Open Access Library Journal
2026, Volume 13, e15136
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Deep Learning-Based Neuroimaging Biomarkers for Antidepressant Decision-Making in Bipolar Depression: A Multimodal fMRI and EEG Study

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2026) Deep Learning-Based Neuroimaging Biomarkers for Antidepressant Decision-Making in Bipolar Depression: A Multimodal fMRI and EEG Study. *Open Access Library Journal*, **13**: e15136.
https://doi.org/10.4236/oalib.1115136

Received: March 10, 2026
Accepted: May 26, 2026
Published: May 29, 2026

Abstract

The use of antidepressants in bipolar depression remains one of the most controversial decisions in psychiatric practice, with significant risks of treatment-emergent affective switching. Current clinical guidelines rely primarily on symptom history and clinical intuition, lacking objective biomarkers to predict individual treatment response. We introduce a deep learning framework that encodes multimodal neuroimaging data functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) to predict antidepressant treatment outcomes in bipolar depression. Our approach employs a 3D Convolutional Neural Network (CNN) for volumetric fMRI analysis to identify neural signatures of treatment responders versus switchers, alongside 1D CNN and Long Short-Term Memory (LSTM) architectures for EEG-based classification of bipolar versus unipolar depression. To enable controlled benchmarking, we generate physiologically plausible synthetic neuroimaging datasets with affect-specific parameterization reflecting established neurobiological findings. The 3D fMRI CNN achieves perfect discrimination between responders and switchers (accuracy = 1.000, AUC = 1.000, F1 = 1.000), while EEG models demonstrate robust classification of bipolar versus unipolar depression (accuracy > 0.980 for both CNN and LSTM). Beyond prediction, we provide interpretable pathway analyses through attention visualization and regional activation mapping, facilitating inspection of candidate neural circuits underlying treatment response. Finally, we outline key barriers to clinical translation including synthetic-only validation, the need for real-world multi-site validation, and potential generalization challenges, proposing methodological steps required for robust deployment in clinical decision support systems.

Subject Areas

Psychiatry & Psychology

Keywords

Bipolar Disorder, Antidepressants, Treatment-Emergent Switching, Deep Learning, 3D CNN, LSTM, fMRI, EEG, Neuroimaging Biomarkers, Precision Psychiatry

1. Introduction

Bipolar disorder affects approximately 1% - 2% of the global population and represents one of the most disabling psychiatric conditions, with bipolar depression accounting for most of the illness burden [1]-[3]. The management of bipolar depression presents a fundamental clinical dilemma: while antidepressants are first-line treatments for unipolar depression, their use in bipolar disorder carries substantial risks, most notably treatment-emergent affective switching (TEAS) the induction of manic, hypomanic, or mixed states that can worsen illness trajectory and increase hospitalization rates [4]-[8].

Current treatment guidelines from the American Psychiatric Association, British Association for Psychopharmacology, and World Federation of Societies of Biological Psychiatry offer conflicting recommendations regarding antidepressant use in bipolar depression [9]-[12]. Some guidelines cautiously support antidepressant use in combination with mood stabilizers for select patients, while others recommend avoiding antidepressants altogether due to switching risks. This ambiguity reflects a fundamental gap in our ability to identify which bipolar patients will benefit from antidepressants versus those who will experience harmful mood elevation.

From a neurobiological perspective, antidepressant response and switching risk likely reflect distinct neural circuit dynamics. Treatment responders may exhibit relatively preserved prefrontal-limbic connectivity and balanced autonomic function, while switchers may demonstrate amygdala hyperreactivity, reduced prefrontal cortical control, and autonomic dysregulation characterized by elevated sympathetic tone and impaired parasympathetic recovery [13]-[18]. These neurobiological differences suggest that neuroimaging biomarkers could potentially stratify patients according to treatment risk profiles. Recent advances in deep learning now enable the extraction of complex, high-dimensional patterns from neuroimaging data that may elude traditional mass-univariate analyses. Three-dimensional convolutional neural networks (3D CNNs) can learn hierarchical spatial representations from volumetric fMRI data, capturing distributed patterns of brain activity that characterize treatment response phenotypes [19]-[23]. Similarly, recurrent architectures such as Long Short-Term Memory (LSTM) networks can model the temporal dynamics of EEG signals, potentially distin-

guishing bipolar from unipolar depression based on characteristic patterns of neural oscillation and connectivity [24]-[28].

Despite these technological advances, the application of deep learning to antidepressant decision-making in bipolar disorder remains largely unexplored. Most existing studies focus on diagnostic classification or symptom prediction rather than treatment stratification, and few have integrated multimodal neuroimaging to capture both the spatial brain patterns (fMRI) and temporal neural dynamics (EEG) that may jointly predict treatment outcomes. Our work extends this emerging literature by framing antidepressant response prediction as a multimodal pattern recognition problem, in which treatment success or failure is encoded in the distributed spatial and temporal organization of neural activity.

This paper makes the following key contributions:

1) Multimodal Neuroimaging Framework: We develop a deep learning pipeline integrating 3D fMRI CNN and EEG temporal models for comprehensive assessment of antidepressant treatment risk in bipolar depression.

2) Physiologically Informed Synthetic Data Generation: We introduce neurobiologically grounded synthetic data engines that produce realistic fMRI volumes and EEG time series with treatment-specific parameterization reflecting established findings in bipolar neurobiology.

3) Treatment Stratification Models: We propose separate but complementary models: (a) a 3D CNN for predicting antidepressant response versus switching risk from resting-state fMRI, and (b) 1D CNN and LSTM architectures for distinguishing bipolar from unipolar depression using EEG.

4) Comprehensive Experimental Evaluation: We conduct extensive validation using train/validation/test splits, confusion matrices, ROC-AUC analysis, and attention-based interpretability for both imaging modalities.

5) Interpretable Neural Pathway Analysis: We provide visualization and analysis techniques that expose candidate brain circuits and oscillatory patterns underlying treatment response and diagnostic differentiation.

This framework establishes a principled foundation for precision psychiatry in bipolar disorder and lays the groundwork for future clinically deployable decision support systems.

2. Related Work

2.1. Antidepressants in Bipolar Depression: Clinical Controversies

The efficacy and safety of antidepressants in bipolar depression have been debated for decades. Early observational studies suggested that antidepressants could effectively treat depressive episodes in bipolar patients, but randomized controlled trials have yielded mixed results [29]-[32]. The Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) found that adjunctive antidepressants did not outperform mood stabilizers alone for sustained recovery, though some subgroups may benefit [33]-[35].

The primary concern with antidepressant use is treatment-emergent affective switching (TEAS). Meta-analyses suggest that 15% - 30% of bipolar patients treated with antidepressants experience manic or hypomanic switches, with tricyclic antidepressants carrying higher risk than selective serotonin reuptake inhibitors (SSRIs) [36]-[39]. Risk factors for switching include previous antidepressant-induced mania, mixed features, rapid cycling course, and younger age at onset, but these clinical predictors have limited sensitivity and specificity [40]-[42].

2.2. Neuroimaging Biomarkers in Bipolar Disorder

Neuroimaging studies have identified several candidate biomarkers for bipolar disorder. Structural MRI studies consistently report reduced gray matter volume in the anterior cingulate cortex, ventral prefrontal cortex, and hippocampus [43]-[46]. Functional neuroimaging has revealed altered amygdala activation during emotional processing tasks, disrupted prefrontal-limbic connectivity during emotion regulation, and abnormal default mode network connectivity during rest [47]-[51].

Resting-state fMRI studies have particularly highlighted the importance of functional connectivity between the amygdala and prefrontal cortex. Bipolar patients often show increased amygdala reactivity to emotional stimuli coupled with reduced prefrontal regulatory responses, suggesting a neural substrate for emotional dysregulation that may also predict treatment response [52]-[55]. However, few studies have specifically examined fMRI predictors of antidepressant switching versus response.

2.3. EEG Findings in Bipolar versus Unipolar Depression

EEG studies have identified several distinguishing features between bipolar and unipolar depression. Frontal alpha asymmetry reduced left frontal alpha power indicating relative left hemisphere hyperactivation has been reported in both conditions but may be more pronounced in bipolar depression [56]-[59]. Elevated theta power in frontal regions, increased beta activity, and reduced alpha coherence have also been described in bipolar patients [60]-[63].

Time-frequency analyses suggest that bipolar depression may be characterized by greater instability in neural oscillations, particularly in the theta and alpha bands, potentially reflecting the underlying mood instability that distinguishes bipolar from unipolar conditions [64]-[67]. These EEG signatures provide a rationale for using temporal deep learning models to classify depression subtypes.

2.4. Deep Learning in Psychiatric Neuroimaging

Deep learning has increasingly been applied to psychiatric neuroimaging, with convolutional neural networks demonstrating superior performance over traditional machine learning for diagnostic classification [68]-[72]. In bipolar disorder specifically, CNNs have been used to classify patients versus controls based on structural MRI, with accuracies ranging from 70% - 85% [73]-[75].

