PROBLEM STATEMENT: 1

Classification of Schizophrenia Using EEG Signals

Introduction:

Humans and many other animals can suppress their brain's responses to sensory events caused by their actions. This happens through the corollary discharge forward model, where an "efference copy" of a planned movement is sent to the sensory cortex, representing the expected sensory outcome. For example, when you move your eyes, your brain knows the world isn't moving, and when you speak, your brain expects the sound of your voice.

Schizophrenia, a chronic mental illness affecting about 1% of people worldwide, may involve problems with this corollary discharge process. This makes it hard for patients to distinguish between self-generated and external stimuli. Studying this process can help us understand abnormal brain functions in schizophrenia.

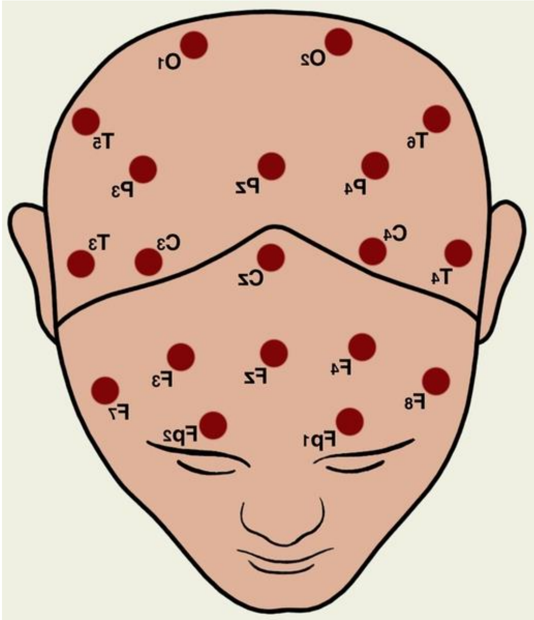
Dataset:

The dataset includes EEG records from two groups: healthy adolescents (n = 39) and adolescents with schizophrenia symptoms (n = 45). Each subject's EEG data is in a separate TXT file, with columns representing EEG samples from 16 channels. Each number is an EEG amplitude (mkV) at a distinct sample. The first 7680 samples are for the first channel, the next 7680 for the second channel, and so on. The sampling rate is 128 Hz, meaning 7680 samples represent 1 minute of EEG recording.

The dataset is available at: [MSU EEG Schizophrenia Data](http://brain.bio.msu.ru/eeg_schizophrenia.htm).

The topographical positions of the channel numbers are illustrated in the table and image below.

1 - F7
2 - F3
3 - F4
4 - F8
5 - T3
6 - C3
7 - Cz
8 - C4
9 - T4
10 - T5
11 - P3
12 - Pz
13 - P4
14 - T6
15 - O1
16 - O2



Objectives:

1. Implement a classification algorithm to differentiate between healthy individuals and those with schizophrenia.

2. Calculate classification metrics such as accuracy, confusion matrix, precision, recall, F1 score, etc.

3. Compare your technique with at least one other method.

4. Use Convolutional Neural Networks (CNNs) and report on their efficiency with small datasets.

PROBLEM STATEMENT-2

Real-Time Traffic Accident Detection in Smart Cities Using Video Data

Introduction:

A novel dataset comprising approximately 5,700 video files has been curated to enhance the development of real-time traffic accident detection systems in smart city environments. This dataset captures a diverse range of traffic scenarios through Traffic/Surveillance Cameras (Trafficam) and Dash Cameras (Dashcam), as well as additional external data sources. The data has been meticulously organized into three segments: Training, Validation, and Testing, each offering a unique mix of traffic and dashcam footage across various scenarios.

Dataset:

The dataset is classified into eight distinct classes:

1. Backend
2. Backend Rollover
3. Frontend
4. Frontend Rollover
5. No Accident Normal Traffic
6. Sidehit
7. Sidehit Rollover
8. General Augmented Crash

This rich classification covers a spectrum of real-world conditions, from regular traffic flow to intricate accident scenes. The dataset distribution is as follows:

- Training: 3,912 files

- Validation: 1,054 files

- Testing: 725 files

The videos are segmented into five-second non-overlapping clips to focus on the rapid dynamics of accidents, providing concise and relevant data for model training and evaluation.

