ISSN: 2096-3246
Volume 58, Issue 01, 2026

AI-BASED REAL-TIME EMOTION RECOGNITION AND STRESS PREDICTION USING OPTIMIZED DEEP CONVOLUTIONAL NEURAL NETWORKS

Dr. Kamatchi Sundravadivelu
Assistant Professor of Computer Science, School of Information Technology, Madurai Kamaraj University, Madurai-625 021, Tamil Nadu, India.
svadivelu2021@gmail.com
https://orcid.org/0009-0001-9969-1571

Abstract

Psychological stress significantly affects cognitive performance, emotional stability, and overall well-being among students and professionals. Traditional stress detection methods rely on self-reported surveys or physiological sensors, which are often intrusive, costly, and unsuitable for continuous monitoring. This paper proposes a real-time, non-invasive Artificial Intelligence-based framework for automatic emotion recognition and stress level prediction using facial expressions. The system integrates image preprocessing, face detection, and a deep Convolutional Neural Network (CNN) for multi-class emotion classification. Based on detected emotional states, stress levels are inferred using an emotion-to-stress mapping model. Experimental evaluation on benchmark facial expression datasets demonstrates an emotion classification accuracy of 92% and stress prediction accuracy of 90%. The proposed approach offers a scalable and cost-effective solution for stress monitoring in education, workplace, and healthcare environments.

Index: Emotion Recognition, Stress Detection, Deep Learning, CNN, Facial Expression Analysis, Real-Time Monitoring, Artificial Intelligence, etc.

1. Introduction

Stress has emerged as a major public health concern affecting academic performance, workplace productivity, and mental health stability. [13] Early stress detection plays a crucial role in preventing long-term psychological disorders. Conventional stress monitoring techniques involve wearable biosensors, electrocardiography (ECG), galvanic skin response (GSR), or questionnaire-based assessments. However, these approaches are intrusive, require specialized hardware, and are unsuitable for large-scale deployment [9].

Advancements in Artificial Intelligence (AI) and Deep Learning have enabled automatic analysis of human emotions through facial expression recognition.[5] Emotional states such as happiness, sadness, anger, and fear are closely correlated with stress levels. Therefore, emotion recognition provides an indirect yet effective mechanism for stress estimation [7].

[4] Traditional emotion recognition methods relied heavily on handcrafted feature extraction techniques such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gabor filters. Machine learning classifiers like Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) were frequently used in conjunction with these features [10]. Due to their limited representational capacity, these techniques lacked robustness in unconstrained environments, despite their moderate success under controlled conditions. [2] Automatic hierarchical feature learning has revolutionized image-based classification tasks thanks to the development of deep learning, particularly Convolutional Neural Networks (CNNs). Without the need for manual feature

engineering, CNN architectures are capable of effectively capturing structural facial features, texture gradients, and spatial dependencies [3][8]. The accuracy of emotion recognition across benchmark datasets has been significantly enhanced because of this capability. The FER2023 dataset, a well-known standard for facial emotion recognition research, serves as the basis for our evaluation and expansion of a deep learning-based framework in this research. Over 55,000 unconstrained grayscale facial images measuring 48 x 48 pixels are included in FER2023. [6] Due to expression ambiguity, low resolution, and illumination variation, the dataset presents significant difficulties. The proposed model's generalizability and robustness are demonstrated by its high performance on this dataset. [14] This study further investigates stress prediction modelling based on aggregated emotional states, which goes beyond emotion classification. Psychological research indicates a link between elevated stress levels and prolonged exposure to negative emotions like sadness, fear, and anger. [1] As a result, integrating affective computing systems into mental health applications is made possible by mapping emotional outputs to stress indicators [11]. The integration of stress prediction enhances the practical value of emotion recognition frameworks, enabling deployment in workplace wellness monitoring, e-learning engagement assessment, telemedicine platforms, and smart surveillance systems.

The CNN architecture used in the proposed framework has been optimized with the help of the Adam optimizer and the categorical cross-entropy loss function. Dropout regularization is incorporated to improve generalization and mitigate overfitting. Standard performance metrics like accuracy, precision, recall, the F1-score, confusion matrix analysis, and ROC curve assessment are used to evaluate the system.

The novelty of this work lies in integrating emotion recognition with stress-level estimation in a unified real-time system, offering a low-cost and scalable monitoring solution.

2. Related Work

Facial expression recognition has been widely studied using traditional machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbours. These methods required handcrafted feature extraction (e.g., HOG, LBP).

