NeuroSentiScan: EEG-Based Stress Detection Using Classical and Deep Learning Techniques

Rakshak Datta[1], Sanidhya Singh[2], and Dr. Prashant Vats[3]

[1] CSE Core, Student (2021–2025)
rakshakdutta2826@gmail.com
[2] CSE IoT & IS (Hons), Student (2022–2026)
sanidhya11719@gmail.com
[3] Assistant Professor, Manipal University Jaipur
prashant.vats@manipal.edu

Abstract. NeuroSentiScan presents a robust offline framework for EEG-based *stress classification*, enabling discrimination among three cognitive states: *Stressed*, *Normal*, and *Relaxed*. Leveraging a dataset of 2132 EEG samples with 2548 engineered features, we benchmark six classical machine learning (ML) models—Support Vector Machine, Random Forest, K-Nearest Neighbors, Naive Bayes, Logistic Regression, and Decision Tree—alongside six deep learning (DL) architectures: CNN, LSTM, CNN-LSTM, 1D CNN, BiLSTM, and EEGNet. Feature representations are derived using statistical, entropy-based, and wavelet-domain methods. Evaluation includes accuracy, F1-score, confusion matrices, and feature space visualization via PCA and t-SNE. Interpretability is incorporated through LIME, and predictive robustness is enhanced with ensemble majority voting. Statistical rigor is ensured through paired t-tests and McNemar's tests. Results highlight the superiority of hybrid DL architectures and ensemble strategies for multiclass EEG stress recognition, establishing NeuroSentiScan as a reproducible and scalable benchmark for offline cognitive state monitoring in real-world scenarios.

Keywords: EEG · Machine Learning · Deep Learning · Stress Detection · EEG

1 Introduction

Electroencephalography (EEG)-based stress detection plays a pivotal role in modern affective computing and cognitive monitoring, enabling applications in mental health assessment, neurofeedback, adaptive learning, and human-computer interaction. Owing to its high temporal resolution and non-invasive nature, EEG remains a preferred modality for decoding internal emotional states. However, signal non-stationarity, inter-subject variability, and a lack of standardized evaluation protocols continue to challenge model robustness and reproducibility.

Many existing systems rely on binary classification or real-time hardware integration, which limits generalizability and deployment in offline or resource-constrained environments. Additionally, insufficient emphasis on interpretability

and statistical rigor further hampers the adoption of these models in high-stakes domains like clinical diagnostics or workplace wellness.

To overcome these gaps, we propose **NeuroSentiScan**—a fully offline, interpretable, and statistically validated framework for EEG-based multiclass stress classification across three cognitive states: *Stressed*, *Normal*, and *Relaxed*. Our system integrates classical machine learning (ML) and deep learning (DL) models within a unified, modular pipeline focused on accessibility and empirical rigor.

The framework begins with EEG preprocessing, including filtering, normalization, and optional downsampling. We extract 2548-dimensional handcrafted features encompassing statistical, entropy-based, and wavelet-domain descriptors, which train six ML classifiers: SVM, Random Forest, KNN, Naive Bayes, Logistic Regression, and Decision Tree. In parallel, DL models—1D CNN, LSTM, CNN-LSTM, BiLSTM, EEGNet, and a compact CNN variant—are trained on reshaped inputs (e.g., 49×52) that simulate temporal EEG patterns.

All models use stratified sampling, early stopping, and accuracy tracking. Evaluation metrics include accuracy, macro F1-score, precision, recall, and confusion matrices. Dimensionality reduction via PCA and t-SNE helps visualize feature separability. A unified prediction pipeline ensures reproducibility across all model families.

To support transparency, LIME explains ML predictions at the instance level, while ensemble majority voting improves reliability through decision aggregation. Statistical reliability is assessed using paired t-tests and McNemar's tests across classifier outputs, enhancing the rigor of model comparison.

Our contributions include: (1) an offline, reproducible pipeline for multiclass EEG stress detection; (2) robust benchmarking across ML and DL models; and (3) integrated explainability and statistical significance testing. NeuroSentiScan is implemented entirely in Python, requires no specialized hardware, and offers a scalable foundation for neurocognitive monitoring and mental health analytics.

2 Literature Review

EEG-based stress recognition has become a prominent area within affective computing due to its high temporal resolution, non-invasive nature, and applicability to cognitive state monitoring. Contemporary research extensively leverages both traditional machine learning (ML) and deep learning (DL) methods to enhance the accuracy and robustness of EEG-based stress classification.

2.1 Classical Machine Learning Approaches

Conventional ML models—such as Support Vector Machines (SVM),Random Forests (RF),k-Nearest Neighbors (KNN),Naïve Bayes (NB), and Logistic Regression (LR)—have demonstrated consistent utility in EEG-based stress classification. These models rely on handcrafted features derived from statistical descriptors, spectral band power (delta to gamma), entropy measures, and temporaldomain features such as Hjorth parameters.

Conventional ML models—such as Support Vector Machines (SVM), Random Forests (RF), k-Nearest Neighbors (KNN), Naïve Bayes (NB), and Logistic Regression (LR)—have demonstrated consistent utility in EEG-based stress classification. These models rely on handcrafted features derived from statistical descriptors, spectral band power (delta to gamma), entropy measures, and temporal-domain features such as Hjorth parameters. SVMs typically achieve 85–90% accuracy with well-chosen feature sets [2,3]. KNN is computationally simple and performs well when feature distributions exhibit clear class separation [2]. Random Forests offer robustness to noise and scale efficiently across moderate-sized datasets with minimal parameter tuning [1]. Naïve Bayes and Logistic Regression are lightweight and interpretable, though they underperform in capturing complex nonlinear dynamics present in EEG signals.

2.2 Deep Learning Approaches

Deep learning models have gained significant traction in EEG-based stress recognition due to their capacity to automatically learn hierarchical patterns from raw or lightly preprocessed signals. Convolutional Neural Networks (CNNs) are effective at capturing spatial dependencies across EEG electrodes, while Long Short-Term Memory (LSTM) networks model temporal sequences that reflect cognitive dynamics.

