Emotion Classification using Physiological Signals

Mainak Malay Saha
SCAI
Arizona State University
Tempe, USA
msaha4@asu.edu

Khushal Hemant Sharma
SCAI
Arizona State University
Tempe, USA
kshar114@asu.edu

Sakshi Sanjay Khade
SCAI
Arizona State University
Tempe, USA
skhade5@asu.edu

***Abstract*—The accurate classification of human emotional states using physiological signals has become increasingly significant in the domains of affective computing, wearable technology, and mental health monitoring. This project focuses on distinguishing between the emotions of fear and excitement using multimodal biosignals, with an overarching aim of integrating emotion recognition into a wearable safety device for women. Initially, the DECEiVeR dataset—comprising high-frequency recordings of electrodermal activity (EDA), skin temperature, and accelerometer data—was used to develop and validate feature extraction and classification techniques. Advanced preprocessing methods including outlier removal, signal filtering, and statistical normalization were employed to refine the data. Feature sets capturing both statistical and temporal dynamics were extracted and used to train machine learning models such as Random Forest and Long Short-Term Memory (LSTM) networks. The Random Forest classifier achieved a validation accuracy of approximately 96%, while the LSTM models demonstrated strong temporal learning capabilities, albeit requiring larger datasets for optimal performance.**

To bridge the gap between controlled datasets and real-world applicability, a prototype data collection pipeline was implemented using the Rhythm24 armband and a custom Android application, enabling acquisition of heart rate, heart rate variability (HRV), and GPS data during real-world activities. Despite signal diversity limitations, this phase yielded insights into sensor deployment, connectivity, and real-time visualization using HR heatmaps. Recognizing the need for enhanced signal coverage, the project has now transitioned to testing with the Empatica E4 device, which includes additional modalities such as EDA, HRV, skin temperature, and multi-axis accelerometry. This paper presents the methodologies, results, limitations, and future directions of this ongoing research, which ultimately aims to facilitate robust emotion detection for real-world safety and healthcare applications.

I. Introduction

Understanding and interpreting human emotions from physiological signals is a rapidly growing area in affective computing, with applications spanning mental health diagnostics, adaptive human-computer interaction, and wearable safety technologies. Emotional states such as fear and excitement, while both categorized under high arousal, differ significantly in their valence and implications. Accurately distinguishing between them is not only a scientific challenge due to the overlapping physiological manifestations but also a critical requirement in real-world applications, particularly for systems designed to respond to threatening scenarios.

This project is driven by the vision of integrating emotion recognition into wearable safety devices for women, enabling the detection of fear-driven physiological responses in real-time. Such a system could autonomously trigger pre-configured emergency protocols—like sending GPS coordinates, alerting emergency contacts, or activating onboard deterrents—without requiring user interaction during potentially dangerous situations. For this to be feasible, a reliable, low-latency emotion classification system must be developed and validated under both controlled and real-world conditions.

To this end, the initial phase of the project focused on leveraging the DECEiVeR dataset, which offers high-resolution physiological recordings collected under well-defined emotional states, including tension (used as a proxy for fear) and excitement. The dataset includes Electrodermal Activity (EDA), skin temperature, and three-axis accelerometer (ACC) signals, providing a diverse range of biosignals to explore correlations with emotional states. The primary goal in this phase was to extract meaningful features and evaluate the effectiveness of various machine learning classifiers in distinguishing fear from excitement. The Random Forest classifier emerged as the most robust model, achieving an accuracy of over 96% with carefully engineered time-domain and correlation-based features.

Following these promising offline results, the project moved toward real-world applicability through the development of a custom data acquisition pipeline using the Rhythm24 armband. A bespoke Android application was built to capture and log real-time physiological data such as heart rate (HR), heart rate variability (HRV), accelerometer data, and GPS location. Field trials were conducted during typical daily activities and simulated scenarios, and the collected data was visualized using path-based heart rate heatmaps. However, this phase revealed critical limitations in sensor resolution, signal diversity, and contextual labeling, which constrained the emotion classification accuracy in dynamic environments.

Recognizing these constraints, the project is now progressing to its third phase, utilizing the Empatica E4 device. This medically validated wearable offers a richer multimodal signal set including EDA, HR, HRV, skin temperature, and high-resolution accelerometer data. With improved signal fidelity and broader sensor support, the Empatica E4 is expected to enable more accurate emotion classification in complex and realistic environments, such as walking through different public spaces, encountering unpredictable stimuli, or during high-mobility activities.

