Artistic Visualization of Dream using EEG

This project aims to decode dreams from EEG data. The goal is reconstructing visual dream content using EEG Signals.

By Pavan Konam, Ashlesha Ahirwadi

Background Research

- EEG (Electroencephalography) captures **brain activity** as **electrical signals**.

- Types of EEG waves

1. Delta (0.5–4 Hz) → Deep sleep, unconscious states
2. Theta (4–8 Hz) → Dreaming, creativity, memory processing
3. Alpha (8–12 Hz) → Relaxation, calm wakefulness
4. Beta (12–30 Hz) → Active thinking, problem-solving
5. Gamma (30+ Hz) → High-level cognition, perception

- Current Solution - DreamDiffusion model.

Dataset Source & Format

The following datasets were used for this project. They consist of EEG data in EDF format accompanied by text files containing dream descriptions. Each dataset originates from a different study and language.

Source	Format	Language
Zhang & Wamsley 2019	EDF and Text files	English
Oudiette_N1Data	EDF and Text files	French
LODE	EDF and Text files	Italian
TWC_USA	EDF and Text files	Conversation-English
Donders	EDF and Text files	Conversation-English

Datasets & Synthetic Data Creation

Diverse EEG Data

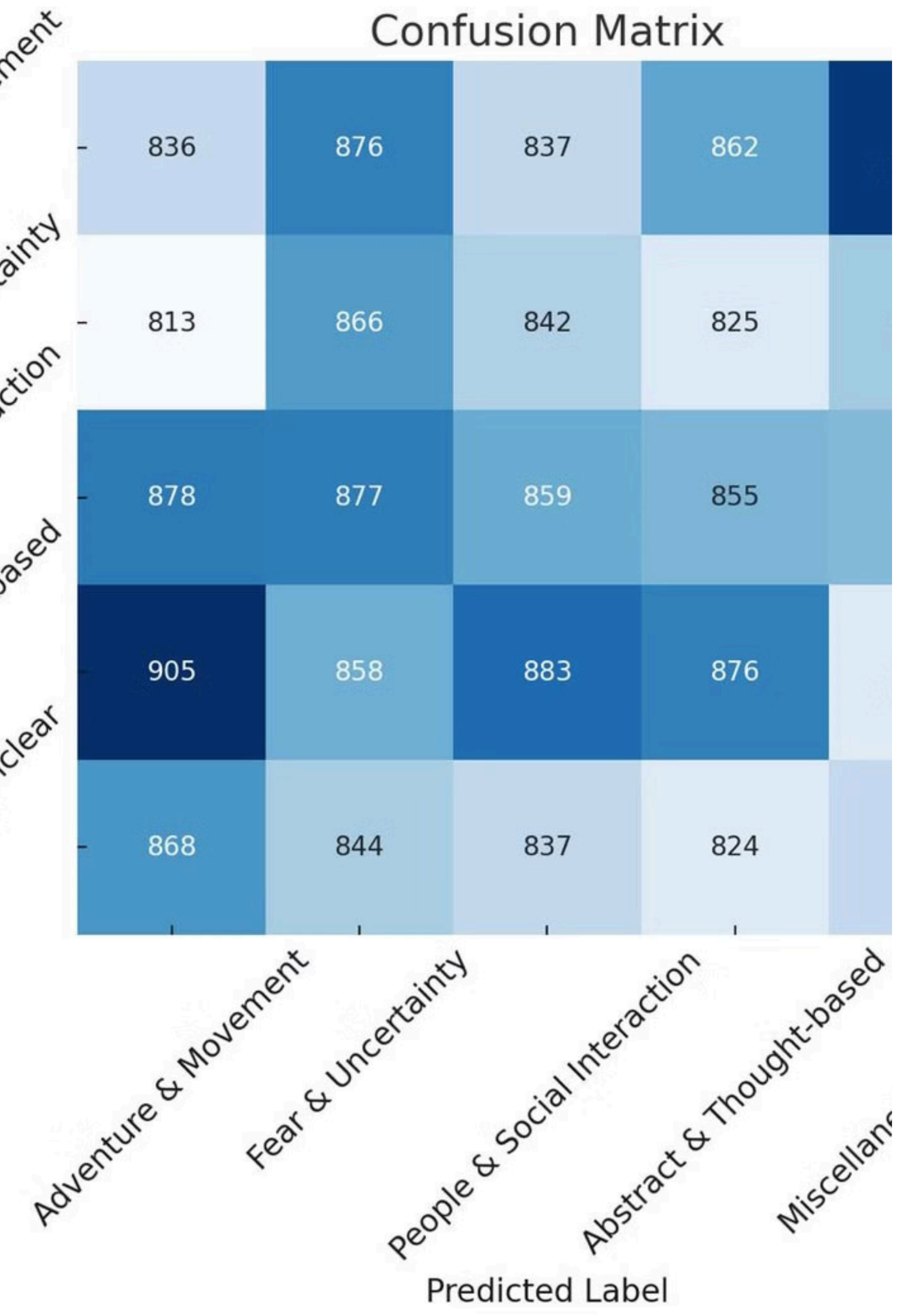
Utilizing 5 different datasets capturing sleep, dream reports, and visual imagery.

