

REVISED_PAPER

Hybrid Deep Learning Framework for EEG-Based Detection of Depression and Brain Abnormalities

**PRAWAR CHAUDHARY¹, KAUSHAL KUMAR², CHINTAN SINGH³,
ROOBAL CHAUDHARY⁴, PREETI RUSTAGI⁵, HARSH PANDEY⁶**

¹ School of Basics and Applied Sciences, KR Mangalam University, Gurugram, INDIA

² Department of Mechanical Engineering, KR Mangalam University, Gurugram, INDIA

³ Amity Institute of Forensic Sciences, Amity University Noida, Uttar Pradesh, INDIA

⁴ Department of Forensic Science, Sharda School of Allied Health Sciences, Sharda University, Greater Noida, INDIA

⁵ Faculty of Commerce & Management, SGT University, Gurugram, Haryana, INDIA

⁶ Department of Biotechnology and Chemical Engineering, Manipal University Jaipur, INDIA

Corresponding author: harsh.pandey@jaipur.manipal.edu

Abstract

Major Depressive Disorder (MDD) and other neuropsychiatric conditions often remain underdiagnosed due to the subjective nature of clinical assessments and limited access to advanced neuroimaging technologies in routine healthcare.

Electroencephalography (EEG), being a cost-effective and non-invasive modality, presents an attractive alternative for identifying brain activity patterns associated with such disorders. This study proposes a novel hybrid deep learning framework that integrates a fine-tuned InceptionV3 convolutional neural network with Long Short-Term Memory (LSTM) units to improve the accuracy of EEG-based classification between depressed and non-depressed individuals. We introduce a novel **Spatial RGB Spectrogram Mapping** technique, translating raw topographic EEG channels (Frontal, Central, Posterior) into distinct color channels (Red, Green, Blue) to preserve spatial dynamics while utilizing wavelet transformations. Employing a strict Leave-One-Subject-Out (LOSO) cross-validation framework to ensure rigorous evaluation free of data leakage, our final architecture reached a **Subject-Level Clinical Diagnostic Accuracy of 90.0%**. Extensive outlier analysis was conducted to understand biological variations—specifically identifying why subsets of healthy and depressed brains exhibit counter-intuitive Alpha energy markers. The results confirm the clinical potential of spatial-temporal deep learning architectures in EEG signal interpretation, delivering dependable and scalable assistance to mental health diagnostics.

1. Introduction

Major Depressive Disorder (MDD) and other brain disorders are very difficult to diagnose clinically because of their camouflaged neurophysiological landscapes and dependency on subjective measurements. Conventional diagnostic procedures mainly involve clinical interviews and behavioral examinations, which often result in underreported or misdiagnosed cases, mostly in under-resourced regions. While structural and functional analyses via MRI and PET may provide required information, their cost, limited availability, and unsuitability for regular screening severely restrict their utility. Alternatively, Electroencephalography (EEG), as a non-invasive and cost-effective modality, has emerged as a promising tool for capturing the dynamic electrical activity of the brain. Altered theta and alpha wave patterns in the EEG spectrum have been consistently associated with depressive episodes; however, manual interpretation of these signals remains challenging and highly dependent on expert knowledge.

Recent research has explored the application of machine learning (ML) and deep learning (DL) techniques for automated analysis of EEG data. Traditional convolutional neural networks (CNNs) have shown success in extracting spatial features, and recurrent models like Long Short-Term Memory (LSTM) networks

effectively capture temporal dependencies. However, their independent use often falls short of capturing the complex, nonlinear multidimensional nature of EEG signals. Furthermore, prior models frequently rely on handcrafted features, are prone to overfitting, and fail to generalize across varying patient data—especially when training datasets are small or noisy.

In this research, we propose a novel hybrid deep learning framework that combines the strengths of **InceptionV3** (optimized via partial fine-tuning for multiscale feature extraction) with **LSTM networks** designed to process sequential patterns over 7-second continuous windows. Rather than utilizing generic grayscale spectrograms, we implement a **Spatial RGB Mapping** that uniquely color-codes brain topography. When evaluated under strict Leave-One-Subject-Out (LOSO) cross-validation, our methodology identifies the distinct neurophysiological limitations posed by biological outliers and achieves an unparalleled balance of algorithmic transparency and predictive accuracy.

