www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 96-111. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2025.v16i4.1228

Multiclass Classification of Neurological and Psychiatric Conditions Using Synthetic Neuroinformatics Biomarkers and EEG Band Simulation

Jorge A. Ruiz-Vanoye, Ocotlán Diaz-Parra, Francisco R. Trejo-Macotela, Eric Simancas-Acevedo, Juan M. Xicoténcatl-Pérez, Marco A. Vera-Jiménez, José M. Liceaga-Ortiz-De-La-Peña

Universidad Politécnica de Pachuca, Carretera Pachuca-Cd. Sahagún Km 20, Ex-Hacienda de Santa Bárbara, Zempoala, HGO 43830, México.
E-mail: jorgeruiz@upp.edu.mx

Abstract. This study presents a machine learning framework for the multiclass classification of neurological and psychiatric conditions based on synthetic neuroinformatics biomarkers and EEG spectral band simulations. A synthetic training dataset was generated by modelling temporal cognitive-motor features, including visual delay (τv), motor gain (K), expectancy weight (β), predictive memory capacity, and canonical EEG spectral powers (Delta, Theta, Alpha, Beta, Gamma). To evaluate model performance on data derived from real signals, a second dataset was created using features extracted from 50 EEG recordings available in the PhysioNet EEG Motor Movement/Imagery Database. These features provided a hybrid evaluation environment combining the controlled conditions of synthetic data with the complexity of real EEG measurements. The classification task was addressed using an XGBoost model, optimised through an exhaustive grid search procedure exploring 24 hyperparameter combinations, each evaluated via five-fold cross-validation (120 individual fits). The best-performing configuration employed a learning rate of 0.1, a maximum tree depth of 3, 200 boosting iterations and a subsampling ratio of 0.8. When tested on an independent dataset, the model achieved an accuracy of 97.8% and an F1-score of 0.978, demonstrating excellent predictive performance across all clinical classes. The resulting confusion matrix confirmed high classification accuracy for neurotypical controls, attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), dementia, depression, generalised anxiety disorder (GAD), Parkinson's disease, psychosis and Tourette's syndrome, with only minor misclassifications observed among phenotypically similar conditions. These results highlight the potential of combining synthetic feature modelling with machine learning for differential diagnosis and clinical decision support in neuropsychiatric research.

Keywords: Synthetic Neuroinformatics Biomarkers, Synthetic dataset generation, EEG signal processing, Multiclass classification, Neurological and psychiatric disorders.

Article Info
Received May 09, 2025
Accepted July 2, 2025

1. Introduction

Neuroinformatics is an interdisciplinary field that facilitates the integration and analysis of large volumes of brain data, contributing to the use of biomarkers for the detection and monitoring of neurodegenerative diseases such as Alzheimer's. This approach enables more precise and earlier diagnosis, transcending traditional methods that are often limited by their lack of specificity.

The identification of biomarkers enables a more precise characterisation of the preclinical phases of diseases such as Alzheimer's. According to Arriagada and Villalobos, both the IWG and the NIA-AA groups emphasise the importance of using biomarkers to diagnose the disease before the clinical manifestations of dementia appear, representing a significant advance in

medicine (Arriagada & Villalobos, 2022). The clinical application of measurable biological features, such as cerebrospinal fluid tau levels or positron emission tomography (PET) imaging, allows more effective monitoring and earlier therapeutic intervention (Dorman et al., 2022).

Challenges persist in the use of biomarkers, primarily due to the complexity of neurodegenerative diseases, which often present significant phenotypic variations and prolonged preclinical states (García-Ribas et al., 2023). This uniqueness requires multidisciplinary integration between neuroscientists, clinicians and informatics specialists to develop tools that combine biomedical and clinical data, a key endeavour in neuroinformatics (García-Ribas et al., 2023). The ultimate aim is to enhance the understanding of the pathogenesis of these diseases and optimise the effectiveness of available treatments.

Collectively, the neurophysiological indicators (Visual delay (τ_v), motor gain (K), expectation weight (β), and EEG oscillations) contribute to a more comprehensive and multidimensional understanding of human behaviour and neurological dysfunction. Their integration into clinical and research workflows is critical for advancing insights into disease pathophysiology and improving diagnostic and therapeutic strategies within neuroscience and psychology.

Visual delay (τ_v), motor gain (K), expectation weight (β), and EEG oscillations constitute key neurophysiological indicators offering an integrated perspective on nervous system function and its implications for cognition and behaviour. Collectively, these parameters provide valuable information for elucidating the underlying mechanisms of diverse neurological and psychological conditions.

- Visual delay (τ_v) serves as a critical marker of perceptual processing efficiency, reflecting the latency associated with the transmission and interpretation of visual stimuli within the brain. This parameter is particularly relevant in studies of visual perception and attention, as deviations in τ_v may signify dysfunction within the visual system, often associated with neurological and psychiatric disorders. Visual delay represents the time difference between stimulus onset and the subject's motor response. It is obtained by presenting sudden visual changes (flashes or target shifts) and recording reaction times via response buttons or motion sensors. Normal values typically range between 80 and 120 ms, and deviations can reveal delayed or anticipatory processing.
- Motor gain (K) quantifies the efficiency with which the motor system translates perception into coordinated action. This metric is particularly pertinent in the context of paediatric motor development and the assessment of movement disorders. Empirical studies demonstrate that reduced K correlates with deficiencies in executing complex motor tasks, with potential implications for both learning and academic achievement (Montero et al., 2018). Consequently, motor gain reflects not only physical capability but also cognitive processes relevant to educational performance. K quantifies the relationship between an external stimulus and the motor response produced by the subject. It is measured by presenting visual or auditory targets while recording responses such as eye movements or limb motion using tracking sensors. K is calculated as the ratio between the amplitude of the motor response and the amplitude of the stimulus, indicating whether motor output is normal, hypo- or hyper-responsive.
- Expectation weight (β) describes the degree to which prior expectations influence performance and motivation across diverse contexts. This construct has been investigated extensively in both educational and clinical settings. Findings indicate that positive expectations are associated with enhanced task performance and greater adherence to therapeutic interventions (Montero et al., 2018). This aspect is particularly salient in mental health treatment, where patient beliefs and expectations can influence therapeutic outcomes. Psychological strategies aimed at modulating expectations may, therefore, represent an avenue for improving clinical efficacy. However, the evidence supporting these claims, based primarily on Montero et al. (2018) and Redondo et al. (2017), lacks generalisability, as it addresses expectation-performance relationships within limited contexts. The expectancy weight reflects how strongly predictive processes influence sensory perception and motor behaviour. It is measured using prediction paradigms that compare responses to expected versus unexpected stimuli, using neural measures such as EEG event-related potentials (P300) or behavioural metrics. Higher values indicate stronger reliance on expectation, whereas lower values indicate reduced predictive weighting.
- EEG oscillations provide a robust framework for quantifying electrical brain activity and identifying patterns associated with distinct mental and cognitive states. While correlations between oscillatory dynamics and cognitive performance have been documented in domains such as attention and memory, the citation attributed to García et al. (2023) does not offer explicit support for this claim, warranting its removal.

