SR (EEG + Speech)

Step 1: Data Preprocessing (The "Cleaning" Engine)

- **EEG:** Apply a **0.5–50 Hz Bandpass Filter** to remove slow drifts and power-line hum. Perform **ICA (Independent Component Analysis)** to isolate and remove EOG (eye-blink) and EMG (muscle) artifacts.
- **Audio:** Remove silence segments and normalize amplitude. Segment audio into **5-sentence chunks** (from Paper 5) to capture temporal prosody.

Step 2: Feature Extraction (Domain Transformation)

- **Linear Features:** Extract **Power Spectral Density (PSD)** for Alpha (8-13Hz), Beta (13-30Hz), and Theta (4-8Hz) bands across 128 channels.
- **Nonlinear Features (from Paper 2):** Calculate **Sample Entropy** and **Hurst Exponents** to measure brain signal complexity.
- **Audio Features:** Extract **MFCCs (Mel-frequency cepstral coefficients)** and **Prosodic features** (pitch variance, jitter, shimmer).

Step 3: Feature-Level Fusion & Dimensionality Reduction

- **Concatenation:** Join the EEG and Audio feature vectors into a single bimodal vector $\mathbf{X}$.
- **Recursive Feature Elimination (RFE):** Since 128 channels provide too many variables for a standard logistic model, use RFE to select the **Top-k** most "depression-relevant" brain regions (Frontal/Temporal).

Step 4: Model Training & Evaluation

- **Model:** Logistic Regression with **L2 Regularization (Ridge)** to prevent overfitting on the small MODMA dataset.
- **Cross-Validation:** Use **Stratified 10-Fold Cross-Validation** to ensure the model generalizes across different ages and genders (as identified in Paper 4).

Part 2: Mathematical Formulation

1. The Logistic Function (Sigmoid)

The model predicts the probability P that a subject has Major Depressive Disorder (MDD) given the bimodal feature vector $\mathbf{X}$:

$$P(y = 1|\mathbf{X}) = \sigma(\mathbf{w}^T\mathbf{X} + b) = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{X}+b)}}$$

Where:

- $\mathbf{w}$: Weights assigned to EEG and Audio features (representing feature importance).
- b: The bias term.

2. Multi-Paradigm Feature Fusion

The feature vector $\mathbf{X}$ is a concatenated representation:

$$\mathbf{X} = [\mathbf{F}_{EEG(EO)}, \mathbf{F}_{EEG(EC)}, \mathbf{F}_{Audio}]$$

Where $\mathbf{F}_{EEG(EO)}$ is the feature set from the **Eyes-Open** paradigm and $\mathbf{EC}$ is from **Eyes-Closed**.

3. Cost Function with L2 Regularization

To avoid "memorizing" the small MODMA dataset, we minimize the Cross-Entropy Loss with a penalty:

$$J(\mathbf{w}) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(P^{(i)}) + (1 - y^{(i)})\log(1 - P^{(i)})] + \frac{\lambda}{2m}\|\mathbf{w}\|^2$$

- The term $\frac{\lambda}{2m}\|\mathbf{w}\|^2$ (L2 Penalty) shrinks the weights of "noisy" EEG channels, focusing the model on the most robust biomarkers.

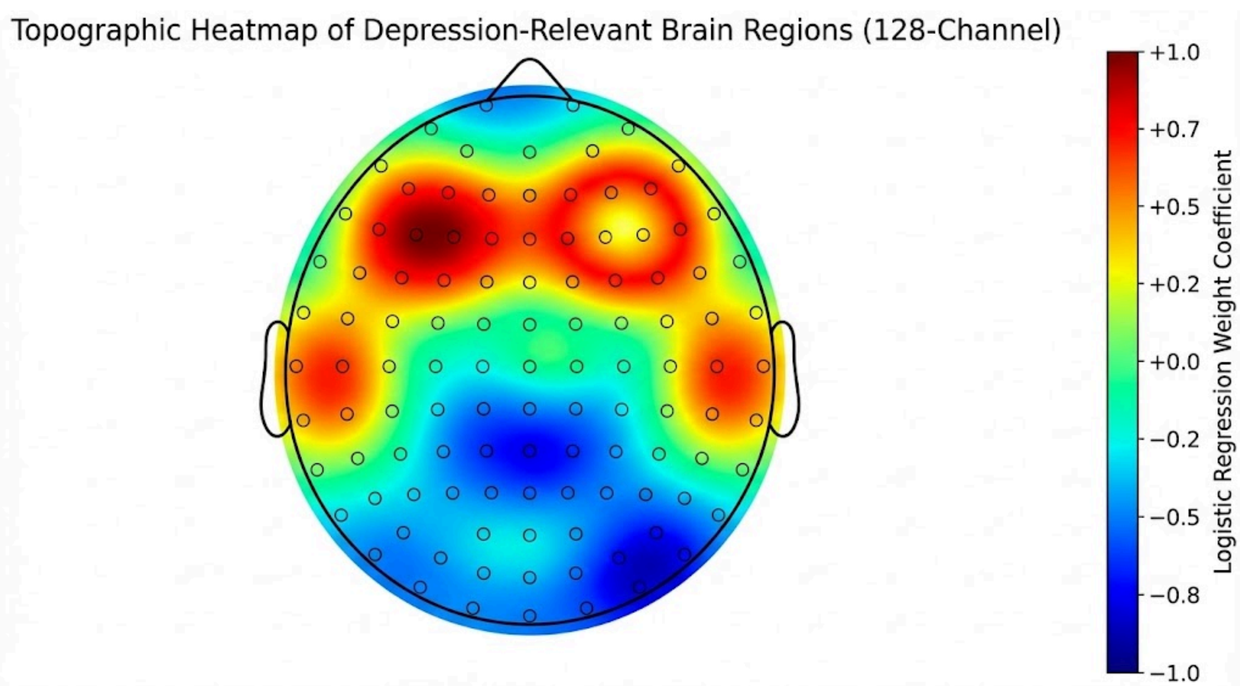
Part 3: Implementation Code (Python - Conceptual)

```python
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import confusion_matrix, classification_report

# 1. LOAD & STANDARDIZE (MODMA Dataset Features)
# Features include Alpha Power, Beta Power, and Mel-frequency coefficients
X_train, X_test, y_train, y_test = load_modma_split()
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# 2. IMPLEMENT ADAPTIVE LOGISTIC REGRESSION (with L2)
# C=0.1 provides strong regularization to handle high-density EEG noise
model = LogisticRegression(penalty='l2', C=0.1, solver='lbfgs',
max_iter=1000)
model.fit(X_train_scaled, y_train)

# 3. EVALUATE
y_pred = model.predict(X_test_scaled)
print(classification_report(y_test, y_pred))

# 4. FEATURE IMPORTANCE ANALYSIS (Viva Insight)
# Higher absolute weights point to the most predictive brain regions
importance = model.coef_[0]
for i, v in enumerate(importance):
    print(f'Feature: {i}, Score: {v:.5f}')
```

Part 4: Real-World Clinical Deployment Model

To bring this to the real world (as suggested by the "villager" consultation idea), the system would operate as a **Cloud-Edge Hybrid**:

1. **Edge (Village Clinic):** A portable 128-channel net collects raw data. A mobile app records the patient's voice during a standard 5-minute interview.
2. **Processing (Cloud):** The raw signal is uploaded. The **ICA Artifact Removal** and **STFT/Mel-transformation** happen on a high-performance server.
3. **Diagnosis:** The **Logistic Regression model** generates a probability score and a **Severity Level (PHQ-8 mapping)**.
4. **Reporting:** The doctor receives a "Topographic Attention Map" (from Paper 4 concepts) showing exactly which brain regions (e.g., Left Frontal) were inhibited, allowing for a personalized treatment plan.



