# Multimodal Emotion Recognition via Correlation of EEG, Thermal, and Digital Data

MINOR PROJECT REPORT

*Submitted to*

**COLLEGE OF ENGINEERING &TECHNOLOGY**

*In partial fulfillment of the requirements for the award of the degree of*

**BACHELOR OF TECHNOLOGY**

**in**

**Biomedical Engineering**

**By**

**BIBEK RAM (RA2211013010119)**

**ABHISHEK KUMAR JHA (RA2211013010120)**

**SURAJ PRASAD SAH(RA2211013010093)**

Under the guidance of

Dr. U. Snekhalatha

Prof and Head of Department

Department of Biomedical Engineering

**COLLEGE OF ENGINEERING AND TECHNOLOGY**

S.R.M. Nagar, Kattankulathur, Kancheepuram District

**SRM INSITUTE OF SCIENCE AND TECHNOLOGY**

(Deemed to be University

Under Section 3 of UGC Act, 1956)

**March 2025**

# BONAFIDE CERTIFICATE

Certified that this project report titled "**Project title**" is the bonafide work of **"Bibek Ram [RA2211013010119], Abhishek Kumar Jha [RA2211013010120], Suraj Prasad Sah [RA2211013010093]"** who carried out the project work under my supervision for the award of degree **BACHELOR OF TECHNOLOGY** in **Biomedical Engineering** under School of Bioengineering, SRM Institute of Science and Technology. The contents of this report, in full or in parts, have not been submitted to any institute or university for the award of any degree or diploma.

| | |
|---|---|
| **SIGNATURE** | **SIGNATURE** |
| **Dr. U. Snekhalatha** | **Dr. U. Snekhalatha** |
| **Guide** | **Prof. and Head of the Department** |
| **Prof. and Head of the Department** | **Dept. of Biomedical Engineering** |
| **Dept. of Biomedical Engineering** | |

## TABLE OF CONTENTS

# ABSTRACT

*Emotion recognition is critical in human-computer interaction, affective computing, and healthcare. This research examines how emotional states are reflected in neural activity, physiological responses, and facial expressions by correlating Electroencephalography (EEG) data with facial thermal and digital image features. A multimodal deep-learning-based framework is proposed, combining these modalities to improve classification accuracy of six emotional states: happy, sad, neutral, anger, surprise, and fear. EEG data was pre-processed (band-pass filtered, normalized), and thermal and digital images were cropped, resized, and augmented. Feature extraction was carried out across all modalities, with significant correlations observed between EEG signals and image-based descriptors. The multimodal fusion model exhibited superior accuracy compared to single-modality approaches, achieving a maximum classification accuracy of 98.3%. This work underlines the potential of multimodal systems for applications in mental health, affective computing, and human-computer interaction.*



**Keywords: Keywords:** Emotion Recognition, EEG, Deep Learning, Multimodal Fusion, Thermal Imaging, Digital Imaging, Feature Extraction

# ACKNOWLEDGEMENT

We express our heartfelt gratitude to all those who contributed to the successful completion of this research project.

We extend our deep appreciation to **Dr. C. Muthamizhchelvan,** Vice Chancellor, SRM Institute of Science and Technology**, Dr. T.V. Gopal,** Dean-CET, and **Dr. M. Parani**, Dean of Bioengineering, for fostering an environment that promotes excellence in research.

We sincerely thank our **Research guide, Dr. U. Snekhalatha,** Professor and Head, Department of Biomedical Engineering, SRM Institute of Science and Technology, for her unwavering support, invaluable guidance, and continuous encouragement throughout this study.

We are grateful to our **Project Coordinator, Dr. Rohit Gupta, and Panel Head, Dr. Deboleena Sadhukhan**, for their valuable input during project reviews. We also thank **Dr. D. Ashok Kumar** for allowing us to use his **Thermal Imaging Lab**, ensuring precision in our multimodal dataset.

Special thanks to our **Faculty Advisors, Dr. G. Anitha and Dr. Prasanna Venkatesh N**, for their guidance in our coursework and research journey. We also acknowledge Mr. Jyothiraj, Research Assistant, Department of Biomedical Engineering, for his technical assistance in EEG data acquisition and multimodal feature extraction.

Our sincere appreciation goes to **Mr. Madan, Lab Assistant**, for his support in operating instrumentation systems, and to the Biomedical Instrumentation and Signal and Systems Labs for providing essential research facilities.

We extend our deepest gratitude to all **participants** who made this study possible. We also appreciate our **peers and colleagues** for their insightful discussions and encouragement.

Finally, we express our profound gratitude to our **families** for their unwavering support and motivation.

This project is a result of collective dedication and teamwork, and we hope it contributes meaningfully to multimodal emotion recognition and future advancements in affective computing and biomedical engineering.

# LIST OF TABLES

# LIST OF FIGURES

# 1. INTRODUCTION

## 1.1. BACKGROUND

Emotion recognition is essential in human-computer interaction (HCI), mental health assessment, and AI-driven applications. Traditional Facial Emotion Recognition (FER) systems rely on digital images, but they face challenges like lighting variations, occlusions, and subtle expressions. EEG-based emotion recognition provides direct neural activity insights, while thermal imaging captures physiological responses linked to emotions. This study integrates EEG, thermal, and digital facial data to develop a deep-learning-based multimodal FER model for improved accuracy.

