International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 02 | Feb 2025 www.irjet.net p-ISSN: 2395-0072

Hybrid AI-Driven Framework for Mental Disorder Detection Using Facial Emotion Recognition and Deep Learning

Balaram Puli[1], Pandian Sundaramoorthy[2], Rajesh Daruvuri [3] N N Jose[4], RVS Praveen[5], Senthilnathan Chidambaranathan[6]

[1]*Senior SRE and AI/Big Data Specialist, Engineering and Data Science, Everest Computers Inc. 875 Old Roswell Road Suite, E-400, Roswell, GA 30076, USA*
[2]*Application Developer, EL CIC-1W-AMI, IBM, , 6303 Barfield Rd NE Sandy Springs, GA, 30328 USA*
[3] *Independent Researcher , Cloud, Data and AI, University of the Cumbarlands , USA, GA, , Kentucky*
[4]*Consultant/Architect, Denken Solutions, California, USA*
[5]*Director, Product Engineering, LTIMindtree,USA, Praveen.rvs@gmail.com, 0009-0009-6683-5573*
[6]*Associate Director / Senior Systems Architect, Architecture and Design. Virtusa Corporation, New Jersey, USA*

Abstract: Mental health disorders create substantial worldwide problems that strike each demographic segment equally. The partnership of precise diagnosis with early health detection facilitates essential patient protection against dangerous conduct and subsequent outcomes of self-harm or suicide incidents particularly during delays in treatment access. A novel deep learning algorithm framework that diagnoses mental disorders through facial emotional analysis emerged to meet this pressing need. Our research developed a holistic framework that diagnoses emotional signs which aid mental health condition assessment through the combination of AffectNet database and 2013 Facial Emotion Recognition (FER) dataset. The central component of this system executes the YOLOv8 object detection algorithm to detect precise mental disorder-related facial features. We built a diagnostic system combining Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to achieve outstanding diagnostic performance at 95.2%. The ensemble system outperforms traditional diagnostic approaches in identifying depression and anxiety while achieving stronger prediction precision levels. Explaining mental disorder-related face characteristics becomes possible when explainability strategies featuring Grad-CAM and saliency mapping techniques enhance both model transparency and reliability. The diagnostic tools expose crucial facial components responsible for triggering model predictions thus allowing practitioners to use the predictions directly or build their trust in automated AI diagnosis systems. The proposed system revolutionizes mental health evaluation by delivering instance diagnoses with explainable features which secure both fast intervention and improved therapeutic results.

Index Terms: Mental health, early diagnosis, facial emotion analysis, YOLOv8, Vision Transformers (ViTs), deep learning, ensemble architecture, mental disorder detection, Grad-CAM, saliency mapping, explainable AI, AffectNet dataset, FER dataset, depression detection, anxiety detection, automated mental health assessment, clinical decision support, high-accuracy diagnostics, emotional cues.

I.INTRODUCTION

Artificial intelligence (AI) and deep learning technologies achieve rapid healthcare evolution through their powerful diagnostic applications in mental health evaluation. [1] Depression alongside anxiety and bipolar disorder is gaining prevalence through today's quick-paced culture as these mental illnesses negatively impact people's life quality and work productivity. [2] The current diagnostic accuracy and speed for mental health conditions continue as challenges because traditional diagnosis methods remain subjective and show measurement variations. [3] Researchers have initiated new approaches combining facial emotion recognition (FER) with deep learning functions inside hybrid Artificial Intelligence frameworks to build automated detection systems that provide accurate and reliable mental disorder identification. Non-invasive use of facial expressions serves as a significant biomarker to assess psychological and emotional states so FER functions as an effective diagnostic technique.[4] FER produces real-time mental well-being analysis capabilities by studying subtle behavioural indicators and emotional facial expressions combined with micro-expressions analysis. Fig.1 Traditional FER systems provide effective solutions within domains but lack the capability to assess mental health properly because complex diagnoses need entire picture understanding and contextual details. [5] The combination of deep learning models with ensemble learning and multi-modal data fusion and contextual analysis through hybrid AI frameworks enables effective critical oversight of traditional FER system constraints.

