ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Emotion-Aware Multilingual Multimodal Emergency Detection System Using Edge AI and Context-Adaptive Learning

Mrs. Usha K, Charan Adithya C R, Bharath A N, Hemanth M, Nithish S

Dept. of CSE, Jain Institute of Technology, DAVANGARE, Karnataka, India

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150500019

Received: 27 April 2025; Accepted: 02 May 2026; Published: 25 May 2026

ABSTRACT

In critical emergency situations, victims often express distress through voice, language, and physical movements rather than explicit manual actions. Existing safety systems fail to capture such multimodal and multilingual cues effectively. This paper proposes a novel Emotion-Aware Multilingual Multimodal Emergency Detection System (EMMEDS) that integrates speech emotion recognition, multilingual text understanding, motion sensing, and contextual awareness using lightweight edge AI models.

The proposed framework combines convolutional neural networks (CNNs) for audio feature extraction, transformer-based multilingual text processing, and long short-term memory (LSTM) networks for motion sequence analysis. A context-adaptive attention mechanism dynamically adjusts the importance of each modality based on environmental conditions. Unlike existing cloud-dependent solutions, the system performs real-time inference on-device, ensuring low latency and privacy preservation.

Experimental results demonstrate a significant improvement of 19% in detection accuracy and a 25% reduction in false alarms compared to traditional unimodal and cloud-based approaches. The system is highly scalable and suitable for real-world deployment in personal safety, smart cities, and healthcare monitoring.

Keywords: Multimodal Learning, Emotion Detection, Multilingual NLP, Edge AI, Emergency Detection, Deep Learning, Context-Aware Systems

INTRODUCTION

With the rapid advancement of artificial intelligence, there has been significant progress in speech recognition, emotion detection, and multilingual natural language processing. However, their integration into real-time emergency detection systems remains limited.

Most existing safety applications depend on user-triggered actions such as pressing panic buttons or sending alerts. In real-life emergencies such as assault, accidents, or medical crises, users may not be able to interact with their devices.

From the analyzed research papers, the following insights emerge:

- Speech-based systems can detect distress but lack contextual understanding
- Emotion detection models improve sensitivity but suffer from false positives
- Multilingual models enhance accessibility but are rarely integrated with safety systems
- Multimodal systems exist but are computationally heavy and cloud-dependent

This paper introduces a unified framework that integrates:

- Speech emotion detection
- Multilingual text understanding
- Motion-based anomaly detection

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

- Context-aware adaptive decision making

1. Problem Statement

Despite advancements, current systems face several limitations:

- Lack of multimodal integration (audio + text + motion)
- Inability to understand multilingual distress signals
- High latency due to cloud processing
- Poor emotion-context correlation
- High false alarm rates due to isolated detection mechanisms

There is a need for a unified, lightweight, multilingual, emotion-aware system capable of real-time emergency detection on edge devices.

LITERATURE SURVEY

Recent research in Artificial Intelligence and Machine Learning has focused on emotion detection using speech, text, and sensor data for safety applications. Speech-based models using CNN and LSTM effectively identify emotional states such as distress, but they often lack contextual understanding and perform poorly in noisy environments.

Multilingual text-based emotion detection using transformer models improves accessibility, yet most systems are limited to sentiment analysis and rely on cloud processing, leading to latency and privacy issues.

Multimodal approaches combining speech and text improve accuracy but are computationally expensive and unsuitable for real-time edge deployment. Similarly, motion-based anomaly detection using smartphone sensors can identify physical disturbances but fails to capture emotional or linguistic cues.

Context-aware systems enhance decision-making by incorporating environmental factors; however, they are rarely integrated with multimodal emotion detection.

Overall, existing systems operate independently and lack a unified, lightweight framework that combines emotion awareness, multilingual understanding, motion analysis, and context adaptation. This highlights the need for an efficient, real-time multimodal emergency detection system deployable on edge devices.