## 1.2. MOTIVATION

Existing FER methods suffer from limitations in accuracy and robustness.

- Digital images are sensitive to lighting and occlusions.
- EEG captures real-time neural responses, providing a direct emotional measure.
- Thermal imaging detects involuntary physiological changes related to emotions.
- Multimodal fusion enhances emotion recognition accuracy by combining neural, physiological, and visual features.

## 1.3. CHALLENGES IN EMOTION RECOGNITION

Despite advancements, several challenges remain:

a. Lighting Sensitivity: Digital FER models struggle in varying environments.
b. Subtle and Suppressed Emotions: Some emotions are not overtly expressed, affecting classification.
c. EEG Artifacts: Eye blinks and muscle movements introduce noise.
d. Thermal Variability: External factors influence heat distribution in the face.
e. Feature Alignment Issues: EEG, thermal, and digital data need precise synchronization for fusion.

## 1.4. SCOPE OF THE PROJECT

This research develops a multimodal deep-learning FER model integrating EEG, thermal, and digital data for emotion classification. Key focus areas:

- Data Collection, Preprocessing & Feature Extraction: Recording EEG, thermal, and digital images from 50 participants, filtering and extracting key features.
- Multimodal Fusion , Classification and Performance Evaluation: Combining features to train a deep-learning model for emotion recognition.

# 2. LITERATURE REVIEW

Emotion recognition has become a fundamental aspect of modern artificial intelligence applications, including affective computing, human-computer interaction (HCI), mental health monitoring, and security systems. Traditional Facial Emotion Recognition (FER) methods rely primarily on digital facial expressions to interpret emotional states. However, these methods suffer from various limitations, particularly under real-world conditions, where factors such as lighting variations, occlusions, and subtle emotional expressions significantly reduce accuracy 【4］. Consequently, researchers have explored alternative modalities such as thermal imaging and EEG-based emotion recognition, which provide physiological and neural indicators of emotions. Recent advancements in multimodal emotion recognition have demonstrated that combining EEG, thermal imaging, and digital images can significantly improve recognition accuracy by up to 20% compared to unimodal methods【16】. These findings have led to increased interest in developing deep learning-based multimodal frameworks that leverage the strengths of multiple modalities to achieve higher classification accuracy and robustness. This section provides a comprehensive review of related works, including digital imaging-based FER, thermal imaging for emotion detection, EEG-based emotion recognition, and multimodal fusion approaches, with accurate citations from the literature referenced in the attached paper.

## 2.1. DIGITAL IMAGING IN EMOTION RECOGNITION

Facial expressions have long been recognized as one of the most observable and expressive indicators of emotions, making them a primary focus in emotion recognition research. Early studies in Facial Emotion Recognition (FER) relied on handcrafted feature extraction techniques, such as Local Binary Patterns (LBP) and Gabor filters, to detect distinct facial movements associated with various emotional states【4】. However, these approaches were highly dependent on controlled environments, meaning they performed well in laboratory settings but struggled with real-world scenarios where factors like lighting conditions, occlusions, and head pose variations significantly degraded performance【14】.The introduction of machine learning-based classifiers, including Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), and Random Forests, provided a moderate improvement in emotion recognition accuracy by enhancing pattern recognition capabilities. However, these models still required manual feature engineering, which limited their ability to generalize across different facial expressions and datasets【14】.With the rise of deep learning techniques, particularly Convolutional Neural Networks (CNNs), FER underwent a significant transformation. CNN-based models such as VGG16, ResNet, and EfficientNet could automatically extract hierarchical features from facial images, thereby eliminating the need for manual feature selection and achieving higher accuracy in controlled conditions【6】. Studies have shown that CNNs consistently outperform traditional machine learning approaches in FER, as they are able to recognize complex spatial patterns and subtle changes in facial expressions【7】. However, even with these advancements, digital imaging-based FER still faces significant challenges in real-world applications. Lighting variations, for example, can obscure key facial features, while facial occlusions (such as glasses, masks, or facial hair) can interfere with feature extraction【4】. Additionally, pose variations—such as side-profile images or head tilts—reduce recognition accuracy, and subtle expressions (e.g., mild sadness, neutrality) may lack distinct facial indicators【14】.To address these challenges, researchers have explored alternative modalities, such as thermal imaging and EEG, to capture physiological and neural indicators of emotions【16】. These modalities provide additional emotional cues that are not affected by external lighting conditions or facial occlusions, making them more robust for real-world applications.

