ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)

Available Online at www.ijetcr.org

Volume 14; Issue 03; 2026; Page No. 66-70

EEG-Based Mental State Prediction Using Hybrid Machine Learning Techniques

Suhani Sharma[1], Sakshi Mishra[1], Ketan Anand[1]

[1]Dept. of CSE, Sharda University, Greater Noida, Uttar Pradesh

2023500400.suhani@ug.sharda.ac.in , 2023302424.sakshi@ug.sharda.ac.in, ketan.anand@sharda.ac.in

Conflicts of interest: Nil

Corresponding author: Ketan Anand

Abstract

This paper presents a Brain-Computer Interface (BCI) system that reads EEG signals and classifies a person's internal state as either Concentrated or Relaxed. A complete pipeline is constructed covering raw EEG data loading to predictions using machine learning. The system uses the MNE-Python library to read EDF files, applies a bandpass filter (8–30 Hz), and extracts frequency band power features from five standard EEG bands: Delta, Theta, Alpha, Beta, and Gamma. These features are used to train a Random Forest classifier. Four different models — Random Forest, SVM, LSTM, and 1D CNN — are compared to evaluate traditional machine learning against deep learning on this time-series classification problem. Results show that Random Forest achieves 93–96% accuracy on the standard task, and deep learning models reach similar levels. The work provides an open-source, reproducible framework suitable for EEG research in healthcare, education, and neuroscience.

Keywords: EEG, Brain-Computer Interface, Mental State Classification, Random Forest, LSTM, 1D CNN, Bandpower Features, Signal Processing.

Introduction

Understanding how the human brain works has always been a complex challenge. One of the more practical questions researchers are asking today is: can we detect what mental state a person is in — concentrated, relaxed, or stressed — just by reading their brain signals? [1] The answer is yes, and it has genuinely useful applications.

Electroencephalography (EEG) is one of the most widely used tools for reading brain activity. It measures electrical signals produced by neurons through sensors placed on the scalp. EEG is non-invasive, relatively affordable, and captures changes in the brain very quickly [2] — making it ideal for real-time mental state monitoring.

This paper builds a complete EEG classification system that takes raw brain signal data, processes it, extracts meaningful features, and uses machine learning to decide whether the person was concentrated or relaxed when the recording was made. Four different classification models are also compared to understand the trade-offs between simpler approaches like Random Forest and more complex deep learning methods.

A. Motivation

Current methods for measuring mental state — such as questionnaires or behavioral observation — are slow, subjective, and difficult to automate. There is a real need for objective tools that can measure cognitive states from physiological signals in real time [3]. EEG-based systems can fill this gap, but using them requires a structured pipeline that is not always straightforward to build from scratch. This work attempts to provide exactly that.

B. Objectives

The main goals of this work are to:

1. Load and preprocess EEG signals from standard EDF files; [4]

2. Extract frequency band power features using established signal processing techniques;
3. Train and evaluate multiple classifiers for binary mental state prediction; and
4. Compare traditional machine learning with deep learning approaches on EEG classification.

II. RELATED WORK

EEG-based mental state classification is a well-studied area. Several groups have shown that frequency band features — particularly Alpha and Beta power — can reliably distinguish between resting and active cognitive states. Table I summarizes the key papers that influenced this work.

Yin et al. [5] showed that the ratio of Alpha to Beta power is a strong discriminating feature for focus detection, achieving 88% accuracy with an SVM classifier. Subasi et al. [4] found that Random Forest consistently outperforms SVM and KNN on EEG tasks. Roy et al. [7] reviewed deep learning models and found that LSTM and CNN architectures outperform traditional ML when enough data is available. Our work draws from all of these findings and brings them together into a single, unified pipeline.

III. SYSTEM DESIGN AND METHODOLOGY

The system is divided into five sequential stages: data loading, preprocessing, feature extraction, model training, and prediction. Each stage is implemented as a separate Python module, making it easy to modify or replace individual components without disrupting the rest of the pipeline.

A. EEG Data Acquisition

The Physionet EEG Motor Movement/Imagery Dataset was used, which provides 64-channel EEG recordings in EDF (European Data Format) files sampled at 160 Hz. The data was loaded using the MNE-Python library [5], which reads EDF files and provides easy access to channel data, timestamps, and metadata such as sampling frequency.

Figure 1 Electrodes For collecting Raw EEG

B. Preprocessing

Raw EEG signals contain a substantial amount of noise — low-frequency drift from body movement, high-frequency noise from muscle activity, and other artifacts. A zero- phase bandpass filter between 8 Hz and 30 Hz was applied to retain only the frequency range that carries neurologically meaningful information (the Alpha and Beta bands). After filtering, the signal was segmented into fixed one-second windows for feature extraction.

