes (Arithmetic and Rest)
    - Relaxed (180s)
    - Arithmetic(60s)
  - Took limited data of Relaxed (60s)
- Custom Dataset
  - 10 Subjects
  - 4 Classes (300s each)
  - Only took limited data (60s) of "Relaxed with Eyes Closed" and "Arithmetic with Eyes Closed"

Design Development (Classification Results)

Model Test Accuracies

Models	5 Second Window (80% Overlap)	
	Online Data	Custom Data
Logistic Regression	91.3	70.1
Decision Tree	85.9	83.2
Random Forest	99.0	95.8
SVM (linear)	90.8	70.7
SVM (rbf)	99.5	89.1
KNN	99.5	98.9

Design Development (Pre-processing)

Independent Component Analysis(ICA)

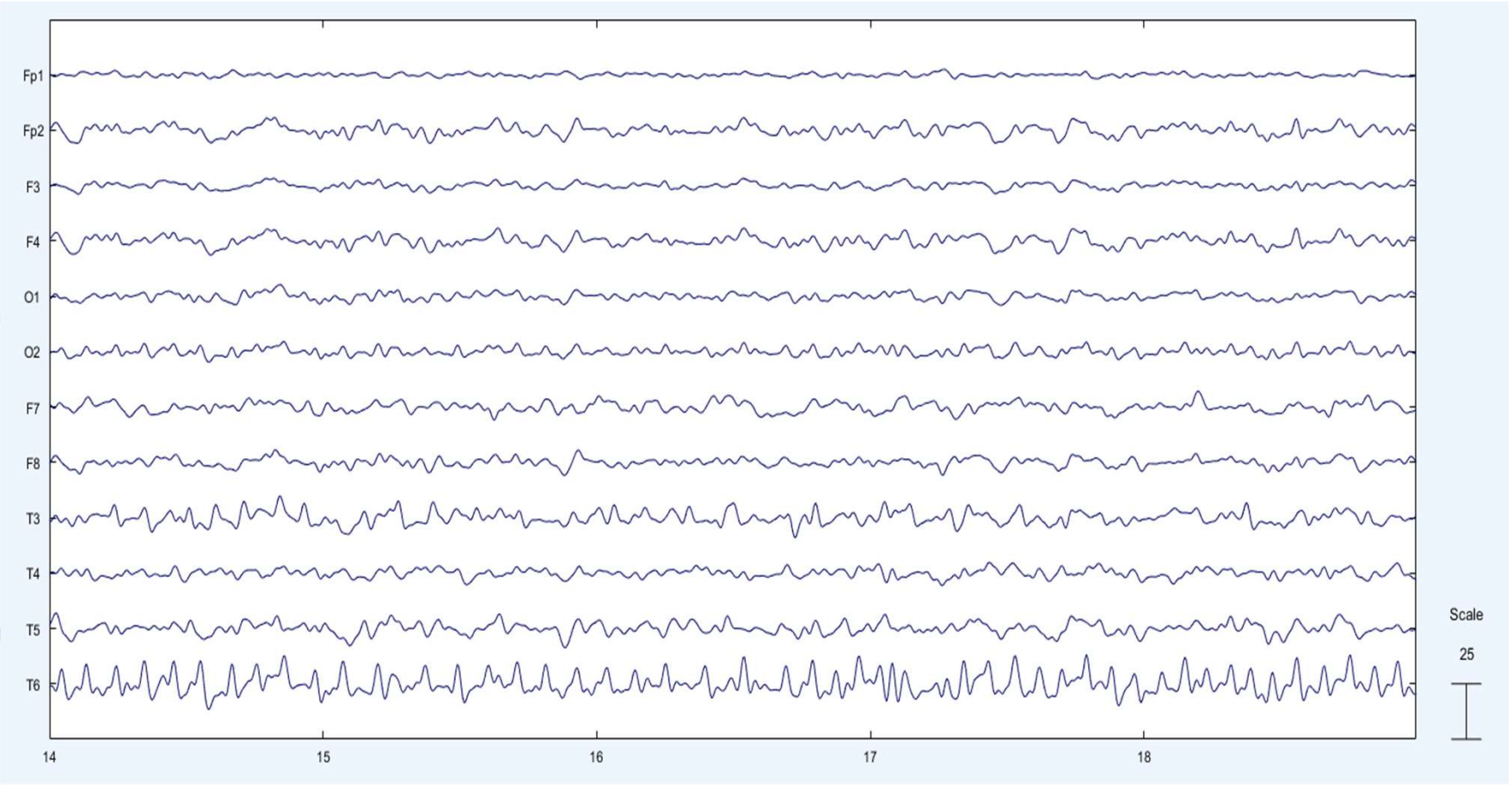
- EEG signals contains artifacts from muscle movements, eye blinks, etc
- ICA decomposes the signal into independent components using higher order statistics.
- Components removed
  - Muscle Activity
  - Eye Activity
  - Cardiac Activity

Design Development (Classification Results)

Model Test Accuracies (ICA)

Models	5 Second Window (80% Overlap)		
	Visual	Mental	Multiclass
Logistic Regression	85.0	67.2	67.1
Decision Tree	93.0	86.2	78.3
Random Forest	99.7	99.9	99.8
SVM (linear)	86.5	72.1	79.8
SVM	99.8	99.8	99.7
KNN	100	100	100

Design Development (Pre-processing)



RAW Data After ICA

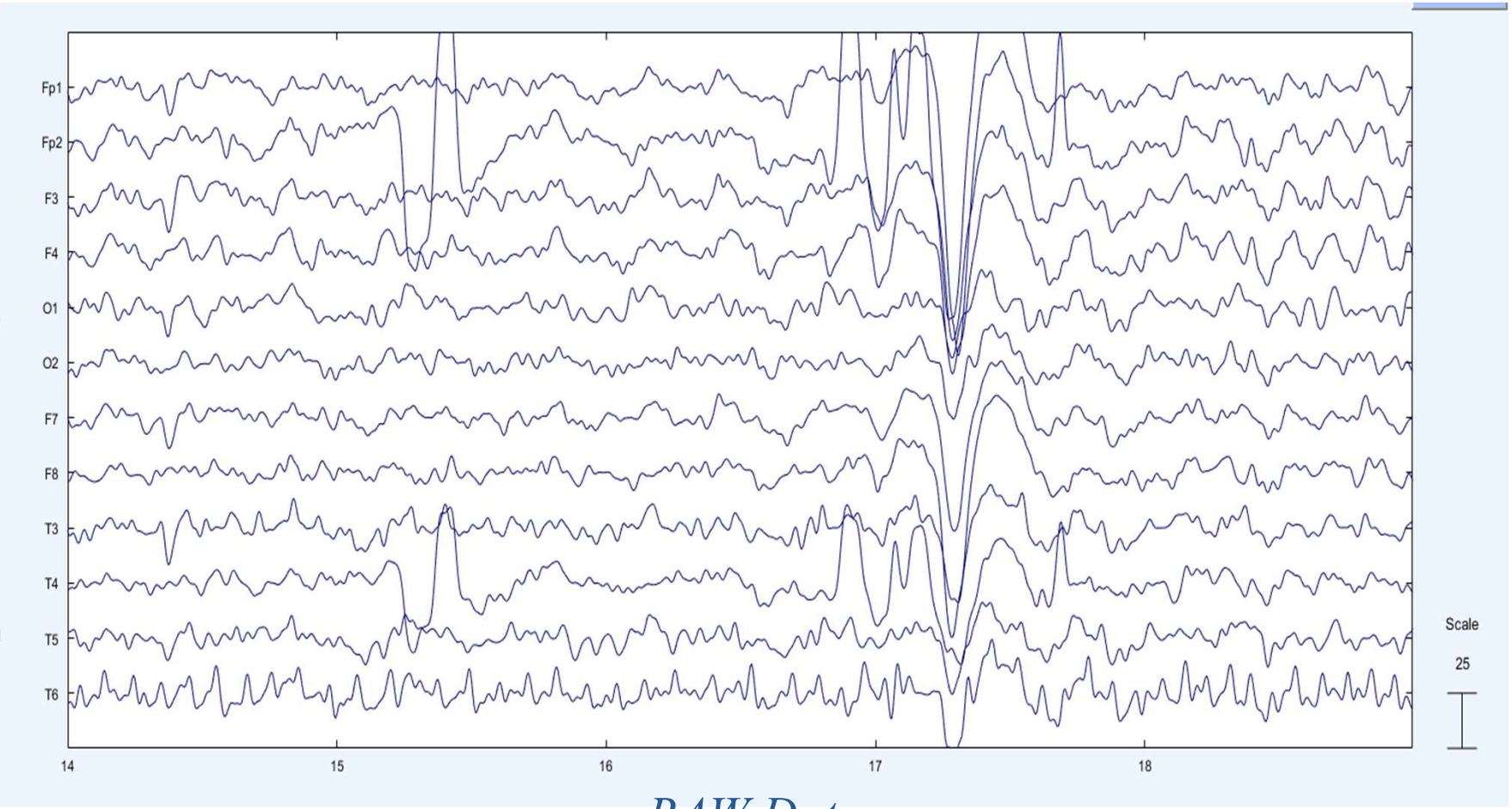
Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Design Development (Continued.)

