Region-Aware Attention with Hybrid Angular-Cosine Margin Loss for Prosopagnosia Face Recognition

Bhavana Nagaraj * and Rajanna Muniswamy

Department of Information Science and Engineering, Vemana Institute of Technology, Bengaluru, India
Email: bhavananagaraj03@vemanait.edu.in (B.N.); rajannam@vemanait.edu.in (R.M.)
*Corresponding author

***Abstract*—Prosopagnosia diminishes an individual's ability to recognize familiar faces, under pose variations, occlusions, and when faces appear visually similar. This study develops a deep learning framework, ResRAMACL, with ResNet-101, Region-Aware Attention Mechanism (RAAM), and a novel Hybrid Angular-Cosine Margin Loss (HACML), for supporting prosopagnosia patients to recognize human faces. Facial images are aligned using 2D facial landmark plotting, which separates key semantic regions from the images. ResNet-101 extracts feature maps for each region, which are weighted by RAAM to highlight cognitively crucial areas. These features are fused into a single embedding vector and trained using HACML, and the facial embeddings are stored in a Structured Query Language (SQL) database. The prediction phase generates an embedding using a trained model, checks the similarity with the stored embeddings, and predicts the final result. Compared to the existing loss functions, the proposed HACML achieved up to 2.16 units higher inter-class separation and up to 0.44 units lower intra-class variance. Ablation study proved higher performance compared with variations of the pipeline. The generalizability analysis proved minimal variation in performance when tested with other datasets. Overall, this proves the significance of the approach for integration with Augmented Reality (AR)-based or mobile cognitive assistance technologies, enabling prosopagnosia patients to recognize faces.**

***Keywords*—face recognition, prosopagnosia, Region-Aware Attention Mechanism (RAAM), ResNet-101, Structured Query Language (SQL) database**

I. Introduction

Prosopagnosia, which refers to face blindness, is a visual impairment causing the patient to be deficient in face recognition. Prosopagnosia affects more than 2% of the population worldwide over the past 20 years [1, 2]. Prosopagnosia is characterized as Developmental Prosopagnosia (DP) or Acquired Prosopagnosia (AP). DP cases occur since birth or have a genetic basis, while AP cases occur due to a variety of conditions, such as trauma, neurodegenerative diseases, stroke, brain surgery, neuroinfectious diseases, and, less frequently, malignancies [3, 4]. Prosopagnosia confuses face recognition of previously met people, particularly unfamiliar faces, and also affects recognition of familiar faces, sometimes even with their own face recognition [5, 6]. Prosopagnosia patients depend on non-facial indicators like, voice, clothing, or hairstyle for identification [7]. The consequences include severe psychological difficulties, including social anxiety, embarrassment, and complications in interpersonal relationships [8–10]. This affects the quality of life of prosopagnosia patients and affects their mental well-being. These consequences lead to the significance of developing an effective face recognition system for prosopagnosia patients to enhance everyday facial recognition, thereby optimising daily activities and quality of life [11, 12]. Face recognition comprises four main stages, such as initial face outline detection, face alignment with different components, feature extraction from these components, and comparison of face features [13]. Feature extraction of facial components is a significant task in the face recognition method. To enhance the face recognition accuracy, it is essential to improve the model's capability to extract the various facial features [14, 15].

Mukhiddinov and Chao [16] developed a supporting technology for the visually impaired persons. The study used Deep Learning (DL) algorithms to design a smart glass system for improving the quality of life of visually impaired patients. However, the model incorrectly detects the small objects. Jian *et al.* [17] developed an Augmented Reality (AR)-based visual aid system to help prosopagnosia patients. The system used a Convolutional Neural Network (CNN) to generate a 128-dimensional feature vector for precise face matching and to improve recognition from multiple angles. Mann *et al.* [18] presented an assistive technology to deliver a real-time face recognition model that delivers audio and visual notification of individuals. The model merges the features of mnemonic and perceptual training to support prosopagnosia rehabilitation. Bardak and Temurta [19] developed a framework for identifying known and

Manuscript received December 1, 2025; revised December 24, 2025; accepted January 28, 2026; published May 13, 2026.

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 improvements to enhance the accuracy through implementing DL models. The margin Softmax loss function is generally used in face recognition tasks. However, it fails to achieve higher inter-class separability and minimal intra-class compactness [20, 21]. Moreover, existing face recognition models supporting prosopagnosia patients lack robustness in handling multi-view scenarios and fail to focus on the cognitively significant facial regions [22]. Prosopagnosia patients struggle to recognize familiar faces, particularly under pose variations, occlusions, and when faces are visually similar.

Existing DL models process the entire face uniformly without giving special attention to cognitively important regions such as the eyes, nose, mouth, and eyebrows. Additionally, existing models struggle to differentiate visually similar faces, which is problematic for prosopagnosia patients and recognition accuracy declines under pose and illumination variations.

Furthermore, traditional margin-based loss functions also fail to balance the factors, inter-class separability, and intra-class compactness. To overcome these challenges, we have proposed this study on face recognition to assist people with prosopagnosia. This study aims to develop a scalable, real-time face recognition system that can be integrated into future AR-based or cognitive assistance technologies, enabling prosopagnosia patients to recognize faces more effectively in everyday social interactions.

The core contributions of the proposed study are summarized as follows:

- This study develops a novel ResRAMACL face recognition framework tailored for prosopagnosia patients, that combines ResNet-101 for feature extraction and a Region-aware Attention Mechanism (RAAM) for highlighting cognitively important facial areas.
- We introduce a Hybrid Angular-Cosine Margin Loss **(**HACML) loss function during training to improve inter-class separability and intra-class compactness, particularly for differentiating visually similar faces.
- The proposed ResRAMACL framework is integrated with a practical assistive pipeline by incorporating a Structured Query Language (SQL) database backend for efficient storage and retrieval of facial embeddings. This enabling scalable real-time deployment in AR-based or mobile cognitive assistance systems.

The remaining portion of the paper is structured as follows: Section II defines the materials and methods of the proposed framework, Section III includes results analysis and discussions on the implementation of the approach, and Section IV concludes the paper.

II. Materials and Methods

A. Dataset Description

The proposed face recognition framework needs a high-quality image dataset with facial landmarks. In this study, we utilize the AFLW2000-3D dataset [23] for face recognition of prosopagnosia patients. The dataset includes 2000 images that are annotated with 68-point image-level facial landmarks. The dataset includes different head poses of people in the dataset and comprises the facial landmarks, as represented in Fig. 1.