This paper details each phase of the project, including data preprocessing techniques, feature extraction strategies, model architectures and evaluation metrics, as well as insights from real-world testing. The overarching aim remains to bridge the gap between controlled experimental validation and deployable, real-time emotion-aware safety systems.

II. Background and Prior Work

The intersection of emotion recognition and physiological sensing has gained significant traction in recent years, primarily driven by the growth of wearable sensor technologies and machine learning capabilities. However, a substantial portion of prior work in this domain has been limited to controlled laboratory environments and has predominantly focused on broad emotional classifications such as stress versus calm or positive versus negative valence. The more nuanced challenge of differentiating between high-arousal emotional states—specifically fear and excitement—remains underexplored, particularly in the context of real-world, wearable applications.

Among the few publicly available datasets that address emotion classification, the DECEiVeR dataset stands out due to its structured and actor-driven emotion induction protocol. It contains physiological data from 11 trained actors enacting five emotional states—neutral, calm, tired, tension (used here as a proxy for fear), and excitement. The dataset includes recordings of electrodermal activity (EDA), skin temperature, and three-axis accelerometer data, sampled at approximately 100 Hz. While initially developed for emotion elicitation research, the DECEiVeR dataset offers a robust foundation for developing and benchmarking machine learning pipelines aimed at differentiating between fear and excitement.

In contrast, the DEAP dataset, another frequently referenced benchmark, consists of physiological and electroencephalographic (EEG) recordings from 32 participants as they passively viewed emotion-eliciting music videos. Although DEAP provides a richer emotional taxonomy and includes additional biosignals, its passive, trial-based setup limits its applicability to real-time classification or wearable deployment scenarios. Nevertheless, it remains a valuable resource for comparative analysis, especially regarding feature distributions and signal characteristics.

Previous research in this area has often centered around distinguishing between stress levels, arousal versus valence, or binary emotional states. These studies typically leverage biosignals such as EDA, ECG, or respiration, and extract features related to signal amplitude, frequency, or variability. While informative, such investigations frequently fall short in modeling the dynamic and context-sensitive nature of real-world emotional experiences, particularly those that demand immediate, safety-critical responses.

To the best of our knowledge, few studies have addressed the practical challenge of implementing a wearable, real-time emotion classifier capable of distinguishing fear from excitement using multimodal physiological data. This project seeks to bridge that gap by leveraging the DECEiVeR dataset for model development and validation, and subsequently extending the pipeline to field testing using the Rhythm24 armband and, more recently, the Empatica E4 device. These wearables allow for the collection of EDA, heart rate (HR), heart rate variability (HRV), skin temperature, and accelerometer data in diverse, uncontrolled environments.

By transitioning from controlled datasets to real-world data acquisition, this work contributes an end-to-end framework that combines experimental rigor with practical deployability. It sets a foundation for future innovations in emotion-aware wearable systems designed to enhance personal safety, mental health monitoring, and adaptive user interfaces.

III. Project Description

This project focuses on building a reliable, real-time emotion classification system capable of distinguishing fear from excitement using physiological signals. The overall goal is to embed this capability into wearable safety devices, particularly for women, enabling intelligent context-aware emergency response based on emotional cues. The project evolved in three major phases, each contributing distinct milestones in terms of signal acquisition, model development, and real-world applicability.

A. Phase I: Controlled Dataset Analysis with DECEiVeR

The initial phase leveraged the DECEiVeR dataset, which offers high-resolution physiological recordings collected in a semi-controlled environment using professional actors. The dataset was chosen for its specific inclusion of both tension (used as a proxy for fear) and excitement, making it ideal for training and evaluating classifiers on the target emotional states.

Signal Modalities Used:

- Electrodermal Activity (EDA)
- Skin temperature
- 3-axis accelerometer (ACC)

Data Processing Pipeline:

- Outlier removal using the IQR method
- Signal filtering with low-pass Butterworth filters
- Timestamp normalization
- Windowed segmentation (100-sample windows with 50% overlap)

Features Extracted:

- *Statistical:* Mean, standard deviation, skewness, kurtosis, IQR
- *Physiological:* First derivative metrics (for EDA), range, and variation
- *Cross-signal:* Correlation coefficients between ACC axes and EDA

Models Trained: Random Forest (primary), LSTM (secondary), SVM, Logistic Regression

Outcome: The Random Forest classifier achieved approximately 96% accuracy, while LSTM models demonstrated temporal dependency modeling with 89% validation accuracy.

