Contents

List of Figures

List of Tables

Abstract

Gender-based Violence (GBV) poses a significant global challenge, necessitating innovative technological interventions for women's safety. This project introduces Raksha, an AI-powered system designed to detect and respond to threats in real-time. Central to this system is the Raksha watch, a wearable device that captures physiological signals such as heart rate, galvanic skin response, and skin temperature, alongside environmental audio. These inputs are processed by an AI model trained on the Women and Emotion Multimodal Affective Computing (WEMAC) dataset, which uses virtual reality generated data to simulate real-life emotional responses. The Raksha watch continuously monitors these inputs, detecting emotions associated with potential threats. Upon identifying a risk, the system triggers an automatic protection protocol, including sending an SOS message with GPS coordinates, recording audio and video of the incident, and providing real-time safety guidance through a mobile application. The app also serves as the system's control center, allowing users to manually activate safety features and utilize geofencing for added protection. The project uses key technologies, including AI-based threat detection, physiological sensors, real-time audio processing, and GPS tracking. Robust encryption ensures data security and privacy throughout the process. By integrating wearable technology with AI and mobile applications, Raksha offers a proactive, intelligent approach to mitigating GBV, empowering women with real-time tools to safeguard themselves in dangerous situations.

Chapter 1

Introduction

Gender-based Violence (GBV) is any act of physical, sexual, or psychological violence directed toward the female gender. It persists worldwide, occurring in every region, country and culture and cuts across income, class, race and ethnicity. It impedes development and prevents women and girls from enjoying their human rights and fundamental freedoms. A datum that helps ponder its impact on society shows that more than 27% of ever-partnered women aged between 15 and 49 experienced physical or sexual violence by intimate partners from 2000 to 2018. Another worrying statistic is that, in 2020, approximately 47,000 women and girls were killed worldwide by their intimate partners or other family members, meaning a woman or girl is killed by someone in her own family every 11 minutes. In India, the situation is especially dire, with nearly half of women in certain regions reporting experiences of violence. National crime statistics reveal over 89,000 cases of cruelty by husbands or relatives and more than 21,000 rape cases annually. These figures highlight the urgent need for a more intelligent, proactive approach to personal safety for women and girls. From a sociological point of view, particular emphasis on education is essential to eradicate, combat, and prevent GBV, but it requires several generations to produce a change in society. In the meantime, technology can be fundamental for helping prevent and combat GBV and empowering women.

With this focus, we propose Raksha, an autonomous system powered by AI. The goal of this project is to automatically or manually report when a woman is in a risk situation to trigger a protection protocol. This risk situation identification is performed by detecting threat-related emotions in the user through a multi-modal intelligent engine feeding with physiological and audio data captured by a wearable device, which in our case is a watch. The collected data is then sent to a mobile application that acts as the control center for managing the system's features. It allows users to customize alerts, access safety tools, receive real-time safety guidance, and collect evidence that may assist law enforcement in case of an incident. To develop this autonomous system, we will follow a systematic process, beginning with the following key steps:

1. **Dataset and Model Training**: In our project, we will be utilizing the Women and Emotion Multimodal Affective Computing (WEMAC) dataset belonging to the UC3M4Safety database, which offers several unique advantages. This dataset was created using immersive technology, specifically virtual reality, to evoke emotions that closely resemble real-life experiences. With a sample size of 100 volunteers, it includes more participants than similar datasets, allowing for more robust insights. Additionally, the dataset has been designed with a gender-sensitive perspective, using a modified labeling system to ensure that different emotional responses are accurately captured. The use of multiple sensory systems also adds

depth, providing a rich, diverse set of data that aligns well with our project's goals. Our model will be trained on this dataset to identify real-time threat-related emotions that indicate risk situations.

2. **Wearable Device**: Following model training, the system will be integrated into a wearable device - the Raksha watch, which serves as the central edge device for physiological monitoring and audio capture to detect fear and threatening situations. The device is equipped with a range of components, which can be classified into four key groups:

- ***Physiological Sensors***: The watch is equipped with sensors for monitoring heart rate (HR), galvanic skin response (GSR), and skin temperature (SKT). These sensors were chosen for their reliability in emotion recognition and ease of integration into wearable devices.
- ***Audio Capture System***: A built-in microphone captures both audio and speech signals to help detect verbal threats.
- ***Power Management Elements***: Efficient power management to ensure long-lasting battery life.
- ***Microprocessor Unit***: This unit handles all the processing tasks, including physiological signal analysis and real-time response triggers.

The watch continuously monitors both physiological and contextual data, transmitting it to the mobile application for real-time processing. Upon detecting a potential threat, the system activates a protection protocol, which triggers the following functionalities:

- ***SOS Message***: This emergency message consists of our current location tracked by GPS and sent to the GSM module.
- ***Video and Audio Recording***: When the Raksha app is activated, the watch records the incident using its built-in camera and microphone. This data can later be sent to law enforcement for further investigation.
- ***Real-time Safety Guidance***: Utilizes Google Maps APIs to overlay escape routes and safe spots, guiding the user to nearby police stations or safe locations.

In case the automatic system fails to detect a threat, the user can manually trigger the protection protocol by pressing the panic button, which activates the same emergency procedures as the automatic detection system.

3. **Mobile Application**: The mobile application serves as the central hub for processing and responding to the data collected from the wearable device, including physiological and contextual information. It manages all core safety features,

ensuring comprehensive protection for the user. In cases where the AI algorithm or manual panic button fails to detect or trigger a response during a threat, the user can directly access all the protection protocol functionalities through the mobile application. These features include audio and video capture, SOS messaging, location tracking, safety routes, and incident response. Additionally, the app uses geofencing to create virtual boundaries around safe locations. If the user leaves these boundaries, the system can send alerts or trigger the protection protocol, using the watch's GPS to track the user's location. This feature provides an extra layer of security, especially in unfamiliar or potentially dangerous areas. A notable feature of the app is its incident response capability. When the protection protocol is activated, the wearable device tracks the user's real-time location, and the app shares this information and safety alerts with users in the vicinity. This creates a community of individuals who can respond and assist the person in danger. Additionally, all data collected - physiological, contextual, and audio - visual during an incident is stored and can be used as evidence by law enforcement agencies.

4. **Security**: Data privacy and security are extremely important, which cannot be ignored. We implement encryption techniques to ensure that any data sent during emergencies is secure and protected from hacking, preserving user privacy at all times.

In summary, Raksha introduces a novel approach to personal safety by blending advanced technology with real-time threat detection. Its sophisticated features ranging from SOS alerts and geofencing to real-time tracking and encrypted incident recording empower individuals with unparalleled protection. By ensuring immediate response and legal safeguards, Raksha redefines the standard for proactive and intelligent safety solutions.

Chapter 2

Literature Survey

Calero et al. (2022) introduced the "WEMAC (Women and Emotion Multi-modal Affective Computing) dataset" [1] a multi-modal resource aimed at detecting emotions, particularly fear, related to Gender-based Violence (GBV). The dataset contains physiological signals (blood volume pulse, galvanic skin response, and skin temperature) and speech data from 100 women exposed to 42 curated audiovisual clips in virtual reality. The stimuli were designed to evoke both fear and non-fear emotions, and data was synchronized across modalities for comprehensive analysis. This dataset enables the training of AI models for real-time emotion detection in critical situations, where recognizing fear can inform rapid safety interventions. What makes WEMAC unique is its gender-specific focus, addressing the gap in emotion datasets by emphasizing the need for accurate fear detection in women. The modified Self-Assessment Manikin (SAM) for emotion labeling, combined with physiological and speech data, supports robust multimodal analysis. The dataset's technical validation demonstrated strong correlations between targeted and self-reported emotions, particularly for fear, making it a valuable resource for developing AI-driven systems capable of providing real-time responses to emotional distress in high-risk environments. Hosseini et al. (2022) introduced a multimodal sensor dataset titled "A Multimodal Sensor Dataset for Continuous Stress Detection of Nurses in a Hospital". [2] This dataset, aimed at detecting stress among hospital nurses during the COVID-19 pandemic, captured real-world biometric signals such as electrodermal activity (EDA), heart rate, and skin temperature using wearable devices like the Empatica E4. It plays a significant role in understanding stress patterns in high-pressure environments by leveraging machine learning models for real-time stress detection. This helps improve healthcare worker well-being by facilitating early stress detection, which can be extended to similar applications in stress monitoring. Additionally, the study developed a continuous stress detection model using a Random Forest-based algorithm to analyze physiological signals and detect stress episodes. The integration of real-time data collection and end-of-shift surveys provides an innovative approach to stress monitoring, demonstrating how AI-powered solutions can enhance detection accuracy. This system's ability to monitor and respond to stress events in real time has broader implications, particularly for AI-driven safety systems where continuous monitoring and immediate intervention are crucial. Bamanikar et al. (2022), in their paper "Stress & Emotion Recognition Using Sentiment Analysis with Brain Signal" [3] developed a system that combines EEG data and sentiment analysis to detect human stress and emotional states like calmness, fear, and happiness. The research applies machine learning techniques to brain signals, providing a more accurate and real-time approach to identifying stress patterns. By analyzing EEG data, the system helps in recognizing emotions, with potential applications in mental health monitoring and behavior analysis.The paper outlines methodologies for EEG signal acquisition, feature extraction, and classification using algorithms like Linear Discriminant Analysis (LDA) and K-Nearest Neighbors (KNN). The authors propose this as a more effective alternative to traditional stress detection methods, which often lack accuracy and are costly. Their framework, using a single-channel EEG device, could be extended to medical diagnostics and human-computer interaction systems, making it a valuable tool for real-time emotional and stress recognition. Schmidt et al. (2018) presented "Wearable Affect and Stress Recognition: A Review" [4] offering a comprehensive overview of wearable-based affect and stress recognition technologies. This review focuses on the use of wearable sensors, primarily measuring physiological parameters such as heart rate, electrodermal activity (EDA), and respiratory patterns to detect affective states. The paper emphasizes the relevance of wearable systems for long-term affect recognition in real-life scenarios, outlining their ideal use cases for mental well-being monitoring and decision support systems. The review also highlights the advantages of multimodal setups, noting that affect recognition systems utilizing multiple sensor modalities achieve nearly 10% higher accuracy compared to unimodal approaches. The paper covers the physiological responses associated with different emotional states and provides a detailed discussion on common sensors such as photoplethysmography (PPG), electrocardiogram (ECG), and electromyogram (EMG), frequently used in wearable setups. The authors also describe the standard data processing pipeline for affect and stress recognition, including preprocessing, feature extraction, and classification methods. Additionally, the review provides guidelines for designing laboratory and field studies, contributing to the development of more robust and accurate wearable affect recognition systems. These insights are crucial for designing AI-driven systems that monitor emotional well-being and stress in real-time. Wu et al. (2021) in their paper "Transformer-based Self-supervised Multimodal Representation Learning for Wearable Emotion Recognition" [5] proposed a novel transformer-based self-supervised learning (SSL) framework aimed at improving wearable emotion recognition through the fusion of multimodal physiological signals. The system employs a temporal convolution-based modality-specific encoder along with a transformer-based shared encoder to capture both intra and inter-modal correlations. The model is pre-trained on a large dataset of physiological signals using a signal transformation recognition task to assign labels. This pretext task enhances the ability to extract generalized multimodal representations, thus boosting performance in emotion-related downstream tasks. The evaluation demonstrated state-of-the-art performance in several emotion recognition benchmarks by leveraging SSL techniques to mitigate the overfitting issues caused by limited labeled data. Their approach addresses one of the key challenges in emotion recognition from wearable devices: how to effectively utilize the relatively small amounts of labeled physiological data available for training. The method incorporates signal transformations that automatically generate labels for large amounts of data, allowing the model to learn more generalized representations. This not only improves classification accuracy but also increases the system's robustness when applied to real-world scenarios with limited or noisy data. Kim et al. (2024), in their paper "Health-LLM: Large Language Models for Health Prediction via Wearable Sensor Data" [6] propose a novel framework that integrates large language models (LLMs) with wearable sensor data for health predictions. This framework addresses the challenge of processing multimodal, time-series data, such as heart rate and sleep metrics, by enhancing the models with contextual information like user demographics and health knowledge. The authors evaluated eight state-of-the-art LLMs, including GPT-4 and Med-Alpaca, across thirteen health prediction tasks, covering domains such as mental health, activity tracking, and cardiac assessments. Their experiments revealed that context-enhanced prompts could improve LLM performance by up to 23.8%, with the fine-tuned Health-Alpaca model achieving the best performance in five out of thirteen tasks. The significance of this work lies in its demonstration that LLMs can effectively process both linguistic and non-linguistic data for health-related tasks. By incorporating user-specific, temporal, and health knowledge contexts into the prompting process, the authors improved the LLMs' ability to interpret complex time-series data. This study bridges the gap between traditional health models and LLMs, showing that even smaller models like Health-Alpaca can perform comparably to larger models like GPT-4 in specific tasks, making it a pioneering step toward more personalized and accurate health predictions using wearable technology. Al Sahili et al. (2023) in their paper "Multimodal Machine Learning in Mental Health: A Survey of Data, Algorithms, and Challenges" [7] provide a comprehensive overview of the application of multimodal machine learning (ML) to detect, diagnose, and treat mental health disorders. This survey discusses various data types used in the field, including text, audio, video, and physiological signals, and explores how integrating these modalities enhances the robustness and accuracy of mental health assessments. The authors highlight that multimodal ML offers improved insights into behavioral patterns, allowing for a more comprehensive understanding of mental health conditions such as depression, stress, bipolar disorder, and PTSD. The paper reviews state-of-the-art ML techniques, such as RNN/CNN-based, transformer-based, and graph neural network-based algorithms, which have shown promise in combining these diverse data types for accurate mental health detection. The authors also discuss the challenges faced in this emerging field, such as data availability, privacy concerns, and the need for robust benchmarking and evaluation methods. While multimodal ML has significant potential, its practical application is still hindered by these complexities. Additionally, the survey emphasizes the importance of addressing bias and ensuring fairness in mental health assessments to avoid inaccurate predictions that could affect underrepresented groups. Despite these challenges, the potential benefits of integrating diverse data types through multimodal machine learning are substantial, offering a path toward more effective and equitable mental health diagnostics and treatments. Reyner-Fuentes et al. (2022) presented "Detecting Gender-based Violence After effects from Emotional Speech Paralinguistic Features" [8] a study focused on identifying post-traumatic effects of gender-based violence (GBV) using paralinguistic cues in speech. Utilizing data from the WEMAC dataset, the study collected physiological and emotional speech data from women who experienced GBV and those who had not. Through the analysis of paralinguistic features like pitch, energy, and spectral characteristics, the study found measurable differences in how GBV victims express emotions, particularly fear, compared to non-victims. This differentiation is critical for developing AI-driven systems aimed at detecting emotional distress through speech. The study employed statistical tests and machine learning models, including a Multilayer Perceptron (MLP), to classify GBV victims based on their speech patterns. Feature extraction was performed using tools like librosa, with a focus on capturing frequency and temporal features. The results demonstrated that the MLP model could distinguish between victims and non-victims with high accuracy using a subject-dependent approach. The findings highlight the potential of leveraging paralinguistic speech features for real-time monitoring and response systems, particularly for assisting victims in high-risk scenarios. Hyndavi et al. (2020) introduced a "Smart Wearable Device Designed for Women's Safety, using IoT Technology" [9] to detect and respond to emergencies. The device is equipped with a pressure sensor, pulse-rate sensor, and temperature sensor, which together monitor abnormal changes in the user's environment and physical state. When two or more sensors detect irregularities, such as high pressure or increased heart rate, the system automatically sends the user's GPS location via a GSM module to predefined contacts. Additionally, a manual panic button is included for cases where the user can trigger the alert themselves. The device uses an Arduino Uno microcontroller to process sensor inputs and activate an alert mechanism, including a buzzer to attract nearby attention. The system operates without needing an internet connection, making it reliable in areas with limited connectivity. It is lightweight, cost-effective, and adaptable to various wearable forms like bracelets or watches. Future enhancements include integrating video and audio recording to provide evidence during emergencies, making it a comprehensive and practical solution for personal safety. Monisha et al. (2016) introduced "FEMME: A Women Safety Device and Application" [10] a comprehensive system designed to protect women in distress. The device incorporates GPS tracking, GSM communication, and an ARM controller, enabling users to send real-time location and distress messages with a single click. It features advanced functions such as a hidden camera detector, audio recording for evidence collection, and emergency calls with location updates sent every two minutes. The system synchronizes with an Android application via Bluetooth, allowing activation through both the device and a smartphone, ensuring reliable functionality even in areas without internet access. A major advantage of FEMME is it's all-in-one design, which eliminates the need to carry multiple safety devices. The system operates independently of GPRS, enhancing its effectiveness in low-network areas, and offers dual-mode operation through hardware or mobile software. Its hidden camera detector adds an extra layer of privacy protection. These features make FEMME a versatile tool for real-time response, monitoring, and forensic capabilities, offering a significant leap forward in personal safety technology for women. Miranda Calero et al. (2022) introduced "Bindi: Affective Internet of Things to Combat Gender-Based Violence" [11] an advanced AI-driven system designed to autonomously detect violent situations linked to Gender-Based Violence (GBV). Bindi leverages a multimodal approach to recognize fear-related emotions and initiate protective actions without requiring manual intervention. The system integrates technologies such as the Internet of Bodies (IoB), affective computing, and cyber-physical systems through a combination of wearable devices, smart sensors, and edge-fog-cloud architecture. The wearable devices, a pendant and a bracelet, capture physiological signals such as heart rate (HR), skin temperature (SKT), and galvanic skin response (GSR), as well as auditory signals. These are processed using a lightweight machine learning model on the bracelet, which detects possible fear-related emotions in real time. If a risky situation is identified, the pendant records auditory data, which is further analyzed for violent acoustic events. Bindi uses a hierarchical multisensorial information fusion strategy, achieving a fear detection accuracy of 63.61% in subject-independent evaluations. The system's unique architecture spans across three layers: edge computing on the wearable devices, fog computing via a smartphone application, and cloud computing for secure data storage and decision-making. This layered approach ensures efficient data processing, real-time responses, and long-term monitoring, making Bindi a cutting-edge solution for enhancing safety and protection in GBV scenarios.