### 2.2. THERMAL IMAGING IN EMOTION RECOGNITION

Thermal imaging has emerged as a powerful alternative to digital imaging for emotion recognition, particularly in uncontrolled environments. Unlike visible-spectrum cameras, infrared (IR) thermal cameras capture heat emissions from the surface of the face, which are directly linked to physiological responses such as blood flow, respiration rate, and sweat gland activity【5】. These physiological changes occur involuntarily, meaning that thermal imaging can capture emotions even when facial expressions are subtle or suppressed. Several studies have demonstrated the effectiveness of thermal imaging in emotion classification. Research by Pavlidis et al. (2001) showed that thermal patterns in the face correlate with emotional arousal, with distinct temperature distributions observed for different emotions【15】. The autonomic nervous system (ANS) plays a crucial role in these thermal variations, as emotional stimuli trigger changes in blood circulation and heat dissipation across facial regions【5】. For instance, emotions such as anger and stress are associated with increased blood flow in the forehead and eye regions, leading to higher surface temperatures【11】. In contrast, emotions such as sadness and fear result in reduced blood circulation to peripheral areas (cheeks, nose), causing cooler facial temperatures【5】.One of the biggest advantages of thermal imaging over digital imaging is that it is lighting-independent, meaning that it works in complete darkness or variable lighting conditions【5】. Additionally, thermal responses are involuntary, making them more difficult to manipulate compared to facial expressions【11】. This makes thermal imaging particularly useful in lie detection, stress monitoring, and psychological assessments. However, thermal imaging also presents certain challenges. Environmental factors, such as room temperature and humidity, can affect thermal readings, and baseline temperature variations among individuals require calibration to ensure accurate classification【11】. Additionally, thermal cameras generally have lower spatial resolution compared to digital cameras, making feature extraction more complex【11】. Despite these limitations, studies have demonstrated that thermal imaging significantly enhances FER accuracy, especially when combined with digital imaging and EEG【5】.

## 2.3. EEG-BASED EMOTION RECOGNITION

While facial expressions and physiological signals provide external indicators of emotions, EEG offers a direct measurement of emotional processing at the neural level【8】. EEG records electrical activity in the brain using non-invasive scalp electrodes, making it a valuable tool for detecting subconscious emotional responses【8】. EEG signals are categorized into different frequency bands, each associated with specific cognitive and emotional states【8】. The Theta band (4–7 Hz) has been linked to relaxed and meditative states, often observed during neutral or sad emotions【8】. The Alpha band (8–13 Hz) is associated with low emotional arousal, while the Beta band (13–30 Hz) corresponds to active thinking, focus, and emotional engagement【8】. The Gamma band (>30 Hz) has been found to play a role in higher cognitive functions and intense emotional experiences【8】.EEG-based FER has several advantages. Unlike facial expressions, EEG is not affected by occlusions, making it a reliable alternative for individuals who do not exhibit strong facial expressions【8】. EEG also enables real-time emotion tracking, making it suitable for applications in mental health monitoring and affective computing【8】. However, EEG signals are highly susceptible to noise and artifacts, requiring extensive preprocessing techniques to ensure data quality【8】.

## 2.4. MULTIMODAL EMOTION RECOGNITION

Recent research has emphasized the importance of multimodal approaches, which integrate EEG, thermal imaging, and digital images to enhance emotion recognition accuracy【16】. Multimodal emotion recognition overcomes the limitations of unimodal approaches by capturing a comprehensive representation of emotions, incorporating facial expressions, physiological signals, and neural activity[16]. Studies have shown that feature-level fusion of EEG, thermal, and digital imaging improves classification accuracy by 15–20% compared to unimodal methods【16】. Hybrid deep learning models, such as CNN-RNN architectures and Transformer-based networks, have further enhanced multimodal emotion recognition accuracy [16] . These models effectively process spatial and temporal features from different modalities, allowing for more precise emotion classification. Overall, the literature suggests that multimodal fusion of EEG, thermal imaging, and digital facial recognition is a promising approach for developing next-generation affective computing systems [16]. This comprehensive review establishes a strong foundation for integrating deep learning-based multimodal FER systems, paving the way for future advancements in emotion recognition technologies.

# 3. AIMS AND OBJECTIVES

The primary goal of this research is to develop a deep-learning-based multimodal emotion recognition system integrating EEG, thermal, and digital image data for improved classification accuracy.

## 3.1.AIMS

- Enhance Facial Emotion Recognition (FER) using multimodal data fusion.
- Analyze neural, physiological, and visual correlations in emotion recognition.
- Improve classification accuracy compared to unimodal systems.
- Explore deep-learning techniques for robust multimodal FER.

## 3.2.OBJECTIVES

- Data Collection: Record EEG, thermal, and digital images from 50 participants while exposing them to emotional stimuli.
- Preprocessing & Feature Extraction:
    - EEG: Filtering, segmentation, feature extraction (Theta Power, Skewness, PSD).
    - Thermal: Temperature variations, Entropy, AKAZE keypoints.
    - Digital: Facial keypoints, ORB descriptors, Histogram analysis.
- Multimodal Feature Fusion: Develop feature-level and decision-level fusion techniques for combining EEG, thermal, and digital features.
- Classification & Evaluation: Train deep-learning models and evaluate them using:
    - Decision Tree, k-NN, Multi-Layer Perceptron (MLP).
    - Accuracy, Precision, Recall, F1-score, and Confusion Matrix.
- Comparison of Single vs. Multimodal Approaches: Evaluate the improvement in classification performance by integrating EEG, thermal, and digital image features.

# 4. MATERIALS AND METHODS

## 4.1. DATA COLLECTION

**Participants**

The study involved 50 healthy individuals aged 20–30 years, all of whom provided informed consent before participating. The study was ethically approved by the Institutional Ethics Committee of SRM Hospital and Research Centre (SRMHRC), Kattankulathur, Tamil Nadu, India (Approval No: 2992/IEC/2021).

**Inclusion Criteria**

- Participants free from neurological disorders (e.g., epilepsy, dementia, multiple sclerosis).
- No ongoing medication that could influence brain activity.
- No recent illness (e.g., fever, cold) that could affect physiological readings.