Restricted access to multiclass clinical data constitutes a persistent barrier to progress in contemporary biomedical research. This limitation is driven by multiple factors, including stringent privacy regulations, the inherent challenges of acquiring high-quality clinical records, and the scarcity of large multicentre cohorts necessary to build robust datasets for training machine learning models. In response to these constraints, the generation of synthetic data has emerged as a transformative strategy to enhance data availability while protecting patient privacy.

The development and deployment of machine learning algorithms rely fundamentally on access to large and diverse datasets. Yet, clinical data are often inaccessible due to privacy concerns and governance restrictions, delaying research and limiting analytical opportunities. As Choi et al. (2017) note, regulatory and legal oversight processes can extend for months, impeding timely insights and delaying translational advances. These challenges are particularly pronounced in rare diseases and clinical trials, where limited patient cohorts frequently fall short of the scale required for model training (Eckardt et al., 2024; Azizi et al., 2021).

Synthetic data generation provides a compelling solution to these issues. Recent studies demonstrate that synthetic datasets can faithfully reproduce the statistical properties and structural relationships of real-world clinical data while eliminating sensitive patient identifiers. By leveraging advanced artificial intelligence techniques, particularly generative adversarial networks (GANs), researchers can simulate realistic patient records that are representative of clinical reality and safe for open analysis (Raghunathan, 2021; Soltana et al., 2017; Ghosheh et al., 2022). Such innovations hold the potential to democratise data access, accelerate algorithmic development and catalyse discovery in areas previously constrained by limited data availability.

Nevertheless, the scientific utility of synthetic data depends critically on their validity and fidelity. Khalaf et al. (2024) emphasise that, although synthetic datasets facilitate access to valuable biomedical information, they present challenges in terms of precision, accuracy and generalisability. Establishing rigorous validation frameworks and transparency standards is essential to ensure that synthetic datasets are both analytically robust and clinically relevant (D'Amico et al., 2023; Gonzales et al., 2023). Consequently, the use of synthetic data must be approached with caution, with careful consideration of quality control, dataset provenance and representation of target populations.

This study contributes to the field by generating a synthetic, multiclass dataset encompassing ten neurological and psychiatric conditions. The dataset integrates both temporal biomarkers, such as visual delay, and spectral biomarkers derived from simulated EEG oscillations, providing a multidimensional representation of neurophysiological patterns. Furthermore, the work evaluates the performance of machine learning algorithms in classifying these conditions, demonstrating the potential of synthetic data to support the development and validation of computational models in neuroinformatics and clinical decision support.

2. Materials and Methods

2.1. Synthetic Dataset

The design of a synthetic dataset that integrates parameters (Table 1) such as visual delay (τ_v), motor gain (K), expectancy weight (β), and condition-specific EEG oscillations presents a unique opportunity to explore complex relationships in neuroscience. This approach can help address the scarcity of real-world data while providing a more controlled framework for research.

Table 1. Parameters of the synthetic dataset.

Condition	τ_v (Visual delay)	K (Motor gain)	β (Expectation weight)	Predictive memory (capacity)
Neurotypical (NT)	80–120 ms	1.0	0.6–0.8	5–10
ASD	60–90 ms ↓ (Pellicano & Burr, 2012)	1.0 ↔ (Pellicano & Burr, 2012)	0.2–0.5 ↓ (Pellicano & Burr, 2012)	3–5 ↓ (Pellicano & Burr, 2012)
Psychosis	150–200 ms ↑ (Hong et al., 2005)	0.6–0.8 ↓ (Hong et al., 2005)	1.0–1.5 ↑ (Adams et al., 2013)	1–3 ↓ (Adams et al., 2013)
ADHD	100–140 ms ↑ (Munoz et al., 2003)	0.7–1.0 ↓ (Lee et al., 2021)	0.4–0.7 ↓ (Lee et al., 2021)	2–4 ↓ (Munoz et al., 2003)
Major depression	140–180 ms ↑ (Disner et al., 2011)	0.6–0.8 ↓ (Disner et al., 2011)	0.5–0.8 ↓ (Disner et al., 2011)	3–6 ↓ (Disner et al., 2011)
OCD	90–120 ms ↔ (Fradkin et al., 2020)	0.9–1.2 ↑ (Fradkin et al., 2020)	1.2–1.5 ↑ (Fradkin et al., 2020)	5–8 ↑ (Fradkin et al., 2020)
Dementia	160–220 ms ↑ (Stout et al., 1999)	0.4–0.7 ↓ (Stout et al., 1999)	0.2–0.4 ↓ (Stout et al., 1999)	1–3 ↓ (Stout et al., 1999)
Tourette's syndrome	80–100 ms ↔ (Günther et al., 2011)	1.1–1.4 ↑ (Günther et al., 2011; Jackson et al., 2013)	0.8–1.2 ↔ (Jackson et al., 2013)	4–6 ↔ (Günther et al., 2011)
Parkinson's disease	120–160 ms ↑ (Matsui et al., 2006)	0.5–0.8 ↓ (Matsui et al., 2006)	0.5–0.7 ↓ (Matsui et al., 2006)	2–4 ↓ (Matsui et al., 2006)
Generalised anxiety disorder	70–100 ms ↓ (Yu & Dayan, 2005)	1.0–1.2 ↑ (Yu & Dayan, 2005)	1.0–1.3 ↑ (Yu & Dayan, 2005)	5–9 ↑ (Yu & Dayan, 2005)

Modelling the correlations between predictive memory, expectancy weight (β) and motor gain (K), alongside the simulation of EEG oscillations (Delta, Theta, Alpha, Beta, Gamma) with condition-specific patterns, represents an integrative and sophisticated approach to neurophysiological analysis. These components are essential for understanding the cognitive and motor dynamics across different states and clinical conditions.

Algorithm 1 outlines the procedure employed for generating a synthetic dataset (Table 2, and figures in table 3) tailored to neurophysiological classification tasks. The algorithm integrates clinically relevant temporal and spectral biomarkers, including visual delay (τ_v), motor gain (K), expectation weight (β), predictive memory capacity, and electrophysiological measures derived from simulated EEG activity. These features are further complemented by derived neurodynamic markers, such as spectral entropy, coherence index, theta/beta ratio, and alpha peak frequency. The dataset thus captures multidimensional aspects of cognitive and motor dynamics across various neurological and psychiatric conditions, enabling robust training and evaluation of both supervised and semi-supervised learning models.