Independent Component Analysis(ICA)

- **EEGLAB toolbox** in MATLAB was used for manual inspection and removal of artifacts.
- EEG data was loaded into EEGLAB, ICA was applied using the *runica* algorithm.
- Components were visually inspected based on:
  - Scalp Topography
  - Time Series
  - Power Spectral Density
- Artifactual components were manually selected and removed

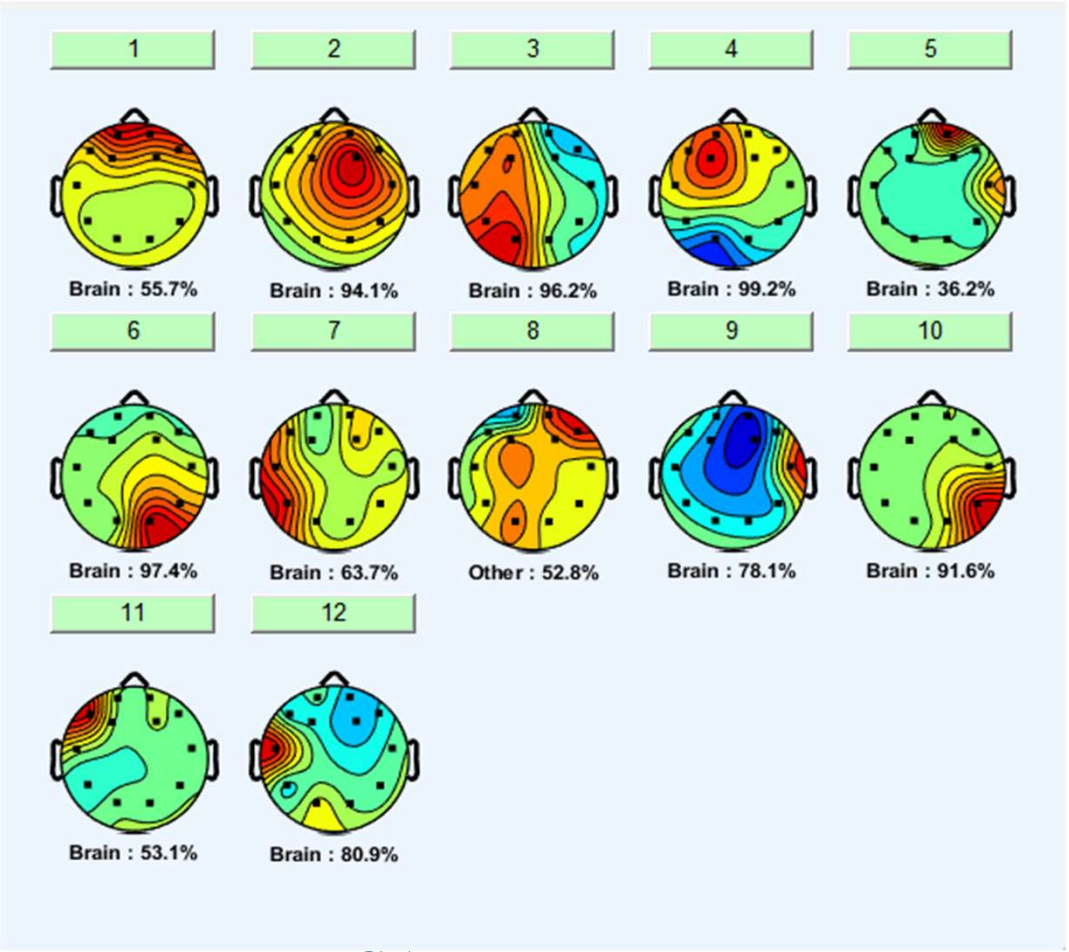
Design Development (Pre-processing)



RAW Data

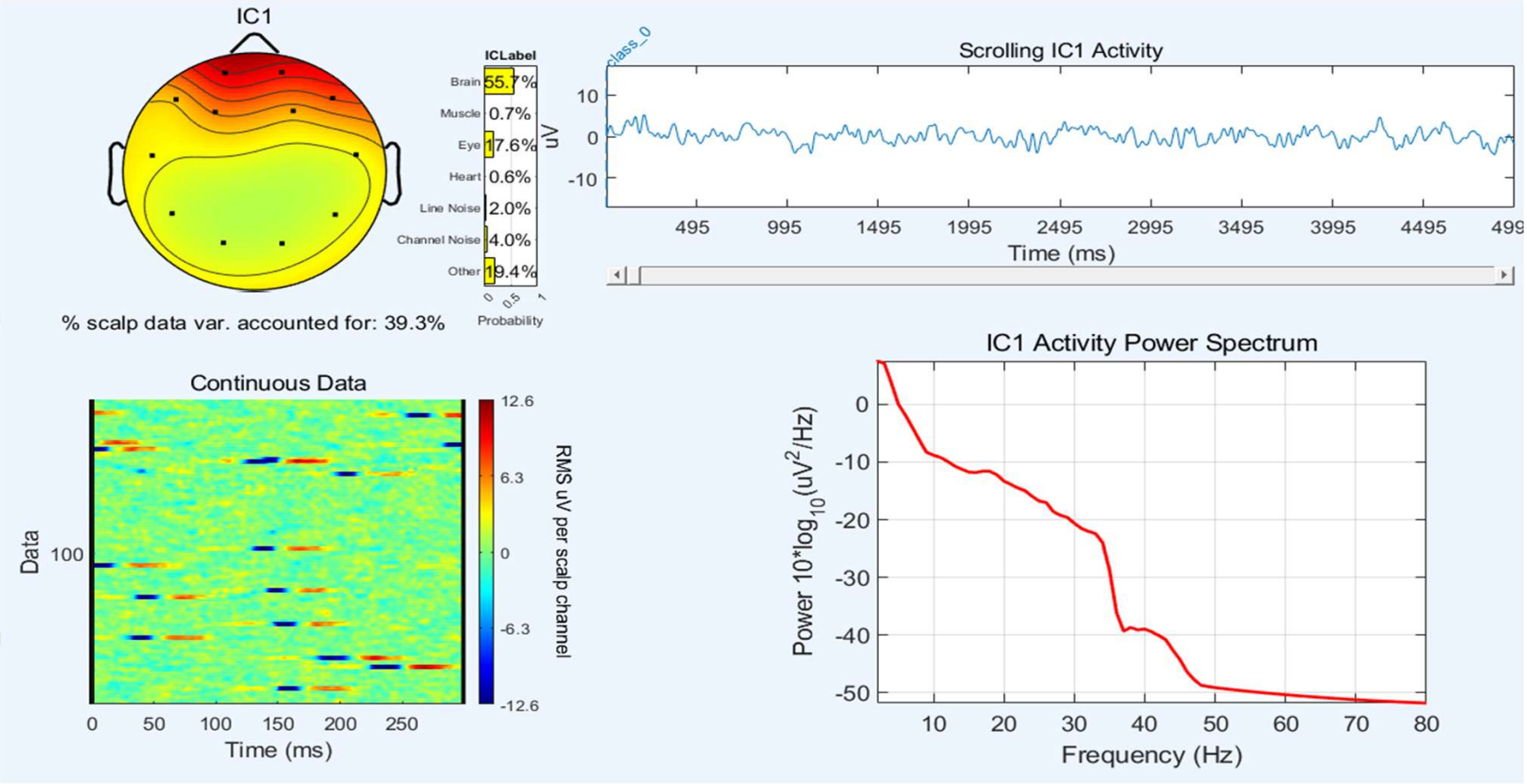
Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Design Development (Pre-processing)