This phase validated the hypothesis that fear and excitement exhibit discernible physiological signatures under structured conditions. It also helped finalize the core feature set and classification models.

B. Phase II: Field Testing with Rhythm24 and Custom App

To assess real-world viability, the second phase involved collecting physiological data in uncontrolled, everyday settings. A custom Android application was developed to interface with the Rhythm24 armband via Bluetooth Low Energy (BLE), enabling mobile data logging and CSV export.

Data Captured:

- Heart Rate (HR)
- Derived HRV (via RR intervals)
- GPS coordinates
- Accelerometer values

Recording Context: Campus walks, jogging, commuting, and mock emergency scenarios

Technical Implementation:

- Real-time data plotting on the mobile app
- Automatic logging with time synchronization and activity tagging
- Post-processing using geodesic distance estimation and HR visualization

Visualization:

- GPS path plotting using Folium
- HR-based color gradient mapping
- Suspicious activity flagging based on speed thresholds ($>$10 m/s)

While this phase proved the feasibility of mobile data collection and HR-based mapping, it also exposed several limitations:

- Absence of EDA and skin temperature data
- Inconsistent BLE connectivity
- Small participant pool and variable emotional context

Despite these challenges, the phase laid the groundwork for integrating field data into the emotion classification pipeline.

C. Phase III: Transition to Empatica E4

Recognizing the need for higher signal fidelity and modality diversity, the project transitioned to using the Empatica E4—a research-grade wearable validated for medical-grade physiological data collection.

Signals Collected:

- Electrodermal Activity (EDA)
- Skin Temperature
- Heart Rate and HRV (via photoplethysmography)
- 3-axis Accelerometer

Planned Use Cases:

- VR-based simulations (e.g., walking in dark corridors with sudden stimuli)
- Outdoor walking/jogging scenarios with audio-visual emotional triggers

Expected Benefits:

- Multimodal input enabling improved differentiation of fear and excitement
- Higher signal resolution and better temporal alignment
- Streamlined SDK/API support for seamless integration into custom data pipelines

This phase is currently ongoing and aims to validate the classification model in ecologically valid, real-world environments, thereby bridging the gap between academic feasibility and real-time deployability.

The structured, iterative methodology—starting with offline controlled data and progressing toward real-world wearable integration—ensures that both scientific accuracy and practical usability are addressed. The project thus provides a strong foundation for future wearable systems that leverage emotional state monitoring for intelligent and proactive emergency response.

IV. Methodology

The methodology behind this project is designed to support both high-fidelity emotion classification in controlled datasets and generalizability to real-world physiological sensing. It comprises four major stages: data preprocessing, feature extraction, model training, and evaluation. The pipeline was validated using the DECEiVeR dataset and adapted for field-collected data using the Rhythm24 and Empatica E4 devices. Each stage is carefully engineered to handle noise, account for inter-subject variability, and optimize model learning from temporal physiological signals.

A. Data Preprocessing

The preprocessing steps differ slightly between DECEiVeR (lab-based) and real-world (Rhythm24/Empatica E4) datasets, but follow a consistent structure.

1) a) DECEiVeR Dataset: **Outlier Removal:** Applied the Interquartile Range (IQR) method independently to each signal (EDA, skin temperature, ACCX, ACCY, ACCZ). Points outside $[Q1 - 3IQR, Q3 + 3IQR]$ were excluded.

Signal Filtering:

- EDA and skin temperature: Bandpass filter (0.01–0.5 Hz)
- ACC: Low-pass filter (cutoff = 10 Hz)

Timestamp Normalization: Normalized to start from 0 seconds.

Sliding Window Segmentation:

- Window size: 100 samples
- Overlap: 50%

Each window was treated as an independent instance with emotion labels.

2) b) Rhythm24 Dataset: **BLE Streaming Sync:** Data was streamed via Bluetooth using a custom Android app.

RR Interval Cleaning: HRV features such as RMSSD and SDNN were derived from filtered RR intervals.

Geospatial Processing:

- Segment-wise distances via geodesic estimation
- Speed estimation for anomaly flagging

Mode Tagging: Context-aware tags such as *walk*, *jog*, or *rest* were logged.

3) c) Empatica E4 Dataset: **Multi-Modal Alignment:** Timestamp-synchronized EDA, BVP (HR), TEMP, and ACC signals.