Data Combination & Categorization

Matching EEG segments to dream content keywords and categorizing them under 5 classes

Different Approach For Data Collection

Data collection approaches varies in terms of - language of text data, method of conversations and approach of Data Sampling



Data Categories

- **Adventure & Movement** (e.g., "going," "went," "see," "where") – Dreams about traveling, exploring, or movement.
- **Fear & Uncertainty** (e.g., "no," "think," "thought," "didn’t") – Dreams involving fear, confusion, or distress.
- **People & Social Interaction** (e.g., "she," "him," "people") – Dreams with conversations, relationships, or interactions.
- **Abstract & Thought-based** (e.g., "thinking," "know," "say") – Dreams focused on thoughts, realizations, or abstract ideas.
- **Miscellaneous & Unclear** (e.g., "some," "something," "things") – Dreams that don’t fit neatly into one category.

EEG Data Preprocessing

1. ## Feature Extraction

   - Highpass and Bandpass filter applied to extract waves
   - Filename,Channel,Band,PSD_Mean,PSD_Std,Mean,Variance,Skewness,Kurtosis

2. ## Data Cleaning

3. ## Handling Missing Values

Text Data Preprocessing

1. Converted txt file to CSV
2. Translate text to english
3. Extracted Keywords to Classify them into 5 categories
4. Combined all the 5 csv dataset in single csv

Model Development: Baseline (Random Forest)

1

Data Split

Train/Test/ Validation split: Training Set: (85521, 10) Validation Set: (21381, 10) Test Set: (11878, 10)

2

Model

Random Forest: simple, interpretable, handles non-linearity.

3

Metrics

```
✅ Validation Accuracy: 0.8819
✅ Test Accuracy: 0.8797

Classification Report (Validation Set):
               precision    recall  f1-score   support

           0       0.83      0.97      0.90      9019
           1       0.95      0.75      0.84      2712
           2       0.99      0.67      0.80       458
           3       0.95      0.80      0.87      3709
           4       0.91      0.87      0.89      5483

    accuracy                           0.88     21381
   macro avg       0.93      0.81      0.86     21381
weighted avg       0.89      0.88      0.88     21381

✅ Model saved successfully!
```

