Real-Time Anxiety Detection Using Emotion Echo: A Multi-Modal and Embedded Approach

Dipanjan Saha[1], Swagata Kundu[1], Ishika Rana[1], Meenakshi Sharma[1], and Anjali Kumari[1]

Department of Information Technology,
B. P. Poddar Institute of Management and Technology,
137, VIP Rd, Mali Bagan, Poddar Vihar, Rajarhat, Kolkata, West Bengal 700052, India
dipanjansaha2003@gmail.com, swagata.kundu@bppimt.ac.in, ishikaishaan12@gmail.com, meenasharma282003@gmail.com, anjalikri2003ak@gmail.com

Abstract. Anxiety disorders are among the most prevalent mental health conditions which is often remain undiagnosed due to limited awareness and insufficient real-time diagnostic tools. Existing solutions are constrained by single-source data reliance, classifier inconsistencies, and inadequate responsiveness in real world applications. In this paper, we propose Emotion Echo, a multi-modal system designed for real-time anxiety detection. The system integrates facial emotion recognition using a custom convolutional neural network (CNN), physiological monitoring through a MAX30102 heart rate sensor, and psychological assessment via standardized questionnaires. These modalities are synthesized into a unified anxiety score using a weighted algorithm. Emotion Echo gives 82% accuracy in emotion classification and demonstrates superior performance over transformer-based models in terms of diagnostic precision and computational efficiency. It is optimized for deployment on mobile and embedded platforms, enabling accessible and scalable mental health monitoring.

Keywords: Anxiety Detection · Emotion Recognition · Multi-modal System · Deep Learning · Mental Health Monitoring .

1 Introduction

Depression and anxiety are among the most prevalent mental health disorders glob-ally, affecting individuals across all age groups. Rising academic, professional, and social pressures have contributed to increasing psychological distress, particularly among younger populations. Despite their widespread occurrence, anxiety-related conditions often go undiagnosed and untreated due to societal stigma, limited mental health literacy, and lack of real-time diagnostic tools. Prolonged psychological stress impairs emotional well-being and is associated with reduced cognitive per-formance, lower productivity, physical health issues, and increased suicide risk.

Conventional diagnostic methodologies, such as psychometric questionnaires and structured clinical interviews, offer valuable insights but are inherently static, episodic, and reliant on subjective self-reporting. These approaches are not designed to capture real-time emotional fluctuations or support timely interventions. Although machine learning (ML) and artificial intelligence (AI) have shown promise in mental health applications, existing systems often suffer from issues like reliance on unimodal data sources, inconsistent classifier performance, and limited real-time capability.

Recent years have seen a significant rise in research on machine learning (ML)-based mental health monitoring systems, driven by the need for early and objective diagnosis of anxiety and depression. The effectiveness of using classical Machine Learning algorithms such as Random Forest and Naïve Bayes in predicting stress-related conditions is demonstrated in [17] for the DASS-21 responses. In [7] wearable-based heart rate variability (HRV) and deep learning models are employed to forecast mental health outcomes with notable accuracy. Biosignal data - specifically electrodermal activity (EDA), blood volume pulse (BVP), and EEG - for stress and anxiety classification using ensemble methods like SVM and Random Forests are explored in [8]. In [11] AnxoRelief, an ML-driven mobile application that incorporates standardized psychiatric scales (e.g., PHQ-9, GAD-7) and conversa-tional AI is introduced which reports high accuracy using Random Forest classifiers. The use of social media and online behavioral data for mental health prediction are investigated in [2][9]. In [4][1][22], CNN and Random Forest models are applied to university and adolescent populations, achieving high accuracy (up to 96.8%) in identifying anxiety and depression, especially when incorporating sociodemographic and behavioral inputs[19]. In [16] a survey of physiology-based stress detection has been conducted based on combination of ECG, EMG, and EEG which improves the robustness of diagnosis [6].Transformer-based FER model[**?**] that rely solely on visual cues.