Fig. 1. Sample images from the AFLW2000-3D dataset. (The dataset holds images of different poses of persons.).

The AFLW2000-3D dataset consists of images from neurotypical individuals. This dataset focuses on pose variation, visual similarity discrimination, and region-aware feature learning, which supports assistive deployment. We utilize this dataset to evaluate the performance of our proposed framework for assistive face recognition support in patients with prosopagnosia.

B. Methods

1) The ResRAMACL framework

This study proposes a face recognition framework (ResRAMACL) tailored for prosopagnosia patients. In the training phase, the input images are detected and aligned using a 2D facial landmark plotting method. The detected landmarks are divided into key semantic regions: eyes, nose, mouth, eyebrows, and global face. ResNet-101 extracts feature maps for each region, which are weighted by RAAM to emphasize cognitively important areas. The weighted regional features are fused into a single embedding vector, which is trained using HACML to enforce strong inter-class separability and intra-class compactness, even under pose variations or visually similar faces, and the facial embeddings are stored in an SQL database. The prediction phase generates embeddings using the trained model and then compared, with the stored embeddings in the SQL database. This checks the similar embeddings and predicts the face. Here, we have involved SQL, to store the recognition related data for future applications like AR glass or mobile cognitive assistance technologies. This enables prosopagnosia patients to recognize faces effectively. The illustration of this pipeline is shown in Fig. 2.

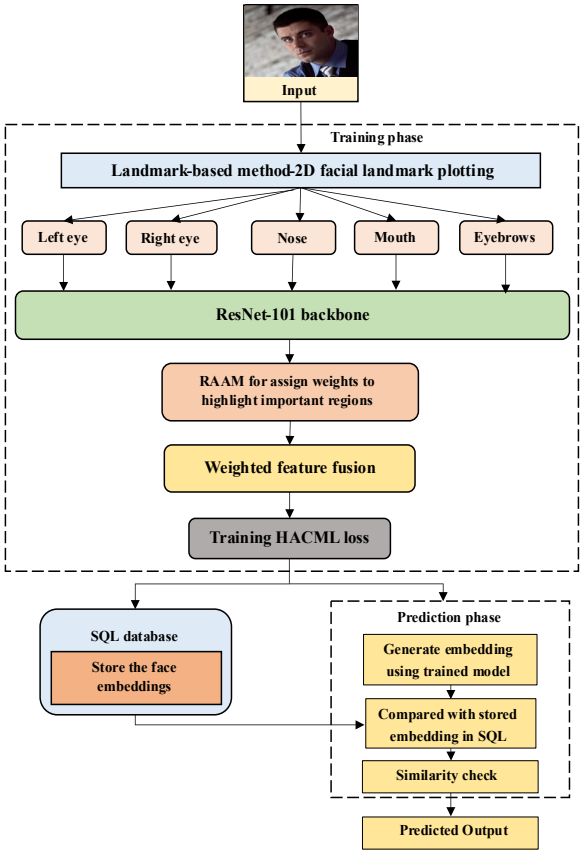

Fig. 2. Schematic diagram of the proposed ResRAMACL framework.

2) Preprocessing through landmark-based method for detecting facial regions

Facial landmark detection is a significant step to enhance face recognition tasks. A landmark-based method for detecting facial regions first uses a face detector to find the entire face, followed by a model that identifies key regions. In this study, we detected a group of facial landmarks such as, the nose tip, nose corners, and eye corners, using the 2D facial landmark plotting method [24]. A 2D facial landmark detector is applied to estimate 68 predefined landmark points corresponding to key facial structures. Using the detected landmarks, Region-of-Interest (ROI) bounding boxes are defined for each facial region based on fixed geometric relationships among landmark points. The eye region is localized using landmarks around the orbital area, the nose region using the nasal bridge and tip landmarks, and the mouth region using lip contour landmarks. Each ROI is resized to a fixed spatial resolution and spatially aligned with respect to the global face coordinate system to ensure consistent region correspondence across identities and pose variations. These landmarks are then used to express and segment specific facial regions for prosopagnosia patients' facial recognition applications. The landmark-based face detection approach is expressed in Eq. (1).

$$LM = \{G_{k,l}, F_{k,l} | 1 \le k \le K, 1 \le l \le L\} \quad (1)$$

where LM represents the set of landmarks, $(G_{k,l}, F_{k,l})$ denotes the 3D coordinates of each landmark, and K and L denotes the number of landmarks and frames. All extracted regions are resized to a uniform input resolution and aligned to maintain spatial consistency across samples. These detected regions are fed to the proposed ResRAMACL model for accurate face recognition of prosopagnosia patients. The extracted facial regions are forwarded to the RAAM, which adaptively weights each region according to its discriminative contribution.

3) The training module

The detected and aligned face regions using a landmark-based method ensure spatial consistency. The aligned images are divided into key semantic regions such as eyes, nose, mouth, eyebrows, and the entire face, which is passed to the proposed ResRAMACL framework. The proposed ResRAMACL framework integrates ReNet-101 with the RAAM and HACML, as illustrated in Fig. 3. ResNet-101 [25] is a CNN family that includes a depth of 101 layers. In this study, we utilize the ResNet-101 that extracts feature maps for each region.

The pooling layer decreases the dimensions of the feature maps and number of parameters, and the latter layers generate a single long feature vector. Thus, the multidimensional input is reduced to a single dimension while moving from convolution layers to a Fully Connected (FC) layer. The FC layer is made up of a number of layers. Each region R_x is processed through the ResNet-101 backbone to obtain deep feature maps $F_x \in \mathbb{R}^{H \times W \times C}$, represented as:

$$F_x = f_{ResNet101}(R_x) \quad (2)$$

In Eq. (2), $f_{ResNet101}(R_x)$ represents the feature extraction function of the ResNet-101 network. To highlight cognitively important facial regions, the extracted features are passed through the RAAM [26]. RAAM assigns attention weights to each region feature map F_x. The attention weight ω_x is calculated as:

$$\omega_x = \frac{e^{(W_x^T g_t)}}{\sum_{c=1}^{N} e^{(W_x^T f_k)}} \quad (3)$$

In Eq. (3), W_x indicates the weight parameter, g_t denotes the region-level feature vector combined via Global Average Pooling (GAP), f_k denotes the channel vector, W_x^T represents the transpose of the weight vector, and N denotes the total number of facial regions. The features of a weighted region are combined using a single vector, as defined in Eq. (4).

