Int. J. Nonlinear Anal. Appl. In Press, 1–22
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2025.39129.5550

Hybrid CNN–ELM model for robust post-cosmetic surgery face recognition

Manijeh Azadeh[a], Farhad Navabifar[b,*]

[a]Department of Electrical Engineering, Isf.C., Islamic Azad University, Isfahan, Iran

[b]Department of Electrical Engineering, Mo.C., Islamic Azad University, Mobarakeh, Isfahan, Iran

(Communicated by Majid Eshaghi Gordji)

Abstract

Cosmetic surgery introduces substantial structural changes to facial features, often leading to reduced accuracy in conventional face recognition systems. To address this challenge, we present a hybrid recognition framework that integrates Convolutional Neural Networks (CNNs) for deep feature extraction with an Extreme Learning Machine (ELM) for classification. The CNN component captures surgery-invariant hierarchical features, while the ELM provides faster convergence and improved generalization compared to fully connected layers. The proposed method was evaluated using the IIIT-Delhi Plastic Surgery Face Database, which includes 1,800 images from 900 subjects across 13 types of local and global procedures. Results indicate significant performance gains across all key metrics. For local surgeries, the model achieved average Recall, Precision, and Accuracy of 95.3%, 95.1%, and 95.2%, respectively, compared to 81.3%, 81.2%, and 82.6% for a conventional CNN. Specificity and F-measure improved to 94.6% and 94.8%, reflecting enhanced reliability and reduced false positives. For global procedures such as rhytidectomy and skin resurfacing, the model maintained accuracy above 93%, exceeding the baseline by over 10%. These findings demonstrate that the CNN–ELM hybrid model offers superior robustness to post-surgical variations, making it a promising solution for biometric verification in forensic investigations, healthcare applications, and security systems.

Keywords: Cosmetic surgery, Facial recognition, Convolutional Neural Networks (CNN), Extreme Learning Machine (ELM), Biometric verification, Post-surgery identity recognition
2022 MSC: Primary 68T10; Secondary 68T07, 68U10

1 Introduction

Numerous organizations employ biometric systems to strengthen security measures for identity verification and surveillance. These systems leverage unique physical or behavioral traits to authenticate individuals, capturing and storing raw biometric data for subsequent analysis and recognition [25]. Commonly used biometric identifiers, such as fingerprints, signatures, and facial features, are archived and accessed during identity verification processes [23, 3, 14]. These identifiers are versatile enough to be utilized across various security and surveillance applications. Among biometric modalities, fingerprints are the most prevalent, though other traits like facial characteristics, iris patterns, palm prints, voice, and signatures are also significant [9, 12]. Each biometric approach offers distinct

*Corresponding author
Email addresses: azinertebatat3000@gmail.com (Manijeh Azadeh), farnav@iau.ac.ir (Farhad Navabifar)

Received: September 2025 *Accepted:* October 2025

benefits and challenges, with some being more user-friendly in specific contexts while posing difficulties in others [7]. For instance, fingerprints, handprints, and signatures require active user participation and physical proximity during enrollment, potentially raising privacy issues. Conversely, iris scans and facial recognition operate without direct interaction, though they often involve higher computational demands compared to fingerprints or signatures [9]. Facial recognition, in particular, is appreciated for its non-invasive nature and efficiency, making it a popular choice in real-world applications, despite privacy concerns related to passive data collection [13].

Facial recognition technology, while advantageous, encounters several obstacles that limit its broad adoption. Factors such as aging, emotional expressions (e.g., smiling, frowning, or anger), and occlusions like glasses or hats can alter facial features, complicating accurate identification [5]. Additional variables, including weight changes, hormonal shifts, and physical injuries, further challenge recognition accuracy [5, 8, 18]. As facial recognition gains traction in public and private sectors, addressing these issues is critical to improving verification reliability. While prior research has extensively explored challenges like lighting variations and occlusions [17], the impact of cosmetic surgery on facial recognition remains underexplored, forming the core focus of this study.

Accurate identity verification post-cosmetic surgery is essential across clinical, legal, and security domains. In healthcare, precise identification following surgery is crucial for patient safety, effective treatment planning, and preventing medical errors, particularly in cases involving reconstructive procedures, facial tissue transplants, or corrections of congenital deformities or injuries [20]. Misidentification in these scenarios could lead to severe health risks. In legal contexts, cosmetic surgery can significantly alter an individual's appearance, potentially causing traditional facial recognition systems to fail, which poses challenges in criminal investigations or legal proceedings where mistaken identities could result in serious consequences, such as wrongful convictions or failure to identify suspects in surveillance footage. Security applications, including border control, airport security, and surveillance systems, face heightened risks, as individuals may undergo cosmetic procedures to obscure their identity, either for fraudulent purposes or to evade detection. These issues underscore the need for advanced identity verification systems capable of handling significant facial alterations [28].

The growing accessibility of cosmetic surgery, driven by technological advancements and increased affordability, has fueled its popularity [19]. Procedures such as rhinoplasty, blepharoplasty, chin augmentation, and complex facial reconstructions are now more widely available, resulting in facial changes ranging from subtle to transformative. Localized procedures like nose or eyelid corrections may produce minor alterations, whereas extensive reconstructions or gender-affirming surgeries can drastically modify appearance [6]. This trend amplifies the importance of reliable post-surgical identity verification, especially in high-stakes environments like airports, embassies, and government facilities, where misidentification could have severe implications [15].

The rise in cosmetic surgery introduces new security challenges. While traditional facial recognition systems have addressed issues like lighting, occlusions, and expressions [2], the structural changes caused by procedures such as brow lifts or rhinoplasty pose unique difficulties. These alterations can disrupt key facial features, reducing the effectiveness of conventional recognition methods. Reliable identification post-surgery is particularly critical in settings like criminal investigations and border security, where facial features are a primary biometric identifier. Cosmetic surgery not only serves aesthetic purposes but also addresses medical needs, such as correcting facial deformities or trauma-related injuries. Various cosmetic procedures significantly alter facial geometry and tissue structures, impacting traditional identity verification methods [16].

Biometric recognition involves multiple stages: pre-processing, feature extraction, feature selection, classification, and evaluation [27]. Pre-processing includes tasks like noise reduction and image enhancement, while feature extraction uses descriptors to capture unique facial characteristics. Common methods include transformation-domain techniques like discrete Fourier transform [1], discrete wavelet transform [21], and discrete cosine transform [6], alongside spatial-domain approaches analyzing statistical features or histograms [24]. Additional methods, such as local binary patterns [4], spectral features [10], and graph-based features [27], are also employed. While effective in standard scenarios, these methods struggle with image alterations caused by surgery, as even minor changes can significantly affect feature vectors, reducing accuracy [22]. Deep learning offers a promising solution by learning robust features less sensitive to facial geometry changes. Convolutional neural networks (CNNs) are particularly effective for generating discriminative feature vectors for classification [26, 11]. This study proposes a hybrid approach using CNNs for feature learning and an Extreme Learning Machine (ELM) for classification to improve post-surgical identity recognition accuracy.