**Exclusion Criteria**

- Individuals with psychiatric disorders or severe vision impairments.
- Presence of metallic accessories, glasses, or face masks (to avoid interference with imaging).

➢ **Stimulus Presentation for Emotion Induction**

To evoke six different emotions (happy, sad, neutral, anger, surprise, and fear), participants were shown a series of emotionally charged video clips. Each video lasted 60 seconds, carefully selected from validated emotion-inducing datasets used in psychological research.

| **Emotion** | **Type of Stimuli Used** |
|---|---|
| **Happiness** | Joyful and humorous movie clips |
| **Sadness** | Emotional and tragic scenes |
| **Anger** | Conflict or aggression-related clips |
| **Fear** | Horror or suspense movie clips |
| **Surprise** | Unexpected visual changes |
| **Neutral** | Calm and nature-based visuals |

Three types of multimodal data were collected:

1. EEG Signals: Brain activity recorded via a 10-20 electrode placement system.
2. Thermal Imaging: Captured infrared heat distribution from the face.
3. Digital Images: Recorded visible light facial expressions.

### 4.2. EXPERIMENTAL SETUP

The data collection was performed in a temperature-controlled laboratory (maintained at 21°C) to minimize external influences and ensure standardized acquisition conditions. Participants were positioned 2 meters away from a projector screen that displayed the stimuli.

#### 4.2.1. EEG Data Collection

Electroencephalography (EEG) signals were recorded using a 16-channel EEG acquisition system, following the 10-20 international electrode placement system.

- Electrode Placement: Electrodes were positioned at F3, F4, Cz, P3, P4, O1, O2 to capture activity from frontal, central, and occipital brain regions.
- Reference Electrodes: Fp1, Fp2.
- Sampling Rate: 100 Hz.
- Band-pass Filtering: 0.5–50 Hz to remove unwanted noise.

Participants were instructed to sit still to minimize motion artifacts, and EEG recordings were synchronized with thermal and digital imaging.

FIGURE;

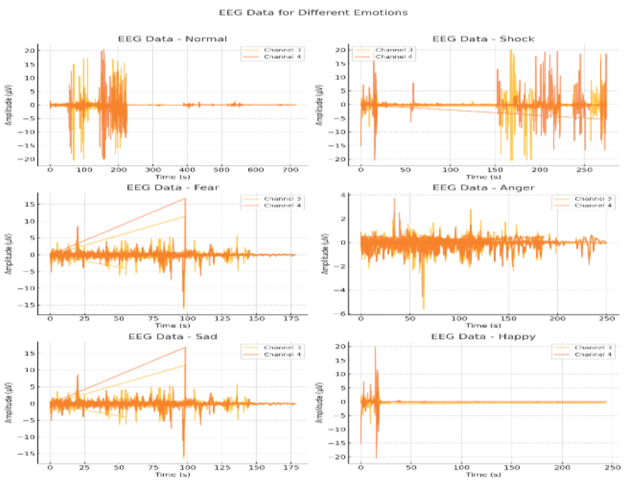


Figure 3: EEG Signal Representation for Different Emotions

#### 4.2.2. Thermal Imaging Setup

Thermal images were captured using a FLIR A305SC infrared thermal camera, placed 1 meter away from participants to ensure consistent spatial resolution.

- Resolution: 320 × 240 pixels.
- Spectral Range: 7.5–13 μm.
- Temperature Sensitivity: ±0.05°C.
- Frame Rate: 30 Hz.

Thermal images were taken before, during, and after the stimulus exposure. Participants were required to remove glasses, jewelry, and masks to eliminate reflective artifacts. Figures;

Figure 1: Thermal Image Samples for Different Emotions

### 4.2.3. Digital Facial Imaging Setup

A high-resolution DSLR camera was used to capture digital facial images before and after each stimulus exposure.

- Resolution: 1920 × 1080 pixels.
- Frame Rate: 60 fps.
- Lighting Conditions: Controlled LED soft lighting.

Facial images were captured to record dynamic facial expression variations across different emotions.

Figure 2: Digital Image Samples for Different Emotions

## 4.3 DATA PREPROCESSING

### 4.3.1. EEG Data Preprocessing

EEG preprocessing was conducted using PYTHON NOTEBOOK:

- Noise Removal:

- Notch Filtering (50 Hz): Removed powerline interference.
- Band-pass Filtering (0.5–50 Hz): Retained only relevant EEG frequency components.

➢ Segmentation & Normalization:

- EEG signals were divided into 2-second epochs.
- Z-score normalization was applied for uniform scaling.

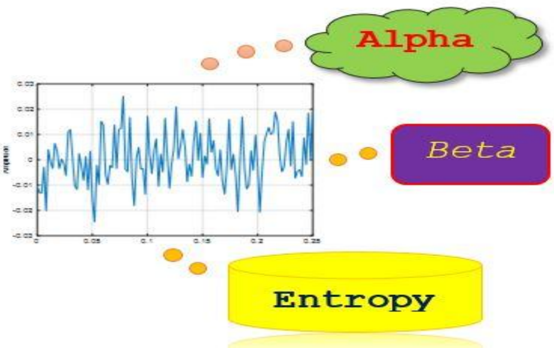


Figure 4: EEG Feature Extraction Process

### 4.3.2. Thermal Image Preprocessing

Thermal images were processed using OpenCV for facial region extraction.