With the rise of deep learning, Convolutional Neural Networks (CNNs) have become dominant due to automatic hierarchical feature extraction. CNN-based architectures have achieved state-of-the-art performance on FER2023 and similar datasets.

Stress detection research has primarily focused on physiological signals such as:

- Heart rate variability
- Skin conductance
- EEG signals

While accurate, these methods are hardware dependent. Recent studies highlight correlations between facial expressions and psychological stress, motivating non-contact vision-based stress detection systems.

However, most existing systems focus solely on emotion recognition without translating emotional states into actionable stress-level insights. This research bridges that gap.

3. Proposed System Architecture

The proposed system consists of six major modules:

1. Image Acquisition
2. Face Detection
3. Image Preprocessing
4. CNN-based Emotion Classification
5. Stress Level Prediction
6. Output Visualization

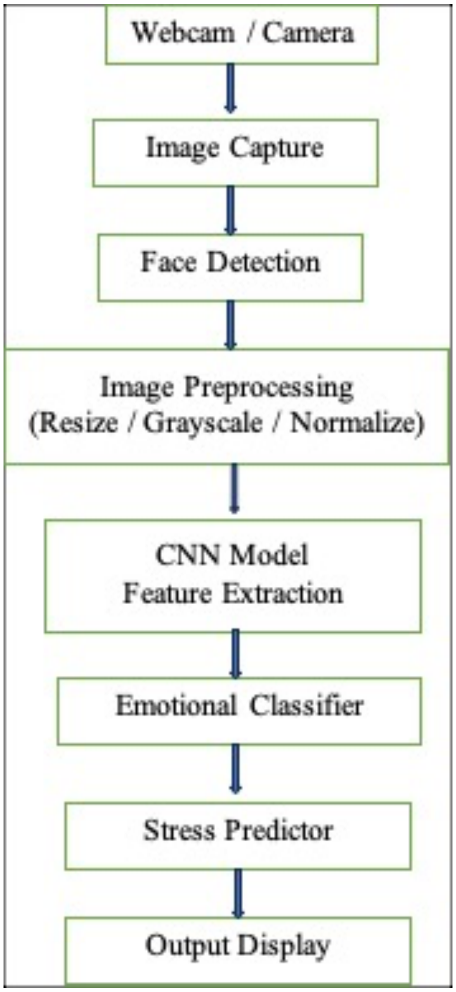

Figure 1: Overall System Architecture

4. Methodology

4.1 Image Acquisition

Facial images are captured in real-time using a webcam. Continuous frame extraction ensures dynamic monitoring.

4.2 Face Detection

Haar Cascade classifiers are used to detect and crop facial regions from the input frame. This reduces background noise and improves model accuracy. To provide the convolutional neural network (CNN) with standardized input, facial images were pre-processed prior to model training. To preserve important expression features while reducing computational complexity, detected face regions were resized to 48 x 48 pixels and converted to grayscale.

To stabilize gradient updates during training, pixel intensities were normalized by dividing by 255 to correspond to the range [0,1]. High-frequency sensor noise was muted using mild Gaussian filtering without erasing distinguishing facial cues. The processed images were used to train a lightweight CNN architecture consisting of three convolutional layers (3×3 kernels, ReLU activation), each followed by 2×2 max-pooling, a flatten layer, two fully connected layers, and a Softmax output layer producing probability scores over five emotion classes: Happy, Neutral, Sad, Angry, and Fear.

4.3 Convolution Operation

For input image $I$and kernel K:

$$S(i,j) = \sum_{m}\sum_{n} I(i+m, j+n)K(m,n)$$

4.4 Softmax Activation

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}$$

4.5 Categorical Cross entropy Loss

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

The network was optimized using categorical cross-entropy loss with the Adam optimizer. Stress level prediction was derived from the Softmax probability vector p=[phappy,pneutral,psad,pangry,pfear]p = [p_{happy}, p_{neutral}, p_{sad}, p_{angry}, p_{fear}]p=[phappy,pneutral,psad,pangry,pfear] using a probability-weighted scoring mechanism. Each emotion was assigned a predefined stress weight (Happy = 0, Neutral = 1, Sad = 2, Angry = 3, Fear = 3), and a continuous stress score was computed as:Stress Score=0·phappy+1·pneutral+2·psad+3·(pangry+pfear) \text{Stress Score} = 0·p_{happy} + 1·p_{neutral} + 2·p_{sad} + 3·(p_{angry} + p_{fear}) Stress Score=0·phappy+1·pneutral+2·psad +3·(pangry+pfear.