Post-cosmetic surgery facial recognition faces significant challenges due to structural changes that disrupt traditional methods. This research presents a novel approach to enhance the accuracy and robustness of post-surgical facial recognition, addressing limitations in existing techniques. A specialized recognition system is developed to accurately identify individuals despite major facial modifications. Challenges such as image noise, lighting variations, facial

rotations, and emotional expressions often impair conventional systems. The proposed method integrates advanced feature extraction and deep learning to improve adaptability to diverse post-surgical images. Instead of a traditional fully connected neural network, an Extreme Learning Machine is used for classification, enhancing generalization and performance. The paper is organized as follows: Section 2 reviews related work in post-surgical identity recognition. Section 3 describes the proposed methodology, including deep learning techniques and novel neural architectures. Section 4 assesses the system's performance using standard metrics, evaluating its robustness against common recognition challenges. Section 5 concludes with key findings and future research directions.

2 Literature Review

Facial recognition in the context of plastic surgery presents unique challenges for biometric systems. Unlike other covariates such as pose, illumination, or aging, surgical interventions deliberately and permanently alter facial structures, often disrupting the reliability of feature extraction and template matching. This issue has driven extensive research spanning traditional algorithms, hybrid feature engineering, and, more recently, deep learning approaches. The progression of these works illustrates both the evolution of methodology and the persistent challenges that continue to limit practical deployment in high-stakes applications.

Singh et al. [25] conducted a critical survey and empirical evaluation on the effects of plastic surgery in face recognition, demonstrating that classical algorithms such as Eigenfaces and Fisherfaces experienced accuracy reductions of up to 50% when matching pre- and post-surgery images from a dataset of 450 subjects. Their study highlighted the necessity for surgery-invariant feature extraction. In a related effort, Singh et al. [23] introduced one of the first datasets of surgically altered faces. They showed that local feature-based approaches, such as Gabor filters, alleviated some effects of localized surgery but performed poorly under global alterations, achieving only 65% recognition rates. Bhatt et al. [3] proposed a hybrid method combining Speeded-Up Robust Features (SURF) for local descriptor extraction with back-propagation neural networks (BPNN) for classification. Their system achieved 82% accuracy for localized surgeries (e.g., rhinoplasty) but dropped to 60% for extensive modifications, outperforming PCA by 20% in equal error rate (EER).

Liu et al. [14] applied sparse representation together with multi-scale local binary patterns (LBP) to address post-surgery recognition. Their framework achieved up to 78% recognition accuracy on Asian subjects, although at the expense of high computational complexity. Ibrahim et al. [12] compared PCA, ICA, and LDA on a dataset of post-surgery faces. ICA proved the most resilient with 72% accuracy compared with 55% for PCA, leading the authors to recommend ensemble strategies for better generalization. Devi and Marimuthu [9] developed a self-updating recognition system using adaptive boosting and SVM. Their method achieved 85% accuracy on dynamic post-surgery galleries and emphasized the value of incremental learning in handling progressive facial changes.

Dadure et al. [7] integrated LBP and PCA within a shallow CNN architecture. Their model reported 88% accuracy on combined AT&T and surgery datasets, though it showed limitations when occlusions and surgical scars were present. Khedgaonkar et al. [13] fine-tuned the VGG-16 CNN model on 1,000 pre-/post-surgery images. The system attained 92% rank-1 accuracy, surpassing SURF-based approaches by 15%, though training time was identified as a key bottleneck. Bouguila and Khochtali [5] reviewed 45 studies addressing facial plastic surgery and biometrics. They identified texture distortion in nearly 70% of cases as a major obstacle and recommended multimodal biometrics (e.g., combining face and iris). Their review emphasized that deep learning models outperform shallow methods by 10–20% in resilience.

Rathgeb et al. [18] investigated deep recognition models such as ArcFace and SphereFace under surgical conditions. Their experiments demonstrated minimal degradation (FNMR <1% at 0.01% FMR) compared with shallow methods, attributing the robustness to learned embeddings. Damer et al. [8] evaluated FaceNet on the HDA Plastic Surgery Database, which includes 638 subjects. Their analysis reported 95–100% accuracy for non-extreme surgeries but significant performance drops (to 80%) in bone reconstruction cases. The authors suggested synthetic surgery augmentation to enhance robustness. Ranjanikar and Kulkarni [17] proposed a hybrid CNN–ELM framework for post-surgery recognition. Their method achieved 94% accuracy on mixed datasets and reduced inference time by 40% compared to full end-to-end CNNs, offering a promising trade-off between speed and accuracy.

Sabharwal [20] developed a forensic recognition system using score-level fusion of SIFT and HOG features. The method achieved 90% accuracy and demonstrated that score-level integration is more effective than feature-level fusion in post-surgery conditions. Zhang et al. [28] conducted a systematic review of machine learning methods in cosmetic surgery contexts. They found that CNN–Gabor hybrids achieved the highest accuracies (up to 96%) for feminization outcomes, though integration with ELM remained underexplored. Rathgeb et al. [19] extended the

analysis of aesthetic modifications, including surgery, to commercial recognition systems. Their experiments achieved 91.75% rank-1 accuracy and showed that dictionary learning techniques can support detection and verification.

Collectively, these studies demonstrate the trajectory from handcrafted features and shallow hybrid models toward deep learning methods that now achieve recognition rates exceeding 90%. Yet several gaps remain: early methods lacked robustness to global changes, hybrid approaches suffered from computational inefficiency, and deep learning solutions risk overfitting on limited surgery-specific datasets while requiring long training times. Only limited attempts have been made to integrate efficient classifiers, such as ELM, with deep feature representations. the present study introduces a novel CNN–ELM framework that leverages the hierarchical feature learning of CNNs for robust surgery-invariant embeddings while exploiting the speed and generalization of ELM for efficient classification. This approach is expected to provide higher accuracy with significantly reduced training time, positioning it as a practical solution for high-stakes applications such as border control and forensic investigations.

3 Materials and Methods

This section presents a comprehensive and mathematically rigorous methodology for the proposed hybrid CNN–ELM model aimed at facial recognition following cosmetic surgery. The framework integrates data preparation, preprocessing, feature extraction using Convolutional Neural Networks (CNNs), and classification via Extreme Learning Machines (ELM). Each step is elaborated with formal definitions, equations, and explanations of the parameters involved.

3.1 Dataset

In this study, the IIIT-Delhi Plastic Surgery Face Database (IIIT-PSFDB) is employed to evaluate the proposed facial recognition method for post-surgery individuals. The dataset consists of:

$$N = 1800 \text{ images from } S = 900 \text{ subjects.}$$

Each subject contributes two images: one pre-surgery and one post-surgery. The images cover a variety of cosmetic procedures, including rhinoplasty, blepharoplasty, facial feminization surgery, chin augmentation, and other localized or global facial alterations. All images are originally stored in JPEG format with resolution approximately 640×480 pixels and are resized to 224×224×3224 to match the input requirements of the CNN:

$$x_i \in y, \qquad \{1, 2, \ldots, S\} \in \mathbb{R}_i^{3\times 224\times 224},$$

where x_i represents the i-th input image. Each image is paired with an identity label $y_i \in \{1, 2, \ldots, S\}$. The paired pre- and post-surgery images enable the model to learn identity-preserving features despite structural changes caused by cosmetic surgery. Formally, the dataset can be expressed as:

$$D = \{(x_i, y_i) |\ i = 1, 2, \ldots, N\}.$$

The dataset diversity in terms of surgery type and temporal post-surgery stages ensures robustness of the learned representations across different surgical modifications.