1. Paper Implementations (Skeletons)

Paper 1: Modified DenseNet121 (Bimodal CNN)

Focus: Transfer learning with concatenated feature vectors.

```
from tensorflow.keras.applications import DenseNet121
from tensorflow.keras.layers import Dense, Concatenate, Input,
GlobalAveragePooling2D
from tensorflow.keras.models import Model

# EEG and Audio branches
input_eeg = Input(shape=(224, 224, 3))
input_audio = Input(shape=(224, 224, 3))

base_model = DenseNet121(weights='imagenet', include_top=False)

# Shared or separate backbones
x1 = base_model(input_eeg)
x1 = GlobalAveragePooling2D()(x1)

x2 = base_model(input_audio)
x2 = GlobalAveragePooling2D()(x2)

# Feature-level Concatenation
merged = Concatenate()([x1, x2])
output = Dense(1, activation='sigmoid')(merged)

model_p1 = Model(inputs=[input_eeg, input_audio], outputs=output)
```

Paper 3: Vision Transformer (ViT)

Focus: Global attention across multi-frequency EEG bands.

```
import torch
from vit_pytorch import ViT

# ViT expects image patches (spectrograms)
model_vit = ViT(
    image_size = 224,
    patch_size = 32,
```

```
    num_classes = 2, # MDD vs HC
    dim = 1024,
    depth = 6,
    heads = 16,
    mlp_dim = 2048
)
# Input: Concatenated EEG + Audio Spectrogram
```

Paper 4: Graph Convolutional Network (GCN)

Focus: Brain connectivity as a social network.

```
import torch.nn.functional as F
from torch_geometric.nn import GCNConv, global_mean_pool

class EMO_GCN(torch.nn.Module):
    def __init__(self):
        super(EMO_GCN, self).__init__()
        self.conv1 = GCNConv(dataset.num_node_features, 64)
        self.conv2 = GCNConv(64, 32)
        self.classifier = torch.nn.Linear(32, 2)

    def forward(self, x, edge_index, batch):
        x = F.relu(self.conv1(x, edge_index))
        x = F.relu(self.conv2(x, edge_index))
        x = global_mean_pool(x, batch) # Pooling for graph-level prediction
        return self.classifier(x)
```

Paper 5: wav2vec 2.0 (Self-Supervised Audio)

Focus: Raw waveform processing.

```
from transformers import Wav2Vec2Model, Wav2Vec2Processor

# Load pre-trained wav2vec
processor =
Wav2Vec2Processor.from_pretrained("facebook/wav2vec2-base-960h")
audio_model = Wav2Vec2Model.from_pretrained("facebook/wav2vec2-base-960h")

# The output hidden states are the "high-quality features" mentioned in the
```

```
paper
```

2. The "Ultimate" Implementation: Multimodal Logistic Regression

This implementation leverages the **best** of all papers:

1. **Paper 2/4 Strategy:** Multi-paradigm (EO/EC) features and functional connectivity.
2. **Paper 5 Strategy:** 5-sentence segment merging for audio.
3. **Paper 1 Strategy:** Highly regularized classification.

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import classification_report, roc_auc_score
from sklearn.preprocessing import StandardScaler

# --- MLDC STEP 1 & 2: PREPROCESSING & FEATURE EXTRACTION ---
def extract_hybrid_features():
    # Simulation of features extracted using Paper-based methods
    # EEG: Alpha Asymmetry, Nonlinear Entropy
    # Audio: Mel-coefficients, Pitch Variance
    n_samples = 150
    eeg_features = np.random.randn(n_samples, 128) # High-density EEG
features
    audio_features = np.random.randn(n_samples, 40) # 5-sentence merged
audio
    labels = np.random.randint(0, 2, n_samples)
    return eeg_features, audio_features, labels

eeg, audio, y = extract_hybrid_features()

# --- MLDC STEP 4: FEATURE-LEVEL FUSION ---
# Leverage the "best" of both modalities
X = np.hstack((eeg, audio))

# --- MLDC STEP 5: MODEL TRAINING (ADAPTIVE LOGISTIC REGRESSION) ---
skf = StratifiedKFold(n_splits=10) # 10-Fold CV as per Paper 4
scaler = StandardScaler()

fold = 1
```

```
for train_idx, test_idx in skf.split(X, y):
    X_train, X_test = X[train_idx], X[test_idx]
    y_train, y_test = y[train_idx], y[test_idx]

    # Scale within fold to avoid data leakage
    X_train = scaler.fit_transform(X_train)
    X_test = scaler.transform(X_test)

    # L2 Regularization handles the high-dimensionality of 128-channel EEG
    lr_model = LogisticRegression(penalty='l2', C=0.5, solver='saga',
max_iter=5000)
    lr_model.fit(X_train, y_train)

