Exploring the Diagnostic Potential of LLMs in Schizophrenia Detection through EEG Analysis

1st Michele Guerra
Mosaic Research Center - DiBT
University of Molise
Pesche, Italy
michele.guerra@unimol.it
0009-0005-9990-234X

2nd Roberto Milanese
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
roberto.milanese@polito.it
0009-0009-8758-753X

3rd Michele Deodato
Division of Science
New York University
Abu Dhabi, United Arab Emirates
md5050@nyu.edu
0000-0002-2624-1430

4th Madalina G. Ciobanu
Department of Computer Science
University of Salerno
Fisciano(SA), Italy
mciobanu@unisa.it
0009-0006-0212-3911

5th Fausto Fasano
Mosaic Research Center - DiBT
University of Molise
Pesche, Italy
fausto.fasano@unimol.it
0000-0003-3736-6383

***Abstract*—Schizophrenia is a psychiatric disorder that presents significant diagnostic challenges due to its complex neurophysiological characteristics. This paper investigates the potential of Large Language Models (LLMs), such as OpenAI's GPT-4 and GPT-o1, in detecting schizophrenia through electroencephalography (EEG) analysis. Using the LMSU public ScZ EEG dataset, we conducted a series of experiments involving different types of input data, including raw EEG signals, frequency band summaries, and graphical representations of brain activity. Our findings demonstrate that LLMs can accurately classify schizophrenic and healthy individuals while offering interpretable, clinically relevant insights aligned with established EEG markers. By integrating these models into the diagnostic workflow, we explore the concept of Symbiotic AI, where LLMs act as cognitive collaborators, enhancing clinicians' ability to analyze complex data efficiently and transparently. This approach not only improves diagnostic accuracy but also facilitates real-time decision-making, paving the way for earlier and more precise detection of schizophrenia in clinical settings.**

***Index Terms*—Schizophrenia Detection, Large Language Models, EEG Analysis, Explainable AI, Clinical Decision Support**

I. INTRODUCTION

The integration of artificial intelligence (AI) in healthcare is transforming the way we approach complex medical diagnostics, offering new opportunities to enhance precision and efficiency in patient care [1]. One of the most challenging conditions to diagnose is schizophrenia (ScZ) [2] [3], a severe psychiatric disorder characterized by disruptions in thought processes, perceptions, and emotional responsiveness [4]. Despite its widespread impact, with over 24 million people affected worldwide[1], schizophrenia lacks definitive biomarkers [5], and its diagnosis often relies on subjective clinical assessments and neuroimaging, which may miss subtle neurophysiological cues.

Electroencephalography (EEG) has gained attention as a non-invasive tool capable of capturing brain activity patterns associated with schizophrenia. However, interpreting the intricate signals from multiple brain regions requires expert analysis and remains a complex task [6] [7] [8]. Recent advances in AI, particularly in large language models (LLMs), have shown potential beyond text-based tasks, extending into structured data interpretation across domains, including healthcare [9]. Although initial research has started to explore the application of LLMs to biomedical signal analysis, their potential in interpreting EEG data for psychiatric diagnoses, particularly schizophrenia, remains underexplored.

This study seeks to evaluate the capability of LLMs in the complex task of schizophrenia diagnosis by interpreting EEG data. Moving beyond traditional AI classification tasks, we assess whether LLMs can generate clinically relevant explanations and engage in high-level reasoning using multimodal data based on EEG inputs. Specifically, we leverage the publicly available ScZ EEG dataset collected at Lomonosov Moscow State University (LMSU) [10] [11], comprising EEG recordings from 84 subjects, including both schizophrenic and healthy controls. A balanced subset of 17 participants was chosen based on prior studies with strong validation accuracy in machine learning [12] [13], ensuring a robust foundation for our analysis.

We conducted a series of experiments using two state-of-the-art LLMs from OpenAI[2]: `GPT-4o`, which can process raw EEG files and graphical data, and `GPT-o1`, a more recent model with superior reasoning capabilities, but limited to text-based inputs. The experiments progressively increased in complexity, ranging from raw EEG data to graphical represen-

[1] https://www.who.int/news-room/fact-sheets/detail/schizophrenia

[2] https://openai.com/

tations of brain activity, enabling a comprehensive comparison of the models' strengths and limitations in clinical contexts.

Central to this research is the concept of Symbiotic AI (SAI), where AI acts as a cognitive collaborator, assisting healthcare professionals in interpreting complex diagnostic data. The integration of LLMs with EEG analysis has the potential to enhance clinicians' ability to detect subtle neurophysiological markers associated with schizophrenia, thereby facilitating earlier and more accurate diagnoses.