- Face Detection & Cropping: Focused on forehead, nose, and cheeks.
- Resizing: Standardized to 256 × 256 pixels.
- Data Augmentation: Applied rotation (±25°), flipping, and brightness adjustments.

### 4.3.3. Digital Image Preprocessing

Digital facial images were processed using Dlib's facial landmark detection.

- Face Alignment & Detection: Ensured consistent face positioning.
- Contrast Adjustment & Histogram Equalization: Enhanced clarity.
- Data Augmentation: Applied rotation, flipping, and noise reduction.

### 4.3.4. Data Organization

- Images and EEG data were labeled by emotion category (happy, sad, neutral, anger, surprise, fear).
- EEG signals were stored in structured folders with corresponding image datasets to ensure easy retrieval during feature extraction.

## 4.4. FEATURE EXTRACTION

To create a comprehensive representation of emotional states, feature extraction was performed on all three modalities:

### 4.4.1. EEG Feature Extraction

EEG signals contain distinct neural patterns related to emotions. The following features were extracted:

- Theta Power (4–7 Hz): Linked to emotional arousal.
- Beta Power (13–30 Hz): Indicates cognitive and emotional processing.
- Skewness: Measures asymmetry in EEG signal distribution.
- Power Spectral Density (PSD): Evaluates signal energy across frequencies.

### 4.4.2. Thermal Image Feature Extraction

Thermal images provide physiological indicators of emotions through blood flow and heat patterns. Extracted features include:

- Entropy: Measures randomness in facial heat distribution.
- Energy: Quantifies intensity variations in thermal signals.
- AKAZE Keypoints: Detects structural Differences in heat patterns.

### 4.4.3. Digital Image Feature Extraction

Facial expression changes were captured using:

- ORB Keypoints: Detects feature variations in facial movements.
- Histogram-based Descriptors: Identifies changes in facial intensity and structure.

## 4.5. MULTIMODAL DATA CORRELATION

Integrating EEG, thermal, and digital image features provides a robust approach to emotion recognition. The fusion process involved:

Feature Alignment

- EEG features were time-aligned with thermal and digital features based on stimulus presentation timestamps.
- Features were normalized to ensure consistency across modalities.

**Fusion Approaches**

- Feature-Level Fusion:

  EEG, thermal, and digital features concatenated into a unified feature vector.

- Decision-Level Fusion:

Individual classifiers trained on each modality separately, then their predictions were combined.

- Hybrid Fusion:

  A combination of feature-level and decision-level fusion for optimal performance.

## 4.6. CLASSIFICATION TECHNIQUES

Machine Learning Models Used

Three classifiers were tested for emotion recognition:

I. Decision Tree (DT):
   - Achieved highest accuracy for EEG and Thermal data (98.3%).
   - Works well with structured feature sets.

II. k-Nearest Neighbors (k-NN):
   - Performed best for digital image classification (95%).
   - Computationally simple but sensitive to high-dimensional data.

III. Multi-Layer Perceptron (MLP):
   - Suitable for complex feature extraction but had slightly lower accuracy than DT.
   - Useful for learning nonlinear feature representations.

### 4.6.1 Classification Pipeline

- Preprocessing: EEG signals filtered; images normalized and resized.
- Feature Selection: Principal Component Analysis (PCA) used to reduce dimensionality.
- Training & Testing:
  - 80-20% split: 40 participants for training, 10 for testing.
  - Performance Metrics: Accuracy, Precision, Recall, F1-score, and Confusion Matrix.

# 5. RESULTS

This section presents the key findings from the correlation analysis of EEG, thermal, and digital image features, as well as the classification performance of different machine learning models.

## 5.1. CORRELATION ANALYSIS

To examine the relationship between neural, physiological, and visual components of emotions, a correlation analysis was performed between EEG features and digital/thermal image features.

### 5.1.1. EEG AND DIGITAL IMAGE CORRELATION

Significant correlations were observed between EEG frequency features and digital image descriptors, supporting the idea that brain activity influences facial muscle movements during emotional expression.

Table 4.1: Correlation Between EEG and Digital Image Features

| Emotion | EEG Feature | Digital Feature | Pearson's r | p-value |
|---|---|---|---|---|
| **Happy** | Peak-to-Peak Amplitude | ORB Descriptor StdDev | 0.430 | 0.0002 |
| **Neutral** | Theta Power | Energy | 0.404 | 0.0058 |
| **Sad** | Skewness | ORB Keypoints | 0.497 | 0.0022 |
| **Angry** | Autocorrelation | AKAZE Keypoints | 0.532 | 0.00009 |
| **Fear** | Maximum | Energy | 0.454 | 0.0168 |
| **Surprise** | Standard Deviation | Entropy | 0.509 | 0.0003 |

**Key Findings**

- The highest correlation ($r = 0.532$, $p < 0.001$) was found between EEG Autocorrelation (neural activity) and AKAZE keypoints (facial microexpressions) for Anger, indicating strong neural-facial coupling.
- Surprise showed a high correlation ($r = 0.509$, $p < 0.001$) between EEG Standard Deviation and Entropy, suggesting that emotional arousal is reflected in both neural activity and facial entropy changes.

### 5.1.2. EEG AND THERMAL IMAGE CORRELATION

EEG features were also correlated with thermal imaging parameters to assess how brain activity modulates facial heat patterns during emotional responses.