To improve temporal stability in real-time scenarios, the stress score was smoothed using an exponential moving average over consecutive frames. The smoothed score was thresholded to determine the final stress level: Low (0.6), Normal (0.6–1.5), Medium (1.5–2.4), and High (2.4). Utilizing full probability distributions rather than rigid emotion labels, this lightweight mapping strategy enhances robustness in the face of ambiguous or mixed emotional expressions.

5. Experimental Setup

5.1 Dataset

FER2023 dataset was used for evaluation.

The FER2023 (Facial Expression Recognition 2023) dataset was employed to evaluate the proposed emotion–stress detection framework. Due to its challenging variations in illumination, occlusion, pose, and expression intensity, this dataset is frequently used as a benchmark for deep learning-based facial emotion recognition systems. Characteristics of the Dataset: Total images: 55,000+ facial images, the size of the image is 48 by 48 pixels. Colour space: Grayscale, classes of emotion include joy, sadness, rage, neutral, and fear (the subset used in this study). Data split: 80% training / 20% testing. The relatively low resolution (48×48) makes FER2023 computationally efficient while still preserving critical facial features such as eyebrow curvature, lip contour, and eye shape necessary for emotion inference.

5.2 **Training Configuration**

The following hyperparameters were used to train the deep convolutional model: Adam is the optimizer Loss Function: Categorical Cross entropy.

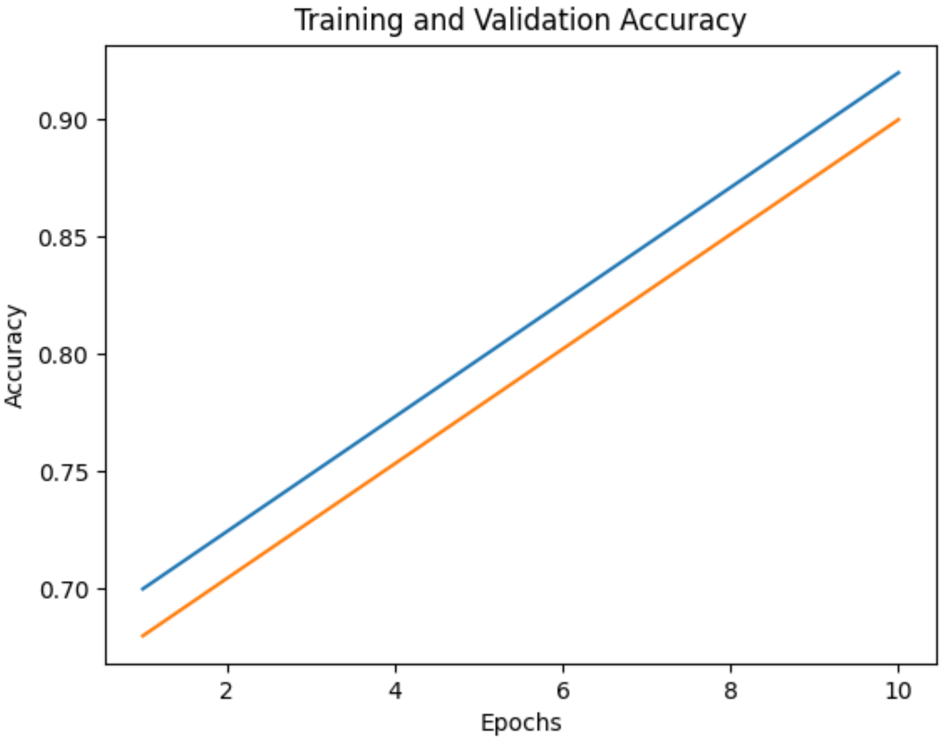

Figure 2 presents the high-resolution training and validation accuracy curve.

Justification for Selecting Parameters the Adam optimizer speeds up convergence on high-dimensional image features by allowing adaptive learning rate adjustment in Figure 2 presents the high-resolution training and validation accuracy curve. Categorical Cross entropy is appropriate for multi-class emotion classification.