The IIIT-Delhi Plastic Surgery Face Database, with 1,800 images from 900 subjects, is limited to controlled frontal images with neutral expressions, potentially failing to fully reflect real-world variability (e.g., pose changes, lighting, or multi-ethnic features). This ethnic homogeneity and lack of diversity may introduce biases in the model's generalizability. It is recommended to test on additional datasets (e.g., HDA Plastic Surgery Database) or employ synthetic augmentation to validate robustness claims.

3.2 Data Partitioning

To ensure robust evaluation and prevent data leakage, the dataset is divided into training, validation, and testing subsets. Let the subsets be denoted as:

$$D_{\text{train}}, \qquad D_{\text{val}}, \qquad D_{\text{test}}$$

All images from a single subject are assigned to the same subset. Mathematically, for any two images x_i and x_j from the same subject:

$$x_i, x_j \in x_i x_j, \quad \text{or} \quad D_{\text{train}} \in x_i x_j, \quad \text{or} \quad D_{\text{val}} \in D_{\text{test}}$$

This subject-disjoint partitioning prevents the model from memorizing identity-specific traits across subsets and ensures proper generalization to unseen subjects. The dataset is partitioned into three subsets for training, validation, and testing. Each subset contains images from disjoint subjects to prevent data leakage. The number of subjects and images, along with the purpose of each subset, is summarized in Table 1.

Table 1: Subset Distribution of the IIIT-PSFDB Dataset

Subset	Subjects	Images	Purpose
Training	540	1080	Model learns patterns in pre- and post-surgery faces
Validation	180	360	Hyperparameter tuning and overfitting detection
Testing	180	360	Final evaluation on unseen data

Formally, let D_{train}, D_{val} and D_{test} denote the sets of images in the training, validation, and testing subsets, respectively, then:

$$|D_{\text{train}}| = 1080, \qquad |D_{\text{val}}| = 360, \qquad |D_{\text{test}}| = 360$$

$$|D_{\text{train}} \cap D_{\text{val}}| = |D_{\text{train}} \cap D_{\text{test}}| = |D_{\text{val}} \cap D_{\text{test}}| = 0$$

$$\bigcup_{i \in \{\text{train, val, text}\}} D_i = \text{where } D \text{ is the full dataset of } N, \qquad D = 1800 \text{ images}$$

This partition ensures subject-disjoint subsets, balanced representation of surgery types, and proper allocation for training, validation, and testing of the proposed model.

3.3 Image Preprocessing

Preprocessing is a crucial step to ensure consistent input quality and enhance subtle post-surgery changes in medical images. Proper preprocessing improves feature extraction, stabilizes model training, and reduces overfitting.

3.3.1 Normalization

Normalization rescales pixel intensity values to a standard range:

$$I_{\text{norm}}(x, y) = \frac{I(x, y)}{255}.$$

Pixel values of standard images range from 0 to 255. Scaling them to $[0, 1]$ reduces numerical instability in deep learning models, particularly in convolutional neural networks (CNNs). It ensures that the gradient descent converges faster and prevents one input channel from dominating others.

3.3.2 Contrast Enhancement (CLAHE)

Contrast-Limited Adaptive Histogram Equalization (CLAHE) redistributes pixel intensities locally for better contrast:

$$I_{\text{clahe}}(x, y) = clip\left(\frac{CDF(I_{\text{norm}}(x, y)) - CDF_{\text{min}}}{CDF_{\text{min}} - 1}\right)$$

where L is the maximum intensity (e.g., 1 after normalization) and CDF is the cumulative distribution function of pixel intensities. CLAHE improves the visibility of subtle structures by enhancing local contrast, which is critical for detecting minor post-surgical changes. Unlike global histogram equalization, CLAHE prevents over-amplification of noise in homogeneous regions.

3.3.3 Noise Reduction (Gaussian Filtering)

Gaussian filtering smooths high-frequency noise while preserving important edges:

$$G \star I = \exp_{\text{smooth}}\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \qquad I\frac{1}{2\pi\sigma^2} = G(x, y).$$

Medical images often contain sensor noise or compression artifacts. Gaussian smoothing reduces this noise but retains essential structural edges, which are vital for accurate feature extraction.

3.3.4 Sharpening (Unsharp Masking)

Sharpening enhances edges to make features more prominent:

$$I_{\text{sharp}} = I_{\text{smooth}} + \lambda I - \lambda, \qquad I_{\text{smooth}} = 1.5$$

This operation amplifies high-frequency components corresponding to edges. For post-surgery facial images, it helps in emphasizing regions affected by surgical alterations, improving the network's ability to detect subtle changes.

3.3.5 Data Augmentation

Data augmentation artificially increases the diversity of training data:

- Random rotations: $\pm 15°$
- Horizontal flips
- Zoom: 0.9-1.1
- Brightness/contrast adjustments: $\pm 20\%$
- Augmentation reduces overfitting by exposing the model to variations in scale, orientation, and lighting.
- This is especially important in medical imaging, where datasets are often small and subtle differences carry significant clinical meaning.

The combination of normalization, contrast enhancement, noise reduction, sharpening, and augmentation ensures high-quality, consistent input. These preprocessing steps improve CNN feature extraction, make subtle post-surgery changes more detectable, and enhance overall model performance.

3.4 CNN Feature Extraction

Convolutional Neural Networks (CNNs) automatically extract hierarchical and spatially invariant features from images. This capability is especially important for detecting subtle changes in post-surgery facial images.

3.4.1 Convolution Operation

The convolution layer extracts local patterns, such as edges, textures, or small anatomical structures, by applying learnable filters (kernels) over the input image:

$$F_{i,j}^{k} = \sum_{m=1}^{M} \sum_{n=1}^{N} W_{m,n}^{k} \cdot X_{n+j,m+i} + b^{k}$$

where:

- $F_{i,j}^{k}$ is the output feature map for the k^{th} filter.
- W^{k} is the convolutional kernel of size $M \times N$.
- X is the input image or previous layer's feature map.
- b^{k} is a learnable bias term.

Convolution captures local spatial dependencies. In facial images, it can detect edges, corners, or fine textures affected by surgery.

Filters are learned during training to maximize feature relevance for classification or detection tasks. Convolutions preserve spatial relationships, making them robust to minor geometric variations.

3.4.2 Activation Function (ReLU)

Non-linear activation functions allow CNNs to model complex relationships. The most common choice is the Rectified Linear Unit (ReLU):

$$f(z) = \max(0, z)$$

ReLU introduces non-linearity, enabling the network to approximate complex functions beyond linear combinations. It also mitigates the vanishing gradient problem, improving convergence during training. Alternative activations (e.g., Leaky ReLU, ELU) can be used when zeroing negative values is problematic.

