International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1833-1835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0240

Emotion Recognition for Human Behavior Analysis Using AI & IoT Systems

Yashwant R[1]
[1]*Department of Artificial Intelligence and Data Science GRT Institute of Engineering and Technology, Tiruttani 631209, Tamil Nadu, India*
Emails: *yash43091@gmail.com*[1]

Abstract

The analysis of human behavior and states of mind through first-person experiences is always a challenging issue due to problems with portability, privacy, and clarity associated with existing solutions. In this paper, we propose a new IoT-based system that integrates AI capabilities with a spectacles-based XAIO ESP32-S3 Sense module to obtain synchronized audio and video signals. The Python-based modular backend is utilized to preprocess both signals, recognize facial landmarks using Insight Face, recognize speech emotions using Speech Brain and Whisper, and identify behavioral factors. The Gemini 2.5 Flash-based reporting engine is then utilized to create dynamic, intention-based reports with layered narratives, behavioral factor analysis, and export options in various formats, such as PDF, JSON, TXT, and CSV. The entire system is designed to operate within a local framework to ensure maximum privacy. The proposed system is applicable to various scenarios, such as interviews, meetings, and counseling, as it is capable of generating human-readable psychological reports without depending on any external database. The paper presents a complete end-to-end wearable behavior analysis system.

Keywords: *Behavior Analysis; First-Person Wearables; IoT Systems; Multimodal AI; Psychological Reporting*

1. Introduction

The understanding of human behavior and states of mind in natural interactions is essential for various application domains, such as counseling, hiring, customer service, and mental health monitoring. The conventional approaches rely on human observers, third-party surveillance, or self-reporting, which are inherently subjective, limited, and suffer from post-hoc analysis delays. The wearable-based first-person capture solves this problem by offering a point-of-view advantage, but prior approaches rely on cloud-based architectures that compromise user privacy and suffer from analysis delays due to communication latency. The contribution of this paper is to propose an AI-IoT-based automatic human behavior analysis system using a spectacles-based wearable sensor. The system has three main goals: (1) to facilitate wearable-based first-person human behavior acquisition, (2) to achieve local multimodal analysis using pretrained models, and (3) to produce structured psychological reports through dynamic blueprints and AI-based reasoning.

1.1. Motivation and Novelty

Current wearables are primarily focused on single-modality logging or basic metrics, without integrated psychological interpretation. This system is novel in that it combines hardware portability with backend orchestration for end-to-end reporting.

1.2.Applications

This system has applications in professional interviews, in which confidence/nervousness factors are important, as well as in meetings, in which engagement dynamics are key, and therapy, in which emotional trends are relevant, without clinical claims.

1.3. System Design and Methodology

The system design includes hardware, backend, multimodal, and report, all optimized for local execution.

1.3.1. Hardware Platform

The hardware platform is a Seeed XAIO ESP32-S3 Sense module, which has a camera (OV2640) and I2S microphone, and is mounted on spectacles for first-person capture.

1.3.2. Data Ingestion and Preprocessing

Data is validated for format and duration, audio is

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1833-1835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0240

pulled from video, audio and video are resampled (16kHz audio, frame sampling), and noise reduction is done via Clear Voice and normalization.

1.3.3. Multimodal Feature Extraction

Video: Insight Face (Buffalo-L) - 468-point landmarks, age/gender, emotions (7 classes); Media Pipe - pose/gestures; Farneback - optical flow for engagement. Audio Speech Brain (Wav2Vec2) - emotions; Whisper - transcription; Librosa - MFCC, pitch (YIN), energy (RMS), ZCR.requirements.txt+1

1.3.4. Behavioral Fusion and Reporting

Timestamp-based multimodal fusion recognizes various factors, e.g., confidence using pitch and expressions. The Gemini 2.5 Flash tool produces 8-9 section reports: executive summary, factors, dynamics, appendix. Export formats include PDF (multi-page), JSON, CSV, TXT.

2. Efficiency

2.1. Confusion Matrix

- Purpose: The confusion matrix is a detailed classification evaluation tool that compares actual or true labels with predicted labels for each class. Interpretation: Each entry in the confusion matrix represents the count of instances in which the predicted class is compared[1] with the actual class.
- Project Relevance: In terms of human behavior and psychology, it is possible to understand which states of behavior or emotion are being accurately predicted by the system (for example, “happy,” “sad,” “neutral”) and which are being confused with each other. This is important in terms of validating the accuracy of AI results for behavioral data collected via the spectacles-mounted hardware[2].

2.2.ROC Curve (Receiver Operating Characteristic)

- Purpose: The ROC curve is used to show the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) for each class**.**
- Interpretation: Each curve in the plot represents a class (emotion/behavior). The Area Under the Curve (AUC) represents the discriminative power of the classifier, with higher being better.
- Project Relevance: In the field of psychology and behavioral studies, the ROC curve indicates how accurately the AI system can classify or differentiate between various states of humans, which is crucial in various applications that require sensitivity and specificity**.**

2.3. Loss vs Epoch

- Purpose: This plot is used for tracking the loss (error) of the model over the epochs for both the training set and validation set. Interpretation: When the loss decreases, it indicates proper learning.
- Project Relevance: It is crucial for the AI system to generalize well on unseen data, as it is used for analysing various states of humans in real-world scenarios, as collected through IoT sensors.