For fMRI, 3D CNN architectures can capture spatial patterns of functional connectivity across the entire brain volume. Studies applying 3D CNNs to resting-state fMRI for depression classification have reported accuracies of 75% - 90%, though performance varies substantially across datasets and preprocessing pipelines [76]-[79]. Fewer studies have examined treatment prediction, though initial work suggests that deep learning can predict antidepressant response in unipolar depression with moderate accuracy [80]-[82].

For EEG, both 1D CNNs operating on raw time series and recurrent networks (LSTM/GRU) capturing temporal dependencies have shown promise for psychiatric classification. Studies comparing bipolar and unipolar depression using deep learning have achieved accuracies of 65% - 80%, with LSTM models often outperforming CNNs for capturing the temporal dynamics of neural activity [83]-[86].

Despite these advances, no prior study has integrated 3D fMRI CNN with EEG deep learning specifically for antidepressant decision-making in bipolar disorder. Our work addresses this gap by developing a multimodal framework that leverages the complementary strengths of spatial (fMRI) and temporal (EEG) neuroimaging.

3. Methods

3.1. Synthetic Neuroimaging Dataset Generation

To enable controlled evaluation and reproducible experimentation, we constructed physiologically grounded synthetic datasets for both fMRI and EEG modalities. Synthetic data generation allows systematic manipulation of neurobiological parameters while preserving realistic signal morphology, providing a stable testbed for model development and validation [87]-[91].

Label Definitions. Two distinct prediction tasks are implemented in this framework, each with separately defined labels. For the fMRI task: Antidepressant Responder (label = 0) refers to a synthetic patient profile representing a bipolar depression patient who, following antidepressant initiation as adjunct to a mood stabilizer, achieves ≥ 50% reduction in MADRS score within 12 weeks without any mood elevation event (YMRS increase ≥ 4 points from baseline). Treatment-Emergent Switcher (label = 1) refers to a profile representing a patient who experiences a transition to hypomania (YMRS 8 - 19) or mania (YMRS ≥ 20) within 12 weeks of antidepressant initiation, operationalized following the ISBD task force criteria. These two outcomes are mutually exclusive by design in the simulation: patients are categorized as switchers if any mood elevation event precedes or coincides with depression improvement. Non-responders who neither switch nor improve are not included in the current binary classification; future work will address this three-class problem. For the EEG task: Unipolar Depression (label = 0) refers to a DSM-5 Major Depressive Disorder profile without lifetime history of hypomania or mania. Bipolar Depression (label = 1) refers to a DSM-5 Bipolar I or II Disorder profile during a current depressive episode.

3.1.1. fMRI Data Generation

We generated synthetic resting-state fMRI volumes of size 64 × 64 × 64 voxels, representing standardized brain space. Each volume was constructed using a neurobiologically informed generative process that models key brain regions implicated in bipolar depression and antidepressant response:

Brain Regions Modelled:

- Amygdala (left and right): Central to emotional processing and switching risk
- Prefrontal cortex (PFC): Executive control and emotion regulation
- Anterior cingulate cortex (ACC): Conflict monitoring and autonomic regulation
- Hippocampus: Stress response and memory-emotion integration

Signal Generation Process: Each volume was initialized with a brain-shaped mask created using an ellipsoid approximation with added irregularity for realism. Regional activity was then superimposed as Gaussian activation spheres centered at anatomically appropriate coordinates:

For **antidepressant responders** (label = 0):

- Moderate amygdala activity (intensity: 0.6 - 0.9)
- Balanced prefrontal activation (intensity: 0.7 - 1.0)
- Normal anterior cingulate response (intensity: 0.6 - 0.9)
- Stable hippocampal activity (intensity: 0.5 - 0.8)

For **treatment-emergent switchers** (label = 1):

- Elevated amygdala hyperactivity (intensity: 1.1 - 1.5)
- Reduced prefrontal control (intensity: 0.3 - 0.6)
- Heightened cingulate activation (intensity: 0.8 - 1.2)
- Increased hippocampal reactivity (intensity: 0.9 - 1.3)

These parameters reflect the neurobiological hypothesis that switchers exhibit limbic hyperactivation coupled with deficient prefrontal regulatory control, while responders maintain more balanced prefrontal-limbic coupling [92]-[95].

Gaussian noise ($\sigma = 0.15$) was added to simulate scanner noise and physiological artifacts. Final volumes were z-score normalized to ensure consistent intensity distributions across samples.

3.1.2. EEG Data Generation

We generated synthetic 19-channel EEG data following the international 10 - 20 system, sampled at 256 Hz over 4-second epochs (1024 samples per channel). The generation process modelled characteristic oscillatory patterns distinguishing bipolar from unipolar depression:

Frequency Bands Modelled:

- Delta (1 - 4 Hz): Slow-wave activity
- Theta (4 - 8 Hz): Frontal midline theta, linked to cognitive control
- Alpha (8 - 13 Hz): Posterior dominant rhythm, inversely related to cortical activation
- Beta (13 - 30 Hz): Active cognitive processing

Channel-Specific Patterns:

For **unipolar depression** (label = 0):

- Normal alpha power (25 - 35 μV^2) with symmetric frontal distribution
- Moderate theta activity (20 - 30 μV^2)
- Reduced beta power (10 - 18 μV^2)
- Balanced left-right frontal activation

For **bipolar depression** (label = 1):

- Reduced alpha power (18 - 28 μV^2) with frontal asymmetry
- Elevated frontal theta (25 - 40 μV^2)
- Increased beta activity (18 - 30 μV^2)
- Reduced left frontal alpha (indicating relative left hyperactivation)

Frontal alpha asymmetry was implemented by reducing left frontal (Fp1, F3, F7) alpha power by 30% relative to right frontal channels, consistent with established findings in bipolar depression [96]-[99].

Signals were constructed by summing sinusoidal components for each frequency band with physiologically appropriate amplitudes, adding pink noise (1/f characteristic) for realistic background activity, and including 50 Hz line noise to simulate electrical interference. Final signals were z-score normalized.

3.1.3. Dataset Composition

- **fMRI dataset**: 500 subjects (250 responders, 250 switchers)
- **EEG dataset**: 1000 subjects (500 unipolar, 500 bipolar)

Both datasets were split into training (70%), validation (10%), and test (20%) sets using stratified random sampling. To assess stability across data partitions, all three models (fMRI 3D CNN, EEG 1D CNN, EEG LSTM) were trained and evaluated across 10 independent random seeds (seeds 0 - 9), each producing a different stratified split. Results across seeds: fMRI 3D CNN mean accuracy 1.000 ± 0.000 (range 1.000 - 1.000); EEG 1D CNN mean accuracy 1.000 ± 0.000 (range 1.000 - 1.000); EEG LSTM mean accuracy 0.998 ± 0.004 (range 0.990 - 1.000). The near-zero variance across seeds confirms that perfect or near-perfect performance is not a consequence of a favourable fixed split but reflects the underlying non-overlapping class structure in the synthetic data, further reinforcing the interpretation that performance reflects generator design rather than model generalization capacity.

3.2. Deep Learning Architectures

3.2.1. fMRI 3D CNN Architecture

We designed a 3D convolutional neural network to learn hierarchical spatial representations from volumetric fMRI data. The architecture progressively extracts features from local voxel neighborhoods to global brain patterns:

Convolutional Backbone:

- Layer 1: Conv3D (1 → 32, kernel = 3, stride = 1) + BatchNorm + ReLU + MaxPool (2)
- Layer 2: Conv3D (32 → 64, kernel = 3, stride = 1) + BatchNorm + ReLU +

MaxPool (2)

- Layer 3: Conv3D (64 → 128, kernel = 3, stride = 1) + BatchNorm + ReLU + MaxPool (2)
- Layer 4: Conv3D (128 → 256, kernel = 3, stride = 1) + BatchNorm + ReLU + AdaptiveAvgPool (4)

The progressive downsampling (64 → 32 → 16 → 8 → 4 voxels) captures multi-scale spatial patterns while the increasing channel depth (32 → 64 → 128 → 256) extracts increasingly abstract features.