C. Feature Extraction

For each one-second segment, the power in five standard EEG frequency bands was calculated using Welch's Power Spectral Density (PSD) method. [1] Welch's method reduces noise in spectral estimates by averaging multiple overlapping periodograms. The five bands used are Listed Below in table:

Table I Bands with their Frequency Range

Frequency	Range
Delta	0.5-04 Hz
Theta	04-08 Hz
Alpha	08-13 Hz
Beta	13-30 Hz
Gamma	30-45 Hz

This yields a 5-dimensional feature vector per window per channel. Bandpower is calculated by numerically integrating the PSD over each band's frequency range.

D. Marker Generation

Since the Physionet dataset was not originally designed for mental state labeling, a rule-based approach was used. If Alpha power (8–13 Hz) was greater than [6] Beta power (13–30 Hz) for a given window, the label was 0 (Relaxed); otherwise, the label was 1 (Concentrated). This approach is grounded in established neuroscience — Alpha suppression is a known marker of increased cognitive engagement [8].

E. Machine Learning Models

A Random Forest classifier (100 trees) was trained on the bandpower feature vectors. The trained model was saved using Joblib for reuse without retraining. An SVM with an RBF kernel was also tested on the same features for comparison.

F. Deep Learning Models

Two deep learning architectures were implemented in PyTorch and evaluated on a synthetic dataset of sinusoidal and square-wave sequences representing two distinct signal patterns.

1. LSTM Model: Uses a single LSTM layer with 32 hidden units, followed by a fully connected output layer. The LSTM processes the input sequence step by step, making [5] it well-suited to capture patterns in time-series data.
2. 1D CNN Model: Uses two convolutional layers (16 and 32 filters) with ReLU activations and MaxPooling, followed by a fully connected classifier. The CNN learns local temporal patterns by applying learned filters along the time axis.

Figure 2 Model Accuracy

Both models were trained for 10 epochs using the Adam optimizer and cross-entropy loss.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

The EEG pipeline was run on subject S001R01 from the Physionet dataset, a 64-channel recording at 160 Hz. For the deep learning comparison, 2000 synthetic time-series samples (1000 sinusoidal, 1000 square-wave) of 50 timesteps each were generated. [7] All models were trained on an 80/20 train-test split. All code was written in Python 3.8 and run in Google Colab.

B. Results

Table II shows the accuracy of all four models on the time- series classification task. Note that deep learning models were evaluated only on synthetic data in this study, while Random Forest was also tested on real EEG features.

Table II Models With their Accuracy

Model	Architecture	Data Type	Accuracy (%)
Random Forest	Ensemble Trees	Band Features	93–96
SVM (RBF)	Kernel Classifier	Band Features	90–94
LSTM	Recurrent Net	Synthetic TS	88–95
1D CNN	Conv Network	Synthetic TS	90–96

Random Forest achieved the highest accuracy on real EEG bandpower features. On the synthetic task, all four models performed comparably, with the 1D CNN achieving a slight edge in some runs. Exact figures vary across random seeds and training runs, which is expected with small datasets.

C. Temporal Prediction

The trained Random Forest model was applied to the full EEG recording using a sliding one-second window. The resulting prediction sequence showed a consistent state of Concentrated (label 1) across the entire recording. This is expected because the Physionet motor imagery dataset primarily captures active cognitive [8] states. A separate resting-state recording would be needed to observe transitions between Concentrated and Relaxed.

Figure 3 Graphs predicting Mental states

D. Signal Processing Validation

FFT analysis before and after filtering confirmed that the bandpass filter effectively removed out-of-band noise. The filtered signal retained the 8–30 Hz components and showed significantly reduced low-frequency drift, clearly [6] [9] visible in the spectral plots.

Figure 4 Signal Processing

Figure 5 Complete EEG Mental State Classification Pipeline.

V. DISCUSSION

The results confirm that EEG-based mental state classification using frequency band features is both practical and reliable. The Alpha-to-Beta power ratio is a simple but effective feature for distinguishing focused versus relaxed states, and it is well-supported by decades of neuroscience research.

Random Forest performed best among the four models, which aligns with findings from Subasi et al. [4]. It is robust, does not require large quantities of data, and provides easily interpretable feature importance scores. However, deep learning models [10] offer a key advantage: they can operate directly on raw or minimally processed[11] time-series data, without any manual feature engineering. For large-scale or real-time deployments, this makes LSTM and CNN models more attractive.

The most significant limitation of this study is marker quality. The Physionet dataset was designed for motor imagery classification, not cognitive state monitoring. The rule-based markers derived [12]from the Alpha/Beta ratio are physiologically reasonable but have not been validated against subjective self-reports or behavioral measures. Future studies should use datasets specifically designed for cognitive state labeling, such as SEED or MAHNOB-HCI.