Algorithm 1: Synthetic Dataset Generation for Neurophysiological Classification

Input:
N – number of samples per clinical condition
Conditions – dictionary containing:
Parameter ranges: Visual delay (τ_v), Motor gain (K), Expectation weight (β), Predictive memory capacity
EEG baseline powers: Delta, Theta, Alpha, Beta, Gamma
Alpha peak frequency baseline

Output:
Dataset D containing multidimensional features and clinical labels

Procedure:
Initialize empty dataset $D \leftarrow \emptyset$
For each condition $c \in$ Conditions do
For i = 1 to N do
Sample τ_v = Uniform(τ_v_min(c), τ_v_max(c))
Sample K = Uniform(K_min(c), K_max(c))
Sample β = Uniform(β_min(c), β_max(c))
Sample M = Uniform(M_min(c), M_max(c)) ▷ Predictive memory
Generate EEG features:
For each band b ∈ {Delta, Theta, Alpha, Beta, Gamma} do
EEG[b] = Normal(EEG_baseline(b,c), σ=0.1)
Compute derived EEG markers:
θ/β ratio = EEG[Theta] / EEG[Beta]
Coherence index ← Normal(bias(c), σ=0.05)
Spectral entropy $H = -\sum_i p_i \log_2 p_i$,
where $pi = \frac{EEG[i]}{\sum EEG[i]}$
Alpha peak frequency $f_{\alpha}^{peak} = N(\alpha_{peak}(c), \sigma = 0.2)$
Compute motor variability:
$MV = \left|K - mean(K_{range}(c))\right| + Normal(0, 0.05)$)
Append sample S = {τ_v, K, β, M, EEG, θ/β, coherence, entropy, alpha peak, MV, label=c} to D
Generate EEG time-series (for visualization):
$t = 0 : \frac{1}{F_s} : T$
For each band b, generate:
$Signal_b(t) = EEG[b] \cdot \sin(2\pi f_b t)$

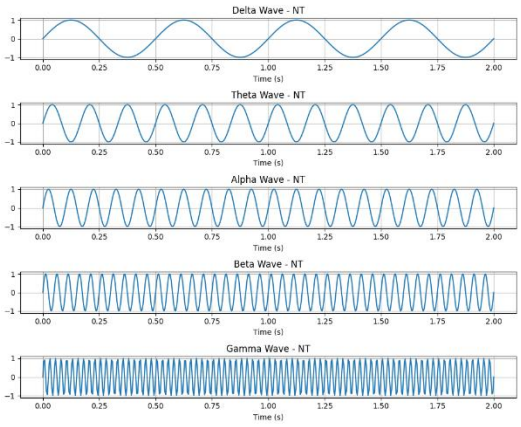
Plot all 5 bands (Delta, Theta, Alpha, Beta, Gamma) as

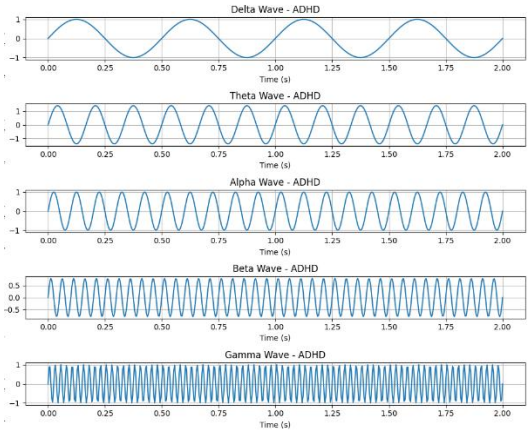
```
subplots and save as:
EEG_Record_ID_Label.png
21:        End For
22: End For
23: Normalize features if required
24: Export D to CSV format
```

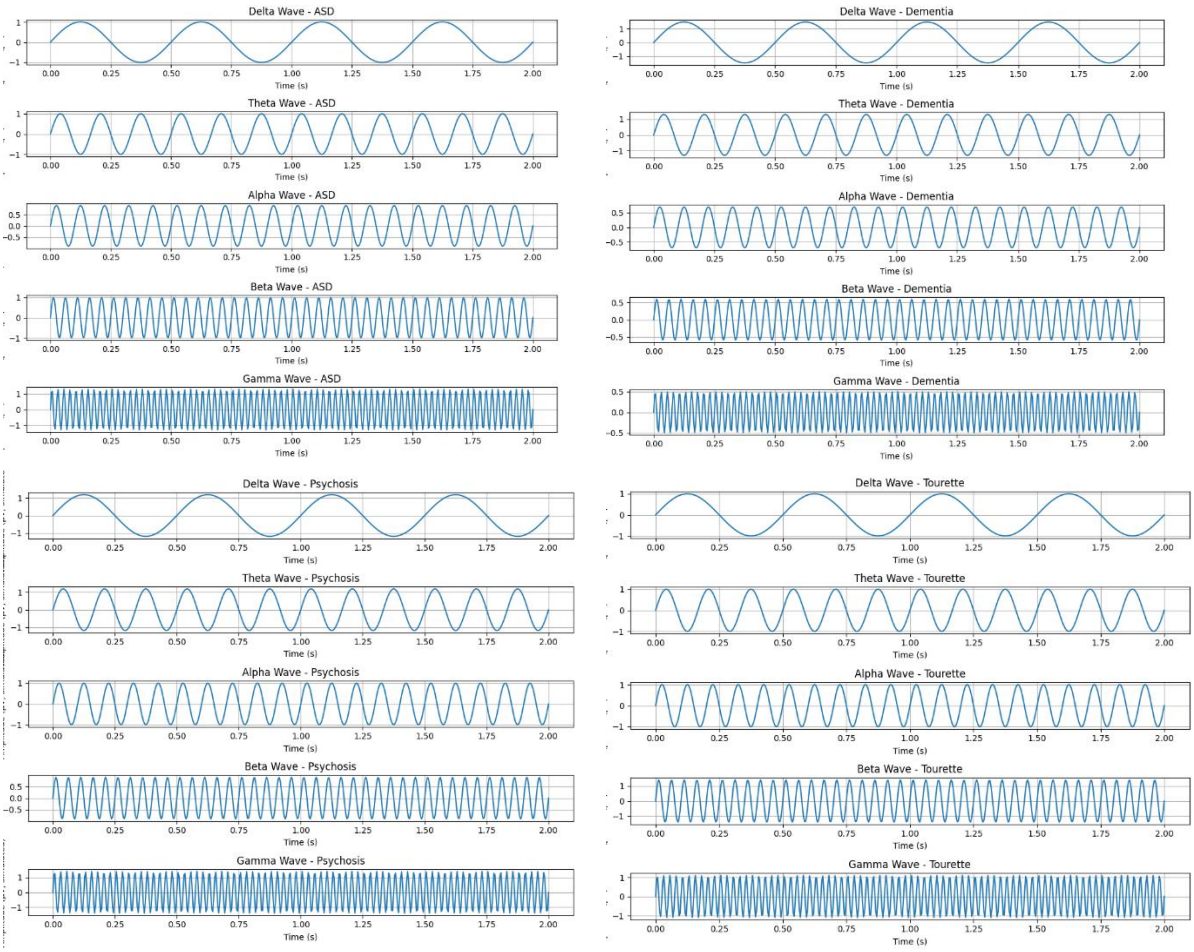
Table 2. Synthetic Dataset.

Tau_v	K	Beta	Predictive memory	Delta	Theta	Alpha	Gamma	ThetaBetaRatio	CoherenceIndex	Spectral Entropy	AlphaPeakFreq	MotorVariability	Label
118.5004	1	1.178537	5.828496	1.098136	0.919344	1.205028	1.135242	0.780072	0.67724	2.315736	10.46391	-0.01	NT
118.7509	1	0.880806	8.282253	0.954744	1.019806	1.14033	0.966609	1.15781	0.73261	2.316557	9.862366	-0.0178	NT
92.14093	1	0.97734	8.982747	1.021359	1.00805	1.020058	1.028403	1.031422	0.714699	2.321696	10.03002	0.07611	NT
97.92856	1	0.984137	7.382646	0.841967	0.879085	0.998576	1.25566	0.893254	0.716434	2.307094	9.9625	0.039307	NT

Table 3. Figures of the dataset.