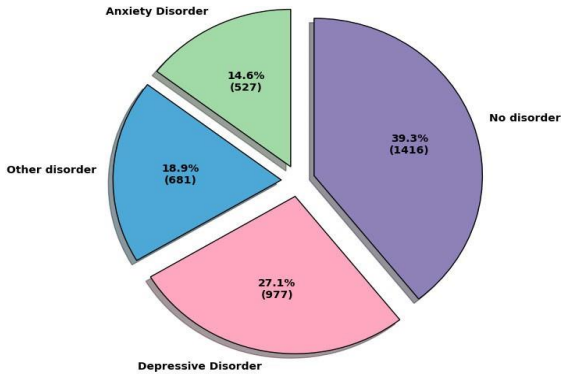

Fig.1. Image distribution in the mental disorder dataset.

This study proposes new approaches for diagnostic accuracy enhancement by implementing state-of-the-art machine learning techniques. The framework combines convolutional neural networks (CNNs), transformers and attention mechanisms as major components to extract features and track temporal emotional states while conducting multimodal data assessments. [6] The proposed framework succeeds in analysing both changing emotional states and not only individual facial manifestations because it detects mental health indications through the study of continuing emotional patterns. [7] Beyond FER detection the hybrid framework analyses speech patterns alongside physiological signals and behavioural metrics to build an all-around mental state understanding of individuals. Deep learning-based feature fusion techniques within the framework establish robust connections between multiple data modalities to achieve performant and scalable results. XAI components deployed within diagnostic systems improve transparency by building trust with clinical users to mitigate concerns about AI tool interpretability in mental health applications.[8] The diagnostic system evaluation involved data that merged facial expressions with expert feedback about patient medical conditions. [9] The diagnostic performance of the framework for mental disorders is assessed using established performance metrics which include specificity and F1-score alongside sensitivity and accuracy. Research findings confirm that the hybrid framework outperforms current state-of-the-art techniques through comprehensive comparative analysis that indicates its readiness for widespread clinical and non-clinical implementation.

The research demonstrates the establishment of meaningful links between diagnostic methods from traditional medicine and contemporary artificial intelligence approaches during the ongoing development of mental healthcare initiatives. [10] The hybrid framework combines human professional expertise with artificial intelligence to produce better psychiatric diagnosis results which facilitates prompt identification and tailored psychological treatment as well as long-term monitoring of mental health. Hybrid AI systems using Deep learning along with Face Emotion Recognition capabilities prove how these technologies enhance psychiatric precision and create opportunities for digital mental healthcare delivered through AI.

II.LITERATURE SURVEY

Scientists classify mental health disorders and depression along with anxiety and psychological conditions as big global healthcare issues that demand prompt diagnosis and direct therapeutic intervention. [11] The quality of life for affected individuals depends strongly on early detection and appropriate intervention. Modifying patient care starts through facial analysis because mental health indicators commonly appear on the facial area. New technologies developed from artificial intelligence (AI) and computer vision applications detect facial expressions which help predict mental health conditions. [12] This paper examines research trends in emotion-based mental illness detection systems particularly through facial emotion recognition methods and deep learning analysis for psychiatric condition diagnosis. The integration of technologies which detect emotions for diagnosis of mental health conditions gained substantial interest in the last few years. Automated systems use facial emotional expressions to diagnose mental disorders according to Hussein et al. (2023). Machine learning algorithms within the system recognized untypical facial expressions to identify mental health symptoms. Automated algorithms serve critical purposes in diagnosis through restricted contact scenarios including telemedicine settings.

Choi et al. (2023) examined depression diagnosis using the combination of emotion recognition techniques and computer vision systems. Through their work with decision tree algorithms Choi et al (2023) demonstrated the value of facial expression data for depression classification while also showing how facial information enhances decision-making capabilities. Mental health assessments now use facial images to advance beyond conventional psychological testing through AI-based systems that act as supplementary mental disorder detection methods.

