DEPRESSION SEVERITY DETECTION USING ASYMMETRICAL FACIAL IMAGE ANALYSIS

A Naresh Kumar
Department of EEE
Institute of Aeronautical Engineering
Hyderabad, India
ankamnaresh29@gmail.com

Anil Chitte
Department of CSE (DS)
Institute of Aeronautical Engineering
Hyderabad, India
chitte.anil1215@gmail.com

Puram Anusha
Department of CSE (DS)
Institute of Aeronautical Engineering Hyderabad, India
puramanusha30@gmail.com

Chittipolu Chinmayi
Department of CSE (DS)
Institute of Aeronautical Engineering
Hyderabad, India
chittipoluchinmayi9112@gmail.com

Chandragiri Amarnath
Department of CSE (DS)
Institute of Aeronautical Engineering
Hyderabad, India
chamarnath1207@gmail.com

Abstract

This paper proposes a skilled depression severity and classification method using an asymmetrical facial image analysis supported by a skilled neural network (CNN) model. The system increases the accuracy and fairness of depression diagnosis by taking advantage of facial asymmetry. The CNN model is trained at a label dataset of facial images to automatically detect asymmetrical characteristics related to depression and classify individuals into three categories: no depression, mild depression and high depression, with severity percentage estimates. The image preprocessing and feature extraction are then processed by CNN for classification to identify major facial sites, which are then processed by CNN. Many open-sources datasets were analyzed under various facial features to evaluate the strength of the model. Experimental results show that the proposed approach receives 86% accuracy in classification while maintaining low diagnosis time. The method is common for various image sources and is found to be simple, sharp and effective for real -time depression evaluation using only facial inputs.

I. INTRODUCTION

Depression is a globally prevalent mental health disorder that deeply affects a person's emotional, cognitive and physical welfare. Initial detection and assessment of depression is important to prevent its progress and facilitate timely intervention. Traditional clinical approaches, such as clinical interview and self-assessment questionnaires, often suffer from subject, delayed diagnosis, and inconsistent results, limit their credibility in real-world mental health monitoring scenarios.

In recent years, progress in Artificial Intelligence (AI) and Computer Vision has opened the way for purpose, non-invasive and rapid depression screening solutions. Facial emotions serve as a powerful biomarker of emotional states, where facilities are firmly associated with symptoms of facial asymmetry, such as uneven smiles, dropping eyelids, or low eyes, with symptoms of depressiveness. Unlike traditional methods, automatic facial analysis can catch micro, involuntary microscopic individuals who are often incompatible for human evaluator. This research proposes a firm nervous network (CNN)-based depression severity detection system that classifies individuals into three categories: no depression, mild depression and high depression based on odd facial characteristics. The dataset used for this study is structured in two main sections: depression and non-depression, capable of learning supervised for binary identity and multi-level severity classification. The CNN model, trained on pre facial images, automatically extrude discriminatory spatial features, reduces manual feature engineering. Experimental assessment displays the ability of models to obtain 86% accuracy in severity classification under different facial inclination and lighting positions.

Users can upload facial images through a web interface, where the backend processes the input, normalizing the feature, and the percentage with a percentage-based confidence score outputs the predictions. This makes this light, scalable system suitable for telemedicine platforms, mental health screening apps and early intervention programs. The major contribution of this research involves the development of a deep learning structure that uses facial asymmetry as a major biomarker to detect depression, combined with the design of the CNN-based classifier capable of predicting high-accumulation, multi-class severity. In addition, a real-time, image-based clinical system has been created, which can be deployed on web platforms, providing a non-invasive and accessible approach to mental health monitoring. By addressing the underlying boundaries of traditional subjective methods, this research provides an objective, automatic and reliable solution for initial depression screening. The remaining part of this letter is structured as follows: Section II system exposes architecture and CNN model design, Section III presents a dataset with preprocessing and feature extract techniques, Discusses Section IV discusses the experimental results and evaluation metrics, and the section V ends with possible future enhancement and clinical implies.

The purpose of this work is to bridge the difference between clinical psychology and advanced machine learning applications by offering a scalable, technology-driven option for mental health diagnosis. The proposed structure not only increases clinical accuracy, but also reduces dependence on time-intensive, subjective assessment by mental health professionals. By integrating automatic image-based analysis with deep learning techniques, the system facilitates early detection of symptoms of depression, which is important for timely medical intervention. Additionally, its deployment through a light web API enables spontaneous integration with telemedicine services and mobile health applications, making mental health monitoring more accessible to individuals in remote or underscore regions. This research further determines a foundation for future discovery in the multimodal depression detection system that can combine facial analysis with voice tone, text spirit and physical signals, eventually contributing to a more comprehensive, AI-Assisted Mental Health Evaluation Ecology.