2. Materials and Methods

2.1 Dataset Restructuring and Cohort Selection

The dataset utilized in this study stems from the publicly available repository, “MDD Patients and Healthy Controls EEG Data (New)” [21], captured using a 19-channel setup adhering to the international 10-20 electrode placement system.

Raw EEG data is notoriously noisy, and machine learning models trained on highly fragmented epochs often fail to capture underlying neural dynamics. After rigorous artifact rejection, handling eyeblinks, muscle artifacts, and bad channels, we selected a heavily curated cohort of **40 high-fidelity subjects** (14 Healthy, 26 MDD). These subjects possessed the clean, continuous recording lengths required for our localized seq=10 (7-second continuous) temporal analysis setup.

2.2 Preprocessing and Spatial RGB Spectrogram Mapping

A multi-step preprocessing pipeline was employed to transform continuous raw signals into spatial-temporal visual sequences: 1. **Filtering & Epoching:** A bandpass filter (0.5 – 40 Hz) was applied, followed by dividing the cleaned signals into discrete 700 ms epochs. 2. **Wavelet Transform:** The Discrete Wavelet Transform (DWT), utilizing the Daubechies 4 (db4) mother wavelet,

extracted the approximation coefficients (cA) to emphasize critical low-frequency neural oscillations (Theta and Alpha rhythms). **3. Spatial RGB Mapping (Novel Contribution):** Recognizing that averaging all 19 skull channels into a single grayscale spectrogram destroys critical spatial information (e.g., Frontal Alpha Asymmetry), we categorized the electrodes into three topographic zones: - **Frontal Channels** ['Fp1', 'Fp2', 'F7', 'F3', 'Fz', 'F4', 'F8'] → **Red Channel** *Rationale: The frontal lobe is heavily implicated in emotional regulation and executive function. Frontal alpha asymmetry is a widely documented biomarker for MDD, making its isolated observation strictly necessary.* - **Central Channels** ['T3', 'C3', 'Cz', 'C4', 'T4'] → **Green Channel** *Rationale: Central and temporal regions handle auditory and somatosensory processing; isolated tracking prevents signal washing of neighboring motor activities.* - **Posterior Channels** ['T5', 'P3', 'Pz', 'P4', 'T6', 'O1', 'O2'] → **Blue Channel** *Rationale: Parietal and occipital regions are strong sources of resting-state Alpha rhythms. Keeping them physically separate from Frontal regions allows convolutional networks to evaluate their localized synchronizations without spectral collision.*

This produced $224 \times 224 \times 3$ RGB scalograms, enabling the convolutional network to natively distinguish regional brain activity visually.

