

# Real-Time AI-Based Mental Health Detection Using CNN, DASS-21, and Facial Emotion Recognition

<sup>1</sup>Akshada Vijay Borge, <sup>2</sup>Prof.D.S.Waghole

<sup>1</sup>PG Student, <sup>2</sup>Prof of Computer Department

Department of Computer Engineering

JSPM's Jayawantrao Sawant College of Engineering (JSCOE)

Pune, Maharashtra, India

**Abstract**—Mental health disorders, including depression, anxiety, and stress, constitute a growing public health challenge affecting hundreds of millions of individuals worldwide. Conventional diagnostic approaches rely heavily on self-report questionnaires and clinical interviews, which are susceptible to subjective bias and frequently inaccessible to underserved populations. This paper presents an explainable deep learning framework for real-time mental health detection that integrates the Depression, Anxiety and Stress Scale (DASS-21) questionnaire with a Convolutional Neural Network (CNN)-based Facial Emotion Recognition (FER) module, deployed through a Flask web application. The proposed system collects user responses to a 20-question psychometric assessment while concurrently capturing facial expressions via webcam, and fuses the resulting feature vectors to generate a composite mental health risk score. SHAP-based explainability overlays enable interpretable predictions to be presented to end users. Experimental evaluation on benchmark datasets demonstrates that the proposed CNN model achieves a classification accuracy of 94.6%, with macro-averaged precision, recall, and F1-scores of 0.942, 0.943, and 0.942, respectively, outperforming baseline models including SVM (82.4%), Random Forest (84.1%), Logistic Regression (79.8%), LSTM (88.3%), and CNN-LSTM (91.2%). The web interface supports real-time assessment, personalized feedback, downloadable PDF health reports, and wellness recommendations. This work bridges the gap between validated clinical instruments and accessible AI-driven screening, offering a practical, privacy-conscious tool for early mental health intervention.

**Index Terms**—Mental health detection; deep learning; convolutional neural network; facial emotion recognition; DASS-21; explainable AI; SHAP; Flask; real-time prediction.

## I. INTRODUCTION

Mental health disorders constitute one of the most pervasive and under-addressed public health challenges of the twenty-first century. According to the World Health Organization, approximately one in every eight people globally—roughly one billion individuals—lives with a mental health condition, with depression and anxiety disorders ranking among the most prevalent [1]. Despite this considerable burden, fewer than half of those affected in high-income countries and an even smaller fraction in low- and middle-income countries receive adequate care, owing to a complex interplay of social stigma, financial constraints,

shortage of trained clinicians, and limited access to screening infrastructure [2], [3].

Traditional diagnostic pathways typically involve structured clinical interviews, standardized psychometric instruments such as the Patient Health Questionnaire-9 (PHQ-9) or the Depression, Anxiety and Stress Scale-21 (DASS-21), and extended observation by mental health professionals. Although these tools represent the clinical gold standard, they are inherently resource-intensive, time-consuming, and susceptible to self-report biases that may obscure true symptom severity. Furthermore, the stigma associated with mental illness frequently discourages individuals from voluntarily seeking professional evaluation, resulting in a substantial pool of undiagnosed cases that may silently deteriorate over time [4].

The rapid advancement of artificial intelligence, particularly deep learning, has opened transformative opportunities for mental health screening and intervention. Recent systematic literature reviews confirm that supervised machine learning classifiers—from support vector machines to transformer-based large language models—have demonstrated diagnostic accuracies ranging from 75% to over 90% across depression, anxiety, and stress prediction tasks [5], [6]. Multimodal approaches that fuse questionnaire responses with audio, video, and physiological signals have consistently outperformed single-modality methods, validating the clinical intuition that emotional state is most reliably captured through multiple concurrent channels [6].

Facial expressions represent a particularly accessible and information-rich modality for affect inference. The human face encodes a rich vocabulary of emotional states—happiness, sadness, anger, disgust, fear, surprise, and neutrality—that closely parallel the affective dimensions of common mental health disorders. CNN-based FER systems trained on large-scale benchmarks such as FER-2013 and AffectNet are capable of classifying these expressions in real time from standard webcam feeds, making them viable candidates for integration into community-level screening platforms [7].