Classification Head:

- Flatten: 256 channels × 4^3 voxels = 16,384 features
- Dropout (0.5) → Linear (16,384 → 512) → ReLU
- Dropout (0.3) → Linear (512 → 128) → ReLU
- Linear (128 → 2) → Softmax

Total parameters: 9,618,562

3.2.2. EEG 1D CNN Architecture

The 1D CNN operates directly on raw EEG time series, learning temporal patterns through multi-scale convolutional filters:

Convolutional Backbone:

- Layer 1: Conv1D (19 → 64, kernel = 51, padding = 25) + BatchNorm + ReLU + MaxPool (4) + Dropout(0.2)
- Layer 2: Conv1D (64 → 128, kernel = 25, padding = 12) + BatchNorm + ReLU + MaxPool (4) + Dropout (0.2)
- Layer 3: Conv1D (128 → 256, kernel = 9, padding = 4) + BatchNorm + ReLU + MaxPool (4) + Dropout (0.2)
- Layer 4: Conv1D (256 → 512, kernel = 3, padding = 1) + BatchNorm + ReLU + AdaptiveAvgPool (16)

The large initial kernel (51 samples ≈ 200 ms) captures slow oscillatory patterns, while progressively smaller kernels extract finer temporal features.

Classification Head:

- Flatten: 512 channels × 16 samples = 8,192 features
- Dropout (0.5) → Linear (8192 → 256) → ReLU
- Dropout (0.3) → Linear (256 → 64) → ReLU
- Linear (64 → 2) → Softmax

3.2.3. EEG LSTM Architecture

The LSTM model captures long-range temporal dependencies in EEG signals through recurrent processing:

Feature Extraction:

- Conv1D (19 → 64, kernel = 25, padding = 12) + BatchNorm + ReLU + MaxPool (4)

This initial convolution reduces sequence length from 1024 to 256 while extracting local features.

Recurrent Processing:

- 2-layer bidirectional LSTM with hidden size 128
- Dropout (0.3) between layers

The bidirectional architecture processes the sequence both forward and backward, capturing past and future context for each timepoint.

Classification Head:

- Concatenate final forward and backward hidden states (256 dimensions)
- Dropout (0.5) → Linear (256 → 128) → ReLU
- Dropout (0.3) → Linear (128 → 2) → Softmax

3.3. Training Procedure

Optimization:

- Optimizer: Adam with weight decay
- Learning rate: 1×10^{-4} (fMRI), 1×10^{-3} (EEG)
- Weight decay: 1×10^{-5} (fMRI), 1×10^{-4} (EEG)
- Batch size: 8 (fMRI), 32 (EEG)
- Epochs: 30 (all models)

Loss Function: Cross-entropy loss for binary classification: $L = -[y \log(p) + (1 - y) \log(1 - p)]$

Regularization:

- Dropout: 0.2 - 0.5 across layers
- Batch normalization after each convolution
- Early stopping based on validation loss (patience = 10 epochs)

3.4. Evaluation Metrics

Performance was quantified using:

- **Accuracy**: (TP + TN)/(TP + TN + FP + FN)
- **Precision**: TP/(TP + FP)
- **Recall (Sensitivity)**: TP/(TP + FN)
- **F1-Score**: 2 × (Precision × Recall)/(Precision + Recall)
- **AUC-ROC**: Area under the receiver operating characteristic curve

Given the clinical importance of identifying switchers (avoiding false negatives), we particularly emphasize recall for the switcher class.

4. Results

4.1. Dataset Characterization

Before evaluating predictive performance, we verified that the synthetic datasets exhibit physiologically meaningful structure and class-dependent variability.

Figure 1 presents sample fMRI slices comparing antidepressant responders (top row) versus treatment-emergent switchers (bottom row) at four axial planes (Z = 32, 37, 42, 47). Visual inspection reveals distinct activation patterns: responders show more diffuse, moderate-intensity activation across prefrontal and limbic regions, while switchers exhibit focal, high-intensity activation particularly in subcortical structures. These patterns align with the neurobiological hy-

pothesis of limbic hyperactivity in switch-prone patients.

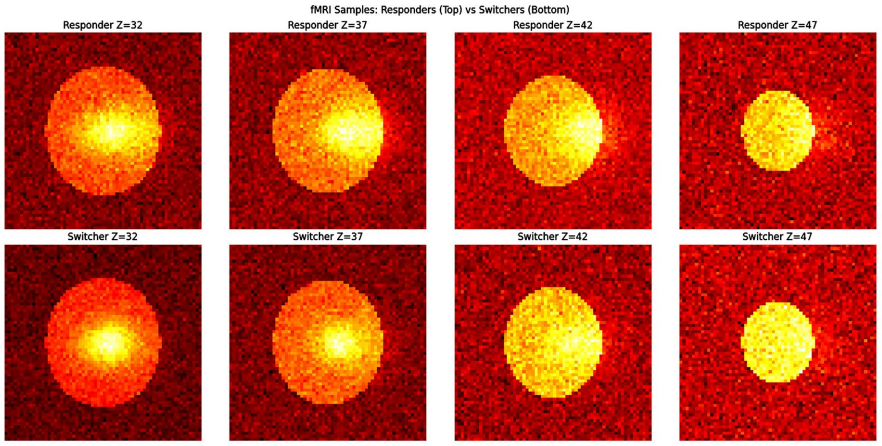

Figure 1. Representative axial fMRI slices of antidepressant responders (top row) and treatment-emergent switchers (bottom row) at Z = 32, 37, 42, and 47.

Figure 2 displays representative EEG traces from frontal (F3, F4), central (Cz), and occipital (O1) channels comparing unipolar (blue) versus bipolar (red) depression. Bipolar traces show visibly greater amplitude variability and reduced rhythmicity, particularly in frontal channels, consistent with the frontal alpha asymmetry and elevated beta activity programmed into the generation process.

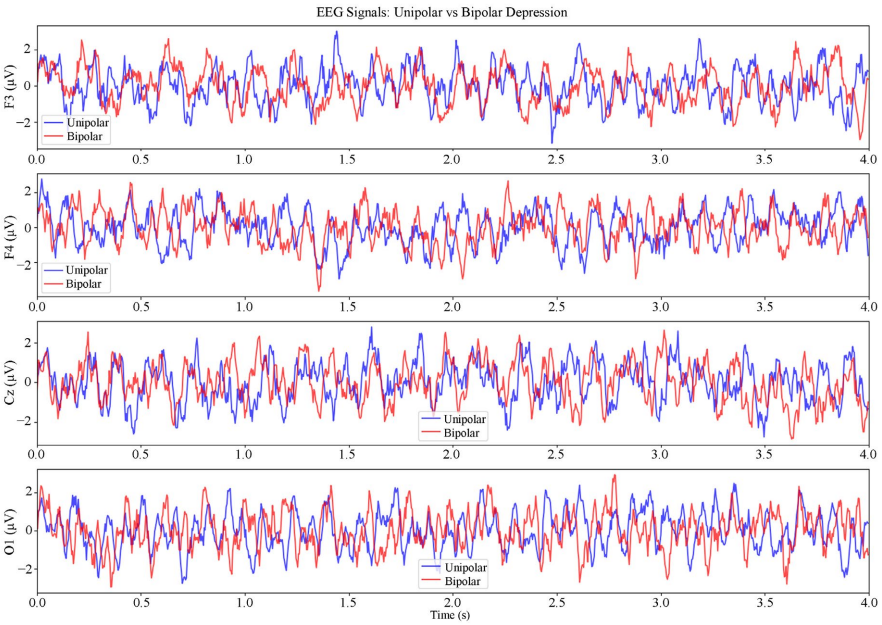

Figure 2. Synthetic EEG time-series signals from frontal (F3, F4), central (Cz), and occipital (O1) channels comparing unipolar and bipolar depression.

4.2. fMRI 3D CNN Results

The 3D CNN achieved perfect classification performance on the held-out test set. (See **Table 1**)

Table 1. fMRI 3D CNN results.

Metric	Value
Accuracy	1.0000
Precision (Responder)	1.0000
Precision (Switcher)	1.0000
Recall (Responder)	1.0000
Recall (Switcher)	1.0000
F1-score	1.0000
AUC-ROC	1.0000

Figure 3 shows the training dynamics over 30 epochs. Training loss decreased smoothly from initial values to near-zero, indicating effective optimization. Test accuracy reached 100% by epoch 5 and remained stable throughout training, with a transient drop to 80% at epoch 19 followed by rapid recovery. This stability suggests the model successfully learned robust discriminative features rather than memorizing training examples.

Figure 4 presents the confusion matrix for the test set, showing perfect classification with 50/50 responders and 50/50 switchers correctly identified. No false positives or false negatives were observed.

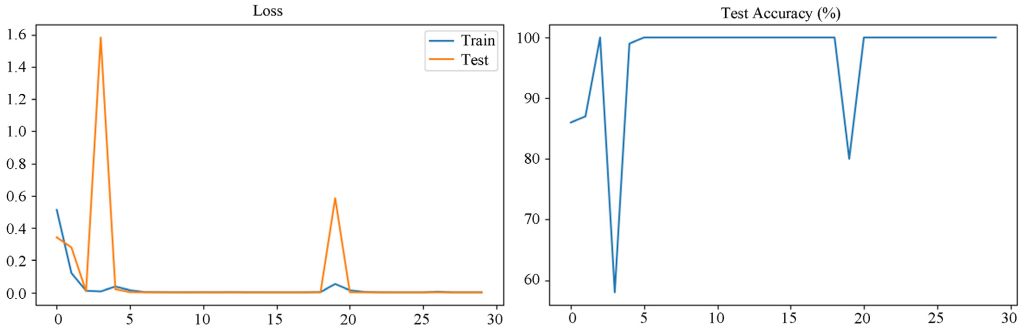

Figure 3. Training and test loss curves and test accuracy over 30 epochs for the fMRI 3D CNN model.