ICA Decomposition

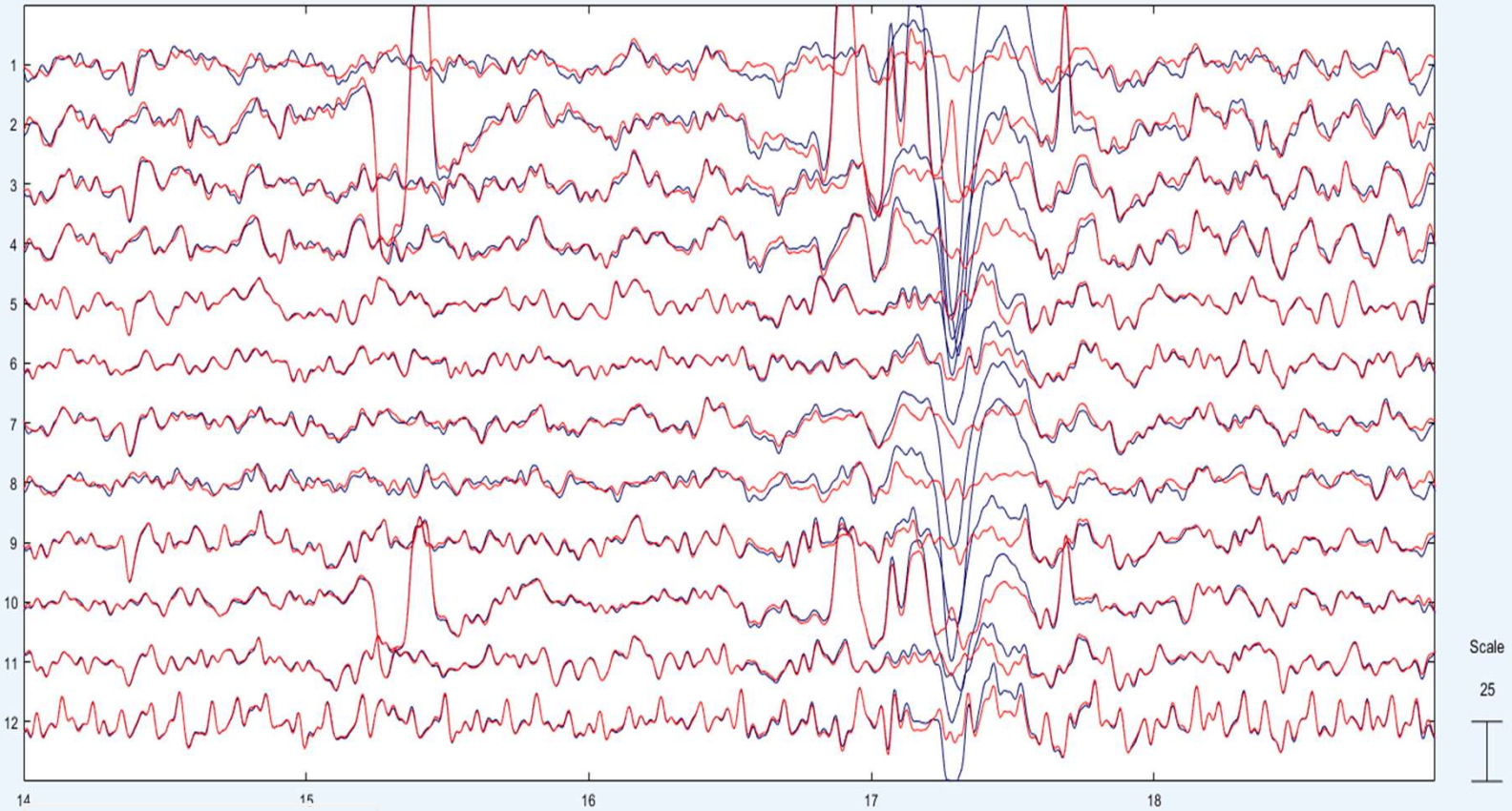
Design Development (Pre-processing)



Artifact Identification (IC1)

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

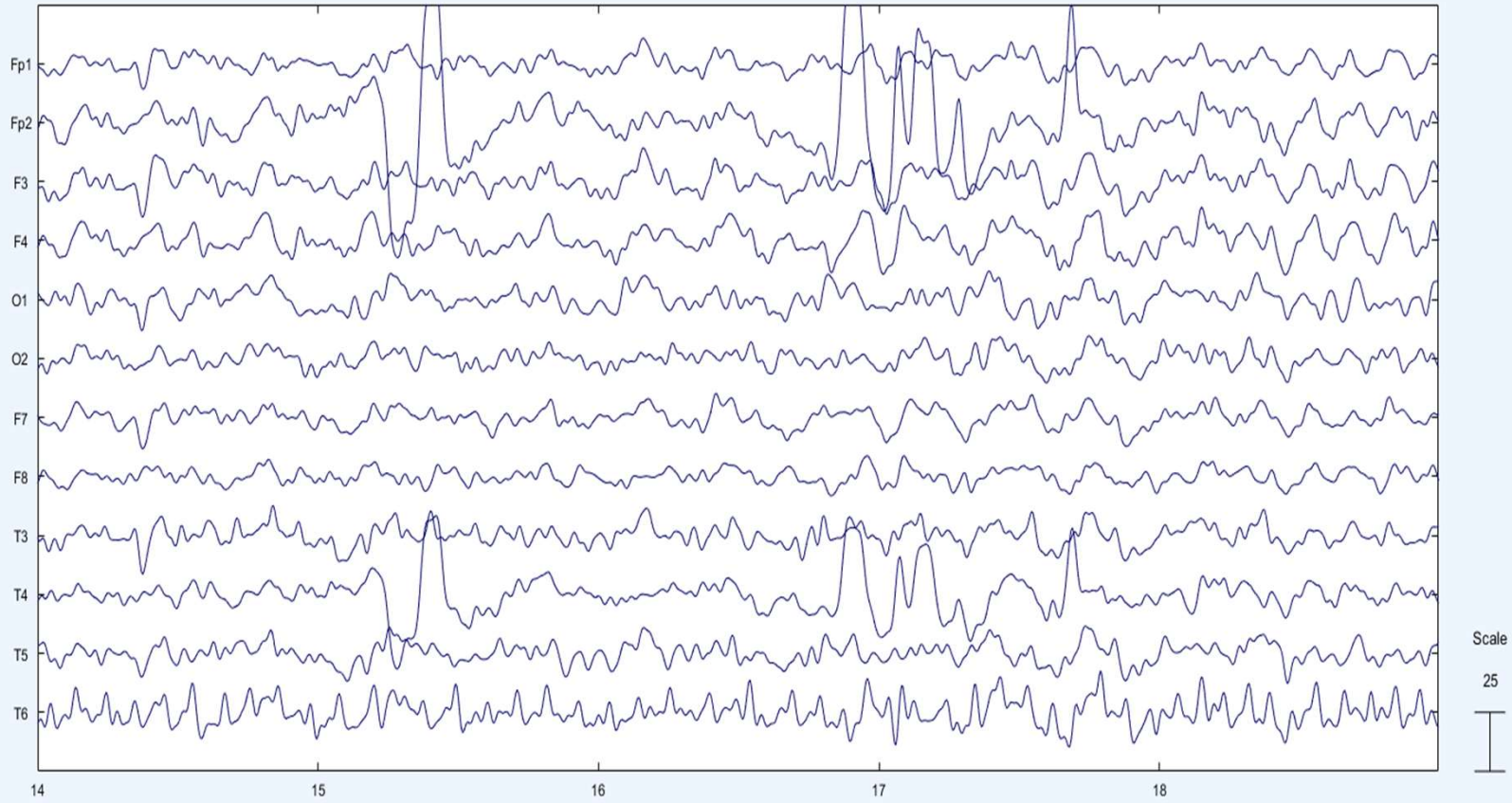
Design Development (Pre-processing)



Artifact Removal (IC1)