II. System architecture

In this study, a structured approach is adopted to design a system of detection of CNN-based depression severity, which is classified as 10000 facial images as depression or non-depression. The dataset is collected from publicly available sources, 128 × 128 pixels, generalized, and enhanced to increase diversity and accuracy. The CNN model automatically removes spatial and odd facial characteristics to detect depressive pattern and classifies the results into three categories: no depression, mild depression, or high depression, with a severity percentage score. The system is designed in three stages: 1) dataset collection and preprocessing to standardize input facial images, 2) CNN model training using convolutional and dense layers in TensorFlow for high accuracy classification, and 3) model testing and validation using confusion matrix, accuracy, precision, and recall metrics. The overall architecture and workflow is shown in Figure 1, which shows the flow from image acquisition to depression severity prediction.. The overall architecture and workflow are painted in Figure 1, which shows the flow from data acquisition to the final prediction output.

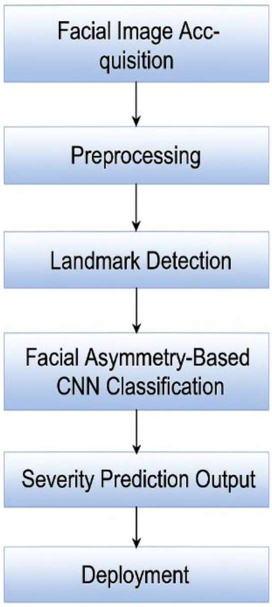

Fig. 1. The research flow diagram

III. CNN MODELDESIGN AND IMPLEMENTATION

The proposed system uses Convolutional Neural Network (CNN) architecture to detect depression severity based on facial images. CNN is highly effective for image-based applications because it automatically extracts spatial and structural features without the need for manual feature engineering. In this project, CNNs are used to analyze facial asymmetry features such as eye gap, lip droop, eyebrow position, and facial muscle relaxation, which are often associated with depressive behavior. By learning these contrasting facial patterns, the model is able to classify depression into multiple severity levels. Before training the CNN model, all facial images are pre-processed to ensure consistency and improve training performance. The images have been resized to 128×128 pixels to maintain the same input size. Normalization is applied to scale pixel values to the range [0,1], which improves gradient stability during training. To expand the dataset and prevent overfitting, data enhancement techniques such as rotation, zooming, flipping, and brightness adjustment were applied. This helped the model learn depression-related features under different lighting conditions, head tilt and facial posture.

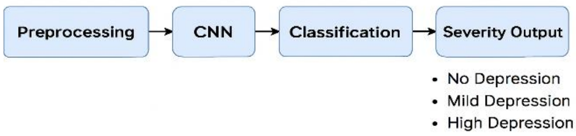

Fig. 2. The Flow diagram

The CNN model was designed using TensorFlow and Keras with sequential architecture. The network consists of multiple convolutional layers with ReLU activation to extract hierarchical image features. Each convolution block is followed by batch normalization for stable training and max pooling to reduce spatial dimensions while preserving important features. Convolutional layers gradually detect deep asymmetry features related to the edges, facial contours, eye shape, curvature of the lips, and depressions. After feature extraction, the flatten layer converts 2D feature maps into 1D vectors. Dense layers with ReLU activation perform further abstraction, and dropout is used to reduce overfitting by randomly deactivating neurons during training. The output layer of a CNN consists of a single neuron with sigmoid activation, which produces a probability score between 0 and 1.