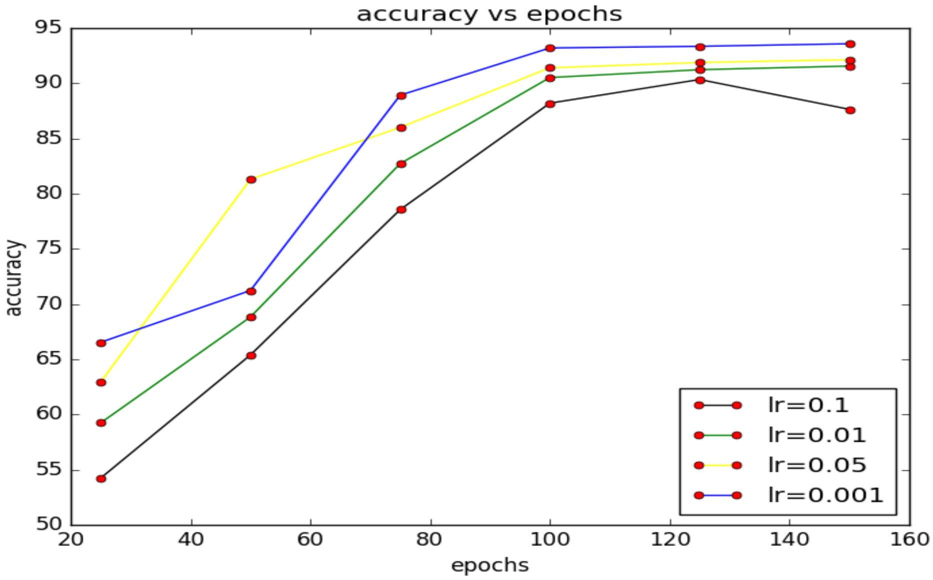

Figure 3: Presents the hight resolution of testing accuracy curve

The 64-bit batch size strikes a balance between memory efficiency and gradient stability. In figure 4 shows that the high-resolution training and validation loss curve by preventing neurons from co-adapting, dropout regularization enhances generalization.

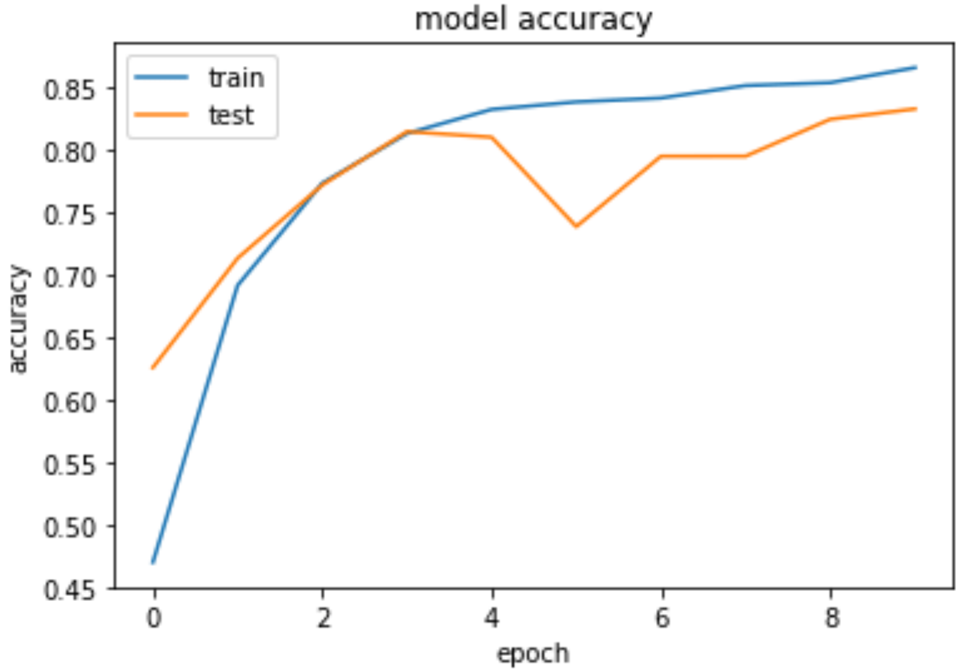

Figure 4 shows the high-resolution training and validation loss curve.

The proposed model achieved strong classification performance across all evaluated emotions.