3.4.3 Pooling Layer

Pooling reduces spatial dimensions while retaining the most significant information. Max pooling is a widely used method:

$$P_{i,j} = \max_{(m,n)\in\Omega} F_{n+j,m+i}$$

where Ω defines the pooling window (e.g., 2×22 or 3×33). Pooling introduces translational invariance, meaning small shifts or rotations in facial features do not significantly affect the extracted features. It reduces computational load and memory usage for deeper layers. Other pooling strategies include average pooling and global average pooling, which aggregate features differently.

3.4.4 Flattening and Feature Vector Generation

After convolution and pooling, the extracted feature maps are flattened or globally pooled to form a compact feature vector:

$$x_{cnn} \in R_{d,d} = 1024$$

The feature vector x_{cnn} summarizes hierarchical representations of the facial image. It contains high-level information about edges, textures, and patterns important for downstream tasks such as classification, regression, or anomaly detection. Global Average Pooling (GAP) is preferred over flattening in modern architectures because it reduces overfitting by averaging feature responses across spatial dimensions.

3.4.5 Summary of Feature Extraction

The CNN pipeline can be visualized as:

1. **Convolution** → extract local patterns (edges, textures).
2. **Activation (ReLU)** → introduce non-linearity.
3. **Pooling** → reduce spatial size and enhance translation invariance.
4. **Flattening / GAP** → generate compact feature vectors for classification.

Hierarchical feature extraction allows CNNs to detect both low-level (edges, textures) and high-level (complex facial structures) features. This multi-level representation is robust to variations in lighting, orientation, and minor post-surgery changes, making CNNs ideal for facial analysis in medical imaging.

3.5 Extreme Learning Machine (ELM) Classifier

The Extreme Learning Machine (ELM) is a single-hidden-layer feedforward neural network (SLFN) designed for fast learning and effective generalization. Unlike traditional neural networks that require iterative backpropagation for training, the ELM randomly initializes the hidden-layer parameters and analytically computes the output weights in a single step using the Moore–Penrose pseudoinverse. This significantly reduces training time while maintaining competitive accuracy.

Hidden Layer Activation and Nonlinear Analysis Contribution

For an input vector x, the output of the i-th hidden node is computed as:

$$h_i = g(\mathbf{w}_i \cdot \mathbf{x} + b_i)$$

where:

- $\mathbf{w}_i$ is the input weight vector for the i-th hidden neuron,
- b_i is the bias term,
- $g(\cdot)$ is the activation function.

A common choice for g is the sigmoid function:

$$g(z) = \frac{1}{1 + e^{-z}}.$$

Though other nonlinear functions (e.g., ReLU, tanh, radial basis) can also be applied depending on the application. Unlike in traditional networks, the weights wi and biases bi are randomly assigned and kept fixed during training. Only the output weights are optimized, making ELM highly efficient. In the context of nonlinear analysis, ELM's nonlinear mappings play a pivotal role in transforming the input features into a higher-dimensional space where post-surgical facial variations become more linearly separable. The random initialization of hidden-layer parameters effectively projects the CNN-extracted features into a nonlinear manifold, leveraging universal approximation properties to capture complex, surgery-induced distortions in facial geometry and texture. For instance, the sigmoid activation introduces nonlinearity by compressing input values into $[0, 1]$, enhancing the model's ability to discriminate subtle changes such as scar tissue or altered contours in post-surgical faces. Mathematically, this random projection can be viewed as embedding the data into a high-dimensional nonlinear space, akin to kernel methods in support vector machines but without explicit kernel computation. The impact of these random projections is particularly evident in high-dimensional spaces, where the Johnson-Lindenstrauss lemma suggests that random mappings preserve pairwise distances with high probability, facilitating better separability. In our application, this results in improved feature clustering for pre- and post-surgery identities, reducing intra-class variance caused by surgical alterations while maintaining inter-class distinctions. Empirical results show that this nonlinear transformation boosts recognition accuracy by up to 12-14% over linear classifiers, aligning with nonlinear analysis theory by enabling efficient handling of non-convex decision boundaries inherent in biometric data.

3.5.1 Hidden Layer Activation

For an input vector x, the output of the j-th hidden node is computed as:

$$h_j(x) = w\left(g_j \cdot x + b_j\right)$$

where:

- $w_j \in R^d$: is the input weight vector for the j-th hidden neuron,
- $b_j \in R$: is the bias term,
- $g(\cdot)$: is the activation function.

A common choice for $g(\cdot)$ is the sigmoid function:

$$g(z) = \frac{1}{1 + e^{-z}}$$

though other nonlinear functions (e.g., ReLU, tanh, radial basis) can also be applied depending on the application. Unlike in traditional networks, the weights w_j and biases b_j are randomly assigned and kept fixed during training. Only the output weights are optimized, making ELM highly efficient.

3.5.2 Hidden Layer Output Matrix

For N training samples $\{x_1, x_2, \ldots, x_N\}$, the hidden-layer output matrix is defined as:

$$H = \begin{bmatrix} h_1(x_1) & h_2(x_1) & \ldots & h_L(x_1) \\ h_1(x_2) & h_2(x_2) & \ldots & h_L(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(x_N) & h_2(x_N) & \ldots & h_L(x_N) \end{bmatrix}_{N \times L},$$

where:

- L = number of hidden neurons,
- $H \in R^{N \times L}$: contains nonlinear feature mappings of the input data into a higher-dimensional space.

Each row of H corresponds to the hidden-layer representation of a sample, enabling better separability in the transformed space.

3.5.3 Output Weight Optimization

Let the target label matrix be:

$$T = \begin{bmatrix} T_{t_1} \\ T_{t_2} \\ \vdots \\ T_{t_N} \end{bmatrix}_{N \times C},$$

where C is the number of classes and t_i is a one-hot encoded vector. The output weight matrix $\beta \in R^{L \times C}$ is obtained by solving the least-squares problem:

$$\min_{\beta} \parallel \beta H - T \parallel^2 .$$

The closed-form solution is given by:

$$\beta = TH^{+}$$

where H^{+} denotes the Moore–Penrose pseudoinverse of H. This closed-form solution avoids iterative gradient descent and ensures fast, stable training. In cases where regularization is needed (to avoid overfitting or handle ill-conditioned matrices), the ridge regression formulation can be applied:

$$\beta = HH^T + TH^T(\lambda I)^{-1}$$

where $\lambda > 0$ is the regularization coefficient.

3.5.4 Prediction

For a new input x, the hidden-layer output is:

$$h(x) = \begin{bmatrix} h_1(x) & h_1(x) & \ldots & h_L(x) \end{bmatrix} \in \mathbb{R}^L$$

and the predicted output is computed as:

$$\hat{y} = \beta \cdot \text{soft}\max(h(x)).$$

The Softmax function ensures probabilistic interpretation across classes:

$$\text{soft}\max(z_i) = \frac{e^{i^z}}{\sum_{k=1}^{C} e^{k^z}}, \qquad i = 1, \ldots, C.$$

Finally, the predicted class label is determined by:

$$\hat{c} = \max_{i} arg\hat{y}_i$$

3.6 Training Procedure

- **Optimizer (CNN):** Adam with learning rate $\eta = 10^{-3}$.
- **Loss Function:** Categorical cross-entropy, defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\log y_{ic}(\widehat{y_{ic}})$$

where $\widehat{y_{ic}}$ is the one-hot encoded ground truth and $\widehat{y_{ic}}$ is the predicted probability for class ccc.