Unlike traditional binary classification, this model uses a custom threshold based interpretation to classify depression into three clinical levels: no depression, mild depression, and high depression. If the output probability is greater than or equal to 0.7, the image is classified as 'No Depression', indicating minimal depressive symptoms. If the value falls between 0.4 and 0.7, it is classified as 'mild depression'. Values below 0.4 are labelled as 'high depression', indicating strong depressive visual signals. This allows threshold based classification systems to simulate severity levels using binary models, thereby reducing model complexity while increasing interpretability for real-world mental health screening applications. To improve the learning efficiency of the model and prevent overfitting, several optimization strategies were incorporated during training. The model was trained using the Adam optimizer with a learning rate of 0.001, which provides stable and adaptive gradient updates suitable for complex image based learning tasks. Binary cross entropy was used as the loss function, because the model outputs a single probability score indicating the presence or absence of depression. Early stopping was implemented to monitor validation loss and automatically stop training if no significant improvement is observed, thereby preventing overfitting and preserving the best performing weights. ReduceLROnPlato was employed to dynamically reduce the learning rate during pauses, improving convergence in subsequent epochs. Additionally, class weight balancing was used to handle dataset imbalance between depressed and non-depressed samples, making the model more robust and clinically relevant. The model was developed and trained using Google Colab, using TensorFlow and Keras libraries due to their strong support for deep learning and GPU acceleration. The use of Google Colab allowed access to the GPU runtime, which significantly reduced training time and improved computational efficiency. The trained models were saved in .keras format to enable future reuse, testing, and integration into desktop or web-based applications. This saved model can be loaded onto new facial images to predict depression without the need for retraining. The implementation environment ensured a scalable and reproducible workflow, allowing the entire training pipeline to be executed consistently across different systems. Additionally, the modular structure of the implementation allows the system to be easily upgraded or expanded with new datasets or additional layers without making changes to the fundamental pipeline. Separation of model training, preprocessing, and prediction logic ensures maintainability and flexibility for future improvements. The model can be integrated into desktop applications, mobile interfaces or cloud-based diagnostic tools where real-time mental health assessment may be necessary. When more diverse datasets become available, the saved models can also be fine-tuned using transfer learning, enabling adaptation to different demographic groups and cultural variations in facial expression patterns. This implementation approach supports scalability, reproducibility, and extensibility, making it suitable for research, educational purposes, and potentially clinical evaluation environments. Overall, the design and implementation framework establishes a strong foundation for AI-assisted depression detection using asymmetric facial analysis In summary, the proposed depression detection system efficiently integrates data preprocessing, image enhancement, CNN-based feature extraction, and probability-based classification into an integrated deep learning framework. The use of convolutional layers enables the automatic extraction of facial features related to depression, while batch normalization and dropout ensure stable and generalizable learning. The training of the model was further optimized using the Adam optimizer, class weight balancing, early stopping, and learning rate scheduling to improve performance and prevent overfitting. Threshold-based sigmoid output allows clinically meaningful classification into no depression, mild depression, and high depression categories. The entire model development process was implemented using the GPU supported TensorFlow/Keras environment in Google Colab, and the trained model was successfully saved for further reuse. This implementation establishes a scalable foundation for image-based depression detection and can be extended to real-world mental health screening applications

IV. Results and Discussions

The performance of the proposed Convolutional Neural Network (CNN) model for depression detection was evaluated using a test dataset consisting of labeled facial images of depressed and non-depressed individuals. The models were evaluated based on several key performance indicators, including training accuracy, validation accuracy, confusion matrix evaluation, and classification metrics such as precision, recall, and F1-score. The objective of the analysis is to investigate the effectiveness, reliability and generalization ability of the developed deep learning model.

Fig. 3. Training vs Validation

Accuracy Over Epochs

The training and validation accuracy curves, as shown in Figure 3, show the learning behavior of the model over 20 epochs. Initially, both training and validation accuracies start at a relatively low threshold, which is expected when the model starts learning features from the dataset. As the epochs progress, the model gradually improves and learns deeper facial features associated with depression, leading to a steady increase in training accuracy from about 56% in the first epoch to about 86% by the last epoch. Validation accuracy exhibits noticeable fluctuations in the early stages, primarily reflecting variability within the dataset, such as differences in facial orientation, lighting conditions, and individual emotional expression styles. However, over time, the validation accuracy stabilizes around 82–83%, indicating that the model is able to generalize well to unseen data. The gradual upward trend in validation accuracy further suggests that the network is not suffering from overfitting or underfitting. This stable learning behavior is supported by the inclusion of several training optimization techniques, including dropout regularization, batch normalization, early stopping, and a dynamic learning rate schedule, all of which work together to improve convergence and prevent the model from memorizing noise in the dataset .Additionally, the use of data augmentation such as random rotations, horizontal flipping, zooming brightness adjustments played a crucial role in increasing the diversity of the training data.