$$S = \sum_{n=1}^{N} \omega_x F_x \quad (4)$$

where S represents the fused single vector that highlights the significant and the cognitively relevant features. N denotes the total number of facial regions and F_x signifies the feature vector extracted from the x-th facial region. The fused vector is normalized using the L_2-normalization to ensure scale-invariant I_x, as denoted in Eq. (5).

$$S = L_2 norm(\sum_{n=1}^{N} I_x) \quad (5)$$

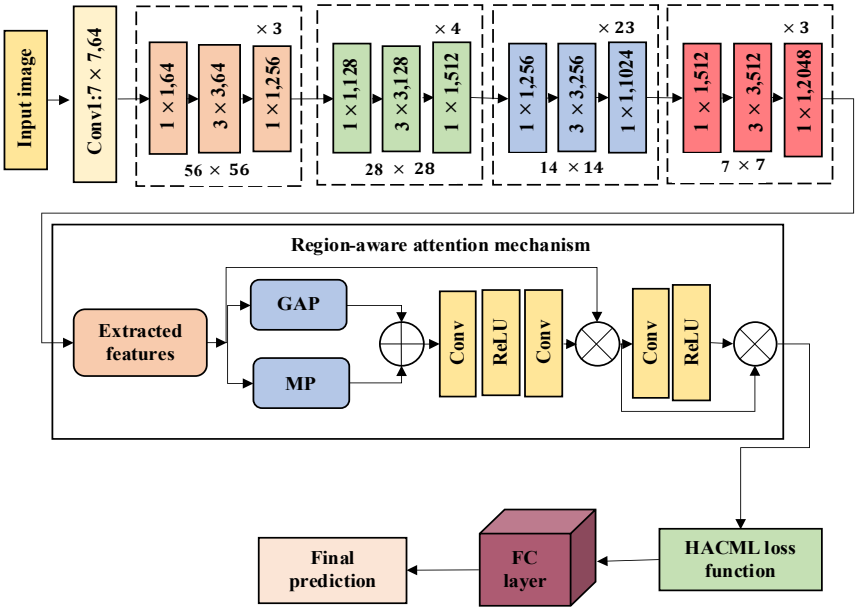

Fig. 3. Architecture diagram of the ResRAMACL model.

This vector is trained using the novel HACML function to apply inter-class separability and intra-class compactness, even under pose variations or visually similar faces. In the HACML, softmax loss normalizes [27] the gradient logarithm of a fixed element discrete probability distribution. This helps to divide features of different regions by maximizing the subsequent probability of the ground truth class. The mathematical formulation behind it is denoted in Eq. (6).

$$L = -\frac{1}{M}\sum_{u=1}^{M} log q_u = -\frac{1}{M}\sum_{u=1}^{M} log \frac{e^{W^S_{b_u} a_u + y_{b_u}}}{\sum_{v=1}^{m} e^{W^S_v a_u + y_v}} \quad (6)$$

where q_u represents the predicted probability of feature vector a_u . a_u signifies the u -th sample's feature representation, b_u signifies the equivalent class of a_u, W_v represents the v-th column of the weight matrix W, and y_v represents the bias term. $W^S_{v^{a_u+y_v}}$ represents the activated value of the a_u of the last FC layer merged with the weight vector W_v and y_v. M and m denote the number of training samples and regions, respectively. The various values of the a_u are determined through the cosine distance of the vectors. To enhance the face detection outcome, a_u is scaled to a finite value y through L_2 normalization. After normalization, the face detection outcome depends on the angle between the W_v and the a_u. The improved softmax loss is denoted as Normalized Softmax (NS) loss. The improved $W^S_{v^{a_u}}$ and the NS loss function are expressed as in Eqs. (7) and (8).

$$W^S_{v^{a_u}} = ycos\theta_v \quad (7)$$

$$L_{NS} = -\frac{1}{M}\sum_{u=1}^{M} log \frac{e^{ycos\theta_{bv}}}{e^{ycos\theta_{bv}} + \sum_{v=1, v\neq b_u}^{M} e^{cos\theta_v}} \quad (8)$$

Here, there is a chance for the NS loss function to lose its ability to perform feature learning for facial discrimination. M denotes the number of samples in a mini-batch and θ_v represents the learnable parameters of the network. Hence, we apply a margin to the softmax loss function to strengthen intra-class compactness and inter-class separability. To optimize both inter-class separability and intra-class compactness, the proposed HACML combines the properties of the Angular Margin (ArcFace) and Cosine Margin (CosFace) approaches. The proposed HACML loss function is represented in Eq. (9).

$$L_{HACML} = -\frac{1}{M}\sum_{u=1}^{M} log \frac{e^{(ycost\theta_{bv} - a)}}{e^{(ycost\theta_{bv} - m)} + \sum_{v=1, v\neq b_u}^{M} e^{cos\theta_v}} \quad (9)$$

where, t represents the multiplicative angular margin parameter, which signifies the multiple of the angle extension, a denotes the additive cosine margin parameter, which signifies the degree of increase of the cosine. The traditional ArcFace enforces an additive angular margin to explicitly control class boundary geometry. The traditional CosFace introduces an additive cosine margin to penalize similarity overlap in the normalized feature space. Recently, the hybrid margin approaches combine these strategies; however, most existing methods depend on heuristic linear combinations of multiple margin terms. This introduces conflicting gradients and unstable optimization behavior. The proposed HACML integrates angular and cosine margins within a single unified objective. This formulation allows angular margin constraints to regulate inter-class boundary geometry, while cosine margin constraints simultaneously suppress intra-class variance within the optimization step. The joint formulation ensures smooth convergence and stable decision boundaries, leading to improved robustness and generalization. The parameters t and a increase the angle margin, which enhances the training. This combination of parameters contracts the target angle using accepting both t and a penalties concurrently. By combining these two margins, HACML ensures that visually similar faces are projected, while those of the same identity remain grouped. Finally, the trained face recognition data are

stored in a structured SQL database. This allows efficient retrieval and integration into real-time applications like assistive systems for prosopagnosia patients.