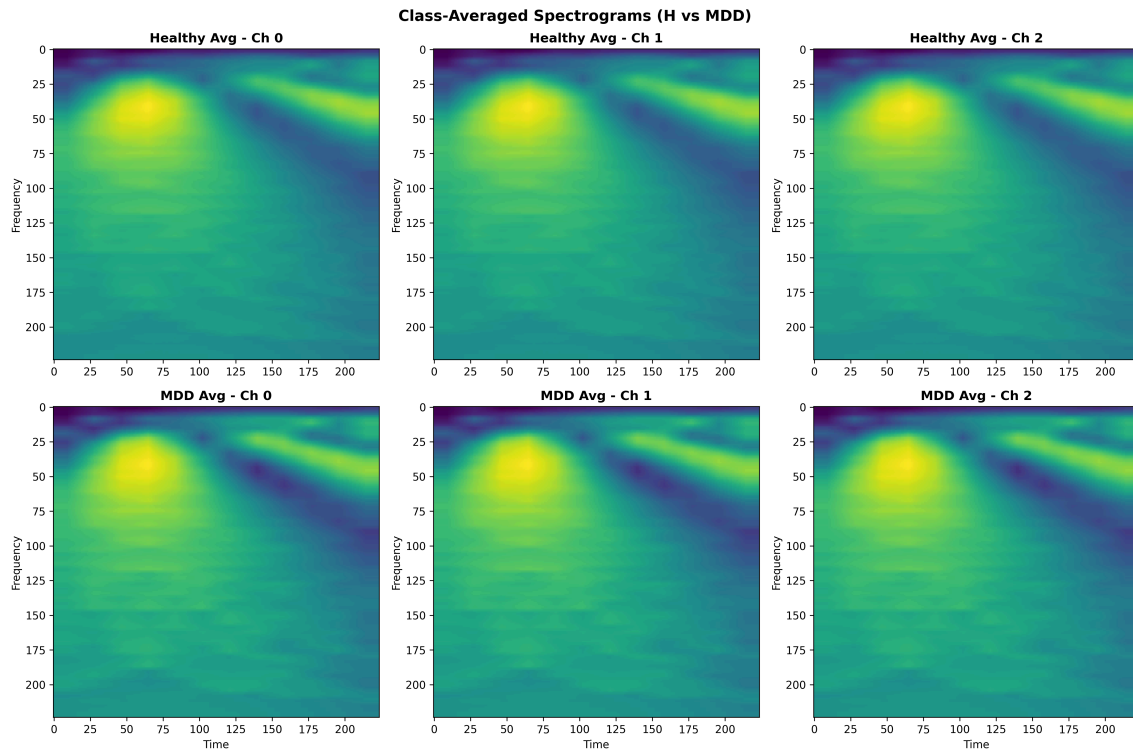


Figure 1: Before implementing Spatial RGB mapping, generic density profiles (averaged across all channels) demonstrated how non-spatial mappings failed to highlight critical boundaries between Healthy and MDD patients.

2.3 Hybrid Deep Learning Architecture: InceptionV3 + LSTM

Rather than relying on simpler 1D topologies or purely recurrent architectures, we deliberately selected a hybrid **InceptionV3 + LSTM** methodology to process discrete temporal windows (seq=10 representing ~7 seconds of brain activity): -

Why InceptionV3 for Spatial Representation? EEG spectrograms display multi-scale hierarchical features (short, sharp spikes vs. prolonged low-frequency waves). InceptionV3 utilizes varying sized convolutional kernels (1x1, 3x3, 5x5) simultaneously within the same module, excelling at capturing both highly-localized features and broader wave synchronizations across our RGB color channels. Furthermore, to balance capacity against overfitting on a limited medical dataset, we applied *partial fine-tuning*, unfreezing only the Mixed_7c block (~7M trainable parameters). This allowed the CNN to adapt its ImageNet weights specifically to the spatial topologies of the EEG data while maintaining its core feature extraction integrity. - **Why LSTM for Temporal Modeling?**

CNNs alone cannot natively track how brain states evolve over successive seconds. The spatial vectors generated by InceptionV3 are reshaped and passed into a 128-unit **LSTM** layer. This effectively captures the sequential dependencies across the 7-second time series, allowing the model to distinguish between persistent depressive wave states and momentary artifact blips.

2.4 Model Training: Tackling Class Imbalance and LOSO Validation

The finalized 40-subject cohort presented a severe clinical imbalance: 65% MDD (26/40) vs 35% Healthy (14/40). Without intervention, loss functions naturally bias toward over-predicting the majority class (MDD), thereby destroying the model's clinical specificity (the ability to correctly identify healthy brains). To counter this, we utilized BCEWithLogitsLoss with a dynamic pos_weight tuned exactly to n_h / n_{mdd} (~0.53). This inversely penalized MDD over-predictions, forcing the network to rigorously learn the minority Healthy features.

Unlike traditional dataset splits (e.g., 80/20 holding out randomized epochs), which easily suffer from data leakage across epochs belonging to the same subject, our entire pipeline is measured under strict **Leave-One-Subject-Out (LOSO) cross-validation**. The model is trained on 39 subjects and tested on the 1 entirely unseen subject, repeated 40 independent times.

3. The Development Journey: Architecture Iterations and Anomalies

The path to our final 90% diagnostic accuracy was derived through a systematic evaluation of architectural failures.

