Cognitive Assistance for Prosopagnosia Patients Using Landmark-Based Identity and Emotion Recognition

Cognitive Assistance System for Prosopagnosia Patients

Bhavana Nagaraj*, Rajanna Muniswamy
Department of Information Science and Engineering, Vemana Institute of Technology, Kormangala, Bengaluru, 560 034, India

***Abstract*—Prosopagnosia is a neurological condition that significantly affects the social interaction and quality of life of individuals. Existing assistive systems mainly focus on either face identity recognition or face emotion recognition, limiting their effectiveness in cognitive assistive scenarios. To overcome these, an integrated framework is needed that jointly addresses identity and emotion recognition to efficiently support prosopagnosia patients, as discussed in this study. The proposed system includes two separate modules: a face identity recognition module and a face emotion recognition module. The proposed system detects and aligns faces using Multi-Task Cascaded Convolutional Networks (MTCNN) with five-point landmark alignment. Face identity recognition is performed using an EfficientNet-B3 backbone to extract 512-dimensional facial embeddings, which are matched against a Structured Query Language (SQL) database using cosine similarity. Then, facial landmarks are detected along with emotion recognition, using the Dlib library, and are structured as a graph for High-order Graph Attention Network (HoGAN)-based relational interactions detection. The proposed system is trained using a joint loss function to efficiently provide real-time assistive feedback. The system achieves high recognition performance, with AUC values of 99.8% and 99.5% in both face identity recognition and face emotion recognition modules.**

***Keywords*—Cognitive assistance system; EfficientNet-B3; emotion recognition; face identity recognition; prosopagnosia**

I. Introduction

Prosopagnosia is an impaired ability to identify individuals by their faces, which complicates recognition of previously met individuals' faces, particularly unfamiliar faces, and also recognition of familiar faces [1],[2]. Individuals with prosopagnosia struggle to identify friends, family, colleagues, and, in some cases, their own face in the mirror. This results in psychosocial difficulties, including social anxiety, embarrassment, and problems in interpersonal relationships [3],[4]. These negative psychosocial consequences impact the quality of life of individuals due to their face recognition difficulties [5],[6]. Prosopagnosia patients also suffer from other aspects of cognition, including interpreting facial expressions and emotional cues, which are crucial for effective social communication [7],[8]. Furthermore, facial expressions are a significant form of non-verbal communication that can be utilized to gather someone's emotional state and likely intentions [9],[10]. These consequences lead to the significance of developing an effective face recognition and emotion recognition system for prosopagnosia patients, thereby optimizing daily activities and quality of life [11],[12].

Advanced neurocognitive models of face processing suggest that face identity and expression analyses proceed along distinct pathways. Individuals who experience acquired prosopagnosia are unable to recognize identity and are able to recognize expression, and those who experience impaired processing of expressions are able to recognize identity [13],[14]. Early approaches focused on feature-based and holistic face recognition techniques, while recent systems leverage Deep Learning (DL) models to extract discriminative facial embeddings for identity matching [15],[16]. Computer vision research is also helpful for facial emotion recognition performed through Deep Convolutional Neural Networks (DCNNs) [17],[18]. However, there is a limited emotion recognition system in a large sample of prosopagnosics using sensitive tests in recent years [19],[20]. Moreover, existing emotion recognition systems are typically designed for generic applications and are not optimized for real-time cognitive assistance mechanisms [21].

Existing DL-based approaches analyzed the face recognition and facial emotion recognition separately. Face identity recognition approaches focus on extracting discriminative embeddings using a Convolutional Neural Network (CNN). Existing emotion recognition approaches classify the facial expression through spatial and temporal facial features [22]. Despite these advancements, identity recognition and emotion recognition are treated as separate problems in current assistive technologies [23],[24]. The lack of an integrated framework jointly providing identity and emotional information limits the practical utility of these systems. Furthermore, existing solutions lack robustness analysis under unconstrained conditions, such as visually similar faces, partial occlusion, and non-frontal poses. To overcome these, this study aims to develop an integrated system that simultaneously performs face identity and emotion recognition to efficiently support prosopagnosia patients. The key contributions of the proposed work are summarized as follows:

- Unified Cognitive Assistance Framework: The study develops a supervised DL framework that jointly performs face identity recognition and facial emotion recognition, designed to support prosopagnosia patients in real-world social interactions.

- Efficient Identity Recognition Module: An EfficientNet-B3-based embedding extraction approach is employed to generate discriminative facial representations, enabling accurate identity retrieval through cosine similarity matching with a Structured Query Language (SQL)-based identity database.
- Landmark-Based Graph Emotion Module: A facial emotion recognition module is proposed that leverages a High-order Graph Attention Network (HoGAN) to model spatial and relational dependencies among emotion-relevant facial regions, enabling accurate and real-time emotion classification.
- Joint Learning Strategy: The system is trained using a combined loss function incorporating triplet loss for identity discrimination and categorical cross-entropy loss for emotion classification, ensuring balanced optimization of both tasks.
- Robust Real-Time Assistive Feedback: The proposed framework provides real-time integrated feedback combining identity and emotional information and demonstrates robustness under challenging conditions such as pose variation, occlusion, and visually similar faces.

The remaining part of the study is organized as follows: Section II reviews the existing works related to face identity and emotion recognition, Section III describes the workflow of the proposed unified deep learning framework, Section IV delivers the experimental results of the proposed framework, and Section V concludes the study.

II. Literature Survey

This section reviews existing works related to face identity recognition and face emotion recognition, which aim to support the proposed framework in cognitive assistive systems to support prosopagnosia patients.

A. Face Identity Recognition Approaches

Volfart et al. [25] validated two behavioral tests of Face Identity Recognition (FIR) using natural and ambient face images. The approach assessed recognition of famous identities without requiring verbal output and evaluated face identity matching across different views of familiar and unfamiliar faces. However, the approach needed pictures of famous faces for the population test. Olivares et al. [26] investigated how individuals with prosopagnosia process unfamiliar faces, focusing on the different and common neurocognitive mechanisms. The approach examined external and internal facial features in face-feature matching tasks and identified face perception and memory formation. Li [27] developed an integrated framework that combines Multi-feature Local Binary Pattern (MLBP), Histogram of Oriented Gradients (HOG), and Gray Level Co-occurrence Matrix (GLCM) features into a fusion algorithm (MLBP-HOG-G) to develop an accurate face recognition model. The framework addressed challenges like lighting changes, shadows, facial occlusion, and pose variations in unconstrained environments. Jain et al. [28] created an Augmented Reality (AR)-based 3D face recognition system as a social aid for individuals with face blindness that recognizes and identifies people in social contexts using a CNN with augmented reality. The approach supported prosopagnosia patients by providing real-time identification of faces during social interactions. Mukhiddinov et al. [29] introduced an assistive technology for the visually impaired individuals. The study utilized DL algorithms to design a smart glass system to enhance the quality of life of face-blindness patients. Still, the model incorrectly detects the small objects.