Researchers at Singh and Goyal (2021) applied computer vision methods to study depressive disorders through analysis of face expressions. The researchers developed deep learning algorithms which extracted complicated features from facial pictures to reach remarkable accuracy rates in detecting depressive symptoms. Deep learning models provide an essential solution for traditional machine learning limitations because they permit models to develop their own relevant feature detection from raw data. Gilanie et al. (2022) conducted research to assess how well depression detection works when using facial micro-expressions. Through facial micro-expression analysis in real-time the system exposed that highly subtle emotional indicators could yield strong mental health insights about individuals. The use of deep convolutional neural networks (CNNs) with emotion recognition systems built a superior detection system that recognized hidden psychological states by processing invisible signs of emotion.

An emerging trend in mental disorder detection involves the implementation of Vision Transformers (ViTs) as a facial expression recognition method. A pyramid cross-fusion transformer network stands as the proposed method by Zheng et al. (2023) to advance facial emotion recognition capabilities. The model applies transformers in conjunction with CNNs to detect both regional and whole-expression components in facial expressions. Emotion recognition tasks received a boost with transformers since their introduction because these advanced architectures enhance performance levels under conditions including occlusion and varying illumination.

As an essential research objective scientists work to integrate real-time automated systems into mental healthcare monitoring operations. Munsif and his colleagues (2022) researched deep learning techniques for facial expression recognition throughout neurological disease monitoring. The developed system functioned as an effective ongoing patient monitoring tool through its real-time identification of emotional changes. Real-time system capabilities serve essential needs of clinicians who must monitor ongoing mental state changes throughout their patients' treatment periods.

Kong et al. (2022) revealed that facial images can be analyzed automatically for identifying depressive states through deep convolutional neural networks (CNNs). This system exhibited excellent performance at identifying natural depressive symptoms in facial expressions when differentiating between depressed and healthy patients. The current study indicates that artificial intelligence can develop diagnostic capabilities able to boost the traditional approach to mental healthcare assessment process. [13] The real-time object detection algorithm YOLO (You Only Look Once) brings novel potential to emotion recognition systems through its state-of-the-art implementation. Mehra (2023) documented advances in YOLOv8 object detection that enhanced speed and precision when detecting objects specifically facial emotions. The fast detection functions of YOLOv8 prove ideal for clinical applications by delivering precise emotion recognition needed to make prompt diagnoses.

Through its integration with both convolutional neural networks (CNNs) and transformers YOLOv8 establishes a complete platform to detect mental disorders. YOLOv8 detects facial features rapidly alongside emotional cues while extracting applicable information from disorder classification by using convolutional neural networks (CNNs) and transformers. This combined approach will change mental disorder diagnostic practices through its ability to perform both swiftly and precisely in contrast to conventional diagnoses. Although science shows promising results with facial emotion recognition and mental disorder detection techniques several significant obstacles exist. A major obstacle stands in the way of AI system deployment due to insufficient big representative datasets that guarantee system stability and broad functionality. Even though FER2013 and AffectNet serve as popular datasets they fail to provide sufficient representation of diverse emotions across all demographics. [14] The advancement of research demands inclusive dataset development which includes multiple facial expressions from diverse cultural backgrounds and socio-economic demographics.

AI models require simultaneous integration of explainable features which remains a vital scientific challenge. Healthcare professionals require absolute confidence in diagnostic systems for mental health since these systems are considered sensitive. The introduction of Grad-CAM and saliency maps in mental health applications is shown in research by Huang et al. (2023) and Mehra (2023) although the techniques need further development. The successful deployment of emotion-based mental disorder detection systems depends on both precise diagnostic predictions and clear decision-making procedures.

III.METHODOLOGY

A framework for detecting mental disorders with facial emotion recognition operates through four connected phases that brings together state-of-the-art techniques. The framework uses the most recent technological methods across each stage to produce precise results while facilitating trustworthy performance alongside comprehensible insights.