Chapter 3

Proposed System

3.1 System Overview

3.1.1 Datasets

Database	Datasets	Conditions	Participants
UC3M4Safety Database	Audiovisual Stimuli: Videos Audiovisual Stimuli: Emotional Ratings	Online crowd-sourcing	General public and expert judges
	WEMAC: Biopsychosocial Questionnaire WEMAC: Physiological Signals WEMAC: Audio Features WEMAC: Self-reported Emotional Annotations	Laboratory	Women volunteers

Table 3.1: Overview of the UC3M4Safety Database and its datasets, conditions, and participants.

The proposed system utilizes six datasets derived from the UC3M4Safety WEMAC (Women and Emotion Multi-modal Affective Computing) database, created under the EMPATIA-CM project funded by the Comunidad de Madrid and co-financed by European programs. This dataset initiative aims to develop affective computing solutions for enhancing the safety of gender-based violence (GBV) victims by detecting emotions such as fear through physiological and audio signals. WEMAC is a gender-sensitive, open-access dataset collected during laboratory experiments involving 100 Spanish-speaking women participants aged between 20 and 77 years. Participants were exposed to validated audiovisual stimuli using immersive Virtual Reality (VR) technology, ensuring a realistic and intense emotional experience. Data acquisition followed strict ethical guidelines approved by the Ethics in Research Committee at Universidad Carlos III de Madrid (UC3M) and complies with GDPR standards. The dataset is licensed under Creative Commons Attribution 4.0 International License (CC BY 4.0), and access to encrypted files is provided upon signing a Data Usage Agreement.

- **Audiovisual Stimuli: Videos**
  The sixth dataset focuses on the audiovisual stimuli themselves. A curated selection process led to 42 emotional video clips, carefully balanced between fear and non-fear categories. Two batches of 14 clips were used during experiments, with emphasis on strong emotional elicitation, minimal ambiguity, and balanced representation across Valence-Arousal space dimensions. Videos were chosen based on expert evaluations and large-scale crowdsourcing to ensure emotion specificity and reliability.

- **Audiovisual Stimuli: Emotional Ratings**
  The fifth dataset contains all relevant experimental documentation, such as

participant informed consent forms, detailed experiment instructions, and sensor setup diagrams. These documents ensure ethical transparency, reproducibility, and offer valuable metadata for secondary analysis.

- **WEMAC: Biopsychosocial Questionnaire**
  The second dataset captures biopsychosocial background information provided before the experiments through a structured questionnaire. It contains 15 features for each participant, including age group classification, stress levels, prior traumatic experiences, fears, tendencies toward anxiety, substance usage, and physical activity history. These contextual variables are crucial for understanding the individual variability in emotional responses.

- **WEMAC: Physiological Signals**
  The fourth dataset involves physiological signals recorded during video exposure, including Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), and Skin Temperature (SKT). Acquired via the BioSignalPlux research system at 200 Hz sampling rate, the signals underwent preprocessing steps such as filtering, downsampling, and smoothing to enhance quality. Data is organized in MATLAB (.mat) format, with a clear structure for each participant and stimulus.

- **WEMAC: Audio Features**
  The first dataset consists of speech-derived audio features extracted after participants verbally responded to two oral questions following each video stimulus. Instead of sharing raw audio, six types of features were extracted: 38-dimensional librosa features (MFCCs, spectral and temporal properties), 88-dimensional eGeMAPS low-level descriptors from openSMILE, 6373-dimensional ComParE features, DeepSpectrum embeddings based on ResNet50 (2048 dimensions) and VGG19 (4096 dimensions), 128-dimensional VGGish embeddings, and 256-dimensional PASE+ embeddings. Each feature set is structured in one CSV file per user and stimulus, with second-wise alignment based on timestamp.

- **WEMAC: Self-reported Emotional Annotations**
  The third dataset encompasses self-reported emotional labels collected after each stimulus. Each of the 1400 entries (100 participants × 14 clips) includes dimensional emotion ratings (arousal, valence, dominance), liking of the video, familiarity with the emotion and situation, prior exposure to the clip, and discrete emotion categories experienced. Participants selected one out of 12 basic emotions: joy, sadness, anger, surprise, fear, calm, tenderness, disgust, contempt, hope, attraction, and tedium. This self-evaluation data acts as ground truth for validating emotion detection models.

3.1.2 System Architecture

Raksha is structured into three primary layers:

- **Wearable Device (Raksha Watch):** Captures physiological and environmental data.
- **Mobile Application:** Acts as the central hub for data processing, user interaction, and communication with the wearable device and cloud infrastructure.
- **Cloud Infrastructure:** Manages data storage, advanced processing, and facilitates long-term analytics and model training.

These layers work in unison, leveraging AI models to process data in real-time and trigger appropriate protection protocols when potential threats are detected.

3.1.3 System Components

- **Raksha Watch (Wearable Device):**
  - Functionality: Captures physiological signals such as heart rate (HR), skin tem perature (SKT), and galvanic skin response (GSR). Records environmental audio to detect verbal threats.
  - Hardware: Equipped with sensors (HR, SKT, GSR), a microcontroller, and a Bluetooth module for communication.
- **Mobile Application:**
  - Functionality: Serves as the user interface, offering features like real-time threat detection, SOS messaging, geofencing, and user preference management. Processes AI model outputs and communicates with both the wearable device and the cloud server.
  - Components: Integrates AI models (RNN, CNN, LSTM), handles data processing, and manages user interactions.
- **Cloud Server:**
  - Functionality: Facilitates data storage, backup, and advanced analytics. Ensures secure transmission and storage of sensitive information.
  - Components: Utilizes AWS services, MySQL for data storage, and serverless functions (AWS Lambda) for processing tasks.

3.1.4 Technology Stack

- **Programming Languages:** Python (AI Models), Dart (Mobile Application), Embedded C (Wearable Device Firmware).
- **Machine Learning Models:** Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM).
- **Frameworks and Libraries:** TensorFlow, Keras (Deep Learning Model Development), Scikit-learn (Data Preprocessing and Label Encoding), Pandas, NumPy, SciPy (Data Handling), Librosa (Audio Signal Processing).

- **Backend Services:** Amazon Web Services (AWS) for cloud storage and computing.
- **Database:** MySQL for structured data storage.
- **Communication Protocols:** Bluetooth Low Energy (BLE) (Planned for Wearable Communication), HTTPS (Secure Data Transmission), Twilio API (Real-time SMS Alerts).

3.1.5 High-Level Architecture Diagram

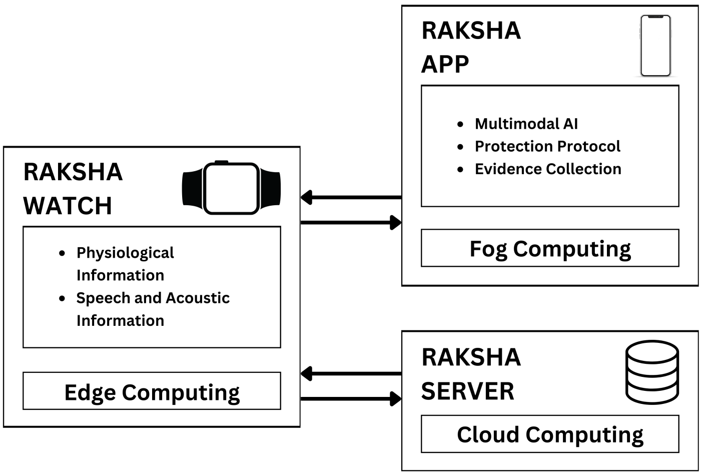


Fig. 3.1: High-Level Architecture

3.2 Architectural Design

3.2.1 High-Level Architecture

Raksha's architecture is divided into three distinct tiers:

- **Wearable Tier**: The Raksha Watch continuously captures physiological and environmental data, transmitting it to the mobile application.
- **Mobile Application Tier**: The mobile app processes incoming data using embedded AI models, manages user interactions, and initiates protection protocols upon threat detection.