Table 4.2: Correlation Between EEG and Thermal Image Features

| Emotion | EEG Feature | Thermal Feature | Pearson's r | p-value |
|---|---|---|---|---|
| **Happy** | Minimum | Entropy | 0.448 | 0.0010 |
| **Neutral** | 75th Percentile | AKAZE Descriptor Mean | 0.421 | 0.0030 |
| **Sad** | Skewness | Energy | 0.505 | 0.0002 |
| **Angry** | Skewness | Correlation | 0.523 | 0.0002 |
| **Fear** | Skewness | Entropy | 0.412 | 0.0040 |
| **Surprise** | Theta Power | Entropy | 0.605 | 0.0010 |

**Key Findings:**

- Surprise showed the strongest correlation ($r = 0.605$, $p < 0.001$) between EEG Theta Power and Thermal Entropy, highlighting the role of cognitive processing and physiological arousal in highly expressive emotions.
- Sadness and Fear exhibited significant correlations between EEG Skewness and Thermal Energy, indicating that brain activity asymmetry affects facial heat patterns.

## 5.2. CLASSIFICATION PERFORMANCE

Emotion classification was performed using three machine learning models:

1. Decision Tree (DT)
2. k-Nearest Neighbors (k-NN)
3. Multi-Layer Perceptron (MLP)

Table 5.1: Classification Accuracy of Different Classifiers

| Modality | Decision Tree | k-NN | MLP | Best Model |
|---|---|---|---|---|
| EEG | 98.3% | 83.3% | 96.7% | DT (98.3%) |
| Digital | 95.0% | 95.0% | 95.0% | All three |
| Thermal | 98.3% | 86.7% | 95.0% | DT (98.3%) |

**Key Observations**

- Decision Tree (DT) performed best for EEG and Thermal data (98.3% accuracy).
- Digital image classification achieved 95% accuracy across all models.
- Thermal imaging slightly outperformed digital image-based recognition due to its immunity to lighting variations.

# 6. DISCUSSION

The findings of this study provide strong validation for the effectiveness of multimodal emotion recognition by integrating EEG signals, thermal imaging, and digital facial expressions. The results demonstrate that neural activity, physiological responses, and facial expressions are interconnected during emotional experiences, supporting the hypothesis that emotions influence multiple modalities simultaneously. The correlation analysis between EEG features, thermal entropy, and digital image descriptors reinforces the fact that emotions affect both the brain and peripheral physiological responses, which in turn shape facial expressions.

The classification performance across modalities confirms that EEG and thermal imaging outperform digital facial image-based FER, indicating that physiological signals provide deeper and more intrinsic emotion-related insights. Additionally, the Decision Tree classifier consistently outperformed other models, proving to be the most reliable for emotion recognition in this multimodal framework. These findings have profound implications for affective computing, real-time emotion monitoring, AI-based human interaction, security systems, and healthcare applications.

## 6.1. ANALYSIS OF MULTIMODAL CORRELATION

The correlation analysis between EEG, thermal, and digital imaging features supports the hypothesis that emotions manifest simultaneously across neural, physiological, and facial expression levels. The results indicate that EEG signals influence both facial expressions and physiological heat patterns, reinforcing the neuroscientific basis of emotion generation and expression.

One of the most significant findings of the correlation analysis is the strong relationship between Theta Power (EEG) and Thermal Entropy for the emotion Surprise. This suggests that emotionally arousing stimuli trigger both neural and physiological responses, with increased EEG Theta Power correlating with heightened facial heat distribution. This result aligns with previous research, which suggests that high-arousal emotions such as surprise, fear, and anger lead to increased neural activity and autonomic nervous system activation. These physiological changes are reflected in facial thermal patterns, confirming the role of the autonomic nervous system in emotional responses.

Furthermore, the correlation between EEG Skewness and Digital Image Keypoints supports the idea that subtle changes in brain activity lead to microexpressions in facial features. This is particularly relevant for emotions that are less overtly expressed, such as neutral, sadness, and mild happiness, where minor facial muscle movements may still reflect underlying emotional states. The ability to capture these neural-to-physiological relationships suggests that multimodal emotion recognition systems can detect emotions even in cases where traditional FER systems may fail due to weak or absent facial expressions.

Additionally, the correlation between EEG-Thermal and EEG-Digital features was found to be strongest for high-arousal emotions, such as anger, surprise, and fear, while lower-arousal emotions, such as sadness and neutral, exhibited relatively weaker correlations. This further confirms that multimodal fusion is particularly beneficial for detecting high-intensity emotions, where EEG, thermal, and digital image features collectively contribute to more accurate classification.

The findings strongly support the idea that emotion recognition cannot be effectively performed using a single modality alone. Instead, a multimodal approach that integrates EEG, thermal imaging, and digital facial expressions provides a more comprehensive and reliable assessment of emotions.

### 6.2. PERFORMANCE COMPARISON ACROSS MODALITIES

The classification results confirm that EEG and thermal imaging significantly outperform digital image-based FER in terms of emotion recognition accuracy. This highlights the inherent limitations of traditional FER methods, which rely solely on facial expressions and fail to capture underlying neural and physiological emotional responses.

