

**SDM COLLEGE OF ENGINEERING & TECHNOLOGY,  
DHAVALAGIRI, DHARWAD – 580002**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



**Report  
On  
“Metadata-Driven Synthetic EEG Signal Generator for  
Multi-Channel Brainwave Simulatio**

# Table of Contents

Problem Statement.....	3
Introduction .....	3
System Overview .....	5
Metadata Architecture .....	5
Signal Generation Pipeline .....	6
Acknowledgement.....	16

## **Problem Statement**

Create a robust, configurable synthetic EEG dataset generator that produces physiologically plausible multi-channel EEG recordings tied to participant metadata (age, gender, mental state, long/short-term issues, music exposure).

## **Introduction**

Electroencephalography (EEG) is one of the most widely used non-invasive neuroimaging techniques for studying brain activity. By capturing the electrical potentials generated by synchronized neuronal firing, EEG provides valuable insight into cognitive processes, emotional states, neurological conditions, and the influence of external stimuli such as music or stress. EEG signals are typically categorized into distinct frequency bands—delta, theta, alpha, beta, and gamma—which correspond to different functional and physiological states of the brain. For example, alpha activity increases during relaxation, beta is associated with alertness or stress, and theta is elevated during fatigue or emotional imbalance.

Despite its importance, **obtaining large, diverse, and high-quality EEG datasets remains a significant challenge**. Real EEG acquisition requires specialized hardware, controlled laboratory environments, strict participant protocols, and careful handling of ethical and privacy considerations. Moreover, EEG signals are highly subject-specific: age, gender, emotional state, health conditions, and even short-term disturbances such as fatigue or infection can substantially alter brainwave properties. As a result, collecting representative datasets for machine learning model training or algorithm benchmarking is often expensive, time-consuming, and limited in scale.

To overcome these barriers, synthetic EEG generation has emerged as an important complementary strategy. A well-designed synthetic generator can simulate realistic,

physiologically inspired EEG signals without requiring actual human subjects. Crucially, synthetic datasets enable rapid experimentation, help validate analytical pipelines, and support model development before real EEG is available.

The **Synthetic EEG Dataset Generator** developed in this project addresses this need by providing a metadata-driven, multi-channel EEG synthesis framework. Instead of producing generic or random EEG signals, the system incorporates **demographic attributes (age, gender), cognitive state (relaxed, stressed, anxious, depressed, drowsy), long- and short-term medical conditions, and music-based modulation** to shape the characteristics of the generated brainwave data. Frequency-band amplitudes and artifacts are influenced by established neurophysiological research, ensuring that the resulting signals resemble patterns observed in real EEG recordings.

Furthermore, the generator produces multi-band data for each standard EEG channel, injects realistic artifacts such as eye blinks and muscle noise, and applies a post-processing RMS alignment technique to ensure that the final band energies closely match the metadata-driven expectations. The output is a raw CSV dataset containing channel-band waveforms, ready for machine learning, visualization, statistical analysis, or feature extraction.

By offering a configurable, reproducible, and extensible framework, this synthetic EEG generator serves as a valuable tool for researchers, students, and developers working in computational neuroscience, affective computing, biomedical signal processing, and EEG-based emotion or health monitoring systems.

## System Overview

Create a robust, configurable synthetic EEG dataset generator that produces physiologically plausible multi-channel EEG recordings tied to participant metadata (age, gender, mental state, long/short-term issues, music exposure).

**The system is a Streamlit-based EEG signal generator with:**

- Metadata-driven control
- Editable multiplier mappings (mental, medical, demographic)
- Band-wise signal synthesis
- Artifact injection
- Music transient effects
- RMS-based post-scaling for band accuracy
- CSV export with all channel-band signals

**It is designed for:**

- Machine learning dataset creation
- Benchmarking classification pipelines
- Research simulations
- Teaching and demonstration purposes

## Metadata Architecture

Category	Parameters
Demographic	Name, Age, Gender
Cognitive / Emotional	Mental state (e.g., relaxed, stressed, anxious, depressed, happy, drowsy)
Health Profile	Long-term issues, Short-term issues
Music-Related	Listening preference, Genre, Onset time, Duration

Recording Settings	Duration, Sample rate
--------------------	-----------------------

## Signal Generation Pipeline

The generator creates EEG data in five major steps:

### 1 Time-base creation

A time vector is produced according to duration and sample rate:

```
time = 0 → duration (resolution = 1/srate)
```

### 2 Band-specific waveform synthesis

For each channel-band pair:

- 1–3 sinusoids chosen randomly within band frequency range
- Amplitude modulation simulates natural variability
- Low-level Gaussian noise added
- Channel scaling introduces inter-channel variability

### 3 Music transient envelope

If music is enabled:

- Time-local amplitude boosts applied during onset → duration
- Genre-specific bands enhanced (e.g., classical: alpha↑, rock: beta↑/gamma↑)

### 4 Artifact injection

Simulated eye-blanks and muscle artifacts are inserted into channel-level signals:

- Eye blinks → large, slow Gaussian bumps
- Muscle bursts → short, high-frequency noise bursts

Artifacts are then redistributed proportionally into band components.