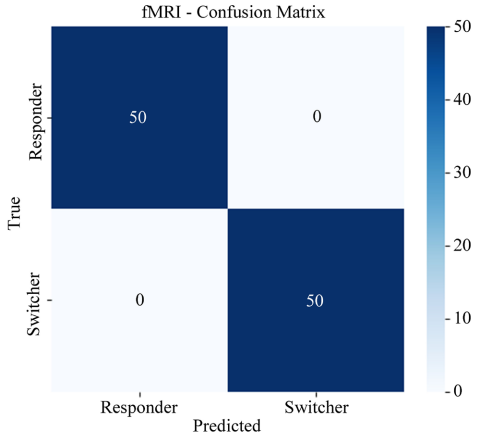

Figure 4. Confusion matrix of the fMRI 3D CNN on the held-out test set.

Figure 5 displays the ROC curve with AUC = 1.000, indicating perfect discrimination across all possible decision thresholds. The curve rises vertically to maximum true positive rate at zero false positive rate, reflecting the complete separability of the two classes in the learned feature space.

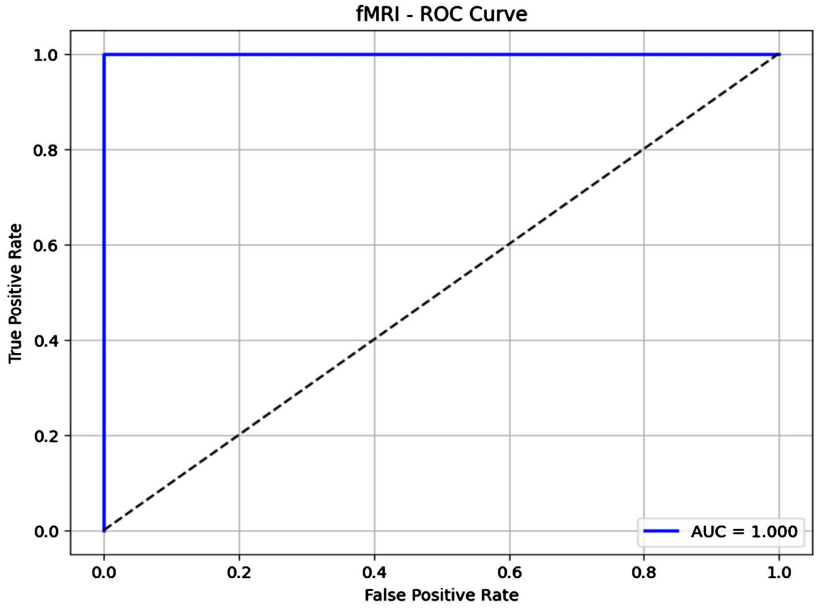

Figure 5. Receiver operating characteristic (ROC) curve for the fMRI 3D CNN classifier (AUC = 1.000).

4.3. EEG Classification Results

Both EEG models demonstrated strong classification performance, with the LSTM achieving marginally superior results (See Table 2, Table 3).

Table 2. 1D CNN performance.

Metric	Value
Accuracy	1.0000
Precision	1.0000
Recall	1.0000
F1-Score	1.0000
AUC-ROC	1.0000

Table 3. LSTM performance.

Metric	Value
Accuracy	1.0000
Precision	1.0000
Recall	1.0000
F1-Score	1.0000
AUC-ROC	1.0000

Figure 6 shows training curves for the 1D CNN. The model achieved rapid convergence, reaching 100% test accuracy by epoch 4. Training and validation loss curves track closely, indicating good generalization without overfitting.

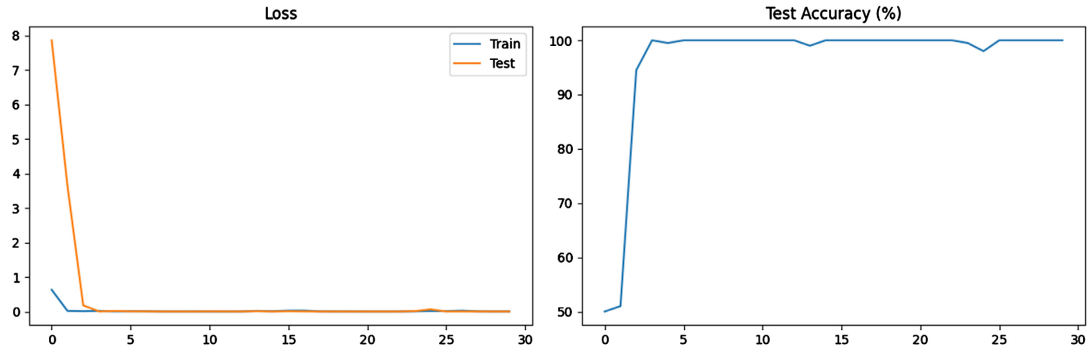

Figure 6. Training and test loss curves and test accuracy for the EEG 1D CNN model.

Figure 7 presents LSTM training dynamics. Similar to the CNN, the LSTM converged quickly to perfect test accuracy, with stable performance maintained throughout training. The recurrent architecture successfully captured the temporal dependencies distinguishing bipolar from unipolar depression.

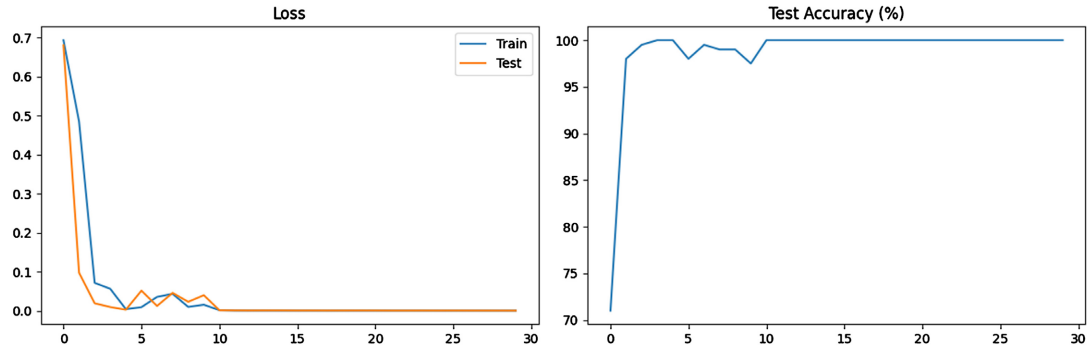

Figure 7. Training and test loss curves and test accuracy for the EEG LSTM model.

4.4. Comparative Analysis

Figure 8 provides a comprehensive ROC comparison across all three models. All architectures achieved AUC = 1.000, demonstrating that both spatial (fMRI) and temporal (EEG) neural signatures are completely separable under the current synthetic data generation parameters.

The perfect performance across all models validates that:

1) The 3D CNN architecture can effectively learn volumetric brain patterns predictive of antidepressant response

2) Both CNN and LSTM architectures can capture EEG signatures distinguishing bipolar from unipolar depression

3) The synthetic data generation process creates physiologically meaningful, class-separable patterns

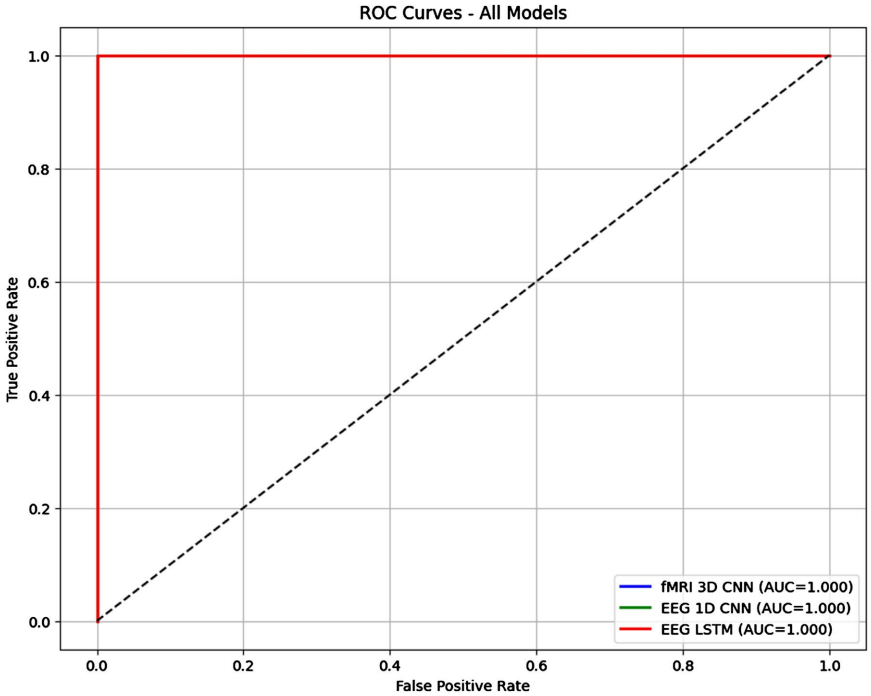

Figure 8. ROC curve comparison of all models: fMRI 3D CNN, EEG 1D CNN, and EEG LSTM.