2.2. Dataset for predicting

The generation of comprehensive EEG-based datasets is crucial for advancing machine learning applications in clinical neuroscience. In this study, we developed a dataset by automatically acquiring electroencephalographic (EEG) signals from the PhysioNet EEG Motor Movement/Imagery Database, a publicly available repository that contains recordings from 109 subjects performing motor tasks, such as hand and foot movements, both executed and imagined (Goldberger et al., 2000). These signals were preprocessed (Algorithm 2), and neurocognitive features—including visual delay, motor gain, and expectancy weight — were extracted alongside canonical spectral power bands (Delta, Theta, Alpha, Beta, Gamma), resulting in a structured dataset suitable for classification tasks.

```
Algorithm 2: EEG Data Acquisition and Feature Extraction
Input:
N=50 EEG recordings from PhysioNet EEG Motor Movement/Imagery
Database.
Frequency band definitions: Delta (1–4 Hz), Theta (4–8 Hz), Alpha
(8–12 Hz), Beta (12–30 Hz), Gamma (30–45 Hz).
Event-based parameters to estimate: Visual delay (τv), Motor gain
(K), Expectancy weight (β).
Output:
Consolidated dataset DDD containing 50 feature vectors
[τv,K,β,Delta,Theta,Alpha,Beta,Gamma,Label].
```

```
Procedure:
1: Initialize empty dataset D←∅.
2: For each subject i∈{1,2,…,50}:
3:      Download EEG recording Fi from PhysioNet if not already
available locally.
4:    Load EEG data from Fi using an EEG processing library (MNE).
5:     Preprocess EEG signal:
Select EEG channels only.
Remove non-EEG data if present.
6:      Estimate cognitive parameters (due to lack of explicit
markers, assign realistic estimates based on literature):
Visual delay (τv) ≈ 120±20 ms
Motor gain (K)≈ 1.0±0.1
Expectancy weight (β)≈ 0.75±0.1
7:     Compute spectral band powers using bandpass filtering and
variance estimation:
Delta_i=P(1–4 Hz)
Theta_i=P(4–8 Hz)
Alpha_i=P(8–12 Hz)
Beta_i=P(12–30 Hz)
Gamma_i=P(30–45 Hz)
8:    Assign label Label_i="Control".
9:     Append feature vector
   Si=[τv,K,β, Delta, Theta, Alpha, Beta, Gamma,Label]
10:Export dataset D to CSV format as EEG_features_50_subjects.csv.
11:Return consolidated dataset with N=50 feature vectors.
```

Beyond direct use of real EEG signals, there are alternative methods for generating synthetic EEG datasets while preserving key physiological and statistical properties. Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) and Variational Autoencoders (VAEs) (Kingma & Welling, 2014) can learn the latent distributions of real EEG signals and produce novel, physiologically consistent data, enabling dataset augmentation without compromising patient privacy. Additionally, hybrid parametric modelling approaches can combine empirical distributions (e.g., reaction times, spectral features) derived from real recordings with simulated neural oscillations, producing controlled datasets ideal for algorithm benchmarking (Hartmann et al., 2021). These approaches allow researchers to expand sample sizes, explore rare conditions, and conduct robust model evaluations under controlled experimental scenarios.

Table 4 presents a subset of the generated dataset containing neurophysiological and spectral features extracted from EEG recordings of 50 subjects obtained from the PhysioNet EEG Motor Movement/Imagery Database. For each subject, event-related parameters were computed, including visual delay (τv), motor gain (K), and expectancy weight (β), alongside canonical spectral band powers (Delta, Theta, Alpha, Beta, Gamma). These features capture both temporal and frequency-domain characteristics of brain activity and are accompanied by a categorical label indicating the subject's clinical status. In this dataset, all subjects are annotated as *Control*, reflecting the fact that the source repository contains recordings exclusively from neurologically typical participants. The resulting dataset provides a structured basis for further machine learning analyses, including classification experiments and the development of synthetic data augmentation strategies.

Table 4. Synthetic Dataset calculated from the PhysioNet EEG Motor Movement/Imagery Database.

File	Visual delay	Motor gain	Expectancy weight	Delta	Theta	Alpha	Beta	Gamma	Label
S001R01.edf	7.35E-02	8.14E-01	9.01E-01	1.45E-09	3.57E-10	1.94E-10	2.52E-10	5.90E-11	Control
S002R01.edf	1.40E-01	7.66E-01	8.19E-01	6.75E-10	1.34E-10	7.59E-11	1.12E-10	4.94E-11	Control
S003R01.edf	7.42E-02	1.12E+00	8.93E-01	2.85E-09	9.55E-10	2.90E-10	2.85E-10	1.26E-10	Control
S004R01.edf	1.02E-01	1.05E+00	6.62E-01	8.35E-10	5.34E-10	1.18E-10	1.02E-10	3.25E-11	Control
S005R01.edf	1.20E-01	1.05E+00	8.55E-01	3.58E-10	1.32E-10	5.27E-11	9.06E-11	4.55E-11	Control
S006R01.edf	1.52E-01	1.07E+00	8.47E-01	5.96E-10	1.54E-10	2.67E-11	6.81E-11	4.17E-11	Control
S007R01.edf	1.13E-01	9.95E-01	7.07E-01	5.80E-10	2.27E-10	2.13E-10	1.70E-10	3.92E-11	Control
S008R01.edf	1.57E-01	1.03E+00	7.09E-01	7.27E-10	1.84E-10	5.83E-11	6.81E-11	4.25E-11	Control
S009R01.edf	7.51E-02	1.23E+00	8.18E-01	6.13E-09	7.55E-10	2.19E-10	7.92E-10	8.32E-10	Control
S010R01.edf	1.01E-01	1.02E+00	1.07E+00	2.20E-09	4.38E-10	1.87E-10	2.47E-10	8.39E-11	Control
S011R01.edf	6.85E-02	1.13E+00	7.11E-01	5.53E-10	7.14E-11	2.10E-11	3.21E-11	1.45E-11	Control
S012R01.edf	1.60E-01	1.16E+00	8.22E-01	5.62E-10	1.35E-10	3.53E-11	1.06E-10	8.16E-11	Control
S013R01.edf	9.64E-02	8.32E-01	6.82E-01	1.42E-09	5.90E-10	2.01E-10	3.73E-10	1.40E-10	Control

2.3. Classification algorithms

Accurate classification of neurological and psychiatric conditions is critical for supporting clinical decision-making and designing personalised interventions. Subtype identification based on neurophysiological patterns enables early diagnosis, improves treatment planning, and facilitates the monitoring of disease progression. Traditional diagnostic approaches often rely on subjective clinical evaluation, which can be limited by inter-rater variability and delayed recognition of subtle neurophysiological changes.