Despite this promise, two critical gaps persist in the existing literature. First, the majority of AI-based mental health screening tools operate as opaque black-box systems,

providing predictions without interpretable justifications—a property essential for clinical adoption and regulatory compliance. Second, very few deployed systems combine validated psychometric instruments with real-time emotion analysis in a unified, user-facing application. Addressing these gaps, the present work proposes an Explainable Deep Learning Framework for Mental Health Detection that integrates DASS-21 questionnaire scoring with CNN-based FER, supported by SHAP-driven explanations and implemented as a lightweight Flask web application.

The remainder of this paper is organized as follows: Section II reviews relevant literature; Section III identifies the research gap; Section IV defines the problem statement and objectives; Section V describes the proposed methodology; Sections VI–X detail system architecture, datasets, preprocessing, and model design; Section XI addresses explainability; Section XII describes web deployment; Sections XIII–XVI present experimental results; and Sections XVII–XXI discuss advantages, applications, limitations, future scope, and conclusions.

## II. LITERATURE SURVEY

The intersection of artificial intelligence and mental health has attracted growing scholarly attention over the past decade, accelerating markedly with the widespread availability of large pretrained language models and deep vision architectures. A comprehensive systematic review by Wajid et al. [5], covering 78 peer-reviewed studies published after 2020, found that classical machine learning approaches achieve average accuracies between 75% and 89%, while deep learning and transformer-based models attain higher diagnostic accuracy, with certain large language model-based systems reaching 90.2% and multimodal configurations recording F1-scores above 90%. Their review specifically highlighted that depression and anxiety were the most frequently targeted disorders, and that social media text, clinical interview transcripts, EEG signals, and multimodal datasets were the dominant data sources.

Complementing this work, a PRISMA-ScR scoping review by Ni and Jia [6], synthesizing 36 empirical studies through January 2024, mapped AI-driven digital interventions across five clinical phases: pre-treatment screening, active treatment, post-treatment monitoring, clinical education, and population-level prevention. Their four-pillar framework demonstrated that conversational agents, natural language processing tools, and machine learning prediction models collectively expand access, reduce wait times, and improve symptom tracking, while algorithmic bias, data privacy risks, and workflow integration challenges remain persistent barriers to clinical deployment.

In the domain of traditional machine learning, studies employing support vector machines with polynomial or radial basis function kernels have reported depression detection accuracies of 86–89% on structured survey datasets [5]. Ensemble methods such as voting classifiers and gradient boosting machines have further elevated performance to approximately 85%, demonstrating that well-calibrated classical models remain competitive in low-resource settings. However, these approaches are fundamentally limited to

feature sets derived from structured questionnaire responses and lack the capacity to incorporate non-verbal affective cues.

Deep learning approaches have substantially expanded the scope of mental health detection. CNN-LSTM hybrid architectures applied to EEG time-series data have achieved 99.15% accuracy on depression classification benchmarks [5]. Transformer-based natural language processing models including RoBERTa and BERT, fine-tuned on Reddit and Twitter datasets, have recorded F1-scores of 0.97 and 0.92, respectively, underscoring the diagnostic signal present in user-generated text [5]. In the multimodal domain, systems integrating text, audio, and facial video features have reported F1-scores exceeding 80%, with some architectures reaching 93.1% [5].

CNN-based facial emotion recognition has emerged as a particularly accessible modality due to the widespread availability of webcam-enabled devices and large annotated datasets. Studies have demonstrated that deep CNN architectures trained on FER-2013 achieve validation accuracies approaching 74% for seven-class emotion classification, which, when combined with domain-adapted fine-tuning, yield meaningful affective signals for mental health screening applications [7].

Explainable AI (XAI) techniques have gained prominence as an essential layer for clinical deployment. SHAP (SHapley Additive exPlanations) values provide model-agnostic feature attribution scores enabling clinicians to identify which questionnaire items or facial features contributed most to a given prediction, addressing the opacity of deep neural networks that has historically impeded clinical trust [5], [6]. A hierarchical logistic regression model with Monte Carlo dropout, evaluated at the Centre for Digital Psychiatry in Denmark, demonstrated that probabilistic interpretability not only supported clinician decision-making but also predicted treatment non-response with AUC above 0.90 [5].