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

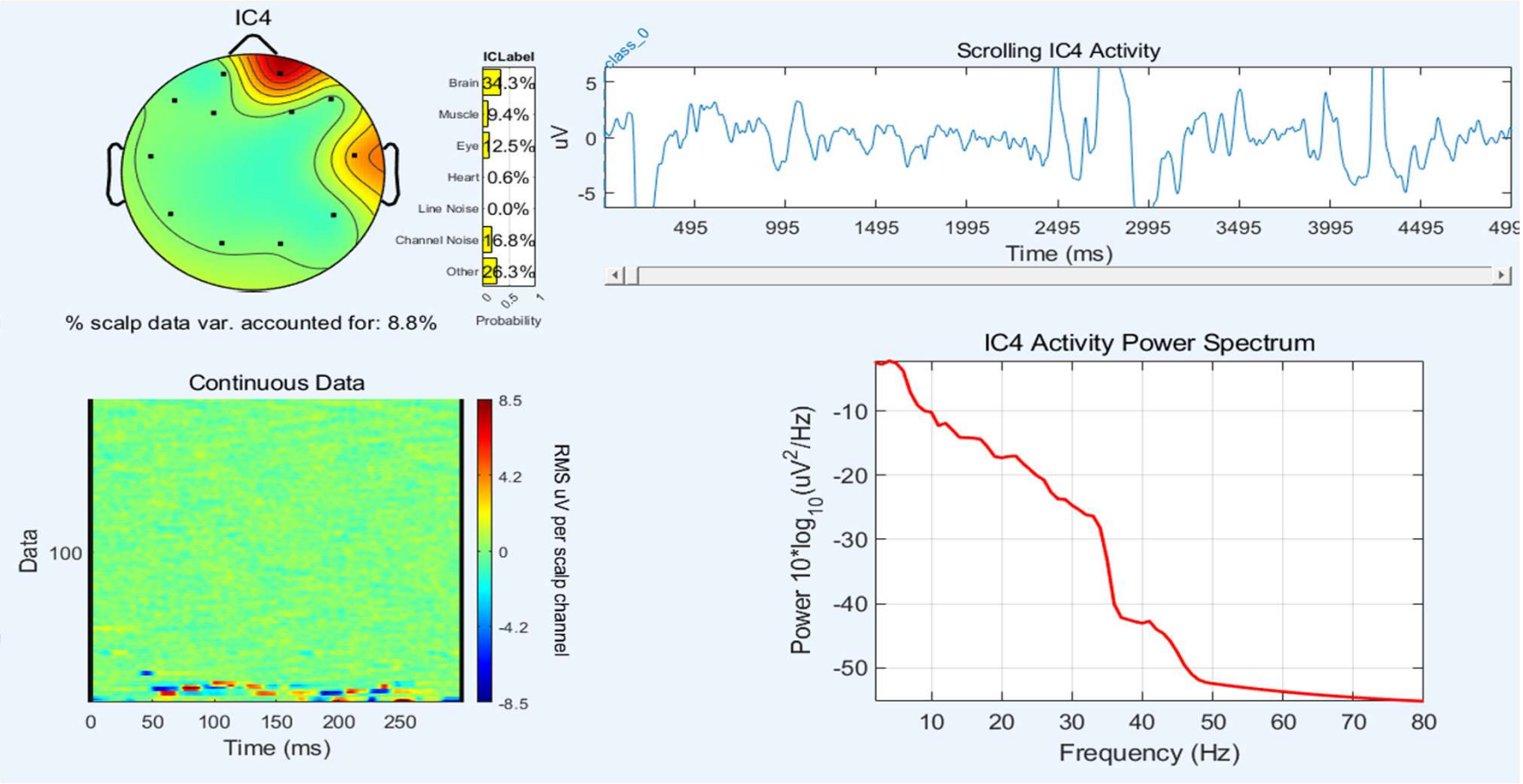
Design Development (Pre-processing)



Data After Removal of IC1

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

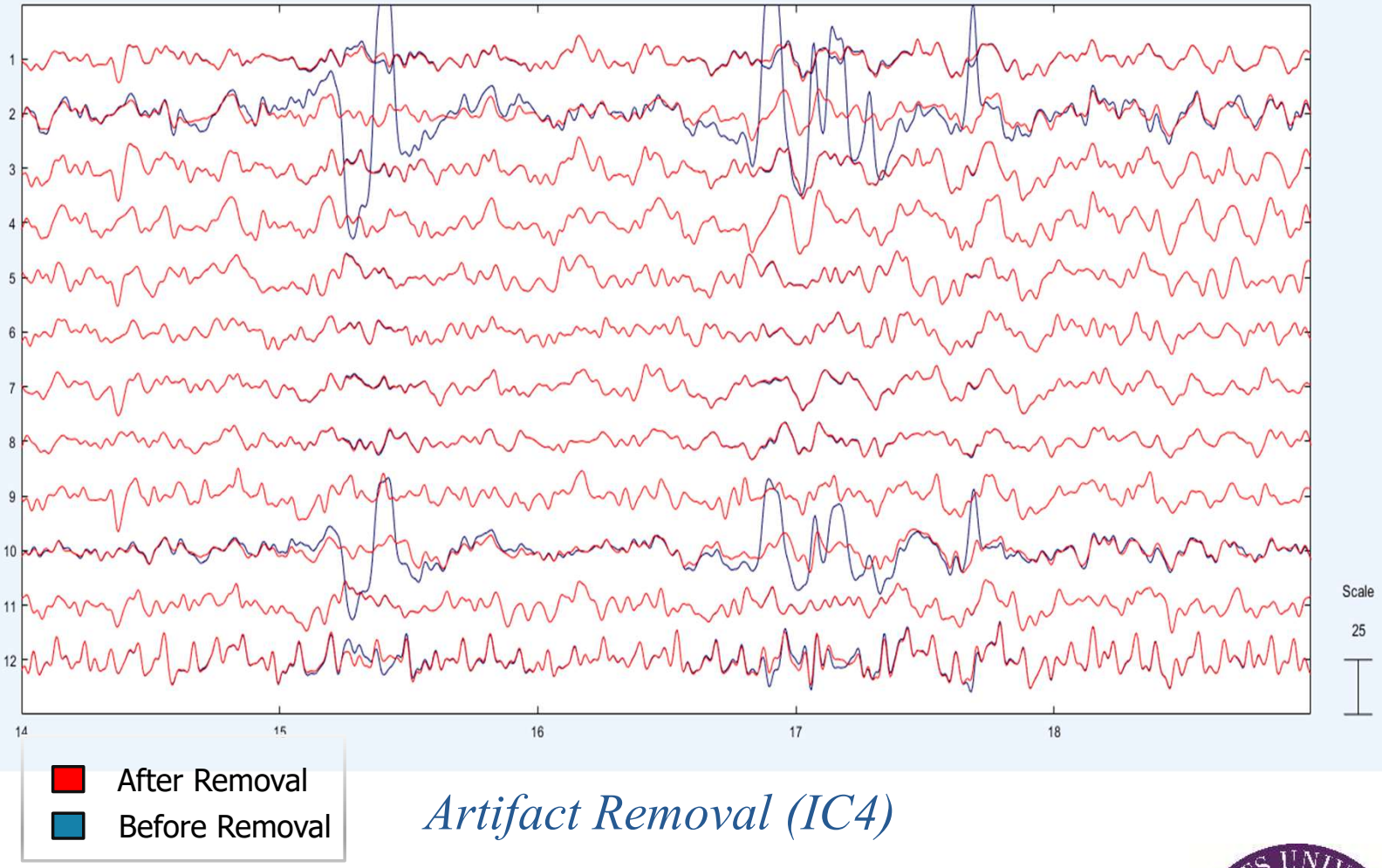
Design Development (Pre-processing)



Artifact Identification (IC4)