- **Cloud Tier**: Provides data storage, backup solutions, advanced analytics, and supports long-term AI model training and improvements.

3.2.2 Detailed Component Design

3.2.2.1 Wearable Module

- **Purpose**:
  - Capture real-time physiological data, including heart rate (HR), skin temperature (SKT), and galvanic skin response (GSR).
- **Interfaces**:
  - Bluetooth Low Energy (BLE) connection to transmit data to the mobile application.
- **Components**:
  - Sensors: HR sensor, SKT sensor, GSR sensor.
  - Microcontroller: Processes raw sensor data.
  - Bluetooth Chip: Facilitates wireless communication with the mobile device.
- **Operation Flow**:
  - Sensors continuously monitor physiological parameters.
  - Data is processed by the microcontroller.
  - Processed data is transmitted via BLE to the mobile application at defined intervals.

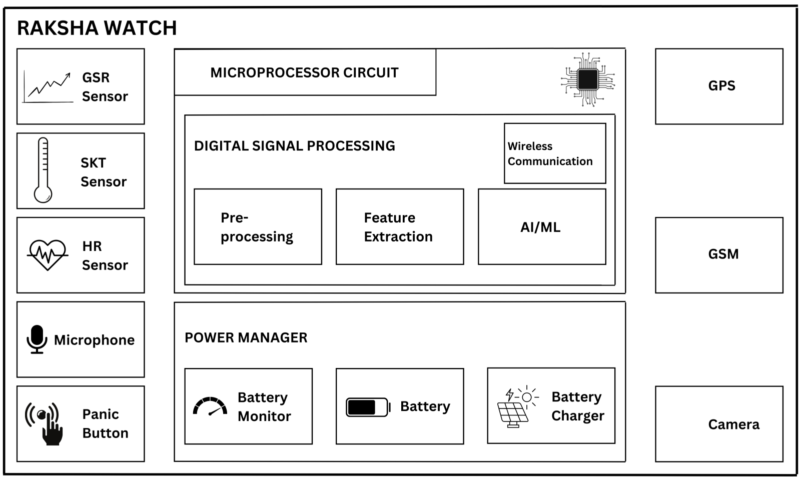


Fig. 3.2: Wearable Module

3.2.3 AI Threat Detection Module

- **Purpose**:
  - Analyze physiological and audio data to detect emotions and sounds indicative of danger.
- **Interfaces**:
  - Receives data from the Wearable Module.
- **Components**:
  - RNN and LSTM Models: Analyze temporal patterns in physiological data to detect fear or stress.
  - CNN Model: Processes audio data to identify distress signals or alarming sounds.
- **Operation Flow**:
  - Receives and preprocesses data from the wearable device.
  - Inputs data into AI models for analysis and determines threat levels based on model outputs.
  - Initiates protection protocols if a threat is detected.

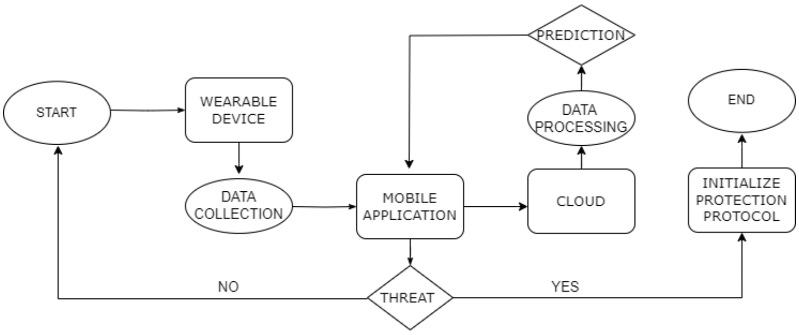


Fig. 3.3: AI Threat Detection Module

3.2.4 Mobile Application

- **Purpose**:
  - Facilitate immediate response actions during detected threats, including sending SOS messages, activating recordings, guiding users to safe zones, community support, and geofencing.
- **Components**:
  - Google Maps API Integration: Provides geolocation services and maps.
  - Notification Systems: SMS, email, and push notifications for SOS alerts.
  - Audio/Video Recording Module: Captures evidence during an incident.
- **Operation Flow**:
  - Upon threat detection, gathers current GPS location.

- Sends SOS messages to predefined contacts with location details.
- Activates audio/video recording to capture evidence.
- Guides the user to the nearest safe zone using map data.

3.2.5 Cloud Processing Module

- **Purpose**:
  - Manage long-term data storage, perform advanced threat analytics, and facilitate continuous AI model improvement through retraining with real-world data.
- **Components**:
  - MySQL Database: Stores structured data, including user profiles, historical data, and incident reports.
  - AWS Lambda Functions: Handles serverless processing tasks such as data ingestion, preprocessing, and triggering analytics workflows.
  - Data Security Tools: Ensure data is encrypted and access-controlled.
- **Operation Flow**:
  - Receives data transmitted from the mobile application.
  - Stores data securely in the MySQL database.
  - Performs batch analytics and aggregates data for reporting.
  - Utilizes collected data to retrain and improve AI models periodically.

3.2.6 Workflow Diagram

The proposed system begins with data collection through wearable devices, capturing physiological data (e.g., skin temperature, GSR, heart rate) and contextual information like speech and acoustic signals. This data is sent to a server for processing, where an AI model analyzes it to detect potential threats. If no threat is identified, the system continues monitoring, but if a threat is detected, a protection protocol is activated. This protocol includes sending SOS messages with GPS coordinates to emergency contacts, initiating audio and video recording for evidence, providing safety guidance, alerting nearby community members for support, and setting up geofencing to monitor movement. All collected data is securely stored for post-incident analysis, ensuring real-time protection and enabling evidence collection.The following figure describes how the users will interact and navigate through the Raksha system.

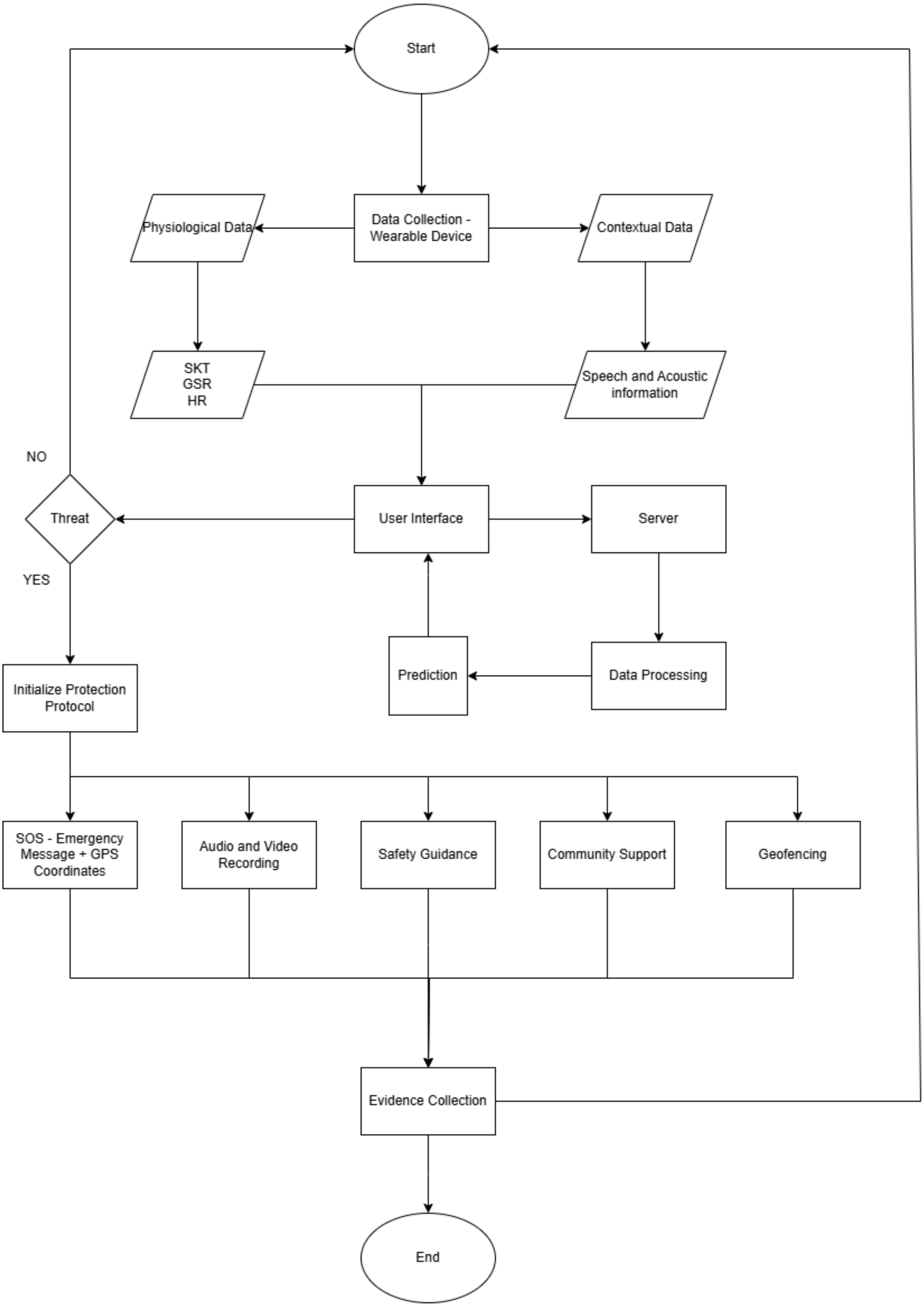


Fig. 3.4: Project Workflow

3.2.7 Data Management

3.2.8 Data Sources

- **Physiological Data:** Collected from Raksha Watch sensors measuring heart rate, skin temperature, and galvanic skin response.
- **Audio Data:** Captured through the wearable device's microphone to identify verbal threats or distress signals.
- **Location Data:** GPS data obtained via the mobile device to track the user's real-time position.
- **User Preferences and Profiles:** Includes user settings, emergency contacts, and personalized configurations.

3.2.9 Data Flow

- **Data Capture:** Raksha Watch continuously captures physiological and audio data.
- **Data Transmission:** Captured data is transmitted via Bluetooth to the mobile application.
- **Data Processing:** The mobile app processes incoming data using embedded AI models to assess threat levels.
- **Emergency Protocol Activation:** If a threat is detected, the mobile app initiates SOS messaging, activates audio/video recording, and provides navigational assistance.
- **Data Storage:** All data, including physiological readings, audio recordings, and incident reports, are securely transmitted to the cloud server for storage and future reference.
- **Data Access:** Authorized users and law enforcement can access relevant data through secure channels for incident resolution and analysis.

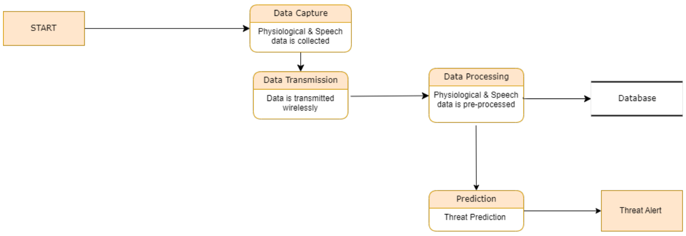

START
Data Capture
Physiological & Speech data is collected
Data Transmission
Data is transmitted wirelessly
Data Processing
Physiological & Speech data is pre-processed
Database
Prediction
Threat Prediction
Threat Alert

Fig. 3.5: Data Flow Diagram (DFD)

3.3 Algorithm and Model Integration

3.3.1 Model 1: Recurrent Neural Networks (RNN)

- **Purpose:** Detect real time emotional states from physiological data, identifying signs of fear or stress.
- **Integration:** Embedded within the mobile application to process continuous physiological signals from the Raksha Watch.
- **Implementation Details:** Trained on time-series data to recognize patterns in dicative of distress. Utilizes frameworks like TensorFlow/ Pytorch for deployment on mobile platforms.

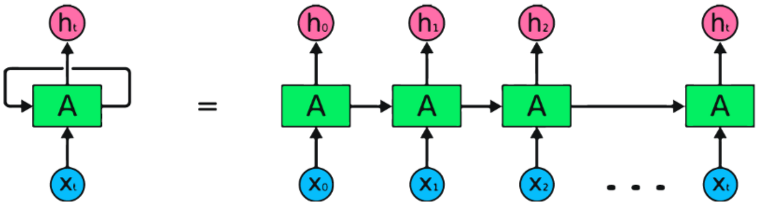


Fig. 3.6: Recurrent Neural Networks (RNN)

3.3.2 Model 2: Convolutional Neural Networks (CNN)

- **Purpose:** Analyze audio data to detect vocal stress or alarming sounds indicative of danger.
- **Integration:** Works alongside the RNN to process multimodal inputs (audio and physiological data) within the mobile application.
- **Implementation Details:** Trained on audio datasets containing distress calls and environmental sounds. Optimized for real-time processing to ensure swift threat detection.