Dataset Structure:

The dataset is organized into three primary directories: Training, Validation, and Testing. Each directory contains zipped files for different classes of traffic scenarios, ensuring a structured and organized approach to data management.

The dataset can be obtained from [IEEE Dataport](https://ieee-dataport.org/documents/traffic-accident-detection-video-dataset-ai-driven-computer-vision-systems-smart-city).

Objectives:

1. Implement a traffic accident detection algorithm using the provided video dataset.

2. Calculate performance metrics such as accuracy, confusion matrix, precision, recall, F1 score, etc.

3. Compare your technique with another method to establish benchmark performance.

4. Evaluate the efficiency of different models, including Convolutional Neural Networks (CNNs), in handling the diverse scenarios presented in the dataset.

PROBLEM STATEMENT-3

Optimizing Wireless Communication: The Role of Spectrum Sensing in Efficient Spectrum Management

Details of a Challenging Problem

In our increasingly connected world, the radio spectrum has become a crucial resource, underpinning the functioning of countless wireless communication systems. From mobile phones to Wi-Fi networks, GPS to satellite communications, our daily lives depend on the efficient use of the radio spectrum. However, as the demand for wireless communication grows, so does the need for innovative solutions to manage and utilize this finite resource more effectively. This is where spectrum sensing comes into play. Spectrum sensing refers to the ability of a cognitive radio to detect unused spectrum bands and adapt its transmission strategy to use those bands without causing interference to licensed users. It is akin to finding an empty lane on a busy highway, allowing traffic to flow more smoothly without causing congestion. Effective spectrum sensing can significantly enhance the efficiency of spectrum usage, enabling higher data rates, reduced interference, and more reliable communication.

In real-life applications, dynamic spectrum access allows devices to dynamically access available spectrum bands in environments where spectrum demand fluctuates, such as in emergency response scenarios, ensuring reliable communication when it is needed most. In densely populated areas where multiple wireless devices operate simultaneously, spectrum sensing helps mitigate interference, ensuring that communication remains clear and uninterrupted. As the Internet of Things (IoT) expands, the need for efficient spectrum management becomes critical, and spectrum sensing enables IoT devices to communicate effectively without overwhelming the available spectrum.

Identifying whether a signal is present in a given frequency band is the first step in spectrum sensing, known as signal detection. It ensures that cognitive radios do not interfere with ongoing transmissions by licensed users. Once a signal is detected, classifying its type (e.g., BPSK, QPSK, 8PSK) helps determine the appropriate action for the cognitive radio, such as avoiding interference or adapting its transmission parameters. However, spectrum sensing is inherently challenging due to the dynamic and unpredictable nature of the radio environment. Factors such as noise, fading, and interference make reliable signal detection and classification a complex problem. Additionally, the need for real-time processing adds to the computational burden, requiring efficient and robust algorithms.

Objectives

- Students will develop advanced signal detection and classification algorithms using the RadioML.2018.01a dataset. This task requires them to preprocess the data, select and train appropriate machine learning models, and optimize these models for accurate and efficient performance in detecting and classifying various signal types.

- Through the competition, students will gain hands-on experience with real-world radio signal data, enhancing their technical expertise in signal processing and machine learning. They will learn to apply theoretical knowledge to practical challenges, improve their data analysis skills, and understand the intricacies of cognitive radio systems and their applications in communication technologies.
- By tackling complex problems presented in the competition, students will develop critical problem-solving skills and innovative thinking. They will learn to handle dynamic and noisy environments, refine their approach to algorithm development, and gain insights into advanced spectrum sensing techniques that are crucial for the future of wireless communications.

Dataset

The competition will use the RadioML.2018.01a dataset, a comprehensive collection of radio signal data widely used in research on automatic modulation classification and spectrum sensing. The dataset includes samples of various modulation types, such as BPSK, QPSK, 8PSK, AM-DSB, AM-SSB, CPFSK, PAM4, QAM16, QAM64, and WBFM. Each sample is represented as a complex baseband IQ (In-phase and Quadrature) signal. This dataset provides a rich resource for developing and testing spectrum sensing algorithms.