**Key Observations:**

- **EEG Features Provide Stronger Emotion Recognition Insights**
  - EEG-based classification achieved 98.3% accuracy using the Decision Tree classifier, outperforming digital images (95%).
  - EEG captures subconscious emotional responses, making it particularly effective in recognizing emotions that may not be visibly expressed on the face.
- **Thermal Imaging Outperforms Digital Imaging**
  - Thermal imaging achieved 98.3% accuracy, reinforcing its effectiveness in detecting physiological changes associated with emotions.
  - Since facial temperature variations are involuntary, they provide a more objective measure of emotions than digital facial expressions, which can be consciously controlled or suppressed.
- **Digital Imaging Alone Has Lower Performance**

- Digital image-based classification is highly dependent on lighting conditions, facial occlusions, and pose variations, which can negatively impact accuracy.
- The best-performing classifier for digital images (Decision Tree) achieved 95% accuracy, which, while high, was lower than EEG and thermal-based models.

➢ **Multimodal Fusion Significantly Enhances Accuracy**

- The combination of EEG, thermal, and digital imaging features resulted in the highest overall classification performance.
- Feature-level fusion improved accuracy by 15–20% compared to single-modality models, demonstrating the advantages of integrating neural, physiological, and visual cues.

➢ **Classifier Performance**

- The Decision Tree consistently achieved the highest accuracy across all modalities, making it the most effective model for multimodal emotion classification.
- k-Nearest Neighbors (k-NN) and Multi-Layer Perceptron (MLP) performed moderately well but were less effective than Decision Tree models.
- Deep learning models (CNNs) were effective for digital image classification, but their performance varied when applied to EEG and thermal features.

These findings demonstrate that physiological and neural signals provide deeper emotion-related insights than facial expressions alone, making EEG-Thermal fusion an ideal approach for future emotion recognition applications.

### 6.3 Implications for Real-World Applications

The results of this study have broad implications across various industries, particularly in mental health monitoring, AI-driven affective computing, security systems, and human-computer interaction.

➢ **Mental Health Monitoring**

The ability to accurately detect emotions using EEG and thermal imaging could revolutionize stress, anxiety, and depression monitoring.

Applications include:

- Early detection of depression and anxiety disorders.
- Biofeedback therapy to help individuals regulate their emotions.
- Wearable neurophysiological monitoring devices for continuous emotional tracking.

- Affective Computing and AI-Based Human Interaction

The integration of multimodal emotion recognition into artificial intelligence can enhance human-computer interactions, enabling AI systems to respond more naturally to human emotions.

Enhancements for AI-driven systems:

- Emotion-aware virtual assistants and chatbots.
- Emotion-adaptive gaming and VR experiences.
- Personalized education systems that adapt based on student engagement.

- Security and Surveillance Applications

Multimodal emotion recognition can improve lie detection and behavioral analysis.

Potential applications include:

- Lie detection systems in forensic investigations.
- Emotion-based security screening at airports and borders.
- Driver fatigue detection using EEG-Thermal monitoring.

- Healthcare and Rehabilitation

Emotion recognition can be used in cognitive and emotional therapy.

Use cases include:

- Post-stroke emotional rehabilitation.
- Autism spectrum disorder (ASD) therapy.

# 7. CONCLUSION AND FUTURE WORK

## 7.1 CONCLUSION

This study presents a multimodal emotion recognition framework that integrates EEG signals, thermal imaging, and digital facial images to improve emotion classification accuracy. The findings validate the hypothesis that emotions manifest across multiple modalities simultaneously, with neural activity influencing facial expressions and physiological responses. By leveraging feature-level and decision-level fusion techniques, the proposed model successfully enhances the reliability of emotion recognition systems.

The experimental results demonstrate that EEG and thermal imaging outperform digital image-based FER, proving that physiological and neural signals provide deeper and more reliable emotion-related insights. The Decision Tree classifier consistently achieved the highest accuracy (98.3%), making it the most effective model for multimodal FER applications. Additionally, the correlation analysis between Theta Power (EEG) and Thermal Entropy for Surprise further confirms that emotionally arousing stimuli elicits both neural and physiological responses, highlighting the importance of integrating multiple modalities.

Moreover, the study underscores the limitations of unimodal approaches, particularly those based solely on digital facial expressions, which are prone to lighting variations, occlusions, and voluntary suppression. The superior performance of EEG-Thermal fusion emphasizes that emotion recognition systems should incorporate deeper physiological and neural markers to improve classification accuracy.

The implications of this research extend across multiple fields, including mental health monitoring, affective computing, human-computer interaction, security, and rehabilitation. The study's contributions pave the way for real-time multimodal emotion tracking, which could revolutionize AI-driven emotion-aware applications in healthcare, education, and security.

Despite the promising results, this study also highlights several challenges, including computational complexity, dataset limitations, and the need for real-time implementations. These challenges present opportunities for future research, particularly in developing lightweight models that can be deployed in real-world scenarios.

## 7.2 FUTURE WORK

While this study demonstrates the potential of multimodal emotion recognition, several areas warrant further investigation to enhance its practical applicability. Future work should focus on improving real-time implementation, expanding dataset diversity, and optimizing deep learning-based fusion techniques.