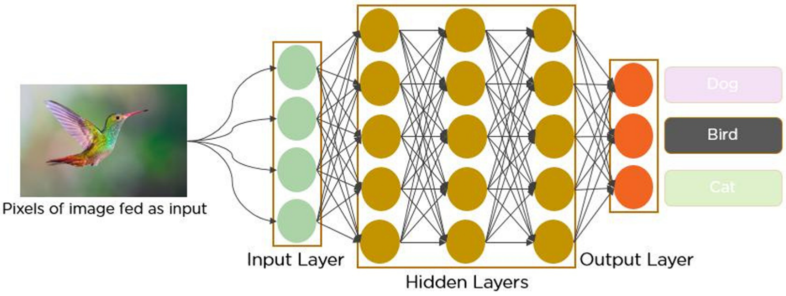


Fig. 3.7: Convolutional Neural Networks (CNN)

3.3.3 Model 3: Long Short-Term Memory (LSTM)

- **Purpose:** Capture long-term dependencies in physiological data to detect prolonged emotional distress.
- **Integration:** Integrated within the AI model of the mobile application for continuous and comprehensive data analysis.
- **Implementation Details:** Utilizes LSTM layers to remember past data points, improving detection accuracy over extended periods. Facilitates better differentiation between temporary spikes and sustained distress signals.

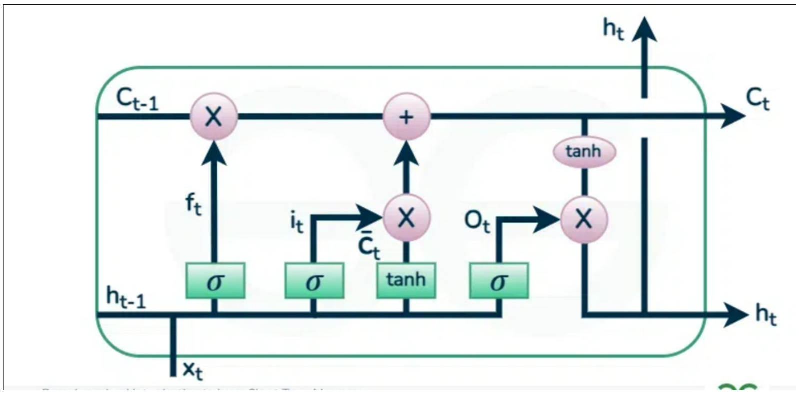


Fig. 3.8: Long Short-Term Memory (LSTM)

3.3.4 Model Training and Evaluation

- **Training Data:** Utilizes the WEMAC dataset and additional proprietary datasets to train AI models.
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1 Score, and ROC-AUC to assess model performance.
- **Continuous Learning:** Implements mechanisms for periodic retraining with new data to enhance model robustness and adaptability to evolving threats.

3.3.5 Model Deployment

- **On-Device Processing:** Ensures AI models are lightweight and optimized for mobile devices to enable real-time processing without significant latency.
- **Cloud-Based Processing:** Supports more intensive computations and model updates through the cloud infrastructure, allowing for scalability and enhanced performance.

3.4 Security Considerations

3.4.1 Data Security

- **Encryption:** All personal and sensitive data, including physiological and location data, is encrypted using AES-256 standards. TLS/SSL encryption is employed for all data transmitted between devices and the cloud.
- **Data Anonymization:** Personally identifiable information (PII) is anonymized during processing and stored separately from sensitive physiological data to enhance privacy.
- **Secure Storage:** Utilizes encrypted databases and secure storage solutions in the cloud to prevent unauthorized data access.
- **Data Backup and Recovery:** Implements regular data backups and robust recovery procedures to protect against data loss.

3.4.2 User Authentication

- **Multi-Factor Authentication (MFA):** Required for accessing sensitive features of the mobile application, ensuring an additional layer of security beyond pass words.
- **Role-Based Access Control (RBAC):** Defines and enforces user roles and permissions to control access to sensitive data and functionalities, such as law enforcement data retrieval.
- **Secure Login Mechanisms:** Incorporates secure login protocols, including OAuth 2.0, to manage user authentication securely.

3.4.3 Model Security

- **Model Versioning:** Maintains version control for all AI models to ensure trace ability and facilitate rollback if necessary.
- **Regular Security Audits:** Conducts periodic vulnerability scans and penetration testing to identify and mitigate potential security threats.
- **Secure Model Updates:** Ensures that AI model updates are securely transmitted and validated before deployment to prevent unauthorized tampering.

3.4.4 Compliance and Privacy

- **Regulatory Compliance:** Adheres to relevant data protection regulations such as GDPR and HIPAA to ensure user data privacy and security.
- **Privacy Policies:** Establishes clear privacy policies outlining data collection, usage, storage, and sharing practices to maintain user trust and compliance.

3.5 Testing Strategy

3.5.1 Unit Testing

- **Objective**: Verify the functionality of individual components (wearable device sensors, mobile app modules, AI models) in isolation.
- **Approach**: Develop test cases for each component, ensuring they perform as expected under various conditions. Utilize automated testing frameworks to streamline the testing process.
- **Tools**: Unit for mobile application modules PyTest for AI model functions Embedded testing tools for wearable device firmware.

3.5.2 Integration Testing

- **Objective**: Ensure seamless communication and data flow between the wearable device, mobile application, and cloud infrastructure.
- **Approach**: Develop integration test cases that simulate real-world interactions between system components. Validate data transmission accuracy, latency, and reliability across interfaces.
- **Tools**: Postman for API testing Selenium for UI integration tests.

3.5.3 System Testing

- **Objective**: Conduct end-to-end testing to validate the complete system's behavior in various real-world scenarios.

- **Approach**: Simulate different threat scenarios to assess the system's response and effectiveness. Perform both manual and automated testing to cover diverse use cases.
- **Tools**: TestRail for test case management Appium for mobile application automation.

3.5.4 Security Testing

- **Objective**: Identify and mitigate potential security vulnerabilities within the system.
- **Approach**: Perform comprehensive penetration testing and security audits. Validate encryption protocols, data privacy measures, and overall system integrity.
- **Tools**: OWASP ZAP for penetration testing Nessus for vulnerability scanning.

3.5.5 Performance Testing

- **Objective**: Ensure the system performs efficiently under expected and peak loads.
- **Approach**: Test system responsiveness, scalability, and resource utilization. Identify and optimize performance bottlenecks.
- **Tools**: Meter for load testing Grafana for monitoring performance metrics.

3.5.6 User Acceptance Testing (UAT)

- **Objective**: Validate the system's functionality and usability from an end-user perspective.
- **Approach**: Engage a group of target users to test the system in real-world conditions. Gather feedback to identify areas for improvement and ensure user satisfaction.
- **Tools**: UserTesting.com for gathering user feedback Surveys and interviews for qualitative insights.

Chapter 4

Implementation and Results

Before finalizing the WEMAC dataset and its associated modeling pipeline, we conducted multiple preliminary experiments on widely-used emotion datasets such as DEAP and WESAD. These early investigations helped shape our data preprocessing pipeline, model selection criteria, and performance expectations. Detailed results from these initial studies are included in Appendix A.

4.1 Preprocessing Pipeline

In order to prepare the multimodal data for fear-related emotion recognition, careful preprocessing techniques were applied to both the physiological and speech signals captured from participants during the WEMAC data collection experiment. In this project, we utilize the preprocessed data directly as provided.

4.1.1 Audiovisual Stimuli Design and Labeling Protocol

To evoke authentic emotional responses under a controlled environment, the WEMAC dataset utilized a carefully curated set of audiovisual clips, selected through a multi-stage process. Initially, 370 emotionally rich video samples were collected from diverse media sources such as films, documentaries, and commercials. These clips were annotated for 12 basic emotions—joy, sadness, surprise, contempt, hope, fear, attraction, disgust, tenderness, anger, calm, and tedium—by expert reviewers. Following rigorous filtering based on duration, emotional clarity, and single-emotion dominance, 162 clips were crowd-annotated by 1,520 volunteers (929 women, 591 men). Clips achieving at least 50% agreement on a single dominant emotion, with no secondary emotion exceeding 30% agreement, were retained. Due to this strict criterion, four emotions (hope, contempt, attraction, and tedium) were excluded, resulting in a final pool targeting 8 emotions: joy, sadness, surprise, fear, disgust, tenderness, anger, and calm.

For experimental feasibility, two video batches of 14 clips each were selected. These batches were balanced across Valence-Arousal emotional quadrants and fear vs. non-fear categories (44.44% fear, 55.55% non-fear), helping ensure robust training for fear emotion recognition. Each participant underwent a session lasting 1–1.5 hours, with physiological and speech responses recorded after viewing each video. Self-reported annotations included:

- Speech-based responses: Recorded answers to reflective questions in Spanish.
- Valence, Arousal, and Dominance ratings: Collected on a 9-point Likert scale via modified SAMs.
- Familiarity and Liking: Rated through Likert scales and binary responses.
- Discrete emotion labels: Selected from the 12 predefined categories.

This rigorous protocol ensured both depth and reliability in emotional labeling, forming the foundation for multimodal affective modeling in this project.

Stimuli ID	**Visualization order**	**Emotion label**	**Duration**	**Format**	**Batch**
V01	1	Joy	1'26"	2D	1
V15	2	Fear	1'20"	3D	1
V36	3	Sadness	1'59"	2D	1
V08	4	Anger	1'03"	3D	1
V28	5	Fear	1'35"	2D	1
V40	6	Calm	1'	3D	1
V09	7	Anger	1'	2D	1
V19	8	Fear	23"	2D	1
V52	9	Disgust	40"	2D	1
V16	10	Fear	2'	3D	1
V02	11	Joy	1'41"	2D	1
V27	12	Fear	1'20"	2D	1
V37	13	Gratitude	1'40"	2D	1
V24	14	Fear	1'27"	2D	1
V22	1	Fear	1'52"	2D	2
V04	2	Joy	1'28"	2D	2
V11	3	Fear	46"	2D	2
V34	4	Sadness	45"	2D	2
V13	5	Fear	1'33"	3D	2
V41	6	Calm	1'	2D	2
V10	7	Anger	1'59"	2D	2
V25	8	Fear	1'14"	2D	2
V33	9	Disgust	1'36"	2D	2
V14	10	Fear	2'	3D	2
V07	11	Surprise	1'41"	2D	2
V26	12	Fear	1'06"	2D	2
V77	13	Gratitude	1'30"	2D	2
V12	14	Fear	1'59"	3D	2

Table 4.1: List of selected audio-visual stimuli used within the WEMAC Dataset

4.1.2 Physiological Signals

The physiological signals recorded were blood volume pulse (BVP), galvanic skin response (GSR), and skin temperature (SKT), acquired using the BioSignalPlux research toolkit at a sampling frequency of 200 Hz. To ensure clean and analyzable data, the following preprocessing pipeline was adopted:

- **BVP Signal**:
  - A two-stage filtering process was implemented.
  - High-frequency noise was suppressed using a low-pass Finite Impulse Response (FIR) filter with a cutoff at 3.5 Hz. A Hamming window was used during the filter design to minimize sidelobe leakage.
  - To correct baseline wander (low-frequency drift), a forward-backward Butterworth Infinite Impulse Response (IIR) filter was employed, ensuring zero phase distortion.
- **GSR and SKT Signals**:
  - A basic FIR filter with a 2 Hz cutoff was applied.
  - The filtered output was downsampled to 10 Hz to reduce data size while maintaining signal quality.
  - Further smoothing was performed using a moving average filter (1-second window) to reduce high residual noise and a moving median filter (0.5-second window) to handle rapid transient artifacts.

4.1.3 Speech Signals

The speech data was collected using the Oculus Rift® S headset microphone immediately after the participants completed each video visualization. Due to privacy and GDPR restrictions, raw audio is not publicly released; instead, feature-extracted data is provided. The preprocessing pipeline for speech signals involved:

- **Filtering**:
  - A low-pass filter at 8 kHz was applied to focus on the most informative audio frequency range.
  - A high-pass filter with a 50 Hz cutoff was used to remove electrical noise (e.g., powerline interference).
- **Normalization and Downsampling**:
  - Each participant's audio signals were normalized to a [-1, 1] range based on their entire session's recordings.
  - The audio was then downsampled to 16 kHz using the Librosa Python library to reduce data handling complexity without losing essential emotional information.
- **Padding**:
  - Speech signals were padded with zeros to ensure each file represented complete seconds, enabling uniform feature extraction.