## 5 RMS alignment (post-processing)

After synthesis:

1. Compute observed RMS per band across channels.
2. Compare with metadata-derived target band distribution.
3. Apply conservative amplitude scaling (0.7–1.4 $\times$ ) to nudge RMS closer to expected values.

This ensures metadata meaningfully influences the generated data.

## Raw CSV Dataset (Final Output)

The final output is a **single CSV file** containing all synthesized EEG samples.

### CSV Structure

The CSV includes:

#### Time column

`time`

#### Channel-band signal columns

Format:

`<Channel>_<Band>`

**Example:**

`Fp1_delta, Fp1_theta, Fp1_alpha, Fp1_beta, Fp1_gamma  
Fp2_delta, Fp2_theta, ...`

**These contain floating-point amplitudes ( $\mu$ V) for each timestamp.**

#### Metadata columns (repeated on each row)

```
patient_name  
age  
gender  
mental_state  
long_term_issues  
short_term_issues  
wants_to_listen  
music_genre  
music_onset  
music_duration
```

### Example header

```
time, Fp1_delta, Fp1_theta, Fp1_alpha, Fp1_beta, Fp1_gamma,  
Fp2_delta, Fp2_theta, ... , F7_theta, F7_alpha, F7_beta,  
...  
patient_name, age, gender, mental_state, long_term_issues,  
short_term_issues, wants_to_listen, music_genre,  
music_onset, music_duration
```

### Units

All voltage samples are in **microvolts ( $\mu$ V)**.

### Row count

**Number of rows = duration  $\times$  sample\_rate.**

## Deterministic Generation & File Naming

### Reproducible seeds

If no seed is provided, a deterministic seed is computed from:

**name + age + gender + mental\_state + music\_genre**

so regenerating with identical metadata yields identical synthetic EEG.

### Filenames

Files follow the pattern:

**synthetic\_eeg\_<Name>\_YYYYMMDD\_HHMMSS.csv**

## **Batch Generation**

A batch metadata CSV can be uploaded to generate multiple EEG datasets at once:

- One output CSV per participant
- All files stored in generated\_batch/
- Optional deterministic seeding (`seed = 1000 + i`) ensures reproducibility

## **Experimental Behavior & Observations**

### **Band distribution patterns**

Metadata strongly influences RMS distribution:

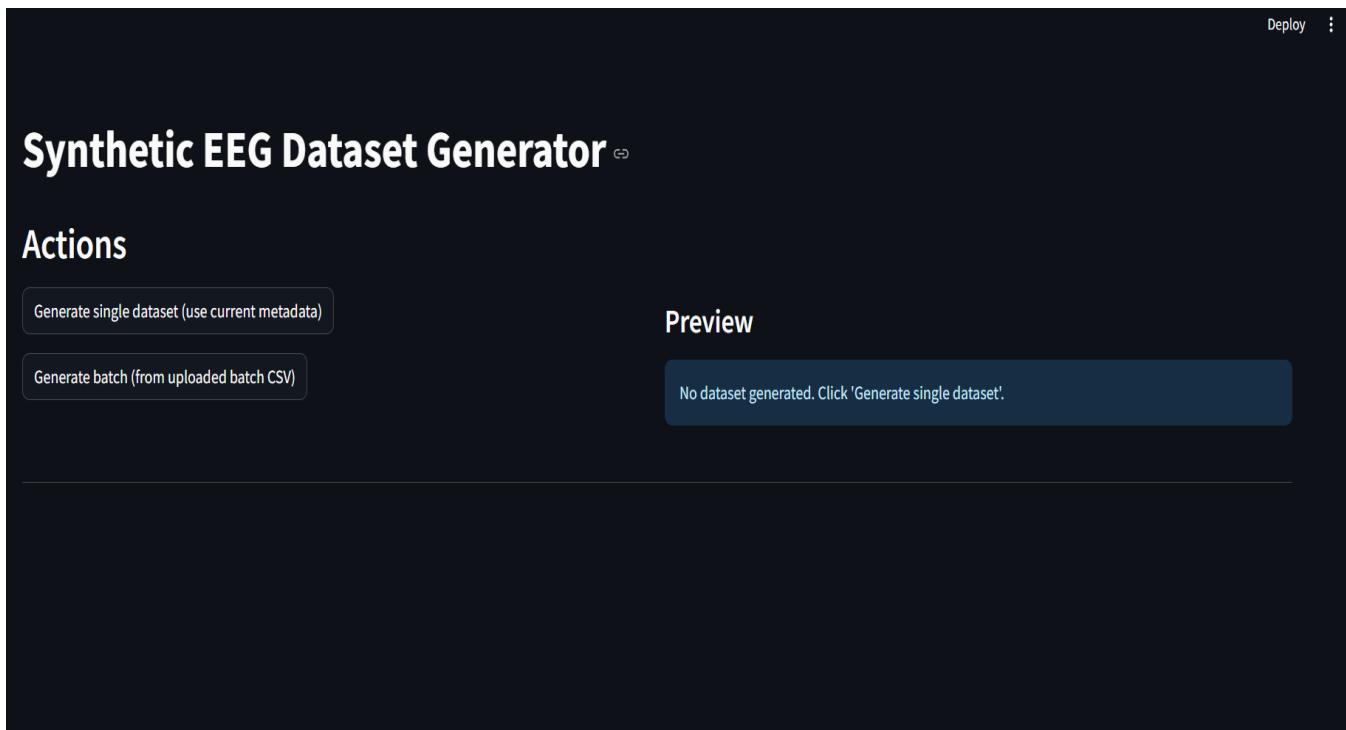
- Relaxed → ↑ alpha
- Stressed → ↑ beta
- Drowsy → ↑ delta/theta
- Music (classical) → ↑ alpha/theta during onset-duration
- Long-term issues (e.g., epilepsy) → broad-band increases

### **Artifact characteristics**

Artifacts appear as:

- Sharp amplitude peaks (eye-blink-like)
- Short high-frequency distortions (muscle bursts)

These resemble real EEG noise sources.



*Figure 1 : Home Interface of the Synthetic EEG Dataset Generator*

<

## Participant metadata (enter values then press Generate)

Name

Age

25 - +

Gender

Male ▼

Mental state

relaxed ▼

Add mental state (type & press Add)

