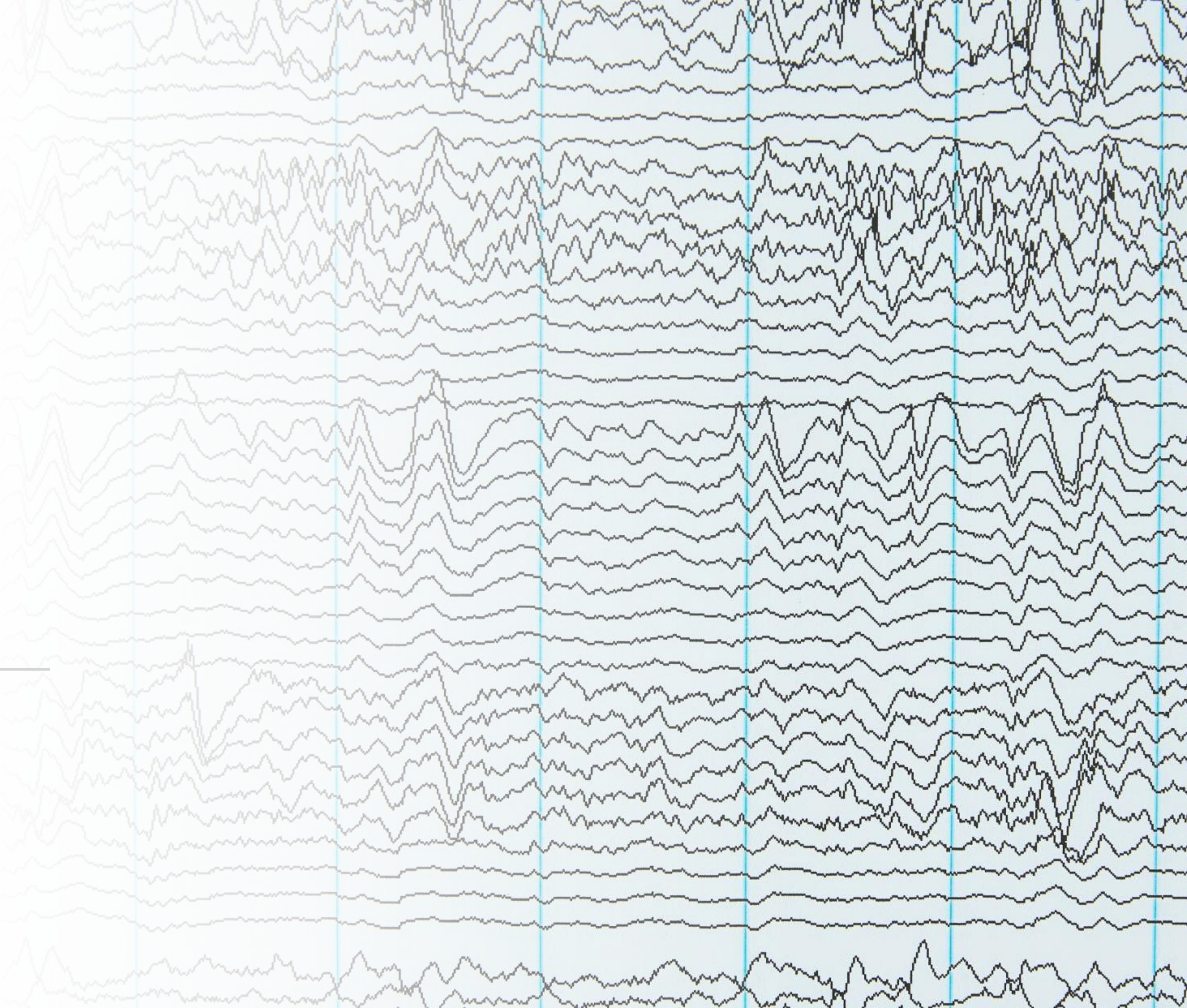
EEG Signal Classification Using Machine Learning



By: Muneezé Malik

Instructor: Dr. Adnan Amin

Course: Machine Learning Lab

Date: April 22, 2025

Introduction to EEG

🧠 What is EEG?

- EEG stands for Electroencephalography
- A non-invasive technique to record brain activity
- Measures electrical signals using electrodes placed on the scalp

📈 EEG Signals:

- Represent voltage fluctuations from neuron activity
- High temporal resolution — used in neuroscience, sleep studies, BCI, and more

Project Overview

📊 What Are We Trying to Do?

- We aim to classify EEG signals into different mental states (eye states)

- Use machine learning models to recognize patterns in EEG waveforms

🧪 Approach:

1. Preprocess raw EEG data from .edf format
2. Visualize and understand data distribution
3. Train ML models with and without feature engineering
4. Apply SMOTE to handle class imbalance
5. Compare model performance using F1-score

Why It Matters

Importance of EEG Classification:

- Helps in diagnosing neurological conditions
- Enables development of brain-computer interfaces (BCI)
- Enhances real-time cognitive state detection

ML + EEG:

- Manual EEG analysis is complex and time-consuming
- Machine Learning helps automate and improve accuracy

Introduction

🔍 **Objective:**

- Classify EEG signals using machine learning
- Evaluate models with and without:
  - Feature Engineering
  - Data Balancing (SMOTE)

🎯 **Goal:**

- Achieve the highest possible F1-scores using preprocessing techniques

Dataset Description

📦 EEG Dataset (Mendeley Data)

- 6 folders containing .edf EEG recordings

- Eye state classification (4 classes)

📐 Input Shape:

- x_train: (5499, 19, 500)

- 19 channels, 500-time samples per signal

Data Preprocessing

🔄 **Preprocessing Steps:**

- Converted .edf to .npy using MNE-Python
- Flattened EEG signals for raw modelling
- Extracted statistical features (mean, std, max, min)

🧹 **Libraries Used:**

- MNE, NumPy, Scikit-learn, SMOTE

Visualization

📉 **Class Imbalance Observed:**

- Class 3 had very few samples

📈 **EEG Signal Plot:**

- Time-series plots for 3 sample channels

- Visual verification of signal patterns

📊 Used Countplot & Line Plot

Machine Learning Models

🧠 Models Used:

1. Logistic Regression
2. Random Forest
3. Support Vector Machine

🧪 Evaluation Metric:

- Weighted F1-Score
- Confusion Matrix

Results – Raw Data

⚙️ **Input:** Flattened EEG signals (9500 features)

📊 **F1 Scores:**

- Logistic Regression: 0.4541
- Random Forest: 0.8732
- SVM: 0.6815

⚠️ Class 3 was not recognized well by any model

Feature Engineering & SMOTE

📌 **Feature Engineering:**

- Mean, Std, Max, Min → 76 features

⚖️ **Data Balancing:**

- SMOTE is used to oversample minority classes

📈 **Result**: Balanced class distribution

Results – After Feature Engineering

📊 F1 Scores (Balanced Data):

- Logistic Regression: 0.8904
- Random Forest: 0.9375
- SVM: 0.8961

🎯 Observation:

- Accuracy improved significantly
- Best performance by Random Forest

Final Comparison Table

🏆 Best Model: Random Forest (F1 = 0.9375)

Model	Raw F1 Score	Balanced F1 Score
Logistic Regression	0.4541	0.8904
Random Forest	0.8732	0.9375
SVM	0.6815	0.8961