Model Development: CNN with Spectrograms

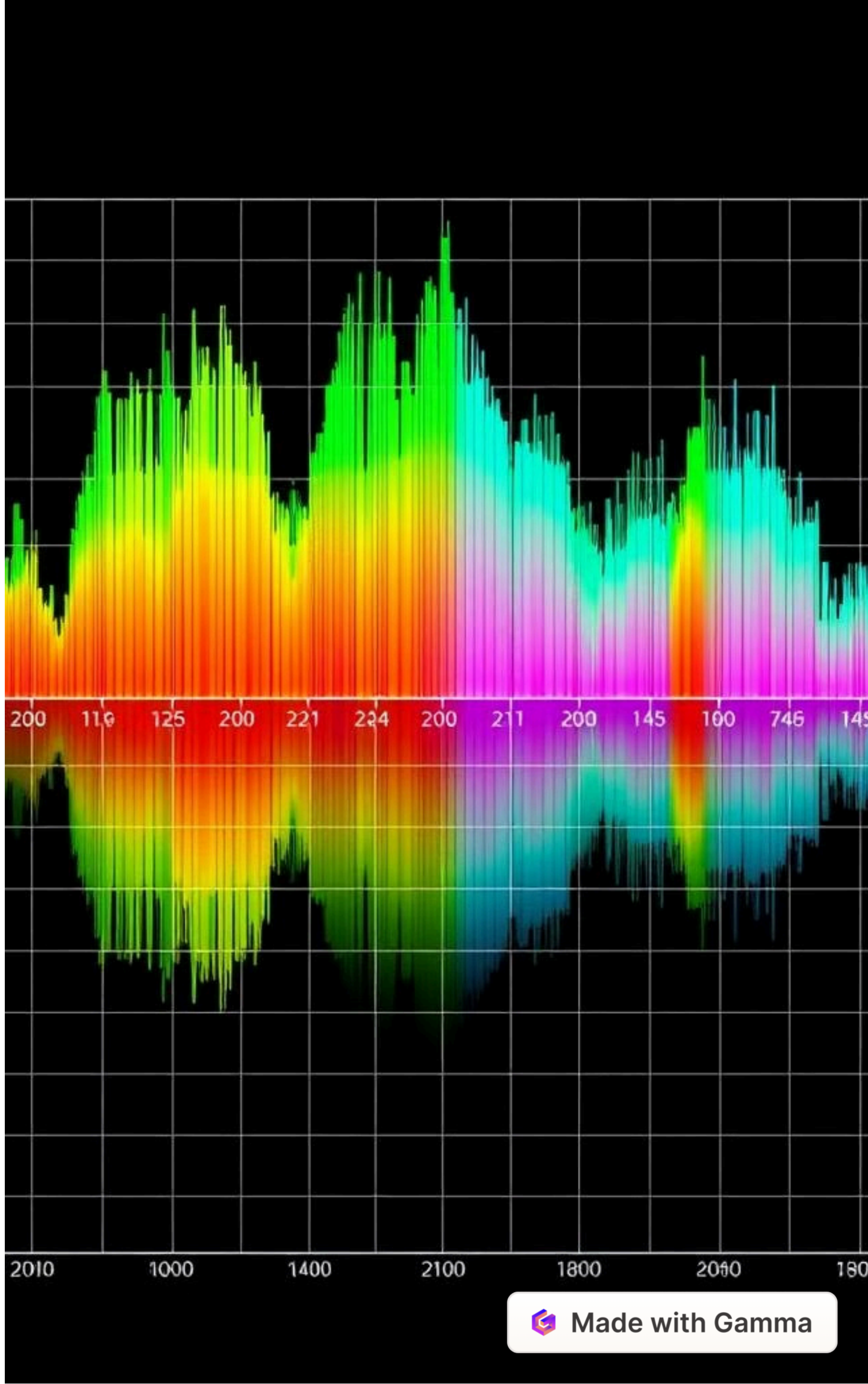
- **EEG Transformation**: Converted EEG time-series to time-frequency representations using STFT or Wavelet Transform.
- **Spectrogram Purpose**: Visualizes how signal frequency content changes over time, revealing hidden patterns in EEG signals.
- **Non-Stationary Nature**: EEG signals change over time; spectrograms show how frequency bands behave during dreams.
- **Dream State Classification**: Different dream states may have distinct frequency distributions, aiding classification.

Dataset Changes for CNN

1. Spectrogram
   - Converting EEG signals into spectrogram images.
   - Each EDF file can create multiple spectrograms, one for each EEG channel.
   - If an EDF file contains 32 EEG channels, it will generate 32 spectrograms. -
   - Total number of spectrograms = number of EDF files × number of EEG channels per file.