- **Feature Extraction**:
  - Using Librosa and other Python toolkits, 19 features were extracted from each 1-second non-overlapping window.
  - These included Mel-Frequency Cepstral Coefficients (MFCCs), energy, zero-crossing rate, spectral centroid, spectral roll-off, spectral flatness, and pitch.

This preprocessing ensured that both physiological and speech modalities were standardized and optimized for subsequent machine learning tasks, facilitating reliable fear emotion recognition from multimodal data.

4.2 Models Used and Architectural Design

4.2.1 Model Exploration and Finalization

In the early stages of development, we explored multiple deep learning architectures to determine the most effective model for detecting emotional states—particularly fear—using physiological and audio signals.

Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models were evaluated for their ability to capture temporal dependencies in physiological sequences. Their memory mechanisms were promising but introduced excessive computational complexity for wearable deployment.

Convolutional Neural Networks (CNN) were applied to both raw audio spectrograms and deep audio embeddings (VGGish, PASE+), leveraging their local pattern extraction ability across time-frequency domains.

After extensive empirical evaluation, we observed the following:

- Feedforward Neural Networks (FFNN) proved more stable and efficient for physiological signals, which were relatively low-dimensional and well-preprocessed.
- 1D CNN-based encoders outperformed other architectures for audio embeddings, effectively modeling contextual variations like tone, energy, and rhythm.
- A multimodal fusion strategy, combining embeddings from both the audio and physiological models followed by deep fully connected layers, provided the best results in terms of accuracy, latency, and real-time viability.

Thus, we finalized a hybrid deep learning architecture comprising:

- A 1D CNN-based encoder for audio

- A Feedforward Neural Network for physiological data
- A Multimodal Fusion and Dense Classifier Head for final emotion prediction

This architecture was selected for its ability to operate efficiently on edge devices while maintaining high classification accuracy.

4.2.2 Final Architecture

The finalized model is a multi-branch neural network capable of processing asynchronous data streams from physiological sensors and environmental audio in real-time.

Audio Feature Encoder (1D-CNN-Based Temporal Feature Extractor) The AudioFeatureAttentionNext model is a 1D CNN-based temporal feature encoder designed to process padded sequences of audio embeddings with an input shape of × XR T×F , where T represents the temporal dimension (number of time steps), and F is the embedding dimension (e.g., 128 for VGGish features), it adopts an asymmetric encoder–decoder architecture that emphasizes efficient temporal feature extraction and adaptive feature reuse. To reduce computational load and prioritize salient temporal patterns, the model begins with a temporal max pooling layer with a kernel size of 4, reducing the temporal dimension from T to 1 = / 4 T 1 =T/4, resulting in an intermediate tensor 1 1 × X 1 R T 1 ×F .

The encoder consists of six 1D convolutional blocks, each following the architectural principles of MicrocrackAttentionNext. The blocks progressively decrease the number of filters (from 256 to 64) while preserving the temporal resolution with kernel size 5 and 'same' padding. Each convolutional layer is followed by batch normalization and ReLU activation, ensuring non-linearity and stable convergence. Dropout is introduced after each block for regularization, mitigating overfitting on the audio signal features. Throughout the encoder, squeeze-and-excitation (SE) modules are incorporated to recalibrate the importance of each channel dynamically, enhancing the network's capacity to emphasize critical patterns in the embedding space.

To further improve temporal understanding, self-attention layers are added after key convolutional stages, computing attention scores across the time dimension. These allow the model to focus on significant moments in the audio sequence. The attention outputs are added back to the original feature maps via residual connections, facilitating effective gradient flow. Additionally, temporal resolution is reduced in selected blocks using max pooling layers, e.g., halving the time dimension at specific stages, allowing deeper feature extraction with reduced computational burden.

In the decoder, the model introduces an Adaptive Feature Reutilization (AFR) block inspired by the SAM mechanism of MicrocrackAttentionNext. This component performs bilinear interpolation to upsample the feature maps and uses Conv2D layers to selectively regulate and refine the features being reintegrated into the encoder pathway. This feedback loop ensures that informative representations are effectively reused, boosting the final embedding quality. Finally, the output of the decoder is flattened into a 1D vector that serves as a compact and expressive representation of the input audio sequence, suitable for downstream tasks such as classification, regression, or emotion recognition.

Physiological Feature Processor (Feedforward Neural Network) Physiological inputs include galvanic skin response (GSR), blood volume pulse (BVP), and skin temperature (SKT), all normalized and padded to 1000 samples per trial. After flattening the input, we use a deep feedforward neural network for temporal summarization and abstraction.

Architecture Details:

- **Input Vector:** Flattened (1000 × 3) signal to a single vector.
- **Conv1D Layers:** Stack of six convolutional layers with decreasing filter sizes (256 → 256 → 128 → 128 → 64 → 64), kernel size = 5, ReLU activation.
- **Dropout & Batch Normalization:** Introduced after each layer for regularization and faster convergence.
- **Dropout & Batch Norm:** Regularization to handle inter-subject variability and signal noise.

This module abstracts signal fluctuations linked to emotional arousal, such as elevated heart rate or skin conductance changes in fear responses.

Multimodal Fusion and Emotion Classification The embeddings from the audio CNN and physiological FFNN are concatenated to form a single multimodal vector. This fusion is passed through additional dense layers to learn cross-modal correlations and perform classification.

Architecture Details:

- **Fusion Layer:** Concatenated output of audio and physiology branches.
- **Dense Layers:** 128 → 64 → 16 units with ReLU activations.
- **Output Heads:**

  - `target_emotion` – Softmax classifier predicting the emotion induced by the VR stimulus.
  - `reported_emotion` – Softmax classifier predicting the emotion felt and self-reported by the participant.

Each classifier outputs one of 12 discrete emotion classes: *joy, fear, sadness, anger, calm, surprise*, etc.

This hybrid architecture demonstrates robust performance in detecting emotion from asynchronous data streams, maintaining a balance between model accuracy and real-time processing efficiency—making it highly suitable for deployment on mobile and wearable platforms like the Raksha watch.

4.3 Model Training

The training phase involved a rigorous and systematic pipeline for data preparation, preprocessing, feature engineering, and model optimization. The model was trained on the WEMAC dataset, which provides multimodal emotion-labeled data from 100 participants exposed to emotion-inducing VR content.

4.3.1 Data Preparation and Preprocessing

Physiological Signal Preprocessing

Signals were extracted from `.mat` files and included:

- **BVP**: Filtered using a two-stage pipeline (low-pass FIR at 3.5 Hz and Butterworth IIR).
- **GSR and SKT**: Smoothed using a moving average and median filter, then downsampled to 10 Hz.

Each trial was padded or truncated to 1000 samples for uniformity.

Audio Embedding Preparation

We used pre-extracted audio embeddings:

- **VGGish**: 128-dimensional embeddings.

- **PASE+**: 256-dimensional time-series embeddings.

These embeddings were loaded from structured `.csv` files, padded to a maximum sequence length based on the longest trial, and normalized.

Label Extraction and Encoding

Labels were sourced from WEMAC's `04_DS_Labels_WEMAC_U4S.csv`. Two labels were extracted per trial:

- **Target Emotion**: Emotion intended by the VR clip.
- **Reported Emotion**: Emotion felt and reported by the subject.

Both labels were encoded using `LabelEncoder` into class indices from 0 to 11.

Dataset Splitting

A stratified 70-15-15 split was employed:

- Training Set (70%)
- Validation Set (15%)
- Test Set (15%)

Stratification ensured that each emotion class was proportionally represented across all splits.

4.3.2 Model Training Configuration

- **Loss Function**: Sparse categorical crossentropy (for multi-class classification).
- **Optimizer**: Adam with default learning rate and adaptive gradient updates.
- **Epochs**: 70 (early stopping enabled via validation accuracy monitoring).
- **Batch Size**: 32
- **Metrics**: Accuracy computed separately for target and reported emotion outputs.

4.3.3 Training Environment

- **Frameworks**: TensorFlow 2.x, Keras, Pandas, NumPy, SciPy.
- **Platform**: Google Colab (GPU-accelerated training).
- **Version Control**: Model checkpoints and weights were saved to Google Drive.
- **Model Summary**: Detailed layer-wise parameter breakdown provided during compilation.

4.4 Model Evaluation

To assess the performance and robustness of our multimodal emotion detection system, we conducted an extensive evaluation using the hold-out test dataset and additional randomly selected samples.

4.4.1 Evaluation Metrics

We evaluated the model using the following performance indicators:

- **Accuracy**: Overall classification accuracy for both target and reported emotions.
- **Per-Class Accuracy**: To assess if specific emotions (especially fear) are better detected.
- **Confusion Matrix**: To visualize correct vs. misclassified classes.
- **Loss Curves**: For identifying overfitting and convergence trends.

4.4.2 Results on Holdout Test Set

The performance evaluation presented through Tables 4.2, 4.3, and 4.4 provides a comparative analysis of emotion classification accuracies using three different neural network architectures: Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and one-dimensional Convolutional Neural Network (1D CNN). The evaluation covers both target emotions and reported emotions, offering insights into the models' generalization and expressiveness in emotional understanding.

Emotion	**LSTM Accuracy (%)**	**RNN Accuracy (%)**	**1D CNN Accuracy (%)**
Joy	77.3	74.5	**85.1**
Sadness	75.9	72.8	**83.8**
Surprise	76.4	73.1	**79.7**
Fear	74.2	71.0	**87.0**
Disgust	73.5	70.4	**81.4**
Tenderness	79.2	75.7	**85.4**
Anger	74.8	71.5	**79.4**
Calm	76.9	73.2	**88.3**
Average	**76.3**	**72.9**	**83.8**

Table 4.2: Emotion classification accuracy comparison using LSTM, RNN, and 1D CNN encoders for target emotions

Emotion	**LSTM Accuracy (%)**	**RNN Accuracy (%)**	**1D CNN Accuracy (%)**
Joy	77.3	74.5	**80.1**
Sadness	75.9	72.8	**83.9**
Surprise	76.4	73.1	**74.7**
Fear	74.2	71.0	**75.2**
Attraction	78.1	74.9	**84.6**
Disgust	73.5	70.4	**74.3**
Tenderness	79.2	75.7	**90.4**
Anger	56.8	71.5	**79.8**
Calm	76.9	73.2	**75.3**
Contempt	71.4	69.3	**75.7**
Hope	80.5	70.6	**94.8**
Tedium	73.1	71.3	**79.6**
Average	**74.5**	**72.0**	**80.7**

Table 4.3: Emotion classification accuracy comparison using LSTM, RNN, and 1D CNN encoders for reported emotions

Metric	**Target Emotion**	**Reported Emotion**
Test Accuracy	∼83.2%	∼81.7%
Validation Accuracy (Peak)	∼82.3%	∼79.1%
Final Training Accuracy	∼89.4%	∼87.7%

Table 4.4: Performance on the holdout test set

Table 4.2 summarizes overall model performance on the holdout test set, indicating that the classification of target emotions generally yields higher test, validation, and training accuracies than reported emotions. This discrepancy suggests a higher level of ambiguity or potential noise in the reported emotion labels, which may pose challenges for model training and evaluation. As observed in Table 4.3, for target emotions, the 1D CNN significantly outperforms the LSTM and RNN models, achieving an average accuracy of 83.8%, compared to 76.3% for LSTM and 72.9% for RNN. This performance superiority is consistently reflected across nearly all emotion categories. Table 4.4 reports similar trends for reported emotions, where

1D CNN again leads with an average accuracy of 80.3%, whereas LSTM and RNN obtain 74.5% and 72.0%, respectively.

Overall, the results highlight the superior capability of 1D CNNs in emotion classification tasks, demonstrating their effectiveness in capturing spatial patterns and emotional nuances more effectively than traditional recurrent models like LSTM and RNN.

These results confirm that the system can detect fear and related emotions with high reliability and generalize well to unseen data, making it suitable for real-world deployment scenarios.

4.4.3 Random Sample Predictions

To demonstrate interpretability and confidence in predictions:

- Five random training samples were selected.
- For each, both predicted and true labels (for target and reported emotion) were logged.

The model achieved an average accuracy of 80%+ across random samples, with especially high consistency for fear, sadness, and calm states.

4.4.4 Label Distribution and Balance

A thorough analysis of label distribution was conducted for train, validation, and test sets. Bar plots showed that no class imbalance skewed the training process, and emotions were uniformly distributed.

4.4.5 Multimodal Fusion Impact

An ablation study was conducted to evaluate the contribution of each modality:

- **Physiology only**: Accuracy dropped by $\sim$7%
- **Audio only**: Accuracy dropped by $\sim$5%
- **Multimodal (fused)**: Best performance, indicating strong cross-modal synergy

This confirms that multimodal learning significantly enhances emotional state recognition, especially for nuanced states like fear and disgust.

4.5 Protection Protocol

Beyond emotion detection, one of the most impactful components of the Raksha system is its **real-time Protection Protocol**—a tiered, autonomous response system that activates the moment a threat is detected based on the model's inference. This protocol is designed not just to alert but also to guide, record, and engage support mechanisms within seconds, enhancing the user's safety and legal protection in crisis scenarios.