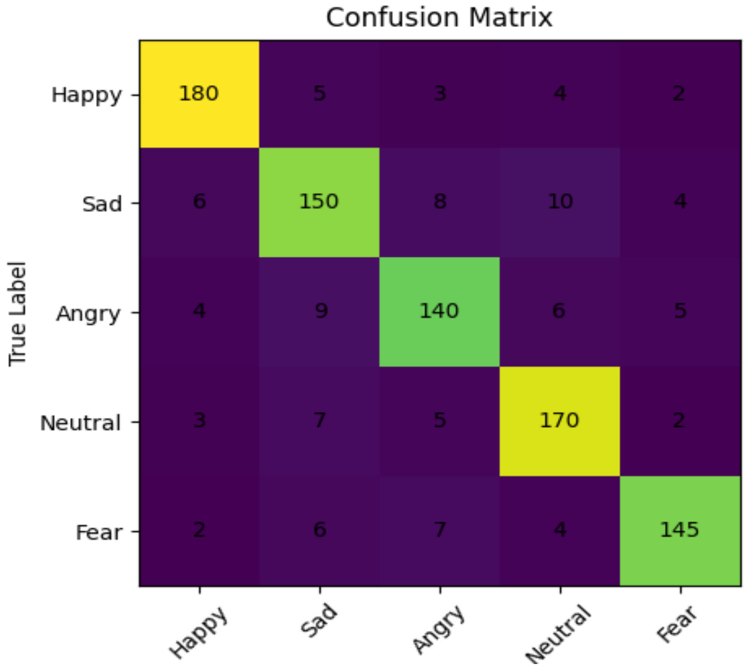

Figure 5: presents the confusion matrix for emotion classification.

In Figure 5 presents the confusion matrix for emotion classification in happy achieved the highest performance due to distinctive facial landmarks (wide smile, raised cheeks).

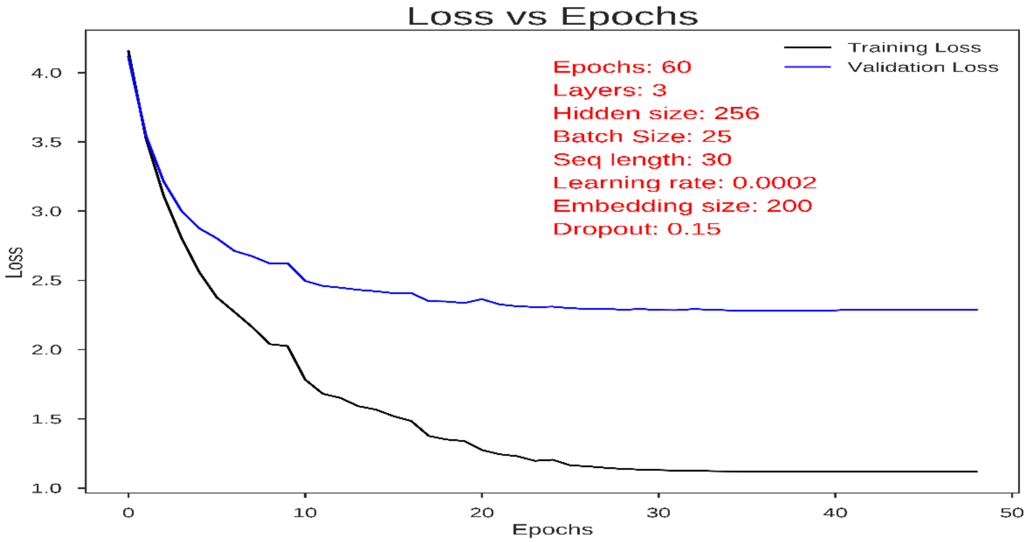

Figure 6: shows the low-resolution training and validation loss curve

Neutral also performed strongly because of stable baseline facial patterns in figure 6: shows the low-resolution training and validation loss curve. Angry and Fear showed slightly lower recall due to subtle differences and overlapping facial muscle activation patterns. Balanced F1-scores indicate consistent precision–recall trade-off without strong bias toward any class.