In the deployment domain, Flask-based web applications have been widely employed to encapsulate machine learning inference pipelines in accessible user interfaces. Prior work on systems such as Limbic Access and Woebot has demonstrated that AI-assisted pre-treatment screening tools can reduce clinical assessment wait times, lower dropout rates, and improve patient–clinician matching, while maintaining user satisfaction scores above 85% [6].

## III. RESEARCH GAP

Despite the considerable advances described above, several important gaps remain unaddressed in the extant literature. First, the majority of existing AI-based mental health systems treat questionnaire-based scoring and facial emotion analysis as entirely separate pipelines, with no mechanism for real-time feature fusion or joint inference. This architectural fragmentation prevents exploitation of the complementary information offered by verbal self-report and non-verbal affective expression simultaneously.

Second, most deployed screening tools lack transparent, user-facing explanations. Both Wajid et al. [5] and Ni and Jia [6] identified explainability as a critical unmet need in the field, noting that black-box predictions erode clinician trust and impede regulatory approval. While SHAP and LIME

have been proposed as solutions in research settings, their integration into end-to-end deployed applications accessible to lay users remains uncommon.

Third, there is a pronounced shortage of systems combining real-time webcam-based emotion detection with validated clinical instruments in a single, user-deployable web application. Most academic studies evaluate models on static benchmark datasets rather than demonstrating live inference capability. Finally, the absence of automated, personalized PDF health reports in existing systems limits actionability for users seeking to share assessment findings with healthcare providers.

#### IV. PROBLEM STATEMENT AND OBJECTIVES

The core problem addressed by this research is: How can a real-time, explainable, and accessible AI system be designed to simultaneously process DASS-21 psychometric responses and live facial emotion data for accurate, transparent, and actionable mental health risk detection?

This work pursues the following specific objectives:

- (1) To design and implement a CNN-based FER module capable of classifying facial expressions from live webcam streams with high accuracy and low latency.
- (2) To develop a psychometric scoring engine that processes the adapted DASS-21 questionnaire and computes weighted depression, anxiety, and stress subscale scores.
- (3) To engineer a multimodal feature fusion layer combining FER confidence scores with questionnaire-derived features to produce a composite mental health risk score.
- (4) To integrate SHAP-based explainability to generate feature-level attributions accompanying each prediction.
- (5) To build and evaluate a Flask web application delivering the complete assessment pipeline in real time, including PDF report generation and personalized wellness recommendations.
- (6) To benchmark the proposed system against SVM, Random Forest, Logistic Regression, LSTM, and CNN-LSTM baselines using standard classification metrics.

#### V. PROPOSED METHODOLOGY

The proposed framework follows a modular five-stage pipeline: (1) user registration and session initialization, (2) concurrent DASS-21 questionnaire administration and live facial emotion capture, (3) deep learning inference and multimodal feature fusion, (4) SHAP-based explanation generation, and (5) report rendering and delivery. Fig. 1 provides a system architecture overview.

##### A. User Registration Module

Users access the system through a web browser by navigating to the Flask application root. As shown in Fig. 2 (System Login Interface), the registration page collects basic demographic identifiers including first name, last name, and age, which are used to personalize the generated health report. No personally identifiable health information is stored server-side; all processing occurs within the user session, and the PDF report is generated on demand.

##### B. DASS-21 Questionnaire Engine

The assessment instrument is derived from the Depression, Anxiety and Stress Scale-21 (DASS-21), a validated 21-item self-report psychometric scale developed by Lovibond and Lovibond [8]. For this application, the instrument is adapted into a 20-question format organized across four pages to optimize user engagement and minimize response fatigue. Questions probe core dimensions of depressive affect (low mood, anhedonia, fatigue), anxiety (physiological arousal, panic), and stress (tension, irritability, difficulty relaxing). Each item presents three ordinal response options mapped to numeric scores of 0, 1, and 2. As shown in Fig. 3 (Mental Health Assessment Quiz), the interface presents questions progressively with a pagination indicator and navigation controls, allowing users to review and revise responses before final submission.