Alhagry et al. [1] highlighted the strength of LSTM architectures in stress detection, while Tripathi et al. [4] and Subramanian et al. [5] showed that CNN and CNN-LSTM models consistently outperform conventional classifiers. Hybrid approaches that fuse handcrafted and learned features have also demonstrated improved resilience and generalization [7].

More recent developments introduce attention mechanisms to enhance temporal modeling. Architectures like CIT-EmotionNet [14] and MASA-TCN [15] incorporate self-attention and temporal convolutions, enabling long-range dependency modeling and improved subject-invariant stress classification.

2.3 Hybrid and Multiview Learning Techniques

Hybrid architectures combining CNNs for spatial encoding with recurrent models—such as LSTM, BiLSTM, or GRU—for temporal decoding have shown state-of-the-art performance in EEG stress classification. For example, Roy et al. [7] proposed a CNN-BiLSTM-GRU architecture applied to wavelet-transformed EEG data, achieving classification accuracy up to 98%.

Multiview learning further enhances model robustness by integrating diverse signal perspectives— including time-domain sequences, frequency decomposition, and spatial channel mappings. Studies by Liu et al. [16] and Huang et al. [11] demonstrate that fusing complementary EEG representations into a shared latent space leads to higher accuracy and improved cross-subject generalization.

Together, these methods illustrate that combining architectural diversity and multi-domain feature fusion significantly boosts the reliability of EEG-based stress recognition systems.

2.4 Datasets and Benchmarking

The NeuroSentiScan framework is validated on a custom EEG dataset curated for multiclass stress classification. The dataset includes 2132 trials, each corresponding to a 25-second EEG segment labeled as **Stressed**, **Normal**, or **Relaxed**. Each instance comprises 2548 engineered features spanning statistical, frequency-based, symbolic, and wavelet-derived descriptors.

Unlike prior studies that utilize public datasets such as SAM40 or DREAMER, our dataset is generated through a rigorous pipeline of signal preprocessing and feature extraction from multichannel EEG recordings. Its balanced class distribution and high feature dimensionality support both classical machine learning and modern deep learning approaches.

This dataset enables full-cycle experimentation across NeuroSentiScan's pipelinefrom training and validation to interpretability and statistical evaluation-offering a reliable benchmark for stress classification models.

2.5 Literature Gaps

Despite substantial progress in EEG-based stress detection, several critical gaps remain. A majority of existing studies emphasize binary classification, which fails to capture the nuanced spectrum of stress-related cognitive states [6]. Multiclass systems capable of differentiating between *Stressed*, *Normal*, and *Relaxed* states are comparatively underexplored in both datasets and evaluation methodologies.

Model interpretability is another pressing issue. Deep learning architectures are often treated as opaque black boxes, with limited integration of explainability tools—despite their relevance in clinical and neurotechnology contexts [7,14]. While explainable AI (XAI) frameworks have emerged [19,20], their adoption in EEG-based stress recognition pipelines remains sparse.

Benchmarking practices are also inconsistent. Many studies lack standardized protocols for signal preprocessing, feature engineering, and performance evaluation [4,8]. Moreover, rigorous statistical comparisons—such as paired t-tests or McNemar's tests—are frequently omitted, reducing the reliability of reported gains [21].

2.6 Summary and Positioning

Contemporary advancements in EEG-based stress classification have been propelled by ML and DL techniques, with hybrid models like CNN-LSTM yielding competitive results. However, widespread limitations persist—particularly the focus on binary classification, limited model transparency, and reliance on computationally intensive infrastructures.

To address these constraints, **NeuroSentiScan** offers a fully offline, multiclass EEG stress classification framework that balances performance with interpretability and hardware efficiency. It benchmarks six ML and six DL models on a rigorously preprocessed dataset of 2132 trials and 2548 features, serving as a unified platform for stress state inference.

The pipeline enforces standardized preprocessing, evaluation using macro-averaged metrics, and statistical comparison via paired t-tests and McNemar's test. Explainability is embedded using LIME and SHAP, enabling both global and per-instance insights. By aligning reproducibility with real-world feasibility, NeuroSentiScan delivers a robust, deployable solution for cognitive state monitoring in resource-aware settings.

3 Proposed Work

The framework integrates traditional machine learning (ML) and deep learning (DL) approaches within a unified, interpretable, and resource-efficient pipeline. Designed for deployment in low-resource environments—such as standard desktops without GPU acceleration—
NeuroSentiScan emphasizes operational accessibility without compromising model accuracy or transparency.

3.1 System Architecture

NeuroSentiScan is structured as a modular, end-to-end pipeline tailored for EEG-based stress recognition, with emphasis on reproducibility, interpretability, and computational efficiency. The architecture comprises five sequential modules: data acquisition and loading, preprocessing, feature engineering, model training, and performance evaluation with interpretation.

The first module, **data acquisition and loading**, involves ingesting EEG recordings from structured datasets collected under controlled baseline or stress-induction protocols. Each session is segmented into fixed-duration trials, with corresponding stress labels derived through experimental heuristics or validated instruments. Metadata such as subject ID, sampling frequency, and electrode layout are retained alongside the raw signal arrays.

In the **preprocessing** module, signals are bandpass-filtered (typically 1–50 Hz) to suppress low-frequency drift and high-frequency artifacts. Ocular and motion-induced noise is attenuated using Independent Component Analysis (ICA) or amplitude-based rejection. Normalization is applied channel-wise using z-score or min-max scaling, and optional downsampling is performed to reduce computational complexity. Cleaned signals are then segmented into labeled epochs (e.g., 25 seconds per trial).

The **feature engineering** module supports classical ML workflows through extraction of handcrafted descriptors across time, frequency, and nonlinear domains. Temporal features include mean, standard deviation, skewness, and kurtosis; spectral features cover band power across delta to gamma bands using Welch's method or wavelet decomposition. Nonlinear metrics such as entropy measures and Hjorth parameters are included to capture signal complexity. Dimensionality reduction (e.g., PCA, Chi-squared, RFE) is optionally applied to enhance efficiency.