This paper presents a comprehensive evaluation of LLMs for EEG-based schizophrenia diagnosis, assessing both their diagnostic performance and their ability to contribute to a more transparent and explainable AI in healthcare. The findings provide insights into the development of AI-driven diagnostic tools that can enhance patient care by fostering a symbiotic relationship between clinicians and AI systems.

In summary, the contributions of this work are twofold:

- We provide an in-depth evaluation of LLMs in the interpretation of EEG data for schizophrenia diagnosis, highlighting their strengths in different input formats and experimental setups.
- We explore the potential of Symbiotic AI, showcasing how LLMs can support clinicians by providing interpretable, real-time insights that improve diagnostic accuracy and efficiency.

II. Related Work

A review of the literature reveals several studies that have applied AI techniques to the assisted analysis of EEG signals. A survey by Chen et al. [14] presents a number of interesting approaches that have been applied to both diagnosis and monitoring of neurological disorders. A recent review by Rahul et al. [15] focuses specifically on the automated classification of ScZ based on EEG recordings. Their analysis covers 40 research papers published between 2013 and 2023. Of these, some use Machine Learning (ML) models, including Support Vector Machines, Decision Trees, and Random Forests. The rest use Deep Learning (DL) models, including Convolutional Neural Networks, which are commonly used in image classification, and Recurrent Neural Networks, which excel at processing sequential data. Comparison of the two approaches shows that DL-based solutions achieve better results in terms of classification accuracy, but neural networks require significantly more computational resources to train. Another problem highlighted is that datasets are often too small and contain unbalanced data, with negative samples often outnumbering positive ones. Also, as Amrani et al. [16] point out, the black-box nature of DL models makes it difficult to understand how they make their decisions, raising trust issues within the medical community.

For these reasons, research interest has recently shifted to LLMs. A survey by Wang et al. [17] presents the latest developments and applications of automated EEG data analysis. A number of the different approaches presented use general-purpose LLMs, although in most cases, they focus on generic spatiotemporal series and not specifically on EEG tracings. An exception is the work of Kim et al. [18], which explores the ability of GPT-3 to detect generic anomalies in EEG recordings. A fine-tuned version of the model is found to be competitive with other DL-based solutions in terms of performance. However, in the zero-shot context, the accuracy falls even below that of straightforward ML-based approaches. In addition, a significant limitation is that the EEG signals are not used directly in the evaluation, but a set of manually extracted quantitative features. Conversely, one of the advantages of using DL and LLM models is that no feature extraction is required, allowing for a faster and more comprehensive analysis of the collected data. This is exactly the strategy followed by Cui et al. [19] who trains the GPT transformer to automatically classify EEG chunks in the context of Brain-Computer Interaction. Lastly, Hu et al. [20] investigate the use of multi-modal LLMs in two different tasks, the recognition of human emotions and the diagnosis of depressive disorders. In addition to EEG data, they use audio recordings for the former and facial expression images for the latter. Their study shows that classifications based on multi-modal data significantly outperform those based on single-modal data.

At present, our work is the first to evaluate the performance of automatic classification of ScZ from EEG traces using LLMs.

III. STUDY DESIGN

The primary objective of this study is to evaluate the potential of LLMs to interpret and classify EEG data for the diagnosis of ScZ, going beyond simple classification to assess their capacity for generating clinically relevant explanations. To achieve this, we designed four distinct experiments, each involving progressively more complex input formats and prompts to explore different aspects of LLMs' interpretive abilities.

A. Dataset

The EEG dataset used in this study was collected by Lomonosov Moscow State University (LMSU) and includes recordings from 84 subjects, of which 45 are diagnosed with ScZ and 39 are healthy controls (HC). The EEG data were collected in a resting eye-closed state for 60 seconds across 16 channels: F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2, with a sample rate of 128 Hz. All ScZ patients were diagnosed at the Mental Health Research Center (MHRC) following the ICD-10 classification for schizophrenia (F20, F21, F25).

For this study, we selected a balanced subset of 17 subjects from a prior work [12], [13]. The 'group 3' designation originates from these earlier studies where machine learning models demonstrated the highest validation accuracy when applied to this specific subset. By using this group, we ensured a robust and reliable dataset, as it represents a well-validated and challenging sample set, making it ideal for testing the capabilities of large language models in schizophrenia detection.

B. Large Language Models

Two versions of LLMs developed by OpenAI were used in this study: `GPT-4o` and `GPT-o1`. These models differ significantly in their capabilities, particularly in handling data formats and reasoning complexity:

- **GPT-4o:** A robust LLM capable of processing both textual data and file inputs, such as the raw EEG data provided in the .eea format. This model was used in experiments involving the direct analysis of EEG recordings and graphical representations, allowing us to test its ability to process multi-modal biomedical data.
- **GPT-o1:** The latest iteration of OpenAI's language models, released in September 2024, which has demonstrated superior performance in various STEM-related benchmarks, including physics, chemistry, and biology. `GPT-o1` surpasses its predecessors in tasks that require deep reasoning, due to its capacity to engage in step-by-step problem-solving processes. However, it cannot directly process file inputs, limiting its role to text-based data analysis. We used this model in experiments where the EEG data was simplified into frequency band information, testing its reasoning abilities with reduced and structured inputs.