Though these approaches[15][5][23] provide valuable insights, many are constrained by limitations such as classifier inconsistency, reliance on unimodal data, or lack of real-time performance. The need for noninvasive, scalable, and real-time systems has been emphasized in [10][3], particularly for applications beyond clinical settings. To address these limitations, in this paper, we propose Emotion Echo, a lightweight, real-time, multimodal framework that combines facial emotion recognition via a custom CNN, heart rate monitoring using the $MAX30102$ sensor, and depression assessment through standardized self-report questionnaires. The system synthesizes these inputs into a unified anxiety score, offering enhanced diagnostic accuracy and user adaptability.

The rest of the paper is organized as follows: The next section presents system model. In the section 3 we present our proposed framework Emotion Echo with de-tail description of its each components.The experimental setup is presented in sec-tion 4 followed by results and discussion in section 5 and finally concluding remarks in section 6.

2 System Model

Emotion Echo employs a three-stream data pipeline for comprehensive anxiety assessment:
Facial Emotion Recognition (FER): Live video feed from a webcam is processed using a custom CNN model. The model classifies facial expressions into one of seven emotional categories. These classifications are converted into a normalized emotion score that reflects the user's emotional intensity over time.
Heart Rate Monitoring: A MAX30102 pulse oximeter sensor continuously records heart rate (BPM). The heart rate is normalized to generate a physiological stress score, capturing the user's current arousal or rest state.
Psychological Questionnaire: The user answers a set of clinically standardized questions from the Handbook of Psychiatric Rating Scales. Based on the responses, a depression score is computed and categorized (e.g., mild, moderate).

These three scores—emotional(E), physiological(H), and psychological(D) are fused using a weighted formula:

$$A = \alpha \cdot E + \beta \cdot H + \gamma \cdot D \tag{1}$$

Where α, β,γ are three weight factors which reflect the importance of each modality and are designed to balance real-time emotional fluctuation with underlying psychological and physiological states.

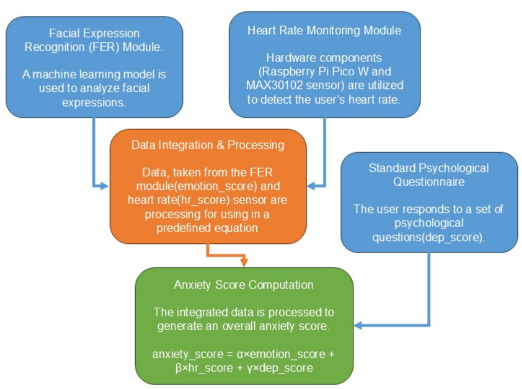

Fig. 1. Block Diagram of Emotion Echo

Figure1 presents the block diagram of our proposed Emotion Echo.

3 Proposed Technique

In this section, we present our proposed technique that introduces a Machine learn-ing based multimodal real-time system for anxiety detection, focusing on: *Real-time image processing with lightweight computation Real-time physiological data using hardware Psychological analysis by giving standard psychological questionnaires*

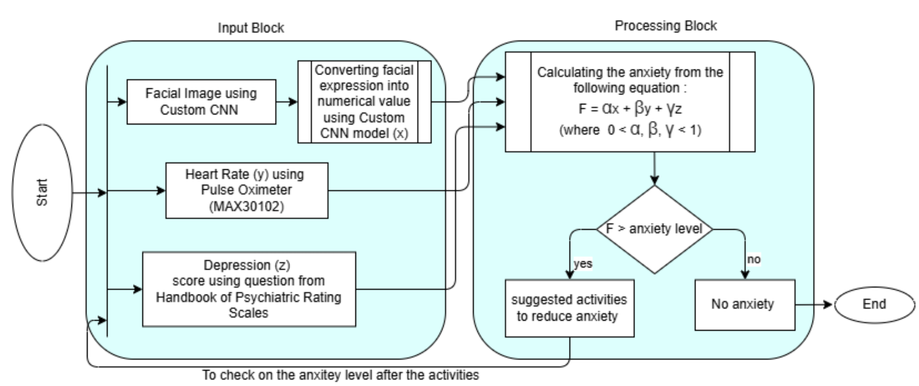

Fig. 2. Emotion Echo Framework

3.1 Real-time image processing

In this section we present how live image in captured by webcam and emotion calculation is done from facial expression.