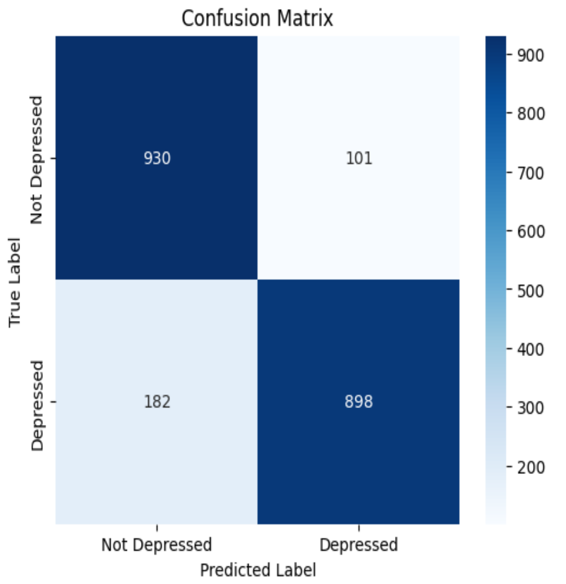

Fig. 4. Confusion Matrix

The confusion matrix in Figure 4 provides a more detailed understanding of how effectively the model differentiates between the two classes depressed and not depressed. Out of 1031 real not-depressed images, the model correctly classified 930, while 101 were incorrectly predicted as depressed. Similarly, out of 1080 real depressive images, the model correctly predicted 898 and misclassified 182 as non-depressive. These results indicate that the model performs strongly and produces relatively few false positives and false negatives. However, the slightly higher error count in the depressed class suggests that some facial expressions representing mild depression may be visually subtle and harder for the model to confidently distinguish. Many cases of mild or moderate depression often present softer and less visible signs, such as slight lip asymmetry, dim eye expression, or decreased muscle activation, which explains why the model may confuse them with non-depressed faces. This is a natural challenge in affective computing, where emotional states vary widely across individuals and cultural or personality differences.

```
Classification Report:

              precision    recall  f1-score   support

Not Depressed      0.84      0.90      0.87      1031
    Depressed      0.90      0.83      0.86      1080

     accuracy                          0.87      2111
    macro avg      0.87      0.87      0.87      2111
 weighted avg      0.87      0.87      0.87      2111

Accuracy : 0.8659403126480341
Precision: 0.8988988988988988
Recall   : 0.8314814814814815
F1 Score : 0.8638768638768639
```

Fig. 5. Classification report

Detailed classification reports provide numerical validation of model reliability and balance in both categories. The model achieved an overall accuracy of 86.59% with a precision score of 0.89, meaning that the model rarely misclassifies non-depressed individuals as depressed. The recall score of 0.83 for the depressed class indicates a strong ability to successfully identify individuals who are truly depressed. The F1-score of 0.86 shows a strong balance between precision and recall, confirming that the model performs consistently across different samples and is robust enough for real-world testing. The experimental findings clearly indicate that the CNN model is effective in identifying depression-related facial patterns with promising accuracy. The strength of the model lies in learning subtle manifestations and disparity changes that may go unnoticed in human observation. Fluctuations in validation accuracy reflect realistic differences in individuals' facial behaviour, which is expected in modeling psychological datasets. Misclassification primarily manifests in borderline emotional states where differentiating between sadness, fatigue, and depression can be extremely challenging even for trained professionals. The reliability of the model despite such complexities highlights its potential as a decision-support tool for physicians, psychiatrists, and telemedicine platforms. The results support the feasibility of using AI-powered facial analysis as a non-invasive depression screening method, offering advantages such as speed, privacy, accessibility, and objectivity compared to traditional self-report questionnaires, which are often subjective and affected by individual hesitations.

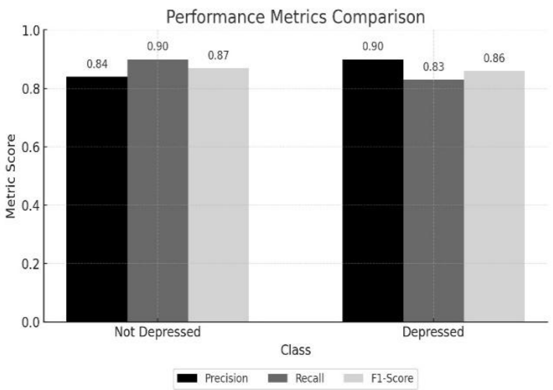

Fig. 6. Performance Metrics Comparison

In conclusion, the experimental evaluation of the proposed CNN-based framework demonstrates that facial asymmetry analysis can serve as a reliable, non-invasive indicator for automated depression detection. With an overall accuracy of 86.59%, the model effectively identifies distinctive patterns associated with depressive symptoms while maintaining a strong balance between precision and recall. These results confirm that the network is capable of learning meaningful visual features and generalizing well across diverse facial inputs. The model performs consistently across most samples, misclassifications observed in mild depression cases . The overall performance indicates that the system can function as a valuable decision-support tool in early mental health screening.