2. Mapping with Categories

3. New .csv file for CNN



CNN Model Training

Architecture : CNN

- 3 convolutional layers (extracting spatial patterns in spectrograms).
- Max pooling layers (reducing spatial dimensions while preserving features).
- Fully connected (dense) layers (classifying the image into 5 categories).
- Dropout layer (preventing overfitting).

Optimizer & Loss

- Model is trained using the Adam optimizer and CrossEntropy loss for 30 epochs.
- Training accuracy and loss are plotted to analyze model performance.

Training Time

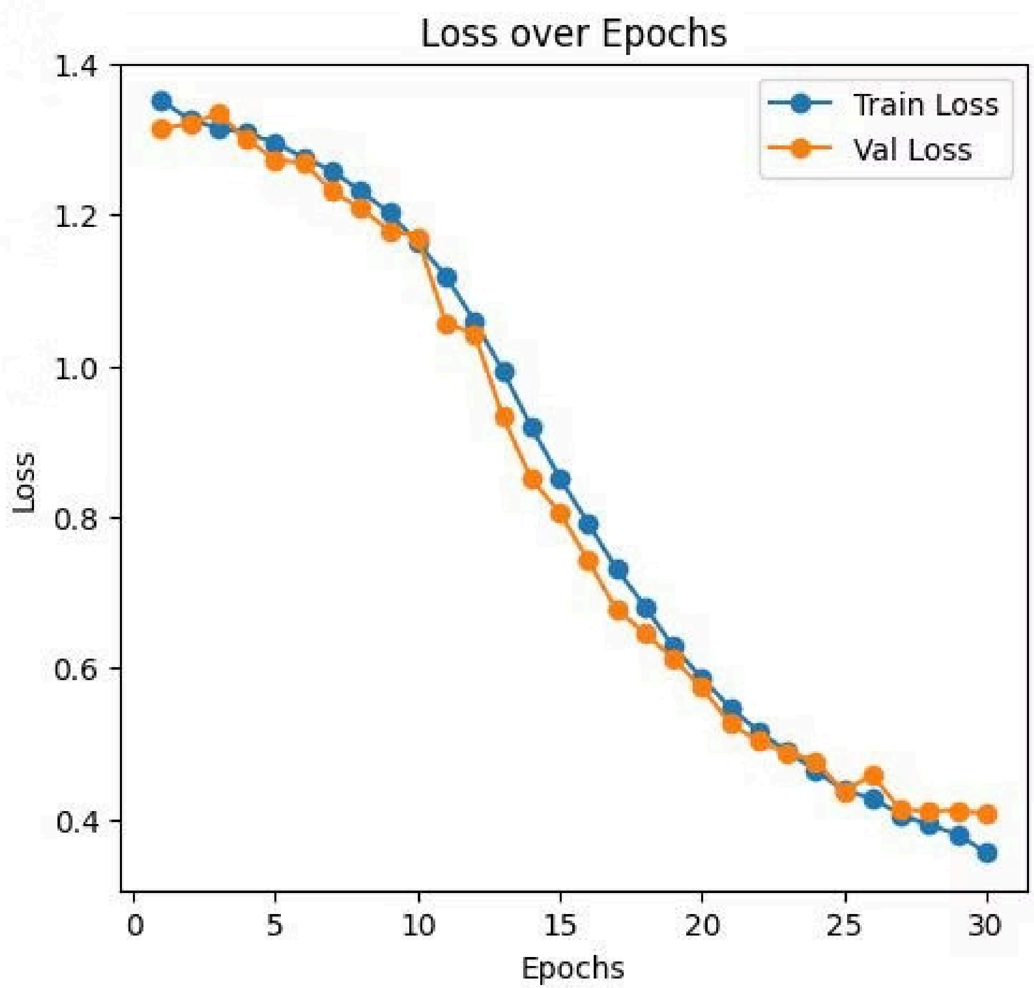
Approximately 4 hours for 30 epochs.