The choice of these models is motivated by their distinct abilities: while `GPT-4o` allows for direct interaction with complex biomedical data, `GPT-o1`'s strength in logical reasoning and its real-time evaluation of each step of a task offers a unique perspective on its application to the medical diagnostic process.

C. Input Types and Data Formats

To evaluate the models across different conditions, we used various types of input data:

- **Raw EEG data:** For the experiments requiring direct file input, we used EEG files in the .eea format. This format preserves the full complexity of the 16-channel EEG recordings, allowing us to test the LLMs' capability to process and interpret raw neurophysiological signals.
- **Frequency band data:** In some experiments, we extracted and provided only the frequency band data from the EEG signals (e.g., delta, theta, alpha, beta, gamma bands). This reduction in data complexity allowed us to assess the model's ability to interpret simplified, summary-level information.
- **Graphical representations:** Using MNE-Python[3], a powerful tool for processing and visualizing EEG/MEG data, we generated scalp maps and EEG signal traces. These graphical inputs represent brain activity's spatial and temporal characteristics, simulating the type of data clinicians observe in real-time during EEG monitoring. These inputs were used to test whether `GPT-4o` could derive meaningful insights from visual patterns typically used in clinical practice.

[3]https://mne.tools/stable/index.html

D. Experimental Procedure

Each experiment explored different input types and varying levels of guidance, progressively testing the models under various conditions. Across all experiments, the output produced by the models followed the same structure, as reported below.

Output format

Subject: ⟨*subject name*⟩The identifier of the subject.
Classification: ⟨*Schizophrenic/Healthy*⟩A binary classification as either "Schizophrenic" or "Healthy."
Reasoning: A natural language explanation of the classification, including references to relevant EEG patterns, features, or abnormalities that led to the decision.
Frequency Band Analysis: For experiments involving only frequency band data, the model was required to analyze specific frequency bands (e.g., alpha, delta, theta) and comment on any significant findings, such as changes in power or asymmetries between hemispheres.
Image Analysis: For experiments involving graphical EEG representations, the model was tasked with interpreting scalp maps and EEG signal traces, identifying any abnormal spatial or temporal patterns.

Experiment 1. In the *first experiment*, the objective was to establish a baseline for `GPT-4o`'s ability to classify subjects based on raw EEG data without additional guidance. The raw EEG data from the 17 subjects was provided in .eea format with no pre-processing, allowing the model to interpret the data directly. This *zero-shot* approach tested the model's internal understanding of EEG signals and its ability to derive clinical insights from unstructured biomedical data.

Prompt Experiment 1

For each subject, you are required to: Clearly determine whether the subject is affected by schizophrenia, classifying them as either "Schizophrenic" or "Healthy." Provide a detailed explanation of your assessment, specifically citing the features present in the data that support your conclusion.
For each subject, the following are available: Detailed EEG Data Files (.eea) acquired through electrodes on the subject. Each file represents a 60-second EEG recording during an eyes-closed resting state. The recordings were made at a sampling rate of 128 Hz and include 16 EEG channels corresponding to electrode positions. Each number in the file represents an EEG amplitude (in microvolts) for a single sample. There are 7,680 samples per channel (60 seconds at a sampling rate of 128 Hz), and the channels are F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2. All schizophrenic patients were diagnosed at

the Mental Health Research Center (MHRC) according to the ICD-10 diagnostic criteria for schizophrenia (ScZ), specifically codes F20, F21, and F25. During the MHRC examination, the patients had not received any pharmacological treatment and are adolescents.

This experiment tested the model's ability to handle raw biomedical signals, a notoriously complex task due to the variability, noise, and lack of clear patterns in EEG data, especially without prior domain-specific tuning.

Experiment 2. In the *second experiment*, we introduced more specific domain knowledge into the prompt, directing the model to focus on crucial EEG markers associated with schizophrenia. This was done by referencing scientific literature on EEG abnormalities in schizophrenia, such as characteristic deviations in delta, theta, and gamma frequency bands. By providing more explicit guidance, we tested whether the additional context improved the accuracy and quality of the model's reasoning.