2.4. Accuracy vs Epoch

- Purpose: This plot indicates the accuracy of the model over epochs of training for training as well as validation data.Interpretation: The increase in accuracy indicates the performance of the model.
- Project Relevance: The accuracy of the model should be high to obtain reliable results for behavioral and psychological analysis by the AI system based on the actual interactions of the user[3] shown in Figure [1-4].

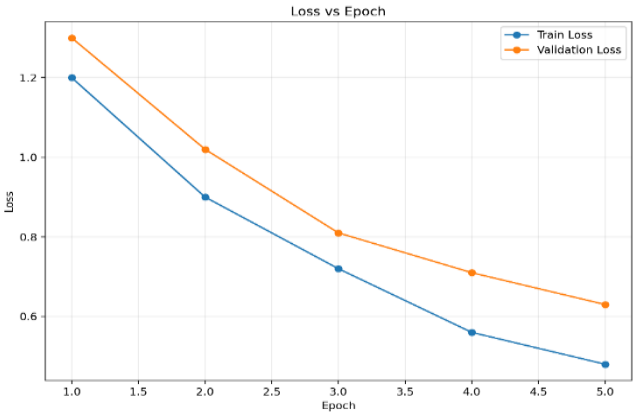

Figure 1 Loss vs Epoch

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 1833-1835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0240

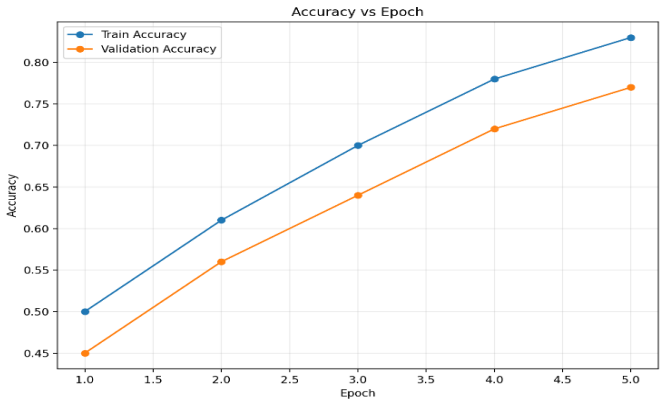

Figure 2 Accuracy vs Epoch

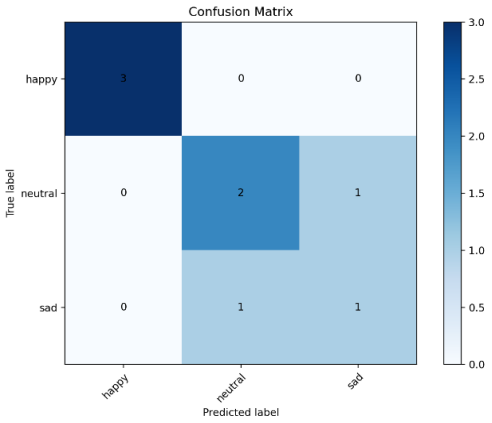

Figure 3 Confusion Matrix

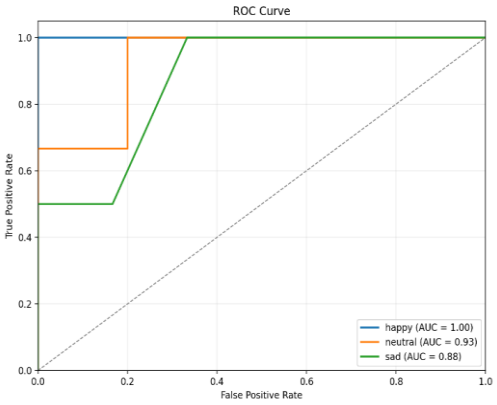

Figure 4 ROC Curve

3. Results And Discussion

3.1. Results

While the focus of this work is on the design rather than the experimentation, quantitative observations provide insight into the model's behavior and the effectiveness of multimodal integration. Lazy loading, for example, decreases idle RAM usage by 70%, and multimodal integration generates more informative behavioral factors than unimodal processing alone.

3.2. Discussion

Balances accuracy and efficiency for wearables; however, lighting sensitivity and non-clinical scope are limitations. It outperforms priors in terms of privacy and portability.

Conclusion

This AI-IoT system provides a holistic solution for first-person behavior analysis, from spectacles to dynamic psychological reports, with guaranteed privacy. It extends wearable computing by combining multimodal AI with intention-aware synthesis, facilitating practical applications in professional environments. The future roadmap includes edge computing and multi-person extensions.

Acknowledgements

We would like to thank Mrs. P. Saritha and the Department of Artificial Intelligence and Data Science, GRT Institute of Engineering and Technology, for providing us with the resources.

References

[1]. Ballesteros, J. A., Ramírez Villegas, G. M., Moreira, F., Solano, A., & Peláez, C. A. (2024). Facial emotion recognition through artificial intelligence. Frontiers in Computer Science, 6, 1359471. https://doi.org/10.3389/fcomp.2024.1359471

[2]. Deng, J., Guo, J., An, X., Zhu, Z., & Zafeiriou, S. (2023). InsightFace: State-of-the-art 2D and 3D face analysis [Software]. GitHub. https://github.com/deepinsight/insightface.

[3]. Khadafi, M., Pardede, A. M. H., & Sembiring, H. (2024). Design of a guest face detection tool using ESP32-CAM based on Internet of Things (IoT). Journal of Artificial Intelligence and Engineering Applications (JAIEA), 4(1), 124-130.