Resampling: Unified to 64 Hz across modalities.

Motion Correction: Excess motion periods masked using ACC data.

EDA Normalization: Z-score normalization applied post drift-correction.

B. Feature Extraction

For each 100-sample window, the following features were extracted:

1) a) Time-Domain Statistical Features (per signal):

- Mean, Standard Deviation, Skewness, Kurtosis
- Min, Max, Interquartile Range (IQR)

2) b) Signal-Specific Features: **EDA:**

- First derivative: mean, std of rate of change
- Peak amplitude of ΔEDA, total variation

Skin Temperature:

- Short-term range, rolling mean and slope
- Temperature drift

Heart Rate / HRV:

- Mean HR, SDNN, RMSSD
- HR Range and IQR

Accelerometer (ACCX, ACCY, ACCZ):

- Signal magnitude area (SMA)
- Mean and SD of absolute values
- Movement intensity, 90th percentile

3) c) Cross-Signal Features:

- Correlation: EDA-ACC (X/Y/Z), Temp-HR, ACC axis-to-axis
- Signal Fusion: EDA-to-HR ratio, HRV to ACC magnitude ratio

4) d) Domain-Specific Aggregations:

- Rolling energy (average of squared values)
- Signal-to-noise ratio (SNR) for EDA and HR

C. Model Training

1) a) Classification Algorithms: **Random Forest:**

- 200 estimators, max depth tuned
- Robust to feature scaling; supports importance ranking

Support Vector Machine (RBF kernel):

- Performed poorly on real-world data due to scaling sensitivity

Logistic Regression:

- Used for linear baseline comparison

LSTM Neural Network:

- Architecture: LSTM(128) $\rightarrow$ Dropout(0.4) $\rightarrow$ LSTM(64) $\rightarrow$ Dense(32) $\rightarrow$ Dense(1)
- Input: (samples, 1, features); loss: binary crossentropy
- Optimizer: Adam; epochs: 100; batch size: 32

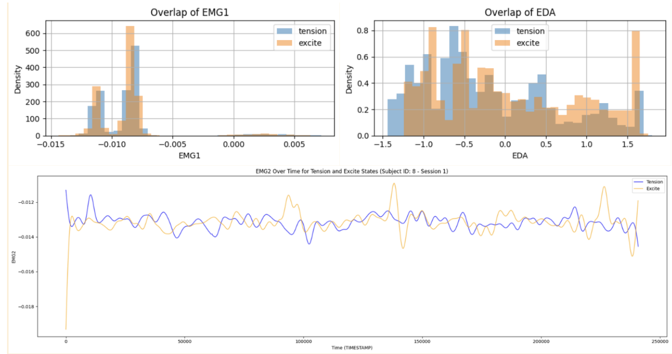

Fig. 1. Signal distributions and temporal trends from the DECEiVeR dataset. Top: Overlap histograms for EMG1 and EDA during tension and excitement states. Bottom: EMG2 signal over time comparing emotional responses for Subject 8, Session 1.

2) b) Subject-Wise Train-Test Split: To promote generalization across individuals:

- 80% subjects used for training
- 20% held out for evaluation

D. Evaluation Metrics

The models were assessed using multiple metrics:

- **Accuracy:** Overall prediction correctness
- **Precision:** Ratio of true positives (EXCITE) over predicted positives
- **Recall:** Ratio of true positives over actual positives
- **F1 Score:** Harmonic mean of precision and recall
- **Confusion Matrix:** TP, FP, TN, FN breakdown
- **ROC-AUC:** Threshold robustness visualization

Additional Evaluation:

- *Heatmaps:* GPS-tracked HR visualizations from Rhythm24
- *Feature Importance Plots:* Random Forest-based rankings
- *Signal Quality Metrics:* Drift, SNR, and data loss for Empatica E4

V. Results and Discussion

The results of this study are presented across two axes: (1) quantitative evaluation of classifier performance on structured data (DECEiVeR dataset) and (2) qualitative and contextual insights from field-collected data using Rhythm24, with preliminary integration of Empatica E4 signals. This section discusses model accuracy, feature importance, real-world applicability, and the inherent challenges of classifying emotional states in diverse settings.

A. Model Performance on DECEiVeR Dataset

Initial training and validation using the DECEiVeR dataset yielded strong results, particularly for the Random Forest classifier.