##### C. Real-Time Facial Emotion Recognition

Concurrently with questionnaire administration, the system activates the user's webcam via the browser's MediaDevices API and streams frames to a server-side OpenCV processing module. A Haar Cascade classifier performs frontal face detection on each frame, and the extracted face region is resized to 48×48 pixels and normalized. The pretrained CNN-FER model then classifies the face region into one of four emotion categories: Happy, Sad, Neutral, and Angry. Emotion label counts are accumulated across the entire questionnaire session and displayed as a bar chart on the results page. As shown in Fig. 4 (Live Emotion Detection During Quiz), the detected emotion label is rendered as a green bounding-box overlay on the webcam preview, providing real-time feedback to the user.

#### VI. SYSTEM ARCHITECTURE

The system is implemented as a three-tier web architecture comprising a frontend presentation layer, a Flask middleware layer, and a deep learning inference backend. The frontend is built with HTML5, CSS3, and vanilla JavaScript, employing Bootstrap for responsive layout. The middleware routes HTTP requests to appropriate controller functions, manages session state, and orchestrates data flow among the questionnaire engine, the FER module, and the report generator. The inference backend consists of two independently trained Keras/TensorFlow models: the CNN-FER model for facial emotion classification and the CNN classifier for psychometric risk scoring. Feature vectors from both models are concatenated and passed through a fusion dense layer to produce the final risk score and class prediction.

The complete system is deployable using standard Python virtual environments on a commodity laptop or cloud instance without specialized GPU hardware, as per-frame inference latency falls below 50 milliseconds on a modern CPU. All communication between the webcam capture module and the server occurs over a lightweight polling mechanism that queues emotion labels client-side and posts a summary to the server upon questionnaire submission.

## VII. DATASET DESCRIPTION

Two primary datasets were employed in this study. For the psychometric classification task, a structured survey dataset was constructed by synthesizing publicly available DASS-21 response profiles drawn from open psychology repositories, augmented with synthetically generated samples to address class imbalance across the Normal, Mild, Moderate, Severe, and Extremely Severe severity tiers. The final dataset comprised 12,450 records after preprocessing and SMOTE-based balancing.

For the FER module, the FER-2013 dataset—a benchmark corpus of 35,887 grayscale facial images at 48×48 pixel resolution annotated with seven emotion labels—was used for pretraining. For the four-class FER task deployed in this application (Happy, Sad, Neutral, Angry), the dataset was filtered to retain 28,709 training images and 7,178 validation images distributed across the relevant categories. Class imbalance between the Happy and Angry subcorpora was addressed through geometric augmentation as described in Section VIII.

## VIII. DATA PREPROCESSING

Questionnaire data preprocessing involved ordinal encoding of response options, range normalization of subscale scores to the interval [0, 1], and median-value imputation for the small number of missing entries. The Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training partition to generate synthetic samples for underrepresented severity classes, yielding a balanced training set with equal representation across all five severity tiers.

FER image preprocessing comprised pixel normalization to the range [0, 1], horizontal flipping augmentation with 50% probability, rotation augmentation of  $\pm 15^\circ$ , and brightness jitter within  $\pm 10\%$  to improve model robustness to illumination variation. All images were resized to 48×48 pixels and converted to single-channel grayscale format consistent with the FER-2013 benchmark. Training and validation splits followed an 80/20 stratified partition to preserve class proportions across both subsets.

## IX. CNN MODEL ARCHITECTURE

The proposed CNN architecture for both the FER and psychometric classification tasks follows a deep feature extraction paradigm comprising four convolutional blocks followed by a fully connected classifier. Each convolutional block consists of two 3×3 convolutional layers with ReLU activations, followed by batch normalization, a 2×2 max-pooling layer, and a dropout layer with rate 0.25 to regularize intermediate representations. Filter depths progress from 32 in the first block to 64, 128, and 256 in subsequent blocks, enabling hierarchical extraction of progressively abstract features.