B. Face Emotion Recognition Approaches

Tsantani et al. [30] established that developmental prosopagnosia impairs facial expression recognition under challenging conditions, delivering new insights into the impairment in face processing. However, the study depended on a relatively small sample size of prosopagnosia participants. Wang et al. [31] investigated how neurotypical adults recognize and imitate dynamic emotional expressions and how these abilities are related to autistic traits. The study investigated emotion recognition and examined expression imitation that provided insights into how subtle variations in autistic traits influence social-emotional processing. Bardak and Temurta [32] created a model for identifying known and unknown faces using a regional brain perspective and simple neural networks. The features were classified using the K-Nearest Neighbors (KNN) algorithm, Probabilistic Neural Networks (PNN), and Support Vector Machine (SVM) to accurately identify the faces. However, the model needed enhancements to improve the accuracy through executing DL models. Guresli et al. [33] developed a Custom Lightweight CNN-based Model (CLCM) that optimized to reduce computational complexity while maintaining strong performance in emotion recognition tasks. However, the model did not accurately predict all the emotional states. Talaat et al. [34] developed a facial emotion recognition framework through DL techniques to accurately detect and classify emotional states. However, the model used only a limited amount of data to classify the emotional states. Mukhiddinov et al. [35] developed an emotion recognition system that utilized CNN-based models to detect and classify emotions. The system assisted visually impaired individuals to enhance social interaction and communication in daily life. However, in this study, facial landmark features were not correctly obtained.

C. Problem Statement

The analysis of existing works shows that existing face identity recognition approaches focus on improving feature discrimination, fail to identify the emotional state, as listed in Table I. Also, the facial emotion recognition systems only classify the emotional state and fail to identify the face. These limit their usefulness for assistive cognitive applications. Furthermore, most optimization strategies do not exploit shared facial representations through joint learning frameworks. Also, many existing methods are evaluated under controlled conditions and lack robustness against real-world challenges such as pose variations, occlusion, visually similar faces, and dynamic social environments. To overcome these, there is a need for an integrated system that simultaneously delivers identity recognition and emotion understanding in real-time. This supports prosopagnosia patients during daily social interactions.

TABLE. I. ANALYSIS OF EXISTING WORKS

References	Models / Approaches	Merits	Demerits
Volfart et al. [25]	Face identity recognition tests using natural and ambient face images	Evaluated face identity recognition without requiring verbal responses	Relied on images of famous faces.
Olivares et al. [26]	Neurocognitive analysis of unfamiliar face processing	Provided insights into face perception in prosopagnosia individuals	It doesn't provide a practical solution for prosopagnosia patients
Li [27]	MLBP-HOG-GLCM	Robust to illumination changes, shadows, occlusion, and pose variations	Features limit scalability and performance.
Jain et al. [28]	3D face recognition using CNN	Enabled real-time social assistance for prosopagnosia patients	High computational complexity
Mukhiddinov et al. [29]	Smart glass system using DL	Improved quality of life for face-blind individuals	Poor detection of small objects
Tsantani et al. [30]	Experimental study on facial expression recognition	Provided insights into emotion recognition impairments in prosopagnosia	Limited sample size
Wang et al. [31]	Dynamic emotion recognition and imitation analysis	Studied the relationship between emotion recognition and autistic traits	Doesn't suitable for prosopagnosia patients
Bardak and Temurta [32]	KNN, PNN, SVM	Attained a better performance	Lack resulting in lower recognition accuracy
Guresli et al. [33]	CLCM	Reduced computational complexity	Failed to accurately recognize all emotional states
Talaat et al. [34]	DL-based emotion recognition framework	Achieved accurate emotion classification	Trained on limited datasets
Mukhiddinov et al. [35]	CNN-based emotion recognition assistive system	Enhanced social interaction for visually impaired individuals	Facial landmark features were not accurately extracted.

III. MATERIALS AND METHODS

The proposed methodology develops a unified deep learning framework to provide cognitive assistance for prosopagnosia patients, as illustrated in Fig. 1. The proposed system was separated into two modules: a face identity recognition module and an emotion recognition module using the VGGFace2 and FER2013 datasets. The proposed system first detects and aligns input images using a Multi-Task Cascaded Convolutional Neural Network (MTCNN).

In the face recognition module, the aligned image is passed through an EfficientNet-B3 backbone that generates 512-dimensional face embeddings. These embeddings are stored in the SQL database, and cosine similarity is used to match the extracted embedding with stored identities. In parallel, the emotion recognition module operates on the same aligned face by detecting the facial landmarks using the Dilb landmark detector. Node features are derived from EfficientNet-B3 feature maps and processed using a HoGAN to learn relational interactions among emotion-relevant facial regions. Then, identity and emotion recognition modules are optimized using a joint loss function. Finally, the system integrates the retrieved identity information with the predicted emotional state to generate real-time assistive feedback that efficiently supports prosopagnosia patients.

A. Dataset Description

In this study, two publicly available benchmark datasets are utilized to support the tasks of face identity recognition and face emotion recognition, as shown in Fig. 2. First, the VGGFace2 dataset [36] is utilized, which is a large-scale face dataset designed to capture substantial intra-class variations. The dataset includes 3.31 million images of 9131 subjects, with wide diversity in pose, age, illumination, ethnicity, and background conditions. This dataset is suitable for learning robust facial embeddings and provides accurate face recognition in unconstrained environments. For emotion recognition, the FER2013 dataset [37] is used to train the emotion recognition module.

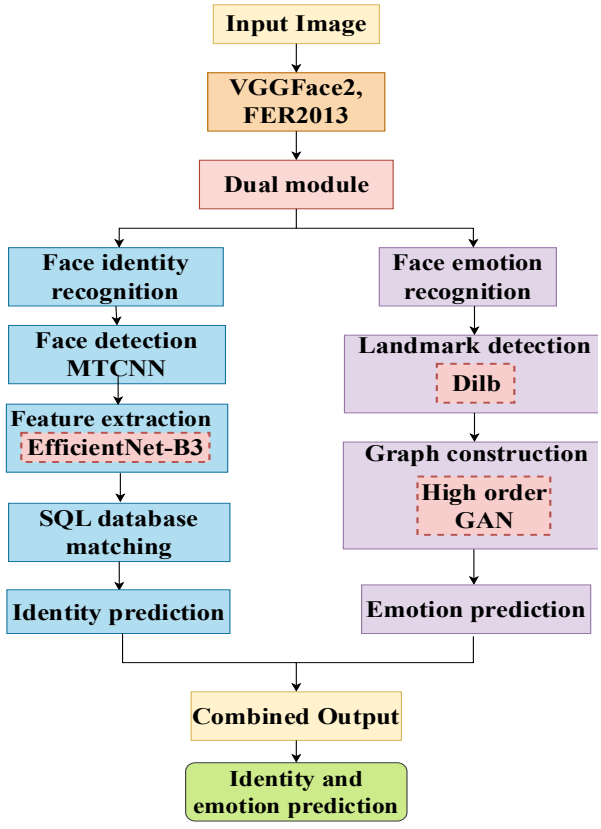

Fig. 1. Schematic representation of the proposed unified deep learning framework.