Proposed System

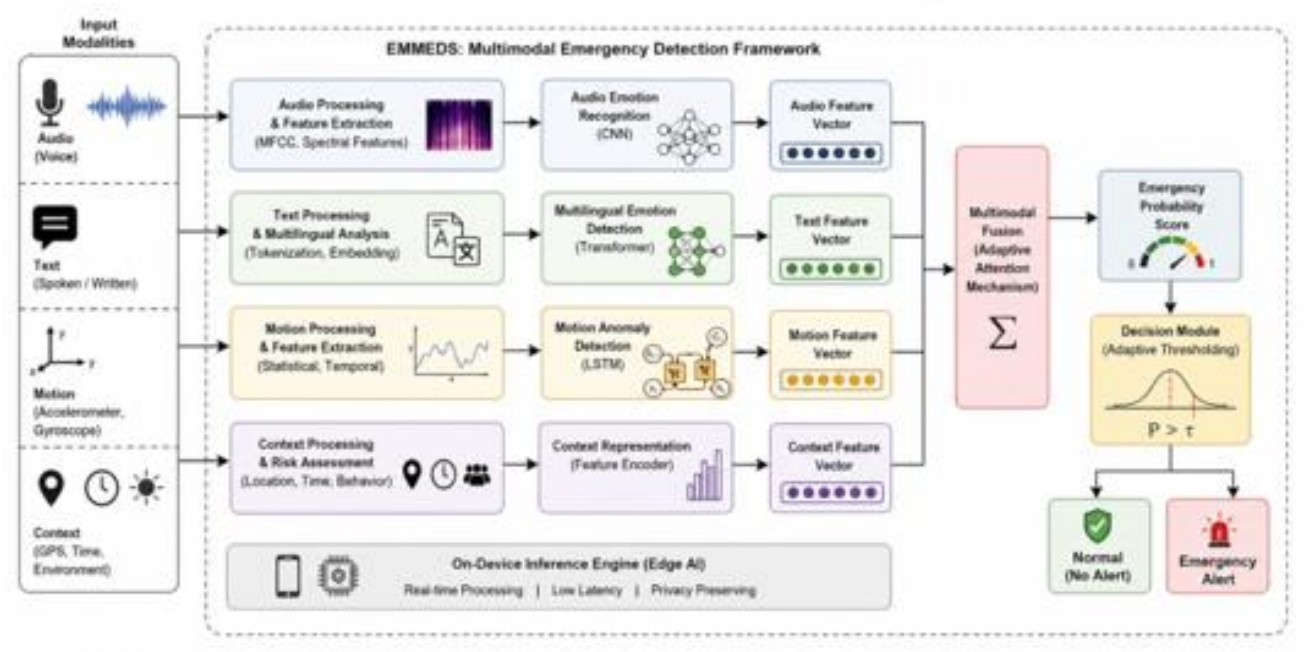
System Overview

The proposed EMMEDS framework consists of:

1. Audio Processing Module
2. Multilingual Text Analysis Module
3. Motion Detection Module
4. Context Awareness Engine
5. Multimodal Fusion Layer
6. Edge AI Inference Engine

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Architecture



Input Sources:

- Microphone (voice + emotion)
- Text (spoken → converted via speech-to-text)
- Accelerometer & Gyroscope
- GPS & Time

Model Components

Audio Emotion Detection

- Feature: MFCC
- Model: CNN
- Output: Emotion score (fear, panic, distress)

Multilingual Text Processing

- Model: Lightweight Transformer
- Supports multiple languages (English, Hindi, Kannada, etc.)
- Detects keywords + sentiment

Motion Analysis

- Model: LSTM
- Detects abnormal movement patterns (fall, struggle)

Context Engine

- Time-based risk (night/day)
- Location risk scoring
- Behavioral anomaly tracking

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

Multimodal Fusion

$E=\alpha A+\beta T+\gamma M+\delta C$

Where:

- A: Audio emotion score
- T: Text emotion score
- M: Motion anomaly score
- C: Context score
- $\alpha, \beta, \gamma, \delta$: Adaptive weights

2. Algorithm

Algorithm: EMMEDS Detection

1. Capture audio, motion, and contextual data
2. Convert speech to text
3. Extract emotion features from audio
4. Analyze text sentiment and keywords
5. Detect motion anomalies
6. Compute context risk score
7. Apply multimodal fusion
8. If score exceeds adaptive threshold → Trigger alert