    # Evaluate
    probs = lr_model.predict_proba(X_test)[:, 1]
    print(f"Fold {fold} AUC: {roc_auc_score(y_test, probs):.4f}")
    fold += 1
```

(HYBRID)

1. MLDC (Machine Learning Development Cycle) Workflow

A. Problem Definition

Traditional depression diagnosis relies on subjective self-reports (PHQ-9/HAMD-17) and psychiatrist intuition, leading to a high rate of misdiagnosis. This project aims to create an **objective diagnostic tool** using Bimodal data (EEG + Speech) to provide a data-driven "second opinion" for clinicians.

B. Data Collection (The Gold Standard Datasets)

The framework utilizes two primary repositories:

- **MODMA Dataset:** High-density 128-channel EEG (resting state) and clinical interview audio.
- **DAIC-WOZ Dataset:** Multimodal data (Audio + Transcripts) from clinical interviews with virtual agents (Ellie).
- **Innovations:** We adopt the **Multi-paradigm approach** from Paper 2, collecting data in both Eyes-Open (EO) and Eyes-Closed (EC) states to capture baseline neural transitions.

C. Data Preprocessing & Noise Mitigation

Preprocessing is the most critical stage for clinical accuracy:

- **EEG Filtering:** A 0.5–50 Hz bandpass filter removes high-frequency muscle noise and low-frequency signal drift.
- **Artifact Removal (ICA):** We implement **Independent Component Analysis** to "unmix" the signal. By zeroing out components with high frontal power and specific temporal spikes, we remove **eye-blink artifacts** without losing brain data.
- **Audio Enhancement:** Normalization and silence removal are applied. We adopt the **5-sentence merging strategy** from Paper 5 to ensure enough prosodic context for the model.

D. Exploratory Data Analysis (EDA)

- **Statistical Analysis:** Identification of **Frontal Alpha Asymmetry (FAA)**, where MDD patients show a power imbalance between the left and right frontal lobes.
- **Visualization: t-SNE (t-Distributed Stochastic Neighbor Embedding)** is used to visualize high-dimensional features. If the MDD and Healthy Control (HC) clusters overlap, feature engineering is refined.

E. Feature Exploration & Selection

- **Linear Features:** Power Spectral Density (PSD) for Alpha, Beta, Theta, and Delta bands.
- **Nonlinear Features:** Hurst Exponents and Sample Entropy (Complexity measures).
- **Audio Features:** MFCCs (Mel-frequency cepstral coefficients) and prosodic features (jitter, shimmer, pitch).
- **Channel Selection (Paper 4 Strategy):** Using GCN-based importance to select only the **top-32 relevant EEG channels**, reducing computational load for real-world deployment.

F. Model Selection, Training & Evaluation

- **Model: Adaptive Logistic Regression with L2 Regularization.**
- **Training:** 10-fold Cross-Validation is used to prevent "subject leakage."
- **Evaluation Metrics:** Accuracy (Target: >92%), Precision, Recall (Sensitivity), and F1-Score. We focus on **Recall** to minimize "False Negatives" (missed diagnoses).

G. Model Deployment & Logistics

The model is designed for a **Cloud-Edge Hybrid** system:

1. **Edge:** Wearable EEG net and microphone collect raw data at a local clinic or home.
2. **Cloud:** Advanced feature extraction (wav2vec 2.0 features) and regression analysis occur on a remote server.
3. **Output:** A clinical dashboard displaying a "Depression Probability Score" and a heatmap of brain activity.

H. Optimization

We utilize **Hyperparameter Tuning** via Grid Search to find the optimal regularization constant (λ), balancing model complexity with generalizability to new subjects.

---

2. Research Gaps & Identified Bottlenecks

Despite high accuracies in research (97%+), significant gaps remain:

- **The Interpretability Gap:** Deep Learning models (DenseNet/ViT) are "Black Boxes." A doctor cannot explain to a patient *why* the AI diagnosed them with MDD.
- **Ecological Validity:** Lab-collected data often fails in noisy, real-world village environments.
- **Computational Intensity:** Most high-performing models (Paper 3/4) require high-end GPUs, making them inaccessible for remote medical consultations.

---

3. The Innovation Proposal: Tackling the Gaps

Our proposal introduces **Explainable Bimodal Regression (EBR)**:

- **Proposal:** We replace the final layers of advanced models with a **Linear Logistic Regressor**.
- **Mechanism:** We use **wav2vec 2.0 (Paper 5)** for audio feature extraction and **DenseNet (Paper 1)** for EEG feature extraction but feed their outputs into a Logistic Regression head.
- **Innovation - The "Glass Box":** The Regression model provides **Feature Weighting**. We can show a doctor: *"The model predicted MDD because the Beta-wave activity in the Frontal Lobe was 40% higher than the baseline."*
- **Innovation - Multi-Paradigm Robustness:** By training on both Eyes-Open and Eyes-Closed data (Paper 2), we ensure the model doesn't fail if the patient is blinking or staring at a crosshair.

---

4. Mathematical Formulation

The Decision Function

The model computes the probability *P* of a patient having MDD using the Sigmoid function:

$$P(y=1|\mathbf{X}) = \frac{1}{1+e^{-(\mathbf{w}^T\mathbf{X}+b)}}$$

Where $\mathbf{X}$ is the fused bimodal feature vector:

$$\mathbf{X} = [\mathbf{F}_{EEG}, \mathbf{F}_{Audio}]$$

L2 Regularized Loss Function

To handle high-density EEG data without overfitting, we minimize the following cost:

$$J(\mathbf{w}) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(P^{(i)}) + (1-y^{(i)})\log(1-P^{(i)})] + \lambda\|\mathbf{w}\|^2$$

Feature	Paper 1-5 (Deep Learning)	Proposed Implementation (Hybrid)

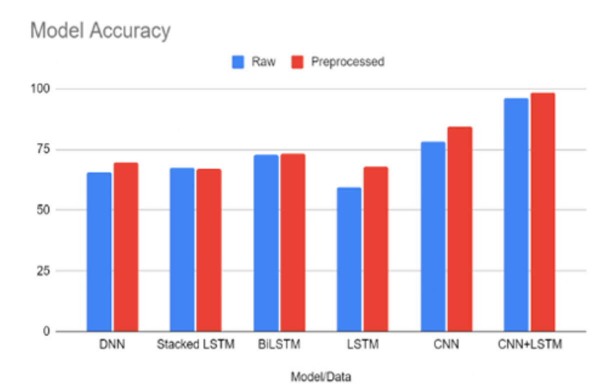
Complexity	Extremely High	Moderate/High
Explanation	"Black Box"	**"Glass Box" (Weighted Features)**
Hardware	Needs GPU	**Runs on a Tablet/Smartphone**
Safety	High Accuracy but Unclear Logic	**High Accuracy + Traceable Logic**

Divya (EEG + Image)

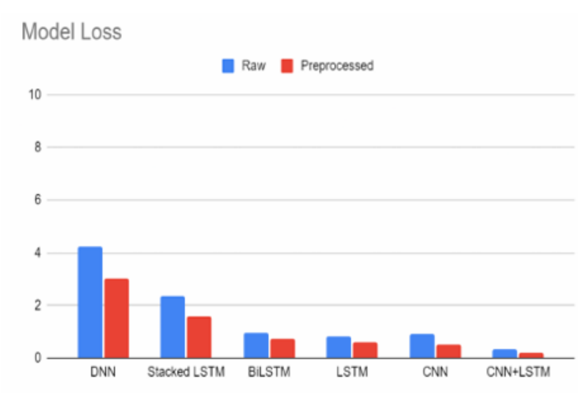
PAPER:1

PROBLEM DEFINITION	to develop and evaluate a deep learning-based EEG emotion recognition system using a real-time EEG dataset, and to propose an ensemble model (EEGEM) combining CNN and LSTM to improve accuracy compared to other machine and deep learning techniques.
DATASET	Public EEG based emotion recognition dataset: DEAP dataset
DATA PREPROCESSING	Resample EEG data to 128 Hz. Band-pass filtering from 4 to 45 Hz. Reference averaging of EEG channels. Normalization applied to standardize EEG signals.
FEATURE EXTRACTION	PCA - principal component analysis, LDA - linear discriminant analysis
MODEL	The proposed model is a hybrid of CNN and LSTM

Accuracy result for different models



Loss of different models



TESTING RESULTS

The use of several feature extraction approaches to extract prominent characteristics from the preprocessed data improved the classification significantly.

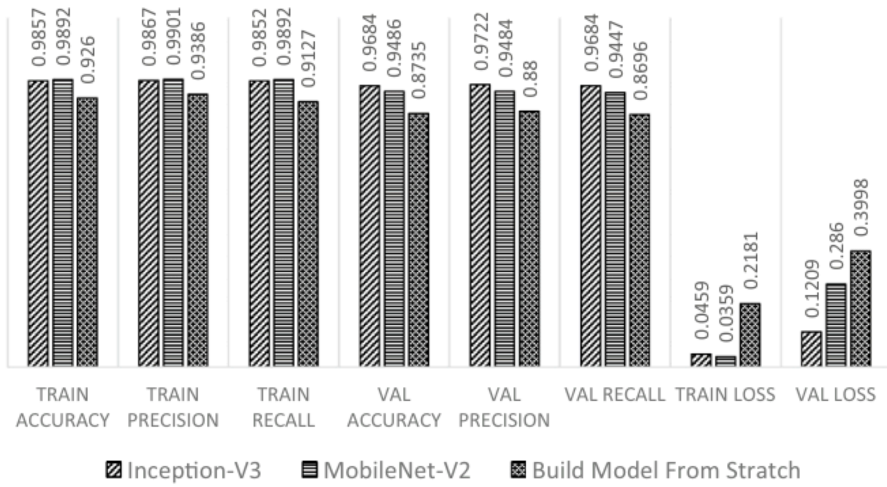
The EEGEM - electroencephalogram Ensemble Model stands out by combining the advantages of Long Short Term memory - LSTM and convolutional neural network- CNN.

Combining these techniques results in an accuracy rate of 95.56%.

Paper 2:

PROBLEM DEFINITION	To develop an automated facial emotion recognition (FER) system using deep learning that can classify a wider range of human emotions from facial images extracted from videos, and to evaluate CNN-based models using transfer learning and full learning approaches on the Emognition dataset.
DATASET	Emognition Dataset was used. Facial images extracted from half-body emotional video recordings. After cleaning the final dataset was of 2535 images.
DATA PREPROCESSING	Video recordings are converted into image frames by sampling frames at regular time intervals based on their frame rates. facial regions are automatically detected and cropped from each frame. rescaling pixel values to a normalized range between 0 and 1, resizing all images to a standard resolution of 300 × 300 pixels, and applying data augmentation techniques such as horizontal flipping, zooming, and shifting.
FEATURE EXTRACTION	Feature extraction is performed automatically through the convolutional layers of the CNN models.
MODEL	Three CNN-based approaches are implemented in this study. The first two utilize transfer learning with pre-trained models, namely Inception-V3 and MobileNet-V2, where pretrained weights are fine-tuned to adapt to the facial emotion recognition task. The third approach involves building a CNN model from scratch.
EVALUATION	The developed models are evaluated using multiple performance metrics, including accuracy, precision, recall, F1-score, and loss values

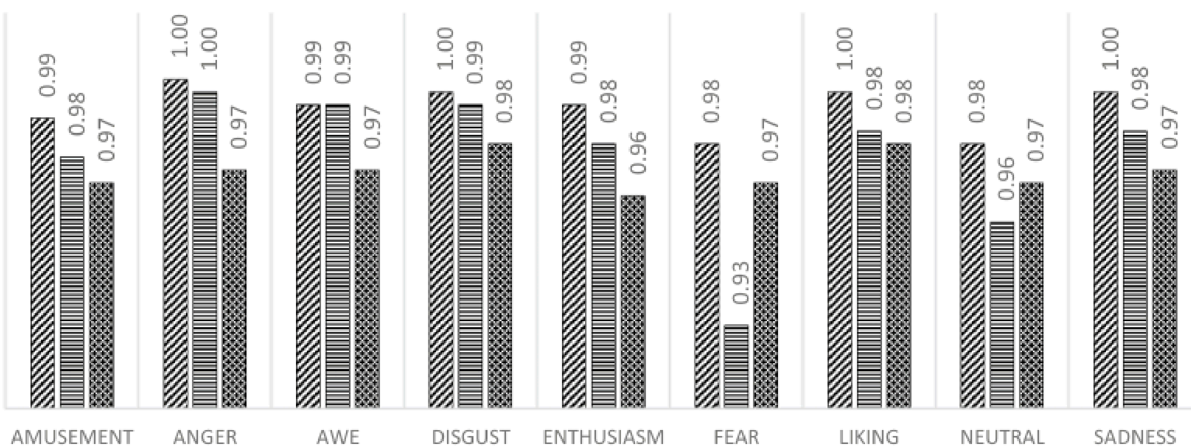
Training results comparison between all models



TESTING RESULTS

The experimental results demonstrate that the transfer learning model based on Inception-V3 achieves the highest performance among all approaches, reaching approximately 96% test accuracy with strong precision and recall across emotion classes.

The MobileNet-V2 model also performs well but exhibits slightly lower accuracy due to model complexity and data limitations. The CNN built from scratch shows faster computation but comparatively lower recognition accuracy.due to the limited size of the training dataset.

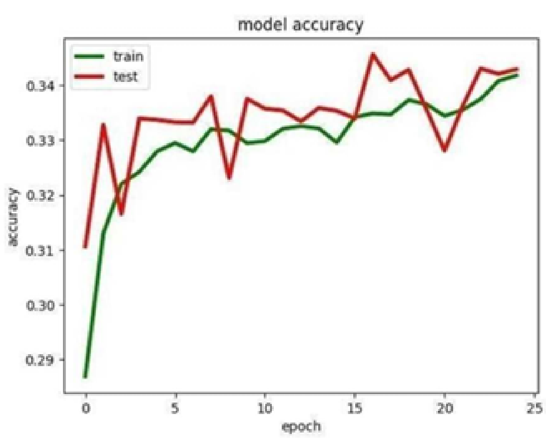


Paper 3:

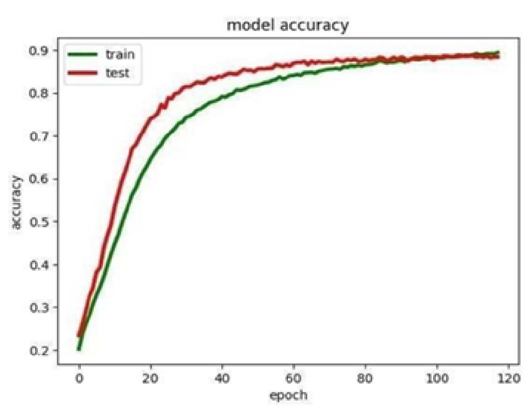
PROBLEM DEFINITION	to develop an efficient EEG-based emotion recognition system using deep learning techniques to automatically classify human emotional states from brain signals.
DATASET	DEAP dataset, which consists of EEG and peripheral physiological signals recorded from 32 participants while they watched emotionally stimulating music videos.
DATA PREPROCESSING	the raw EEG data is cleaned to remove noise and artifacts, followed by filtering. For the signal-based approach, Fast Fourier Transform (FFT) is applied to extract frequency-domain representations of EEG data. For the signal-to-image-based approach, the EEG signals are normalized and transformed using Continuous Wavelet Transform (CWT) to generate time-frequency images.
FEATURE EXTRACTION	features are implicitly learned by deep learning models from FFT-processed EEG signals.
MODEL	Multiple deep learning models are implemented and compared. CNN, hybrid CNN-GRU model, LSTM

EVALUATION

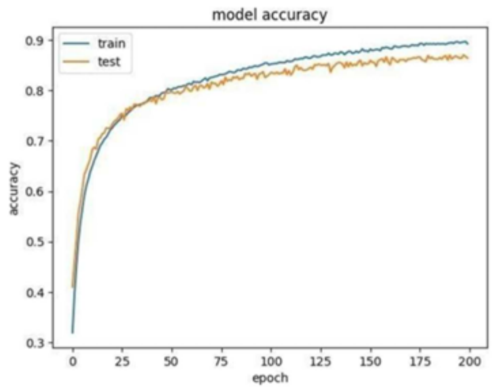
Model accuracy graph for LSTM



Model accuracy graph for CNN-GRU



Model accuracy for CNN