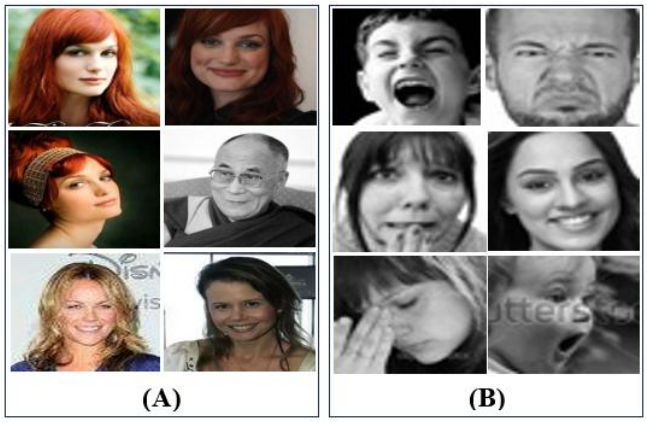

Fig. 2. Sample images from the dataset. (A) represents the sample images from the VGGFace2 dataset, and (B) represents the sample images from the FER2013 dataset.

The data consists of 48x48 pixel grayscale images of faces. The dataset consists of approximately 35,000 grayscale facial images, annotated into seven basic emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. The dataset includes images captured under diverse conditions with variations in facial expressions, occlusion, and illumination, making it suitable for learning generalized emotion representations. In this work, FER2013 is used only during the training phase to enable the model to learn emotion-specific facial patterns.

B. Face Detection and Alignment Using MTCNN

Face detection and alignment in the proposed framework are performed using the MTCNN [38], which consists of three parts: Proposal Network (P-Net), Refinement Network (R-Net), and Output Network (O-Net). The P-Net generates the image at multiple scales to generate a set of candidate face bounding boxes.

These bounding boxes are utilized to classify face and non-face windows, to estimate bounding box regression vectors for face location. R-Net rejects false candidates from the P-Net. Lastly, O-Net outputs five facial landmarks: the left and right mouth corners, the center of the nose, and the centers of the left and right eyes.

These landmarks are used to perform face alignment by estimating a similarity transformation that normalizes facial pose, scale, and rotation. The optimal transformation matrix T, is obtained by minimizing the alignment error between detected landmark positions P^*, as denoted in Eq. (1).

$$min \sum_{k=1}^{5} ||T(i_k, j_k) - (i_k{}^*, j_k{}^*)||_2^2 \quad (1)$$

where, k denotes the index of facial landmarks and (i_k, j_k) represents the detected coordinates of the k-th facial landmark. The aligned face is then cropped and resized to a fixed resolution of 224×224 pixels. This process improves the stability and accuracy of subsequent identity and emotion recognition modules.

C. Face Identity Recognition Module

The face identity recognition module is employed to extract discriminative facial representations of faces and identify individuals in unconstrained environments. It utilizes EfficientNet-B3 to generate facial embeddings. This matched against a stored identity database using cosine similarity to predict known or unknown identities.

1) Feature extraction through EfficientNet-B3: In this study, we utilized an EfficientNet-B3 [39] to extract deep features in the aligned images, as illustrated in Fig. 3. The pretrained model used is trained on ImageNet and is utilized as the convolutional backbone with all its parameters frozen. This increases a network's breadth, depth, and resolution and scales each dimension. The performance of models is improved when each size is scaled. EfficientNet-B3 uniformly balances network depth, width, and input resolution, as expressed in Eq. (2) – (4).

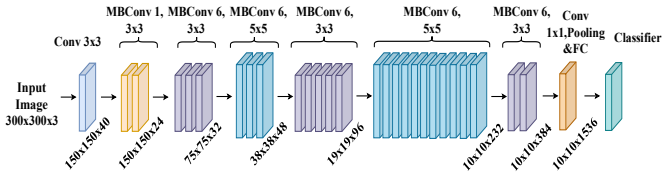

Fig. 3. Architecture diagram of the EfficientNet-B3.

$$depth\_d = \alpha\theta \quad (2)$$

$$width\_w = \beta\theta \quad (3)$$

$$resolution\_r = \gamma\theta \quad (4)$$

where, $\alpha.\beta^2, \gamma^2$, where θ controls the overall model size and α, β, γ are scaling coefficients. The network is composed of Mobile Inverted Bottleneck Convolution (MBConv) blocks augmented with squeeze-and-excitation (SE) attention. This enhances channel-wise feature recalibration. For an intermediate feature map $X \in R^{H*W*C}$, the MBConv block first expands the channel dimension using a pointwise convolution. Then, a depth-wise convolution is used to efficiently model spatial information. Channel importance is then computed through global average pooling, as denoted in Eq. (5).

$$g = \frac{1}{hw} \sum_{u=1}^{h} \sum_{v=1}^{w} X(u, v) \quad (5)$$

where, g denotes the channel-wise descriptor vector, h denotes the height of the feature map X, and w denotes the width of the feature map X. The SE module generates channel attention weights, as mentioned in Eq. (6).

$$SE = \delta(W_2 \vartheta(W_1 g)) \quad (6)$$

where, δ and ϑ denotes the ReLU and sigmoid activation functions, respectively. W_1 and W_2 denotes the weight matrix of the fully connected layer. These weights are used to recalibrate the feature map, enabling the network to emphasize identity-relevant facial regions. The final convolutional feature maps produced by EfficientNet-B3 are aggregated using global average pooling to generate a compact feature vector f. This vector is then projected through a fully connected layer to obtain a fixed-length 512-d embedding (emd), as expressed in Eq. (7).

$$emd = W_f f + b_f \quad (7)$$

where, $W_f f$ denotes the learnable weight matrix and b_f signifies the corresponding bias vector. This encodes high-level facial characteristics suitable for identity discrimination. These embeddings serve as robust feature representations for subsequent identity matching and emotion analysis, balancing accuracy and efficiency in real-time facial recognition tasks.