C. SQL for Storing Face Recognition Data

In the proposed ResRAMACL framework, we have included a SQL database. An SQL database [28] is integrated to efficiently store, manage, and retrieve facial recognition data for future real-time applications. In this study, SQL database supports scalable storage, retrieval, and real-time operation of the proposed face recognition framework in assistive applications. After the model generates facial embeddings through the ResNet-101 with RAAM and HACML modules, these embeddings are stored in structured SQL. Each entry corresponds to an individual, and his/her's features and their normalized embedding vector. This provides a compact digital signature of the face. During real-time operation, the system retrieves embeddings that are stored, and upon receiving a new query embedding. This database supports continuous updates, allowing the system to add new identities or retrain existing profiles dynamically. Thus, SQL provides a secure and scalable data management backbone for the recognition framework, which also provides a smooth integration with real-time assistive technologies that facilitate prosopagnosia patients with their social interactions.

III. Results and Discussion

A. Hyperparameter Settings and Data Splitting

The proposed model utilized several hyperparameters to strike a balance between model performance and computational efficiency, as illustrated in Table I. The AFLW2000-3D dataset is split into 70% for training, 20% for validation, and 10% for testing. This ensures a fair and data leakage-free evaluation. The data split was performed at the identity level rather than the image level. The AFLW2000-3D dataset contains rich 3D pose annotations and multi-view facial variations. This makes it suitable for evaluating robustness to pose and geometric distortions. In this work, AFLW2000-3D is used to assess the proposed framework's ability to handle extreme pose variations. Experiments on VGGFace2 and the Similar Face Dataset during generalization, identities present in the training set were explicitly excluded from the evaluation sets.

TABLE I. Experimental Hyperparameter Settings for Proposed Resramacl Model

Component	Hyperparameter	Value
Backbone ResNet-101	Backbone	ResNet-101
	Attention FC	Linear (2048 → 1) + Sigmoid
	Stack features	Collect features from all regions
RAAM	Stack weights	Collect attention weights for all regions
	RAAM fusion	(16, 2048)
HACML	Margin	0.5
	Similarity	Cosine
	L2-regularization	0.00004
Training	Epochs	50
	Learning rate	0.0001
	Optimizer	Adam
	Batch size	16

B. Performance Analysis of the Proposed Model

The proposed ResRAMACL face recognition framework achieves a high detection accuracy. The proposed ResRAMACL framework was assessed using a few standard metrics. They are, namely, accuracy, precision, recall, F1-Score, specificity, False Positive Rate (FPR), False Negative Rate (FNR), Negative Predictive Value (NPV), and Area Under Curve (AUC). The proposed model achieves an accuracy of 99.57%, precision of 99.23%, recall of 99.63%, F1-Score of 98.46%, and specificity of 99.53%. Also, the proposed model attains a minimal false rate with an FPR of 0.52% and an FNR of 0.37%. The proposed model achieves an NPV of 99.66% and an AUC of 99.61%. These performance metrics highlight that the ResRAMACL framework enables prosopagnosia patients to recognize faces more effectively.

We evaluate the proposed framework using standard face recognition benchmarks, including verification and identification protocols. Face verification performance is assessed using the True Accept Rate (TAR) at fixed False Accept Rate (FAR) thresholds, as well as Receiver Operating Characteristic (ROC) curves at low FAR values. Identification performance is evaluated using Rank-1 identification accuracy. For verification, positive and negative face pairs are constructed following standard protocols, and cosine similarity between normalized embeddings is used for decision making. TAR is reported at FAR levels of 10^{-1}, 10^{-2}, 10^{-3}, and 10^{-4} to reflect high-security operating points, as represented in Table II. Identification performance is measured under a closed-set identification scenario, reporting Rank-1 accuracy.

TABLE II. Face Verification Performance

Model	TAR @ FAR = $1e^{-1}$	TAR @ FAR = $1e^{-2}$	TAR @ FAR = $1e^{-3}$	TAR @ FAR = $1e^{-4}$
ArcFace	97.42	96.82	95.12	94.81
CosFace	98.11	97.32	96.37	94.84
Proposed ResRAMACL	98.42	98.34	97.27	97.58

1) Results of facial landmark detection

Fig. 4 shows the detected key semantic regions, such as eyebrows, eyes, nose, mouth, and global face, using the 2D facial landmark plotting landmark-based method. The green dots represent the 2D facial keypoints detected on key regions of each face. The landmarks on the images define the spatial coordinates that support the model in highlighting the most cognitively relevant facial regions. These precise landmark detections enable the RAAM to extract different region-based features.

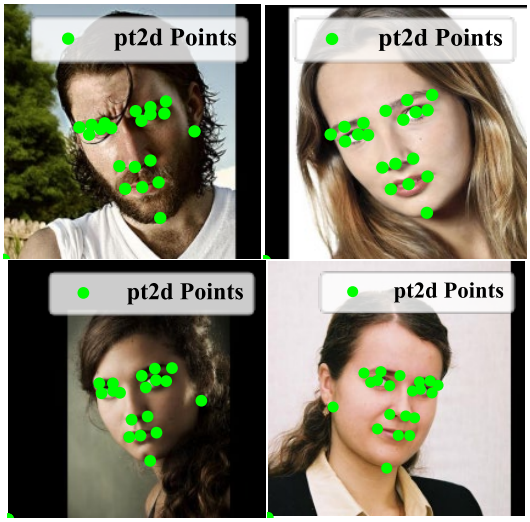

Fig. 4. Visualization of the detected facial regions.

This process allows the model to mimic human visual cognition, focusing on the region's necessary features for face identification. As shown in the figure, the accurate placement of these landmarks on both frontal and slightly angled faces proves that the proposed preprocessing pipeline effectively aligns faces.

2) Results of RAAM for focusing on important facial regions

Fig. 5 illustrates the attention weight distribution across key facial regions learned by the RAAM. The results show that the global face (0.21) and eyebrows (0.18) regions attain the highest attention weights. Nose (0.15), mouth (0.14), and eyes (0.16) regions attain the lower attention weights.

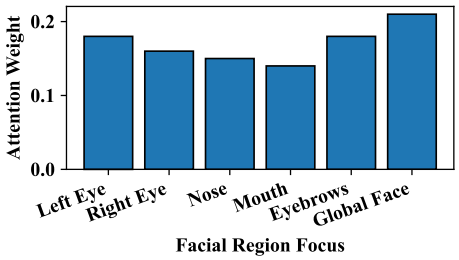

Fig. 5. Distribution of attention weights across facial regions.