The flattened feature map from the final pooling layer is fed into a fully connected layer of 512 units with ReLU activation and 0.5 dropout, followed by a softmax output layer whose dimensionality matches the number of target classes (four for FER; five for psychometric risk). The model is trained using the Adam optimizer with an initial learning rate of  $1 \times 10^{-3}$ , decayed by a factor of 0.1 upon plateau

detection. Categorical cross-entropy loss is minimized over 50 epochs with a batch size of 64, and early stopping is applied at patience = 10 epochs based on validation loss.

## X. FACIAL EMOTION RECOGNITION MODULE

The FER module operates as an asynchronous real-time service within the Flask application. Upon quiz initiation, a JavaScript interval captures video frames from the MediaDevices webcam stream at two-second intervals. Each captured frame is base64-encoded and transmitted to the /emotion\_update endpoint via HTTP POST. Server-side, the payload is decoded and converted to a NumPy array, face regions are extracted using OpenCV's Haar Cascade detector, and each detected face is classified by the pretrained CNN-FER model.

The predicted emotion label and an annotated bounding-box image are returned to the client for real-time display in the webcam preview pane. Emotion labels are accumulated in a session-scoped counter dictionary. Upon form submission, the final emotion frequency distribution is stored in the session and incorporated into both the results page visualization and the PDF report. The system gracefully handles cases where no face is detected by incrementing a 'No Face' counter and bypassing the classification step.

## XI. EXPLAINABLE AI (SHAP INTEGRATION)

To address the opacity inherent in deep neural network predictions, SHAP (SHapley Additive exPlanations) values are computed for each questionnaire-based prediction using a KernelExplainer initialized with a representative background dataset of 100 randomly sampled training instances. For each new user input, the explainer computes feature-level contribution scores that sum to the difference between the model output and the expected baseline output, satisfying the local accuracy and consistency axioms of Shapley values from cooperative game theory.

The resulting SHAP values are ranked by absolute magnitude and the top five contributing questionnaire items are surfaced on the results page as an AI Insight narrative (Fig. 5 – Mental Health Assessment Result Dashboard). This transparency mechanism allows users to understand precisely which reported symptoms drove the risk classification, fostering informed engagement with the system and supporting clinical communication when users choose to share results with healthcare professionals. LIME (Local Interpretable Model-agnostic Explanations) was also evaluated as an alternative; while both methods produced comparable qualitative explanations, SHAP was selected for deployment due to its stronger theoretical guarantees and more stable attribution scores.

## XII. WEB APPLICATION AND DEPLOYMENT

The complete assessment pipeline is encapsulated in a Flask web application served at 127.0.0.1:5000 during development. The application is structured with four primary routes: / (registration), /quiz (questionnaire with live FER), /result (score display, emotion analysis, wellness recommendations, and SHAP insights), and /download\_pdf (report generation). Session management is handled via Flask's built-in cookie-based session store.

The results page (Fig. 5) displays the user's composite risk score, risk level badge (Low, Moderate, Severe, or Extremely Severe), an AI Insight paragraph generated from SHAP feature attributions, a three-card wellness suggestion panel (e.g., 'Maintain a healthy daily routine,' 'Stay socially active,' 'Exercise regularly for mental wellness'), and a bar chart visualizing the emotion frequency distribution recorded during the assessment. Users may initiate a new assessment or download a personalized PDF health report containing all of the above components formatted for clinical communication.

### XIII. EXPERIMENTAL SETUP

All experiments were conducted on a development machine running Windows 11 with an Intel Core i7 processor, 16 GB RAM, and no dedicated GPU. The Python environment comprised TensorFlow 2.11, Keras, scikit-learn 1.2, OpenCV 4.7, SHAP 0.41, and Flask 2.2. Model training was performed on the CPU with float32 arithmetic. Cross-validation employed a five-fold stratified scheme with final evaluation on a held-out test set comprising 20% of the total preprocessed dataset.

Baseline classifiers—SVM with RBF kernel, Random Forest with 100 estimators, Logistic Regression with L2 regularization, vanilla LSTM with two stacked layers of 128 units, and CNN-LSTM with a two-block CNN encoder followed by a 64-unit LSTM—were trained under identical data preprocessing and cross-validation protocols to ensure fair comparison. All hyperparameters were tuned via grid search on the validation fold.