2) SQL database matching for identity prediction: After extracting the 512-dimensional facial embedding from the EfficientNet-B3 feature extractor, identity prediction is performed through a SQL-based identity database [40]. During the enrolment phase, embeddings corresponding to known individuals are computed and stored in the database along with unique identity labels. During inference, the embedding obtained from a query face is compared against all stored embeddings e_s in the database using cosine similarity. The maximum similarity score is selected, and if it exceeds a predefined open-set threshold τ, the corresponding identity is assigned. The optimal threshold was selected by minimizing the Equal Error Rate (EER). This ensures a balanced trade-off between false acceptance and false rejection rates. Based on this analysis, τ was fixed at 0.65, which provided the best separation between known and unknown identities while maintaining stable real-time performance. Otherwise, the face is labeled as unknown. The identity associated with the maximum similarity score is selected as the predicted identity. The SQL database enables efficient storage, indexing, and retrieval of identity embeddings. This allows scalable and real-time identity matching. This structured matching mechanism allows reliable identity prediction while supporting dynamic updates, making it suitable for practical assistive applications for prosopagnosia patients.

D. Face Emotion Recognition Module

The face emotion recognition module is employed to identify the emotional state of an individual. It utilizes a landmark-based graph construction through Dilb and a HoGAN to capture spatial and relational dependencies among facial regions. This enables accurate real-time emotion classification to support prosopagnosia patients.

1) Facial landmark detection using Dilb: In the facial emotion recognition module, the proposed system performs landmark detection using the Dilb facial landmark detector [41]. This identifies 68 anatomically consistent key points distributed across critical facial regions. The detected landmark set is represented in Eq. (8).

$$L = \{(a_x, b_x)\}_{x=1}^{68} \tag{8}$$

where, (a_x, b_x) represents the two-dimensional coordinates of the x-th landmark. These landmarks capture subtle geometric variations in facial muscle movements that are strongly associated with emotional expressions. This normalization ensures consistency across faces of different sizes and orientations. This provides a stable geometric representation for subsequent graph-based emotion recognition.

2) Graph construction through high-order graph attention network (HoGAN): HoGAN [42] extends the conventional Graph Attention Network by modeling higher-order interactions that capture long-range dependencies across the graph. In this study, the facial landmark graph $G = (V, E)$, where V denotes the nodes and E denotes the edges, with node features $h_x \in \mathbb{R}^G$ for each landmark x, a standard attention mechanism computes pairwise attention coefficients between directly connected nodes, as denoted in Eq. (9).

$$e_{xy} = \text{LeakyReLU}(\text{w}^\top[\text{Mh}_x \parallel \text{Mh}_y]) \tag{9}$$

where, M denotes the learnable linear transformation, w represents the attention weight vector, and $\parallel$ signifies the concatenation. These coefficients are normalized using a SoftMax function over the neighborhood $\mathcal{N}_x$ of node x, as denoted in Eq. (10). This allows the network to adaptively weight the contribution of neighboring landmarks based on their relevance.

$$\sigma_{xy} = \frac{\exp(e_{xy})}{\sum_{k \in \mathcal{N}_x} \exp(e_{xk})}, \tag{10}$$

where, attention is extended to capture information from multi-hop neighborhoods by aggregating features from k-order neighbors. This enables the network to model complex facial muscle interactions across distant regions of the face. The high-order feature representation of the node x is computed as in Eq. (11).

$$\text{H}_i^{(k)} = \alpha\left(\sum_{y \in \mathcal{N}_x{}^k} M^k \, h_y\right) \tag{11}$$

where, $\mathcal{N}_x{}^k$ represents the set of nodes reachable within k hops, $\sigma_{xy}{}^k$ denotes the high-order attention coefficients, and $\alpha(.)$ represents the nonlinear activation function. Multiple attention heads are employed to stabilize learning and enhance representational capacity. The HoGAN effectively captures both localized and global facial dynamics that are critical for emotion expression. This capability is particularly beneficial for facial emotion recognition, as emotional cues are often distributed across multiple, spatially distant facial regions. Consequently, the high-order graph attention mechanism enables more discriminative and robust emotion modeling.

E. Joint Learning Strategy through Loss Functions

The proposed framework adopts a joint learning strategy to simultaneously optimize face identity recognition and facial emotion recognition within a unified DL framework. The model formulated the training dataset as D, where a_x represents the input facial image, b_x^{id} denotes the identity label, and b_x^{em} signifies the emotion label. Identity learning is formulated as a metric learning problem using triplet loss ($\mathcal{L}_{tr}$), as expressed in Eq. (12).

$$\mathcal{L}_{tr} = \sum_{i=1}^{N} \max\left(0, \| f(a_x^u) - f(a_x^p) \|_2^2 - \| f(a_x^u) - f(a_x^n) \|_2^2 + \alpha\right) \tag{12}$$

where, $f(\cdot)$ denotes the identity embedding function, N denotes the total images, and a_x^u, a_x^p, and a_x^n represent anchor, positive, and negative samples, respectively, with a margin α. Emotion recognition is optimized using a categorical cross-entropy loss ($\mathcal{L}_{cross}$), as denoted in Eq. (13).

$$\mathcal{L}_{cross} = -\sum_{c=1}^{C} b_c \log(\hat{b}_c) \quad (13)$$

where, C represents the number of emotion classes, b_c signifies ground-truth label, and $\hat{b}_c$ denotes predicted probability. The overall objective function is formulated as a weighted sum of both losses, as shown in Eq. (14).

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{tr} + \lambda_2 \mathcal{L}_{cross} \quad (14)$$

where, $\lambda_1$and λ_2 balance the contribution of identity and emotion tasks. This joint optimization enables the network to learn shared facial representations, improving generalization while ensuring efficient convergence for real-time assistive applications.

IV. Results and Discussion

This section presents the experimental performance of the proposed unified DL framework. The performance is assessed using quantitative metrics, qualitative visualizations, ablation analysis, and comparative analysis to demonstrate the effectiveness of the proposed system.

A. Hyperparameter Settings and Complexity Analysis

The proposed model finetunes hyperparameters to strike a balance between model performance and computational efficiency, as demonstrated in Table II. The proposed unified DL framework results in a balanced computational complexity for the proposed system. EfficientNet-B3 has a computational complexity of approximately $O(H \times W \times D)$, where H and Wdenote the spatial dimensions and D represents the channel depth. The emotion recognition module operates with a complexity of $O(K \times N^2)$, where $N = 68$ landmarks and K denotes the number of attention heads. This makes it suitable for real-world assistive applications for prosopagnosia patients.