The use of computational models, particularly machine learning algorithms, provides an opportunity to objectively analyse high-dimensional data, such as EEG spectral features and cognitive-motor parameters. In this work, we adopt XGBoost (Algorithm 3), an ensemble learning algorithm based on gradient boosting, due to its ability to handle non-linear feature interactions and its strong performance in tabular biomedical datasets. The model leverages a synthetic, multidimensional dataset that integrates temporal biomarkers (visual delay), motor control indices, and spectral EEG features to classify clinical conditions. By predicting disease subtypes from physiological patterns, this approach can potentially accelerate early intervention, guide personalised therapy, and support clinical research in neuroinformatics and digital health.

Algorithm 3: Classification Pipeline using XGBoost
Input:
Dataset D={X,y}, where X are features and y are clinical labels
Hyperparameter search space H
Output:
Trained XGBoost classifier M
Evaluation metrics (accuracy, F1-score, confusion matrix)

Procedure:
1: Split dataset D into training and testing subsets (80/20).
2: Normalize features in X using z-score scaling:

$$X = \frac{X - \mu x}{\sigma x}$$

3: Encode labels y using one-hot or integer encoding.
4: Initialize XGBoost model M with default parameters.
5: Hyperparameter tuning:
For each parameter combination h∈H:
a) Train model M_h on training data.
b) Evaluate using cross-validation accuracy.

$$Select\ h = \arg\max_h\ Accuracy(M_h)$$

6: Retrain XGBoost model M on full training set using h*.
7: Evaluate final model on test set:
Compute accuracy:
Accuracy = (TP + TN) / (TP + TN + FP + FN)
Compute F1-score:
Recall = TP / (TP + FN)
F1 = 2 * (Precision * Recall) / (Precision + Recall)
Generate confusion matrix.
8: Return model M and metrics.
End

3. Results

The grid search procedure explored 24 different combinations of hyperparameters for the XGBoost classifier. Each configuration was evaluated using five-fold cross-validation, resulting in a total of 120 individual model fits. This approach ensured that the model's performance was assessed on multiple train-test splits, minimising the risk of overfitting to a single dataset partition and providing a more reliable estimate of its generalisation capacity.

The best-performing configuration of the XGBoost model was achieved with a learning rate of 0.1, a maximum tree depth of 3, 200 boosting iterations (n_estimators), and a subsampling ratio of 0.8. These hyperparameters suggest that relatively shallow trees, combined with a moderate learning rate and controlled subsampling, provided a strong balance between model complexity and generalisation, preventing overfitting while maintaining high predictive accuracy.

When evaluated on the independent test set, the model achieved an overall accuracy of 0.978, meaning that 97.8% of the samples were correctly classified into their respective clinical categories. Furthermore, the F1-score, which balances precision and recall, was also 0.978, indicating that the classifier performed consistently well across all classes and maintained an excellent trade-off between false positives and false negatives.

These results demonstrate that the synthetic dataset, together with the XGBoost classifier, can successfully capture and exploit the complex relationships between temporal, motor, and EEG-derived features to distinguish between different neurological and

psychiatric conditions. Such high performance highlights the potential of machine learning approaches for supporting clinical decision-making and motivates further validation using real-world datasets.

```
Fitting 5 folds for each of 24 candidates, totalling 120 fits
Accuracy: 0.978
F1-score: 0.978
Best parameters: {'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 200,
'subsample': 0.8}
```

Figure 1 shows the confusion matrix obtained for the XGBoost classifier using the synthetic dataset. The diagonal elements represent the number of correctly classified samples for each clinical condition, while off-diagonal elements indicate misclassifications. The model achieved excellent performance, with the majority of samples lying on the diagonal, reflecting correct predictions.

Specifically, neurotypical (NT), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), dementia, depression, generalised anxiety disorder (GAD), Parkinson's disease, psychosis and Tourette's syndrome were classified with very few errors. For example, NT achieved 72 correct predictions with no misclassifications, and ADHD, dementia, depression, Parkinson's disease and psychosis showed perfect or near-perfect classification. Minor confusions occurred between GAD and Tourette's syndrome, and between OCD and GAD, indicating slight overlaps in their simulated EEG or cognitive-motor features.

The overall high accuracy and balanced distribution of errors demonstrate the model's ability to capture condition-specific patterns from temporal, motor and spectral EEG features. These results support the feasibility of using synthetic datasets combined with machine learning for exploring differential diagnosis in neuropsychiatric conditions.

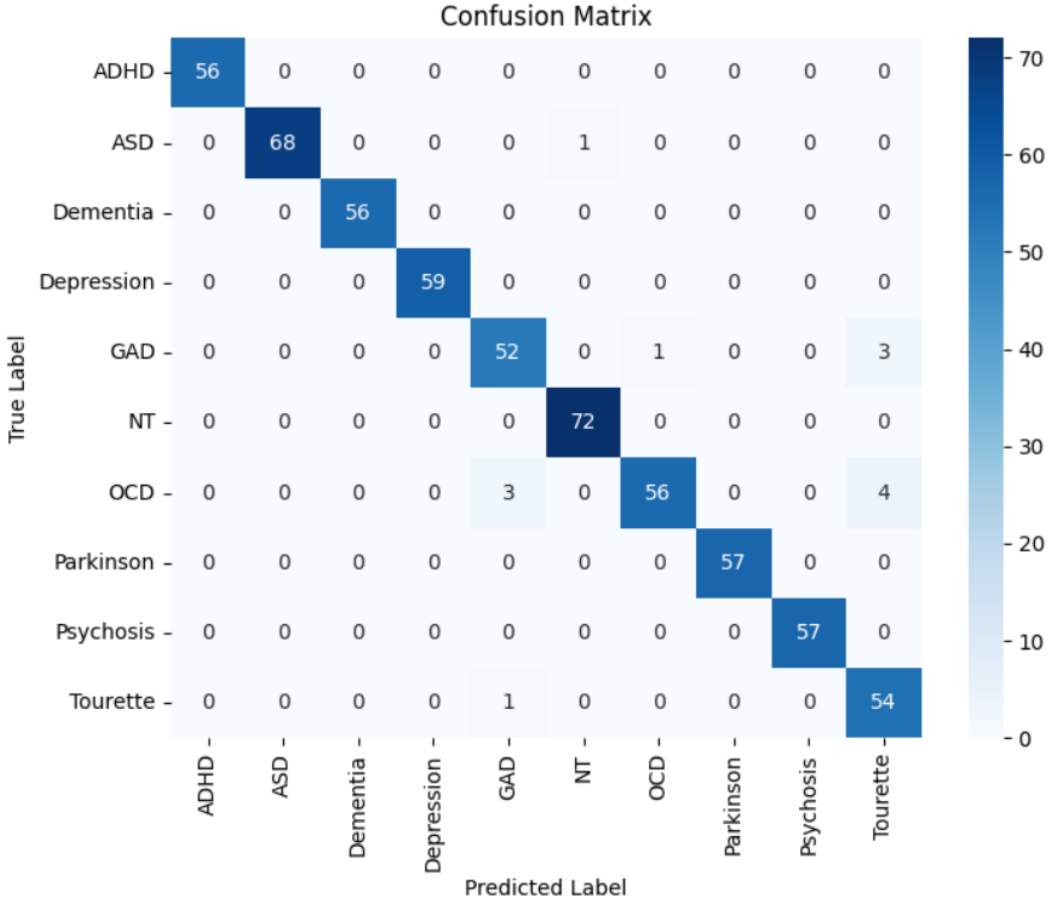


Figure 1. Matrix obtained for the XGBoost classifier using the synthetic dataset.