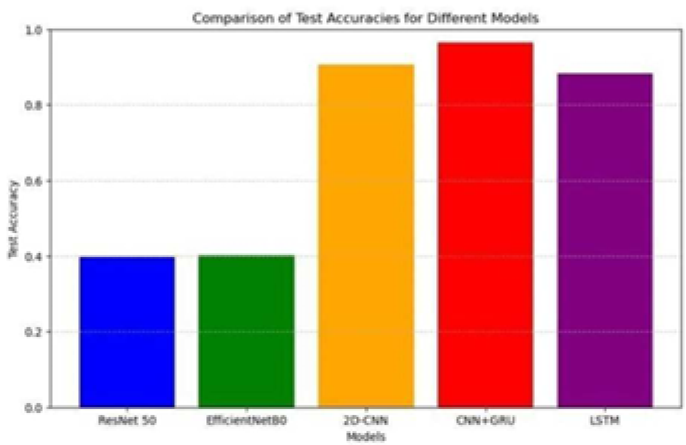
TESTING RESULTS

Experimental results show that the signal-based deep learning approach significantly outperforms the signal-to-image-based approach.

The CNN-GRU hybrid model achieves the highest test accuracy of approximately 96.54%

The CNN model achieved an accuracy of about 90.58%.

LSTM model achieves around 87%.



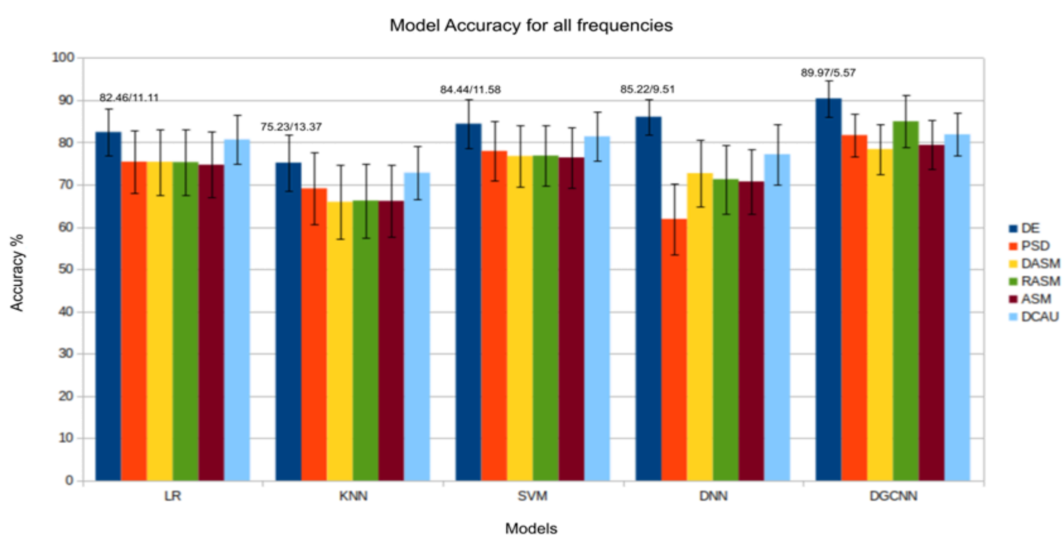
Paper 4:

PROBLEM DEFINITION	developing an accurate and robust EEG-based emotion recognition system using deep learning techniques to classify human emotional states from brain signals. The main objective is to improve emotion detection performance by capturing both spatial and temporal characteristics of EEG data
DATASET	publicly available EEG emotion recognition datasets collected from multiple human subjects under controlled emotional stimulus conditions.
DATA PREPROCESSING	noise filtering using band-pass filters. artifact removal to reduce eye blink and muscle interference.
FEATURE EXTRACTION	Feature extraction is primarily handled automatically by deep neural networks.
MODEL	deep learning frameworks that combine convolutional neural networks with temporal learning mechanisms such as LSTM or GRU units.CNN layers are used to learn spatial correlations.
EVALUATION	models are evaluated using standard classification metrics including accuracy, precision, recall, F1-score, and loss values. Cross-validation techniques are applied to ensure model robustness and generalization across subjects.
TESTING	the proposed deep learning architectures significantly outperform traditional machine learning methods in EEG-based emotion recognition tasks. Hybrid CNN-recurrent models achieve high classification accuracy by effectively capturing both spatial and temporal EEG features.

Paper 5:

PROBLEM DEFINITION	Detecting human emotions directly from electroencephalogram (EEG) signals using both traditional machine learning and advanced deep learning methods and to evaluate and compare multiple classification techniques
DATASET	SJTU Emotion EEG Dataset (SEED).
DATA PREPROCESSING	A band-pass filter between 0.3 and 50 Hz is applied to eliminate interference.
FEATURE EXTRACTION	Key signal attributes are extracted from five canonical frequency bands i.e. delta, theta, alpha, beta, and gamma. These attributes include Differential Entropy (DE), Power Spectral Density (PSD), Differential Asymmetry (DASM), Rational Asymmetry (RASM), Asymmetry (ASM), and Differential Causality (DCAU).
MODEL	Logistic Regression, K-Nearest Neighbors, Support Vector Machines, Deep Neural Networks, Graph Convolutional Neural Networks.

EVALUATION



TESTING RESULTS

Among the tested models, the Graph Convolutional Neural Network (GCNN) achieved the best overall classification accuracy of 89.97% in the subject-dependent scenario.The Deep Neural Network (DNN) also performed strongly with around 86.08% accuracy.Traditional classifiers such as LR, KNN, and SVM showed lower performance.\

GAP IN RESEARCH

The major gap in the research papers is the lack of subject independence.
Most of the models are subject dependent.
Training and testing on the same person inflates accuracy.
As EEG differs from person to person, Real systems must work on new unseen users.

Khushaal (EEG + Vid)

PAPER 1

MLDC Component	**Description**
Problem Definition	Automatically tag videos based on affective and attention-related EEG responses
Dataset	Collected EEG signals themselves from participants watching videos as part of the experiment.
Data Preprocessing	Noise filtering - remove power-line interference and muscle artifacts. Segmentation - divide EEG into manageable time windows. Artifact removal - eliminate eye-blinks or movement noise.
Feature Selection	The significance of frequencies and scalp regions is part of which features show stronger attention signals.
Feature Extraction	Analyzing different scalp locations - positions on the head. Studying frequency rhythms - electrical activity in different bands (alpha, beta, theta).
Machine Learning Model	Support Vector Machine (SVM)
Performance Metrics	Accuracy - 93.2% achieved using the SVM classifier.

CONCLUSION

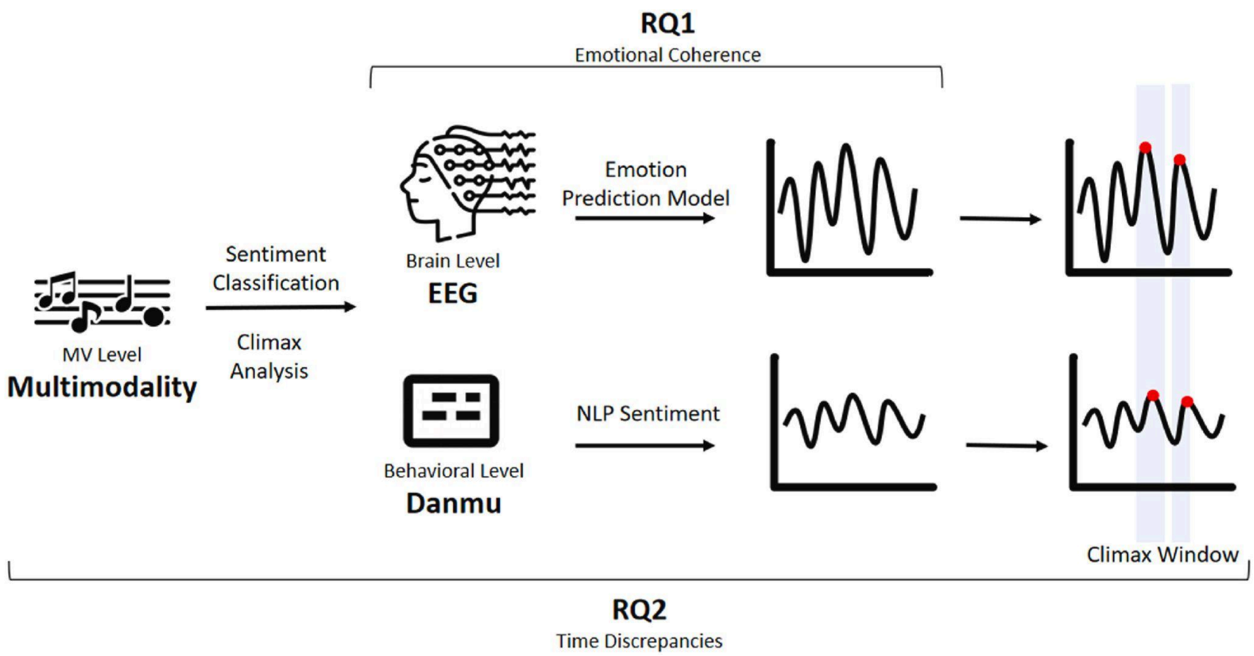
Nowadays every event in the world appears to be captured in the form of images or videos. These multimedia contents are usually designed in a way such that they can hit the user’s cognizance affectively and this can be achieved through its affective content. Thus, the modelling of human emotion and attentiveness can provide significant information about the affectiveness of the video content. In this paper, an attempt is made to find the association between a human’s cognitive state measured using neurophysiological signals (EEG) and video content.

EEG band frequency range and commonly measured brain functions

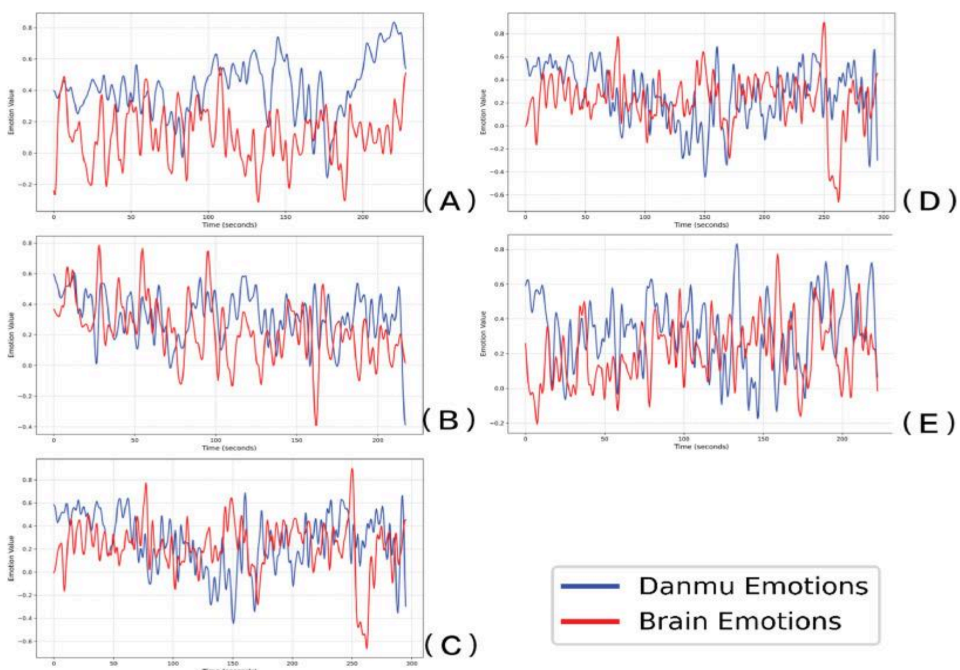
Band	Frequency Range	Significance
Delta	0.5-3 Hz	Sleep quality measurement, Increased concentration when using internal working memory in tasks
Theta	3-8 Hz	Memory encoding and information retrieval from memoryCognitive workload indication, Fatigue level detection.