Add mental state

Figure 2 : Participant Metadata Input Section

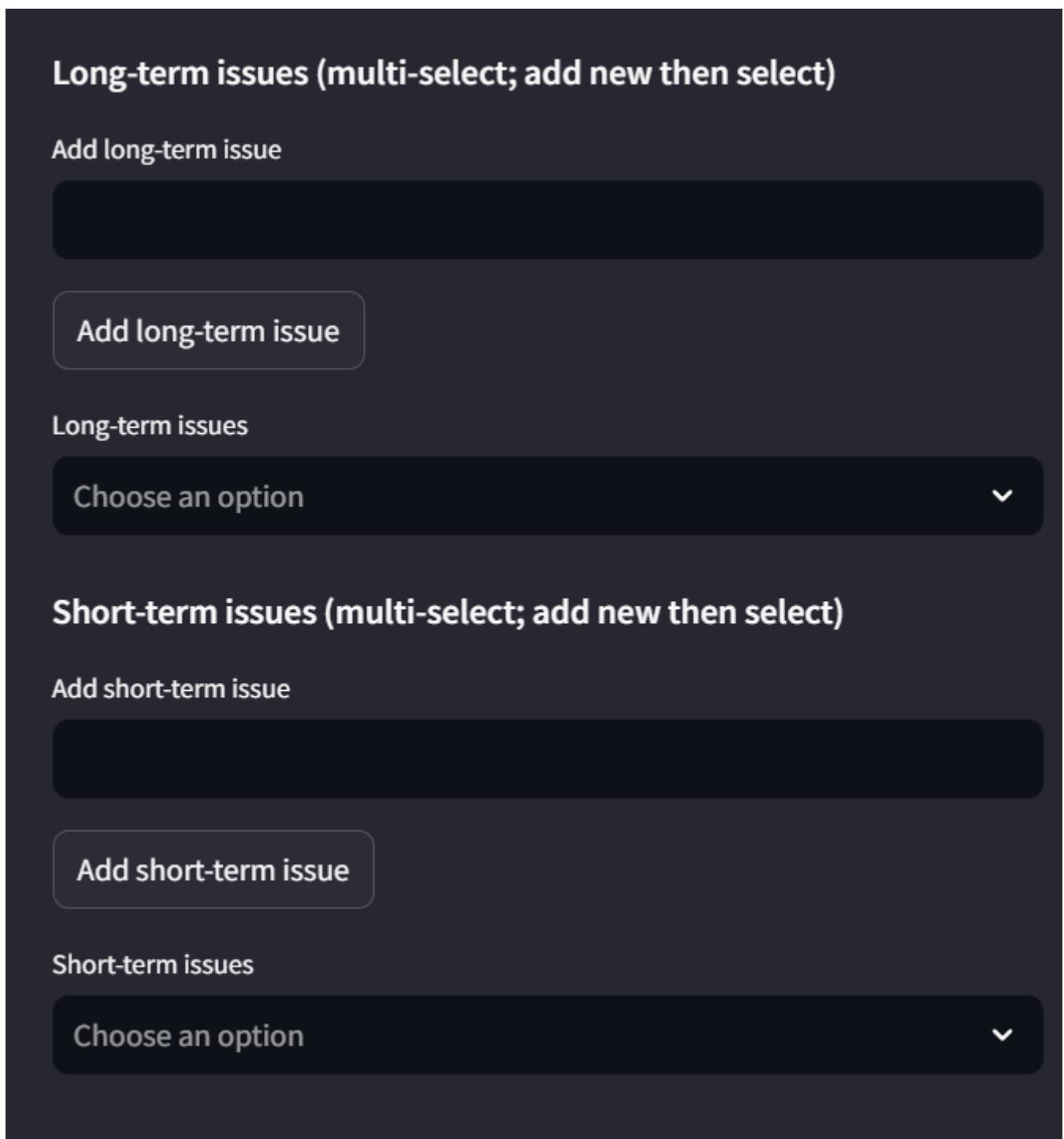


Figure 3 : Long-Term and Short-Term Issue Configuration Panel

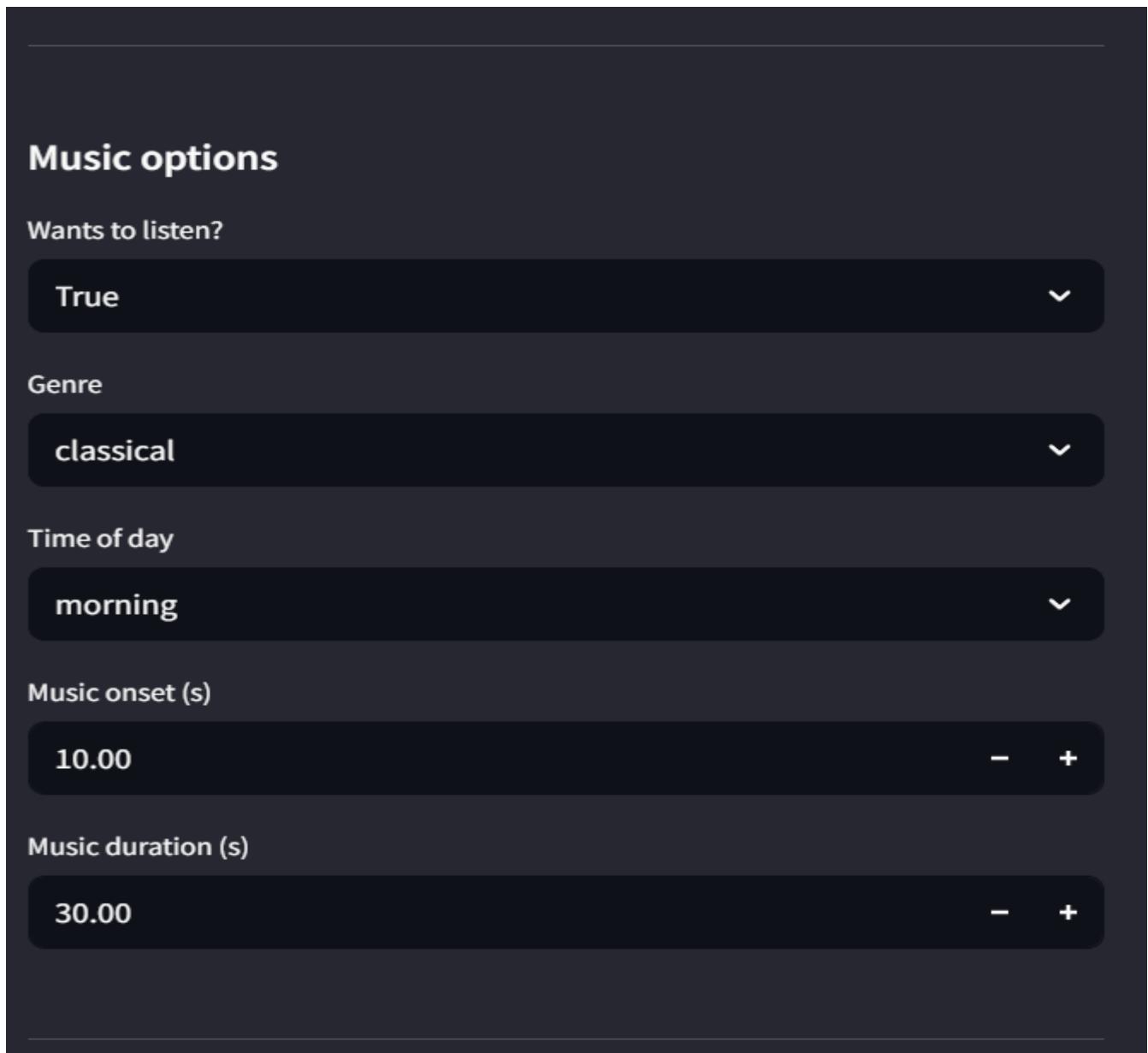


Figure 4 : Music Exposure Configuration Section

## Recording settings

Recording duration (s)

60.00

- +

Sample rate (Hz)

256

▼

Template CSV (optional)



Drag and drop file here

Limit 200MB per file • CSV

Browse files

Batch metadata CSV (optional)



Drag and drop file here

Limit 200MB per file • CSV

Browse files

Create .docx report

Figure 5 : Recording Settings and Template Upload Panel

**Synthetic EEG Dataset Generator**

**Actions**

Generate single dataset (use current metadata)

Generated: synthetic\_eeg\_Darshan\_20251212\_094839.csv

Generate batch (from uploaded batch CSV)

**Preview**

First 6 rows of last generated file:

	time	Fp1_delta	Fp1_theta	Fp1_alpha	Fp1_beta	Fp1_gamma	Fp2_delta	Fp2_theta	Fp2_gamma
0	0	-8.8528	12.3055	9.6394	0.7987	1.9137	-3.333	19.1407	-
1	0.0039	-5.5651	6.8764	5.4803	0.4747	-2.2347	-4.2156	18.7491	-
2	0.0078	-3.2166	3.1908	2.5423	0.182	-1.4017	-3.9383	16.7945	-
3	0.0117	-4.6591	3.9677	3.4871	-0.1304	1.0411	-3.2652	12.5401	-
4	0.0156	-2.5535	1.4297	1.2852	-0.2494	1.4677	-2.6435	9.5335	-
5	0.0195	6.0295	-1.3793	-1.9904	1.0383	-0.4302	-0.9298	2.3182	-

Download CSV

Figure 6 : Dataset Generation Confirmation and CSV Preview

Per-channel RMS (summary):						
	delta	theta	alpha	beta	gamma	
Fp1	12.498	19.061	12.74	4.893	3.516	
Fp2	13.378	9.151	5.225	6.032	1.151	
F3	10.966	15.865	3.857	3.481	2.535	
F4	5.065	9.866	17.279	5.4	3.22	