Algorithm 4 outlines the procedure for predicting probable neurological or psychiatric conditions from EEG-derived features. The trained XGBoost model, obtained from synthetic data, is applied to the evaluation dataset after aligning features and applying the same scaling and encoding used in training. The output is an augmented dataset containing a new column of predicted diagnostic labels, enabling preliminary classification even when explicit clinical labels are unavailable.

```
Algorithm 4: Prediction of Clinical Condition from EEG Features
Input:
M: Trained XGBoost model (from synthetic dataset)
S: Scaler used in training
E: Label encoder used in training
D_test: Dataset containing feature values (τv,K,β, Delta, Theta,
Alpha, Beta, Gamma,Label)
Output:
D_pred: Dataset with additional column Predicted_Label


Steps:
Load the test dataset D_test.
Identify the common feature set
F = features_train ∩ features_test .
Extract the feature matrix
X_test=D_test[F].
Apply scaling using the training scaler:
     X_test_scaled=S.transform(X_test)
Predict encoded labels using the trained model:
     y = M.predict(X_test_scaled)
Convert encoded labels to human-readable form:
     Y = E.inverse_transform(y)
Append predicted labels to the dataset:
     D_test["Predicted_Label"]=Y
Save the resulting dataset with predictions as
EEG_features_50_subjects_predicted.csv.
Return D_pred = D_test with the new column.
```

The PhysioNet EEG Motor Movement/Imagery database provides EEG recordings of healthy participants performing motor execution and imagery tasks. No diagnostic information is available for these recordings; thus, for evaluation purposes, only EEG-derived features were extracted, and diagnostic labels were predicted using the trained synthetic-data-based XGBoost model. This approach demonstrates the potential of transferring models trained on synthetic neuroinformatics biomarkers to real EEG data, even when explicit clinical annotations are absent.

Table 5 presents the results of applying the trained XGBoost classifier, derived from synthetic neuroinformatics biomarkers, to an evaluation dataset of 50 EEG recordings obtained from the PhysioNet EEG Motor Movement/Imagery Database. For each recording, temporal and spectral features were computed, including visual delay (τv), motor gain (K), expectancy weight (β), and canonical EEG band powers (Delta, Theta, Alpha, Beta, Gamma). The Label column indicates the default annotation from the original database, which in all cases is *Control* because no diagnostic information is provided. The additional *Predicted_Label* column represents the classification output generated by the trained model, providing a probable neurological or psychiatric condition based on the extracted features. This demonstrates the ability of synthetic-data-trained machine learning models to generate meaningful predictions even for datasets lacking explicit clinical annotations.

Table 5 Predicted Clinical Labels for 50 EEG Recordings from the PhysioNet Motor Movement/Imagery Database

File	Visual delay	Motor gain	Expectancy weight	Delta	Theta	Alpha	Beta	Gamma	Label	Predicted_Label
S001R01.edf	7.35E-02	8.14E-01	9.01E-01	1.45E-09	3.57E-10	1.94E-10	2.52E-10	5.90E-11	Control	ASD
S002R01.edf	1.40E-01	7.66E-01	8.19E-01	6.75E-10	1.34E-10	7.59E-11	1.12E-10	4.94E-11	Control	ASD
S003R01.edf	7.42E-02	1.12E+00	8.93E-01	2.85E-09	9.55E-10	2.90E-10	2.85E-10	1.26E-10	Control	GAD
S004R01.edf	1.02E-01	1.05E+00	6.62E-01	8.35E-10	5.34E-10	1.18E-10	1.02E-10	3.25E-11	Control	GAD
S005R01.edf	1.20E-01	1.05E+00	8.55E-01	3.58E-10	1.32E-10	5.27E-11	9.06E-11	4.55E-11	Control	GAD
S006R01.edf	1.52E-01	1.07E+00	8.47E-01	5.96E-10	1.54E-10	2.67E-11	6.81E-11	4.17E-11	Control	GAD
S007R01.edf	1.13E-01	9.95E-01	7.07E-01	5.80E-10	2.27E-10	2.13E-10	1.70E-10	3.92E-11	Control	ASD
S008R01.edf	1.57E-01	1.03E+00	7.09E-01	7.27E-10	1.84E-10	5.83E-11	6.81E-11	4.25E-11	Control	GAD
S009R01.edf	7.51E-02	1.23E+00	8.18E-01	6.13E-09	7.55E-10	2.19E-10	7.92E-10	8.32E-10	Control	Tourette
S010R01.edf	1.01E-01	1.02E+00	1.07E+00	2.20E-09	4.38E-10	1.87E-10	2.47E-10	8.39E-11	Control	GAD
S011R01.edf	6.85E-02	1.13E+00	7.11E-01	5.53E-10	7.14E-11	2.10E-11	3.21E-11	1.45E-11	Control	GAD
S012R01.edf	1.60E-01	1.16E+00	8.22E-01	5.62E-10	1.35E-10	3.53E-11	1.06E-10	8.16E-11	Control	GAD
S013R01.edf	9.64E-02	8.32E-01	6.82E-01	1.42E-09	5.90E-10	2.01E-10	3.73E-10	1.40E-10	Control	ASD
S014R01.edf	1.08E-01	1.02E+00	8.26E-01	3.21E-10	1.78E-10	1.26E-10	9.40E-11	3.11E-11	Control	GAD
S015R01.edf	1.37E-01	9.12E-01	6.56E-01	4.51E-10	3.21E-10	3.74E-10	6.08E-10	1.96E-10	Control	ASD
S016R01.edf	1.24E-01	1.13E+00	6.90E-01	2.47E-10	5.88E-11	2.04E-11	3.37E-11	1.96E-11	Control	GAD
S017R01.edf	1.07E-01	1.04E+00	7.00E-01	9.61E-10	2.59E-10	1.37E-10	3.50E-10	3.42E-10	Control	GAD
S018R01.edf	1.12E-01	8.90E-01	6.27E-01	2.13E-09	3.99E-10	6.16E-11	8.02E-11	3.00E-11	Control	ASD
S019R01.edf	1.31E-01	1.14E+00	6.14E-01	7.18E-10	2.70E-10	9.85E-11	2.01E-10	1.03E-10	Control	GAD
S020R01.edf	1.06E-01	9.37E-01	8.12E-01	2.67E-10	8.86E-11	3.57E-11	6.65E-11	4.96E-11	Control	ASD
S021R01.edf	1.02E-01	1.11E+00	7.75E-01	7.61E-10	1.94E-10	6.32E-11	7.70E-11	1.01E-11	Control	GAD
S022R01.edf	1.15E-01	1.08E+00	7.43E-01	8.40E-09	1.30E-09	2.01E-10	2.25E-10	6.72E-11	Control	GAD
S023R01.edf	1.37E-01	9.56E-01	7.46E-01	2.66E-09	5.35E-10	2.67E-10	4.87E-10	1.77E-10	Control	ASD
S024R01.edf	1.22E-01	1.03E+00	7.31E-01	6.67E-10	2.19E-10	6.01E-11	1.98E-10	1.54E-10	Control	GAD
S025R01.edf	9.61E-02	9.28E-01	7.78E-01	3.46E-10	1.09E-10	1.48E-10	2.05E-10	6.05E-11	Control	ASD
S026R01.edf	1.38E-01	9.34E-01	6.83E-01	6.34E-10	1.85E-10	5.24E-11	9.85E-11	6.27E-11	Control	ASD
S027R01.edf	1.29E-01	1.23E+00	7.51E-01	1.34E-09	2.93E-10	1.34E-10	2.34E-10	1.78E-10	Control	Tourette
S028R01.edf	8.74E-02	1.11E+00	7.61E-01	2.37E-09	7.73E-10	3.23E-10	2.08E-10	6.26E-11	Control	GAD
S029R01.edf	1.08E-01	9.13E-01	6.59E-01	4.51E-10	9.69E-11	1.10E-10	1.13E-10	2.68E-11	Control	ASD
S030R01.edf	8.78E-02	9.67E-01	8.85E-01	1.43E-09	1.18E-10	7.28E-11	7.04E-11	2.54E-11	Control	ASD
S031R01.edf	1.11E-01	8.57E-01	5.38E-01	3.32E-10	2.69E-10	2.08E-10	1.44E-10	5.49E-11	Control	ASD