This distribution enables the model to effectively learn to prioritize cognitively significant regions that are most crucial for human facial recognition. Highlighting these discriminative regions, the model reduces confusion between visually similar faces and enhances the interpretability of feature learning. This demonstrates that RAAM improves the quantitative performance and introduces cognitive consistency between the network's attention and human recognition behavior.

Table III shows the results of RAAM that focusing attention on specific cognitive facial regions significantly. All regions are combined adaptively through the RAAM mechanism, and the model achieves the best performance with 99% accuracy and 98.46% F1-Score. This ensures that region-aware learning enhances the discriminative feature representation of the face. This indicates that the RAAM successfully aligns the face, allowing the model to focus on regions most critical for identity recognition, especially in visually similar faces. The attention weight distributions reveal that RAAM does not excessively favor any single facial region across identities. While regions such as the eyes receive higher weights on average, attention allocation dynamically adapts based on feature discriminability. This adaptive behavior prevents bias toward highly variable regions and contributes to stable recognition across diverse facial characteristics.

TABLE III. PERFORMANCE EVALUATION OF REGION-AWARE ATTENTION MECHANISM

Focused Facial Region	Attention Weight	Recognition Accuracy (%)	Visually Similar Face Accuracy (%)	F1-Score (%)
Left eye	0.18	95.5	93.9	94
Right eye	0.16	94.9	92.8	93
Nose	0.15	93.5	91.9	92
Mouth	0.14	94.8	92.6	93
Eyebrows	0.18	93.8	91.5	92
Global Face	0.21	97.2	95.0	96
Combined Regions (RAAM)	Adaptive Weighted Fusion	99.0	97.9	98.46

Fig. 6 explains the ROC curve and its corresponding AUC values to provide a way to evaluate the performance of the ResRAMACL model that marks the true positive rate as opposed to the false positive rate. The AUC values are between the range 0 to 1, with larger values indicating better performance. It highlights the estimation of how well a model can differentiate between positive and negative rates in analyzing the impact of changing thresholds on detection accuracy.

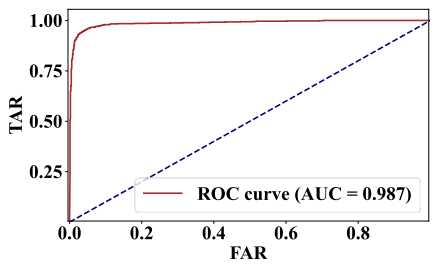

Fig. 6. ROC curve analysis of the proposed framework.

3) Analysis of proposed HACML loss function

Table IV analyses the discriminative strength of the proposed HACML, compared against Softmax, ArcFace, and CosFace loss functions. The proposed model attains the highest inter-class separation (6.93) and lowest intra-class variance (0.88) compared to softmax, ArcFace, and CosFace loss functions. This highlights the superior discriminative capacity of the proposed HACML loss function.

Fig. 7 presents the pose-invariant recognition accuracy of the proposed ResRAMACL model compared with ArcFace and CosFace. Yaw angle represents the difference between an object's heading and its actual direction of travel. This angle ensures the stability and performance of the proposed model. The model was compared with different yaw angles (0°, 30°, 60°, and 90°). The accuracy of ArcFace and CosFace significantly drops as the yaw angle increases, with values from 97.8% to 84.7% and 98.1% to 86.0%, respectively. The proposed ResRAMACL framework maintains a consistently high accuracy, decreasing only from 98.5% at 0° to 94.3% at 90°. This stability demonstrates the better pose and view invariance of the proposed framework. The combination of RAAM and HACML allows the model to extract robust, view-independent features even under large facial rotations.

Fig. 8 illustrates the curve of the HACML loss per epoch size. As shown in this figure, the loss function changes from high to low with the number of training epochs of the proposed ResRAMACL model. The model has a bigger loss value throughout initial training to speed up the gradient descent process of the optimizer to select a better solution in the global domain. As the number of epochs increases, it is found that the loss value decreases. This shows how the loss value decreases smoothly through a HACML loss function.

TABLE IV. COMPARISON OF LOSS FUNCTIONS ON INTER-CLASS SEPARABILITY AND INTRA-CLASS COMPACTNESS

Loss Function	Inter-Class Distance	Intra-Class Variance	Accuracy (%)	F1-Score (%)
Softmax	3.12	1.85	91.1	90.8
ArcFace	4.77	1.32	94.7	94.6
CosFace	4.64	1.27	95.8	95.4
HACML (Proposed)	6.93	0.88	99.0	99.0

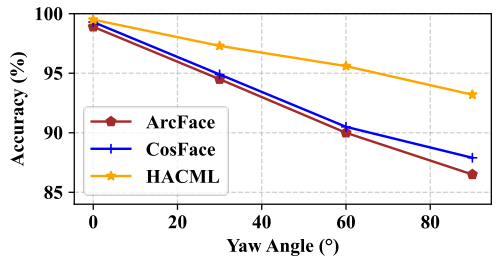

Fig. 7. Pose-invariant recognition accuracy of the proposed HACML functions with ArcFace and CosFace functions.

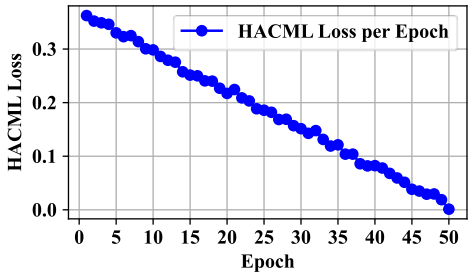

Fig. 8. Visualization of the HACML loss per epochs curve.

To evaluate the robustness of the proposed HACML, sensitivity analysis used by varying its key hyperparameters, such as the angular margin, the cosine margin, and the scaling factor. Table V exhibits stable performance across a wide range of margin and scale configurations. Moderate variations in angular and cosine margins result in marginal performance fluctuations of less than 0.4%, indicating that HACML does not depend on fine-tuned hyperparameter settings. The optimal configuration balances inter-class separation and intra-class compactness, while neighboring configurations maintain comparable performance. This confirms the robustness and practical deployability of the proposed hybrid margin formulation.