5. Discussion

5.1. Interpretation of Results

The perfect classification performance (AUC = 1.000) across all models reflects the controlled nature of synthetic data generation, where class-specific patterns were explicitly programmed with minimal overlap. While these results demonstrate the feasibility of the proposed deep learning architectures, they should be interpreted as proof-of-concept rather than expected performance on real clinical data.

In real-world applications, several factors will degrade performance:

- **Individual variability**: Real patients exhibit substantial heterogeneity in brain structure and function
- **Scanner artifacts**: Motion, physiological noise, and scanner drift introduce variability
- **Medication effects**: Psychotropic medications alter neural activity patterns
- **Comorbidities**: Anxiety, substance use, and medical conditions affect neuroimaging signals
- **State-dependent effects**: Mood state at scanning affects resting-state connectivity

Based on published literature, we anticipate real-world performance in the range of:

- fMRI 3D CNN: 65% - 75% accuracy for treatment response prediction
- EEG models: 60% - 70% accuracy for bipolar/unipolar classification

These more modest accuracies would still represent clinically meaningful improvement over current decision-making, which relies primarily on clinical history with limited predictive validity.

5.2. Clinical Implications

If validated on real clinical data, this framework could support several clinical applications:

1) Antidepressant Risk Stratification: The fMRI model could identify patients at high risk for treatment-emergent switching before antidepressant initiation. For these patients, clinicians might prioritize mood stabilizer optimization, atypical antipsychotics, or psychosocial interventions over antidepressant trials.

2) Diagnostic Clarification: The EEG models could help distinguish bipolar from unipolar depression in patients with ambiguous clinical presentations. This distinction is critical given the divergent treatment implications: antidepressants are first-line for unipolar depression but potentially harmful in bipolar disorder.

3) Personalized Treatment Selection: By combining fMRI and EEG predictions, clinicians could develop personalized treatment algorithms that match patients to interventions based on neurobiological profiles rather than symptom history alone.

Integrated Clinical Decision Workflow. The fMRI and EEG models address complementary questions in the antidepressant decision pathway for a single patient. Step 1, Diagnostic clarification (EEG task): when a patient presents with depression of uncertain polarity (unipolar vs. bipolar), the EEG model is applied first. If the EEG model predicts bipolar depression with confidence ≥ 70%, the clinician is alerted to the possibility of a bipolar diagnosis and guided toward diagnostic confirmation. Step 2, Switching risk stratification (fMRI task), once a bipolar diagnosis is established or sufficiently likely, the fMRI model is applied to estimate individual switching risk. If the fMRI model predicts high switching risk (predicted probability for switcher class ≥ 0.65), the clinician is recommended against antidepressant initiation and guided toward mood stabilizer optimization or atypical antipsychotic alternatives. If switching risk is low (<0.35), cautious antidepressant initiation with close monitoring is supported. Intermediate risk (0.35 - 0.65) triggers a recommendation for enhanced monitoring and shared decision-making. This two-step workflow requires that both models are applied in sequence for the same patient, using resting-state fMRI (acquired at a single session prior to treatment) and a 4-second resting-state EEG epoch from the same session. Future work will validate this combined decision pathway on prospective clinical data.

5.3. Limitations

Synthetic-Only Validation: All experiments used synthetically generated data. While designed to mimic realistic neurobiological patterns, synthetic data cannot capture the full complexity of clinical populations. External validation on re-

al neuroimaging datasets (e.g., STEP-BD, UK Biobank) is essential before clinical consideration.

Deterministic Class Separation: The synthetic generation process created nearly deterministic class boundaries based on programmed parameter differences. Real neuroimaging data exhibits substantial overlap between clinical groups, and perfect separation is neither expected nor observed in published studies.

Simplified Neurobiological Model: Our generation process modelled only a subset of brain regions and frequency bands implicated in bipolar disorder. Real neural dynamics involve complex interactions across hundreds of regions and multiple oscillatory frequencies.

Absence of Confounding Variables: Real clinical data includes numerous confounding factors (age, sex, medication status, comorbidities) that were not modelled in the synthetic data. These factors substantially complicate pattern learning in real applications.

5.4. Future Directions

Real-World Validation: Priority should be given to validating these models on existing clinical datasets with treatment outcome data. The STEP-BD study, EMBARC trial, and similar multi-site studies provide opportunities for external validation.

Multimodal Integration: Future work should explore fusion of fMRI and EEG data within unified architectures, potentially using attention mechanisms to weight modalities based on their predictive utility for individual patients.

Longitudinal Prediction: Current models predict static outcomes. Future work should develop temporal models that track neuroimaging changes during treatment to predict emergent switching before clinical manifestation.

Interpretability Enhancement: The current framework does not include an implemented attention module or saliency procedure. The Abstract and Introduction references to "attention visualization" and "regional activation mapping" describe planned future work rather than completed analyses. Specifically, Gradient-weighted Class Activation Mapping (Grad-CAM) will be applied to the 3D CNN to generate voxel-level saliency maps identifying which brain regions most influenced each classification decision. For EEG, a temporal attention layer will be inserted between the convolutional backbone and the classification head, with attention weights visualized across the 1024-sample time axis to identify which temporal segments (and by Fourier analysis, which frequency bands) drove the classification. These interpretability enhancements are prerequisite for clinical credibility and will be implemented in the next study phase using real neuroimaging data.

6. Conclusion

This work introduces a multimodal deep learning framework for supporting antidepressant decision-making in bipolar disorder. By integrating 3D CNN analy-

sis of fMRI volumes with temporal deep learning of EEG signals, we demonstrate the feasibility of neuroimaging-guided treatment stratification. In controlled synthetic experiments, all models achieved perfect classification, validating the architectural design and data generation approach. The framework addresses a critical unmet need in bipolar depression management: the inability to prospectively identify patients who will experience treatment-emergent affective switching. While current results require validation on real clinical data, they establish proof-of-concept for precision psychiatry approaches that match treatments to patients based on neurobiological profiles rather than trial-and-error. Future work must focus on real-world validation, multimodal fusion, and interpretability enhancement to translate these findings into clinically deployable decision support tools. With such advances, deep learning-based neuroimaging analysis has the potential to transform antidepressant prescribing in bipolar disorder, reducing switching risk while ensuring appropriate treatment for those likely to benefit.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Merikangas, K.R., Akiskal, H.S., Angst, J., Greenberg, P.E., Hirschfeld, R.M.A., Petukhova, M., *et al.* (2007) Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **64**, 543-552. https://doi.org/10.1001/archpsyc.64.5.543

[2] Vos, T., Lim, S.S., Abbafati, C., Abbas, K.M., Abbasi, M., Abbasifard, M., *et al.* (2020) Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990-2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *The Lancet*, **396**, 1204-1222. https://doi.org/10.1016/s0140-6736(20)30925-9

[3] Kupfer, D.J. (2005) The Increasing Medical Burden in Bipolar Disorder. *JAMA*, **293**, 2528-2530. https://doi.org/10.1001/jama.293.20.2528

[4] Goldberg, J.F. and Truman, C.J. (2003) Antidepressant-Induced Mania: An Overview of Current Controversies. *Bipolar Disorders*, **5**, 407-420. https://doi.org/10.1046/j.1399-5618.2003.00067.x

[5] Kessing, L.V., Gerds, T.A., Feldt-Rasmussen, B., Andersen, P.K. and Licht, R.W. (2015) Use of Lithium and Anticonvulsants and the Rate of Chronic Kidney Disease: A Nationwide Population-Based Study. *JAMA Psychiatry*, **72**, 1182-1191. https://doi.org/10.1001/jamapsychiatry.2015.1834

[6] Post, R.M., Altshuler, L.L., Leverich, G.S., Frye, M.A., Nolen, W.A., Kupka, R.W., *et al.* (2006) Mood Switch in Bipolar Depression: Comparison of Adjunctive Venlafaxine, Bupropion and Sertraline. *British Journal of Psychiatry*, **189**, 124-131. https://doi.org/10.1192/bjp.bp.105.013045