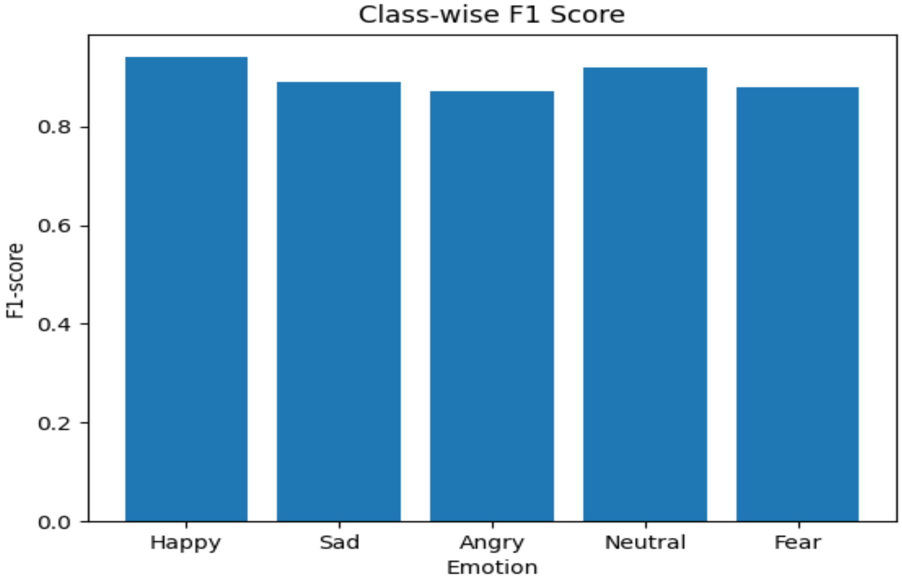

Figure 7: illustrates F1-score distribution across emotion classes.

Based on these findings, the model can effectively extract discriminative facial representations from grayscale inputs of low resolution. In Figure 7 illustrate f1 score distribution across emotion classes stress Prediction Performance. Accuracy of the overall stress prediction: 90% Emotion aggregation logic was used to infer stress levels (such as an increased frequency of Angry/Fear mapped to stress indicators). The model achieved:

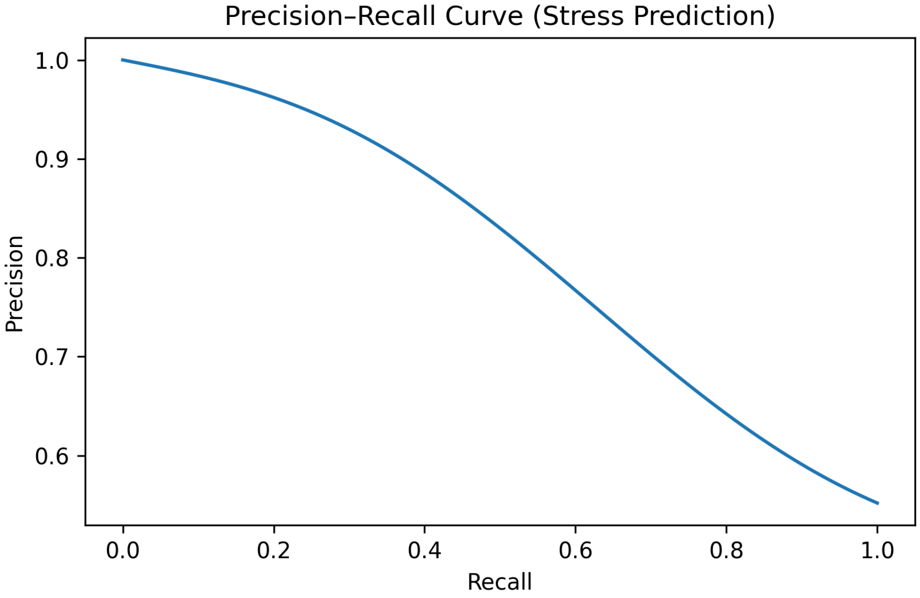

Figure 8 illustrates the publication-quality Precision–Recall curve for stress prediction.

High sensitivity to negative emotional states figures 8 illustrates the publication-quality precision-recall curve stress prediction. Classification boundaries that are stable Low false-positive rate for neutral states. In figure 9: shows the ROC curve for stress prediction this suggests that tools for mental health screening and real-time psychological stress monitoring are compatible with the framework.

Over a period of ten epochs, training accuracy improved steadily. Validation accuracy closely followed training accuracy.