TABLE V. SENSITIVITY ANALYSIS OF HACML HYPERPARAMETERS

Angular Margin	Cosine Margin	Scale Factor (s)	Accuracy (%)	AUC
0.35	0.20	32	98.85	0.9921
0.40	0.30	40	99.12	0.9943
0.50	0.35	64	99.57	0.9967
0.55	0.40	64	99.26	0.9954
0.60	0.45	70	98.74	0.9911

The cosine margin parameter a a vital aspect in HACML, and the performance became optimal when a = 0.25, as shown in Fig. 9. To examine the influence of a in our proposed ResRAMACL framework, experiments were conducted to examine the outcome under the parameter a. The AFLW2000-3D dataset was utilized to train the HACML and assess its performance for a ranging from 0.15 to 0.35. Fig. 9 shows the outcome of angular margin parameter t in HACML. The outcome of the parameter t, the cosine margin parameter was set as $a = 0$.

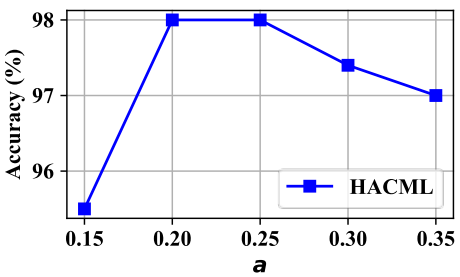

Fig. 9. Accuracy of HACML under different cosine margin values on AFLW2000-3D.

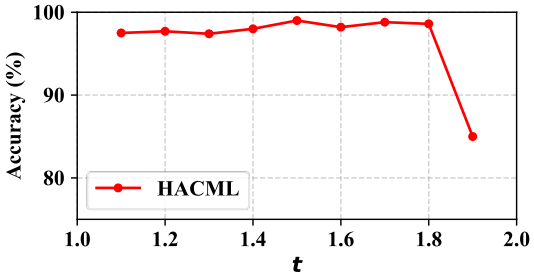

Fig. 10. Accuracy of HACML under different angular margin values on AFLW2000-3D.

As represented in Fig. 10, we found that, for values of t from 1.1 to 1.8, the accuracy is maintained stable. This confirmed that when the degree of the angular margin parameter was altered to a definite value, the recognition outcome was enhanced.

4) A look on the SQL database

Fig. 11 illustrates the SQL database interaction of the proposed ResRAMACL framework to store and manage facial data for recognition. The image on the left represents the input face processed by the ResRAMACL framework, while the table on the right shows the numerical vector generated by the network. These feature values are multidimensional numerical representations that encode the unique characteristics of the face. Each element corresponds to a dimension in the learned feature space, allowing the system to differentiate this identity from others through cosine similarity. These feature vectors are stored in the SQL database. This supports the system to operate in real time. This enables instantaneous identification or verification of the person. This mechanism ensures fast retrieval, efficient storage, and accurate real-time recognition, making it useful for future assistive technologies for prosopagnosia patients.

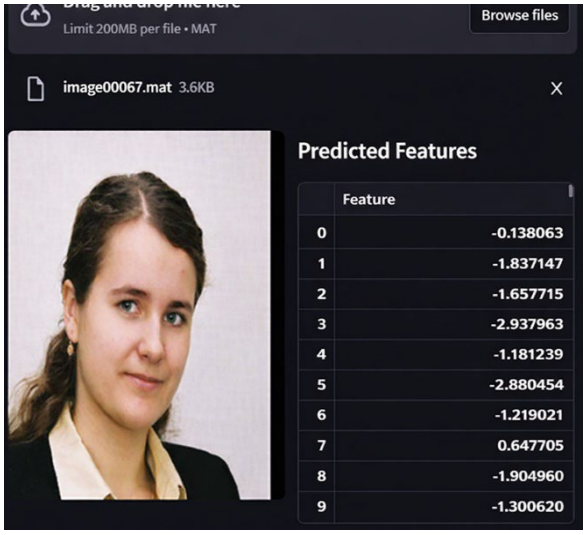

Fig. 11. Visualization of predicted feature values associated with the given image shows the SQL storage to store and manage facial data for recognition.

C. Ablation Study

Table VI analyzes the configuration of ResNet-101 with different blocks and analyzes the integrated components of RAAM and HACML. The table highlights the significant contributions of the proposed pipeline with components ResNet-101, RAAM, and HACML with a highest accuracy of 99.57%, precision of 99.23%, recall of 99.63%, and F1-Score of 98.46%.

TABLE VI. ABLATION STUDY ANALYSIS

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed ResRAM	99.57	99.23	99.63	98.46
ResNet-101 replaced with ResNet50	97.84	96.95	97.10	96.75
ResNet-101 replaced with ResNet108	98.12	97.81	97.95	97.59
ResNet-101 without attention	95.12	96.34	96.24	96.36
ResNet-101 with global attention	97.12	97.12	97.82	97.45
Without RAAM	95.64	94.72	95.11	94.66
RAAM replaced by Convolutional Block Attention Module	96.85	96.25	96.35	95.95
Without angular margin	97.02	96.16	96.49	95.84
Without cosine margin	97.30	96.85	96.70	96.45
Without 2D facial landmark plotting	98.45	98.28	98.14	97.91

D. Generalization Performance Analysis of the Proposed Model

The proposed ResRAMACL framework was evaluated on three different benchmark datasets: Face recognition dataset [29], Similar face dataset [30], and VGGFace2 dataset [31] to ensure generalizability of the proposed model. The Face recognition dataset contains a diverse set of individuals with neutral expressions, lighting, and frontal poses. The Similar face dataset is designed to challenge recognition models with similar faces, individuals who share closely similar facial regions such as eye shape, jawline, and skin tone. The VGGFace2 dataset is a large-scale and widely recognized benchmark for deep face recognition research. Utilization of all three datasets provides a multi-dimensional validation framework that ensures real-world applicability under various conditions.

Table VII analyses the generalizability of the proposed model by validating our proposed model using three other face recognition datasets. This table displays that our proposed ResRAMACK framework is better generalized across several face recognition datasets. This effective validation of our proposed framework through the external dataset emphasizes its potential as a reliable tool for real-time face recognition for prosopagnosia patients.