Prompt Experiment 2

Role: You are an experienced clinical neurophysiologist specializing in EEG data analysis and the diagnosis of schizophrenia. In this task, you will analyze EEG data, with specific guidance on known schizophrenia-related markers. I am providing you with raw EEG data and instructions based on well-established scientific findings regarding EEG abnormalities in schizophrenic patients.
For each subject, you are required to: Clearly determine whether the subject is affected by schizophrenia, classifying them as either "Schizophrenic" or "Healthy." Provide a detailed explanation of your assessment, based solely on the data provided.
Specific Instructions:

- Carefully analyze the EEG data, with attention to the following established diagnostic EEG criteria for schizophrenia:
  - **Increased slow-wave activity:** Notable increases in Delta and Theta bands, especially in the frontal regions (F3, F4), are strongly associated with schizophrenia.
  - **Reduction in Alpha power:** A significant reduction in Alpha activity in the posterior regions (O1, O2) is a typical marker in schizophrenic subjects.
  - **Gamma band abnormalities:** Pay attention to any significant changes in Gamma band activity, as deviations in this band have been linked to cognitive dysfunction in schizophrenia.
  - **Hemispheric asymmetry:** Look for notable asymmetries in the EEG signals between the right and left hemispheres, as they may indicate disrupted neural coordination, a characteristic observed in schizophrenia.
- Highlight specific patterns or anomalies in the data that support your classification.
- Classify each subject as either "Schizophrenic" or "Non-Schizophrenic," providing clear evidence for your decision, and avoid unsupported conclusions.
- Ensure that your explanation is precise, citing relevant EEG markers and abnormalities, and use appropriate clinical terminology.

Subject Data: For each subject, the data will include raw EEG signals, sampled at a rate of 128 Hz, from 16 channels (F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2). ⟨*Raw EEG data*⟩
Important Note: Base your analysis on the known EEG criteria for schizophrenia provided. Do not classify all subjects as schizophrenic by default—diagnose only when supported by clear evidence in the data. Your conclusions should be well-motivated, unambiguous, and grounded in clinical knowledge.

Experiment 3. `GPT-o1` was tested in the *third experiment* using only frequency band data extracted from the EEG recordings. The prompt instructed the model to classify subjects based on the provided bands and explain its reasoning. This experiment aimed to evaluate how well `GPT-o1` could operate on simplified, pre-processed data and whether its advanced reasoning capabilities could compensate for the lack of raw signal data.

Prompt Experiment 3

Role: You are an experienced clinical neurophysiologist specializing in EEG data analysis and the diagnosis of schizophrenia. I am providing you with the average power values for the frequency bands (Delta, Theta, Alpha, Beta, Gamma) for each EEG channel of several subjects.
For each subject, you are required to: [Same as Experiment 2]
Specific Instructions:

- Carefully analyze the frequency band data for each EEG channel.
  [Same as Experiment 2]

Subject Data: For each subject, the data will include the average power values for the Delta, Theta, Alpha, Beta, and Gamma frequency bands for all EEG channels.
⟨*Frequency band data*⟩
Important Note: [Same as Experiment 2]

Experiment 4.

In the *fourth experiment*, `GPT-4o` was given graphical representations of the EEG data, such as scalp maps and traces, instead of raw files or frequency data. The prompt asked the model to classify the subjects based on visual patterns in the EEG data, mirroring the diagnostic process used by human experts who visually inspect EEG signals in clinical settings. This experiment tested the model's capacity for pattern recognition in visual formats.

Prompt Experiment 4

Role: You are an experienced clinical neurophysiologist specializing in EEG data analysis and the diagnosis of neurological conditions, including schizophrenia. In this task, you will analyze visual representations of EEG data, including signal traces and scalp maps, to determine whether the subject is affected by schizophrenia.

For each subject, you are required to:

Clearly determine whether the subject is affected by schizophrenia, classifying them as either "Schizophrenic" or "Healthy." Provide a detailed explanation of your assessment, based solely on the data provided, specifically citing the features present in the images that support your conclusion.

Specific Instructions:

- **Careful Image Analysis:** Thoroughly examine both the brain scalp maps and EEG signal traces provided for each subject.
- **Diagnostic EEG criteria for schizophrenia in visual data:**
  - **Scalp maps:** Look for significant reductions in Alpha power in the occipital regions (O1, O2), and increases in Delta and Theta activity in the frontal regions (F3, F4).
  - **EEG signal traces:** Focus on the amplitude and frequency of the oscillations. Slower waveforms (Delta and Theta) should be more prominent in the frontal lobes (F3, F4), and there should be reduced activity in the Alpha band, particularly in the occipital areas (O1, O2).
  - **Hemispheric asymmetries:** Pay attention to any visible asymmetries in the activity between the left and right hemispheres by comparing electrode pairs (e.g., F3 vs F4, O1 vs O2).
  - **Coherence and synchronization:** Look for disruptions in the coherence or synchronization of brain activity between different regions, especially between frontal and posterior lobes
- Highlight any specific anomalies or patterns in the scalp maps or signal traces that support your conclusion.
- Classify each subject as either "Schizophrenic" or "Non-Schizophrenic," avoiding unsupported generalizations based on the visual data.
- Ensure clarity and precision in your explanations, using appropriate technical terminology.