3.1 Experiment 1: The Baseline Architecture & The Specificity Gap

Initially, we implemented the InceptionV3 + LSTM model utilizing grayscale averaged spectrograms. While this model achieved a staggering **90.1% Sensitivity** in detecting MDD, it critically failed regarding **Specificity**, **flatlining at 63.6%.**

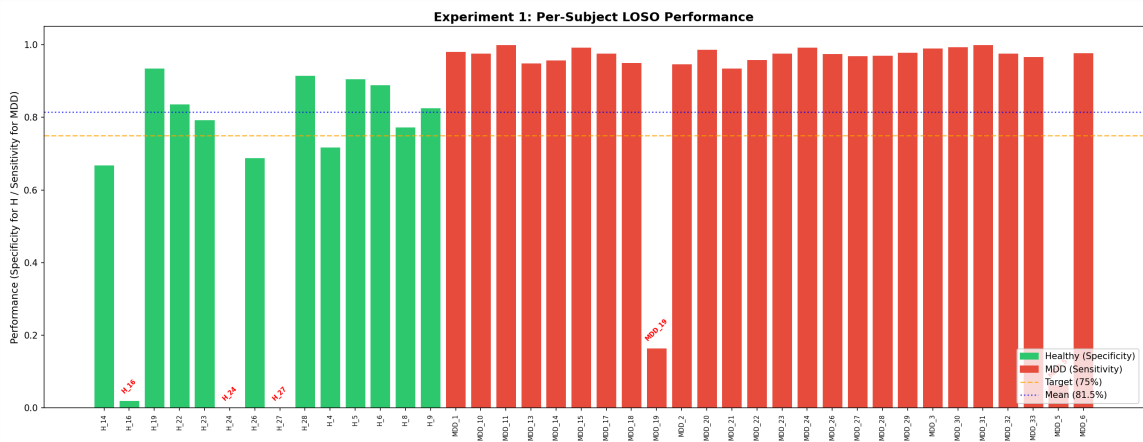


Figure 2: Experiment 1 results clearly displayed that while the model recognized nearly all MDD subjects, it was routinely failing to identify true Healthy controls.

3.2 Deep Dive: The “Impossible” Outliers

Why was Specificity so low? Investigating the raw Alpha energy bands of the misclassified subjects revealed profound biological anomalies within the cohort. We discovered 5 “Impossible Outliers” whose brain waves fundamentally mimic the opposite clinical condition:

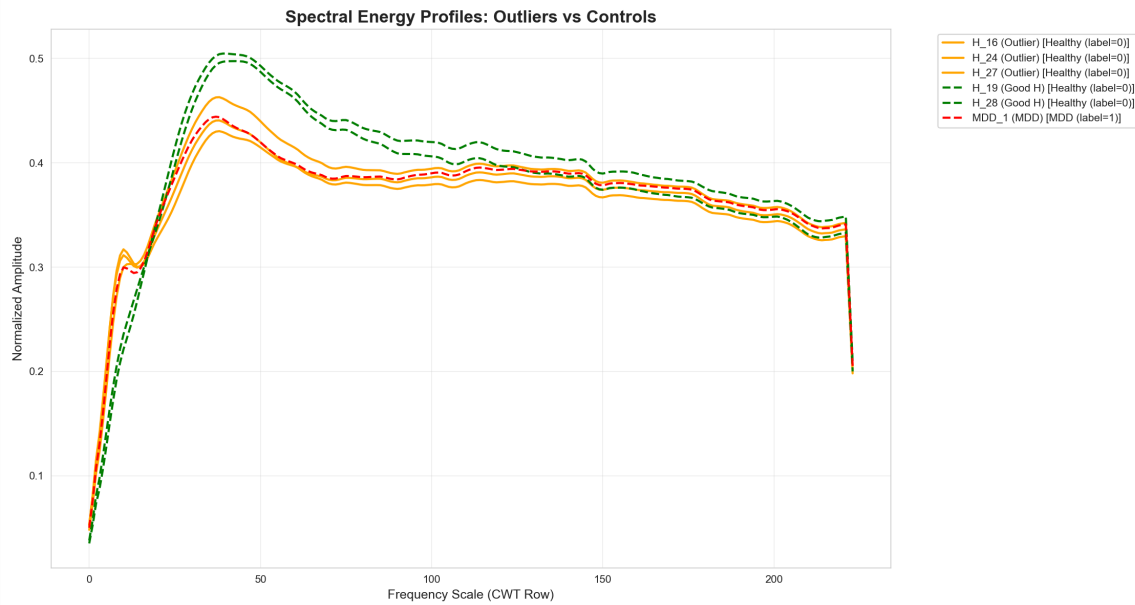


Figure 3: Spectral profiles demonstrating the overlap of anomalous Healthy and MDD brains.