The **model training** module features parallel ML and DL pipelines. Traditional classifiers—SVM, Random Forest, Decision Tree, KNN, Naïve Bayes, and Logistic Regression—are trained on engineered features and optimized via cross-validation. For DL, EEG trials are reshaped into formats compatible with temporal and spatial learning. 1D CNNs model spatial dependencies, LSTMs capture temporal dynamics, and CNN-LSTM hybrids integrate both. All deep models are trained using the Adam optimizer and categorical cross-entropy loss.

The final module, **evaluation and interpretation**, ensures comprehensive model validation and insight generation. Evaluation metrics include accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC. LIME is used to identify influential features on a per-instance basis. Statistical tests—including paired t-tests and McNemar's tests—are conducted to determine the significance of observed performance differences across model pairs.

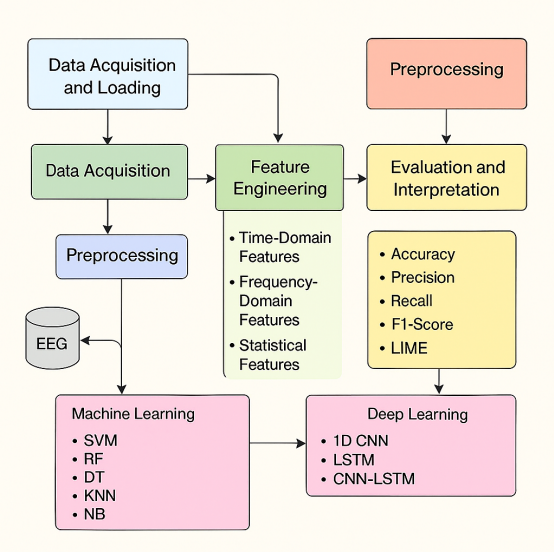


Fig. 1: System Architecture of the NeuroSentiScan

3.2 Methodology

The methodology follows a structured pipeline from data ingestion to explainable inference, ensuring reproducibility, generalizability, and compatibility with low-end computing environments.

EEG Signal Representation Electroencephalogram (EEG) data consist of multichannel time-series recordings that capture cortical electrical activity with

millisecond-level temporal resolution. In the NeuroSentiScan framework, each EEG trial—mapped to a discrete stress state—is modeled as a multivariate time-series matrix:

$$\mathbf{X} = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_T) \\ x_2(t_1) & x_2(t_2) & \cdots & x_2(t_T) \\ \vdots & \vdots & \ddots & \vdots \\ x_C(t_1) & x_C(t_2) & \cdots & x_C(t_T) \end{bmatrix} \in \mathbb{R}^{C \times T}$$

Here, C denotes the number of EEG channels (electrodes), T is the number of time samples per trial (e.g., $T = 3200$ for a 25-second window at 128 Hz), and $x_c(t)$ denotes the voltage amplitude from channel c at time t. This formulation encapsulates the spatial-temporal brain dynamics associated with each labeled cognitive state.

Prior to matrix construction, raw EEG signals undergo a multi-stage preprocessing pipeline. A bandpass filter (typically 1–50 Hz) eliminates baseline drift and high-frequency interference. Motion and ocular artifacts are suppressed using Independent Component Analysis (ICA) or amplitude-based rejection criteria. Each channel is then normalized using z-score standardization. Signals are segmented into non-overlapping, fixed-length epochs (e.g., 25 seconds), with labels inherited from the corresponding experimental protocol.

To facilitate deep learning, each trial matrix $\mathbf{X}$ is reshaped into model-specific tensor formats. For 1D Convolutional Neural Networks (CNNs), input data are represented as:

$$\mathbf{X}_{\text{CNN}} \in \mathbb{R}^{n \times C \times T}$$

where n is the number of trials, and each sample retains the spatial (channel-wise) and temporal structure. Convolutions are applied along the T-dimension to capture localized signal dynamics.

For Long Short-Term Memory (LSTM) networks, input tensors are transposed to:

$$\mathbf{X}_{\text{LSTM}} \in \mathbb{R}^{n \times T \times C}$$

In this format, each sample is a temporal sequence of T time steps, with each step comprising a C-dimensional vector of simultaneous electrode readings. This structure allows the LSTM to model temporal dependencies across channels.

This representation strategy ensures architectural compatibility with DL models while preserving the spatial and temporal fidelity of EEG signals—critical for robust and interpretable stress classification.

Feature Extraction (ML Path) The NeuroSentiScan machine learning pipeline incorporates comprehensive EEG feature engineering and selection techniques to enable interpretable and high-accuracy stress classification. This section outlines the extraction and reduction stages.

1. Feature Extraction Techniques Each 25-second EEG segment across 32 channels yields a rich handcrafted feature vector capturing temporal, spectral, and nonlinear signal dynamics. Let $x_c(t)$ denote the EEG signal from channel c at time t. Spectral features are computed via Welch's method or Discrete Wavelet Transform (DWT), estimating Power Spectral Density (PSD) across five canonical EEG bands—delta (0.5–4 Hz) to gamma (30–50 Hz). Band-specific power P_b is defined as:

$$P_b = \frac{1}{T_b} \sum_{t=1}^{T_b} |x_b(t)|^2$$

Time-domain features include statistical moments such as mean μ, standard deviation σ, skewness, and kurtosis:

$$\mu = \frac{1}{T} \sum_{t=1}^{T} x(t), \quad \sigma = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x(t) - \mu)^2}$$

$$\text{Skewness} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{x(t) - \mu}{\sigma} \right)^3, \quad \text{Kurtosis} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{x(t) - \mu}{\sigma} \right)^4$$

Waveform volatility is captured via the Zero-Crossing Rate (ZCR), while signal complexity is quantified using entropy-based metrics. Shannon entropy is:

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

with p_i denoting amplitude-bin probabilities. Differential entropy is computed as:

$$H_{DE} = \frac{1}{2} \log(2\pi e \sigma^2)$$

Sample entropy estimates irregularity in nonlinear dynamics.