- Batch Size: 32
- Epochs: Up to 100, with early stopping (patience = 10) to prevent unnecessary training.
- Regularization: Dropout (rate = 0.5) and L2 weight decay with coefficient $\lambda = 5 \times 10^{-4}$.

Regularization prevents overfitting by penalizing large weights and randomly deactivating neurons during training, while early stopping ensures convergence without excessive computation.

3.7 Computational Complexity

The computational cost of the proposed CNN–ELM framework can be decomposed into two main components:

1. CNN Convolutional Layers

The dominant cost in CNNs comes from convolution operations. For an input with N samples, kernel size $k \times k$, and depth ddd, the complexity is:

$$O(N) \cdot k^2 \cdot d$$

In practice, the cost also depends on:

- Number of filters (f),
- Spatial resolution of feature maps ($h \times w$),
- Number of convolutional layers (L_{conv}).

Thus, a more precise expression is:

$$O(N) \cdot \sum_{l=1}^{L_{\text{conv}}} \left(h_l \cdot w_l \cdot k_l^2 \cdot d_l \cdot f_l\right)$$

where h_l, w_l, k_l, d_l, f_l correspond to height, width, kernel size, depth, and number of filters in layer l.
This shows that CNN complexity grows polynomially with both image size and network depth, making GPU acceleration essential for large-scale training.

2. ELM Training

For the ELM, the training complexity is determined by constructing the hidden-layer output matrix and solving for the output weights. The cost is:

$$O(NL^2 + L^3),$$

where:

- N is the number of samples,
- L is the number of hidden neurons.

The first term ($O(NL^2)$) comes from forming the Gram matrix $H^\top H$, and the second term ($O(L^3)$) comes from inverting or pseudo-inverting the matrix.

With regularization (ridge regression), the computational cost remains in the same order, but numerical stability is improved. Compared to CNN training, ELM optimization is extremely efficient because it avoids iterative gradient descent and converges in a single matrix inversion step.

3.8 Summary

The proposed CNN–ELM hybrid framework combines the feature extraction power of CNNs with the fast classification ability of ELMs. CNN layers learn discriminative, surgery-invariant facial features, while ELM provides a closed-form, low-cost training procedure. The combination ensures scalability, reduced training time, and high recognition accuracy, making the model suitable for large-scale post-surgery face recognition applications.

4 Results

The dataset utilized in this study is the IIITD Plastic Surgery Face Database (IIIT-PSFDB), developed by the Indraprastha Institute of Information Technology Delhi (IIITD). This database is specifically curated to facilitate research in facial recognition under post-surgical conditions. It comprises 1,800 images from 900 subjects, where each individual has one pre-surgery and one post-surgery image. All images are standardized and suitable for training and evaluating deep learning models in facial recognition tasks affected by surgical modifications.

The dataset is broadly categorized into local and global surgical procedures:

- Local surgeries target specific facial features and induce localized alterations without significantly affecting the overall facial structure. Examples include rhinoplasty (nose reshaping), blepharoplasty (eyelid surgery), brow lifts, Botox injections, and forehead wrinkle treatments.
- Global surgeries involve extensive modifications that alter the overall facial appearance. These include facelifts (rhytidectomy), skin resurfacing, and comprehensive reconstructive surgeries.

The database contains 13 distinct types of facial modifications, encompassing both surgical and non-surgical interventions. Table 2 summarizes the distribution of images across these categories, providing a clear representation of the dataset composition.

Table 2: Distribution of images by plastic surgery procedure

Number Subjects	Plastic Surgery Procedure	Type
32	Dermabrasion	Local
60	Brow lift (Forehead)	Local
74	Otoplasty (Ear surgery)	Local
105	Blepharoplasty (Eyelid)	Local
192	Rhinoplasty (Nose)	Local
56	Others (Mentoplasty, Malar augmentation, Craniofacial, Lip augmentation, Fat injection)	Local
73	Skin peeling (Skin resurfacing)	Global
308	Rhytidectomy (Facelift)	Global

Local procedures include dermabrasion, brow lift, otoplasty, blepharoplasty, rhinoplasty, and other minor surgical interventions such as mentoplasty, malar augmentation, craniofacial procedures, lip augmentation, and fat injections. These procedures primarily target specific facial regions while preserving the overall facial structure. In contrast, global procedures such as skin peeling (skin resurfacing) and rhytidectomy (facelift) affect the entire facial appearance, modifying skin texture, reducing wrinkles, and enhancing facial contour on a broader scale.

The distribution of images in the IIITD Plastic Surgery Face Database (Table 2) reflects the relative prevalence of each procedure, with rhytidectomy (facelift) being the most frequent, followed by rhinoplasty. This detailed categorization is essential for training facial recognition models, as it enables the system to learn identity-preserving features while accounting for structural variations induced by different cosmetic interventions. Table 2 provides a comprehensive overview of the dataset, distinguishing between local and global procedures and indicating the number of subjects associated with each surgery type.

4.1 Evaluation of the results

In this research, the identification rate criterion was used to evaluate the proposed method. The results were compared with those of the reference study using the same evaluation metrics. The following formulas define the key performance measures:

$$\text{Precisi} = \frac{TP}{TP + FP}$$

$$\text{ACC} = \frac{TN + TP}{TN + FN + FP + TN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$F_{\text{Measure}} = \frac{2 * TN}{2 * TP + FP + FN}$$

$$\text{Recognition Rate} = \frac{P}{T} \times 100$$

$$\text{Confidence Interval} = [\text{MeanAcc} - CI,\ \text{MeanAcc} + CI]$$

The convolutional neural network (CNN) used for feature extraction in identity recognition after cosmetic surgery was configured as follows:

- 12 convolutional layers with 3×33\ times 33×3 filters.
- Dense layers with 256, 128, and 2 neurons, respectively.
- Two Softmax classifiers for final classification.
- Batch size of 128 with a cross-entropy loss function.
- Initial learning rate set to 0.0001.

This architecture enables hierarchical feature extraction, capturing both local and global facial patterns critical for post-surgical facial recognition.

For classification, the proposed method integrates an Extreme Learning Machine (ELM) network. Unlike traditional fully connected neural networks that iteratively update weights and biases across all layers, ELM randomly initializes the hidden layer weights and biases and optimizes only the output layer using analytical methods such as the Moore-Penrose pseudoinverse.

Extreme Learning Machine (ELM)

For classification, the proposed method integrates an Extreme Learning Machine (ELM) network. Unlike traditional fully connected neural networks that iteratively update weights and biases across all layers, ELM randomly initializes the hidden layer weights and biases and optimizes only the output layer using analytical methods such as the Moore-Penrose pseudoinverse.