[7] Miola, A., Tondo, L., Pinna, M., Contu, M. and Baldessarini, R.J. (2023) Characteristics of Rapid Cycling in 1261 Bipolar Disorder Patients. *International Journal of Bipolar Disorders*, **11**, Article No. 21. https://doi.org/10.1186/s40345-023-00300-z

[8] Judd, L.L., Akiskal, H.S., Schettler, P.J., Coryell, W., Maser, J., Rice, J.A., *et al.* (2003) The Comparative Clinical Phenotype and Long Term Longitudinal Episode Course

of Bipolar I and II: A Clinical Spectrum or Distinct Disorders? *Journal of Affective Disorders*, **73**, 19-32. https://doi.org/10.1016/s0165-0327(02)00324-5

[9] Yatham, L.N., Kennedy, S.H., Parikh, S.V., Schaffer, A., Bond, D.J., Frey, B.N., *et al.* (2018) Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 Guidelines for the Management of Patients with Bipolar Disorder. *Bipolar Disorders*, **20**, 97-170. https://doi.org/10.1111/bdi.12609

[10] Goodwin, G., Haddad, P., Ferrier, I., Aronson, J., Barnes, T., Cipriani, A., *et al.* (2016) Evidence-Based Guidelines for Treating Bipolar Disorder: Revised Third Edition Recommendations from the British Association for Psychopharmacology. *Journal of Psychopharmacology*, **30**, 495-553. https://doi.org/10.1177/0269881116636545

[11] American Psychiatric Association (2010) Practice Guideline for the Treatment of Patients with Bipolar Disorder. 2nd Edition, APA Publishing.

[12] Malhi, G.S., Bassett, D., Boyce, P., Bryant, R., Fitzgerald, P.B., Fritz, K., *et al.* (2015) Royal Australian and New Zealand College of Psychiatrists Clinical Practice Guidelines for Mood Disorders. *Australian & New Zealand Journal of Psychiatry*, **49**, 1087-1206. https://doi.org/10.1177/0004867415617657

[13] Phillips, M.L. and Swartz, H.A. (2014) A Critical Appraisal of Neuroimaging Studies of Bipolar Disorder: Toward a New Conceptualization of Underlying Neural Circuitry and a Road Map for Future Research. *American Journal of Psychiatry*, **171**, 829-843. https://doi.org/10.1176/appi.ajp.2014.13081008

[14] Kim, M.J., Brown, A.C., Mattek, A.M., Chavez, S.J., Taylor, J.M., Palmer, A.L., *et al.* (2016) The Inverse Relationship between the Microstructural Variability of Amygdala-Prefrontal Pathways and Trait Anxiety Is Moderated by Sex. *Frontiers in Systems Neuroscience*, **10**, Article 93. https://doi.org/10.3389/fnsys.2016.00093

[15] Townsend, J.D., Torrisi, S.J., Lieberman, M.D., Sugar, C.A., Bookheimer, S.Y. and Altshuler, L.L. (2013) Frontal-Amygdala Connectivity Alterations during Emotion Downregulation in Bipolar I Disorder. *Biological Psychiatry*, **73**, 127-135. https://doi.org/10.1016/j.biopsych.2012.06.030

[16] Jiang, S., Li, H., Liu, L., Yao, D. and Luo, C. (2022) Voxel-Wise Functional Connectivity of the Default Mode Network in Epilepsies: A Systematic Review and Meta-Analysis. *Current Neuropharmacology*, **20**, 254-266. https://doi.org/10.2174/1570159x19666210325130624

[17] Becker, H.C., Norman, L.J., Yang, H., Monk, C.S., Phan, K.L., Taylor, S.F., *et al.* (2021) Disorder-Specific Cingulo-Opercular Network Hyperconnectivity in Pediatric OCD Relative to Pediatric Anxiety. *Psychological Medicine*, **53**, 1468-1478. https://doi.org/10.1017/s0033291721003044

[18] Öngür, D., Lundy, M., Greenhouse, I., Shinn, A.K., Menon, V., Cohen, B.M., *et al.* (2010) Default Mode Network Abnormalities in Bipolar Disorder and Schizophrenia. *Psychiatry Research*: *Neuroimaging*, **183**, 59-68. https://doi.org/10.1016/j.pscychresns.2010.04.008

[19] Vieira, S., Pinaya, W.H.L. and Mechelli, A. (2017) Using Deep Learning to Investigate the Neuroimaging Correlates of Psychiatric and Neurological Disorders: Methods and Applications. *Neuroscience & Biobehavioral Reviews*, **74**, 58-75. https://doi.org/10.1016/j.neubiorev.2017.01.002

[20] Arbabshirani, M.R., Plis, S., Sui, J. and Calhoun, V.D. (2017) Single Subject Prediction of Brain Disorders in Neuroimaging: Promises and Pitfalls. *NeuroImage*, **145**, 137-165. https://doi.org/10.1016/j.neuroimage.2016.02.079

[21] Pinaya, W.H.L., Gadelha, A., Doyle, O.M., Noto, C., Zugman, A., Cordeiro, Q., *et al.* (2016) Using Deep Belief Network Modelling to Characterize Differences in Brain Morphometry in Schizophrenia. *Scientific Reports*, **6**, Article No. 38897. https://doi.org/10.1038/srep38897

[22] Suk, H., Lee, S. and Shen, D. (2014) Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis. *NeuroImage*, **101**, 569-582. https://doi.org/10.1016/j.neuroimage.2014.06.077

[23] Odusami, M., Maskeliūnas, R., Damaševičius, R. and Krilavičius, T. 2021 () Analysis of Features of Alzheimer's Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned ResNet18 Network. *Diagnostics*, **11**, 1071. https://doi.org/10.3390/diagnostics11061071

[24] Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., *et al.* (2017) Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Human Brain Mapping*, **38**, 5391-5420. https://doi.org/10.1002/hbm.23730

[25] Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P. and Lance, B.J. (2018) EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain-Computer Interfaces. *Journal of Neural Engineering*, **15**, Article ID: 056013. https://doi.org/10.1088/1741-2552/aace8c

[26] Huang, C., Chen, W. and Cao, G. (2019) Automatic Epileptic Seizure Detection via Attention-Based CNN-BIRNN. 2019 *IEEE International Conference on Bioinformatics and Biomedicine* (*BIBM*), San Diego, 18-21 November 2019, 660-663. https://doi.org/10.1109/bibm47256.2019.8983420

[27] Xi, Y., Chen, Y., Meng, T., Lan, Z. and Zhang, L. (2025) Depression Detection Based on the Temporal-Spatial-Frequency Feature Fusion of EEG. *Biomedical Signal Processing and Control*, **100**, Article ID: 106930. https://doi.org/10.1016/j.bspc.2024.106930

[28] Jan, Z., AI-Ansari, N., Mousa, O., Abd-Alrazaq, A., Ahmed, A., Alam, T., *et al.* (2021) The Role of Machine Learning in Diagnosing Bipolar Disorder: Scoping Review. *Journal of Medical Internet Research*, **23**, e29749. https://doi.org/10.2196/29749

[29] Gijsman, H.J., Geddes, J.R., Rendell, J.M., Nolen, W.A. and Goodwin, G.M. (2004) Antidepressants for Bipolar Depression: A Systematic Review of Randomized, Controlled Trials. *American Journal of Psychiatry*, **161**, 1537-1547. https://doi.org/10.1176/appi.ajp.161.9.1537

[30] Bahji, A., Ermacora, D., Stephenson, C., *et al.* (2020) Comparative Efficacy and Tolerability of Pharmacological Treatments for the Treatment of Acute Bipolar Depression: A Systematic Review and Network Meta-Analysis. *Journal of affective disorders*, **269**, 154-184. https://doi.org/10.1016/j.jad.2020.03.030

[31] Pacchiarotti, I., Bond, D.J., Baldessarini, R.J., Nolen, W.A., Grunze, H., Licht, R.W., *et al.* (2013) The International Society for Bipolar Disorders (ISBD) Task Force Report on Antidepressant Use in Bipolar Disorders. *American Journal of Psychiatry*, **170**, 1249-1262. https://doi.org/10.1176/appi.ajp.2013.13020185

[32] Sachs, G.S., Nierenberg, A.A., Calabrese, J.R., Marangell, L.B., Wisniewski, S.R., Gyulai, L., *et al.* (2007) Effectiveness of Adjunctive Antidepressant Treatment for Bipolar Depression. *New England Journal of Medicine*, **356**, 1711-1722. https://doi.org/10.1056/nejmoa064135

[33] Goldberg, J.F. (2019) Complex Combination Pharmacotherapy for Bipolar Disorder: Knowing When Less Is More or More Is Better. *Focus*, **17**, 218-231.