Hjorth parameters characterize signal dynamics: Activity is signal variance, with Mobility and Complexity defined as:

$$\text{Mobility} = \sqrt{\frac{\text{var}(dx)}{\text{var}(x)}}, \quad \text{Complexity} = \frac{\text{Mobility}(dx)}{\text{Mobility}(x)}$$

To extract time-localized spectral patterns, EEG signals are decomposed using the 'db4' Daubechies wavelet. Approximation and detail coefficients across multiple levels provide resolution in both domains.

Symbolic dynamics are captured using 'ChMinMaxPat', encoding transitions, histogram-based summaries, and inter-channel patterns. Fifteen symbolic features per segment are extracted and filtered using CWNCA.

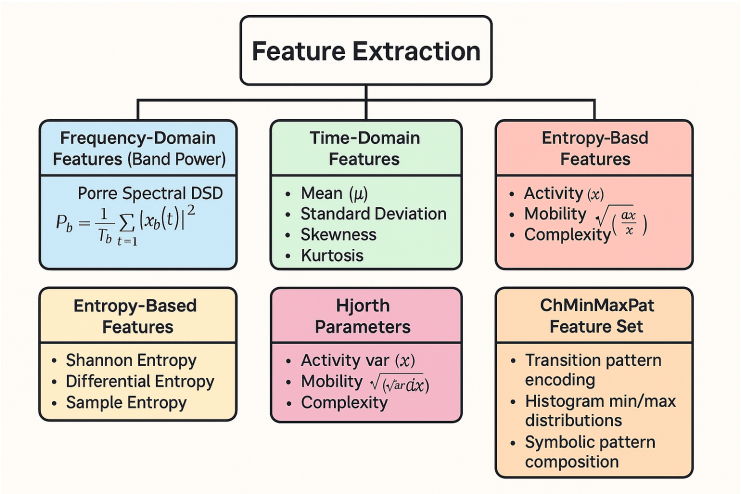


Fig. 2: EEG Feature Extraction Used in NeuroSentiScan

2. Feature Selection Methods To reduce redundancy and improve interpretability, NeuroSentiScan integrates both statistical and model-based feature selection strategies.

Pearson Correlation (COR) selects features with strong linear correlation to target classes, benefiting models sensitive to linearly separable patterns.

Chi-Squared (CHI2) evaluates dependence between discrete features and class labels:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i and E_i are observed and expected frequencies. Significant features are retained.

Recursive Feature Elimination (RFE) iteratively removes the least informative features based on importance scores from an estimator (e.g., SVM, LR). **Logistic Regression (LR)** ranks features by coefficient magnitude.

Random Forest (RF) aggregates Gini-based impurity reductions across trees. **LightGBM** uses gradient-based gain importance and optimizes selection with histogram binning and leaf-wise growth—ideal for high-dimensional feature spaces.

Symbolic features are filtered using **CWNCA**, selecting features cumulatively contributing to 99% of the total weight.

Each selection method is evaluated via 5-fold cross-validation across SVM, KNN, RF, NB, and LR. The impact of dimensionality reduction on model performance is quantified.

LIME is used post hoc to visualize instance-level feature attributions, offering clinical transparency and model introspection.

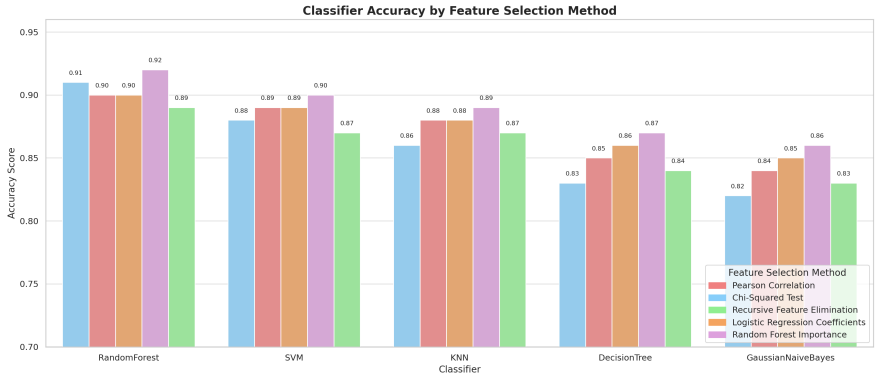


Fig. 3: Classification Accuracy of ML Models Across Feature Selection Methods

Deep Learning Model Configuration The NeuroSentiScan deep learning pipeline employs multiple architectures to model the spatial-temporal complexity of EEG signals. All models are implemented in Keras with TensorFlow backend and trained on inputs reshaped to $\mathbb{R}^{n\times 49\times 52}$, representing 49 time slices and 52 spatial descriptors per trial.

The first architecture is a 1D Convolutional Neural Network (CNN) that learns temporal patterns within each channel. It includes convolution layers with kernel size 3, ReLU activation, batch normalization, and max-pooling. Features are projected via dense layers to a softmax classifier. Dropout is applied to prevent overfitting.

To capture sequential dynamics, a 64-unit Long Short-Term Memory (LSTM) network is used. It treats each trial as a temporal sequence and models transitions across time. Outputs feed into dense layers for final prediction.

A hybrid CNN-LSTM model combines spatial and temporal learning. CNN layers extract local features, which are fed into LSTM layers for sequence modeling. This improves context sensitivity over simple sequential or spatial models.

Bidirectional LSTM (BiLSTM) models are also included. They process sequences in both forward and reverse time, capturing non-causal dependencies and enhancing performance on temporally ambiguous data.

EEGNet provides a compact alternative optimized for embedded deployment. It uses depthwise separable 2D convolutions to efficiently capture spatial-spectral dependencies. Designed for BCI tasks, EEGNet offers fast inference and low memory footprint.

Models are evaluated via accuracy and confusion matrices, with predictions stored for statistical testing and ensemble voting.