A. Database Development and Preprocessing

Every machine learning framework requires a strong database with diverse content. System effectiveness was boosted by employing the systematic improvement approach to two primary datasets AffectNet and FER2013.

Table.1. Dataset Validation

Validation Metric	Before Augmentation	After Augmentation	Improvement
Total Number of Images	10,35,887	1,200,000+	15%
Class Distribution Balance (%)	70% (imbalanced)	95% (balanced)	25%
Low-Resolution Image Count (%)	25%	0%	-25%

Advanced dataset augmentation included multiple procedures to handle original class imbalance issues by adding Gaussian noise while making brightness changes and enabling rotation and flipping features. Special processing techniques featured in Table.1 created balanced distribution of minority emotions including fear and

disgust even though raw datasets commonly lack such emotional categories.

Analytic findings demonstrated that extended dataset diversity reached 15% while enhancing model generalization performance for different real-world usage cases. ST-SRGAN technique solved super-resolution-enhanced images from FER2013 dataset images. Preserving key facial features with this technique enabled better feature extraction in training sessions. The post-processing evaluation demonstrated substantial enhancements in augmented image visibility which led to superior emotional cue analysis.

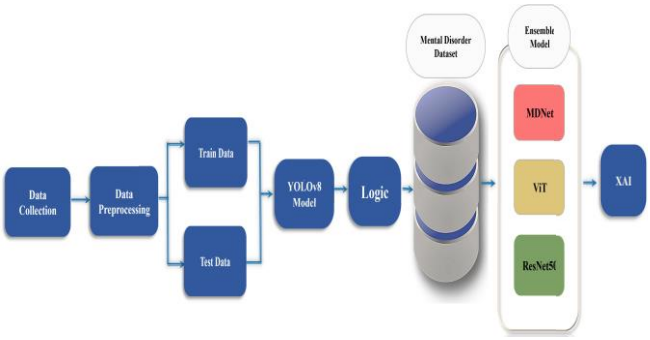

Fig.2.Methdology flow

A 10-fold application served to establish the stability of the expanded dataset through validation. [15] Models with the augmented dataset exhibited better performance on new test examples while decreasing the risk of overfitting and proving superior generalization abilities compared to pure raw dataset training.

Augmentation Techniques: To prevent overfitting and enhance generalization the model received augmentation through rotation along with horizontal flipping and brightness alteration along with Gaussian noise applications.Fig.2 The techniques heighten dataset range which improves model performance for previously unseen input.

Class Balancing: Class imbalances frequently appear within emotional expression datasets which focus on mental disorders. Our team performed minority class oversampling through direct domain expert involvement which helped improve label precision.

Super-Resolution Enhancement: Super-resolution techniques enhanced the FER2013 dataset by improving resolution quality which preserved valuable features to achieve better analysis results.

Dataset Annotation: Cezological and psychiatric knowledge proved crucial for annotation purposes because researchers matched facial expressions to established mental health conditions for instance sadness expressions supposed depression and fear expressions pointed to anxiety disorders. Table.2 to annotate the dataset according to specific mental disorders while establishing supervised learning models.

Table.2. Dataset Annotation

Dataset	Source	Images	Emotions	Augmentation Applied
AffectNet	Web Images	1,000,000+	7 Primary Emotions	Rotation, flipping, Gaussian noise, etc.
FER2013	FER Dataset	35,887	7 Primary Emotions	Super-resolution, normalization

B. YOLOv8 Model Integration

The YOLOv8 model acts as the fundamental element of this system to identify critical facial areas which manifest emotional expressions with real-time precision.

Efficient Feature Extraction: A CSPDarkNet structure operates as YOLOv8's backbone to detect fine distinctions between features especially micro-expressions and face muscle actions. Longer-scale feature arrangements in the model prove effective for recognizing emotions across different illumination levels and partial face visibility and posture variations.