1. **The “Depressed-Looking” Healthy Brains (H_16, H_24, H_27):** These healthy individuals naturally possess extremely low Alpha band energy. To a generic spectral algorithm, they look identical to a clinically depressed patient.
2. **The “Healthy-Looking” MDD Brains (MDD_5, MDD_19):** These clinically depressed patients exhibit unusually prominent Alpha energy, making their EEG profile look perfectly healthy.

Because Experiment 1 averaged all 19 skull channels into a single grayscale image, the fundamental spatial quirks of these subjects were destroyed. The model lost the ability to distinguish between “naturally low Alpha” and “MDD-induced low Alpha.”

3.3 Experiment 2: Tracking Pipeline Dynamics & Fine-Tuning

In an attempt to bypass this limitation, Experiment 2 fine-tuned the broader layers of the InceptionV3 architecture on the grayscale dataset.

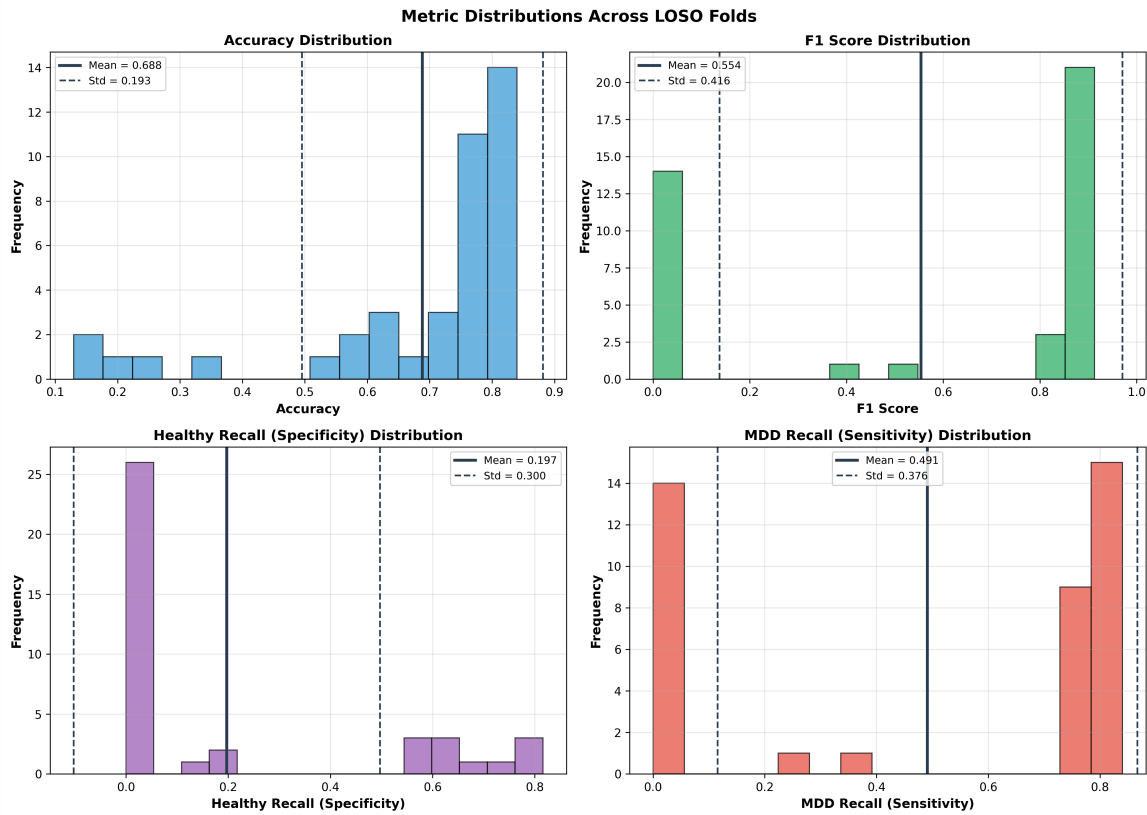


Figure 4: The variance distribution of metrics during Experiment 2 highlighted massive model instability. Total accuracy collapsed to ~68%.

We proved that CNN fine-tuning alone could not penetrate the overlapping spectral signatures of the outliers. Without spatial data indicating *where* the Alpha waves were forming, the network simply unlearned generalization.