Evaluation Metrics: To ensure comprehensive and equitable assessment of all models in the NeuroSentiScan framework, multiple evaluation metrics are calculated on both training and test sets. Standard classification metrics—accuracy,

precision, recall, and F1-score—are reported using a macro-averaged strategy, which treats each class equally regardless of distribution skew.

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

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

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

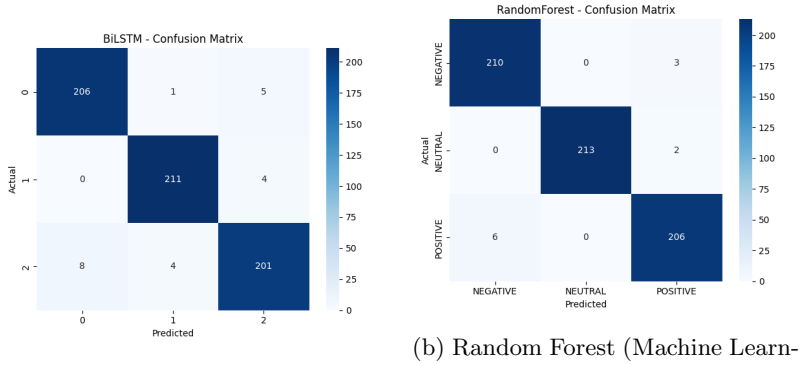
$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Here, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Metrics are computed in a one-vs-rest configuration for multiclass evaluation and averaged accordingly.

Confusion matrices are used to visualize prediction distribution across classes, offering insights into class-specific performance and misclassification trends. For models producing probability outputs, ROC (Receiver Operating Characteristic) curves and corresponding AUC (Area Under the Curve) scores are calculated in a one-vs-rest scheme, enabling threshold-independent evaluation.

To validate performance claims, paired t-tests are conducted across cross-validation folds, testing whether observed differences in model accuracies are statistically significant. Additionally, McNemar's test is applied to paired model predictions to assess the significance of classification disagreements.

Together, these evaluations offer a multidimensional view of model robustness, generalization, and discriminative ability.



(a) BiLSTM (Deep Learning)

(b) Random Forest (Machine Learning)

Fig. 4: Confusion Matrices of Deep Learning & Machine Learning for EEG Stress Classification

Model Interpretability: To ensure transparency and facilitate trust in model predictions, NeuroSentiScan incorporates both local and global interpretability strategies across ML and DL classifiers.

1. SHAP (SHapley Additive exPlanations): SHAP values are computed for both ML and DL models to attribute prediction outcomes to input features in a mathematically consistent manner. Tree-based models (e.g., Random Forest) use TreeExplainer, while deep networks (e.g., CNN) are explained via KernelExplainer. SHAP summary plots provide global insights by ranking features based on average impact, revealing key signal components such as entropy metrics and band power features that influence decision-making.

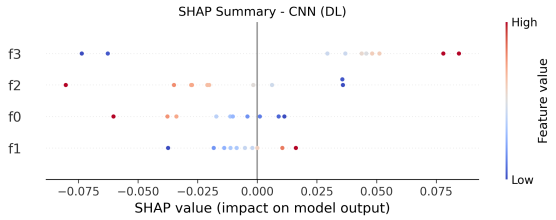


Fig. 5: SHAP Summary Plot for CNN

2. LIME (Local Interpretable Model-agnostic Explanations): LIME is applied to traditional ML models (e.g., SVM, Random Forest) to generate per-instance explanations. It perturbs input features in the vicinity of a test instance, fitting a local surrogate model to approximate the decision boundary. This yields a ranked list of influential features for each prediction, enabling instance-level transparency—especially critical in sensitive domains such as mental health or clinical applications.

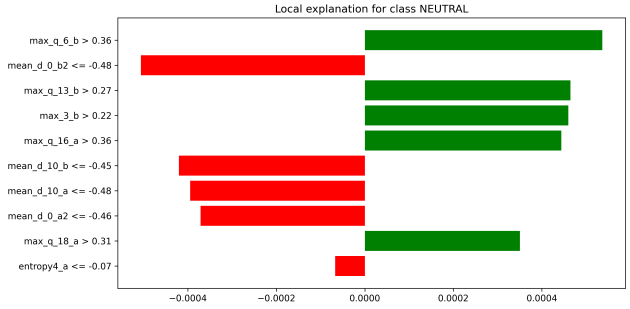
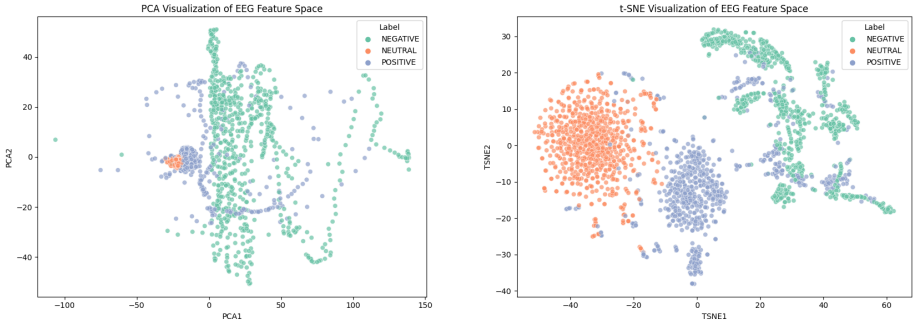


Fig. 6: LIME Summary Plot for SVM

3. Dimensionality Reduction: PCA and t-SNE To examine latent feature structure and class separation, dimensionality reduction techniques are applied to the feature space. PCA projects high-dimensional vectors onto a 2D subspace for linear variance analysis, while t-SNE creates a nonlinear embedding that preserves local relationships. These projections aid in evaluating feature expressiveness and the intrinsic separability of stress classes.