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

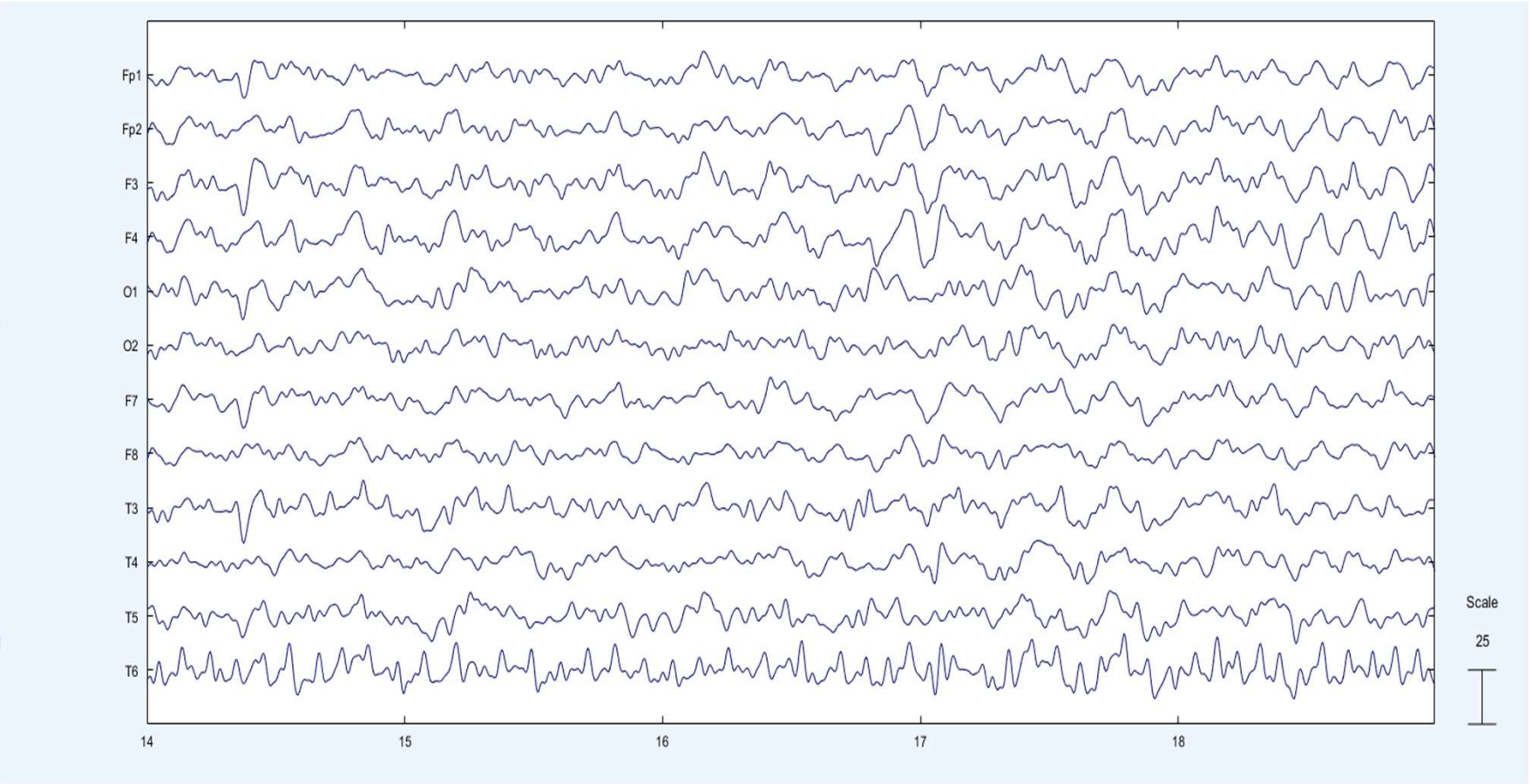
Design Development (Pre-processing)



Artifact Removal (IC4)

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

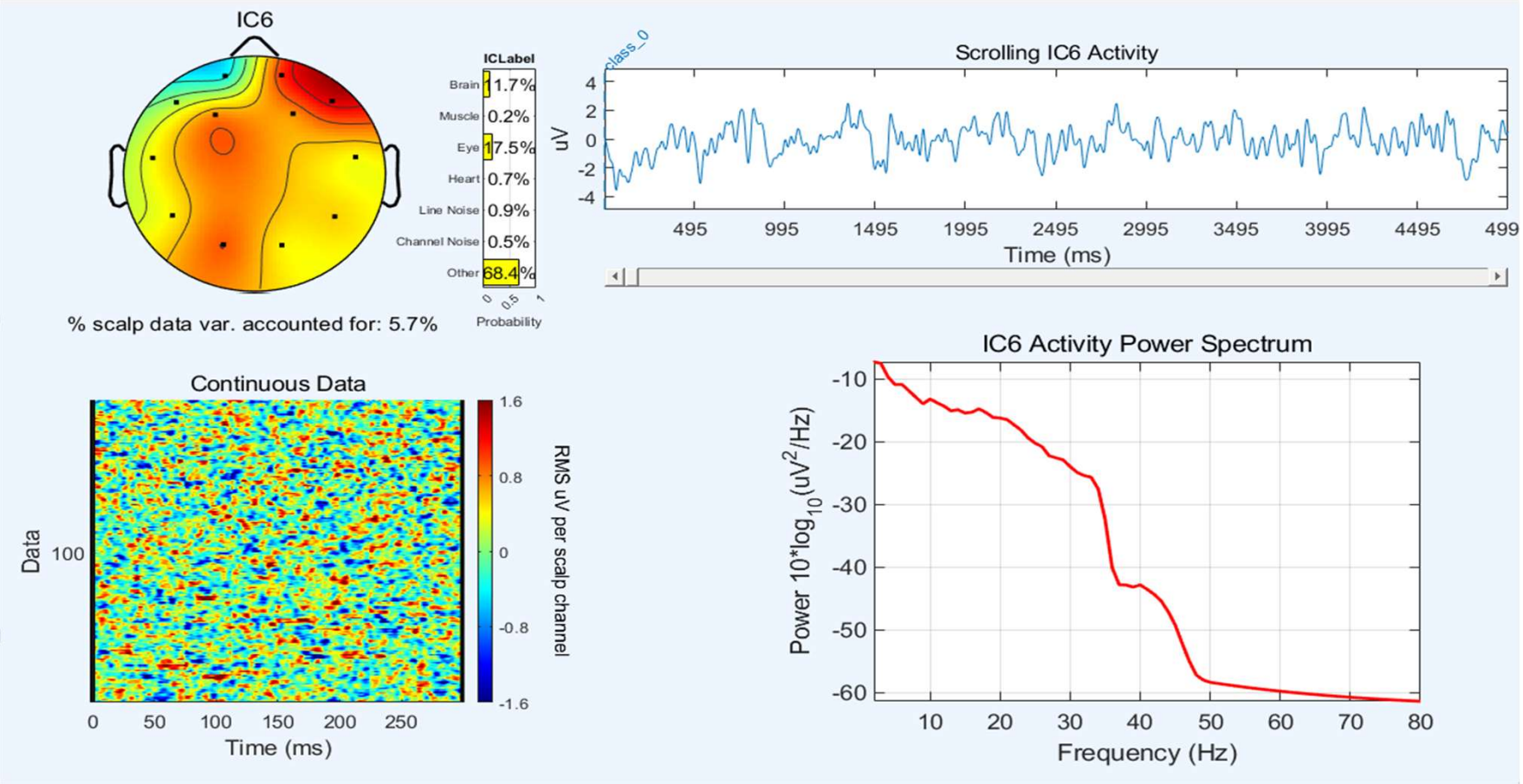
Design Development (Pre-processing)



Data After Removal of IC4

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

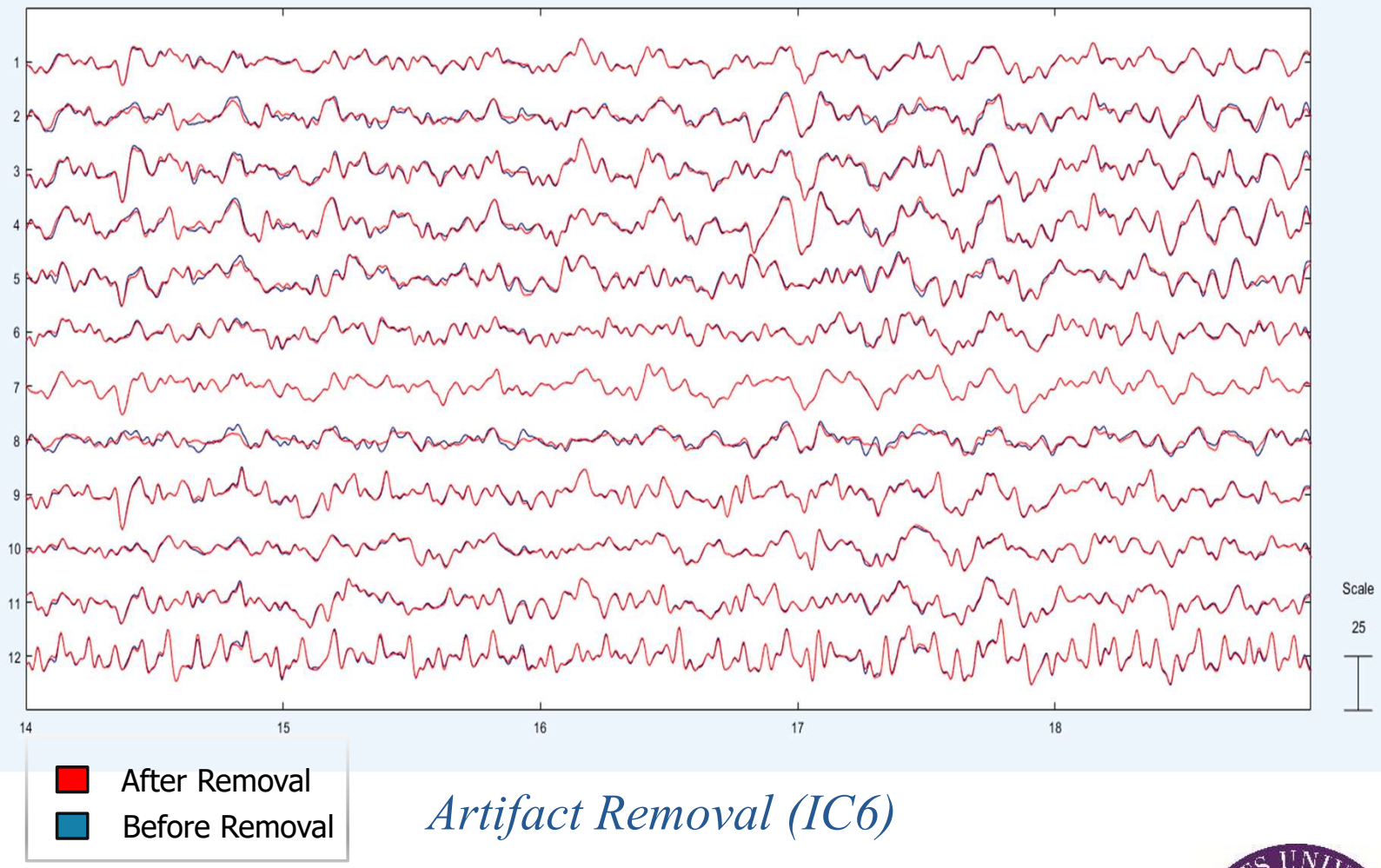
Design Development (Pre-processing)