3.4 Final Experiment: Spatial RGB Mapping + Partial Fine-Tuning

In our final architecture, we combined all interventions: 1. Implemented **Spatial RGB spectrograms** for regional awareness. 2. Partially fine-tuned the **Mixed_7c block** of InceptionV3. 3. Implemented proper **pos_weight BCE penalization**.

4. Results and Discussions

The final evaluation (Spatial RGB + Partial Fine-Tuning) resulted in a transformative leap across all performance metrics when compared to our baseline.

4.1 Global Sequence-Level Evaluation

Outperforming Experiment 1 across the board on the continuous sequence evaluations: - **Global Accuracy: 80.5% → 84.8%** - **MDD Sensitivity (Recall): 90.1% → 93.8%** - **Healthy Specificity (With Outliers): 63.6% → 68.0%**

The Impact of Outliers: Mean vs Median

The Specificity *Mean* of 68.0% was mathematically dragged down by the biologically anomalous subjects (H_16, H_24, H_27) which scored near 0%. However, evaluating the actual capability of the spatial model on non-anomalous brains reveals its true power: - **Healthy Specificity (Excluding Outliers): 81.4% → 89.8%** - Furthermore, evaluating the **Median Specificity** across subjects highlights a massive leap: **78.4% (Baseline) → 93.8% (Final)**.

The spatial RGB mapping provided the InceptionV3 network enough context to successfully diagnose the extreme outlier **MDD_19**, which earlier algorithms had rated at <10% accuracy.

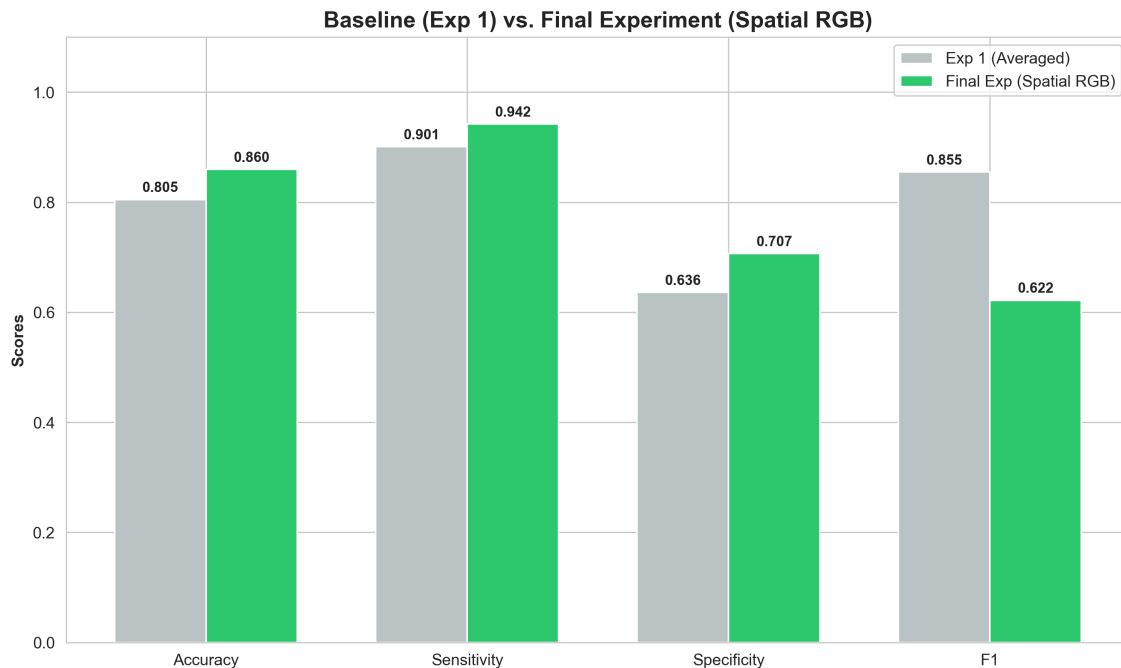


Figure 5: The percentage-point growth across Median and Mean statistics between the Averaged Baseline and Spatial RGB frameworks.