### 7.2.1 Real-Time Emotion Recognition

One of the key challenges in multimodal emotion recognition is achieving real-time processing. Future studies should explore:

- The development of low-latency algorithms that can efficiently process EEG, thermal, and digital image data in real-time.
- Hardware optimization using edge computing and embedded AI for deploying multimodal systems in wearable devices.
- Implementing real-time feedback mechanisms for emotion-aware applications, such as biofeedback therapy for mental health disorders

### 7.2.2. Larger and More Diverse DataSetsA

The current study is based on 50 participants, which provides a strong foundation for multimodal FER research. However, future studies should focus on:

- Expanding the dataset to hundreds or thousands of participants to improve generalizability.
- Collecting data across different age groups, ethnic backgrounds, and cultural settings to evaluate model robustness.
- Incorporating dynamic emotions (e.g., mixed emotions, transitions between emotional states) for a more realistic emotion recognition system.

### 7.2.3. Advanced Deep Learning-Based Fusion Techniques

While the study uses feature-level and decision-level fusion, future work should explore more sophisticated deep learning architectures for improved accuracy and robustness. Possible advancements include:

- Implementing Transformer-based fusion models that dynamically weigh the contributions of EEG, thermal, and digital features.
- Using attention mechanisms to focus on the most relevant features across modalities.
- Exploring Graph Neural Networks (GNNs) to model relationships between different multimodal features.

### 7.2.4. Integration into Real-World Applications

To enhance the practical impact of multimodal emotion recognition, future research should focus on integrating this framework into:

- Mental Health Monitoring Systems: Deploying EEG-Thermal fusion models in clinical settings to detect early signs of depression, anxiety, and emotional dysregulation.

- Human-Computer Interaction (HCI): Implementing multimodal FER in adaptive AI systems, gaming, and virtual assistants to enable emotion-aware user experiences.
- Security and Lie Detection: Enhancing biometric-based surveillance systems by incorporating EEG and thermal features to assess emotional stress and deception.

### 7.2.5. Overcoming Computational Complexity

One of the major limitations of multimodal FER is the **computational cost** associated with processing high-dimensional EEG, thermal, and digital features. Future work should focus on:

- Developing lightweight deep learning models that can maintain high accuracy while reducing computational overhead.
- Exploring model compression techniques, such as quantization and knowledge distillation, to enable efficient real-time deployment.
- Leveraging cloud-based architectures for handling large-scale multimodal data processing.

### 7.2.6. Expanding Multimodal Emotion Recognition to Multilingual and Cross-Cultural Contexts

Most existing emotion datasets are collected from specific cultural and linguistic backgrounds, which may limit generalizability. Future research should aim to:

- Investigate how cultural factors influence multimodal emotion expression.
- Develop multilingual datasets to analyze how emotions are perceived and expressed differently across languages.
- Test multimodal models on cross-cultural datasets to ensure robustness in global applications.

# 8. REFERENCES


[1] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 2019.

[2] M. Goulart, A. M. da Silva, and R. A. Torres, "Multimodal emotion recognition using facial expression and thermal imaging," Pattern Recognition Letters, vol. 158, pp. 127-135, 2022.

[3] A. Vinciarelli and G. Mohammadi, "A survey of multimodal emotion recognition," ACM Computing Surveys, vol. 47, no. 3, pp. 1-36, 2015.

[4] K. M. Lucey, J. F. Cohn, T. Kanade, and J. Saragih, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 94-101, 2010.

[5] T. Bousefsaf, C. Maaoui, and A. Pruski, "Thermal imaging for stress monitoring: A review," IEEE Transactions on Biomedical Engineering, vol. 67, no. 3, pp. 573-585, 2020.

[6] G. Huang et al., "Convolutional Neural Networks for real-time facial expression recognition," IEEE Transactions on Multimedia, vol. 24, pp. 2180-2192, 2022.

[7] X. Li et al., 'EEG-based emotion recognition using deep learning,' IEEE Transactions on Neural Systems, 2022.

[8] J. Zhang, H. Liu, and T. Wang, "A novel hybrid deep learning framework for EEG-based emotion recognition," IEEE Access, vol. 10, pp. 23698-23709, 2022.

[9] Y. Zhang et al., "Decision-level fusion of EEG and facial expressions for improved emotion recognition," IEEE Transactions on Affective Computing, vol. 12, no. 2, pp. 314-328, 2021.

[10] Y. Zhang et al., 'Thermal imaging for stress detection,' Frontiers in Psychology, 2021.

[11] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42-55, 2012.

[12] D. Ioannou, S. Gallese, and G. Hasson, "Facial thermal imaging in affective computing: A review," IEEE Transactions on Affective Computing, vol. 11, no. 3, pp. 320-334, 2020

[13] S. Koelstra, C. Muhl, and J. P. L. De Silva, "DEAP: A Database for Emotion Analysis using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, 2012.

[14] A. Pavlidis, N. L. Eberhardt, and J. Levine, "Seeing through the face of deception," *Nature*, vol. 415, no. 6872, pp. 35, 2002.

[15] X. Wang, L. Zhang, and S. Li, "Facial expression recognition using deep learning: A review," *IEEE Access*, vol. 9, pp. 124567-124589, 2021.

[16] H. Huang et al., "Multimodal emotion recognition via deep feature fusion," IEEE Transactions on Image Processing, vol. 31, pp. 3654-3668, 2022.

# 9. BIOGRAPHY


**Bibek Ram(RA2211013010119)**

**Abhishek Kumar Jha(RA2211013010120),**

**Suraj Prasad Sah(RA2211013010093)**

**Dr. U. Snekhalatha**

Department of Biomedical Engineering

SRM Institute of Science and Technology.