Images Provided:

- **Brain scalp maps:** Visual representations of the spatial distribution of electrical activity across the brain, organized by frequency band (Delta, Theta, Alpha, Beta, Gamma). The scalp mapscalp mapsscalp mapss highlight regions of increased or decreased activity, especially in the frontal and occipital lobes, which are most relevant for schizophrenia diagnosis. Focus on the patterns in these areas to identify any abnormalities.
- **EEG Signal Traces for Each Channel:** Temporal plots showing the electrical activity for each of the 16 EEG channels over the 60-second recording period.

Important Note: [Same as Experiment 2]

E. Evaluation Approach

1) Classification Metrics: We evaluated the models using four key metrics: accuracy, recall, precision, and false positive rate (FPR).

Accuracy ireflects the proportion of correct classifications (schizophrenic and healthy) out of the total predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Although accuracy gives a general idea of the model's performance, it is essential to consider other metrics, especially in cases of class imbalance or where false positives and false negatives carry different weights, as is the case in medical diagnoses.

Recall, measures the model's ability to correctly identify schizophrenic subjects:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

High recall is crucial in medical diagnoses to minimize missed cases.

Precision assesses how many subjects classified as schizophrenic were truly schizophrenic:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Balancing precision and recall helps reduce misdiagnoses, particularly false positives.

Finally, the **False Positive Rate (FPR)** measures the proportion of healthy individuals misclassified as schizophrenic:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Keeping FPR low is essential to avoid unnecessary treatments for healthy individuals. Together, these metrics provide a comprehensive view of model performance, ensuring accuracy, sensitivity (recall), and minimizing misclassifications, which is critical in clinical settings.

2) Explanation Quality Assessment: To assess the quality of the model's explanations, we utilize two complementary methods: comparison with clinical literature and cross-validation across experiments. These strategies enable us to evaluate both the clinical relevance of the model's reasoning and its consistency across different data representations.

Comparison with clinical literature. The first approach directly compares the model's explanations with well-established clinical markers for schizophrenia, frequently reported in the EEG literature. Key features, such as reduced alpha power in posterior regions (especially O1 and O2), increased delta and theta activity in frontal areas (F3 and F4), and hemispheric asymmetry, serve as reference points. We assess whether the model's explanations align with these markers for each subject. This evaluation emphasizes the clinical relevance of the explanations by focusing on whether the model reliably identifies these diagnostic patterns and provides clear justification for the classifications. Explanations that closely match these known markers are deemed reliable and clinically meaningful.

Cross-validation between experiments. The second method involves cross-validating the explanations across multiple experiments, each using different input formats. This process assesses whether the model produces consistent explanations across these varying representations for subjects classified as schizophrenic in at least two experiments. Consistency is key—if the model highlights increased delta and theta activity in frontal regions in one format, we expect to see this same marker appear in the others, ensuring that the model is not arbitrarily selecting features based on the data format but is identifying stable, diagnostic markers. We focus on cases where the model correctly classified subjects, excluding misclassification, to ensure a more accurate evaluation of the model's reasoning. When discrepancies in the explanations arise (e.g., different markers being highlighted for the same subject across experiments), this may indicate potential weaknesses in the model's ability to interpret EEG data consistently. However, consistent explanations reinforce the stability and robustness of the model's interpretative capabilities. By combining these two methods, we strengthen the overall trustworthiness of the model's reasoning across multiple experimental contexts.

IV. Results

This section presents the outcomes of the experiments conducted to evaluate the performance of large LLMs in diagnosing ScZ from EEG data. Two main aspects were analyzed: the classification results in terms of TP, TN, FP, and FN, as well as the overall model performance through Accuracy, Recall, Precision, and FPR. The first subsection focuses on the quantitative results of the classifications, while the second subsection assesses the quality and consistency of the model's explanations.

A. Classification Results

TABLE I
CLASSIFICATION RESULTS AND PERFORMANCE METRICS FOR EXPERIMENTS (EXP) 1–4, SHOWING TRUE POSITIVES (TP), TRUE NEGATIVES (TN), FALSE POSITIVES (FP), FALSE NEGATIVES (FN), ACCURACY (ACC), RECALL (REC), PRECISION (PREC), AND FALSE POSITIVE RATE (FPR) FOR EACH EXPERIMENT.

