iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77500

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Context-Based Emotion Recognition Using EMOTIC Dataset using Deep Learning

Pratiksha Popat Sarode[1], Saniya Santosh Sarode[2], Rutuja Kundlik Yewale[3], Akshay Ankush Jadhav[4]
Computer Enggineering Department, Samarth Polytechnic Belhe

Abstract: ***Emotion recognition from images is a challenging task due to variations in facial expressions, body posture, and environmental context. Facial features alone are often insufficient to accurately identify emotions in real-world scenarios. This paper presents an implementation of a context-based emotion recognition system using deep learning and the EMOTIC dataset. The proposed approach employs a dual-stream convolutional neural network architecture that separately processes human-centric and contextual information. Features extracted from both streams are fused to predict discrete emotion categories and continuous affective dimensions. Experimental results demonstrate that the inclusion of contextual cues significantly improves emotion recognition performance compared to human-only models, validating the effectiveness of the proposed implementation.***
Keywords: Emotion recognition, EMOTIC dataset, deep learning, context-aware learning, CNN, affective computing.

I. INTRODUCTION

Emotion recognition plays a vital role in human–computer interaction, intelligent surveillance, social robotics, and behavioural analysis. Early emotion recognition systems relied primarily on facial expressions; however, in unconstrained environments, facial cues may be unclear due to occlusion, lighting conditions, or non-frontal views. mail. Humans naturally interpret emotions by considering contextual cues such as body posture, surrounding objects, and scene semantics. Motivated by this observation, context-aware emotion recognition has emerged as an effective alternative. The EMOTIC dataset provides a suitable benchmark for this task by offering real-world images annotated with both discrete emotions and continuous affective dimensions. This paper focuses on the implementation of a deep learning-based context-aware emotion recognition system using the EMOTIC dataset.

II. RELATED WORK

Traditional emotion recognition approaches are based on facial expression analysis using handcrafted features or deep CNNs. Although effective in controlled settings, these methods perform poorly in real-world images. Recent research incorporates body posture and contextual information to overcome these limitations. Dual-stream and multi-modal deep learning architectures have shown promising results by combining human-centric and scene-level features. The EMOTIC dataset has enabled significant progress in this direction by providing rich annotations and bounding box information for human subjects.

A. Page Layout

Your paper must use a page size corresponding to A4 which is 210mm (8.27") wide and 297mm (11.69") long. The margins must be set as follows:

- Top = Bottom= 19mm (0.75")
- Left = Right = 14.32mm (0.56")

III.EMOTIC DATASET DESCRIPTION

The EMOTIC (EMOTION In Context) dataset consists of approximately 23,000 images collected from diverse real-world scenarios. Each image includes:

- Human bounding boxes
- Discrete emotion labels (26 categories such as happiness, anger, fear, sadness)
- Continuous emotion dimensions:
  - Valence
  - Arousal
  - Dominance

©IJRASET: All Rights are Reserved

IV. SYSTEM ARCHITECTURE

A. Overview

The implemented system follows a dual-stream deep learning architecture, consisting of:

1) Human Stream Network
2) Context Stream Network

The outputs of both networks are fused to predict emotional states.

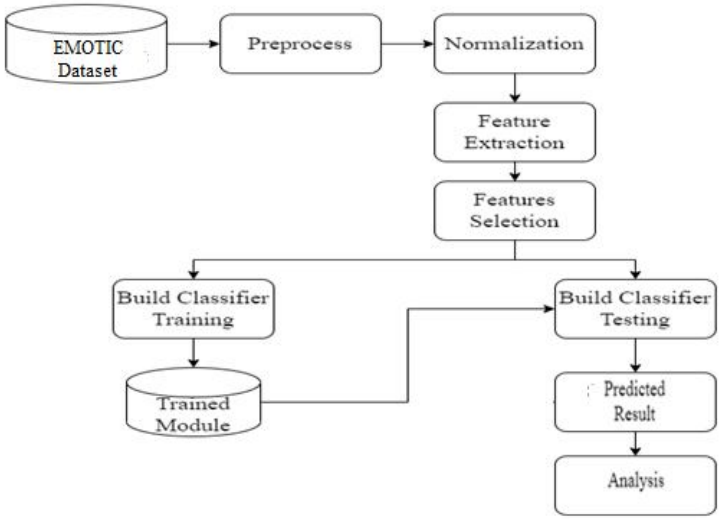


Figure 1: Proposed System Design Implement Module

B. Human Stream

The human stream processes cropped images of detected persons using the bounding boxes provided in the EMOTIC dataset.
This stream focuses on:

1) Body posture
2) Clothing
3) Facial and gesture cues

A pretrained ResNet-50 model is used as the feature extractor. The final fully connected layer is removed, and high-level features are extracted.

C. Context Stream

The context stream processes the full image with the human region masked to avoid redundancy. This stream captures:

1) Scene semantics
2) Surrounding objects
3) Environmental cues

A separate ResNet-50 network is used to extract contextual features.

D. Feature Fusion

Features from both streams are concatenated and passed through fully connected layers. This fusion enables the model to learn relationships between human appearance and contextual information.

E. Emotion Prediction

The fused feature vector is used for:

1) Discrete emotion classification using a Softmax layer
2) Continuous emotion regression (VAD) using fully connected regression layers

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

V. IMPLEMENTATION DETAILS

A. *Preprocessing*

1) Images are resized to 224 × 224
2) Human bounding boxes are cropped for the human stream
3) Data augmentation includes horizontal flipping and normalization
4) Context images are generated by masking the human region

B. *Model Training*

1) Optimizer: Adam
2) Learning Rate: 0.0001
3) Batch Size: 3
4) Epochs: 30

C. *Loss Functions*

1) Cross-Entropy Loss for discrete emotion classification
2) Mean Squared Error (MSE) for continuous emotion regression

The total loss is computed as a weighted sum of classification and regression losses.