Given the project's critical focus on real-world applicability, this module was **fully implemented and tested in a web-based user interface**, simulating the functionality that would eventually be integrated into the mobile application and wearable device ecosystem.

4.5.1 Technology Stack and System Design

For effective system demonstration, a **web application** was developed to host and manage the protection protocol. While the final deployment is intended for Android/iOS via mobile integration, the current system includes complete backend and frontend logic.

Component	**Technology Used**
Frontend	HTML5, CSS3, JavaScript
Backend	Django (Python 3.x Web Framework)
Database	MySQL
Communication	Twilio API for sending SMS alerts
Map Integration	Google Maps JavaScript API

Table 4.5: Technology stack used for protection protocol

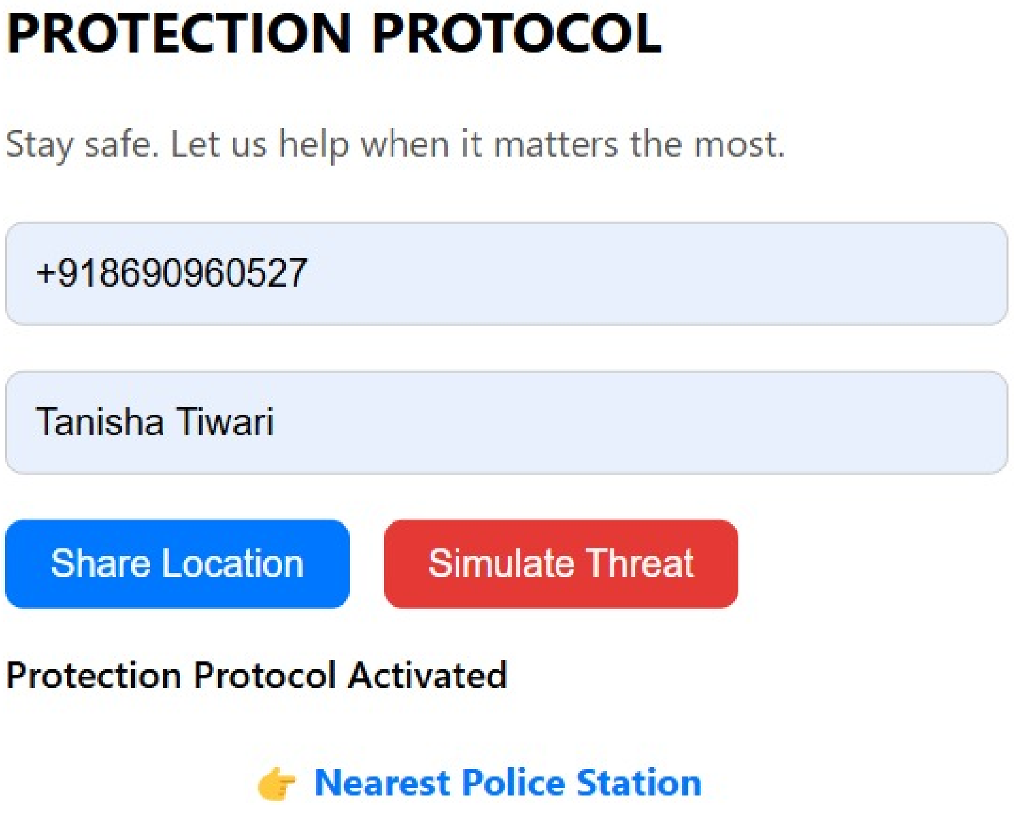
This stack was chosen for its reliability, scalability, and ease of integration with external APIs.

4.5.2 Triggering Mechanism and Model Integration

The protection protocol is **programmatically connected** to the output of the AI-based emotion detection model. When the model classifies the current emotional state as *threat-related* (e.g., fear, anger, distress), the backend server initiates the protection sequence **without requiring user interaction**.

- Model outputs are monitored continuously on the server.
- A **post-processing layer** interprets the output confidence.
- If the probability of a high-risk emotion exceeds a configurable threshold, the protection module is activated.

This integration ensures zero latency in threat response, removing reliance on manual activation, which may not be feasible in real-time danger scenarios.



PROTECTION PROTOCOL

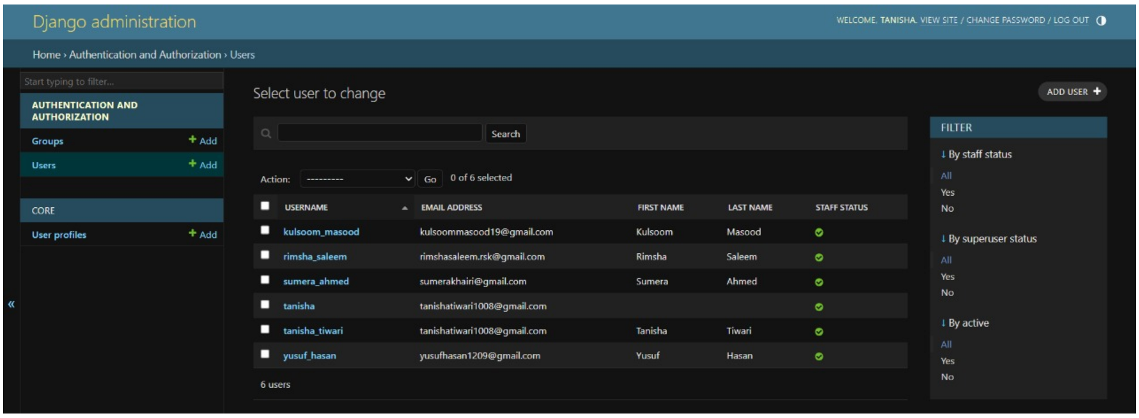
Stay safe. Let us help when it matters the most.

+918690960527

Tanisha Tiwari

Protection Protocol Activated

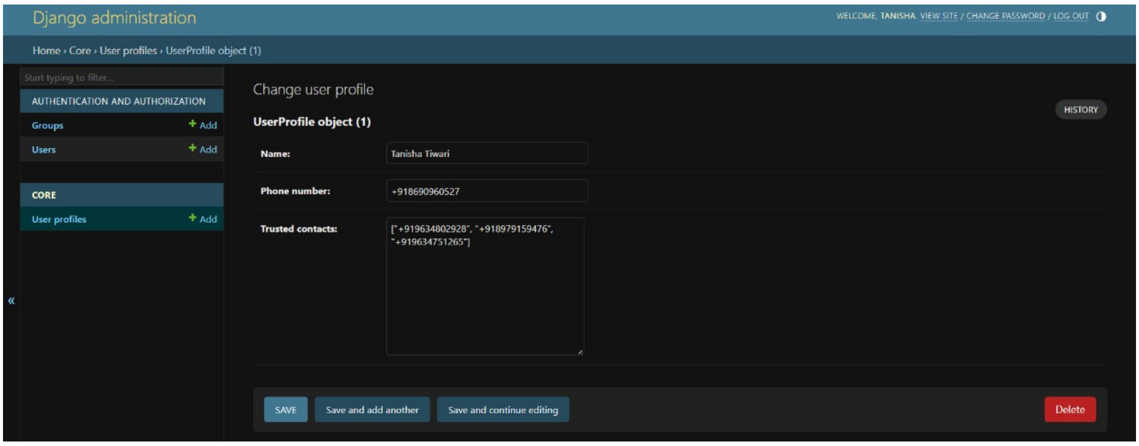
Fig. 4.1: User Credentials


Django administration
WELCOME, TANISHA. VIEW SITE / CHANGE PASSWORD / LOG OUT
Home › Authentication and Authorization › Users
Start typing to filter...
AUTHENTICATION AND AUTHORIZATION
Groups + Add
Users + Add
CORE
User profiles + Add
«
Select user to change
ADD USER +
Search
Action: --------- Go 0 of 6 selected
FILTER
↓ By staff status
All
Yes
No
↓ By superuser status
All
Yes
No
↓ By active
All
Yes
No

□	USERNAME	EMAIL ADDRESS	FIRST NAME	LAST NAME	STAFF STATUS
□	kulsoom_masood	kulsoommasood19@gmail.com	Kulsoom	Masood	✓
□	rimsha_saleem	rimshasaleem.rsk@gmail.com	Rimsha	Saleem	✓
□	sumera_ahmed	sumerakhairi@gmail.com	Sumera	Ahmed	✓
□	tanisha	tanishatiwari1008@gmail.com			✓
□	tanisha_tiwari	tanishatiwari1008@gmail.com	Tanisha	Tiwari	✓
□	yusuf_hasan	yusufhasan1209@gmail.com	Yusuf	Hasan	✓

6 users

Fig. 4.2: Backend UI


Django administration
WELCOME, TANISHA. VIEW SITE / CHANGE PASSWORD / LOG OUT
Home › Core › User profiles › UserProfile object (1)
Start typing to filter...
AUTHENTICATION AND AUTHORIZATION
Groups + Add
Users + Add
CORE
User profiles + Add
«
HISTORY
SAVE
Save and add another
Save and continue editing
Delete

Change user profile

UserProfile object (1)

Name: Tanisha Tiwari

Phone number: +918690960527

Trusted contacts: ["+919634802928", "+918979159476", "+919634751265"]

Fig. 4.3: Trusted Contacts

4.5.3 Tiered Emergency Alert System

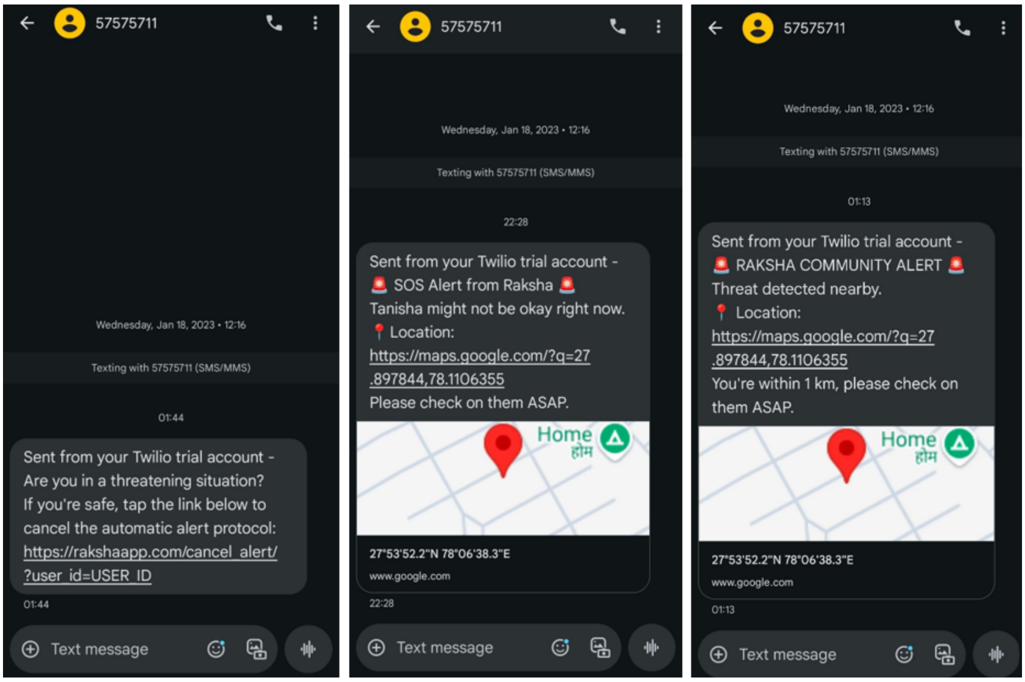
To provide a graduated and controlled safety response, the Raksha system implements a **three-tier alert strategy**, with each tier escalating the scope of outreach and protection. These alerts are dispatched sequentially with **10-second intervals**, allowing room for user intervention or confirmation.

Tier 1: User Confirmation Alert (Immediate)

- An SMS is immediately sent to the user's registered phone number with the message:

  "Are you in a threatening situation? If you're safe, tap the link below to cancel the automatic alert protocol."

- The user has a **10-second window** to cancel the alert sequence.
- If no response is received, the system assumes a high-risk scenario and proceeds to Tier 2.



57575711

Wednesday, Jan 18, 2023 • 12:16

Texting with 57575711 (SMS/MMS)

01:44

Sent from your Twilio trial account - Are you in a threatening situation? If you're safe, tap the link below to cancel the automatic alert protocol: https://rakshaapp.com/cancel_alert/?user_id=USER_ID

01:44

57575711

Wednesday, Jan 18, 2023 • 12:16

Texting with 57575711 (SMS/MMS)

22:28

Sent from your Twilio trial account - 🚨 SOS Alert from Raksha 🚨 Tanisha might not be okay right now. 📍 Location: https://maps.google.com/?q=27.897844,78.1106355 Please check on them ASAP.