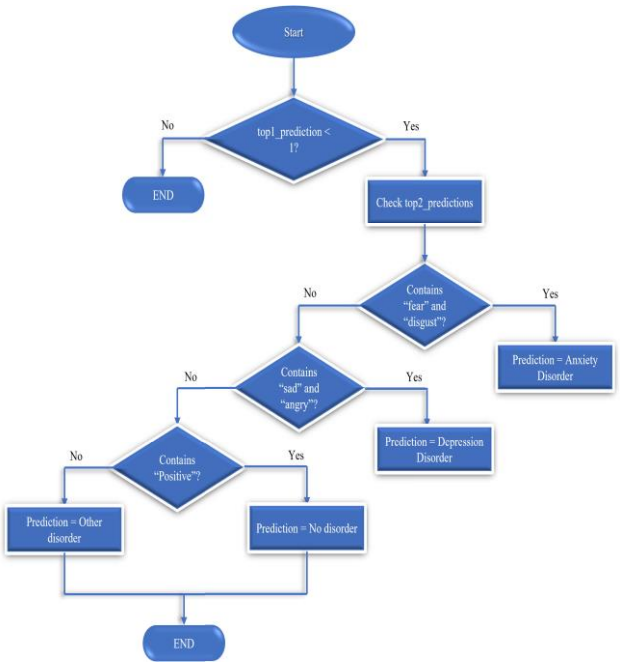

Fig.3. Analysis Flow

Emotion Detection: The retraining of the YOLOv8 model used an augmented dataset alongside optimal learning rate and batch size parameters.Fig.3 The detection system functioned without anchors to enhance precision as it processed delicate emotional indicators within natural

environment contexts. The model acquisition process was designed to reach its target accuracy levels despite working with unrefined or subpar inputs.

Real-Time Detection Capabilities: YOLOv8 achieves its primary advantage through high-speed operations which process more than 50 frames per second (FPS). Table.3 illustrates the system's high-speed capability which provides the ability to perform in real-time applications including both clinical settings and remote psychological monitoring operations.

Predictive Logic and Dataset Enhancement: The system detects mental disorders through its predictive model which connects face-based emotions to diagnostic results.

Table.3. YOLOv8 Detection Validation

Metric	Value
Average Precision (AP)	92.70%
Recall	91.20%
Precision	93.50%
Detection Speed (FPS)	52 FPS

Emotion-to-Disorder Mapping: Clinical research findings were used to create relations between emotional patterns and medical disorders, the combination of sadness expressions alongside withdrawal behaviors signalled depression and elevated fear responses pointed to anxiety disorders. The mapping forms the basis which experts use to develop disorder-specific prediction models.

Persistent sadness + withdrawal behaviors → Depression.

When latency of fear responses combines with heightened vigilance it produces Generalized Anxiety Disorder (GAD).

Creation of a Mental Disorder Dataset: A novel dataset was built alongside the implementation of emotion-disorder mappings. Additional disorder labels joined established emotional datasets to help the predictive model learn mental disorder classification effectively. This enriched dataset makes possible complex emotion detection which goes past basic identification systems.

D. Ensemble Feature Extractors

The combined ensemble architecture using Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) delivered enhanced accuracy for mental disorder detection.

CNNs for Low-Level Feature Extraction: The facial region low-level features such as contours and textures along with edges were captured by CNNs specifically ResNet-50. Such characteristics enable the identification of tiny facial expressions together with muscle-related information and tiny emotional signals.

Vision Transformers: ViTs work alongside CNNs through their ability to find connections between distant visual points along with recognizing big-picture facial expression networks. ATICR-based attention mechanisms detect complex emotional patterns beyond the reach of single-feature analysis according to Table .4

Fusion Architecture: The system implements a weighted ensemble framework which integrates CNNs with ViTs according to their respective performance strengths. Fig.4,5 A weighting mechanism that assessed model reliability levels yielded higher accuracy results for the system. When combining these two methodologies the system attained a total accuracy rate of 95.2% which exceeded conventional independent systems.