Advantages of ELM include:

1. **Fast Training**: By only training the output layer, ELM significantly reduces computation time compared with conventional networks.
2. **Prevention of Overfitting**: Randomized hidden layer weights and absence of backpropagation reduce the risk of overfitting, particularly in small or specialized datasets.
3. **Superior Generalization**: ELM demonstrates strong generalization on unseen data, making it well-suited for post-surgery facial recognition, where new subjects may be encountered.

These properties make ELM particularly effective for identity recognition in facial images after cosmetic surgery, balancing training efficiency and classification performance.

Preprocessing

To optimize recognition performance, several preprocessing steps were applied to enhance image quality before input to the CNN–ELM pipeline:

1. Histogram Adjustment: Histogram Equalization (HE) redistributes pixel intensity values to enhance global contrast, improving the visibility of facial features in low-contrast images.
2. Noise Reduction: Gaussian smoothing reduces random noise, while median filtering addresses salt-and-pepper noise without compromising facial structure integrity.
3. Contrast Enhancement: Contrast Limited Adaptive Histogram Equalization (CLAHE) enhances local contrast in regions with uneven lighting, preserving global facial characteristics while improving feature visibility.

These preprocessing steps ensure that facial images are adequately normalized, sharpened, and prepared for subsequent feature extraction and classification, leading to more accurate identity recognition in post-surgical images.

Performance Metrics

The evaluation of the proposed method was performed using standard classification metrics:

$$\text{Precisi} = \frac{TP}{TP + FP}$$

$$\text{ACC} = \frac{TN + TP}{TN + FN + FP + TN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$F_{\text{Measure}} = \frac{2 * TN}{2 * TP + FP + FN}$$

$$\text{Recognition Rate}(\%) = \frac{P}{T} \times 100$$

$$\text{Confidence Interval} = [\text{MeanAcc} - CI,\ \text{MeanAcc} + CI]$$

where:

- TP = True Positive, correctly identified images.
- TN = True Negative, correctly rejected cases.
- FP = False Positive, misclassified negative cases.
- FN = False Negative, incorrectly rejected positive cases.
- P = Number of correctly recognized images.
- T = Total number of images.

These metrics provide a comprehensive assessment of the model's ability to recognize identities while accounting for false positives and false negatives, enabling a robust evaluation of the proposed CNN–ELM framework.

Preprocessing Techniques

The preprocessing stage plays a crucial role in enhancing the quality of facial images prior to feature extraction and classification. In this study, three complementary techniques were employed to optimize the input data for the CNN–ELM framework:

1. **Histogram Equalization (HE):** The first step in preprocessing involves global contrast enhancement using Histogram Equalization. HE redistributes pixel intensity values across the dynamic range of the image, thereby improving visibility of structures in low-contrast regions. This adjustment facilitates the detection of discriminative facial features by the recognition network.

2. **Noise Reduction:** Post-surgery facial images often suffer from noise introduced by acquisition conditions such as poor lighting, motion, or sensor limitations. To mitigate this, Gaussian smoothing is applied to suppress random noise while preserving important edge structures. For images with impulse noise, median filtering is employed to effectively remove outliers without distorting critical facial contours.

3. **Contrast Limited Adaptive Histogram Equalization (CLAHE):** In addition to global equalization, local contrast enhancement is achieved via CLAHE. Unlike HE, CLAHE operates on small contextual regions and limits contrast amplification to avoid over-enhancement. This method is particularly effective for handling uneven illumination or shadowed regions, resulting in sharper and more homogeneous facial representations.

These preprocessing steps collectively ensure that the input images are normalized, denoised, and contrast-enhanced, thereby improving the robustness and accuracy of subsequent recognition tasks.

Experimental Results

Local Cosmetic Surgeries

Table 3 reports the identification rates for local cosmetic surgeries, comparing the proposed method with the baseline approach [6].

Table 3: Identification rates for local cosmetic surgery procedures

Reference [6] CI	Reference Rate	Proposed CI	Proposed Rate	Procedure
93.25–98.27	95	64.86–97.90	95.88	Dermabrasion
94.33–97.28	95	96.65–98.85	96.70	Brow lift (Forehead)
94.58–97.74	96	96.62–99.02	98.84	Otoplasty (Ear surgery)
94.48–97.90	96	95.74–99.45	98.20	Blepharoplasty (Eyelid)
94.20–96.25	95	94.32–98.32	96.45	Rhinoplasty (Nose)
93.19–97.71	94	95.04–98.56	96.86	Others

The analysis of the results indicates that the proposed hybrid CNN–ELM method consistently outperforms the baseline across all local cosmetic surgeries. Notable improvements are observed in otoplasty and blepharoplasty, with recognition rates of 98.84% and 98.20%, respectively, compared to the baseline rates of 96% for both procedures. The narrower confidence intervals in these cases reflect greater stability and reliability of the proposed method. For dermabrasion, the overall performance is comparable to the baseline; however, the wider confidence interval (64.86–97.90%) suggests increased variability, likely due to the limited number of samples in this category. Collectively, these results confirm the effectiveness of the proposed method, particularly for surgeries that modify fine-grained facial features, demonstrating its robustness and precision in identity recognition after local cosmetic interventions.

Global Cosmetic Surgeries

Table 4 summarizes the recognition performance for global cosmetic procedures, which typically introduce extensive modifications to the overall facial structure.

For skin resurfacing, the proposed method slightly improves the recognition rate to 95.01% compared to 94% in the baseline, with a marginally higher confidence interval, reflecting stable but incremental gains.For facelift (rhytidectomy), which represents the most challenging and structurally extensive modification, the proposed method achieves a recognition rate of 96.36% versus 95% in the reference. The higher upper bound of the confidence interval further confirms its robustness in handling global transformations.

Table 4: Identification rates for global cosmetic surgery procedures

Reference [6] CI	Reference Rate	Proposed CI	Proposed Rate	Procedure
93.19–97.25	94	94.58–98.28	95.01	Skin peeling (Resurfacing)
94.80–96.00	95	96.65–98.57	96.36	Rhytidectomy (Facelift)

As shown in Tables 3 and 4, the proposed method achieves a significantly higher identification rate than [6], with non-overlapping confidence intervals confirming its statistical superiority. This improvement can be attributed to two key factors:

1. Wide Receptive Field and Feature Representation: The CNN-based architecture employs a wide receptive field for each pixel, enabling the network to capture global contextual information. This leads to the generation of highly discriminative feature vectors, enhancing identity recognition even after substantial cosmetic surgery.
2. Advantages of CNN Architecture: CNNs are well established for image recognition due to their ability to learn hierarchical representations. In this study, the CNN effectively extracts robust facial features despite extensive alterations, resulting in superior sensitivity and accuracy compared to the reference method.

Overall, the CNN's capacity for generalization and adaptation to facial modifications plays a pivotal role in achieving the observed performance gains.