Exp	TP	TN	FP	FN	ACC	REC	PREC	FPR
1	4	2	6	5	0.35	0.44	0.4	0.75
2	9	5	3	0	0.823	1	0.75	0.375
3	8	8	0	1	0.941	0.888	1	0
4	9	4	4	0	0.764	0.69	0.69	0.5

Across the four experiments, different input formats were used to assess how well the models could differentiate between ScZ and HC subjects. The results, including TP, TN, FP, and FN for each experiment, are shown in Table I, as well as the overall performance metrics (Accuracy, Recall, Precision, and FPR).

Experiment 1, which served as a baseline, involved providing GPT-4o with raw EEG data using a basic prompt with no domain-specific guidance. The model's performance was relatively poor, as shown in Table I, with 4 correct classifications of ScZ subjects (TP) and 2 correct classifications of HC subjects (TN), but it misclassified 6 HC subjects (FP), resulting in an FPR of 0.75 and a low accuracy of 0.35. These results highlight the model's difficulty in interpreting raw EEG data without any structured guidance, leading to frequent over-diagnosis and a recall of only 0.44, meaning it correctly identified less than half of the ScZ cases.

Experiment 2 demonstrated a significant improvement when domain-specific EEG markers were introduced to GPT-4o alongside the raw EEG data. The model correctly classified all ScZ subjects, achieving a perfect recall of 1.00. Still, it misclassified 3 HC subjects (FP = 3), resulting in a slight reduction in precision (0.75) and an FPR of 0.38 (see Table I). The improved accuracy of 0.82 underscores the model's enhanced ability to interpret EEG data when guided by specific markers associated with ScZ. However, the false positives indicate that the model remained sensitive to benign EEG variations in HC subjects, which it sometimes mistook for pathological features.

In *Experiment 3*, the model GPT-o1 was introduced, and the input data consisted solely of frequency band summaries derived from the EEG recordings. GPT-o1, being a more advanced model, demonstrated superior performance through its ability to apply **chain of thought** (CoT) reasoning. As highlighted in benchmarks by OpenAI, GPT-o1's CoT reasoning enables it to analyze input data more intelligently and specifically, focusing on essential patterns while avoiding extraneous noise. This resulted in a higher performance, with the model achieving an accuracy of 0.94 and a recall of 0.89 (see Table I). It successfully classified all but one ScZ subject (FN = 1) and made no false positive classifications (FP =

0). The perfect precision of 1.00 reflects the model's focused analysis, mainly when provided with simplified input like frequency bands. The absence of FP in this experiment shows how GPT-o1's reasoning capabilities helped avoid overfitting irrelevant data features.

Experiment 4 used graphical representations of the EEG data, such as scalp maps and signal traces, to simulate real-world clinical EEG interpretations. Despite a lower accuracy of 0.76 compared to Experiment 3, the model maintained a perfect recall of 1.00 (see Table I), correctly identifying all ScZ subjects. However, 4 HC subjects were misclassified (FP = 4), leading to an FPR of 0.50 and a reduced precision of 0.69. This result suggests that while graphical data allowed the model to capture ScZ patterns effectively, it also introduced visual complexity that the model struggled to navigate, leading to more false positives compared to Experiment 3.

Certain subjects were repeatedly misclassified, particularly in Experiments 2 and 4, indicating that their EEG data consistently confused the models. This suggests that these HC subjects may exhibit non-pathological anomalies that resemble ScZ markers, particularly when the model analyzes specific patterns or interprets complex visual data.

In summary, the results presented in Table I demonstrate the crucial role of input data format and model sophistication in determining classification performance. GPT-4o performed well when provided with detailed prompts and specific guidance, effectively analyzing complex data. However, when tasked with interpreting raw EEG data without any prior information or context, its performance significantly declined. In contrast, GPT-o1, with its advanced CoT reasoning, excelled in interpreting frequency band data, offering more precise and targeted analysis.

B. Explanation Quality Assessment

In this section, we present the results of the cross-validation process and the comparison of the model's explanations with established clinical markers for schizophrenia. Despite differences in accuracy across experiments, the model's explanations generally aligned with known diagnostic markers of schizophrenia. The following markers were consistently identified across experiments:

- **Increased delta and theta activity in the frontal regions:** This was a stable finding across all subjects and all experiments, reflecting the model's capacity to detect slow-wave activity typical of cognitive dysfunction in schizophrenia.
- **Reduced alpha power in posterior regions:** In most cases, the model reliably identified a reduction in alpha power in the occipital and posterior regions, particularly in *Experiments 2* and *4*. This reduction is a well-documented marker of schizophrenia, reinforcing the clinical validity of the model's interpretations.
- **Hemispheric asymmetry:** In 2 subjects, the model detected hemispheric asymmetry, particularly in the frontal regions, indicating disrupted inter-hemispheric communication—a feature often observed in schizophrenia. This detail was more apparent in scalp map-based explanations and added an extra layer of diagnostic insight.