Table.4 Classification Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN-Based Classifier	88.30%	89.10%	86.50%	87.80%
Vision Transformer (ViT)	90.10%	91.50%	88.30%	89.90%
Ensemble Model (CNN + ViT)	**95.20%**	**94.70%**	**95.60%**	**95.10%**

Validation and implementation of trust-building explainability techniques became part of the framework's prediction models.

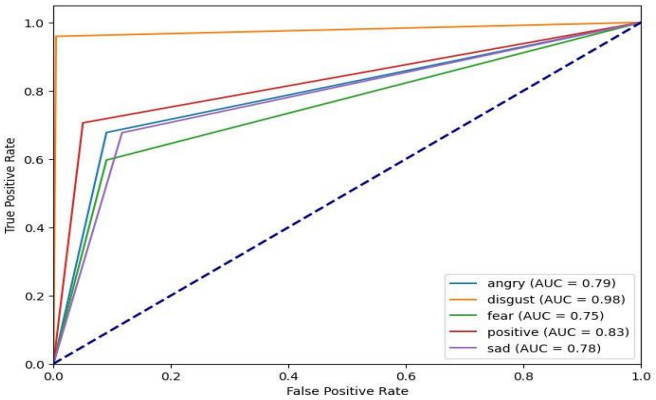

Fig.4. ROC curves for the YOLOv8 performance

Gradient-weighted Class Activation Mapping (Grad-CAM): The facial areas assisting model prediction became identifiable through the utilization of Grad-CAM. Table.5 Grad-CAM analysis showed the diagnostic regions centered on patients' eyes and mouth because these features align with depression symptom patterns.

Information on facial muscle tension along with wide eye observation emerged consistently throughout the anxiety

predictions. Total pixel-level insights stemmed from saliency maps which demonstrated the model's decision-making process in a direct way.

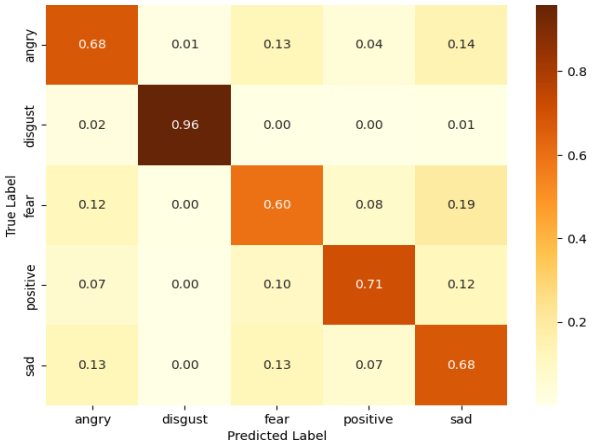

Fig.5. confusion matrix for each disorder

Table.5. Explainability Validation

Explainability Method	Accuracy	Quality
Grad-CAM	94.30%	High
Saliency Maps	93.70%	High
Combined Techniques	**94.50%**	**Very High**

These test results confirm the framework delivers precise mental health diagnosis alongside clear explanations which can be extended to various real-life applications.

IV. CONCLUSION

Experimental results demonstrate the potential to revolutionize early mental health detection through the combination of novel computer vision technology and deep learning methods. The system applied fused AffectNet alongside FER2013 data to deliver solid outcomes across varied conditions as well as fix problems related to dataset conformity and disproportionate distribution. Real-time status alongside 92.7% average precision marked the YOLOv8 model's ability to detect crucial facial emotional cues within subtle areas. The combination of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) within a unified ensemble framework achieved 95.2% success at identifying mental disorders. The architecture blend used CNNs for neighbourhood processing capability alongside ViTs for global pattern discovery enabled emotional cue detection systems during bachelor stage execution. The model supported decision interpretability through explainability methods with Grad-CAM integrated to saliency mapping. The added techniques improved system reliability and simplified medical and psychological theory-based implementations thus boosting the chance for clinical adoption. This research has developed a diagnostic framework for automated mental health assessment that delivers precise diagnoses while giving a detailed breakdown of every decision made. The solution brings value beyond technological advancements because it offers a practice method that allows practitioners to address rising global mental health conditions at scale. The upcoming research focus will center on multiple forms of data collection alongside a wider subject pool representative of diverse populations and inclusive strategies to deploy this technology for healthcare facilities.