TABLE VII. GENERALIZATION PERFORMANCE OF THE PROPOSED MODEL

Metrics	Face Recognition Dataset	Similar Face Dataset (SFD)	VGGFace2 Dataset	Proposed with AFLW2000-3D Dataset
Accuracy (%)	98.29	97.63	98.50	99.57
Precision (%)	97.11	96.88	97.96	99.23
Recall (%)	97.37	96.92	98.15	99.63
F1-Score (%)	97.22	96.81	98.03	99.46

E. Comparative Analysis

The proposed model is compared with the existing state-of-the-art approaches, which were using models like VGGNet, KNN, and Quantum CNN and achieved an accuracy of 92.8%, 98.11%, and 78.70%, respectively. While the proposed model achieved an accuracy of 99.57%, it highlights the proposed ResRAMACL framework's effectiveness in face recognition, as represented in Table VIII.

TABLE VIII. COMPARATIVE ANALYSIS WITH EXISTING APPROACHES

Reference	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)
Mukhiddinov and Chao [16] 2021, with VGGNet	92.8	86.3	81.7	-
Bardak and Temurtas [19] 2025, with KNN	98.11	-	98.08	97.92
Hossain *et al*. [10] 2024, with Quantum CNN	78.70	63.52	61.55	61.37
Proposed with ResNet-101 (ResRAMACL)	99.57	99.23	99.63	99.53

F. Discussion

The proposed ResRAMACL framework overcomes the limitations of existing face recognition systems for prosopagnosia patients. The datasets were partitioned using identity-disjoint splits, ensuring that no subject appears in more than one of the training, validation, or test sets. This identity-level separation prevents implicit memorization and guarantees that performance metrics reflect genuine generalization. Early stopping based on validation loss prevents overfitting to the training data. Additionally, the proposed Hybrid HACML contributes to regularization by enforcing structured embedding constraints. By jointly optimizing angular separation and cosine similarity margins, HACML reduces intra-class dispersion without collapsing embeddings, leading to improved generalization. The robustness of this model is further validated through cross-dataset evaluation, where the model trained on AFLW2000-3D demonstrates consistent performance on larger unseen datasets such as the face recognition dataset, VGGFace2, and the Similar Face Dataset. The VGGFace2 dataset contains millions of images with substantial diversity in age, illumination, and capture conditions. The consistent performance observed across these datasets indicates that the proposed framework does not depend on dataset-specific identity patterns and generalizes effectively to unseen identities and acquisition conditions.

The experimental findings indicate that the proposed ResRAMACL framework is effective in addressing the major issues of unconstrained face recognition, such as high intra-class variance, subtle inter-class similarity, pose variation, and partial occlusion. ResNet-101 is used as the backbone and helps extract features well, and the model can extract low-level features, texture, and high-level semantic features of the face. Nevertheless, the default ResNet-101 model demonstrates a performance weakness in the case of untrained models without the explicit attention or margin constraints. To overcome these problems, the RAAM provides adaptive weights to capture cognitively important facial areas. This effect is quantitatively validated by the observed reduction in intra-class variance. The attention-guided feature aggregation ensures that identity-relevant regions dominate the embedding representation, leading to more stable recognition performance across challenging conditions. In parallel, the proposed Hybrid HACML plays a critical role in enhancing embedding discriminability. Compared to single-margin approaches, HACML jointly optimizes angular separation and cosine similarity constraints. This dual-margin enforces compact intra-class clustering and maximizes inter-class margins. The proposed framework achieves an average inference time of 38ms per image on an NVIDIA GPU, corresponding to over 25 frames per second. The region extraction and attention modules introduce minimal overhead due to parallel processing. These results confirm that the proposed system is computationally viable for real-time assistive applications, including mobile and AR-based identity support systems. The proposed HACML offers improved AUC and F1-Score values, as well as clearer inter-class separation visualized in the embedding space. The proposed framework demonstrates strong generalization capability. The model is tested on external datasets such as the Face recognition dataset, VGGFace2, and the Similar Face Dataset, with high recognition accuracy and minimal decline in performance. This indicates that the learned representations are not dataset-specific and do not depend on identity memorization. The combination of region-aware attention and margin-based supervision enables the model to effectively capture facial characteristics. This confirms that the proposed ResRAMACL framework effectively fulfills the objectives. These findings validate the proposed framework design as a reliable and scalable assistive face recognition system suitable for real-world applications.

IV. CONCLUSION

This study successfully developed a proposed framework for assistive face recognition support in patients with prosopagnosia. The proposed ResRAMACL

framework combines a ResNet-101 backbone with a RAAM and a HACML, effectively producing discriminative feature learning, region-level interpretability, and pose invariance. The proposed HACML enhanced both inter-class separability and intra-class compactness compared to traditional loss functions such as ArcFace and CosFace. The experimental results and ablation studies confirm the effectiveness, robustness, and generalization capability of the proposed framework across multiple datasets. The proposed framework does not model the cognitive mechanisms of prosopagnosia. It is designed as an assistive recognition system that can function as an external cognitive aid. These systems have the potential to support identity recognition in daily interactions, reduce social anxiety, and enhance the quality of life of individuals with prosopagnosia. The study also highlights transparency, robustness, and ethical deployment by addressing data leakage, overfitting, and bias considerations. This ensures the proposed ResRAMACL framework and advances the field of deep face recognition, providing a reliable, interpretable, and practical solution for assisting prosopagnosia patients in recognizing faces in everyday social environments.

The study includes some limitations in this proposed framework. The study faced limitations in real-time data management when integrated with the SQL-based storage system. The database grows with thousands of embeddings, query retrieval time and memory overhead may increase, affecting the real-time response rate of applications such as AR-assisted recognition for prosopagnosia patients. Also, we have now developed a framework for a face recognition system for prosopagnosia patients. The absence of datasets involving individuals diagnosed with prosopagnosia is one of the limitations of this study. Future work will involve collaboration with clinicians and cognitive scientists to assess usability and social effectiveness in real-world assistive scenarios. In the future, we will expand our framework to practical applications for several social conditions that patients with prosopagnosia find themselves in, using AR-assisted technologies or real-time mobile applications. The future enhancements enable the ResRAMACL framework to support adaptive, personalized, and scalable real-time cognitive assistance applications.