F7	12.288	14.172	3.633	7.829	1.502	
F8	11.065	16.904	16.338	4.344	2.726	
Fz	9.141	13.308	7.06	6.826	2.03	
C3	6.287	10.547	13.586	6.19	2.704	
C4	8.512	3.368	14.015	7.099	0.634	
Cz	10.323	4.18	6.044	8.813	1.973	

Figure 7 : Per-Channel RMS Summary Table

## **Acknowledgement**

The team wishes to express its sincere gratitude to all the individuals and organizations that contributed to the successful completion of this project, "Analysing Impact of Music on Human Emotions Using EEG Signals "

We are profoundly indebted to the Vision Group on Science and Technology (VGST), an initiative of the Karnataka Science and Technology Promotion Society (KSTePS), Government of Karnataka, for their generous financial support. This work was carried out under the Centres of Innovative Science and Engineering Education (CISEE) scheme and was made possible by the sanction of ₹30,00,000/- (Thirty Lakhs Only).

We also acknowledge the continuous support, necessary facilities, and conducive environment provided by S.D.M. College of Engineering & Technology, Dharwad, and specifically the Department of Computer Science and Engineering, which were crucial to realizing the project's objectives.

Our deepest thanks go to the Principal Investigator, Dr. U. P. Kulkarni, and Co-Principal Investigator, Dr. Vidyagouri Kulkarni, for their invaluable guidance and leadership throughout this research.

Finally, we acknowledge the collective efforts of our project team members Preetika Panalkar, Rohit Halappanavar, Rohit Badi, Sumpreetsing Rajaput—— whose active participation and commitment contributed significantly to the completion of this work.