3.1.1 Emotion Detection and Classification using FER In this section, we will describe a custom-built CNN to detect, classify, and quantify human emotions from facial expressions in real-time, using webcam feeds. It combines *image preprocessing*, *emotion classification*, and *continuous mood tracking for dynamic emotional analysis*

The proposed technique aims to perform real-time emotion recognition from facial expressions using a custom-designed Convolutional Neural Network (CNN). The system workflow is divided into three primary stages:

Image Preprocessing: Facial expression images are converted to grayscale, normalized, and resized to 48×48×1 pixels to reduce computational complexity and en-sure uniform input for the model.

Emotion Classification: The proposed CNN consists of four convolutional layers (128, 256, 512, 512) with max-pooling and dropout layers for feature extraction and regularization. A softmax layer classifies expressions into 7 categories: Happy, Sad, Angry, Fearful, Surprised, Disgusted, and Neutral

. *Real-Time Deployment:* The trained model is integrated into a webcam-based

system using a Haar Cascade Classifier for face detection. Detected faces are pre-processed and passed to the CNN for real-time emotion prediction.
Our CNN Model is designed specifically for facial emotion recognition, enabling better control over layer structure and performance optimization for real-time use. We use four Progressive Convolution 2D Layers with increasing filter sizes ($128 \rightarrow 512$), allowing effective extraction of both low- and high-level emotional features. A consistent dropout rate of 0.4 is applied after each convolution and dense layer to reduce overfitting, especially important with the FER2013 dataset. Input images are resized to 48×48 and converted to grayscale, minimizing computational load and making the model ideal for mobile or embedded deployment. Two dense layers (512 and 256 units) improve learning capacity without excessive complexity, ensuring high accuracy while maintaining fast inference speed.

3.1.2 Emotion Calculation Based on the live image from webcam feeds, our custom CNN model detects and quantifies emotions. Then, the detected emotion is scored based on its intensity and psychological impact.The system records and aggregates emotion scores every 30 seconds to track mood trends. Emotion scoring can help to identify patterns of emotional distress and help to describe levels of depression [21]in the system.

3.2 Heart Rate Detection and Cloud Integration

A real-time heart rate monitoring is done using a MAX30102 heart rate sensor, connected to a Raspberry Pi Pico W, with cloud integration via Firebase for continuous health monitoring. The sensor detects heartbeats, estimates BPM, and sends data online for analysis as part of an anxiety assessment system. The $MAX30102$ sensor connected to the analog stick receives the pulse signals, which are processed to estimate the BPM using a simple formula (BPM = value / 120). An OLED display (128×64) connected via I2C (pins 0: SDA, 1: SCL) provides real-time visualization of BPM.The Raspberry Pi Pico W connects to WiFi using predefined credentials, allowing time-stamped BPM data to be sent to Firebase via HTTP POST.The system operates in a continuous loop for live monitoring, uploading and displaying data, and has interrupt control for safe reset the device if needed.Figure3presents the hardware of heart rate detection.

3.3 Depression Score Calculation from Questionnaires

The system includes a self-assessment module to evaluate depression levels using a structured questionnaire from the Handbook of Psychiatric Rating Scales[13].Users answer a standardized questionnaire, with each response assigned a predefined score.The total score classifies depression into Minimal (0–9), Mild (10–18), Moderate (19–29), or Severe (30+) categories.A web-based interface records responses instantly, providing immediate feedback.The depression score contributes to anxiety level calculations [18][12]in the system.