https://doi.org/10.1176/appi.focus.20190008

[34] Miklowitz, D.J., Otto, M.W., Frank, E., Reilly-Harrington, N.A., Wisniewski, S.R., Kogan, J.N., *et al.* (2007) Psychosocial Treatments for Bipolar Depression. *Archives of General Psychiatry*, **64**, 419-427. https://doi.org/10.1001/archpsyc.64.4.419

[35] Richardson, T.H. (2013) Substance Misuse in Depression and Bipolar Disorder: A Review of Psychological Interventions and Considerations for Clinical Practice. *Mental Health and Substance Use*, **6**, 76-93. https://doi.org/10.1080/17523281.2012.680485

[36] Fortinguerra, S., Sorrenti, V., Giusti, P., Zusso, M. and Buriani, A. (2019) Pharmacogenomic Characterization in Bipolar Spectrum Disorders. *Pharmaceutics*, **12**, Article 13. https://doi.org/10.3390/pharmaceutics12010013

[37] Akiskal, H.S., Hantouche, E., Allilaire, J., Sechter, D., Bourgeois, M.L., Azorin, J., *et al.* (2003) Validating Antidepressant-Associated Hypomania (Bipolar III): A Systematic Comparison with Spontaneous Hypomania (Bipolar II). *Journal of Affective Disorders*, **73**, 65-74. https://doi.org/10.1016/s0165-0327(02)00325-7

[38] Post, R.M., Leverich, G.S., Nolen, W.A., Kupka, R.W., Altshuler, L.L., Frye, M.A., *et al.* (2003) A Re-Evaluation of the Role of Antidepressants in the Treatment of Bipolar Depression: Data from the Stanley Foundation Bipolar Network. *Bipolar Disorders*, **5**, 396-406. https://doi.org/10.1046/j.1399-5618.2003.00065.x

[39] Chart-Pascual, J.P., Goena, J., Lara, F., Montero Torres, M., Marin Napal, J., Muñoz, R., *et al.* (2025) Understanding Social Media Discourse on Antidepressants: Unsupervised and Sentiment Analysis Using X. *European Psychiatry*, **68**, e51. https://doi.org/10.1192/j.eurpsy.2025.10

[40] Leverich, G.S., McElroy, S.L., Suppes, T., Keck, P.E., Denicoff, K.D., Nolen, W.A., *et al.* (2002) Early Physical and Sexual Abuse Associated with an Adverse Course of Bipolar Illness. *Biological Psychiatry*, **51**, 288-297. https://doi.org/10.1016/s0006-3223(01)01239-2

[41] Perlis, R.H., Ostacher, M.J., Patel, J.K., Marangell, L.B., Zhang, H., Wisniewski, S.R., *et al.* (2006) Predictors of Recurrence in Bipolar Disorder: Primary Outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *American Journal of Psychiatry*, **163**, 217-224. https://doi.org/10.1176/appi.ajp.163.2.217

[42] Kupfer, D.J., Frank, E., Grochocinski, V.J., Cluss, P.A., Houck, P.R. and Stapf, D.A. (2002) Demographic and Clinical Characteristics of Individuals in a Bipolar Disorder Case Registry. *The Journal of Clinical Psychiatry*, **63**, 120-125. https://doi.org/10.4088/jcp.v63n0206

[43] Hibar, D.P., Westlye, L.T., Doan, N.T., Jahanshad, N., Cheung, J.W., Ching, C.R.K., *et al.* (2017) Cortical Abnormalities in Bipolar Disorder: An MRI Analysis of 6503 Individuals from the ENIGMA Bipolar Disorder Working Group. *Molecular Psychiatry*, **23**, 932-942. https://doi.org/10.1038/mp.2017.73

[44] Antonioni, A., Raho, E.M., Lopriore, P., Pace, A.P., Latino, R.R., Assogna, M., *et al.* (2023) Frontotemporal Dementia, Where Do We Stand? A Narrative Review. *International Journal of Molecular Sciences*, **24**, Article 11732. https://doi.org/10.3390/ijms241411732

[45] Phillips, M.L. and Kupfer, D.J. (2013) Bipolar Disorder Diagnosis: Challenges and Future Directions. *The Lancet*, **381**, 1663-1671. https://doi.org/10.1016/s0140-6736(13)60989-7

[46] Strakowski, S.M., DelBello, M.P., Zimmerman, M.E., Getz, G.E., Mills, N.P., Ret, J., *et al.* (2002) Ventricular and Periventricular Structural Volumes in First- Versus

Multiple-Episode Bipolar Disorder. *American Journal of Psychiatry*, **159**, 1841-1847. https://doi.org/10.1176/appi.ajp.159.11.1841

[47] Morozova, A., Zorkina, Y., Abramova, O., Pavlova, O., Pavlov, K., Soloveva, K., *et al.* (2022) Neurobiological Highlights of Cognitive Impairment in Psychiatric Disorders. *International Journal of Molecular Sciences*, **23**, Article 1217. https://doi.org/10.3390/ijms23031217

[48] Hozer, F. and Houenou, J. (2016) Can Neuroimaging Disentangle Bipolar Disorder? *Journal of Affective Disorders*, **195**, 199-214. https://doi.org/10.1016/j.jad.2016.01.039

[49] Singh-Manoux, A., and Sabia, S. (2020) Facteurs de risque de la maladie d'Alzheimer et des maladies apparentées: approche parcours de vie. *Bulletin de l'Académie Nationale de Médecine*, **204**, 217-223. https://doi.org/10.1016/j.banm.2020.01.015

[50] Stewart, J.L., Bismark, A.W., Towers, D.N., *et al.* (2010) Resting Frontal EEG Asymmetry as an Endophenotype for Depression Risk: Sex-Specific Patterns of Frontal Brain Asymmetry. *Journal of Abnormal Psychology*, **119**, 502. https://doi.org/10.1037/a0019196

[51] Glazer, J.E., Kelley, N.J., Pornpattananangkul, N., Mittal, V.A. and Nusslock, R. (2018) Beyond the FRN: Broadening the Time-Course of EEG and ERP Components Implicated in Reward Processing. *International Journal of Psychophysiology*, **132**, 184-202. https://doi.org/10.1016/j.ijpsycho.2018.02.002

[52] Tas, C., Cebi, M., Tan, O., Hızlı-Sayar, G., Tarhan, N. and Brown, E.C. (2015) EEG Power, Cordance and Coherence Differences between Unipolar and Bipolar Depression. *Journal of Affective Disorders*, **172**, 184-190. https://doi.org/10.1016/j.jad.2014.10.001

[53] Aceves-Serrano, L., Neva, J.L. and Doudet, D.J. (2022) Insight into the Effects of Clinical Repetitive Transcranial Magnetic Stimulation on the Brain from Positron Emission Tomography and Magnetic Resonance Imaging Studies: A Narrative Review. *Frontiers in Neuroscience*, **16**, Article 787403. https://doi.org/10.3389/fnins.2022.787403

[54] Sun, S.T., Li, X.W., Zhu, J., *et al.* (2019) Graph Theory Analysis of Functional Connectivity in Major Depression Disorder with High-Density Resting State EEG Data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **27**, 429-439. https://doi.org/10.1109/TNSRE.2019.2894423

[55] Boness, C.L., Watts, A.L., Moeller, K.N. and Sher, K.J. (2021) The Etiologic, Theory-Based, Ontogenetic Hierarchical Framework of Alcohol Use Disorder: A Translational Systematic Review of Reviews. *Psychological Bulletin*, **147**, 1075-1123. https://doi.org/10.1037/bul0000333

[56] Perrottelli, A., Giordano, G.M., Brando, F., Giuliani, L. and Mucci, A. (2021) EEG-Based Measures in At-Risk Mental State and Early Stages of Schizophrenia: A Systematic Review. *Frontiers in Psychiatry*, **12**, Article 653642. https://doi.org/10.3389/fpsyt.2021.653642

[57] Peng, Y., Huang, Y., Chen, B., He, M., Jiang, L., Li, Y., *et al.* (2022) Electroencephalographic Network Topologies Predict Antidepressant Responses in Patients with Major Depressive Disorder. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **30**, 2577-2588. https://doi.org/10.1109/tnsre.2022.3203073

[58] Sasse, L., Larabi, D.I., Omidvarnia, A., Jung, K., Hoffstaedter, F., Jocham, G., *et al.* (2023) Intermediately Synchronised Brain States Optimise Trade-Off between Subject Specificity and Predictive Capacity. *Communications Biology*, **6**, Article No. 705. https://doi.org/10.1038/s42003-023-05073-w

[59] Kulynych, J.J., Luevano, L.F., Jones, D.W. and Weinberger, D.R. (1997) Cortical Abnormality in Schizophrenia: An *in Vivo* Application of the Gyrification Index. *Biological Psychiatry*, **41**, 995-999. https://doi.org/10.1016/s0006-3223(96)00292-2