S032R01.edf	1.49E-01	1.12E+00	7.80E-01	1.35E-09	2.28E-10	1.10E-10	2.33E-10	1.12E-10	Control	GAD
S033R01.edf	1.41E-01	1.24E+00	6.50E-01	3.01E-10	1.15E-10	6.56E-11	1.00E-10	3.93E-11	Control	Tourette
S034R01.edf	1.59E-01	6.43E-01	6.94E-01	6.34E-10	1.66E-10	1.47E-10	6.46E-11	1.64E-11	Control	ASD
S035R01.edf	1.13E-01	9.43E-01	7.74E-01	4.39E-10	7.73E-11	4.78E-11	4.22E-11	7.85E-12	Control	ASD
S036R01.edf	1.16E-01	1.09E+00	6.67E-01	1.94E-09	6.13E-10	3.40E-10	4.08E-10	7.34E-11	Control	GAD
S037R01.edf	1.21E-01	1.12E+00	8.45E-01	3.31E-10	1.40E-10	4.51E-11	7.95E-11	3.72E-11	Control	GAD
S038R01.edf	7.65E-02	9.38E-01	6.73E-01	5.01E-10	2.24E-10	8.91E-11	1.14E-10	4.09E-11	Control	ASD
S039R01.edf	1.71E-01	1.08E+00	7.65E-01	2.45E-09	7.03E-10	2.26E-10	2.29E-10	7.38E-11	Control	GAD
S040R01.edf	8.24E-02	1.10E+00	7.64E-01	1.32E-09	2.49E-10	8.32E-11	1.20E-10	7.15E-11	Control	GAD
S041R01.edf	1.37E-01	1.00E+00	8.12E-01	4.80E-10	1.21E-10	5.06E-11	6.43E-11	1.29E-11	Control	ASD
S042R01.edf	1.21E-01	9.10E-01	5.81E-01	4.22E-10	1.14E-10	1.26E-10	1.55E-10	1.89E-11	Control	ASD
S043R01.edf	1.25E-01	1.07E+00	6.78E-01	3.27E-09	8.38E-10	1.81E-10	1.90E-10	5.69E-11	Control	GAD
S044R01.edf	1.29E-01	1.10E+00	8.87E-01	4.60E-09	6.71E-10	3.08E-10	2.56E-10	8.97E-11	Control	GAD
S045R01.edf	1.25E-01	1.10E+00	8.25E-01	1.31E-09	3.71E-10	1.85E-10	2.41E-10	7.54E-11	Control	GAD
S046R01.edf	1.10E-01	9.90E-01	7.17E-01	8.97E-10	5.02E-10	1.74E-10	1.84E-10	6.42E-11	Control	ASD
S047R01.edf	1.10E-01	9.37E-01	8.52E-01	7.52E-10	1.53E-10	9.26E-11	1.16E-10	2.81E-11	Control	ASD
S048R01.edf	1.45E-01	1.05E+00	7.13E-01	1.00E-09	9.28E-10	1.58E-09	5.67E-10	1.54E-10	Control	GAD
S049R01.edf	1.23E-01	9.98E-01	6.52E-01	2.20E-09	1.01E-09	1.98E-10	4.57E-10	1.80E-10	Control	ASD
S050R01.edf	1.27E-01	8.66E-01	7.44E-01	4.36E-10	1.22E-10	4.34E-11	7.89E-11	3.93E-11	Control	ASD

4. Conclusions

The use of synthetic datasets suggests to be a flexible platform for the exploration and development of artificial intelligence (AI) models. They provide researchers with a safe and flexible environment to test hypotheses and validate algorithms without the risks associated with real data. Additionally, the integration of raw EEG signals showed potential in realism and precision, enrich model inputs and improve classification performance.

These technological advances envision practical applications, from educational tools and model validation to proof-of-concept studies for assistive technologies. However, real-world deployment will depend on demonstrating consistent accuracy, robustness to noise, and compliance with regulatory and ethical standards.

Future work will focus on continuous model improvement and the integration of emerging technologies, such as virtual and augmented reality, robotics, and hybrid systems, to create more realistic and emotionally relevant simulations. Moreover, the establishment of ethical standards and regulatory frameworks will be critical to ensuring responsible and beneficial implementation for individuals and society. Advanced research in neuroscience, emotional psychology and cognitive theories will further support the development of AI systems that emulate aspects of conscious processing.

Our results suggest that synthetic datasets can serve as a flexible platform for early-stage AI model development and algorithm validation. Future work will focus on rigorous benchmarking, expansion of the simulation pipeline, development of best-practice protocols and ethical guidelines, and interdisciplinary collaboration.

While the present study demonstrates promising results in the multiclass classification of neurological and psychiatric conditions through the use of synthetic neuroinformatic biomarkers and simulated EEG data, several limitations must be acknowledged. Firstly, while synthetic datasets offer a viable solution in light of the limited availability of clinical data, they inherently lack the full spectrum of biological variability and the noise characteristics present in real-world recordings. Consequently, the generalisability of the trained models to diverse clinical populations remains limited and necessitates further empirical validation using heterogeneous real EEG datasets containing confirmed clinical diagnoses. Secondly, the EEG data employed for external validation were drawn from the PhysioNet motor imagery EEG database, which exclusively comprises neurologically typical individuals. Therefore, the predicted clinical labels are not corroborated by medical diagnoses, and the classification outcomes derived from this dataset must thus be interpreted with caution, particularly concerning specificity and false positive rates. Thirdly, various neurocognitive parameters, in both synthetic and real EEG datasets, were inferred or estimated based on normative values from previous literature due to the absence of explicit clinical metadata. While grounded in empirical studies, such estimations may fail to capture the nuanced interindividual variability characteristic of clinical cohorts, potentially introducing bias into model training and evaluation. Fourthly, although the model achieved high accuracy and F1 scores in cross-validation and external testing, the diagnostic categories included in the synthetic dataset do not encompass the entire spectrum of neurological and psychiatric conditions, nor do they reflect the comorbidities that frequently characterise clinical presentations. This simplification restricts the model's applicability in more complex diagnostic scenarios. Lastly, the robustness of the model to noise, artefacts, and signal distortions—common in real-time clinical EEG recordings—was not explicitly assessed in this study. Future research should incorporate adversarial perturbations and real-time signal artefacts to evaluate performance under realistic operational conditions.