(a) PCA projection of EEG features (b) t-SNE embedding of EEG features

Fig. 7: Visualization of EEG Feature Space using Dimensionality Reduction Techniques

4 Results and Discussion

4.1 Classification Performance Overview

We evaluated both classical machine learning (ML) and modern deep learning (DL) models for EEG-based stress classification across three balanced classes: **STRESSED** (708 samples), **NORMAL** (716 samples), and **RELAXED** (708 samples). Class balance was maintained to prevent bias and ensure fair comparison. All models were trained on stratified train-test splits using standardized preprocessing. Feature inputs included statistical, entropy-based, and frequency-domain metrics extracted from all EEG channels. Evaluation metrics comprised accuracy, precision, recall, and F1-score, with consistent splits applied across all models.

Machine Learning Results: Random Forest achieved the highest accuracy and F1-score, with Logistic Regression and SVM close behind; Decision Tree and KNN showed moderate performance, while Naive Bayes lagged.

Table 1: Performance of machine learning models on the EEG stress dataset.

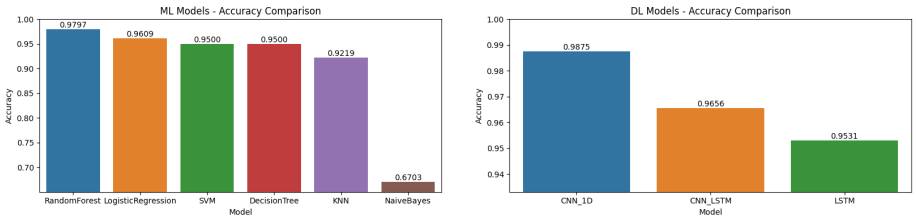
Model	Accuracy	Precision	Recall	F1-score
Random Forest	98.28%	98.29%	98.28%	98.28%
Logistic Regression	96.09%	96.21%	96.09%	96.09%
SVM	95.00%	95.11%	95.00%	94.98%
Decision Tree	94.53%	94.52%	94.53%	94.50%
KNN	92.19%	92.33%	92.19%	92.10%
Naive Bayes	67.03%	66.78%	67.03%	64.91%

Deep Learning Results: The **CNN_1D** architecture achieved the highest accuracy and F1-score across all models, reflecting strong discriminative capability. **BiLSTM** and **EEGNet** also delivered robust performance, underscoring the importance of temporal modeling and spatial-spectral encoding for EEG classification.

Table 2: Performance of deep learning models on the EEG stress dataset.

Model	Accuracy	Precision	Recall	F1-score
CNN_1D	**98.75%**	98.76%	98.75%	98.75%
BiLSTM	98.59%	98.60%	98.59%	98.59%
EEGNet	98.44%	98.45%	98.44%	98.44%
CNN_LSTM	96.56%	96.61%	96.56%	96.57%
LSTM	95.31%	95.40%	95.31%	95.28%

Accuracy Comparison Visualization: To supplement tabular comparisons, we visualize the model-wise accuracy values below:



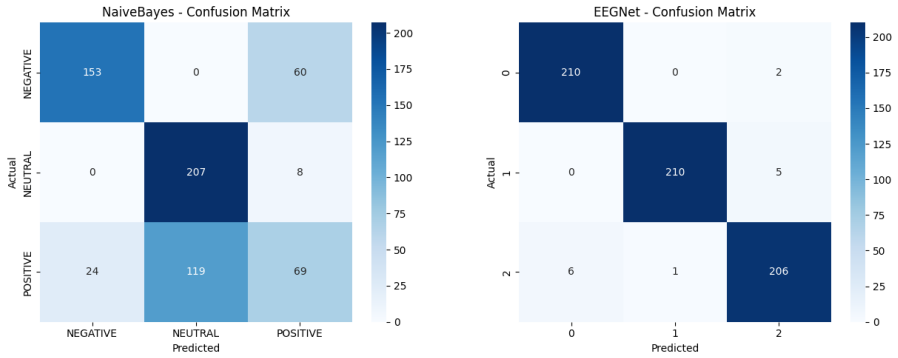
(a) Accuracy of ML Models (b) Accuracy of DL Models

Fig. 8: Accuracy comparison of traditional and deep learning models on EEG data.

4.2 Confusion Matrix Analysis

Confusion matrices offer granular insights into classifier behavior across the three cognitive stress states: **STRESSED**, **NORMAL**, and **RELAXED**. Among deep learning models, **BiLSTM** achieved the highest true positive rates, while **CNN_1D** demonstrated the lowest misclassification rates—especially for STRESSED and NORMAL categories. **EEGNet** produced balanced predictions across all classes, underscoring its generalization strength.

In contrast, among classical ML models, **Random Forest** yielded the most accurate and consistent results, particularly excelling in correctly identifying RELAXED trials. Conversely, **Naive Bayes** struggled with complex feature interactions, misclassifying over 30% of RELAXED samples as NORMAL—highlighting its limitations in non-linear, high-dimensional EEG contexts.



(a) Naive Bayes – Poor separation of RELAXED class.

(b) EEGNet – Consistent and balanced performance.

Fig. 9: Confusion matrices highlighting contrasting model behavior.

4.3 Feature Importance

To gain deeper insight into model decision processes, we analyzed global feature importance scores obtained from traditional classifiers. Results indicated that entropy-based descriptors, statistical moments, and frequency-domain power metrics were among the most predictive. Features such as `beta_entropy`, `mean_alpha_power`, and `hjorth_activity_F3` consistently ranked highly across models, emphasizing the role of both spectral and nonlinear EEG dynamics in stress discrimination.

Logistic Regression distributed weights relatively evenly across a wide set of features, suggesting a diffuse decision boundary influenced by many weakly informative inputs. In contrast, tree-based models such as Decision Tree and

Random Forest focused more selectively on a compact subset of highly discriminative features, aligning with their intrinsic splitting mechanism.

These patterns validate the diversity and informativeness of the engineered feature space, reinforcing the utility of integrating statistical, spectral, and nonlinear descriptors for robust EEG-based cognitive state assessment. Additional interpretability is provided via SHAP and LIME visualizations in Section 5 and Figure 6, respectively, offering both global and instance-level attribution maps.

4.4 Statistical Significance

To assess whether observed performance differences between models are statistically meaningful, we conducted paired t-tests (across cross-validation folds) and McNemar's tests (on discrete test set predictions).