VI. CONCLUSION

This paper presented a complete, modular EEG-based mental state classification system. The pipeline covers everything from raw signal loading to per-second mental state predictions, implemented entirely using open-source tools. Bandpower features from the [13] Alpha and Beta frequency bands were shown to be effective for distinguishing focused and relaxed states, and four classification models were compared to understand their trade-offs.[14] [13]

The system is straightforward to reproduce on standard hardware using publicly available EEG datasets, providing a useful starting point for researchers who wish to explore BCI applications without building everything from scratch.

Future work will focus on real-time streaming with live EEG hardware, multi-class mental state detection, and applying advanced models [9] such as EEGNet and Transformers to real EEG data with proper ground-truth markers.

VII. REFERENCES

[1] A. K. Singh and S. Krishnan, “Trends in EEG signal feature extraction applications,” Front. Artif. Intell., vol. 5, p. 1072801, Jan. 2023, doi: 10.3389/frai.2022.1072801.

[2] J. A. Mendivil Sauceda, B. Y. Marquez, and J. J. Esqueda Elizondo, “Emotion Classification from Electroencephalographic Signals Using Machine Learning,” Brain Sci., vol. 14, no. 12, p. 1211, Nov. 2024, doi: 10.3390/brainsci14121211.

[3] A. Hoffsommer, H. Schneider, S. Pavlitska, and J. M. Zöllner, “DEAP DIVE: Dataset Investigation with Vision transformers for EEG evaluation,” Oct. 01, 2025, arXiv: arXiv:2510.00725. doi: 10.48550/arXiv.2510.00725.

[4] S. A. H. Hussain, I. Raza, S. A. Hussain, M. H. Jamal, T. Gulrez, and A. Zia, “A mental state aware brain computer interface for adaptive control of electric powered wheelchair,” Sci. Rep., vol. 15, no. 1, p. 9880, Mar. 2025, doi: 10.1038/s41598-024-82252-7.

[5] A. Raza and M. Z. Yusoff, “Development of a CNN-LSTM Deep Learning Model for Motor Imagery EEG Classification for BCI Applications,” Eng. Technol. Appl. Sci. Res., vol. 15, no. 3, pp. 22705–22711, Jun. 2025, doi: 10.48084/etasr.9945.

[6] I. H. Elshekhidris, M. B. MohamedAmien, and A. Fragoon, “WAVELET TRANSFORMS FOR EEG

SIGNAL DENOISING AND DECOMPOSITION,” Int. J. Adv. SIGNAL IMAGE

Sci., vol. 9, no. 2, pp. 11–28, Dec. 2023, doi: 10.29284/IJASIS.9.2.2023.11-28.

[7] X. Geng et al., "A method of EEG signal feature extraction based on hybrid DWT and EMD," Alex. Eng. J., vol. 113, pp. 195–204, Feb. 2025, doi: 10.1016/j.aej.2024.11.056.

[8] K. M. Hossain, Md. A. Islam, S. Hossain, A. Nijholt, and M. A. R. Ahad, "Status of deep learning for EEG-based brain–computer interface applications," Front. Comput. Neurosci., vol. 16, p. 1006763, Jan. 2023, doi: 10.3389/fncom.2022.1006763.

[9] U. Talukdar and P. Paul, "OnE: An EEG-based Passive BCI framework for Monitoring Cognitive States During online learning," in Advances In Robotics - 6th International Conference of The Robotics Society, Ropar India: ACM, Jul. 2023, pp. 1–7. doi: 10.1145/3610419.3610422.

[10] "J. E. Moedano-Atristain, Ó. Yáñez-Suárez and E. Bojorges-Valdez, 'Continuous Mental State Estimation Using EEG Band Power Time Series Predictions for BCI Control,' 2023 IEEE EMBS R9 Conference, Guadalajara, Mexico, 2023, pp. 1-4, doi: 10.1109/IEEECONF60929.2023.10525603.,"

[11] M. Zeynali, H. Seyedarabi, and R. Afrouzian, "Classification of EEG signals using Transformer based deep learning and ensemble models," Biomed. Signal Process. Control, vol. 86, p. 105130, Sep. 2023, doi: 10.1016/j.bspc.2023.105130.

[12] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of Artifacts from EEG Signals: A Review," Sensors, vol. 19, no. 5, p. 987, Feb. 2019, doi: 10.3390/s19050987.

[13] M. Y. M. Naser and S. Bhattacharya, "Resting EEG State, an Insight into Motor Imagery Signal Characteristics," in 2023 15th Biomedical Engineering International Conference (BMEiCON), Tokyo, Japan: IEEE, Oct. 2023, pp. 1–5. doi: 10.1109/BMEiCON60347.2023.10321821.

[14] "G. Kaur, M. Gupta and R. Kumar, 'Classifying Human Emotions through EEG data with Machine Learning,' 2024 International Conference Automatics and Informatics (ICAI), Varna, Bulgaria, 2024, pp. 632-636, doi: 10.1109/ICAI63388.2024.10851616.,"