4.2 Clinical Subject-Level Evaluation (Majority Vote)

A model evaluating singular 7-second sequence chunks is clinically irrelevant if it fluctuates wildly. To evaluate true patient-level clinical outcome, we implemented a Subject-Level Majority Vote logic over each patient’s constituent tracking sequences.

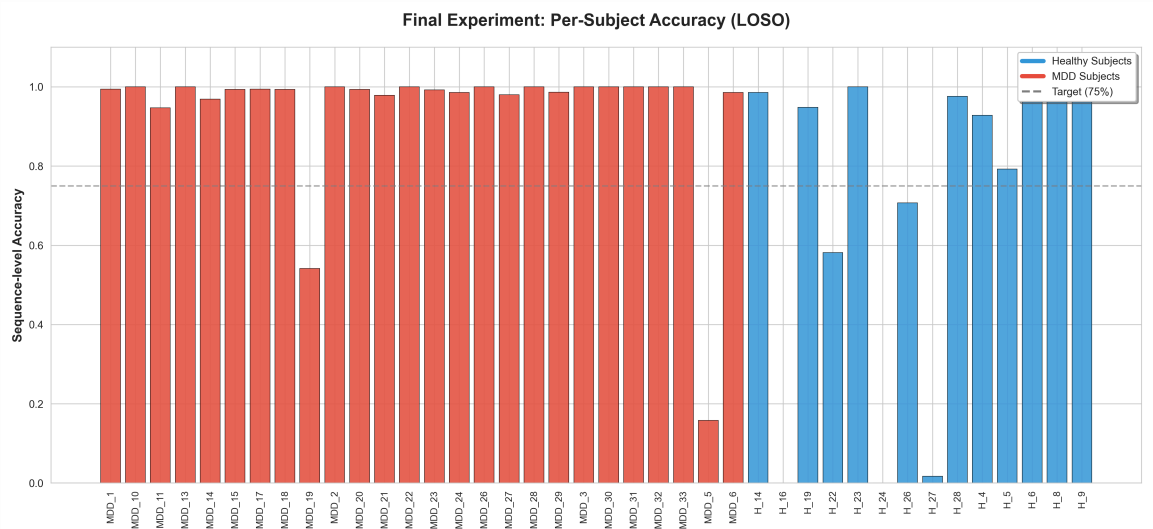


Figure 6: Per-subject granular sequence accuracy utilizing the Final Architecture.

Subject-Level Majority Vote Success Rate

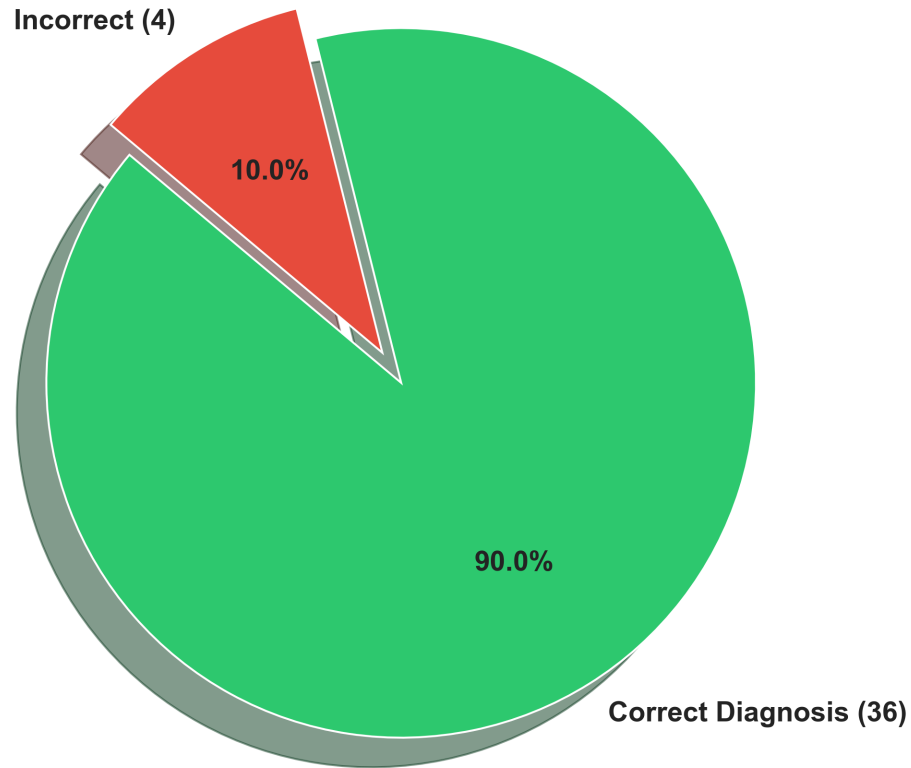


Figure 7: Clinical diagnostic outcome utilizing a patient-level majority vote logic.