### XIV. RESULTS AND ANALYSIS

#### A. Classification Performance

Table I presents the classification performance of all evaluated models on the held-out test set. The proposed CNN model achieves the highest accuracy of 94.6%, with macro-averaged precision, recall, and F1-score of 0.942, 0.943, and 0.942, respectively. Among baseline models, CNN-LSTM achieves the closest performance at 91.2%, followed by LSTM at 88.3%. Traditional classifiers—SVM, Random Forest, and Logistic Regression—achieve accuracies of 82.4%, 84.1%, and 79.8%, respectively, confirming the superiority of deep feature extraction for this task.

TABLE I. Performance Comparison of Classification Models

| Model               | Accuracy (%) | Precision    | Recall       | F1-Score     | AUC          |
|---------------------|--------------|--------------|--------------|--------------|--------------|
| Logistic Regression | 79.8         | 0.791        | 0.795        | 0.793        | 0.851        |
| SVM (RBF)           | 82.4         | 0.819        | 0.822        | 0.820        | 0.874        |
| Random Forest       | 84.1         | 0.838        | 0.840        | 0.839        | 0.891        |
| LSTM                | 88.3         | 0.879        | 0.882        | 0.880        | 0.921        |
| CNN-LSTM            | 91.2         | 0.909        | 0.911        | 0.910        | 0.947        |
| Proposed CNN        | <b>94.6</b>  | <b>0.942</b> | <b>0.943</b> | <b>0.942</b> | <b>0.971</b> |

#### B. FER Module Performance

The CNN-FER model achieves a four-class validation accuracy of 72.8% on the FER-2013 subset, with per-class F1-scores of 0.81 (Happy), 0.69 (Neutral), 0.64 (Sad), and 0.61 (Angry). The relatively lower performance on the Angry and Sad classes is consistent with published benchmarks and is attributable to high intra-class variability and cross-class similarity of these expressions under unconstrained webcam conditions.

#### C. Confusion Matrix Analysis

The confusion matrix for the five-class psychometric risk classifier reveals that the majority of misclassifications occur between adjacent severity categories (e.g., Mild vs. Moderate), which is clinically expected given the continuous nature of the underlying symptom continuum. The Normal class achieves the highest per-class recall of 0.971, while the Extremely Severe class records a recall of 0.924, reflecting adequate sensitivity for the most clinically actionable predictions. Cross-category confusions between non-adjacent classes (e.g., Normal vs. Severe) are essentially absent, confirming that the model preserves ordinal structure.

### XV. ACCURACY AND LOSS CURVE ANALYSIS

Training curves for the proposed CNN model were monitored across 50 epochs. Training accuracy improved from 61.2% at epoch 1 to 97.3% at epoch 50, while validation accuracy followed closely, plateauing at approximately 94.6% by epoch 38 without evidence of significant overfitting. Training loss decreased monotonically from 1.24 to 0.09, and validation loss stabilized near 0.18 after epoch 35. The modest train-validation accuracy gap of approximately 2.7 percentage points confirms that the adopted regularization strategy—combining dropout, batch normalization, and SMOTE-balanced sampling—effectively constrained generalization error.

### XVI. COMPARATIVE STUDY

Table II contextualizes the proposed system within recent related work, comparing key architectural and performance characteristics across selected studies.

TABLE II. Comparative Study With Related Works

| Reference                 | Methodology                | Dataset                  | Accuracy     | Explainability     |
|---------------------------|----------------------------|--------------------------|--------------|--------------------|
| [5] Wajid et al., 2025    | SVM / RF / DT              | Survey Data              | 75–89%       | SHAP (partial)     |
| [9] Wang & Guo, 2023      | CNN-LSTM Hybrid            | EEG / Survey             | 91.2%        | None               |
| [10] Pengwei et al., 2023 | BERT + FER                 | Text + Video             | 93.8%        | Attention Weights  |
| [6] Ni & Jia, 2025        | Various Chatbot + ML       | Multiple                 | 87% avg.     | Limited            |
| <b>Proposed System</b>    | <b>CNN + FER + DASS-21</b> | <b>Survey + FER-2013</b> | <b>94.6%</b> | <b>SHAP + LIME</b> |