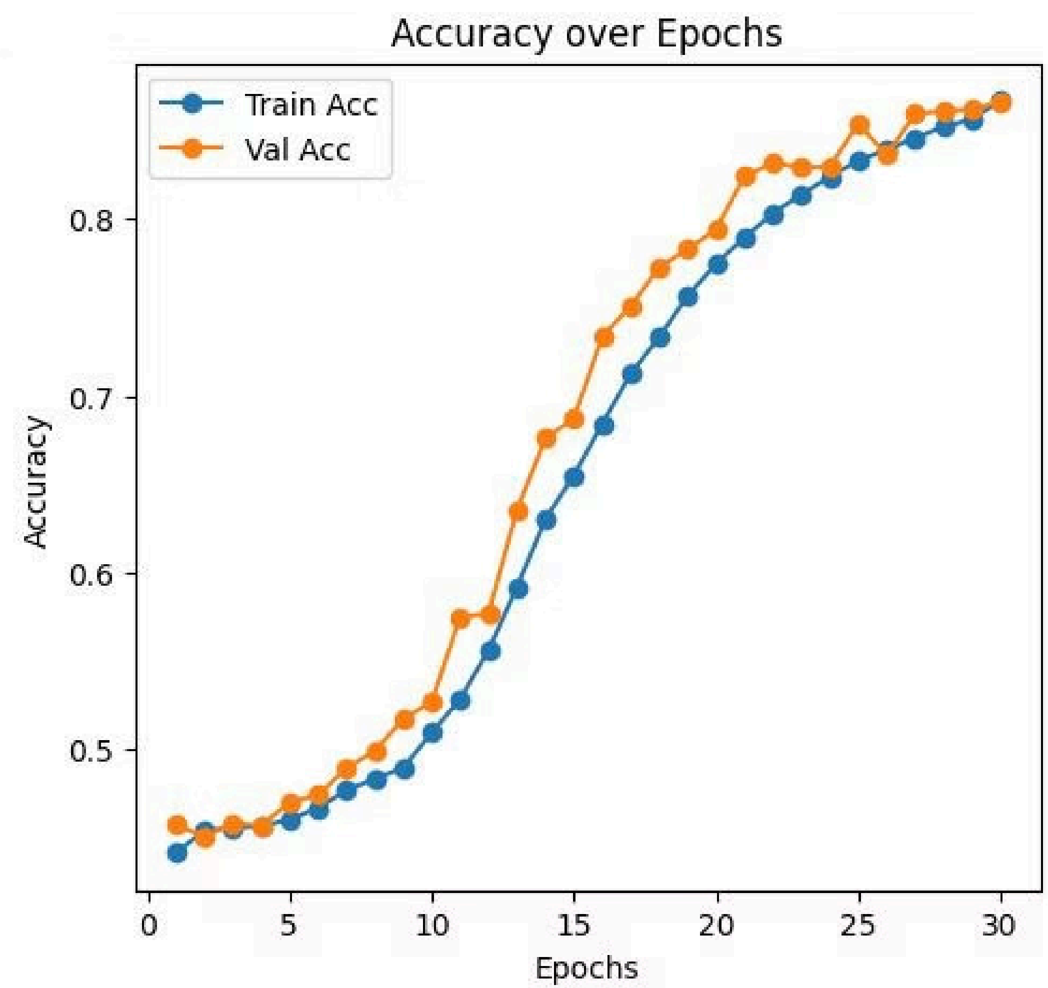
Generated Image

Category : Adventure and Movement

- Characterized by dynamic scenes involving travel, exploration, or physical activity.
- EEG signals fed into the CNN model, which processed spectrograms of the brain activity.
- CNN extracted relevant features, high activity in frequency bands linked to visual processing and motor imagery, suggesting an adventurous and active dream scenario.
- Output from the CNN was then used as input for DALL-E 3, a powerful text-to-image generation model.

Results & Evaluation





Test Accuracy: 87.7895%

Limitations & Future Work

Limitations

Limited EEG resolution, individual variability, data scarcity.

Future Directions

Develop personalized dream decoding models.

Goal

Real-time dream visualization systems.