TABLE. II. Hyperparameter Details of the Proposed Framework

Component	Hyperparameter	Value
MTCNN	Scale factor	0.709
	Number of landmarks	5
EfficientNet-B3	Dropout rate	0.3
	Embedding dimension	512
Dilb	Number of landmarks	68
HoGAN	Number of layers	3
	Attention heads	8
	Hidden dimension	128
	High-order hops	2-hop
Training Settings	Optimizer	Adam
	Learning rate	0.0001
	Weight decay	1×10^{-4}
	Batch size	32
	Epochs	10
	Triplet margin	0.2
	Loss weights (λ_1, λ_2)	1.0, 1.0

B. Performance Analysis of the Proposed Model

This section presents the performance analysis of the proposed unified deep learning framework, as illustrated in Table III. The performance was analyzed through the metrics of accuracy, precision, recall, specificity, f1-score, Matthews Correlation Coefficient (MCC), False Positive Rate (FPR), False Negative Rate (FNR), Equal Error Rate (EER), Cohen's kappa, and AUC values. The results show that the proposed model attains a better performance in both face identity recognition and face emotion recognition. This shows the ability of the proposed model to efficiently support patients with prosopagnosia.

TABLE. III. Performance of the Proposed Unified Deep Learning Framework

Metrics	Face identity recognition	Face emotion recognition
Accuracy (%)	99.58	99.02
Precision (%)	99.08	99.14
Recall (%)	99.11	99.09
Specificity (%)	99.58	99.39
F1-score (%)	99.09	99.21
MCC (%)	99.1	99.01
FPR	0.42	0.61
FNR	0.89	0.92
Equal Error Rate (EER)	0.45	0.87
Cohen's Kappa (%)	99.1	98.9
AUC (%)	99.8	99.5

C. Face Identity Recognition

Fig. 4 shows the results of MTCNN-based face detection and alignment. MTCNN detects faces across different subjects with variations in pose, illumination, facial appearance, and background. The five-point facial landmarks are used to align the faces that are detected, ensuring consistent positioning of key facial components such as the eyes, nose, and mouth. The visual results indicate the improved representation of the facial region following the alignment. This normalization significantly reduces intra-class variations and provides a stable input for the identity recognition module. The findings ensure that MTCNN is effective in preprocessing. This is essential in enhancing the accuracy of recognition in unconstrained and real-time assistive cases.

Fig. 4. Results of face detection and alignment through MTCNN. The first row denotes the original images of the VGGFace2 dataset, and the second row denotes the facial detection images.

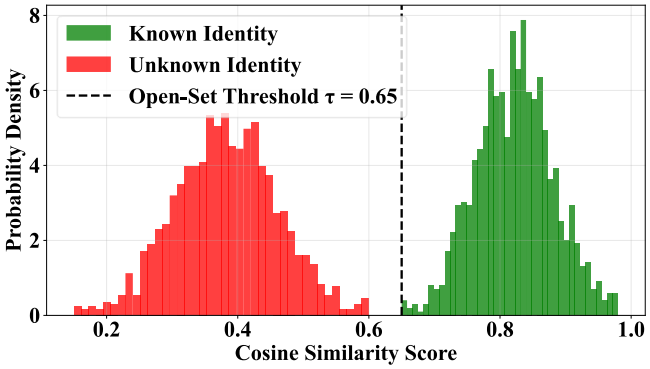

Fig. 5. Distribution of cosine similarity scores for known and unknown identities.

Fig. 5 illustrates the distribution of cosine similarity scores obtained during the identity matching process. The green histogram represents similarity scores for known identity pairs, which are concentrated at higher values. The red histogram corresponds to unknown identity pairs, which exhibit lower similarity scores. This reflects effective separation between different identities. The dashed line denotes the open-set decision threshold set at ($\tau = 0.65$), which serves as the boundary for distinguishing known and unknown identities. The minimal overlap between the two distributions demonstrates the discriminative power of the learned embeddings. This validates the effectiveness of the selected threshold in open-set identity recognition scenarios. This ensures reliable identity prediction while minimizing false acceptances and false rejections.

D. Emotion Recognition

Fig. 6 illustrates the results of facial landmark detection using the Dlib model. The green points represent the detected facial landmarks, which densely capture critical regions including the eyes, eyebrows, nose, mouth, and jawline. These precise geometric representations provide rich structural information, essential for modeling facial dynamics. This serving a reliable foundation for the subsequent graph-based emotion recognition module. This highlights its effectiveness in accurately localizing key facial features across different emotional expressions.

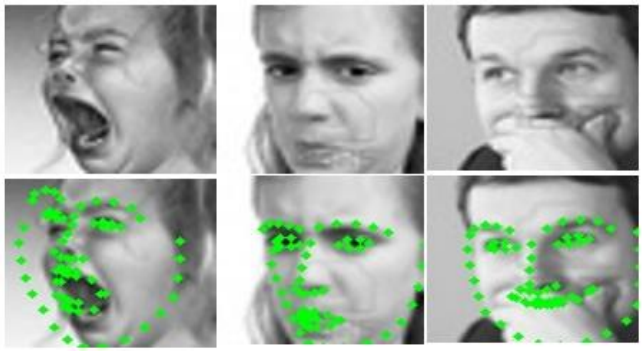

Fig. 6. Results of the facial landmark detector through Dilb. The first row denotes the original images of the dataset, and the second row denotes the facial landmark detection images.

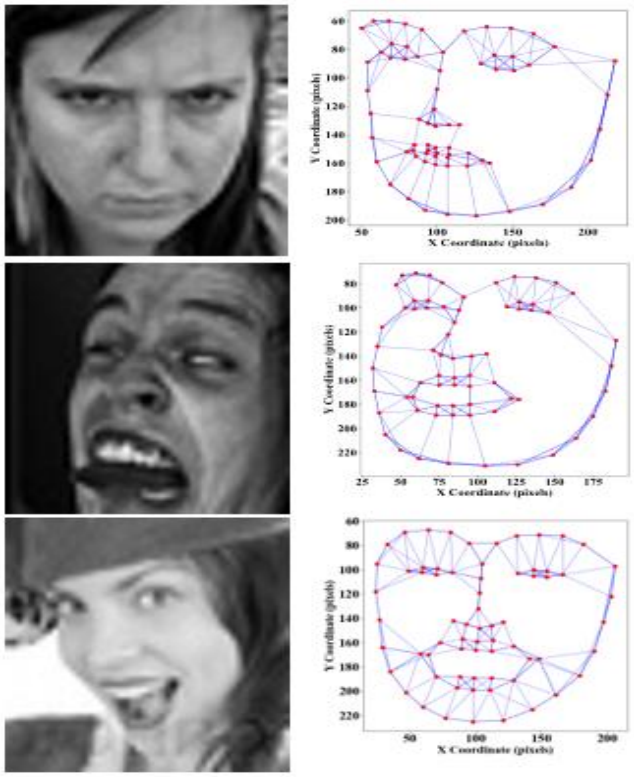

Fig. 7. Graph construction results through a high-order graph attention network.