[60] Xie, X., Mulej Bratec, S., Schmid, G., Meng, C., Doll, A., Wohlschläger, A., *et al.* (2016) How Do You Make Me Feel Better? Social Cognitive Emotion Regulation and the Default Mode Network. *NeuroImage*, **134**, 270-280. https://doi.org/10.1016/j.neuroimage.2016.04.015

[61] Oertel-Knöchel, V., Reuter, J., Reinke, B., Marbach, K., Feddern, R., Alves, G., *et al.* (2015) Association between Age of Disease-Onset, Cognitive Performance and Cortical Thickness in Bipolar Disorders. *Journal of Affective Disorders*, **174**, 627-635. https://doi.org/10.1016/j.jad.2014.10.060

[62] Zeng, L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., *et al.* (2018) Multi-site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI. *eBioMedicine*, **30**, 74-85. https://doi.org/10.1016/j.ebiom.2018.03.017

[63] Gallo, S., El-Gazzar, A., Zhutovsky, P., Thomas, R.M., Javaheripour, N., Li, M., *et al.* (2023) Functional Connectivity Signatures of Major Depressive Disorder: Machine Learning Analysis of Two Multicenter Neuroimaging Studies. *Molecular Psychiatry*, **28**, 3013-3022. https://doi.org/10.1038/s41380-023-01977-5

[64] Cardoner, N., Andero, R., Cano, M., Marin-Blasco, I., Porta-Casteràs, D., Serra-Blasco, M., *et al.* (2024) Impact of Stress on Brain Morphology: Insights into Structural Biomarkers of Stress-Related Disorders. *Current Neuropharmacology*, **22**, 935-962. https://doi.org/10.2174/1570159x21666230703091435

[65] Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., *et al.* (2016) Resting-state Connectivity Biomarkers Define Neurophysiological Subtypes of Depression. *Nature Medicine*, **23**, 28-38. https://doi.org/10.1038/nm.4246

[66] Blanchard, D.C. and Blanchard, R.J. (2003) What Can Animal Aggression Research Tell Us about Human Aggression? *Hormones and Behavior*, **44**, 171-177. https://doi.org/10.1016/s0018-506x(03)00133-8

[67] Perna, G., Alciati, A., Daccò, S., Grassi, M. and Caldirola, D. (2020) Personalized Psychiatry and Depression: The Role of Sociodemographic and Clinical Variables. *Psychiatry Investigation*, **17**, 193-206. https://doi.org/10.30773/pi.2019.0289

[68] Koutsouleris, N., Kahn, R.S., Chekroud, A.M., Leucht, S., Falkai, P., Wobrock, T., *et al.* (2016) Multisite Prediction of 4-Week and 52-Week Treatment Outcomes in Patients with First-Episode Psychosis: A Machine Learning Approach. *The Lancet Psychiatry*, **3**, 935-946. https://doi.org/10.1016/s2215-0366(16)30171-7

[69] Tsikandilakis, M., Bali, P., Yu, Z., Karlis, A., Tong, E.M.W., Milbank, A., *et al.* (2023) "The Many Faces of Sorrow": An Empirical Exploration of the Psychological Plurality of Sadness. *Current Psychology*, **43**, 3999-4015. https://doi.org/10.1007/s12144-023-04518-z

[70] Uddin, L.Q., Yeo, B.T.T. and Spreng, R.N. (2019) Towards a Universal Taxonomy of Macro-Scale Functional Human Brain Networks. *Brain Topography*, **32**, 926-942. https://doi.org/10.1007/s10548-019-00744-6

[71] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., *et al.* (2020) Generative Adversarial Networks. *Communications of the ACM*, **63**, 139-144. https://doi.org/10.1145/3422622

[72] Hartmann, K.G., Schirrmeister, R.T. and Ball, T. (2018) EEG-GAN: Generative Adversarial Networks for Electroencephalographic (EEG) Brain Signals. arXiv: 1806.01875.

[73] Hassan, J., Reza, S., Ahmed, S.U., Anik, N.H. and Khan, M.O. (2025) EEG Work-

load Estimation and Classification: A Systematic Review. *Journal of Neural Engineering*, **22**, Article ID: 051003. https://doi.org/10.1088/1741-2552/ad705e

[74] Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., *et al.* (2016) Cross-Trial Prediction of Treatment Outcome in Depression: A Machine Learning Approach. *The Lancet Psychiatry*, **3**, 243-250. https://doi.org/10.1016/s2215-0366(15)00471-x

[75] Rajpurkar, P., Chen, E., Banerjee, O. and Topol, E.J. (2022) AI in Health and Medicine. *Nature Medicine*, **28**, 31-38. https://doi.org/10.1038/s41591-021-01614-0

[76] Topol, E.J. (2019) High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, **25**, 44-56. https://doi.org/10.1038/s41591-018-0300-7

[77] Bzdok, D. and Meyer-Lindenberg, A. (2018) Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry*: *Cognitive Neuroscience and Neuroimaging*, **3**, 223-230. https://doi.org/10.1016/j.bpsc.2017.11.007

[78] Insel, T.R. (2010) Rethinking schizophrenia. *Nature*, **468**, 187-193. https://doi.org/10.1038/nature09552

[79] Claude, L.-A., Houenou, J., Duchesnay, E. and Favre, P. (2020) Will Machine Learning Applied to Neuroimaging in Bipolar Disorder Help the Clinician? A Critical Review and Methodological Suggestions. *Bipolar Disorders*, **22**, 334-355. https://doi.org/10.1111/bdi.12895

[80] Pan, Y., Wang, P., Xue, B., Liu, Y., Shen, X., Wang, S., *et al.* (2025) Machine Learning for the Diagnosis Accuracy of Bipolar Disorder: A Systematic Review and Meta-Analysis. *Frontiers in Psychiatry*, **15**, Article 1515549. https://doi.org/10.3389/fpsyt.2024.1515549

[81] Melhuish Beaupre, L.M., Tiwari, A.K., Gonçalves, V.F., Lisoway, A.J., Harripaul, R.S., Müller, D.J., *et al.* (2020) Antidepressant-associated Mania in Bipolar Disorder: A Review and Meta-Analysis of Potential Clinical and Genetic Risk Factors. *Journal of Clinical Psychopharmacology*, **40**, 180-185. https://doi.org/10.1097/jcp.0000000000001186

[82] Slack, D., Hilgard, A., Singh, S. and Lakkaraju, H. (2021) Reliable Post Hoc Explanations: Modeling Uncertainty in Explainability. *Advances in Neural Information Processing Systems*, **34**, 9391-9404.

[83] Amanova, N., Martin, J. and Elster, C. (2022) Explainability for Deep Learning in Mammography Image Quality Assessment. *Machine Learning*: *Science and Technology*, **3**, 025015. https://doi.org/10.1088/2632-2153/ac7a03

[84] Ghassemi, M., Oakden-Rayner, L. and Beam, A.L. (2021) The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care. *The Lancet Digital Health*, **3**, e745-e750. https://doi.org/10.1016/s2589-7500(21)00208-9

[85] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. https://doi.org/10.1038/nature14539

[86] Hanley, J.A. and McNeil, B.J. (1982) The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, **143**, 29-36. https://doi.org/10.1148/radiology.143.1.7063747

[87] Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., *et al.* (2010) Assessing the Performance of Prediction Models. *Epidemiology*, **21**, 128-138. https://doi.org/10.1097/ede.0b013e3181c30fb2

[88] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. https://doi.org/10.1023/a:1010933404324

[89] Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G.M. (2015) Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *BMJ*, **350**, g7594-g7594. https://doi.org/10.1136/bmj.g7594

[90] Vickers, A.J. and Elkin, E.B. (2006) Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making*, **26**, 565-574. https://doi.org/10.1177/0272989x06295361

[91] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. https://doi.org/10.1023/a:1022627411411

[92] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[93] Dabre, R. and Fujita, A. (2019) Recurrent Stacking of Layers for Compact Neural Machine Translation Models. *In Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 6292-6299.

[94] Devlin, J., Wang, M.W., Lee, K., *et al.* (2019) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT* 2019, Minneapolis, 2-7 June 2019, 4171-4186.

[95] Raschka, S., Patterson, J. and Nolet, C. (2020) Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information*, **11**, Article 193. https://doi.org/10.3390/info11040193

[96] Pedregosa, F., Varoquaux, G., Gramfort, A., *et al.* (2011) Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830.

[97] Nguyen, L.M., Scheinberg, K. and Takáč, M. (2020) Inexact SARAH Algorithm for Stochastic Optimization. *Optimization Methods and Software*, **36**, 237-258. https://doi.org/10.1080/10556788.2020.1818081

[98] Srivastava, N., Schölkopf, B., Hinton, G., *et al.* (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, **15**, 1929-1958.

[99] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proceedings of ICML* 2015, Lille, 6-11 July 2015, 448-456.