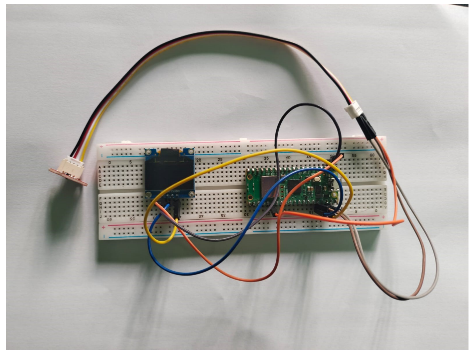

Fig. 3. Realtime Heart rate monitoring system(using MAX30102 sensor)

3.4 Anxiety Scoring by Emotion Echo

Taking facial expression score from custom CNN model, heart rate from MAX30102 heart rate sensor and psychological questionnaire score as inputs , Emotion Echo calculates the final anxeity score by using equation(1).

3.5 Proposed Algorithm

Algorithm 1: Emotion Detection and Scoring Using a CNN Model

Input: Live webcam feed for 30 seconds
Output: Total and average emotion score

Step 1: Import required modules: `sys, io, cv2, time, etc.`
Step 2: Load CNN model from `.json` and `.h5` files using `model_from_json()`, handle exceptions.
Step 3: Initialize Haar Cascade Classifier for face detection.
Step 4: Define `extract_features(image)`: convert to grayscale, reshape to `(1,48,48,1)`, normalize pixel values.
Step 5: Start webcam capture loop for 30 seconds:
  while *time not exceeded* **do**
    Capture frame
    Convert frame to grayscale
    Detect faces using Haar Cascade
    for *each face* **do**
      Extract and resize face ROI
      Predict emotion using CNN
      Display predicted emotion label on frame
    Exit loop if 'q' key is pressed
Step 6: Cleanup: release webcam, close OpenCV windows.
Step 7: Map predicted emotions to predefined scores.
Step 8: Count occurrences, compute total and average emotion scores.

Algorithm 2: Heart Rate Detection Using Microcontroller and Firebase Integration

Input: Pulse sensor data, WiFi credentials
Output: Live BPM data shown on OLED and stored in Firebase

Step 1: Import libraries: `Pin, ADC, I2C, reset, time, network, urequests, SSD1306_I2C`
Step 2: Set constants: `WIFI_SSID, WIFI_PASSWORD, FIREBASE_URL, FIREBASE_KEY`
Step 3: Connect to WiFi: activate and connect interface
Step 4: Initialize OLED via I2C (pins 0, 1, 128x64) and pulse sensor on ADC pin 26
Step 5: `send_to_firebase(bpm)`: create JSON payload, POST to Firebase, handle response
Step 6: `main()`: read sensor, estimate BPM, display on OLED, call Firebase send
Step 7: On interrupt, print message and reset device

4 Result and Analysis

In this section, we discuss, experimental setup, simulation results and comparative analysis of different approaches.

4.1 Experimental Set up

The Emotion Echo system was evaluated using real-time data from multiple participants. For emotion recognition, a custom CNN model was trained on the FER2013 dataset from Kaggle[20], which contains over 35,000 labeled facial images. We split the dataset so that 80% is for training and 20% is for testing. Participants are selected with diverse behavioral and physiological profiles. The final anxiety score was calculated by using normalized values weight fators for facial anxiety, heartrate and psychlogical quesnnarie respectively α, β and γ. We have set α=0.3, β=0.3 and γ=0.4 ensuring accurate and personalized detection.

4.2 Comparison of Custom CNN Model with existing models

Emotion detection requires a balance between accuracy, speed, and efficiency. Among the five models tested—Basic CNN, EfficientNet-B0, VGG16, Mo-bileNetV2, and our Custom CNN—the Custom CNN proved to be the most effective. Unlike pre-trained models, it's trained from scratch to learn emotion-specific features, offering high accuracy (82%) with faster inference and lower computational cost. While VGG16 and EfficientNet-B0 are accurate but heavy, and MobileNetV2 is efficient but slightly less accurate, our model strikes the best balance for real-time use. Its optimized size and performance make it ideal for deployment across devices, including edge and mobile platforms. Figure 4 represents comparison of custom CNN model with existing above said models. From the graph it is clear that custom CNN model has better traning accuracy and validation accuracy also has minimum training and validation loss. Thus our custom CNN model out perform over existing models. To evaluate the effectiveness of the facial emotion