Experiment 1 exhibited lower classification accuracy, which affected the specificity and depth of the model's explanations. While the model correctly identified subjects as schizophrenic in several cases, its explanations were generally generic and focused on broad EEG features, such as increased delta and theta activity in the frontal regions and reduced alpha power in posterior regions. Although these markers are clinically relevant, the explanations lacked the case-specific insights seen in other experiments. This suggests that, in Experiment 1, the model was relying on basic diagnostic indicators, possibly due to the simplicity of the input prompt, which did not provide clinical guidelines or structured steps. As a result, the model's reasoning remained more general and less detailed.

Similar to Experiment 1, *Experiment 4* also exhibited lower accuracy compared to other experiments. However, in *Experiment 4*, the model's explanations were notably more detailed and clinically relevant. Despite the lower accuracy, the model demonstrated a higher level of specificity, particularly in identifying hemispheric asymmetries and describing the distribution of slow-wave activity (delta and theta) across different brain regions. The use of scalp map representations allowed the model to capture more granular patterns, such as frontal asymmetry and refined distinctions between regions of elevated slow-wave activity. This suggests that even when overall classification performance is reduced, the model is still capable of providing insightful and clinically meaningful interpretations when given spatially rich data formats like scalp maps. This highlights the potential for certain data representations to enhance the model's explanatory power, even in scenarios where classification accuracy may otherwise be suboptimal.

An inconsistency was observed in *Experiment 3*, where the model detected elevated alpha power in posterior regions for some subjects, in contrast to the reduced alpha power found in other experiments (*Experiments 2* and *4*). While this discrepancy might reflect variability in the model's sensitivity to different input formats, it is likely due to the frequency band summaries used in Experiment 3, which may have caused the model to overemphasize certain aspects of the alpha band that were less pronounced in other formats. Nevertheless, the broader trend of reduced alpha power in posterior regions, observed consistently in most experiments, confirms this marker as a reliable feature of schizophrenia. The anomaly in Experiment 3 does not significantly undermine the model's overall consistency but underscores the need to consider how different data representations may affect the model's sensitivity to specific features.

V. Conclusions

This study explores the potential of LLMs to assist clinical decision-makers in diagnosing schizophrenia from EEG data and shows how recent models, particularly GPT-o1, excel at this task. However, this model cannot directly process EEG signals but requires structured inputs such as frequency band

summaries. In contrast, the other model considered, GPT-4o, can automatically extract this structured data from raw data without any training, highlighting the versatility of multimodal LLMs in handling different data formats. One of the key contributions of this approach is the promising application of Symbiotic AI, where the model supports healthcare professionals by providing detailed explanations. While LLMs remain inherently black-box systems, their ability to assist in clinical decision-making highlights their potential. However, the challenge of ensuring full interpretability remains open and requires additional tools and frameworks. This can be particularly useful for general practitioners, allowing them to make an initial assessment of suspected cases and then possibly initiate more specialized diagnostic and therapeutic pathways with experts in the field. In this way, even patients who do not usually have access to specialized centers are directed to dedicated channels for quality care. In addition, specialists would benefit from faster and more accurate analysis of EEG recordings. Unlike custom-built models that require extensive training and signal pre-processing, LLMs can save significant time and are highly adaptable to real-time clinical needs, where early detection of neurological conditions such as ScZ is critical. The ability to validate the model's reasoning behind each conclusion through natural language explanations and conversations leads to a more informed and transparent decision-making process. This component of *explainability* is essential because it always leaves the final say to the professional, supporting rather than replacing their decision.

In conclusion, our research highlights the considerable promise of LLMs such as GPT-o1 to aid in EEG-based diagnostics. While further improvements are needed to balance sensitivity and specificity, the potential for these models to serve as practical tools in a symbiotic relationship with clinicians is clear. The future of neurodiagnostic could benefit significantly from the integration of AI technologies that improve not only diagnostic accuracy but also provide actionable insights in real-time, transforming the way we approach complex psychiatric conditions such as ScZ.

ACKNOWLEDGMENT

This publication is part of the project PNRR-NGEU which has received funding from the MUR – DM 118/2023, and part of the UNIMOL DiBT Startup Project INTELLIGENZA_ARTIFICIALE - Dynamic Permission Management in Android Using Artificial Intelligence Models

REFERENCES

[1] S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, A. Aldairem, M. Alrashed, K. Bin Saleh, H. A. Badreldin, M. S. Al Yami, S. Al Harbi, and A. M. Albekairy, "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," *BMC Medical Education*, vol. 23, no. 1, p. 689, Sep 2023. [Online]. Available: https://doi.org/10.1186/s12909-023-04698-z

[2] P. Ruiz-Castañeda, E. Santiago Molina, H. Aguirre Loaiza, and M. T. Daza González, "Positive symptoms of schizophrenia and their relationship with cognitive and emotional executive functions," *Cognitive Research: Principles and Implications*, vol. 7, no. 1, p. 78, Aug 2022. [Online]. Available: https://doi.org/10.1186/s41235-022-00428-z

[3] L. Orsolini, S. Pompili, and U. Volpe, "Schizophrenia: A narrative review of etiopathogenetic, diagnostic and treatment aspects," *J Clin Med*, vol. 11, no. 17, Aug. 2022.