Artifact Identification (IC6)

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

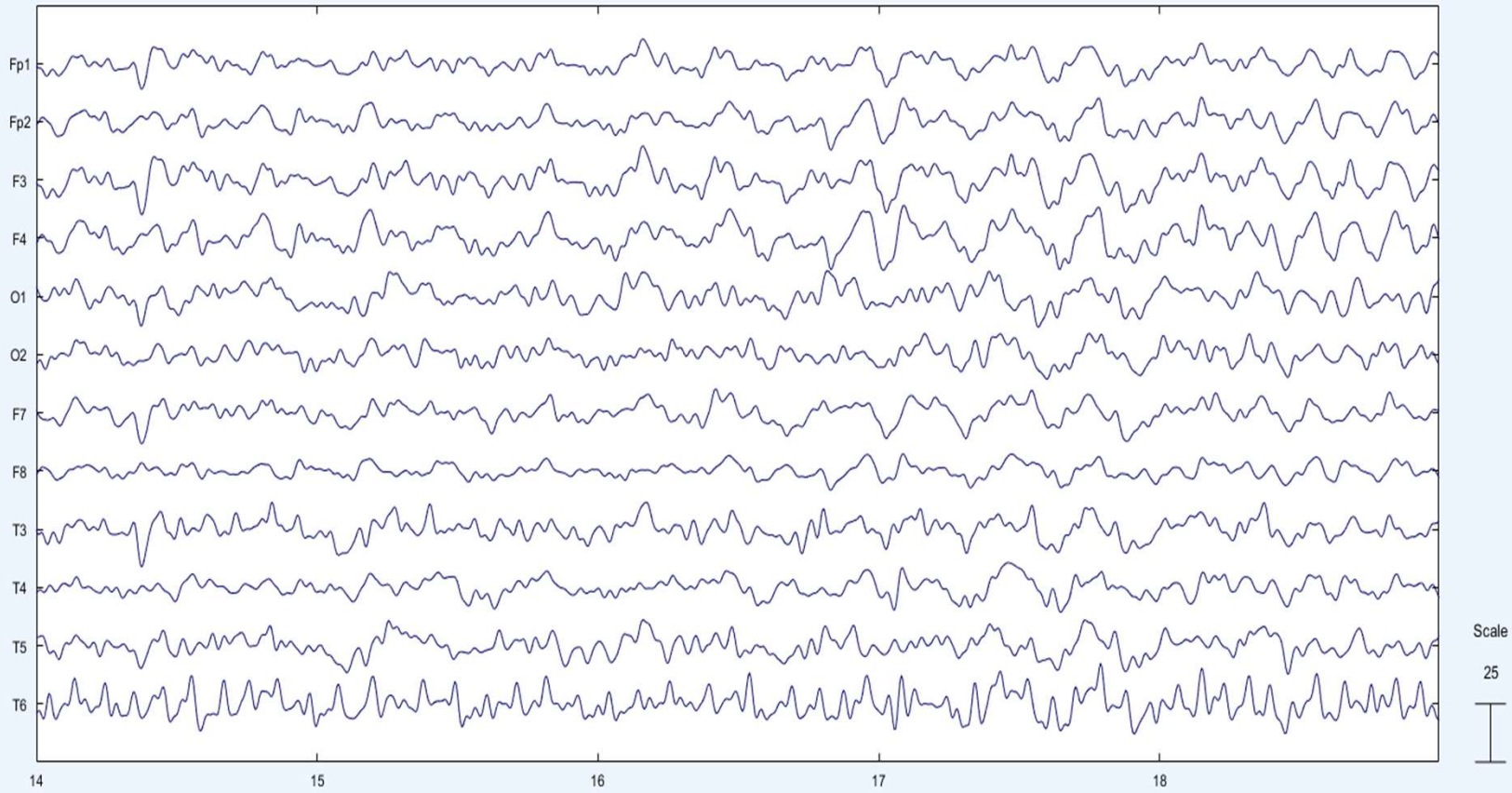
Design Development (Pre-processing)



Artifact Removal (IC6)

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Design Development (Pre-processing)



Clean Data After Removal of IC6

Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Results (Model Test Accuracies)

Model Test Accuracies (Manual ICA)

Models	5 Second Window (80% Overlap)		
	Visual	Mental	Multiclass
Random Forest	98.7	99.4	99.0
SVM (rbf)	97.6	98.3	97.5
KNN	98.4	99.2	98.7

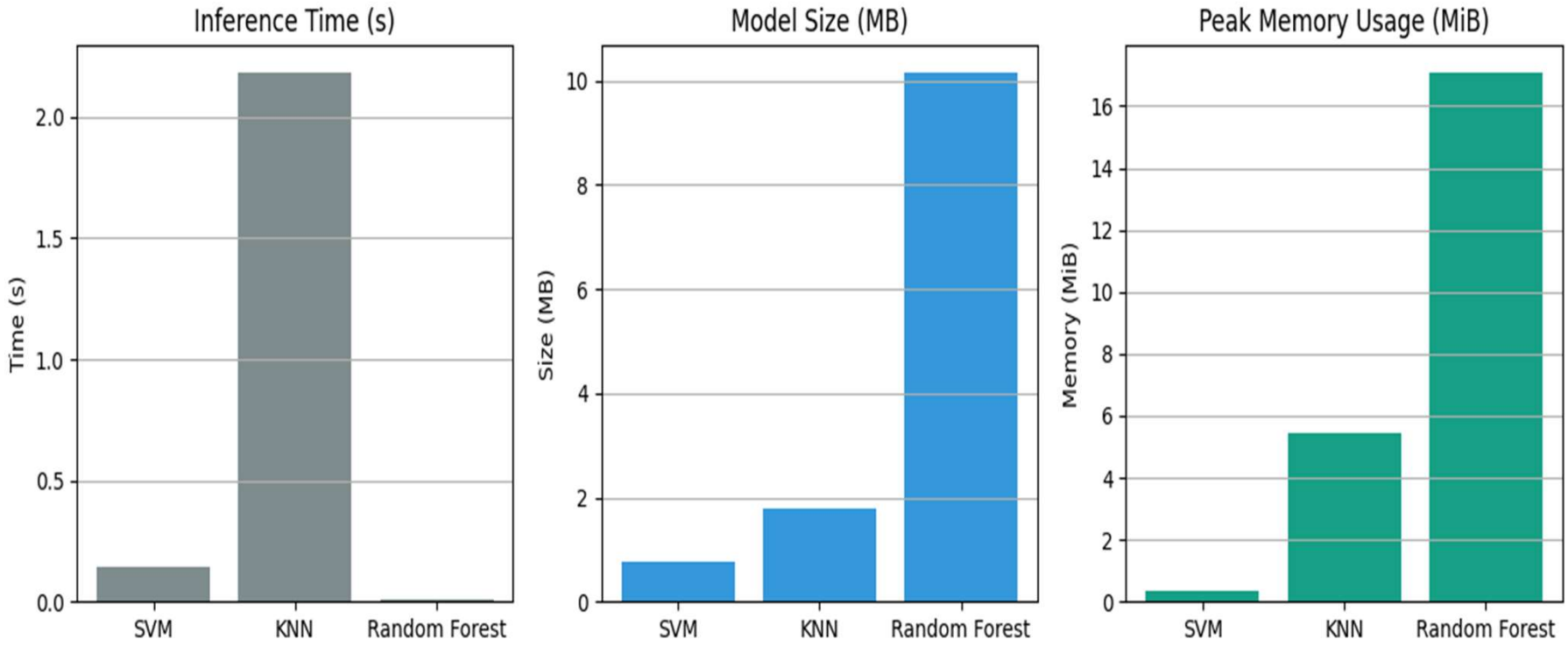
Results (Model Complexities)