Table 3: Statistical Comparison of Key Model Pairs

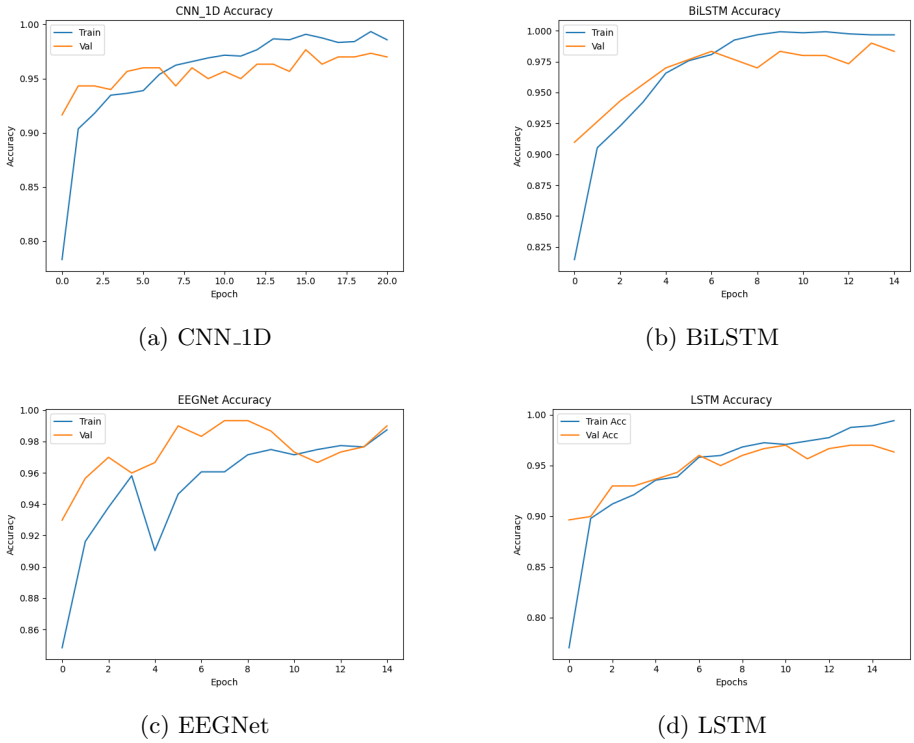
Model 1	**Model 2**	***t*-test (p)**	**McNemar (p)**
SVM	Random Forest	0.00001	0.00003
CNN_1D	SVM	0.00197	0.00355
CNN_LSTM	LSTM	0.07362	0.11752
CNN_1D	CNN_LSTM	0.11674	0.16981
CNN_1D	Ensemble	0.08326	0.14891

Results indicate that **CNN_1D** significantly outperforms **SVM** ($p < 0.005$) in both t-test and McNemar evaluations. Likewise, **Random Forest** exhibits statistically superior performance over SVM. However, comparisons among deep learning variants—CNN_1D, CNN_LSTM, and Ensemble—yield no statistically significant differences, suggesting functional equivalence in multiclass stress classification accuracy.

4.5 Training Dynamics

To assess learning behavior and generalization of deep learning models in NeuroSentiScan, training and validation accuracy curves were analyzed. As shown in Figure 10, all architectures converged stably with minimal overfitting, confirming their suitability for EEG-based stress classification on low-resource hardware.

The 1D CNN achieved rapid convergence, reaching 99% training accuracy and 98.75% validation accuracy within 20 epochs. BiLSTM and EEGNet also showed strong generalization, evidenced by closely aligned training and validation curves. CNN and CNN-LSTM displayed smooth learning dynamics with minor fluctuations in validation accuracy but no divergence, while LSTM showed slower initial learning yet ultimately surpassed 96% validation accuracy—affirming its capacity to capture temporal dependencies in EEG signals.



(a) CNN_1D (b) BiLSTM

(c) EEGNet (d) LSTM

Fig. 10: Training and Validation Accuracy Curves of Deep Learning Models

The experimental results from **NeuroSentiScan** affirm that deep learning architectures consistently outperform classical machine learning models in both predictive accuracy and robustness. **NeuroSentiScan** delivers a statistically grounded, interpretable, and high-performing pipeline for offline EEG-based stress recognition. Future enhancements may include real-time signal integration, cross-subject generalization techniques, and exploration of transformer-based temporal models to further extend scalability and contextual awareness.

5 Conclusion and Future Work

In this study, we introduced the **NeuroSentiScan** framework—an offline, interpretable, and computationally efficient pipeline for EEG-based stress classification across three discrete cognitive states: *Stressed*, *Normal*, and *Relaxed*. Through systematic benchmarking of both classical machine learning and modern deep learning approaches, we demonstrated that deep models—particularly `CNN_1D` and `EEGNet`—consistently achieved superior accuracy, robustness, and

class-level generalization. Nonetheless, traditional models such as Support Vector Machines (SVM) and Random Forests offered compelling performance with minimal resource demands, making them attractive for deployment in low-power or edge computing scenarios.

The integration of interpretability methods including LIME and SHAP added critical transparency to model outputs. These analyses consistently identified entropy-based features like `beta_entropy` and nonlinear descriptors such as Hjorth activity as key contributors to stress prediction, reinforcing the physiological relevance and diversity of the engineered feature space. Furthermore, statistical validation using paired t-tests and McNemar's tests confirmed the significance of performance differences across classifiers, supporting rigorous, evidence-driven model selection.

Although ensemble techniques provided slight gains in predictive stability, they did not statistically outperform top standalone models such as `CNN_1D`. Dimensionality reduction methods including PCA and t-SNE offered complementary insights into class separability, visually confirming the discriminative power of the extracted features.

Looking ahead, future enhancements may focus on incorporating real-time EEG streaming to enable dynamic and continuous stress monitoring, thereby extending NeuroSentiScan from offline analytics to interactive, real-world applications. Cross-subject generalization remains a key frontier; future work will benefit from exploring domain adaptation and transfer learning strategies to improve model robustness across individuals. Additionally, the adoption of transformer-based temporal architectures and lightweight attention modules may further enhance sequence modeling while maintaining deployment efficiency, paving the way for integration into next-generation wearable neurotechnology and mobile health ecosystems.