Fig. 7 presents the graph construction results through HoGAN. The left column shows original facial images, while the right column illustrates the constructed facial graphs derived from the detected landmarks. Each red node represents a facial landmark, and the blue edges indicate the spatial and anatomical connections between landmarks used to form the graph structure. These structured graph representations enable to effectively capture of both local deformations and long-range dependencies between facial regions. This enhances the model's ability to discriminate subtle emotional cues. The visualization demonstrates that the proposed graph-based approach provides an interpretable and expressive geometric representation of facial dynamics.

E. Model Training Performance

Fig. 8 illustrates the training behavior and convergence characteristics of the proposed unified identity and emotion recognition framework. The accuracy curve shows a steady increase in classification accuracy as the number of training epochs progresses. The curve rises from approximately 60% in the initial epoch to nearly 99% by the final epoch. This consistent upward trend indicates effective feature learning and model convergence. This demonstrates that the network successfully captures discriminative facial representations with continued training. The rapid improvement in early epochs reflects efficient optimization, while the gradual saturation in later epochs suggests stable convergence without overfitting.

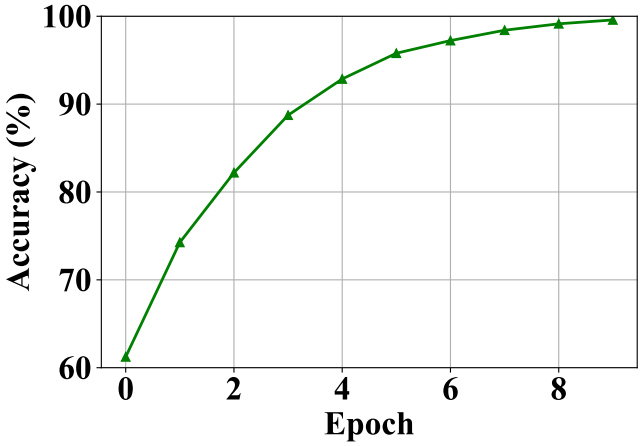

Fig. 8. Training accuracy of the proposed model across epochs, showing improvement and stable convergence during the learning process.

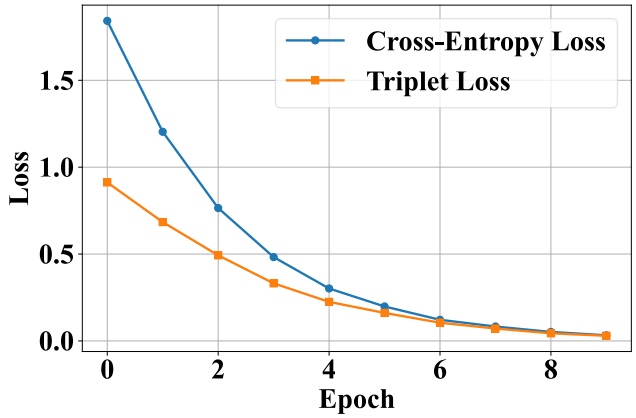

Fig. 9. Training loss curve of the proposed framework showing the reduction of cross-entropy loss for emotion recognition and triplet loss for face identity recognition across epochs.

Fig. 9 illustrates the loss function performance of the proposed framework. Both losses decrease across epochs, with cross-entropy loss reducing sharply during early training. Triplet loss gradually converges as embedding separability improves. The parallel decline of these losses highlights the effectiveness of the joint multi-task learning strategy, where identity and emotion objectives are optimized concurrently. These confirm that the proposed model achieves stable training, balanced optimization across tasks, and robust convergence. This supports its suitability for real-time cognitive assistance applications.

Fig. 10 illustrates the performance of the proposed facial emotion recognition module across seven emotion classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Each emotion class achieves a high number of correct predictions. This indicates high classification accuracy and effective discrimination between emotional states. Misclassifications are minimal and primarily occur between visually similar emotions, such as fear and sadness or anger and disgust. This confirms the reliability and stability of the proposed emotion recognition module, validating its effectiveness for real-world cognitive assistance applications.

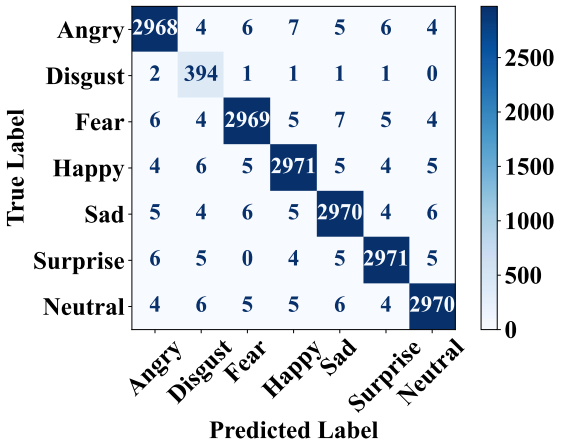

Fig. 10. Confusion matrix of the proposed framework.

Fig. 11. Results of the proposed prosopagnosia assistive system showing face identity and emotion recognition. This predicted identity labels and recognized emotional states.

Fig. 11 illustrates the results of the proposed prosopagnosia assistive system, showing its ability to simultaneously recognize face identity and facial emotion. Each shows an input image processed by the system, with the annotated output displaying the predicted identity label and the recognized emotional state. The examples highlight the system's robustness under varying conditions, including different lighting, facial poses, backgrounds, and emotional intensities. These results confirm the practical applicability of the proposed framework as a real-time cognitive assistance system. This helps to efficiently support individuals with prosopagnosia during everyday interactions.

F. Comparative Analysis

This section presents the comparative analysis of the proposed model, as represented in Table IV. The proposed framework is compared with the baseline models, MLBP-HOG-G, KNN, PNN, SVM, CNN, CLCM, and other approaches. The proposed model attained an accuracy of 99.58% and 99.02% in both modules. This shows that the proposed model attained a better accuracy compared to baseline models, and shows the efficiency in both face identity recognition and face emotion recognition.