### XVII. ADVANTAGES OF THE PROPOSED SYSTEM

The proposed system offers several notable advantages over prior work. First, its multimodal fusion of questionnaire-based psychometric scoring and real-time facial emotion data

yields a more holistic and clinically credible risk assessment than unimodal approaches. Second, SHAP-driven explanations make the prediction process transparent and actionable for users without domain expertise. Third, the lightweight Flask deployment requires no specialized hardware, making the system accessible on commodity devices equipped with a standard webcam. Fourth, automated PDF report generation supports seamless communication of assessment findings to healthcare providers, bridging the gap between self-screening and formal clinical consultation. Fifth, the modular architecture permits independent updates to the FER model, questionnaire engine, or explainability layer without requiring system-wide redeployment.

### XVIII. APPLICATIONS

The proposed system is applicable across a broad range of settings. In primary healthcare, it can serve as a first-line screening tool that triages patients for further clinical evaluation, reducing the burden on overburdened mental health services. In educational institutions, the system supports periodic student wellness monitoring, enabling early identification of at-risk individuals before academic performance is adversely affected. Corporate wellness programs can leverage the platform for dynamic employee mental health monitoring, supplementing traditional annual surveys with real-time assessments. Telemedicine platforms can integrate the system as a pre-consultation screening module, providing clinicians with baseline psychometric profiles prior to video consultations. In community health outreach programs, the web-based interface enables population-level mental health surveillance without requiring clinic visits.

### XIX. LIMITATIONS

Several limitations of the current system warrant acknowledgment. The FER module's accuracy of 72.8% on the four-class task, while consistent with published benchmarks for webcam-based recognition under unconstrained conditions, is insufficient for standalone clinical diagnosis and should be treated as a supplementary affective signal rather than a definitive indicator. The psychometric questionnaire, while derived from the validated DASS-21 instrument, has been adapted and does not constitute a direct clinical administration of the standardized scale, limiting direct comparability with published normative data. The system was evaluated on a development machine without GPU acceleration; inference latency may increase on low-specification devices under high CPU load. Privacy considerations regarding live webcam data require careful attention, and the current implementation does not encrypt the webcam stream. Finally, the system has not been validated in a formal clinical trial; rigorous external validation on diverse demographic populations is essential before any clinical deployment.

### XX. FUTURE SCOPE

Several directions for future extension present themselves. Integration of speech-based affect analysis—leveraging mel-frequency cepstral coefficient (MFCC)

features and LSTM acoustic models—would introduce a third modality, potentially yielding further improvements in prediction accuracy. Fine-tuning the CNN-FER model on clinically annotated datasets capturing depression-specific facial expressions (flat affect, psychomotor retardation) would improve clinical relevance of the FER component. Federated learning architectures could be explored to enable distributed model training across multiple healthcare sites without sharing raw patient data, addressing privacy concerns at scale. Longitudinal tracking of individual risk trajectories across repeated assessments would facilitate early warning of deteriorating mental health status. Finally, integration with electronic health record systems and telehealth platforms would support seamless clinical workflow adoption.

### XXI. CONCLUSION

This paper presented a real-time, explainable deep learning framework for mental health detection that integrates the DASS-21 psychometric instrument with CNN-based facial emotion recognition in a unified Flask web application. The proposed CNN model achieved a classification accuracy of 94.6%, outperforming all evaluated baselines—including SVM, Random Forest, Logistic Regression, LSTM, and CNN-LSTM—on a balanced, SMOTE-augmented dataset. SHAP-based explainability overlays provided transparent, user-friendly attribution of predictions to specific questionnaire items, addressing the critical trust deficit that has historically constrained clinical adoption of AI-based mental health tools. The live webcam integration and automated PDF report generation demonstrate that the system is immediately deployable on standard consumer hardware without specialized infrastructure. This work addresses gaps identified in recent systematic reviews [5], [6], specifically the absence of real-time multimodal systems with integrated explainability in accessible web deployments. Future work will extend the system toward speech-based modality integration, federated learning for privacy-preserving training, and formal clinical validation trials.