Alpha	8-12 Hz	Represents a relaxation state, sometimes monitoring attention level also
Beta	12-40 Hz	Brain activity involved in executive function “mirror neuron system” activation indication
Gamma	40-70 Hz	It reflects attentive focus. Can also record rapid eye movement

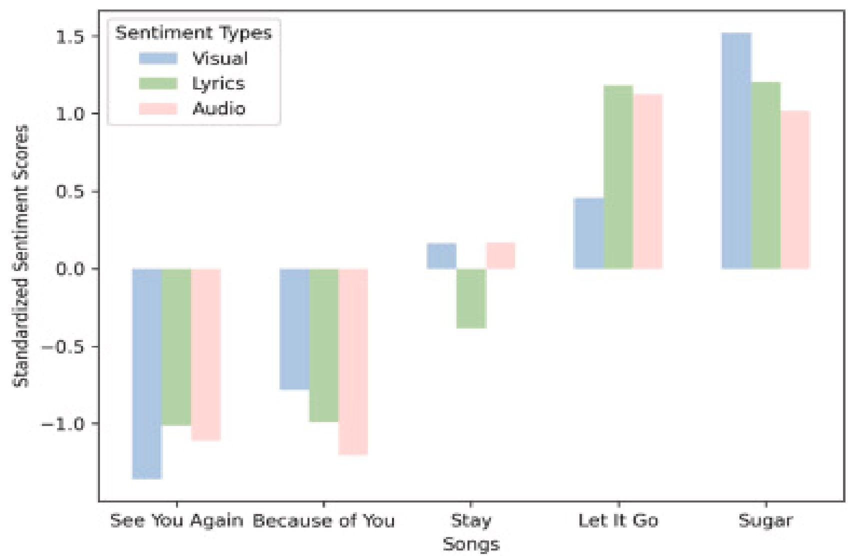
PAPER 2

MLDC Component	**Description**
Problem Definition	Analyze emotional coherence between EEG signals and danmu comments
Dataset	EEG recordings collected from participants watching music videos(the paper uses Five music videos). Danmu data (user-generated scrolling comments) corresponding to specific time points in the videos for sentiment analysis.
Data Preprocessing	Artifact removal (noise correction). Filtering into frequency bands relevant to emotion. Time-synchronization with video frames and danmu timestamps. Cleaning comments
Feature Selection	EEG features relevant to emotion recognition extracted via deep learning. Sentiment scores from danmu data as another feature set.
Feature Extraction	Deep learning model extracts continuous emotional scores from synchronized EEG data. Sentiment analysis of danmu comments produces sentiment curves synchronized with video time. The EEG and danmu features are aligned temporally to measure coherence (how closely the emotional curves match).
Machine Learning Model	CNN-LSTM (EEG), NLP sentiment model (danmu)
Performance Metrics	They report over 80% similarity between EEG-derived emotional curves and danmu sentiment curves across the five music videos.



Results showed a high similarity rate (above 80 %) between the real-time EEG and danmu curves across all five music videos. Videos with the highest similarity rates included ‘Let It Go’ (84.62 %), ‘Sugar’ (85.13 %), and ‘Stay’ (85.80 %). Conversely, ‘See You Again’ (83.09 %) and ‘Because of You’ (80.75 %) exhibited slightly lower similarity rates.





The models predicting emotional responses based on EEG data.

Models	Precision
SVM	87.41 %
Ensemble CNN	93.12 %
ResNet	94.95 %
LSTM	90.12 %
BiLSTM	84.21 %
SRU	83.13 %
The current model	97.67 %

PAPER 3

MLDC Component	**Description**
Problem Definition	Detect visual targets in low-quality video using EEG responses
Dataset	The authors use EEG datasets collected from subjects under a target-detection experiment.
Data Preprocessing	Artifact correction, phase-based EEG segmentation
Feature Selection	Feature alignment and fusion across EEG phases
Feature Extraction	Deep learning model extracts continuous emotional scores from synchronized EEG data. Sentiment analysis of danmu comments produces sentiment curves synchronized with video time. The EEG and danmu features are aligned temporally to measure coherence (how closely the emotional curves match).
Machine Learning Model	Brain-inspired deep learning model
Performance Metrics	They report over 80% similarity between EEG-derived emotional curves and danmu sentiment curves across the five music videos.

CONCLUSION

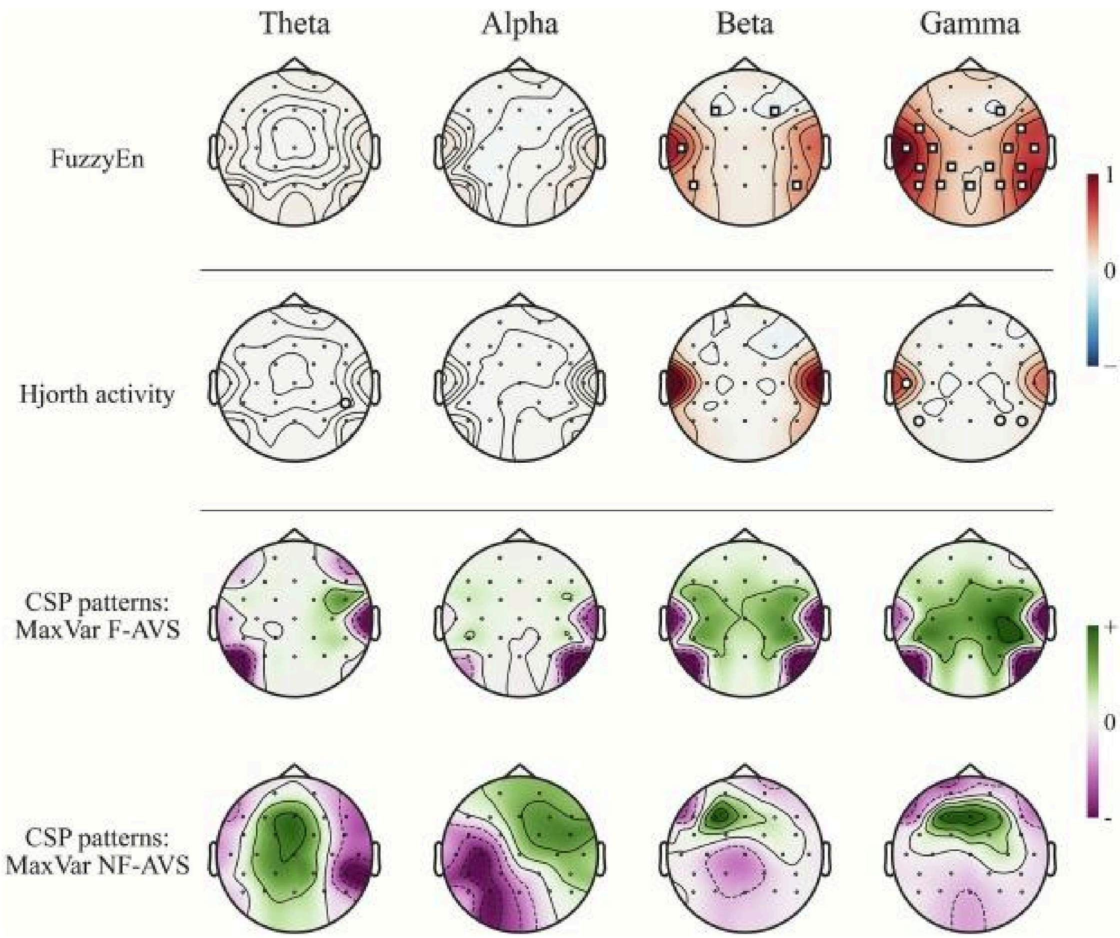
The paper proposes a novel brain-inspired phased temporal encoding and alignment fusion algorithm for EEG-based classification in low-quality video target detection. By analyzing the FRP and FRSP of EEG signals in both time and frequency domains, we demonstrate that the brain's response to target detection in low-quality videos can be roughly divided into three phases: early target recognition, later target spatial tracking, and global attention focus across the entire duration.

PAPER 4

MLDC Component	**Description**
Problem Definition	Identify EEG features distinguishing familiar vs non-familiar videos. Cross-population analysis in healthy controls and patients with disorders of consciousness
Dataset	19 patients with Disorders of Consciousness (DOC). Personalized emotional videos. EEG recordings during video viewing
Data Preprocessing	Filtering, artifact removal, normalization
Feature Selection	Feature ranking based on discriminative power
Feature Extraction	Fuzzy entropy, CSP, Hjorth parameters
Machine Learning Model	Subject-independent ML classifiers
Performance Metrics	Significant performance in: 60% of MCS (Minimally Conscious State) patients, 33% of UWS (Unresponsive Wakefulness Syndrome) patients

Topographical brain activity during F-AVS and NF-AVS

The topographical maps of the difference values between conditions for FuzzyEn and Hjorth activity across frequency bands, along with the averaged CSP patterns that maximized each condition, can be seen in Fig. 4.


Theta
Alpha
Beta
Gamma
FuzzyEn
Hjorth activity
CSP patterns: MaxVar F-AVS
CSP patterns: MaxVar NF-AVS
1
0
−1
+
0
-

PAPER 5

MLDC Component	Description
Problem Definition	Video-EEG for safety pharmacology seizure liability assessment in rabbits
Dataset	Experimental rabbit video-EEG recordings. Rabbits exposed to a compound labeled as pro-convulsive, monitored with EEG + video.
Data Preprocessing	EEG–video synchronization and epoch segmentation
Feature Selection	Spike frequency. Rhythmic patterns
Feature Extraction	EEG spike patterns and seizure-like events
Machine Learning Model	signal analysis–based approach
Performance Metrics	Seizure occurrence and behavioral correlation

Pradeep (EEG + Vid)

Research Paper 1