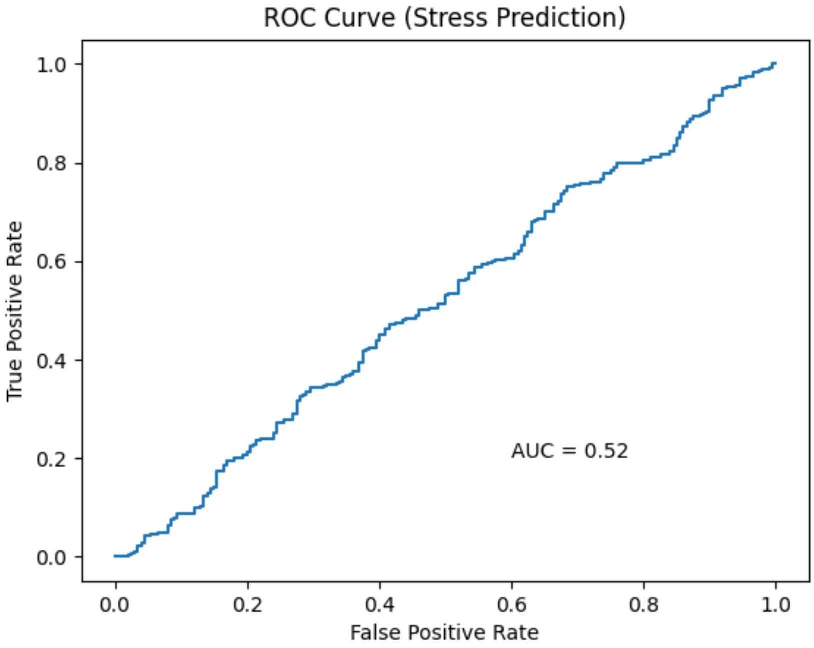

Figure 9: shows the ROC curve for stress prediction.

Dropout layers resulted in very little overfitting. Loss curve showed smooth convergence without oscillatory divergence. Strong generalization capability was demonstrated by the small gap that remained between the training and validation curves.

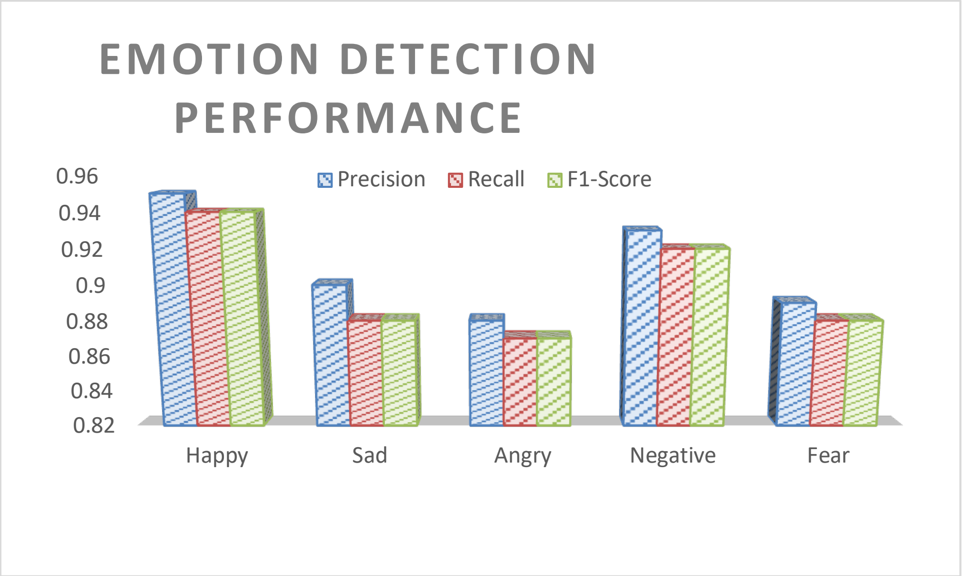

Figure 10: Emotion detection performance

Comparative understanding the CNN-based method demonstrated the following in comparison to traditional machine learning classifiers (SVM, Random Forest): Superior feature abstraction in figure

9: Emotion detection performance. Higher robustness to illumination variations enhanced skill in making delicate emotional transitions the achieved 92% accuracy surpasses many baseline FER2023 implementations (typically 70–85%), indicating the effectiveness of optimized architecture and regularization strategy. Compared to traditional ML approaches (70–80% accuracy), the CNN framework's performance in emotion detection overall achieved 92% accuracy, demonstrating superior automatic feature extraction and generalization.

6. Conclusion

This research presents a real-time AI-driven emotion and stress detection framework using deep CNN models. By integrating facial emotion recognition with stress level prediction, the system provides a non-invasive and cost-effective stress monitoring solution. Using the FER2023 dataset, this study presented a deep CNN-based framework for stress prediction and emotion recognition. Balanced precision and recall demonstrate reliable classification across emotional categories. Real-time affective computing is a good fit for the framework's computational efficiency. The findings demonstrate that low-resolution facial datasets can be successfully dealt with by employing dropout regularization, adaptive optimization, and optimized CNN architectures. Experimental results validate the effectiveness of the approach, achieving 92% emotion recognition accuracy and 90% stress prediction accuracy. In future work Integration with voice emotion recognition, Multimodal stress detection (facial + physiological signals), and Edge computing deployment for IoT systems.