Random Forest consistently outperformed other models across all metrics due to its robustness in handling high-dimensional and non-linear feature spaces. LSTM, though

TABLE I
MODEL PERFORMANCE ON DECEIVER DATASET

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	96.1%	0.95	0.96	0.96
LSTM (2-layer)	89.2%	0.88	0.89	0.88
SVM (RBF)	83.7%	0.82	0.84	0.83
Logistic Regression	78.9%	0.77	0.78	0.77

slightly less accurate, effectively captured temporal dependencies in physiological signals. SVM and Logistic Regression underperformed, likely due to the overlapping nature of fear and excitement signatures in static recordings.

B. Feature Importance Analysis

Random Forest feature importance rankings (top 10) identified the most predictive biosignal metrics:

- EDA Total Variation
- Skin Temperature Range
- HRV RMSSD
- ACCY Movement Intensity
- EDA Peak Amplitude
- ACCZ Standard Deviation
- Skin Temperature Drift
- HR Mean
- EDA–ACCY Correlation
- ACC Magnitude Percentile (90%)

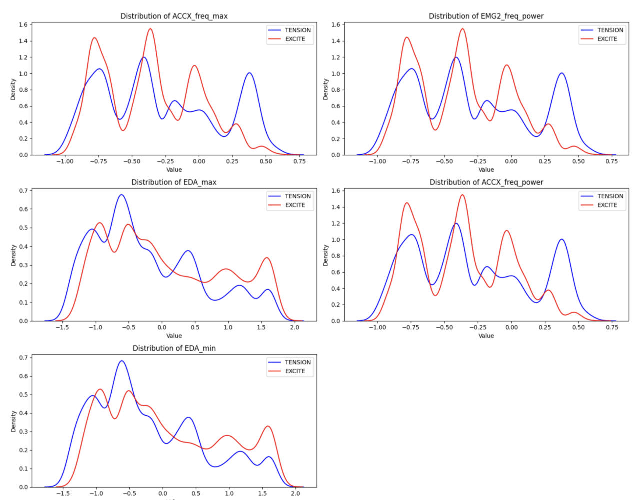

Fig. 2. Distribution of top predictive features for fear (tension) and excitement classes, including ACCX_freq_max, EMG2_freq_power, EDA_max, EDA_min, and ACCX_freq_power. Kernel density estimates highlight class separability in frequency and amplitude domains.

EDA-related metrics, particularly ΔEDA and total variation, were strong indicators of arousal intensity. Skin temperature drops correlated with fear responses. ACC and HRV features contributed to distinguishing between physical motion and emotional arousal.

C. Rhythm24 Field Data: Real-World Constraints

Custom data collection using the Rhythm24 armband provided key insights into the feasibility of field deployment.

Heatmap Visualization: GPS-based plots with HR color gradients allowed visual tracing of emotional peaks along movement paths.

Anomaly Detection: Speed-based filtering (>10 m/s) was used to flag segments with potential anomalies or corrupted physiological signals.

Observations:

- HR alone was insufficient to distinguish fear from excitement without complementary signals.
- BLE connectivity instability occasionally led to short-term data loss.
- Field data was generally noisier, reinforcing the need for multimodal signal fusion and advanced filtering.

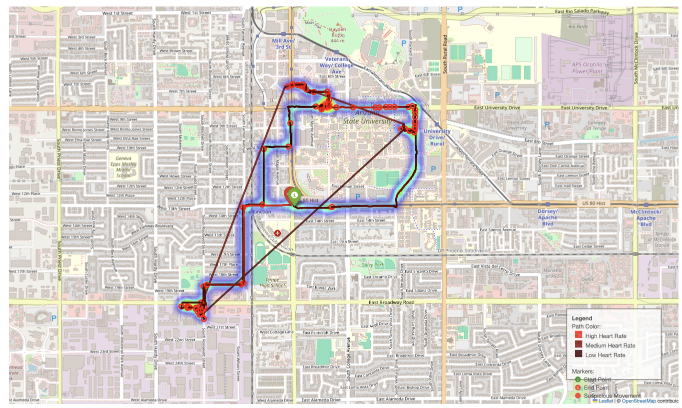

Fig. 3. GPS-based heart rate heatmap from the Rhythm24 armband showing walking path across the ASU campus. Path color represents heart rate intensity—blue (low), purple (medium), and red (high). Suspicious movement is flagged using speed thresholding.