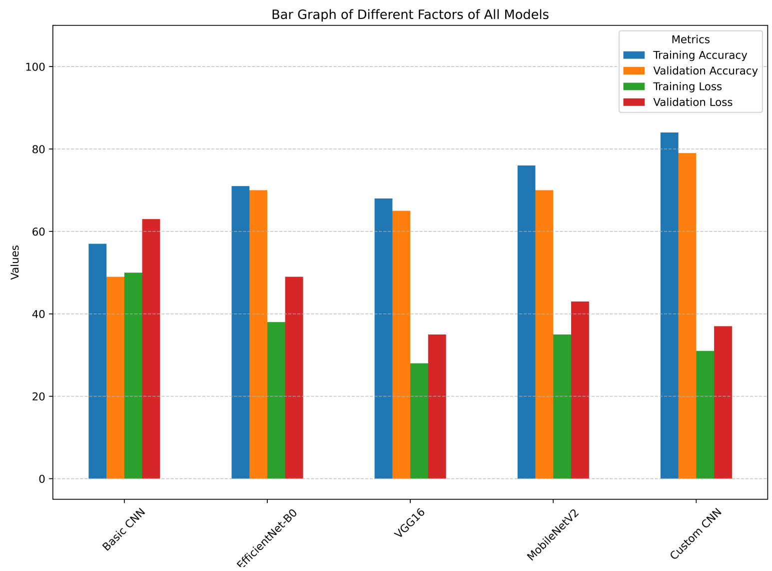

Fig. 4. Accuracy Comparison of Custom CNN model with other existing models

recognition component, we analyzed classification performance on the FER2013 dataset using standard metrics. The results are summarized in th table**??**. The

Table 1. Emotion Detection Performance Metrics

Emotion	**Precision**	**Recall**	**F1-Score**
Happy	0.85	0.88	0.86
Sad	0.81	0.78	0.79
Angry	0.80	0.76	0.78
Fearful	0.76	0.74	0.75
Surprised	0.84	0.87	0.85
Disgusted	0.72	0.70	0.71
Neutral	0.79	0.82	0.80
Average	**0.79**	**0.79**	**0.79**

confusion matrix shows that the Custom CNN model performs well across all emotion classes, with the highest accuracy in detecting Happy and Surprised emo-tions, while minor misclassifications occur between similar expressions like Sad–Neutral and Angry–Fearful.

Table 2. Confusion Matrix for Emotion Classification

Actual / Predicted	Happy	Sad	Angry	Fearful	Surprised	Disgusted	Neutral
Happy	135	4	1	2	6	0	2
Sad	3	125	5	6	2	3	6
Angry	1	6	123	7	1	7	5
Fearful	2	5	6	118	4	4	2
Surprised	5	1	2	2	134	2	4
Disgusted	0	2	5	3	1	115	4
Neutral	2	4	3	2	5	4	130

To ensure the authenticity of the Emotion Echo, we have selected individuals from di-verse backgrounds, attitudes, and personalities. This variation allows us to capture a broad spectrum of emotional and psychological states, ensuring that the data is representative and comprehensive.

Table 3. Emotion, Heart Rate, Depression, and Anxiety Scores of Users

UID	Emotion Score	Heart Rate Score	Depression Score	Anxiety Score
U1	0.78	0.65	0.50 (15/30)	0.629
U2	0.53	0.62	0.23 (7/30)	0.437
U3	0.82	0.72	0.30 (9/30)	0.582
U4	0.92	0.84	0.83 (25/30)	0.860
U5	0.75	0.68	0.86 (26/30)	0.773
U6	0.71	0.62	0.60 (18/30)	0.639
U7	0.24	0.54	0.76 (23/30)	0.538
U8	0.79	0.70	0.40 (12/30)	0.607
U9	0.64	0.63	0.63 (19/30)	0.633
U10	0.78	0.86	0.73 (22/30)	0.789
U11	0.38	0.61	0.20 (6/30)	0.377
U12	0.61	0.74	0.36 (11/30)	0.549
U13	0.89	0.82	0.90 (27/30)	0.873
U14	0.74	0.75	0.70 (21/30)	0.727
U15	0.83	0.78	0.56 (17/30)	0.707

Based on the result matrix from 15 diverse participants, the Emotion Echo system demonstrates consistent and realistic anxiety scoring across varying emotional, physiological, and behavioral profiles. Users with higher depression scores consistently yielded higher overall anxiety scores, even when emotion or heart rate scores were moderate.