Code Availability Statement

The source code used to implement the proposed framework was developed through background coding using open-source libraries. The code is currently maintained in a private repository and can be made available to interested researchers upon request to the corresponding author, subject to ethical guidelines. https://github.com/project12317/project_08s.git

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors contributed to the conception of the problem setting, overall design of the work, and visualization of concepts. BN built the conceptualization and methodology; RM implemented the work, and writing. This version was revised and improved by all authors. all authors had approved the final version.

References

[1] R. J. Bennetts, N. J. Gregory, J. Tree *et al.,* "Face specific inversion effects provide evidence for two subtypes of developmental prosopagnosia," *Neuropsychologia*, vol. 174, 108332, 2022.

[2] N. I. Ivanova, D. M. Kyuchukova, M. E. Tsalta-Mladenov *et al.,* "Prosopagnosia due to metastatic brain tumor: A case-based review," *Cureus*, vol. 16, e55349, 2024.

[3] E. J. Burns, E. Gaunt, B. Kidane *et al.,* "A new approach to diagnosing and researching developmental prosopagnosia: Excluded cases are impaired too," *Behavior research methods*, vol. 55, pp. 4291–4314, 2023.

[4] S. Vaijayanthi and J. Arunnehru, "Facial expression recognition using hyper-complex wavelet scattering and machine learning techniques," in *Proc. the 6th International Conf. on Advance Computing and Intelligent Engineering*, 2022, pp. 411–421.

[5] K. A. Josephs and K. A. Josephs Jr, "Prosopagnosia: Face blindness and its association with neurological disorders," *Brain Communications*, vol. 6, no. 1 fcae002, 2024.

[6] B. Rossion, "Twenty years of investigation with the case of prosopagnosia PS to understand human face identity recognition. Part I: Function," *Neuropsychologia*, vol. 173, 108278, 2022.

[7] K. Imatani, T. Inoue, Y. Oto *et al.,* "Generalized anxiety disorder and depression associated with developmental prosopagnosia: A case report," *Journal of Mental Health & Clinical Psychology*, vol. 7, pp. 50–54, 2023.

[8] J. Lowes, L. M. McGregor, P. J. Hancock *et al.,* "This condition impacts every aspect of my life: A survey to understand the experience of living with developmental prosopagnosia," *PloS One*, vol. 20, e0322469, 2025.

[9] P. E. Emhjellen, R. Starrfelt, R. Raudeberg *et al.,* "Assessment of developmental prosopagnosia in an individual with Tourette syndrome and attention deficit hyperactivity disorder: A case report," *Brain Sciences*, vol. 15, 56, 2025.

[10] S. Hossain, S. Umer, R. K. Rout *et al.,* "A deep quantum convolutional neural network based facial expression recognition for mental health analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 1556–1565, 2024.

[11] S. Klauke, C. Sondocie, and I. Fine, "The impact of low vision on social function: The potential importance of lost visual social cues," *Journal of Optometry*, vol. 16, pp. 3–11, 2023.

[12] S. Byrne and M. Porter, "Rehabilitation and intervention of developmental and acquired prosopagnosia: A systematic review," *Neuropsychological Rehabilitation*, vol. 35, no. 9, pp. 1715–1758, 2025.

[13] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," *IET biometrics*, vol. 10, pp. 581–606, 2021.

[14] W. Ali, W. Tian, S. U. Din *et al.,* "Classical and modern face recognition approaches: A complete review," *Multimedia Tools and Applications*, vol. 80, pp. 4825–4880, 2021.

[15] S. Vaijayanthi and J. Arunnehru, "Deep neural network-based emotion recognition using facial landmark features and particle swarm optimization," *Automatika: Časopis za Automatiku, Mjerenje, Elektroniku, Računarstvo i Komunikacije*, vol. 65, no. 3, pp. 1088–1099, 2024.

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

[17] 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, 1244, 2025.

[18] S. Mann, Z. Pan, Y. Tao *et al.,* "Face recognition and rehabilitation: A wearable assistive and training system for prosopagnosia," in

Proc. 2020 IEEE International Conf. on Systems, Man, and Cybernetics (SMC), 2020, pp. 731–737.
[19] 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, 2025.
[20] J. Xu, X. Liu, X. Zhang *et al.,* "X2-softmax: Margin adaptive loss function for face recognition," *Expert Systems with Applications*, vol. 249, 123791, 2024.
[21] J. Wang, C. Zheng, X. Yang *et al.,* "Enhanceface: Adaptive weighted SoftMax loss for deep face recognition," *IEEE Signal Processing Letters*, vol. 29, pp. 65–69, 2021.
[22] X. Ning, F. Nan, S. Xu *et al.,* "Multi-view frontal face image generation: A survey," *Concurrency and Computation: Practice and Experience*, vol. 35, e6147, 2023.
[23] M. Adly. (2024). AFLW2000-3D Dataset. *Kaggle*. Available: https://www.kaggle.com/datasets/mohamedadlyi/aflw2000-3d
[24] P. Chandran, G. Zoss, P. Gotardo *et al.,* "Infinite 3D landmarks: Improving continuous 2D facial landmark detection," in *Computer Graphics Forum*, vol. 43, no. 6, e15126, 2024.
[25] S. Ali and J. Agrawal, "Automated segmentation of brain tumour images using deep learning-based model VGG19 and ResNet 101," *Multimedia Tools and Applications*, vol. 83, pp. 33351–33370, 2024.
[26] Z. Huang, S. Cheng, and L. Wang, "Medical image segmentation based on dynamic positioning and region-aware attention," *Pattern Recognition*, vol. 151, 110375, 2024.
[27] H. Deng, Z. Feng, G. Qian *et al.,* "MFCosface: A masked-face recognition algorithm based on large margin cosine loss," *Applied sciences*, vol. 11, 7310, 2021.
[28] M. Salahuddin, S. Majeed, S. Hira *et al.,* "A systematic literature review on performance evaluation of SQL and NoSQL database architectures," *Journal of Computing & Biomedical Informatics*, vol. 7, no. 2, 2024.
[29] V. Patel. (2021). Face Recognition Dataset. *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/vasukipatel/face-recognition-dataset
[30] A. Song. (2020). Similar Face Dataset. *Figshare*. [Online]. Available: https://figshare.com/articles/dataset/Similar_Face_Dataset_SFD_/11611071
[31] VGGFace2 Dataset. *Kaggle*. Available: https://www.kaggle.com/datasets/hearfool/vggface2