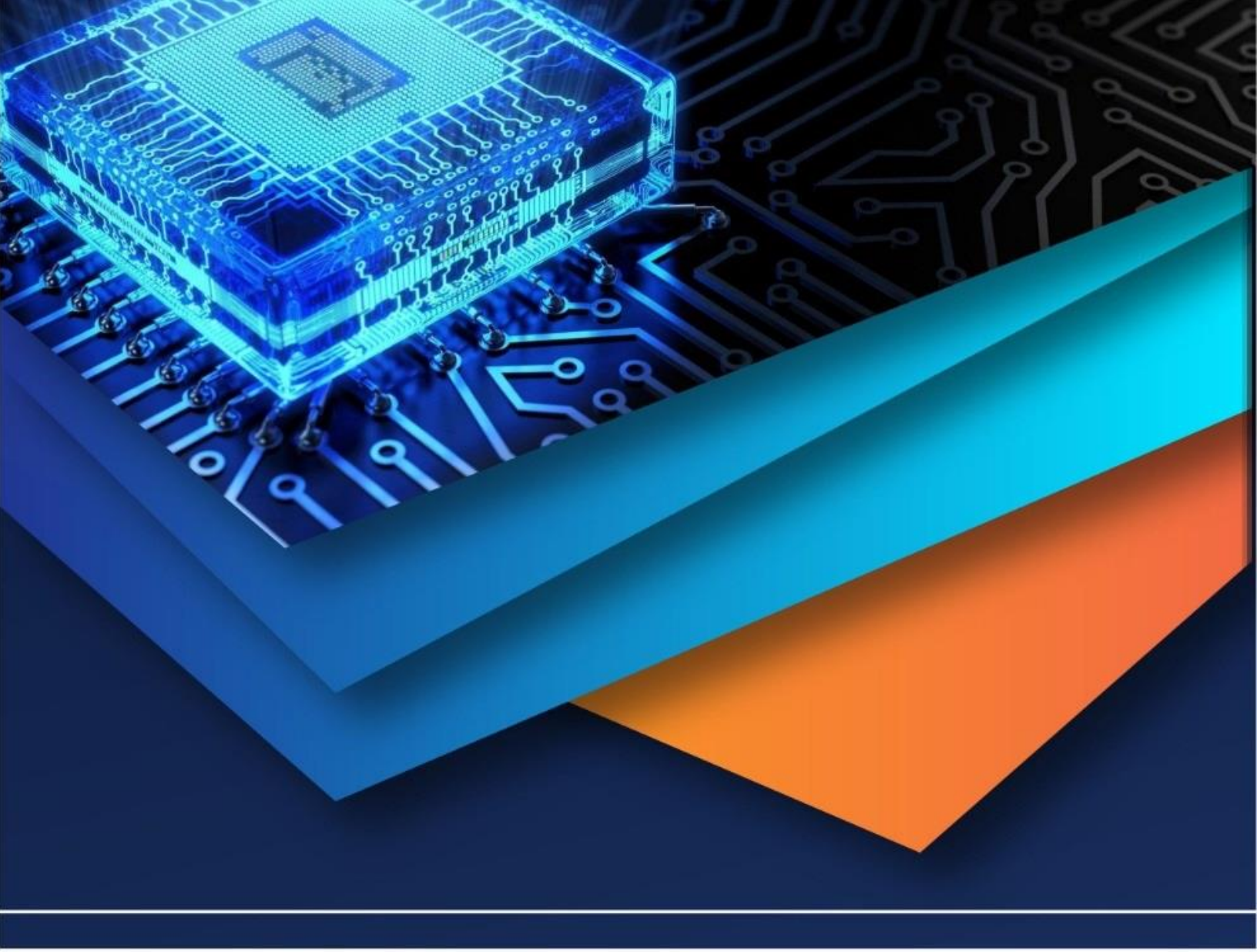
VI. CONCLUSION

This paper presented an implementation of a context-based emotion recognition system using deep learning and the EMOTIC dataset. By combining human-centric and contextual features through a dual-stream architecture, the proposed model achieves improved performance over baseline approaches. Future work will explore attention mechanisms and multi-modal inputs such as audio and text for enhanced emotion understanding.

REFERENCES

[1] Shreya, L., and N. Nagarathna. ”Emotion Based Music Recommendation System for Specially-
[2] Abled.” 2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT). IEEE, 2021.
[3] Lee, Shih-Hsiung, et al. ”A music recommendation system for depression therapy based on EEG.” 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan). IEEE, 2020.
[4] Zhang, Ting-Zheng, Tiffany Chang, and Ming-Hao Wu. ”A Brainwave-based Attention Diagnosis and Music Recommendation System for Reading Improvement.” 2021 IEEE International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC). IEEE, 2021.
[5] Asif, Anum, Muhammad Majid, and Syed Muhammad Anwar. ”Human stress classification using EEG signals in response to music tracks.” Computers in biology and medicine 107 (2019): 182-196.
[6] B˘alan, Oana, et al. ”Emotion classification based on biophysical signals and machine learning techniques.” Symmetry 12.1 (2019): 21.
[7] Santamaria-Granados, Luz, Juan Francisco Mendoza-Moreno, and Gustavo Ramirez-Gonzalez. ”Tourist recommender systems based on emotion recognition—a scientometric review.” Future Internet 13.1 (2020): 2.
[8] Cunha, Joana, et al. ”The Effect of Music on Brain Activity an Emotional State.” Engineering Proceedings 7.1 (2021): 19.
[9] Xu, Baoguo, et al. ”Continuous Hybrid BCI Control for Robotic Arm Using Noninvasive Electroencephalogram, Computer Vision, and Eye Tracking.” Mathematics 10.4 (2022): 618.
[10] Jaafar, Sirwan Tofiq, and Mokhtar Mohammadi. ”Epileptic seizure detection using deep learning approach.” UHD Journal of Science and Technology 3.2 (2019): 41-50.
[11] Kang, Dongwann, and Sanghyun Seo. ”Personalized smart home audio system with automatic music selection based on emotion.” Multimedia Tools and Applications 78.3 (2019): 3267-3276.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)