TABLE. IV. COMPARATIVE ANALYSIS WITH BASELINE MODELS

Models	Face identity recognition			Face emotion recognition		
	Accuracy (%)	Precision (%)	F1-score (%)	Accuracy (%)	Precision (%)	F1-score (%)
MLBP-HOG-G	87.42	86.91	86.99	87.13	86.92	86.96
KNN	88.17	87.64	87.79	86.47	86.12	86.22
PNN	89.54	89.02	89.16	89.92	87.61	87.67
SVM	91.83	91.42	91.51	90.36	90.11	90.17
CNN	94.26	94.08	94.12	93.58	93.34	93.40
CLCM	96.48	96.31	96.35	95.94	95.81	95.85
GAT	98.12	98.01	98.03	97.61	97.49	97.52
Proposed	**99.58**	**99.08**	**99.09**	**99.02**	**99.14**	**99.21**

G. Discussion

This study presented a unified framework for face identity and facial emotion recognition aimed at addressing the real-world challenges faced by individuals with prosopagnosia. The experimental results demonstrate that integrating face detection and alignment with EfficientNet-B3-based identity feature extraction and landmark-driven HoGAN for emotion recognition leads to highly accurate performance. The joint learning strategy enables the model to exploit information between identity and emotion tasks. The study utilized the VGGFace2 and FER2013 benchmark datasets. They may introduce biases related to demographic distribution, illumination conditions, and expression diversity. VGGFace2 primarily contains celebrity images, which may not fully represent real-world population variability. FER2013 consists of low-resolution grayscale images that may limit fine-grained emotional representation. These dataset characteristics influence model generalization in practical deployments. To mitigate this, data augmentation and joint learning were employed to improve robustness. The strong separation observed in cosine similarity distributions indicates that the proposed system is well-suited for real-time assistive applications.

Although the proposed framework achieves very high performance across multiple evaluation metrics, it is important to critically consider potential overfitting effects. The high accuracy values may partially result from training and evaluation on benchmark datasets. To mitigate overfitting, regularization strategies including dropout, joint loss optimization, and validation monitoring were employed during training. The stable convergence behavior observed in training accuracy and loss curves suggests balanced learning. However, the proposed framework acknowledged some limitations. The study depends on publicly available datasets, which do not fully capture the diversity of real-world interactions experienced by prosopagnosia patients. Future work will focus on addressing these limitations by incorporating multimodal information such as speech, contextual cues, and physiological. This provides a foundation for intelligent assistive technologies for advancing cognitive support systems for prosopagnosia patients.

V. CONCLUSION

This study successfully developed a unified DL framework for face identity and facial emotion recognition to support individuals with prosopagnosia in real-world social interactions. The proposed model employed a dual module framework that simultaneously performs face identity recognition and face emotion recognition. The proposed system effectively captured both identity-specific and expression-related facial cues. The joint learning strategy enabled optimization of identity and emotion tasks, leading to high recognition accuracy, stable convergence, and strong generalization performance. The experimental results demonstrated that the proposed approach outperforms conventional task-specific models while maintaining real-time efficiency. The framework provides a scalable solution for cognitive assistance, offering an efficient identity and emotion framework to support the daily social interactions of individuals with prosopagnosia. In the future, we will extend the proposed framework under different real-world conditions, ensuring robust deployment in cognitive, practical assistive environments.

STATEMENTS AND DECLARATIONS

Author contributions: Both authors contributed to the conception of the problem setting, conceptualization, methodology, implementation, testing and writing the final manuscript.

Funding: No funding was received for conducting this study.

Availability of data and materials:

https://github.com/ox-vgg/vgg_face2
https://datarepository.wolframcloud.com/resources/FER-2013

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: The research is original, and the authors of this manuscript created all the figures and tables.

Consent to participate: Not applicable.

REFERENCES

[1] S. Byrne, M. and Porter, "Rehabilitation and intervention of developmental and acquired prosopagnosia: A systematic review," Neuropsychological Rehabilitation, vol. 35, pp. 1-44, Jan 2025.

[2] K. A. Josephs, and K. A. Josephs Jr, "Prosopagnosia: face blindness and its association with neurological disorders," Brain Communications, vol. 6, pp. fcae002, Jan 2024.

[3] E. Nørkær, S. Gobbo, T. Roald, and R. Starrfelt, "Disentangling developmental prosopagnosia: A scoping review of terms, tools and topics," Cortex, vol. 176, pp. 161-193, Jul 2024.

[4] X. Xu, X. He, W. Ren, and X. Zhao, "The network structure of autistic traits, executive function, prosopagnosia and social anxiety," Research in Autism, vol. 131, pp. 202815, Mar 2026.

[5] J. Lowes, L. M. McGregor, P. J. Hancock, B. Duchaine, and A. K. Bobak, "This condition impacts every aspect of my life: A survey to understand the experience of living with developmental prosopagnosia," PloS one, vol. 20, pp. e0322469, Apr. 2025.

[6] B. K. Devisetty, A. Goyal, A. Mishra, M. W. Nijim, D. Hicks, and G. Toscano, "A Hybrid Regression-Based Network Model for Continuous Face Recognition and Authentication," International Journal of Advanced Computer Science & Applications, vol. 15, p30, Oct. 2024.

[7] E. J. Burns, E. Gaunt, B. Kidane, L. Hunter, and J. Pulford, "A new approach to diagnosing and researching developmental prosopagnosia: Excluded cases are impaired too," Behavior research methods, vol. 55, pp.4291-4314, Dec. 2023.

[8] T. Halder, K. Ludwig, and T. Schenk, "Binocular rivalry reveals differential face processing in congenital prosopagnosia," Scientific Reports, vol. 14, pp. 6687, Mar. 6687.

[9] J. J. Barton, and F. E. I. L. Moritz, "Foundations of prosopagnosia: The three classic Austro-German reports," Cortex, vol. 193, pp. 1-15, Dec 2025.

[10] S. Djouab, A. Albonico, S. C. Yeung, M. Malaspina, A. Mogard, R. Wahlberg, and J. J. Barton, "Search for face identity or expression: Set size effects in developmental prosopagnosia," Journal of Cognitive Neuroscience, vol. 32, pp. 889-905. May 2020.

[11] H. T. Wang, K. M. Chuang, T. Rawat, J. L. Lyu, M. Ali, and S. H. L. Chien, "Subjective face recognition ability is linked to objective face memory and face authenticity judgment: validation of the Traditional Chinese version of the 20-Item Prosopagnosia Index," Brain Sciences, vol. 15, pp.1186, Oct. 2025.

[12] R. Fry, X. Li, T. C. Evans, M. Esterman, J. Tanaka, and J. DeGutis, "Investigating the influence of autism spectrum traits on face processing mechanisms in developmental prosopagnosia," Journal of autism and developmental disorders, vol. 53, pp.4787-4808, Dec 2023.