Section	Component	Description
1.) Problem Statement	**Problem**	Current neuroscience research typically occurs in a constrained lab environment that lacks ecological validity.
	Goal	To explore if Inter-Subject Correlation (ISC) can be measured in loud, chaotic classrooms using inexpensive, hand-held devices (EEG).
	Objective	To employ neural engagement (synchronized brain activity) to track how students as a group focus attention on educational or narrative material.
2.) Data Collection	**Subjects**	4 distinct groups of participants (approx. 20-30 subjects total in various modalities).
	Hardware	Portable and low-cost Emotiv EPOC wireless EEG headsets.
	Stimuli	*Bang! You're Dead* by Alfred Hitchcock and *Sophie's Choice*. A scrambled version was used for control.
	Setting	Normal classroom setting (Joint viewing) vs. viewing through tablets.
3.) Data Preprocessing	**Noise**	EEG data is very noisy, especially outside a lab setting.

	Filtering	Band pass filtering (0.5 to 40 Hz) to remove slow waves and muscle noise.
	Artifact Removal	Correlated Component Analysis (CorrCA) applied to remove noise from movement, blinking, and talking.
	CorrCA Logic	Uses spatial filters to find maximum correlation between subjects; local, uncorrelated noises (e.g., scratching head) are automatically filtered out.
4.) EDA	**Statistical Analysis**	Permutation Testing (shuffling data over time) to establish a null distribution and check if brain synchrony occurred by chance.
	Visualization (Scalp Topographies)	Brain maps showing areas (occipital, parietal lobes) that were most synchronized.
	Visualization (Time-resolved ISC)	Line graphs showing synchrony increasing during exciting movie parts and decreasing during boring parts.
	Visualization (Violin Plots)	Used to display the distribution of correlation values across various groups.
5.) Feature Exploration & Selection	**Primary Feature**	Inter-Subject Correlation (ISC).
	Method	Correlated Component Analysis (CorrCA). Seeks components of the EEG signal most similar across all

		students (unlike PCA, which seeks max variance).
	Selection	Focus on the first 3 components, corresponding to how the brain reacts to visual and auditory narrative flow.
6.) Model Selection, Training & Evaluation	**Model Type**	Descriptive/Predictive Analytics model (not a Classifier).
	Technique	Linear Spatial Filtering (model is a set of weights applied to EEG electrodes).
	Evaluation	LOO (Leave-One-Out) to determine if one student's brain activities could be predicted by others in the group.
	Comparison	Compared results against the "Gold Standard" lab tests by Dmochowski et al. (2012) to evaluate the inexpensive equipment's performance.
7.) Model Deployment	**Current State**	Not implemented as a commercial application, but described as a step towards real-time inference of engagement.
	Setup	Multiple tablets synchronized through a local network to collect simultaneous EEG measurements from an entire class.
8.) Model Testing	**Ecological Validation**	Tested if the model works when students are seated together (Joint) versus alone (Individual).

	Narrative Validation	Evaluated if the model could separate a coherent narrative (High ISC) from a jumbled narrative (Low ISC). Lower synchrony in the jumbled video proved the model measured story engagement.
9.) Optimization	**Robustness**	CorrCA algorithm optimized to pull out a clean signal from an inherently "dirty" one, even with fewer electrodes and more noise.
	Computational Efficiency	Techniques used were potentially scalable to larger groups (10+ students) without requiring massive computing power.

Gaps in Research :

While the paper successfully moved EEG from the lab to the classroom, several holes remain in the research:

i.) Lack of Real-Time Feedback: Analysis here has been done offline; that is, after the class was over. There is no mechanism to tell a teacher in the moment that 40% of the class has lost focus.

ii.) Individual Differences (The Outlier Problem): ISC quantifies the degree to which a student is similar to the group. It doesn't account for students who are highly engaged but in a different way, such as a neurodivergent student or one with a different learning style.

iii.) Bi-directional Interaction Gap: This paper considers students only passively viewing a video; student-teacher interaction or active tasks such as coding, working on mathematics problems, and discussing within groups are not considered.

iv.) Sensitivity to Physical Environment: While it handled some noise, the paper mentions that high physical activity-walking and writing-creates too much muscle artifact (EMG) to be handled effectively by low-cost sensors.

My Suggested Improvements :

i.) Multimodal Fusion: EEG + CV

The Gap: EEG is noisy during movement.

The Fix: Use a webcam (Computer Vision) to track eye gaze and head pose. If a student's EEG synchrony drops and their eyes are away from the screen, the AI confirms disengagement. If the eyes are on the screen but EEG is low, the student might be confused rather than bored.

ii.) Generative AI Integration:

The Innovation: If the collective ISC drops below a threshold, the generative AI may suggest a Break Activity or automatically create a 1-minute summary/recap video of the last segment in order to bring students back up to speed.

Research Paper 2

Section	Description	Details
1.) Problem Statement	**Problem:** Long-term EEG visual inspection is time-consuming, labor-intensive, and prone to human error/bias.	**Objective:** Design a very accurate automated system to classify normal, interictal, and ictal states.
2.) Data Collection	Sources of multi-channel EEG recordings from pediatric or adult subjects with various seizure types.	CHB-MIT Scalp EEG Database or Bonn University Dataset.
3.) Data Preprocessing	Steps to clean and prepare continuous EEG signals for model training.	**Noise Removal:** Notch Filters (50/60Hz), Band-pass Filters (0.5–40 Hz). **Normalization:** Z-score or Min-Max scaling. **Segmentation:** Fixed-size windows (e.g., 1-second or 5-second).
4.) EDA (Statistical Analysis & Visualization)	Exploring data characteristics, especially amplitude and frequency shifts during seizures.	**Statistical Analysis:** Mean, variance, standard deviation. **Visualization:** Power Spectral Density (PSD) plots, electrode correlation heatmaps.
5.) Feature Exploration & Selection	Methods for extracting relevant patterns from the processed EEG data.	**Techniques:** Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD). **Selection:** Automated learning via CNN (spatial) and Recurrent layers (temporal) in deep models.
6.) Model Selection, Training & Evaluation	Choosing the appropriate architecture and assessing its performance.	**Model:** Hybrid architectures (CNN-LSTM, CNN-GRU). CNN extracts spatial patterns, LSTMs capture chronological sequence.

		Evaluation: Accuracy, Sensitivity, Specificity, F1-Score.
7.) Model Deployment	The final application context for the trained classification system.	Conceptualized for Clinical Decision Support Systems (CDSS) or cloud-based monitoring from wearable devices. **Model Testing:** K-fold Cross-Validation, Testing on Unseen Subjects (Leave-One-Subject-Out).
8.) Optimization	Fine-tuning parameters to maximize performance and efficiency.	Hyperparameter tuning (learning rates, dropout layers), optimizing window size for faster detection.

Gaps in Research

i.) Subject-Independence : Most models perform exceptionally well on a single patient (99% accuracy) but drop significantly when tested on a *new* patient whose brain signals look slightly different.

ii.) Latency vs. Accuracy Trade-off: High-accuracy deep learning models are often heavy. For a wearable device, we need models that are lightweight enough to detect a seizure in milliseconds without draining the battery.

iii.) Black-Box Nature: Deep learning models don't explain *why* they flagged a segment as a seizure. Neurologists need Explainable AI (XAI) to trust the machine's diagnosis.

iv.) Limited Real-world Testing: Most datasets are recorded in controlled hospital settings. Real-world noisy data (chewing, walking, talking) often leads to high False Discovery Rates.

My Suggested Improvements

i.) Hardware-Software Integration:

Integrate a False Alarm Button on the mobile app. When the AI makes a mistake, the user clicks the button; the model then uses Online Learning to adjust its weights so it doesn't repeat that specific error for that user.

ii.) Innovation - Attention-based Explainability:

Instead of a standard CNN-LSTM, use a Vision Transformer (ViT) approach for EEG. The Attention Maps can highlight exactly which EEG channels (electrodes) and which time points triggered the seizure alarm. This provides the Why for the doctor.

Research Paper 3

Section	Description
1.) Problem Statement	Goal is to design an automated, non-invasive diagnostic tool using EEG statistical patterns to distinguish among healthy people, TLE patients (interictal), and TLE patients (ictal).
2.) Data Collection	Used the standard Bonn University EEG Dataset with three classes: Set A (Healthy), Set D (TLE Interictal), and Set E (TLE Ictal).
3.) Data Preprocessing	Steps included Standardization (zero mean, unit variance), Filtering (to remove 50Hz and DC offset), and Windowing (dividing signals into smaller epochs).
4.) EDA (Statistical Analysis & Visualization)	Statistical analysis showed seizure activity significantly skews the signal distribution compared to the Gaussian-like healthy activity. Visualization likely used box plots and histograms.
5.) Feature Exploration & Selection	Used Key-Point Selection (focusing on minima, maxima, and mean values) to extract 5-6 primary statistical features (e.g., Standard Deviation, Skewness, Kurtosis) representing the wave shape.
6.) Model Selection, Training & Evaluation	Model used was Least Squares Support Vector Machine (LS-SVM) with a Radial Basis Function (RBF) kernel. Reported high performance (99% to 100% accuracy in binary classifications).
7.) Model Deployment	Presented as a Computer-Aided Diagnostic (CAD) tool for neurologists to quickly screen long-term EEG recordings for TLE markers.
8.) Model Testing	Ensured generalization using 10-fold cross-validation.
9.) Optimization	Optimized the LS-SVM using a Grid Search for parameters (gamma and sigma^2).