Model Complexities (Multiclass Classification)

- Data Size: 300 second data of recorded EEG.
- Test Samples: 296
- System Description
  - Processor: 12th Gen Intel(R) Core(TM) i5-1235U
  - RAM: 8 GB
  - OS: Windows 11 Enterprise
- Preprocessing Pipeline Execution Time(Feature Extraction): 0.8s

Results (Model Complexities)

Model Complexities (Multiclass Classification)

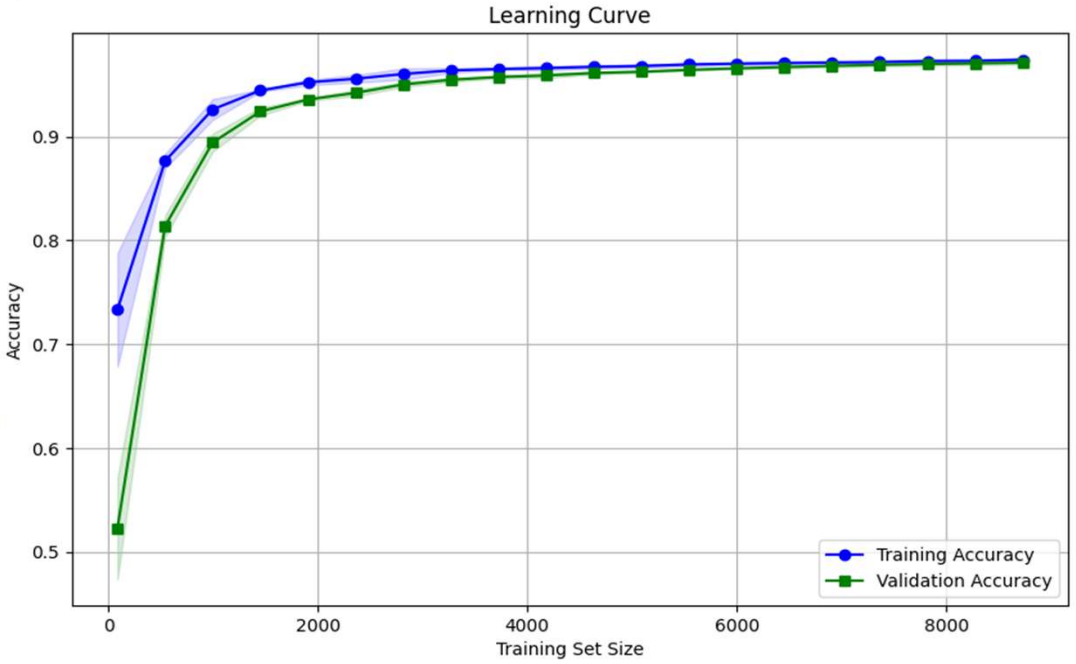


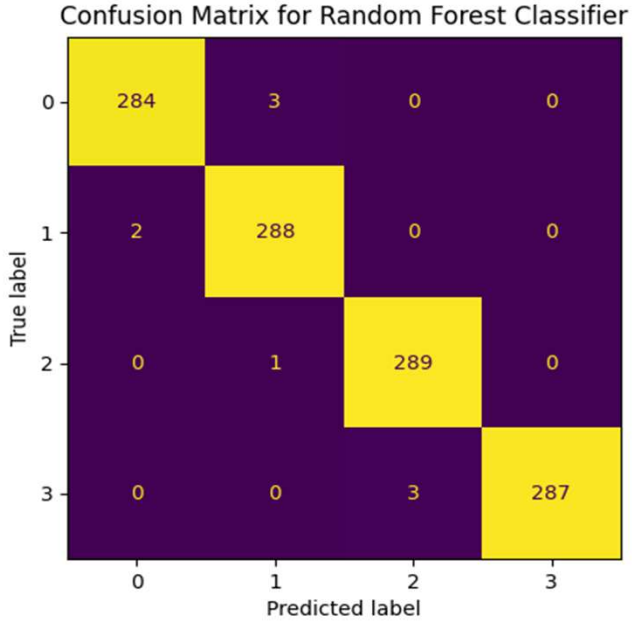
Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Results (Model Selection)

Random Forest

- Test Accuracy: 99.0 %
- Mean 3-fold Cross Validation Score: 98.23



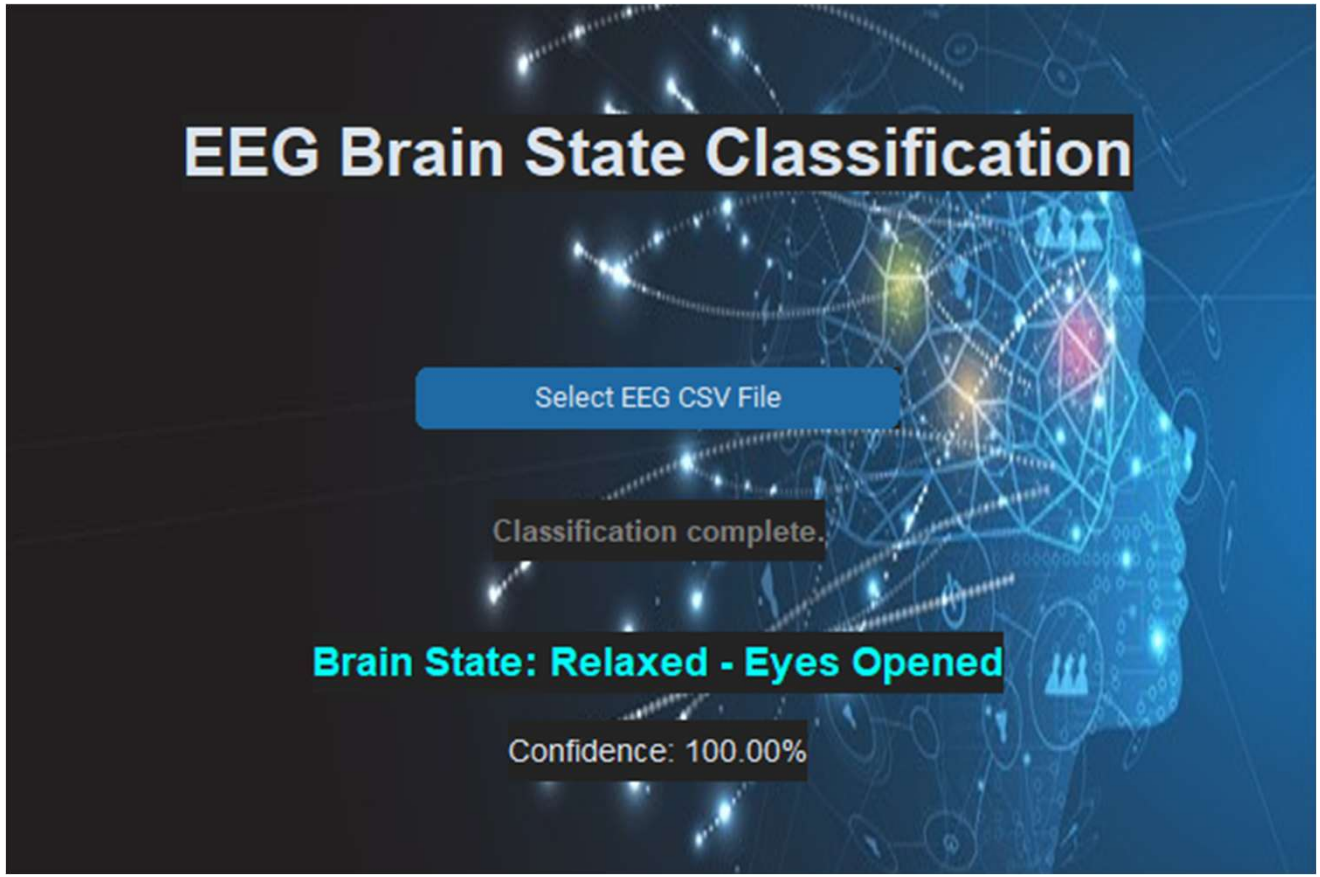


Department of Electrical & Computer Engineering
COMSATS University Islamabad, Abbottabad campus