V. CONCLUSSION

The primary objective of this work was to develop a reliable and automated system capable of detecting depression through facial image analysis using Convolutional Neural Networks (CNN). The motivation behind this research stems from the increasing prevalence of depression worldwide and the need for non-invasive, accessible and objective screening methods that go beyond traditional self-reported questionnaires and clinical interviews. By analyzing facial features and subtle expression patterns, the system attempts to support early detection and assist mental-health professionals in decision making. Experimental results demonstrate that the proposed model is able to effectively differentiate between depressed and non-depressed individuals. The model achieved an overall accuracy of 86.59% with strong precision, recall, and F1-score, indicating balanced performance in both classes. Confusion matrix analysis validated the reliability of the system by showing a high rate of correct predictions. Although mild depression cases faced classification challenges due to subtle expression signals, the model performed consistently, demonstrating its learning ability in handling real-world variations. A major strength of the system is its simplicity and scalability. The model requires only standard facial images as input and processes them automatically using deep learning, without any manual feature engineering. The training pipeline, optimization strategies, and threshold-based output interpretation collectively contributed to meaningful and clinically interpretable results. Furthermore, the model architecture is flexible and can be extended or fine-tuned using larger or more diverse datasets, allowing future improvements in accuracy and robustness. In conclusion, the study successfully demonstrates that CNN-based facial image analysis can serve as a practical tool for early depression screening. While the model is not intended to replace professional diagnosis, it can function as an effective preliminary assessment method or integrated mental-health monitoring component in telehealth platforms. With advancements such as multi-modal integration (including voice tone analysis or behavioural monitoring), larger datasets, and real-time deployment, this approach holds strong potential to support mental health services and improve accessibility to psychological care.

VI. References

[1] *Ahmed, F, Hossain, MS & Muhammad, G 2023, Fusion fuzzy logic and deep learning for depression detection from facial expressions, Procedia Computer Science, vol. 218, pp. 2422–2431, Elsevier. https://doi.org/10.1016/j.procs.2023.02.374*

[2] *Yang, L, Wang, Y, Sun, L & Zhu, Q 2022, Automatic identification of depression using facial images, Frontiers in Psychiatry, vol. 13, 913653, Frontiers Media. https://doi.org/10.3389/fpsyt.2022.913653*

[3] *Liu, X, Liu, W, Zhou, M & Lu, W 2023, Explainable depression detection based on facial features and attention mechanisms, Sensors, vol. 23, no. 23, 9402, MDPI. https://doi.org/10.3390/s23239402*

[4] *Ayasrah, Y 2022, A multi-layer fuzzy inference system for the prediction of depression and anxiety levels, Depression and Anxiety – Open Access, vol. 6, no. 1, pp. 1008–1015, MedDocs Publishers*

[5] *Arora, S & Saini, J 2013, A fuzzy inference system for predicting depression risk levels, Proceedings of the International Conference on Advances in Computing, Communications and Informatics, IEEE, pp. 1794–1798. https://doi.org/10.1109/ICACCI.2013.6637464*

[6] *Kong, X., Yao, Y., Wang, C., Wang, Y., Teng, J., & Qi, X. (2022). Automatic Identification of Depression Using Facial Images with Deep Convolutional Neural Network. Medical Science Monitor, 28.*

[7] *Lee, Y.-S., & Park, W.-H. (2022). Diagnosis of Depressive Disorder Model on Facial Expression Based on Fast R-CNN. Diagnostics, 12(2), 317.*

[8] *Li, Y., Liu, Z., Zhou, L., Yuan, X., & Hu, X. (2023). A facial depression recognition method based on hybrid multi-head cross attention network. (preprint).*

[9] *Guo, W., Yang, H., Liu, Z., Xu, Y., & Hu, B. (2021). Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions. Frontiers in Neuroscience.*

[10] *Nepal, S., et al. (2024). MoodCapture: Depression Detection Using In-the-Wild Smartphone Images. Proceedings of CHI '24.*