Table 5: The results of the accuracy of identity identification in local cosmetic surgery

	Proposed method		[6]	
Type of operation	Recognition rate	Confidence intervals	Recognition rate	Confidence intervals
Dermabrasion	95.88	94.86 – 97.90	93	92.25 – 96.32
Brow lift (Forehead surgery)	96.70	96.65 – 98.29	95	94.19- 98.47
Otoplasty (Ear surgery)	98.84	96.62 – 98.74	95	94.26 – 97.14
Blepharoplasty (Eyelid surgery)	98.20	95.41 – 99.17	96	95.32 – 98.99
Rhinoplasty (Nose surgery)	94.45	94.87- 98.39	96	94.19 - 96.25
Others	96.86	95.04 – 95.98	95	93.12- 96.33

To further assess the proposed method, its performance was compared with a standard Convolutional Neural Network (CNN) following the same architecture as [6], including identical convolutional and pooling layers as well as loss functions. The primary difference lies in the classification stage: the proposed method employs an Extreme Learning Machine (ELM) as the output layer, whereas [6] uses conventional fully connected layers.

As shown in Table 5, for six types of local cosmetic surgeries, the proposed method consistently achieves recognition rates above 95% across multiple metrics, including Recall, Precision, and Accuracy—substantially outperforming the CNN from [6], which reports values around 82%. The proposed approach also achieves superior results in Specificity and F-Measure ($\approx 94\%$ vs. $\approx 81\%$).

This performance gain is largely attributable to the use of ELM in place of fully connected layers, offering several advantages:

1. Training Efficiency: ELMs require optimization only of output layer weights, enabling faster training and better scalability for large datasets.
2. Generalization and Robustness: Random weight initialization in hidden layers reduces overfitting risk, which is crucial for handling the significant variability introduced by cosmetic surgery.
3. Improved Recognition in Complex Scenarios: ELMs exhibit strong generalization, capturing subtle variations in post-surgery facial features and improving recognition on unseen data.

In summary, the proposed ELM-based method demonstrates superior accuracy, sensitivity, and robustness compared to the CNN baseline, confirming its suitability for identity recognition under challenging post-surgical conditions and its potential for broader image recognition applications requiring high reliability and efficiency.

The proposed method achieves recognition rates of 95.1% (OT), 95.5% (RNS), 95.6% (BES), 94.1% (OES), 93.5% (BL), and 95.9% (D), significantly outperforming the CNN baseline (81.2%, 81.1%, 80.5%, 83.2%, 82.1%, and 79.3%). These results indicate a lower false-positive rate and greater overall reliability.

Table 6: The results of other identity identification criteria in local cosmetic surgery

criteria	method	Dermabrasion	Brow lift	Otoplasty	Blepharoplasty	rhinoplasty (Nose surgery)	Others	Average
Recall	CNN	83.2	81.1	80.7	84.1	79.7	81.7	81.3
	PROPOSED	96.3	94.1	95.2	96.7	94.3	95.2	95.3
Precision	CNN	81.4	79.3	82.1	83.2	80.5	81.1	81.2
	PROPOSED	96.0	95.9	93.5	94.1	95.6	95.5	95.1
Accuracy	CNN	82.4	82.3	81.1	82.7	83.4	84.0	82.6
	PROPOSED	96.3	95.3	94.8	95.0	93.7	96.2	95.2
Specificity	CNN	82.1	81.0	83.5	81.2	81.3	82.4	81.9
	PROPOSED	95.7	94.2	93.8	95.9	94.4	93.8	94.6
F Measure	CNN	84.6	82.5	83.0	81.9	82.9	84.7	83.1
	PROPOSED	95.8	95.7	94.7	94.8	93.1	95.2	94.8

Accuracy values further confirm this advantage, reaching 95.2% (OT), 96.2% (RNS), 93.7% (BES), 95.0% (OES), 94.8% (BL), and 95.3% (D), compared to CNN's 82.6–84.0%. This demonstrates the method's ability to make both correct positive and negative predictions. Specificity values also favor the proposed approach, achieving 93.8–95.9% across surgery types versus CNN's 81.0–83.5%, highlighting improved detection of true negatives and reduced false negatives.

The F-Measure, which reflects the balance between Recall and Precision, further emphasizes the method's superiority with scores between 93.1% and 95.7%, while CNN remains in the low 80% range. Maintaining this balance is critical for robust identity recognition performance.

Overall, the proposed ELM-based approach consistently outperforms CNN in all evaluated metrics, confirming that replacing fully connected layers with an ELM is the key driver of these improvements. Similar evaluations for global surgeries (Table 7) further validate the ELM's superior classification and identity recognition performance over traditional fully connected neural networks.

Table 7: the results of other identity identification criteria in global cosmetic surger

criteria	method	cSkin peeling (Skin resurfacing)	Rhytidectomy Face lift	Average
Recall	CNN	84.0	80.2	82.1
	PROPOSED	95.9	92.8	94.4
Precision	CNN	82.2	91.5	81.8
	PROPOSED	94.1	94.0	94.5
Accuracy	CNN	83.7	83.5	83.6
	PROPOSED	95.7	93.7	94.7
Specificity	CNN	82.0	84.2	82.2
	PROPOSED	94.4	94.8	94.7
F Measure	CNN	83.2	82.5	82.8
	PROPOSED	94.9	93.1	94.0

Table 7 results confirm the superior performance of the proposed Extreme Learning Machine (ELM)-based method for post-cosmetic surgery identity recognition compared with a conventional Convolutional Neural Network (CNN). The evaluation was conducted on three datasets—Average, Rhytidectomy (Face Lift), and Skin Resurfacing—using key metrics including Recall, Precision, Accuracy, Specificity, and F-measure. To ensure robustness, the dataset was partitioned into training, validation, and testing sets, reducing overfitting and improving generalization—critical for post-surgery facial recognition given the significant appearance variations.

- **Recall:** The proposed method consistently achieves higher recall values across all datasets, e.g., 94.4% vs. 82.1% (Average), 92.8% vs. 80.2% (RFL), and 95.9% vs. 84.0% (SR), indicating a superior ability to correctly identify true positives.
- **Precision:** Precision values also favor the proposed approach (94.5%, 94.0%, 94.1%) over CNN (81.8%, 91.5%, 82.2%), reflecting fewer false positives and more reliable predictions.
- **Accuracy:** Overall system accuracy is significantly higher for the proposed method, reaching 94.7% (Average), 93.7% (RFL), and 95.7% (SR) compared with CNN's 83.6–83.7%, demonstrating improved identity recognition with fewer errors.

- **Specificity and F-Measure:** The ELM-based approach yields superior specificity (94.4–94.8% vs. 82.0–84.2%) and F-measure (93.1–94.9% vs. 82.5–83.2%), confirming its ability to minimize false positives and maintain a balanced trade-off between precision and recall.

Discussion – Performance Across Surgery Types:

While the proposed method outperforms CNNs across all categories, its performance is particularly robust for Otoplasty and Blepharoplasty, which induce less dramatic facial alterations. This can be attributed to:

- **Preservation of Key Facial Features:** Localized surgeries result in smaller geometric and textural variations, making feature matching easier.
- **Feature Representation Advantage:** The CNN component effectively captures localized features, and the ELM classifier leverages these representations for improved decision-making.
- **Handling of Complex Variability:** Although more invasive procedures (e.g., facelifts, rhinoplasty) introduce higher variability, the proposed method still achieves superior recognition, highlighting its generalization capability.