4.3 Comparison with Benchmark Work

Transformer-based FER model[14] use static facial expressions to infer mental health, relying on architectures like Swin or Vision Transformers. Both the existing Transformer-based FER model and our Custom CNN model use the same FER2013 dataset for emotion classification. However, the accuracy of our model consistently falls in the higher range (74–82%) while Transformer-based FER model ranges between 70–75%. This difference is significant in the context of multi-class emotion classification and especially relevant when these predictions are used for downstream mental health inference. The graph in figure5 shows

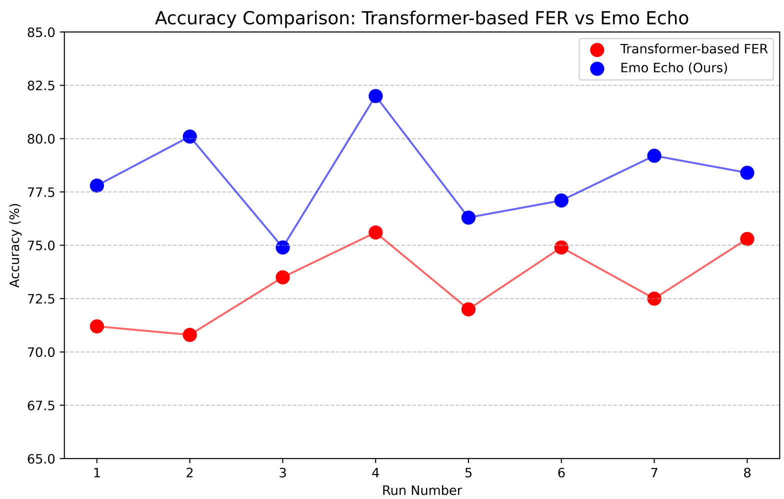

Fig. 5. Comparison of Emotion Echo with Baseline work[14]

that our model significantly outperforms the existing Transformer-based FER system by incorporating a multi-modal approach.

5 Conclusion

In this we have proposed Emotion Echo, a real-time, multi-modal system for anxiety detection and management. By integrating facial emotion recognition, physiological signals, and psychological assessments, the system offers a comprehensive and accurate evaluation of emotional states. Its lightweight and privacy-aware design supports both clinical and personal applications, with superior performance compared to conventional and transformer-based model. As a future work we will focus on application of reinforcement learning and integration with wearable devices . In critical cases, the system aims to facilitate timely professional support through clinical collaboration. These developments aspire to deliver a scalable, intelligent, and user-centric solution for mental health support.