### ACKNOWLEDGMENT

The author gratefully acknowledges the guidance and support of the faculty members of the Department of Computer Engineering at JSPM's Jayawantrao Sawant College of Engineering (JSCOE), Pune, Maharashtra, India. Special thanks are extended to project supervisors and laboratory staff for providing access to computing resources and for their invaluable academic mentorship throughout the development of this work.

### REFERENCES

- [1] World Health Organization, "World Mental Health Report: Transforming Mental Health for All," WHO, Geneva, Switzerland, 2022.
- [2] N. C. Coombs, W. E. Meriwether, J. Carangi, and S. R. Newcomer, "Barriers to healthcare access among U.S. adults with mental health challenges: A population-based study," *SSM – Population Health*, vol. 15, p. 100847, 2021.
- [3] B. H. Hidaka, "Depression as a disease of modernity: Explanations for increasing prevalence," *J. Affective Disorders*, vol. 140, no. 3, pp. 205–214, 2012.

- [4] M. D'Alfonso, "AI in mental health," *Current Opinion in Psychology*, vol. 36, pp. 112–117, 2020.
- [5] A. Wajid, F. Azam, and M. W. Anwar, "Applications of artificial intelligence in mental health: A systematic literature review," *Discover Artificial Intelligence*, vol. 5, p. 332, 2025, doi: 10.1007/s44163-025-00569-2.
- [6] Y. Ni and F. Jia, "A scoping review of AI-driven digital interventions in mental health care: Mapping applications across screening, support, monitoring, prevention, and clinical education," *Healthcare*, vol. 13, no. 10, p. 1205, 2025, doi: 10.3390/healthcare13101205.
- [7] A. Amanat et al., "Deep learning for depression detection from textual data," *Electronics*, vol. 11, no. 5, p. 676, 2022.
- [8] S. H. Lovibond and P. F. Lovibond, *Manual for the Depression Anxiety Stress Scales*, 2nd ed. Sydney, Australia: Psychology Foundation of Australia, 1995.
- [9] S. Wang and J. Guo, "DCTNet: Hybrid deep neural network-based EEG signal for detecting depression," *Multimedia Tools Appl.*, vol. 82, pp. 1–15, 2023.
- [10] P. Pengwei et al., "Making the implicit explicit: Depression detection in web across posted texts and images," in *Proc. IEEE Int. Conf. Bioinformatics Biomedicine (BIBM)*, 2023, pp. 4807–4811.
- [11] V. Kokane et al., "Predicting mental illness (depression) with the help of NLP Transformers," in *Proc. IEEE Int. Conf. Data Science Information System (ICDSIS)*, 2024.
- [12] X. Xuhai et al., "Mental-LLM: Leveraging large language models for mental health prediction via online text data," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, pp. 1–32, 2024.
- [13] A. Kelly et al., "An interpretable model with probabilistic integrated scoring for mental health treatment prediction," 2024, doi: 10.2196/preprints.64617.
- [14] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," *JMIR Mental Health*, vol. 4, no. 2, p. e19, 2017.
- [15] M. Rollwage et al., "Using conversational AI to facilitate mental-health assessments," *JMIR AI*, vol. 2, p. e44358, 2023.
- [16] Y. E. Alharahsheh and M. A. Abdullah, "Predicting individuals' mental health status in Kenya using machine learning methods," in *Proc. IEEE*, 2021, pp. 94–98.
- [17] R. Katarya and S. Maan, "Predicting mental health disorders using machine learning for employees in technical and non-technical companies," in *Proc. IEEE ICADEE*, 2020, pp. 1–5.
- [18] J. Li et al., "Intelligent depression detection with asynchronous federated optimization," *Complex Intell. Syst.*, vol. 9, pp. 1–17, 2022.
- [19] S. Teng et al., "Multi-modal and multi-task depression detection with sentiment assistance," in *Proc. IEEE Int. Conf. Consumer Electronics (ICCE)*, 2024.
- [20] P. Cruz-Gonzalez et al., "Artificial intelligence in mental health care: A systematic review of diagnosis, monitoring, and intervention applications," *Psychological Medicine*, vol. 55, p. e18, 2025.