[13] L. Bell, B. Duchaine, and T. Susilo, "Dissociations between face identity and face expression processing in developmental prosopagnosia," Cognition, vol. 238, pp. 105469, Sep 2023.

[14] K. Minemoto, and Y. Ueda, "Face identity and facial expression representations with adaptation paradigms: New directions for potential applications," Frontiers in psychology, vol. 13, pp. 988497, Dec 2022.

[15] B. Q. Z. Leong, A. M. Hussain Ismail, and A. J. Estudillo, "Persistent task-specific impairment of holistic face processing in acquired prosopagnosia," Scientific Reports, vol. 15, pp. 43115, Dec 2025.

[16] M. A. Hamzah, "Advancing personal identity verification by integrating facial recognition through deep learning algorithms," International Journal of Information Technology, vol. 16, pp. 4381-4386, Oct 2024.

[17] W. Y. Hsu, and Y. C. Chen, "Multi-attribute feature-aware network for facial expression recognition," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 21, pp. 1-20, July 2025.

[18] A. Kazemian, I. Oruc, and J. J. Barton, "Scanning faces: a deep learning approach to studying eye movements in prosopagnosia," Frontiers in Neurology, vol. 16, pp.1616509, Sep 2025.

[19] V. Manippa, and A. Palmisano, M. Ventura, and D. Rivolta, "The neural correlates of developmental prosopagnosia: Twenty-five years on," Brain Sciences, vol. 13, pp.1399, Sep 2023.

[20] S. Albalawi, L. Alamri, J. Atut, S. Albalawi, and R. Haddaddi, "MAHYA: Facial Recognition-Based Pilgrim Identification System for Enhanced Health Monitoring and Assistance," International Journal of Advanced Computer Science & Applications, vol. 16, pp. 928- 941, Mar 2025.

[21] V. Singh, and S. Prasad, "Speech emotion recognition system using gender dependent convolution neural network," Procedia Computer Science, vol. 218, pp.2533-2540, Jan 2023.

[22] J. Ma, X. Wang, and Y. Li, "Specific Deficits in Facial Recognition in Children Aged 7–14 Years With Developmental Prosopagnosia," Journal of Autism and Developmental Disorders, pp.1-14, Nov 2025.

[23] P. Naga, S. D. Marri, and R. Borreo, "Facial emotion recognition methods, datasets and technologies: A literature survey," Materials Today: Proceedings, vol. 80, pp. 2824-2828, Jan 2023.

[24] A. Zhalgas, B. Amirgaliyev, and A. Sovet, "Robust Face Recognition Under Challenging Conditions: A Comprehensive Review of Deep Learning Methods and Challenges," Applied Sciences, vol. 15, pp. 9390, Aug 2025.

[25] A. Volfart, C. Michel, and B. Rossion, "A simple behavioral evaluation test of human face identity recognition with natural images validated with the case of prosopagnosia PS," Scientific Reports, vol. 15, pp. 43149, Dec 2025.

[26] E. I. Olivares, A. S. Urraca, A. Lage-Castellanos, and J. Iglesias, "Different and common brain signals of altered neurocognitive mechanisms for unfamiliar face processing in acquired and developmental prosopagnosia," Cortex, vol. 134, pp. 92-113, Jan 2021.

[27] X. Li, "The Application of Face Recognition Model Based on MLBP-HOG-G Algorithm in Smart Classroom," International Journal of Advanced Computer Science & Applications, vol. 16, p724, Mar 2025

[28] W. H. Jain, B. G. Jhong, and M. Y. Chen, "A Social Assistance System for Augmented Reality Technology to Redound Face Blindness with 3D Face Recognition," Electronics, vol. 14, pp. 1244, Mar 2025.

[29] M. Mukhiddinov, and J. Cho, "Smart glass system using deep learning for the blind and visually impaired," Electronics, vol. 10, pp.2756, Nov 2021.

[30] M. Tsantani, K. L. Gray, and R. Cook, "New evidence of impaired expression recognition in developmental prosopagnosia," Cortex, vol. 154, pp. 15-26, Sep 2022.

[31] H. T. Wang, J. L. Lyu, and S. H. L. Chien, "Dynamic Emotion Recognition and Expression Imitation in Neurotypical Adults and Their Associations with Autistic Traits," Sensors, vol. 24, pp. 8133, Dec 2024.

[32] F. K. Bardak, and F. Temurtaş, "Regional Brain Analysis and Machine Learning Techniques for Classifying Familiar and Unfamiliar Faces Using EEG," Arabian Journal for Science and Engineering, vol. 50, pp. 1-26, Jan 2025.

[33] M. C. Gursesli, S. Lombardi, M. Duradoni, L. Bocchi, A. Guazzini, and A. Lanata, "Facial emotion recognition (FER) through custom lightweight CNN model: performance evaluation in public datasets," IEEE Access, vol. 12, pp. 45543-45559, Mar 2024.

[34] F. M. Talaat, "Real-time facial emotion recognition system among children with autism based on deep learning and IoT," Neural Computing and Applications, vol. 35, pp. 12717-12728, June 2023.

[35] M. Mukhiddinov, O. Djuraev, F. Akhmedov, A. Mukhamadiyev, and J. Cho, "Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people,". Sensors, vol. 23, pp. 1080, Jan 2023.

[36] VGGFace2 dataset link: https://github.com/ox-vgg/vgg_face2.

[37] FER2013 dataset link: https://datarepository.wolframcloud.com/resources/FER-2013

[38] M. Gu, X. Liu, and J. Feng, "Classroom face detection algorithm based on improved MTCNN," Signal, Image and Video Processing, vol. 16, pp.1355-1362, Jul 2022.

[39] M. U. Naveed, M. M. Iqbal, S. Majeed, F. Ali, and Q. G. K. Safi, "Lung Cancer Classification through Transfer Learning and Deep Feature Extraction using EfficientNetB3," Journal of Computing & Biomedical Informatics, vol. 9, Sep 2025.

[40] W. Khan, T. Kumar, C. Zhang, K.Raj, A. M. Roy, and B. Luo, "SQL and NoSQL database software architecture performance analysis and assessments—a systematic literature review," Big Data and Cognitive Computing, vol. 7, pp. 97, May 2023.

[41] J. Wang, C. He, and Z. Long, "Establishing a machine learning model for predicting nutritional risk through facial feature recognition," Frontiers in Nutrition, vol. 10, pp. 1219193, Sep 2023.

[42] L. He, L. Bai, X.Yang, H. Du, and J. Liang, "High-order graph attention network," Information Sciences, vol. 630, pp. 222-234, June 2023.