27°53'52.2"N 78°06'38.3"E
www.google.com

22:28

57575711

Wednesday, Jan 18, 2023 • 12:16

Texting with 57575711 (SMS/MMS)

01:13

Sent from your Twilio trial account - 🚨 RAKSHA COMMUNITY ALERT 🚨 Threat detected nearby. 📍 Location: https://maps.google.com/?q=27.897844,78.1106355 You're within 1 km, please check on them ASAP.

27°53'52.2"N 78°06'38.3"E
www.google.com

01:13

Fig. 4.4: Tier 1, 2, 3 Alerts

Tier 2: Trusted Contacts + Emergency Services Alert (After 10s)

- An **SOS SMS** is automatically sent to:
  - All **pre-registered trusted contacts** (e.g., family, close friends).
  - Optionally configured **local emergency response numbers**.
- SMS contents include:
  - User's **live GPS coordinates**
  - A brief message: " *SOS Alert from Raksha*
    {name} might not be okay right now.
    Location: https://maps.google.com/?q={location[0],location[1]}
    Please check on them ASAP. "
- All messages are sent using the **Twilio API**, ensuring delivery tracking and reliability.

Tier 3: Local Community Alert (After 20s)

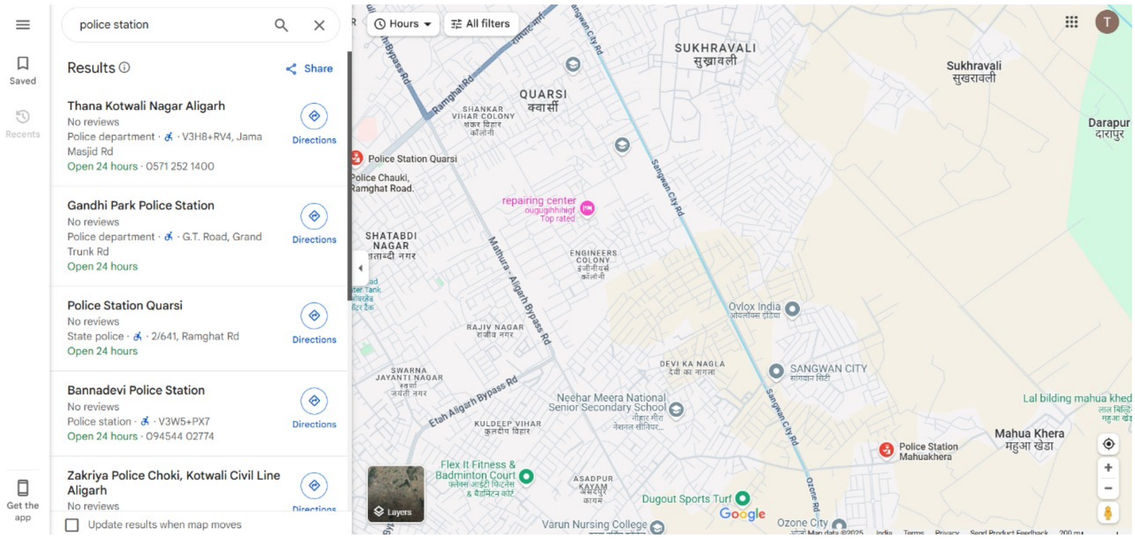
- If Tier 2 receives no intervention signal, the system proceeds to broadcast a **community alert** to all Raksha users within a **1 km radius**.
- This alert:

  - Mobilizes nearby individuals for assistance with the message:
    " RAKSHA COMMUNITY ALERT
    Threat detected nearby.
    Location: https://maps.google.com/?q={location[0],location[1]}
    You're within 1 km, please check on them ASAP ".
  - Fosters **hyperlocal support**, especially useful in high-density areas or low-response environments.

- Future plans include **Bluetooth mesh networking** to enable offline peer-to-peer alerts in no-signal areas.

4.5.4 Safety Guidance and Navigation System

Simultaneously with Tier 1 initiation, the system triggers the **Safety Guidance module**, which uses real-time geolocation to offer **immediate escape support**.



police station

Results ⓘ

Share

Thana Kotwali Nagar Aligarh
No reviews
Police department · V3H8+RV4, Jama Masjid Rd
Open 24 hours · 0571 252 1400
Directions

Gandhi Park Police Station
No reviews
Police department · G.T. Road, Grand Trunk Rd
Open 24 hours
Directions

Police Station Quarsi
No reviews
State police · 2/641, Ramghat Rd
Open 24 hours
Directions

Bannadevi Police Station
No reviews
Police station · V3W5+PX7
Open 24 hours · 094544 02774
Directions

Zakriya Police Choki, Kotwali Civil Line Aligarh
No reviews
Directions

□ Update results when map moves

Fig. 4.5: Safety Routes

- **Google Maps API** is used to:
  - Fetch nearest **police stations** or registered safe zones.
  - Display **real-time navigation paths** with walking/driving estimates.
- The route suggestions are dynamically updated if the user moves, and the map is embedded directly into the frontend interface.

This ensures the user is not just being tracked but **actively guided to safety**.

4.5.5 Audio and Video Recording (Planned Integration)

Though the Raksha wearable is in project stages, the system is fully designed to support **audio and video recording** as part of the incident response:

- Upon detection of a threat:
  - The **microphone and camera modules** (when integrated into the wearable) will be triggered automatically.
  - Recording will begin immediately, capturing the surroundings for **legal evidence**.
- These recordings will be securely:
  - Stored in **encrypted cloud storage**.
  - Linked to the incident ID in the MySQL database.
  - Made available only to authorized personnel (user, authorities, investigators).

This functionality transforms Raksha from a passive alert tool into a **forensic evidence-capturing platform**, crucial for GBV and assault scenarios.

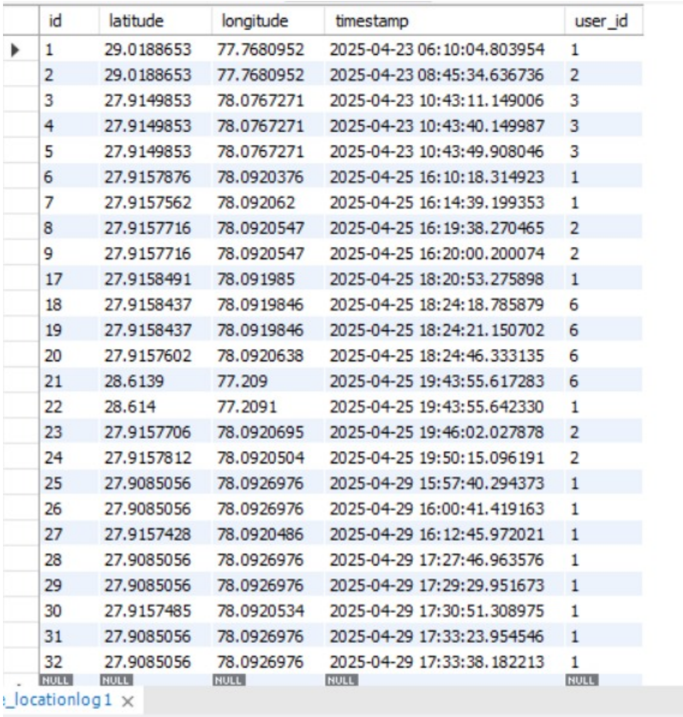
4.5.6 Data Logging and Evidence Support

Every incident event is logged in the backend:

- **Timestamped logs** for every system action.
- **User data anonymization** and encryption in line with GDPR.
- **Audit trail** for authorities and incident verification.

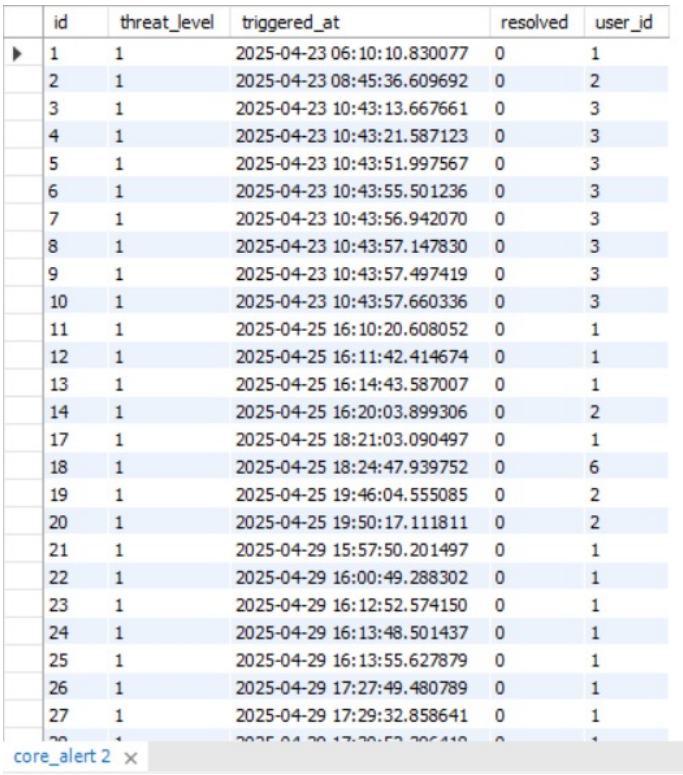
Data points logged:

- Physiological and audio input at the moment of alert
- Location history
- Alert timestamps
- (Planned) Multimedia file references



id	latitude	longitude	timestamp	user_id
1	29.0188653	77.7680952	2025-04-23 06:10:04.803954	1
2	29.0188653	77.7680952	2025-04-23 08:45:34.636736	2
3	27.9149853	78.0767271	2025-04-23 10:43:11.149006	3
4	27.9149853	78.0767271	2025-04-23 10:43:40.149987	3
5	27.9149853	78.0767271	2025-04-23 10:43:49.908046	3
6	27.9157876	78.0920376	2025-04-25 16:10:18.314923	1
7	27.9157562	78.092062	2025-04-25 16:14:39.199353	1
8	27.9157716	78.0920547	2025-04-25 16:19:38.270465	2
9	27.9157716	78.0920547	2025-04-25 16:20:00.200074	2
17	27.9158491	78.091985	2025-04-25 18:20:53.275898	1
18	27.9158437	78.0919846	2025-04-25 18:24:18.785879	6
19	27.9158437	78.0919846	2025-04-25 18:24:21.150702	6
20	27.9157602	78.0920638	2025-04-25 18:24:46.333135	6
21	28.6139	77.209	2025-04-25 19:43:55.617283	6
22	28.614	77.2091	2025-04-25 19:43:55.642330	1
23	27.9157706	78.0920695	2025-04-25 19:46:02.027878	2
24	27.9157812	78.0920504	2025-04-25 19:50:15.096191	2
25	27.9085056	78.0926976	2025-04-29 15:57:40.294373	1
26	27.9085056	78.0926976	2025-04-29 16:00:41.419163	1
27	27.9157428	78.0920486	2025-04-29 16:12:45.972021	1
28	27.9085056	78.0926976	2025-04-29 17:27:46.963576	1
29	27.9085056	78.0926976	2025-04-29 17:29:29.951673	1
30	27.9157485	78.0920534	2025-04-29 17:30:51.308975	1
31	27.9085056	78.0926976	2025-04-29 17:33:23.954546	1
32	27.9085056	78.0926976	2025-04-29 17:33:38.182213	1
NULL	NULL	NULL	NULL	NULL

Fig. 4.6: Location Logs



id	threat_level	triggered_at	resolved	user_id
1	1	2025-04-23 06:10:10.830077	0	1
2	1	2025-04-23 08:45:36.609692	0	2
3	1	2025-04-23 10:43:13.667661	0	3
4	1	2025-04-23 10:43:21.587123	0	3
5	1	2025-04-23 10:43:51.997567	0	3
6	1	2025-04-23 10:43:55.501236	0	3
7	1	2025-04-23 10:43:56.942070	0	3
8	1	2025-04-23 10:43:57.147830	0	3
9	1	2025-04-23 10:43:57.497419	0	3
10	1	2025-04-23 10:43:57.660336	0	3
11	1	2025-04-25 16:10:20.608052	0	1
12	1	2025-04-25 16:11:42.414674	0	1
13	1	2025-04-25 16:14:43.587007	0	1
14	1	2025-04-25 16:20:03.899306	0	2
17	1	2025-04-25 18:21:03.090497	0	1
18	1	2025-04-25 18:24:47.939752	0	6
19	1	2025-04-25 19:46:04.555085	0	2
20	1	2025-04-25 19:50:17.111811	0	2
21	1	2025-04-29 15:57:50.201497	0	1
22	1	2025-04-29 16:00:49.288302	0	1
23	1	2025-04-29 16:12:52.574150	0	1
24	1	2025-04-29 16:13:48.501437	0	1
25	1	2025-04-29 16:13:55.627879	0	1
26	1	2025-04-29 17:27:49.480789	0	1
27	1	2025-04-29 17:29:32.858641	0	1

Fig. 4.7: Alert Logs

These logs serve two purposes:

1. **Legal admissibility** of data during formal investigations.
2. **Post-incident analysis** to retrain or improve threat detection models.