[4] M. M. Picchioni and R. M. Murray, "Schizophrenia," *BMJ*, vol. 335, no. 7610, pp. 91–95, Jul. 2007.

[5] B. Galińska-Skok and N. Waszkiewicz, "Markers of Schizophrenia-A critical narrative update," *J Clin Med*, vol. 11, no. 14, Jul. 2022.

[6] A. Perrottelli, G. M. Giordano, F. Brando, L. Giuliani, and A. Mucci, "Eeg-based measures in at-risk mental state and early stages of schizophrenia: A systematic review," *Frontiers in Psychiatry*, vol. 12, 2021. [Online]. Available: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.653642

[7] G. Marsicano, C. Bertini, and L. Ronconi, "Decoding cognition in neurodevelopmental, psychiatric and neurological conditions with multivariate pattern analysis of eeg data," *Neuroscience & Biobehavioral Reviews*, vol. 164, p. 105795, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763424002641

[8] D. Li, X. Zhang, Y. Kong, W. Yin, K. Jiang, X. Guo, X. Dong, L. Fu, G. Zhao, H. Gao, J. Li, J. Zhai, Z. Su, Y. Song, and M. Chen, "Lack of neural load modulation explains attention and working memory deficits in first-episode schizophrenia," *Clinical Neurophysiology*, vol. 136, pp. 206–218, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245722001584

[9] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nature medicine*, vol. 29, no. 8, pp. 1930–1940, 2023.

[10] N. Gorbachevskaya and S. Borisov, "Eeg data of healthy adolescents and adolescents with symptoms of schizophrenia," *Avaliable online at: http://brain. bio. msu. ru/eeg_schizophrenia. htm*, 2002.

[11] S. Borisov, A. Y. Kaplan, N. Gorbachevskaya, and I. Kozlova, "Analysis of eeg structural synchrony in adolescents with schizophrenic disorders," *Human Physiology*, vol. 31, pp. 255–261, 2005.

[12] M. Shen, P. Wen, B. Song, and Y. Li, "Automatic identification of schizophrenia based on eeg signals using dynamic functional connectivity analysis and 3d convolutional neural network," *Computers in Biology and Medicine*, vol. 160, p. 107022, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523004870

[13] ——, "3d convolutional neural network for schizophrenia detection using as eeg-based functional brain network," *Biomedical Signal Processing and Control*, vol. 89, p. 105815, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S174680942301248X

[14] X. Chen, X. Tao, F. L. Wang, and H. Xie, "Global research on artificial intelligence-enhanced human electroencephalogram analysis," *Neural Computing and Applications*, vol. 34, no. 14, pp. 11 295–11 333, Jul. 2022.

[15] J. Rahul, D. Sharma, L. D. Sharma, U. Nanda, and A. K. Sarkar, "A systematic review of eeg based automated schizophrenia classification through machine learning and deep learning," *Frontiers in Human Neuroscience*, vol. 18, 2024. [Online]. Available: https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2024.1347082

[16] G. Amrani, A. Adadi, M. Berrada, Z. Souirti, and S. Boujraf, "Eeg signal analysis using deep learning: A systematic literature review," in *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*, 2021, pp. 1–8.

[17] P. Wang, H. Zheng, S. Dai, Y. Wang, X. Gu, Y. Wu, and X. Wang, "A survey of spatio-temporal eeg data analysis: from models to applications," 2024. [Online]. Available: https://arxiv.org/abs/2410.08224

[18] J. W. Kim, A. Alaa, and D. Bernardo, "Eeg-gpt: Exploring capabilities of large language models for eeg classification and interpretation," 2024. [Online]. Available: https://arxiv.org/abs/2401.18006

[19] W. Cui, W. Jeong, P. Thölke, T. Medani, K. Jerbi, A. A. Joshi, and R. M. Leahy, "Neuro-gpt: Towards a foundation model for eeg," 2024. [Online]. Available: https://arxiv.org/abs/2311.03764

[20] Y. Hu, S. Zhang, T. Dang, H. Jia, F. D. Salim, W. Hu, and A. J. Quigley, "Exploring large-scale language models to evaluate eeg-based multimodal data for mental health," in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 412–417. [Online]. Available: https://doi.org/10.1145/3675094.3678494