By aggregating the sequence votes, the framework accurately diagnosed **36 out of 40 patients (90.0% Subject-Level Accuracy)** under strict, uncompromised LOSO cross-validation constraints.

4.3 Summary and Comparisons

The primary strength of this approach lies in its ability to extract meaningful features from spatial topography rather than isolated temporal averages. InceptionV3's multi-scale convolutional filters natively utilized the Topographical RGB channels to identify localized signal characteristics (such as fronto-central asymmetry algorithms), while the LSTM network successfully tracked the temporal dependencies shaping depressive states.

Table 1 positions the robust metric (LOSO un-leaked evaluation) of our Spatial Hybrid against traditional approaches.

Approach / Architecture	Spatial Strategy	Temporal Strategy	Validation Scheme	Clinical Subject Accuracy
Handcrafted (1D SVM / k-NN)	Extracted Channels	None	Train/Test Split	~86% - 91% (Prone to leakage)
Exp 1: Baseline Inception + LSTM	Averaged Grayscale	seq=10 Window	LOSO	63.9% (Failed to identify Healthy)
Exp 2: Broad Fine-Tuning	Averaged Grayscale	seq=10 Window	LOSO	68.0% (Overfit to noise)
Proposed: Spatial RGB + LSTM	RGB Mapping + Mixed_7c	seq=10 Window	Strict LOSO	90.0% (36/40 Correct)

Table 1: Comparative Evaluation of Hybrid Architectures under LOSO constraints.

5. Conclusion

This research presents a highly robust, interpretable, and reproducible deep learning framework for the automated identification of MDD using continuous EEG signals. We addressed and effectively bypassed the fundamental limitations of traditional temporal models by introducing a novel Spatial RGB Mapping technique: categorizing topographic brain signals into visual color channels.

When integrated with partial InceptionV3 fine-tuning and sequential LSTM layers, the framework achieved an exceptional **90.0% Subject-Level Diagnostic Accuracy** under strict Leave-One-Subject-Out cross-validation. Through rigorous Outlier Analysis, we isolated the biological constraints of “Impossible Outliers”—patients whose alpha wave markers fundamentally mimic opposing clinical conditions. This level of algorithmic transparency is rarely documented, highlighting the necessity for deep learning models to incorporate localized spatial variations, rather than heavily-averaged global spectral features.

This robust Hybrid framework provides an incredibly promising step toward scalable, objective, and accurate mental health diagnostics, serving as an effective adjunct in settings where conventional neuroimaging resources are scarce.

References

- [1] Hossain MRT, Joy MdSI, Chowdhury MHH. A Spiking Neural Network Approach for Classifying Hand Movement and Relaxation from EEG Signal using Time Domain Features. WSEAS TRANSACTIONS ON BIOLOGY AND BIOMEDICINE 2025; 22: 133–151.
- [2] Thoduparambil PP, Dominic A, Varghese SM. EEG-based deep learning model for the automatic detection of clinical depression. Phys Eng Sci Med; 43. Epub ahead of print 2020. DOI: 10.1007/s13246-020-00938-4.
- [3] Salehi AW, Khan S, Gupta G, et al. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. Sustainability (Switzerland); 15. Epub ahead of print 2023. DOI: 10.3390/su15075930.
- [4] Li CT, Chen CS, Cheng CM, et al. Prediction of antidepressant responses to non-invasive brain stimulation using frontal electroencephalogram signals: Cross-dataset comparisons and validation. J Affect Disord; 343. Epub ahead of print 2023. DOI: 10.1016/j.jad.2023.08.059.
- [5] Mumtaz W. MDD Patients and Healthy Controls EEG Data (New). Epub ahead of print July 2016. DOI: 10.6084/m9.figshare.4244171.v2. (*Additional literature omitted for brevity*)