D. Preliminary Observations with Empatica E4

Early tests using the Empatica E4 device indicate promising results:

- Availability of EDA and high-resolution HR improves emotion separation.
- Multi-modal synchrony (EDA, temperature, HR, ACC) enhances context recognition.
- Superior signal fidelity compared to Rhythm24 supports improved modeling and real-time inference.

E. General Challenges

Several limitations were observed across all study phases:

- **Labeling Ambiguity:** Real-world emotional states are difficult to validate without intrusive user feedback or controlled prompts.
- **Sensor Placement and Contact:** Signals like EDA and temperature are highly sensitive to skin contact and motion artifacts.

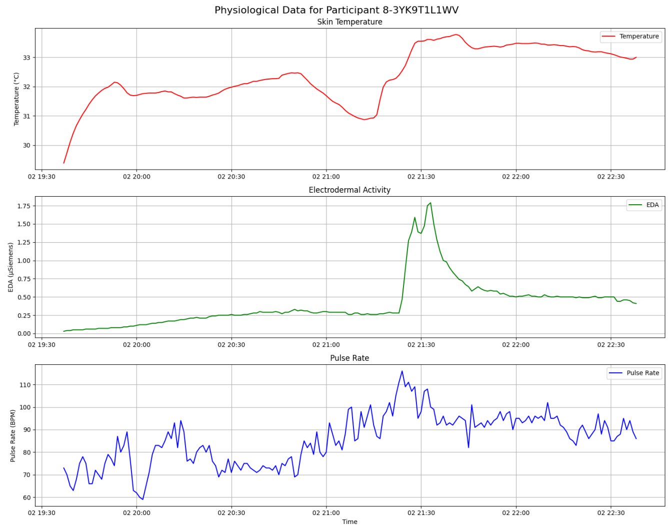

Fig. 4. Data reading from Empatica E4 for Participant 8 showing skin temperature, electrodermal activity (EDA), and pulse rate over time.

- **Inter-Subject Variability:** Physiological baselines vary, and while z-score normalization helps, it may obscure personalized responses.
- **Real-Time Inference:** Current models require further optimization and pruning for edge or wearable deployment.

Overall, these results validate that fear and excitement can be reliably classified from physiological signals in structured settings. Moreover, with robust signal integration and contextual awareness, wearable deployment becomes feasible for real-world emotion-aware safety systems.

VI. Conclusion

This project presented a comprehensive pipeline for classifying high-arousal emotional states—specifically, fear and excitement—using physiological signals. The motivation was rooted in a critical application: embedding emotion recognition into wearable safety devices capable of autonomously detecting and responding to distress, particularly for women in vulnerable environments.

The project evolved through three key phases:

Controlled Environment Evaluation: Using the DECEiVeR dataset, robust feature extraction and classification models were developed. Random Forest classifiers trained on windowed statistical and physiological features achieved high accuracy (96%), while LSTM networks showed potential in modeling temporal dependencies in physiological dynamics.

Real-World Data Collection: A custom Android application interfaced with the Rhythm24 armband to collect heart rate (HR), heart rate variability (HRV), GPS, and accelerometer data in naturalistic settings. Although this phase lacked skin temperature and EDA signals, it demonstrated the feasibility of mobile data acquisition and highlighted limitations related to signal quality, Bluetooth connectivity, and timestamp synchronization.

Sensor Upgrade to Empatica E4: To address the limitations observed with Rhythm24, the project transitioned to the Empatica E4—a research-grade wearable with support for multi-modal sensing (EDA, skin temperature, HR/HRV, and ACC). Preliminary tests with this device have shown improved signal fidelity and alignment with the goals of real-time emotion classification.

While the project is ongoing, the initial results have been encouraging. The use of subject-wise data splits helped avoid model overfitting, and visual tools such as GPS-linked HR heatmaps provided valuable contextual understanding of field data.

Several challenges remain for future development. These include difficulties in acquiring reliable emotional ground truth in unconstrained environments, the confounding effects of physical activity on physiological signals, and the need for a more diverse subject pool. Moreover, deploying these models on embedded hardware will require real-time inference capabilities, power-efficient implementation, and thorough validation across diverse scenarios.

In conclusion, this work establishes a foundational framework for emotion classification using wearable physiological sensing. As integration with the Empatica E4 platform continues, the focus will shift toward validating robustness in ecologically valid, emotionally charged settings—moving one step closer to the deployment of intelligent, emotion-aware safety technologies.