Hardware Deployment (Hardware Details)

- Odroid XU4
  - Processor: Samsung Exynos5 Octa ARM Cortex™-A15 Quad 2Ghz
  - RAM: 2 GB LPDDR3 RAM
  - ROM: 16 GB
  - OS: Ubuntu Mate 6.1
  - Deployed Model: Random Forest

Hardware Deployment (Model Deployment)



Impact of the project

- Societal impact
  - Health Monitoring
    - Early diagnosis of neurological disorders.
    - Accessible EEG-based diagnostics.
  - Cognitive Enhancement
    - Useful applications by tracking focus and attention levels

Impact of the project (continued)

- Environmental impact.
  - Energy Efficiency
    - Low power device
  - Sustainability
    - Sustainable long-term usage.
    - Low environmental footprint.
  - Eco-friendly Practices
    - Reusable sensors and minimal electronic waste

Impact of the project (continued)

Reference: https://www.myclimate.org/en/information/faq/faq-detail/what-are-the-sustainable-development-goals-sdgs/

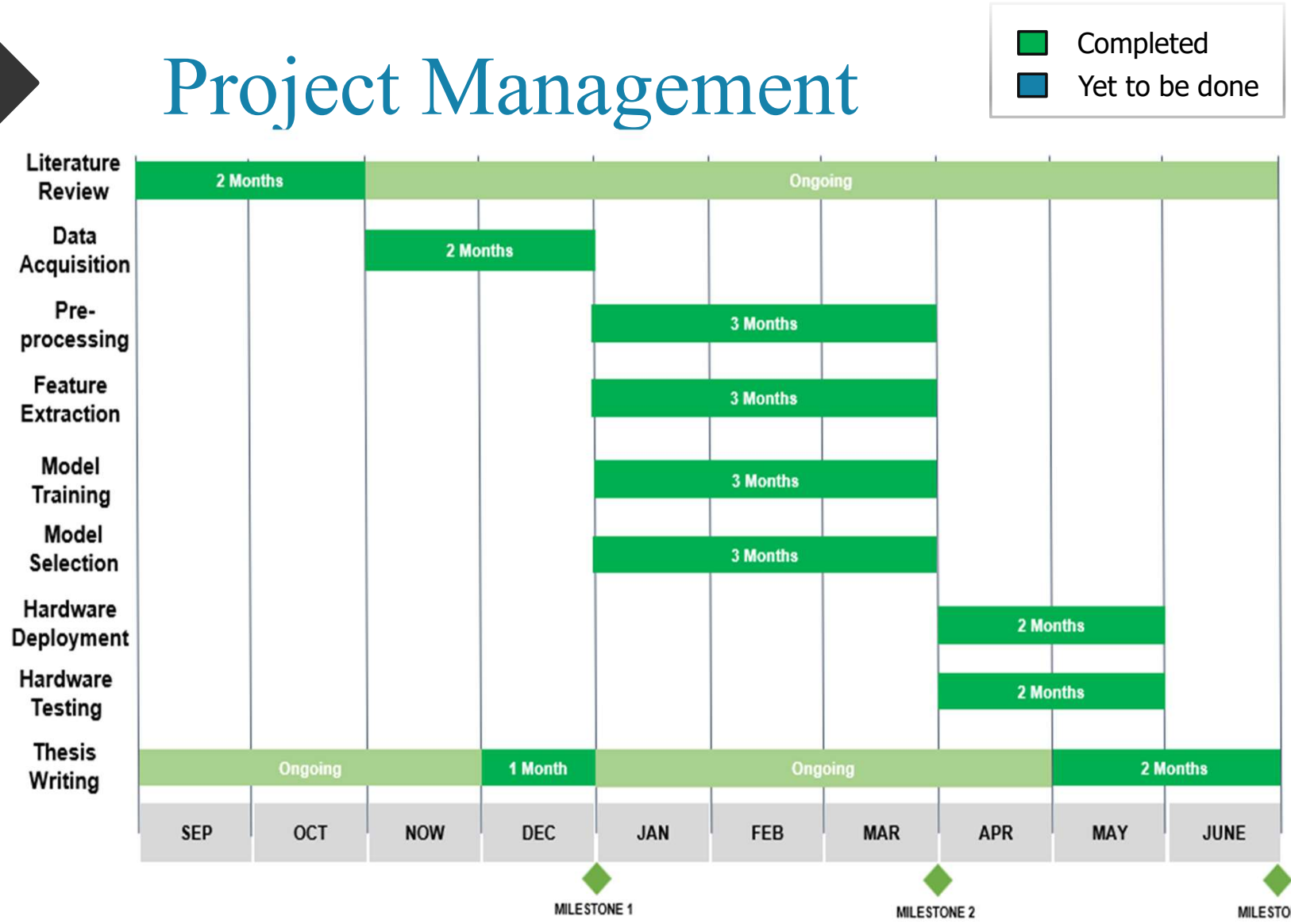
Impact of the project (continued)

- Targeted UN SGDs
  - SDG-3 : Good Health and Well-being
    - Early diagnosis and treatment
    - Accessible healthcare solutions
  - SDG-9 : Industry, Innovation, and Infrastructure.
    - Innovation in healthcare technology
    - Healthcare industry growth

Impact of the project (continued)

- Lifelong Learning
  - Technical Skills
    - Expertise in Machine Learning
    - Proficiency in Python programming language.
  - Standards and Protocols
    - Understanding EEG data processing standards
    - Familiarity with regulatory requirements of medical equipment
  - Problem Solving and Strategies
    - Experience in feature extraction and model optimization

Project Management



Reference: https://www.slideegg.com/free-timeline-chart-template-powerpoint

Project Management (continued)

Teamwork

Task Division			
Tasks	**Ahmed**	**Umer**	**Junaid**
Literature Review	✓✓	✓	✓
Data Acquisition	✓	✓✓	✓
Pre-processing	✓	✓	✓✓
Model Selection	✓✓	✓✓	✓✓
Model Training and Testing	✓✓	✓✓	✓✓
Hardware Deployment	✓	✓	✓✓
Thesis Writing	✓✓	✓✓	✓

Conclusion

- Successful Classification
  - Binary Classification
  - Multi-class Classification
- Potential of Various Applications
  - Healthcare
  - Assistive Technologies
- Future Directions
  - Integrating technology into real-world applications

References

[1] Antoniou, E., Bozios, P., Christou, V., Tzimourta, K.D., Kalafatakis, K., G. Tsipouras, M., Giannakeas, N. and Tzallas, A.T., 2021. EEG-based eye movement recognition using brain–computer interface and random forests. *Sensors*, *21*(7), p.2339.

[2] Sinha, N. and Babu, D., 2016, November. Statistical feature analysis for EEG baseline classification: Eyes Open vs Eyes Closed. In 2016 IEEE region 10 conference (TENCON) (pp. 2466-2469). IEEE.

[3] Mattiev, J., Sajovic, J., Drevenšek, G. and Rogelj, P., 2022. Assessment of Model Accuracy in Eyes Open and Closed EEG Data: Effect of Data Pre-Processing and Validation Methods. *Bioengineering*, *10*(1), p.42.

References (continued.)

[4] Ganguly, B., Chatterjee, A., Mehdi, W., Sharma, S. and Garai, S., 2020, July. EEG based mental arithmetic task classification using a stacked long short term memory network for brain-computer interfacing. In *2020 IEEE VLSI DEVICE CIRCUIT AND SYSTEM (VLSI DCS)* (pp. 89-94). IEEE.

[5] Islam, A., Sarkar, A.K. and Ghosh, T., 2021, July. EEG signal classification for mental stress during arithmetic task using wavelet transformation and statistical features. In *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)* (pp. 1-6). IEEE.

[6] Nirde, K., Gunda, M., Manthalkar, R. and Gajre, S., 2023, March. EEG mental arithmetic task levels classification using machine learning and deep learning algorithms. In 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP) (pp. 1-8). IEEE.

References (continued.)

[7] Zyma, I., Tukaev, S., Seleznov, I., Kiyono, K., Popov, A., Chernykh, M. and Shpenkov, O., 2019. Electroencephalograms during mental arithmetic task performance. Data, 4(1), p.14.