References

1. Ahmed, A.: A machine learning approach to detect depression and anxiety using supervised learning. In: 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (2020)
2. Ahmed, A., Aziz, S., Toro, C., Alzubaidi, M., Irshaidat, S., Abu Serhan, H., Abd-alrazaq, A., Househ, M.: Machine learning models to detect anxiety and depression through social media: A scoping review. Computer Methods and Programs in Biomedicine Update **2**, 100066 (2022). https://doi.org/10.1016/j.cmpbup.2022.100066
3. Ancillon, L., Elgendi, M., Menon, C.: Machine learning for anxiety detection using biosignals: A review. Diagnostics **12**(8), 1794 (2022). https://doi.org/10.3390/diagnostics12081794
4. Bhatnagar, S., Agarwal, J., Sharma, O.: Detection and classification of anxiety in university students through the application of machine learning (2020)
5. Bobade, P., Vani, M.: Stress detection with machine learning and deep learning using multimodal physiological data. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). pp. 51–57 (2020). https://doi.org/10.1109/ICIRCA48905.2020.9183244
6. Bota, P.: A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. IEEE Access (2019). https://doi.org/10.1109/ACCESS.2019.2944001
7. Coutts, L.V., Plans, D., Brown, A.W., Collomosse, J.: Deep learning with wearable based heart rate variability for prediction of mental and general health. Journal of Biomedical Informatics **112**, 103610 (2020). https://doi.org/https://doi.org/10.1016/j.jbi.2020.103610, https://www.sciencedirect.com/science/article/pii/S1532046420302380
8. Gedam, S., Paul, S.: A review on mental stress detection using wearable sensors and machine learning techniques. IEEE Access (2021). https://doi.org/10.1109/ACCESS.2021.3085502
9. Gruda, J., Hasan, S.: Feeling anxious? perceiving anxiety in tweets using machine learning (2019). https://doi.org/10.31234/osf.io/xfmqd
10. Hickey, B., Chalmers, T., Newton, P., Lin, C., Sibbritt, D., McLachlan, C., Clifton-Bligh, R., Morley, J., Lal, S.: Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. Sensors **21**(10), 3461 (2021). https://doi.org/10.3390/s21103461
11. Kumbhar, M., Aleem, Z., Rajput, A., Narejo, S., Rathi, M.: Detecting and relieving anxiety on-the-go using machine learning techniques (2020). https://doi.org/10.13140/RG.2.2.11178.49609
12. Li, J.: Comparison and analysis of deep neural networks in facial expression recognition. Applied and Computational Engineering **21**, 93–98 (2023)
13. Lyerly, S., Abbott, P.: Handbook of Psychiatric Rating Scales (1950–1964). Psychological Assessment Associates, National Institute of Mental Health (1950)
14. Mujiyanto, M., Setyanto, A., Kusrini, K.: Swin transformer with enhanced dropout and layer-wise unfreezing for facial expression recognition in mental health detection. In: Proceedings of the [Conference Name]
15. Nath, R., Thapliyal, H.: Machine learning-based anxiety detection in older adults using wristband sensors and context feature. SN Computer Science **2**(359) (2021). https://doi.org/10.1007/s42979-021-00744-z

16. Panicker, S., Gayathri, P.: A survey of machine learning techniques in physiology based mental stress detection systems. Biocybernetics and Biomedical Engineering **39** (2019). https://doi.org/10.1016/j.bbe.2019.01.004
17. Priya, A., Garg, S., Tigga, N.: Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Computer Science **167**, 1258–1267 (2020). https://doi.org/10.1016/j.procs.2020.03.442
18. Qian, C., Lobo Marques, J., de Alexandria, A., Fong, S.: Application of multiple deep learning architectures for emotion classification based on facial expressions. Sensors **25**(5), 1478 (2025). https://doi.org/10.3390/s25051478, published February 27, 2025
19. Razavi, M., Ziyadidegan, S., Mahmoudzadeh, A., Kazeminasab, S., Baharlouei, E., Janfaza, V., Jahromi, R., Sasangohar, F.: Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: Scoping review. JMIR Mental Health **11** (2024). https://doi.org/10.2196/53714
20. Sambare, M.: Facial expression recognition (fer2013). https://www.kaggle.com/datasets/msambare/fer2013, accessed 2025-06-20
21. Shylaja, R., Megha Sai Sree, P., Rishika, P., Sreeja, T., Srilekha, M.: Facial emotion recognition based scoring system. In: Pandian, A., Palanisamy, R., Narayanan, M., Senjyu, T. (eds.) Proceedings of Third International Conference on Intelligent Computing, Information and Control Systems. Advances in Intelligent Systems and Computing, vol. 1415. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-7330-6_15
22. Tabares Tabares, M., Vélez Álvarez, C., Bernal Salcedo, J., Murillo Rendón, S.: Anxiety in young people: Analysis from a machine learning model. Acta Psychologica **248**, 104410 (2024). https://doi.org/10.1016/j.actpsy.2024.104410
23. Zhang, H., Feng, L., Li, N., Jin, Z., Cao, L.: Video-based stress detection through deep learning. Sensors **20**(5552) (2020). https://doi.org/10.3390/s20195552