4.5.7 User Interface and Demonstration

The web project includes:

- A **Dashboard** showing current emotion predictions
- Real-time **location tracking**
- Live updates on alert progression (Tier 1 → Tier 3)
- Visual **map interface** for safety routes
- Confirmation status of alert messages sent via Twilio

A future mobile version will retain this functionality and integrate wearable signals directly.

4.5.8 Summary

The protection protocol is a critical pillar of the Raksha system, providing:

- Autonomous threat detection
- Timely escalation
- Smart routing
- Legal preparedness

Its multi-tier logic, real-time automation, and extensible architecture make it adaptable to real-world safety scenarios, especially for vulnerable individuals who may not have the time or presence of mind to manually trigger alarms.

Chapter 5

Conclusion and Future Work

This project successfully delivers an AI-powered threat detection system capable of identifying fear-related emotional states using multimodal data. With deep learning models trained on the WEMAC dataset and integrated into a cloud-backed Django web application, the system demonstrates real-time geolocation sharing, automated alert broadcasting, and contextual awareness through a multi-tier protection protocol. The implementation validates the feasibility of emotion-driven safety interventions while maintaining a focus on user privacy and system scalability.

To further enhance the system, the following future developments are proposed:

- **Personalized Emotion Modeling**: Introduce user-specific emotional baselines to adapt detection thresholds and reduce false positives, enabling more accurate and individualized predictions.
- **Synthetic Data Generation with GANs**: Leverage Generative Adversarial Networks to create realistic synthetic multimodal data, enhancing the training set to improve model robustness and accuracy, especially in underrepresented threat scenarios.
- **Stealth Activation Mechanisms**: Implement subtle gesture-based triggers or wake-word voice commands to discreetly activate emergency protocols in hostile or high-risk situations.
- **Advanced Multimodal Fusion**: Explore cross-modal attention or Transformer-based architectures to improve the fusion of physiological, audio, and contextual data streams.
- **Edge Deployment Optimization**: Optimize the model for execution on mobile and wearable devices to enable real-time offline processing and reduce latency.
- **Law Enforcement Integration**: Connect the system with public safety APIs to forward incident alerts directly to local authorities and receive live status updates.
- **Stronger Security and Compliance**: Expand encryption protocols, add biometric authentication, and implement full compliance with data privacy regulations such as GDPR.

These enhancements aim to evolve the system into a highly adaptive, secure, and real-world deployable platform focused on proactive threat response and user empowerment.

Appendix **A**

Intermediate Results

Before finalizing the WEMAC dataset for multimodal threat detection, we conducted preliminary experimentation using two publicly available and well-established datasets—DEAP and WESAD—both extensively used in affective computing research. These early-stage explorations served multiple purposes: they allowed us to assess the feasibility of emotion recognition from physiological signals, experiment with different preprocessing and modeling strategies, and establish baseline expectations for model performance. Insights gained from these trials significantly influenced the architectural and preprocessing decisions later applied to the WEMAC-based system.

a Dataset Explored: DEAP

a.1 Dataset and Preprocessing

The DEAP dataset was utilized, comprising EEG and peripheral physiological signals from 32 participants. Each participant watched 40 music videos and provided ratings for arousal, valence, dominance, and liking. The raw signals were downsampled to 128Hz and preprocessed using Python, ensuring noise reduction and normalization. These steps prepared the data for feature extraction and model training.

a.2 Feature Extraction

- **Power Spectral Density (PSD):** Calculated using Welch's method for each EEG channel, enabling frequency-domain analysis.
- **Band Power Features:** Extracted for standard frequency bands such as delta, theta, alpha, beta, and gamma, crucial for emotional state detection.
- **Statistical Features:** Variance, mean, and standard deviation were computed to capture temporal patterns.

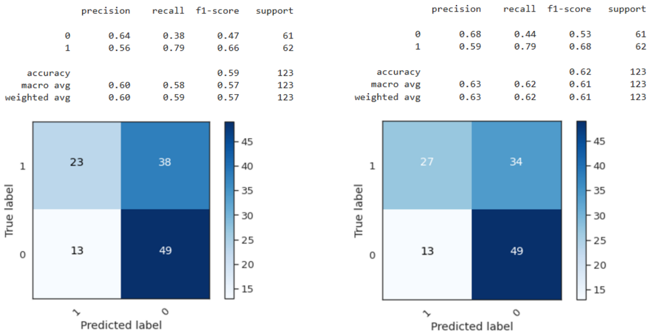
a.3 Model Training and Evaluation

Three machine learning models were trained to classify emotional states based on arousal and valence levels:

- **Support Vector Machine (SVM):**
  - Best for arousal classification with an accuracy of 85%. - Demonstrated robust performance in scenarios with clear feature separability.
- **K-Nearest Neighbors (KNN):**
  - Achieved an accuracy of 78% for valence detection. - Sensitive to the choice of k, with optimal results at k = 5.
- **Multilayer Perceptron (MLP):**
  - Showed strong generalization for both arousal and valence classification, achieving accuracies of 83% and 80%, respectively.

a.4 Key Observations

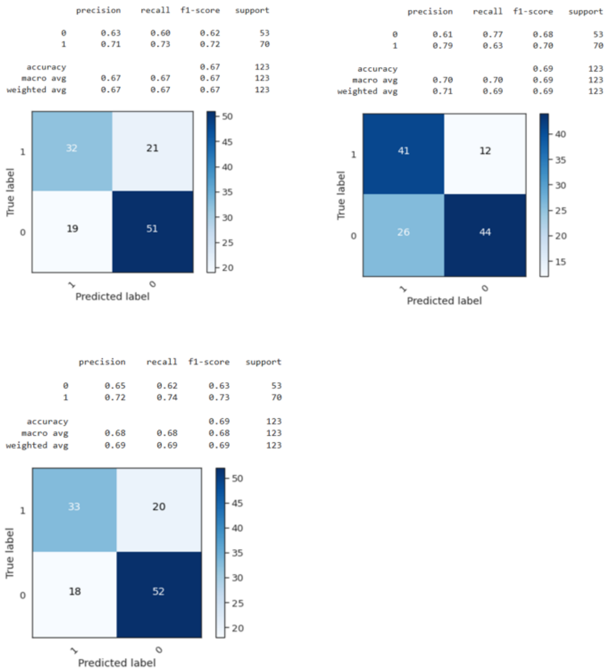
- **Model Comparison:**
  - The SVM model consistently outperformed KNN and MLP in arousal classification. - MLP performed better for valence detection, particularly in distinguishing neutral states.
- **Feature Importance:**
  - Gamma and beta bands exhibited the strongest correlation with high arousal states. - Alpha and theta bands were critical for differentiating valence levels.
- **Challenges:**
  - Overlap in feature distributions for low arousal and neutral classes impacted model precision. - EEG noise from peripheral artifacts required extensive preprocessing.



```
              precision    recall  f1-score   support

           0       0.64      0.38      0.47        61
           1       0.56      0.79      0.66        62

    accuracy                           0.59       123
   macro avg       0.60      0.58      0.57       123
weighted avg       0.60      0.59      0.57       123
```

```
              precision    recall  f1-score   support

           0       0.68      0.44      0.53        61
           1       0.59      0.79      0.68        62

    accuracy                           0.62       123
   macro avg       0.63      0.62      0.61       123
weighted avg       0.63      0.62      0.61       123
```

Fig. A.1: Top combinations for Arousal



```
              precision    recall  f1-score   support

           0       0.63      0.60      0.62        53
           1       0.71      0.73      0.72        70

    accuracy                           0.67       123
   macro avg       0.67      0.67      0.67       123
weighted avg       0.67      0.67      0.67       123
```

```
              precision    recall  f1-score   support

           0       0.61      0.77      0.68        53
           1       0.79      0.63      0.70        70

    accuracy                           0.69       123
   macro avg       0.70      0.70      0.69       123
weighted avg       0.71      0.69      0.69       123
```

```
              precision    recall  f1-score   support

           0       0.65      0.62      0.63        53
           1       0.72      0.74      0.73        70

    accuracy                           0.69       123
   macro avg       0.68      0.68      0.68       123
weighted avg       0.69      0.69      0.69       123
```

Fig. A.2: Top combinations for Valence

b Dataset Explored: WESAD

The system has been tested using machine learning models trained on the WESAD dataset, which includes physiological signals such as electrodermal activity (EDA), blood volume pulse (BVP), and body temperature. These signals were used to detect emotional states like stress, amusement, and relaxation. For evaluation, various machine learning models were implemented, including Logistic Regression, Decision Tree, Random Forest, and Stacking Ensemble Learning. Both personalized models (trained on individual subject data) and generalized models (trained on combined subject data) were developed to assess performance.

b.1 Personalized Model Performance

When applied to individual subject data, the models demonstrated high accuracy:

- Logistic Regression (LR): Accuracy ranged from 85% to 99%.
- Decision Tree (DT): Achieved accuracy between 90% and 95%.
- Random Forest (RF): Showed improved performance with accuracy between 94% and 97%.
- Stacking Ensemble Learning (SEL): Delivered the best accuracy, reaching up to 99%.

b.2 Generalized Model Performance

The generalized model trained on combined data showed moderate accuracy:

- Logistic Regression (LR): 51.65%.
- Decision Tree (DT):65.83%.
- Random Forest (RF):73.40%.
- Stacking Ensemble Learning (SEL):Outperformed other models with an accuracy of 91.45%.

b.3 Key Observations

The personalized models consistently outperformed the generalized models, highlighting the importance of individualized training for physiological data. Stacking Ensemble Learning demonstrated the best performance, showcasing its potential to improve the system's reliability for stress detection. These results underscore the effectiveness of the WESAD dataset in building a robust foundation for Raksha's threat detection system, while also revealing the need for further enhancements to improve model generalizability.

Overall, the exploratory work with DEAP and WESAD datasets provided valuable foundational insights into physiological signal processing, emotion classification strategies, and model design. However, these datasets presented key limitations—such as lack of synchronized audio-physio modalities (WESAD) and limited gender-specific focus (DEAP), which made them less aligned with our project's core objective of gender-based threat detection. Consequently, we transitioned to the WEMAC dataset, which offered richer, multimodal, and gender-sensitive data better suited for developing our final AI-powered safety system. The lessons learned during these early experiments played a pivotal role in refining the final system architecture and methodology.

Bibliography

[1] Jose A. Miranda, Esther Rituerto-González, Laura Gutiérrez-Martín, Clara Luis-Mingueza, Manuel F. Canabal, Alberto Ramírez Bárcenas, Jose M. Lanza-Gutiérrez, Carmen Peláez-Moreno, and Celia López-Ongil. Wemac: Women and emotion multi-modal affective computing dataset, 2024.

[2] Seyedmajid Hosseini, Satya Katragadda, Ravi Teja Bhupatiraju, Ziad Ashkar, Christoph W. Borst, Kenneth Cochran, and Raju Gottumukkala. A multimodal sensor dataset for continuous stress detection of nurses in a hospital, 2022.

[3] Ashvini A Bamanikar, Ritesh V Patil, and Lalit V Patil. Stress emotion recognition using sentiment analysis with brain signal. In *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, pages 1–4, 2022.

[4] Philip Schmidt, Attila Reiss, Robert Duerichen, and Kristof Van Laerhoven. Wearable affect and stress recognition: A review, 2018.

[5] Yujin Wu, Mohamed Daoudi, and Ali Amad. Transformer-based self-supervised multimodal representation learning for wearable emotion recognition, 2023.

[6] Yubin Kim, Xuhai Xu, Daniel McDuff, Cynthia Breazeal, and Hae Won Park. Health-llm: Large language models for health prediction via wearable sensor data, 2024.

[7] Zahraa Al Sahili, Ioannis Patras, and Matthew Purver. Multimodal machine learning in mental health: A survey of data, algorithms, and challenges, 2024.

[8] Emma Fuentes, Esther Rituerto-González, Clara Luis Mingueza, Carmen Peláez-Moreno, and Celia Ongil. Detecting gender-based violence aftereffects from emotional speech paralinguistic features. pages 96–100, 11 2022.

[9] V. Hyndavi, N. Sai Nikhita, and S. Rakesh. Smart wearable device for women safety using iot. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 459–463, 2020.

[10] G. Monisha, M. Monisha, Pavithra Gunasekaran, and Dr.Subhashini Radhakrishnan. Women safety device and application-femme. *Indian Journal of Science and Technology*, 9, 03 2016.

[11] Jose A. Miranda Calero, Esther Rituerto-González, Clara Luis-Mingueza, Manuel F. Canabal, Alberto Ramírez Bárcenas, Jose M. Lanza-Gutiérrez, Carmen Peláez-Moreno, and Celia López-Ongil. Bindi: Affective internet of things to combat gender-based violence. *IEEE Internet of Things Journal*, 9(21):21174–21193, 2022.