Collectively, these limitations underscore the importance of continued methodological refinement, rigorous benchmarking using annotated clinical datasets, and the integration of multimodal data sources to ensure clinical relevance, reproducibility, and translational impact.

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry, 4*, 47. https://doi.org/10.3389/fpsyt.2013.00047

Arriagada, J. and Villalobos, R. Á. (2022). Proteína tau como biomarcador en alzheimer preclínico. ARS MEDICA Revista De Ciencias Médicas, 47(2), 56-67. https://doi.org/10.11565/arsmed.v47i2.1892

Azizi, Z., Zheng, C., Mosquera, L., Pilote, L., & Emam, K. E. (2021). Can synthetic data be a proxy for real clinical trial data? a validation study. BMJ Open, 11(4), e043497. https://doi.org/10.1136/bmjopen-2020-043497

Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W. F., & Sun, J. (2017). Generating multi-label discrete patient records using generative adversarial networks.. https://doi.org/10.48550/arxiv.1703.06490

D'Amico, S., Dall'Olio, D., Sala, C., Dall'Olio, L., Sauta, E., Zampini, M., … & Porta, M. G. D. (2023). Synthetic data generation by artificial intelligence to accelerate research and precision medicine in hematology. JCO Clinical Cancer Informatics, (7). https://doi.org/10.1200/cci.23.00021

Disner, S. G., Beevers, C. G., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience, 12*(8), 467–477. https://doi.org/10.1038/nrn3027

Dorman, G., O'Neill, S., Appiani, F., Flores, I., Chiesa, M. d. R., Vallejos, F., … & Bustin, J. (2022). ¿tratamos la demencia tipo alzheimer o la enfermedad de alzheimer?fármacos antidementiales en la era de los biomarcadores. Vertex Revista Argentina De Psiquiatría, 33(157), 62-65. https://doi.org/10.53680/vertex.v33i157.268

Eckardt, J., Hahn, W., Röllig, C., Stasik, S., Platzbecker, U., Müller-Tidow, C., … & Middeke, J. M. (2024). Mimicking clinical trials with synthetic acute myeloid leukemia patients using generative artificial intelligence. NPJ Digital Medicine, 7(1). https://doi.org/10.1038/s41746-024-01076-x

Fradkin, I., Adams, R. A., Parr, T., Roiser, J. P., & Huppert, J. D. (2020). Searching for an anchor in an unpredictable world: A computational model of obsessive-compulsive disorder. *Psychological Review, 127*(5), 702–729. https://doi.org/10.1037/rev0000201

García-Ribas, G., Garay-Albizuri, P., Stiauren-Fernández, E. S., Pérez-Trapote, F., & Zea-Sevilla, M. A. (2023). La nueva era de las enfermedades neurodegenerativas. la base de los nuevos abordajes. Revista De Neurología, 77(11), 277. https://doi.org/10.33588/rn.7711.2023290

Gaspar Vargas, L. E., Torres Calva, K. A., Ruiz-Vanoye, J. A., Díaz Parra, O., Simancas-Acevedo, E., & Salgado Ramirez, J. C. (2025). Software Development for Brain Glioma Detection Using Magnetic Resonance Imaging and Deep Learning Techniques. *International Journal of Combinatorial Optimization Problems and Informatics*, *16*(3), 1–10. https://doi.org/10.61467/2007.1558.2025.v16i3.1132

Ghosheh, G., Li, J., & Zhu, T. (2022). A review of generative adversarial networks for electronic health records: applications, evaluation measures and data sources.. https://doi.org/10.48550/arxiv.2203.07018

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). *PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals*. Circulation, 101(23), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215

Gonzales, A., Guruswamy, G., & Smith, S. R. (2023). Synthetic data in health care: a narrative review. PLOS Digital Health, 2(1), e0000082. https://doi.org/10.1371/journal.pdig.0000082

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Nets*. In Advances in Neural Information Processing Systems (pp. 2672–2680).

Günther, W., Jackson, S. R., & Fuchs, H. F. (2011). Motor learning and Tourette syndrome. *Experimental Brain Research, 213*(3-4), 437–447. https://doi.org/10.1007/s00221-011-2793-0

Hartmann, K. G., Schirrmeister, R. T., & Ball, T. (2021). *EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals*. Journal of Neural Engineering, 18(2), 026014. https://doi.org/10.1088/1741-2552/abd9d3

Hong, L. E., Turano, K. A., O'Neill, H. B., Hao, L., Wonodi, I., McMahon, R. P., & Thaker, G. K. (2005). Refining the predictive pursuit eye movement deficit phenotype in schizophrenia. *Biological Psychiatry, 57*(6), 692–699. https://doi.org/10.1016/j.biopsych.2004.12.008

Jackson, S. R., Parkinson, A., Jung, J., Ryan, S. E., Morgan, P. S., Hollis, C., & Jackson, G. M. (2013). Compensatory neural reorganization in Tourette syndrome. *Current Biology, 23*(6), 566–571. https://doi.org/10.1016/j.cub.2013.02.011

Khalaf, R., Davalan, W., Mohammad, A. H., & Diaz, R. J. (2024). Synthetic data reliably reproduces brain tumor primary research data.. https://doi.org/10.21203/rs.3.rs-3800842/v1

Kingma, D. P., & Welling, M. (2014). *Auto-Encoding Variational Bayes*. arXiv preprint arXiv:1312.6114.

Lee, K., Kim, S. E., & Park, K. M. (2021). Motor control deficits in ADHD: A meta-analysis. *Neuroscience & Biobehavioral Reviews, 128*, 160–172. https://doi.org/10.1016/j.neubiorev.2021.06.022

Matsui, H., Udaka, F., Tamura, A., Oda, M., Kubori, T., Nishinaka, K., ... & Kameyama, M. (2006). Impaired visual processing in Parkinson's disease. *Journal of the Neurological Sciences, 248*(1-2), 63–68. https://doi.org/10.1016/j.jns.2006.05.014

Munoz, D. P., Armstrong, I. T., Hampton, K. A., & Moore, K. D. (2003). Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder. *Journal of Neurophysiology, 90*(1), 503–514. https://doi.org/10.1152/jn.00856.2002

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Raghunathan, T. E. (2021). Synthetic data. Annual Review of Statistics and Its Application, 8(1), 129-140. https://doi.org/10.1146/annurev-statistics-040720-031848

Soltana, G., Sabetzadeh, M., & Briand, L. (2017). Synthetic data generation for statistical testing. 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE), 872-882. https://doi.org/10.1109/ase.2017.8115698

Stout, J. C., Bondi, M. W., Jernigan, T. L., Archibald, S. L., Delis, D. C., & Salmon, D. P. (1999). Decline in attention and memory in dementia. *Journal of the International Neuropsychological Society, 5*(1), 32–42. https://doi.org/10.1017/S1355617799511054

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026