PROBLEM STATEMENT-4

RSNA 2024 Lumbar Spine Degenerative Classification Challenge

1. Introduction to Problem

The RSNA 2024 Lumbar Spine Degenerative Classification competition seeks to address the critical issue of accurately diagnosing lumbar spine degenerative diseases through advanced image classification techniques. Lumbar spine degeneration is a prevalent condition that affects a significant portion of the population, particularly the elderly. It can lead to chronic pain, reduced mobility, and decreased quality of life. Accurate and early diagnosis is essential for effectively treating and managing these conditions.

Traditional diagnostic methods often rely on radiologists' subjective interpretation of MRI images, which can be time-consuming and unpredictable. This competition challenges participants to develop machine learning models that can automate classification, ensuring more consistent and accurate diagnoses. By leveraging a large dataset of MRI images labeled with various degenerative changes, participants will create models capable of accurately identifying and classifying these changes.

The Radiological Society of North America (RSNA) hosts the competition on Kaggle, a platform renowned for fostering innovation in data science and machine learning. Participants can contribute to the medical community by developing solutions that can potentially enhance the efficiency and accuracy of radiological assessments. This initiative not only aims to improve patient outcomes but also to advance the field of medical imaging through the application of cutting-edge artificial intelligence techniques.

2. Dataset

The dataset provided for this competition consists of MRI images of the lumbar spine. Each image is labeled with various types of degenerative changes, such as disc space narrowing, spondylolisthesis, and spinal stenosis. The dataset is divided into training and test sets. The training set includes labeled images that participants will use to train their models, while the test set includes unlabeled images that will be used to evaluate the performance of the models. The dataset is large and diverse, ensuring that models trained on it can generalize well to different cases of lumbar spine degeneration.

PROBLEM STATEMENT-5

"Tiger Grand Challenge: Transforming Cancer Pathology with AI-Driven Diagnostics"

1. Introduction to the Problem

The Tiger Grand Challenge is an ambitious initiative designed to push the boundaries of what is possible in computational pathology. The central focus of the challenge is to address the growing need for more accurate, efficient, and scalable diagnostic tools in the field of cancer pathology. Cancer diagnosis traditionally relies on the expertise of pathologists who manually examine histopathology slides under a microscope. This process is time-consuming, subject to human error, and often limited by the availability of skilled professionals.

The challenge aims to harness the power of machine learning and artificial intelligence to revolutionize this diagnostic process. By developing advanced algorithms capable of analyzing histopathology images, participants in the Tiger Grand Challenge can contribute to a future where cancer diagnosis is faster, more accurate, and accessible to a broader range of healthcare providers.

Participants are tasked with creating solutions that can:

- ***Automate the Identification and Classification of Cancerous Tissues***: The algorithms should be able to detect and classify different types of cancerous tissues with high accuracy, reducing the burden on pathologists and improving diagnostic consistency.

- ***Enhance Diagnostic Precision and Speed***: By leveraging computational methods, the goal is to significantly reduce the time required for diagnosis, allowing for quicker treatment decisions and better patient outcomes.

- ***Improve Generalizability and Robustness***: Solutions should be robust enough to handle the variability in histopathology images, ensuring reliable performance across different patient populations and clinical settings.

This challenge is critical as it addresses several pressing issues in the medical field:

- ***Increasing Incidence of Cancer***: With the global rise in cancer cases, there is a dire need for efficient diagnostic tools to manage the growing workload on healthcare systems.

- ***Limited Access to Expert Pathologists***: In many parts of the world, there is a shortage of trained pathologists, leading to delays in diagnosis and treatment. Automated solutions can help bridge this gap.

- ***Human Error in Diagnosis***: Manual examination of histopathology slides is prone to human error. Advanced algorithms can assist in minimizing diagnostic errors and ensuring consistent results.

The Tiger Grand Challenge not only seeks to advance technological innovation but also aims to create real-world impact by improving cancer diagnosis and patient care. By participating in this challenge, developers, researchers, and data scientists have the opportunity to contribute to a cause that can save lives and shape the future of medical diagnostics.

2. Dataset

The dataset provided for the Tiger Grand Challenge is a comprehensive collection of histopathology images from diverse sources. These images are annotated by expert pathologists to ensure high-quality and accurate labels. The dataset includes a variety of cancer types and grades, capturing the heterogeneity and complexity of real-world pathological samples. It is divided into training, validation, and test sets to facilitate the development and evaluation of algorithms. Participants are encouraged to use the training set to develop their models, the validation set to tune hyperparameters, and the test set to evaluate the final performance of their solutions.