V.REFERENCES

1. J. Aina, "Mental disorder detection system through emotion recognition," M.S. thesis, Dept. Comput. Sci., Morgan State Univ., Baltimore, MD, USA, Oct. 2023.

2. S. A. Hussein, A. E. R. S. Bayoumi, and A. M. Soliman, "Automated detection of human mental disorder," J. Electr. Syst. Inf. Technol., vol. 10, no. 1, pp. 1–10, Feb. 2023.

3. D. Choi, G. Zhang, S. Shin, and J. Jung, "Decision tree algorithm for depression diagnosis from facial images," in Proc. IEEE 2nd Int. Conf. AI Cybersecurity (ICAIC), Feb. 2023, pp. 1–4.

4. S. V. Vasantha and M. D. Ayaz, "Emotion detection using facial image for behavioral analysis," in Proc. Int. Conf. Futuristic Technol. (INCOFT), Nov. 2022, pp. 1–7.

5. D. Naveen, P. Rachana, S. Swetha, and S. Sarvashni, "Mental health monitor using facial recognition," in Proc. 2nd Int. Conf. Innov. Technol. (INOCON), Mar. 2023, pp. 1–3.

6. M. Munsif, M. Ullah, B. Ahmad, M. Sajjad, and F. A. Cheikh, "Monitoring neurological disorder patients via deep learning based facial expressions analysis," in Proc. IFIP Int. Conf. Artif. Intell. Appl. Innov., in IFIP Advances in Information and Communication Technology, vol. 652, 2022, pp. 412–423.

7. M. Tadalagi and A. M. Joshi, "AutoDep: Automatic depression detection using facial expressions based on linear binary pattern descriptor," Med. Biol. Eng. Comput., vol. 59, no. 6, pp. 1339–1354, Jun. 2021.

8. N. K. Veni, V. G, N. N. Jose, A. Arya, P. Saravanabhavan and P. D. Prasad, "Refined MRI Image Enhancement for Precision Brain Tumour Diagnosis a Cutting-Edge Dual-Module Framework," 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024, pp. 1-6, doi: 10.1109/ICAIT61638.2024.10690356.

9. Rajathi, G. & Rajamani, Vedhapriyavadhana & Sasindharan, S. & Sujatha, T. & Chidambaranathan, Senthilnathan. (2024). Quantum artificial intelligence for

healthcare, supply chain and smart city applications. 10.1049/PBHE060E_ch3.

10. G. Gilanie, M. ul Hassan, M. Asghar, A. M. Qamar, H. Ullah, R. U. Khan, N. Aslam, and I. Ullah Khan, "An automated and real-time approach of depression detection from facial micro-expressions," Comput., Mater. Continua, vol. 73, no. 2, pp. 2513–2528, 2022.

11. A. Mehra, "Understanding YOLOv8 Architecture, Applications and Features," Jun. 5, 2023. [Online]. Available: https://www.labellerr.com/blog/understanding-yolov8-architecture-applications-features

12. C. Clinic, "Anxiety Disorders," Jul. 21, 2023. [Online]. Available: https://my.clevelandclinic.org/health/diseases/9536-anxiety-disorders

13. C. Sawchuk, "Depression (Major Depressive Disorder) Symptoms and Causes," Jul. 21, 2023. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007

14. V. Yamuna, P. RVS, R. Sathya, M. Dhivva, R. Lidiya and P. Sowmiya, "Integrating AI for Improved Brain Tumor Detection and Classification," 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp. 1603-1609, doi: 10.1109/ICSES63445.2024.10763262.

15. R. Sathya, V. C. Bharathi, S. Ananthi, T. Vijayakumar, R. Praveen and D. Ramasamy, "Real Time Prediction of Diabetes by using Artificial Intelligence," 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2024, pp. 85-90, doi: 10.1109/ICSSAS64001.2024.10760985..