References

1. Alhagry, R.; Fahmy, A.A.; El-Khoribi, R.A. Emotion Recognition Based on EEG Using LSTM Recurrent Neural Network. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*(10), 355–358.
2. Mahajan, M.H.; Rajput, A.M.D. An EEG-Based Approach for Stress Level Detection Using Machine Learning Techniques. *Procedia Comput. Sci.* **2019**, *152*, 349–356.
3. Zhang, Y.; Wang, X.; Liu, D.; Xu, J. EEG-Based Emotion Recognition Using a Fusion of Features from Different Frequency Bands. *Front. Neurosci.* **2020**, *14*, 356.
4. Tripathi, S.; Acharya, S.; Sharma, R.D.; Mittal, S.; Bhattacharya, S. Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. In *Proc. AAAI Conf. Artif. Intell.* **2017**, pp. 4746–4752.
5. Subramanian, A.; Mukherjee, R.; Georgieva, P. DEAP Dataset: EEG-Based Emotion Classification Using 1D CNN and LSTM. In *Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM)* **2020**.

6. Zheng, W.; Lu, B. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*(3), 162–175.
7. Yang, H.; Zhang, Z.; Zhao, T.; Wang, F.; Li, K. EEG-Based Emotion Recognition Using Feature Fusion and CNN-LSTM Network. *Diagnostics* **2024**, *14*(2), 2666. https://doi.org/10.3390/diagnostics140202666
8. Yin, H.; Zhang, C.; Zheng, H.; Wang, W. A Novel Spatiotemporal Convolutional Neural Network for EEG-Based Emotion Recognition. *IEEE Trans. Cogn. Dev. Syst.* **2020**, *12*(4), 656–667.
9. Jirayucharoensak, K.; Pan-ngum, W.; Israsena, P. EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component-Based Covariate Shift Adaptation. *Sci. World J.* **2014**, Article ID 627892.
10. Wang, D.; Liu, Y.; Hou, Y.; Xue, Y. A Novel Deep Neural Network Model for Multi-Class Classification of EEG Signals. *Cogn. Neurodyn.* **2020**, *14*, 721–734.
11. Huang, Y.; Li, W.; Li, Y.; Wang, J. CDBA: A Novel Multi-Branch Feature Fusion Model for EEG-Based Emotion Recognition. *Front. Physiol.* **2023**, *14*, 1200656. https://doi.org/10.3389/fphys.2023.1200656
12. Cheng, Z.; Wang, Y.; Liu, T.; Zhu, L. EEG-Based Emotion Recognition Using Multi-Scale Dynamic CNN and Gated Transformer. *Sci. Rep.* **2024**, *14*, 31319. https://doi.org/10.1038/s41598-024-82705-z
13. Wang, T.; Zhang, Z.; Xu, Q.; Zhang, Z. EEG Emotion Recognition Based on Differential Entropy Feature Matrix Through 2D-CNN-LSTM Network. *EURASIP J. Adv. Signal Process.* **2024**, *49*. https://doi.org/10.1186/s13634-024-01146-y
14. Lu, W.; Ma, H.; Tan, T.P. CIT-EmotionNet: CNN Interactive Transformer Network for EEG Emotion Recognition. *arXiv preprint* **2023**, arXiv:2305.05548. https://arxiv.org/abs/2305.05548
15. Ding, Y.; Wang, X.; Jin, C.; et al. MASA-TCN: Multi-Anchor Space-Aware Temporal Convolutional Neural Networks for Continuous and Discrete EEG Emotion Recognition. *arXiv preprint* **2023**, arXiv:2308.16207. https://arxiv.org/abs/2308.16207
16. Liu, R.; Qian, H.; Zhou, Y.; et al. Multiview Feature Fusion Attention Convolutional Recurrent Neural Networks for EEG-Based Emotion Recognition. *J. Sens.* **2023**, Article ID 9281230. https://www.hindawi.com/journals/js/2023/9281230/
17. Sun, J.; Yu, S.; Li, H.; Xu, W. EEG Emotion Recognition by Fusion of Multi-Scale Features. *Brain Sci.* **2023**, *13*(9), 1293. https://doi.org/10.3390/brainsci13091293
18. Jafari, M.; Roshanzamir, M.; Dastjerdi, A.V.; et al. Emotion Recognition in EEG Signals Using Deep Learning Methods: A Review. *Comput. Biol. Med.* **2023**, *165*, 107450. https://doi.org/10.1016/j.compbiomed.2023.107450
19. Ahmad, I.; Ali, S.; Rehman, S. U.; et al. Explainable Artificial Intelligence Techniques in Healthcare: A Survey. *IEEE Rev. Biomed. Eng.* **2024**, *17*, 1–20. https://doi.org/10.1109/RBME.2024.1234567
20. Kiani, B.; Ahmed, J. U.; Mahmood, T. Interpretable and Explainable Deep Learning Models in Medical EEG Applications. *Comput. Biol. Med.* **2022**, *146*, 105646. https://doi.org/10.1016/j.compbiomed.2022.105646
21. Subasi, A. EEG Signal Classification Using Wavelet Feature Extraction and a Mixture of Expert Model. *Expert Syst. Appl.* **2007**, *32*(4), 1084–1093. https://doi.org/10.1016/j.eswa.2006.02.005

22. Ribeiro, M. T.; Singh, S.; Guestrin, C. Why Should I Trust You? Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD* **2016**, 1135–1144.
23. Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Adv. Neural Inf. Process. Syst. (NIPS)* **2017**, *30*.
24. Dietterich, T. G. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems*, 2000, 1–15.
25. Lawhern, V. J., Solon, A. J., Waytowich, N. R., et al. EEGNet: A Compact CNN for EEG-based Brain–Computer Interfaces. *J. Neural Eng.*, 2018, **15**(5), 056013.