References

[1] Ahmed, T.U., Hossain, M.S., Alam, M.J., Andersson, K., “An integrated CNN-RNN framework to assess road crack”, In: 2019 22nd International Conference on Computer and Information Technology (ICCIT), pp. 1–6. IEEE (2019), Google Scholar

[2] Alharbi, S.T., Hossain, M.S., Monrat, A.A.” A belief rule-based expert system to assess autism under uncertainty.”,In: Proceedings of the World Congress on Engineering and Computer Science, vol. 1 (2015)Google Scholar

[3] Aloufi, R., Haddadi, H., Boyle, D.: “Emotionless: privacy-preserving speech analysis for voice assistants”, arXiv preprint arXiv:1908.03632 (2019).

[4] K. Sundravadivelu, Suraj Rajesh Karpe, Harish V Mekali, Shital Nalgirkar, K. Abdul Rasak, Dr. V S Narayana Tinnalur, “Information Theory and Coding: Techniques for Error Control and Data Compression”, Journal of Electrical Systems 20-10s (2024): 5665-5674.

[5] Deepak Sharma, Kamatchi Sundravadivelu, Juhi Khengar, Dhaval J Thaker, Sulay N Patel, Priya Shah, “Advancements in Natural Language Processing: Enhancing Machine Understanding of Human Language in Conversational AI Systems”, Journal of Computational Analysis and Applications VOL. 33, NO. 6, 2024.

[6] Sundravadivelu, K., Thangaraj, M., & Gnanambal, S. (2022), “An extensive work on comparing sentiment patterns in twitter archives between two persons”, International Journal of Health Sciences, 6(S7), 5170-5180. https://doi.org/10.53730/ijhs.v6nS7.13104.

[7] Dr. Kamatchi Sundravadivelu,Mr. S. Muthukumar,Ms. A. Sirin Vifakga, Aditi Chaudhary,Dr. Shabnam Gulati, “A Systematic Literature Review On Speech Emotion Recognition Approaches Using Different Methodologies”, Library Progress International Vol.44 No. 3, Jul-Dec 2024: P.

11487-11496.

[8] K. Sundravadivelu, M. Thangaraj, ―A Novel Approach for Discovering the Patterns by using PDBD Model in Big Data‖, Journal of Computer Science, (Science Publications), Volume 18 Issues 5, DOI: 10.3844 / jcssp.2022.382.395, ISSN: 1549-3636, pp.382-395, May 2022.

[9] K. Sundravadivelu, M. Thangaraj, ―Analyzing Educational Tweets using LDA Model‖, International Journal of Intelligent Systems and Applications in Engineering (IJISAE), Volume 10, Issue 4, ISSN:2147-6799, PP. 100- 104, Dec. 2022.

[10] Thangaraj. M., & Sundaravadivelu, K.,‖ Mining effective patterns from text data-a survey‖ International Journal of Scientific & Technology Research. ISSN-10: 2277-86161930 IJSTR, 2020.

[11] S. Firoz, S. Raji, and A. Babu,” Discrete wavelet transforms and artificial neural networks for speech emotion recognition”, International Journal of Computer Theory and Engineering, vol. 2, no. 3, pp. 319–322, 2010.

[12] S. Livingstone and F. Russo, ―'e Ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, “Multimodal Set of Facial and Vocal Expressions in North American English”, vol. 13, 2018.

[13] U. Raghu Vamsi, B. Yuvraj Chowdhary, M. Harshitha, S. Ravi Eja, and J. Divya Udayan, “Speech emotion recognition (SER) using multilayer perceptron and deep learning techniques”, IEEE Access, 2021.

[14] M. Xu, F. Zhang, and W. Zhang, ―Head fusion: improving the accuracy and robustness of speech emotion recognition on the IEMOCAP and RAVDESS dataset”, IEEE Access, vol. 9, pp. 74539–74549, 2021.

[15] H. Zhang, R. Gou, J. Shang, F. Shen, Y. Wu, and G. Dai, “Pre-trained deep convolution neural network model with attention for speech emotion recognition”, Frontiers in Physiology, vol. 12, Article ID 643202, 2021.