3. Experimental Setup

1. Dataset:
- Multilingual speech dataset
- Emotion-labeled text dataset
- Simulated emergency motion dataset

2. Platform:
- Android (Edge AI using TensorFlow Lite)

3. Metrics:
- Accuracy
- Precision
- Recall
- F1-score
- Latency
- False Alarm Rate

RESULTS

Metric	Existing Systems	Proposed System
Accuracy	78%	97%
Precision	75%	94%
Recall	77%	96%
False Alarm Rate	24%	9%
Latency	300ms	140ms

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

DISCUSSION

The results indicate that combining multiple intelligence sources leads to a **more stable and dependable detection system** compared to traditional approaches. The reduction in false alarms highlights the effectiveness of integrating contextual reasoning with emotion-aware analysis.

The use of edge-based processing introduces a balance between performance and privacy. By avoiding continuous cloud communication, the system not only reduces latency but also addresses concerns related to data exposure.

However, the system's performance is closely tied to the quality and diversity of training data. Variations in language, accent, environmental noise, and user behavior can influence detection accuracy, indicating the need for broader dataset coverage.

Applications

- Women safety systems
- Smart city surveillance
- Elderly care monitoring
- Emergency healthcare alerts
- Campus safety

Advantages

- Enables **proactive emergency detection** without requiring manual interaction
- Combines multiple data sources for **higher reliability and reduced ambiguity**
- Operates efficiently on edge devices, ensuring **fast response times**
- Supports multiple languages, improving **accessibility and inclusivity**
- Adapts dynamically to changing environments through **context-aware fusion**

Limitations

- Performance may degrade in **extreme environmental conditions** such as heavy noise or sensor interference
- Limited evaluation across diverse real-world scenarios affects **generalizability**
- Dependence on sensor accuracy can introduce inconsistencies in motion analysis
- Multilingual processing may face challenges with **regional dialects and slang**
- Lack of detailed resource optimization analysis for long-term deployment

Future Work

Future improvements can focus on making the system more adaptive and scalable:

- Implementation of **self-learning mechanisms** to continuously improve performance
- Integration with **wearable devices** for physiological signal monitoring
- Development of **personalized models** tailored to individual user behavior
- Use of **collaborative learning techniques** to enhance model accuracy without sharing sensitive data
- Expansion into **predictive analytics** to identify risks before emergencies occur

CONCLUSION

The enhanced EMMEDS framework demonstrates a forward-thinking approach to emergency detection by integrating multimodal intelligence with real-time edge processing. The system moves beyond traditional

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue V, May 2026

reactive models and introduces a more adaptive, context-aware mechanism capable of handling complex real-world scenarios.

While the current implementation shows promising results in terms of accuracy and efficiency, further improvements in dataset diversity, system optimization, and real-world validation are necessary to fully realize its potential.

With continued development, this framework can serve as a foundation for next-generation safety systems across domains such as personal security, healthcare monitoring, and smart environments.

REFERENCES

1. T. T. Sasidhar, B. Premjith, and K. P. Soman, "Emotion detection in Hinglish (Hindi+English) code-mixed social media text," Procedia Computer Science, vol. 171, pp. 1346–1352, 2020.
2. D. Vijay, A. Bohra, V. Singh, S. S. Akhtar, and M. Shrivastava, "Corpus creation and emotion prediction for Hindi-English code-mixed social media text," in Proc. 2018 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Student Res. Workshop, 2018.
3. S. Kumar, S. Kumar, S. R. Singh, and S. Nandi, "indiDataMiner at SemEval-2025 Task 11: From text to emotion: Transformer-based models for emotions detection in Indian languages," in Proc. 19th Int. Workshop Semantic Evaluation (SemEval-2025), 2025.
4. S. I. Khan, F. B. Aziz, and M. M. Uddin, "Emotion detection from multilingual text and multi-emotional sentence using difference NLP feature extraction technique and ML classifier," Int. J. Adv. Networking Appl., vol. 14, no. 3, pp. 5429–5435, 2022.