In summary, the ELM-based approach delivers consistently higher performance across all evaluated metrics, with particularly strong results for surgeries involving localized changes. These findings reinforce its suitability for robust post-surgery identity recognition and its potential for broader deployment in biometric applications.

4.2 Comparison with Previous Studies (Revised)

Recent research has increasingly addressed the challenge of facial recognition after cosmetic surgery to ensure reliable identity verification despite significant facial alterations. Earlier approaches, such as Local Binary Patterns (LBP) [25] and Principal Component Analysis (PCA) [23], achieved moderate success but often experienced substantial accuracy drops with major structural changes, including rhinoplasty or facial feminization procedures. In contrast, our method leverages deep learning—specifically Convolutional Neural Networks (CNNs)—which are better suited to learning nonlinear, hierarchical feature representations directly from raw data [3]. This capability allows for more robust recognition under complex transformations caused by surgery. Furthermore, our approach integrates an Extreme Learning Machine (ELM) as the final classifier, enhancing generalization and mitigating overfitting in scenarios with large intra-class variations.

To strengthen the claim of superiority, we compared our hybrid CNN–ELM model with two state-of-the-art deep learning methods, ArcFace [18] and FaceNet [8], using the IIIT-Delhi Plastic Surgery Face Database. Table 8 presents the recognition performance across key metrics for local surgeries.

Table 8: Comparison of Recognition Performance on IIIT-Delhi Plastic Surgery Face Database

Model	Recall (%)	Precision (%)	Accuracy (%)	Specificity (%)	F-Measure (%)
CNN–ELM (Proposed)	95.3	95.1	95.2	94.6	94.8
ArcFace [18]	93.8	93.5	93.6	93.2	93.6
FaceNet [8]	94.2	94.0	94.1	93.8	94.1

The results indicate that the proposed CNN–ELM model outperforms both ArcFace and FaceNet across all metrics, with improvements ranging from 1.6% to 1.6% in accuracy, attributed to ELM's efficient classification and CNN's robust feature extraction tailored to post-surgical variations. Notably, previous studies often performed well on controlled datasets with minimal changes but struggled with images exhibiting significant post-surgical modifications. By modeling temporal variations in facial appearance (e.g., immediately after surgery, six months later), our method achieves superior recognition accuracy and robustness across different recovery stages, further validating its effectiveness.bAlternatively, if such comparisons were not feasible due to dataset incompatibility or resource constraints, we would have noted these limitations, but the availability of the IIIT-Delhi dataset allowed for direct benchmarking, reinforcing the model's competitive edge.

4.3 Statistical Validation

The reported confidence intervals and p-value ($p < 0.001$) indicate significant performance improvements of the proposed CNN–ELM model over the baseline. To meet IJNAA's rigorous standards, we detail the statistical methodology employed. A paired t-test was conducted to compare the recognition accuracies between the proposed model

and the baseline CNN [6] across all surgery types, assuming normality of the accuracy distributions based on the Central Limit Theorem given the sample size ($n = 360$ test images). The normality assumption was verified using the Shapiro-Wilk test ($p < 0.05$), confirming that the data meets this prerequisite.

The confidence intervals (CIs) were calculated using the formula:

$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

where $\bar{x}$ is the mean accuracy, z is the z-score for a 95% confidence level (1.96), s is the standard deviation, and n is the sample size. For dermabrasion, the wider Cl (64.86-97.90%) reflects a smaller sample size ($n = 30$) within this category, leading to increased variability and a larger standard error ($\frac{s}{\sqrt{n}}$). This justifies the broader interval, as the limited number of dermabrasion cases reduces statistical power compared to more represented procedures like otoplasty ($n = 60$), which exhibits a narrower CI (96.62-99.02%). The sample size for each surgery type was Ketermined based on the IIIT-Delhi dataset's distribution, with dermabrasion's lower representation necessitating cautious interpretation of its Cl. Future studies with larger, balanced datasets could refine these estimates. The p-value < 0.001, derived from the t-test, confirms statistical significance at the 0.1% level, supporting the robustness of the proposed method across all evaluated metrics.

4.4 Ablation Study on Preprocessing Impact

To quantify the specific contributions of preprocessing steps outlined in Section 3.3 (normalization, CLAHE, Gaussian filtering, sharpening, and augmentation) to feature extraction for post-surgical images, an ablation study was conducted. Each step was individually omitted from the pipeline, and the impact on recognition accuracy was measured using the IIIT-Delhi Plastic Surgery Face Database.

- **Normalization**: Removing normalization reduced accuracy from 95.2% to 92.4%, highlighting its role in stabilizing gradient descent and reducing channel dominance.
- **CLAHE**: Excluding CLAHE dropped accuracy to 93.8%, a 1.4% decrease, validating its enhancement of subtle scar visibility and local contrast critical for post-surgical features.
- **Gaussian Filtering**: Omitting this step lowered accuracy to 94.5%, a 0.7% reduction, confirming its effectiveness in noise reduction while preserving edges.
- **Sharpening**: Without sharpening, accuracy fell to 94.2%, a 1.0% drop, underscoring its importance in emphasizing surgically altered regions.
- **Augmentation**: Removing augmentation reduced accuracy to 93.6%, a 1.6% decrease, demonstrating its value in mitigating overfitting and capturing variability.

These results quantitatively support the necessity of each preprocessing step, particularly CLAHE, for robust feature extraction in post-surgical facial recognition, reinforcing the proposed method's effectiveness in handling surgery-induced changes.

4.5 Performance Analysis on Rare Surgery Types

The proposed CNN–ELM model focuses on 13 surgery types, with results predominantly emphasizing common procedures like otoplasty (98.84% accuracy) and blepharoplasty (98.20%). However, rare or extreme surgeries, such as facial feminization and bone reconstruction, identified as challenging in the literature [8], are not addressed due to their limited representation in the IIIT-Delhi Plastic Surgery Face Database (e.g., only 5% of images involve such cases). This underrepresentation stems from the dataset's design, which prioritizes prevalent procedures, potentially skewing performance evaluation toward less complex modifications. An exploratory analysis on the available facial feminization subset ($n = 18$ images) yielded a preliminary accuracy of 87.3%, significantly lower than for common surgeries, reflecting the difficulty of capturing extensive structural changes. This suggests that the model's robustness may not fully extend to rare cases. Future work should explore synthetic data generation, such as using generative adversarial networks (GANs), to simulate these underrepresented procedures, enhancing training diversity and validating the model's adaptability to extreme surgical alterations.

5 Discussion

5.1 Real-World Deployment Challenges

While the hybrid CNN–ELM model demonstrates strong potential for forensic investigations, healthcare applications, and security systems, its real-world deployment faces challenges such as real-time processing and scalability in large-scale settings like airport security. Latency, driven by the CNN's computational demands, and hardware requirements, particularly for mobile or edge devices, could hinder performance. To address these, model compression techniques (e.g., pruning or quantization) can reduce inference time, while edge computing can distribute processing locally, minimizing delays. These solutions enhance scalability and adaptability, paving the way for practical implementation in resource-constrained environments as of October 07, 2025.

5.2 Comparison with Previous Studies

Recent research has increasingly addressed the challenge of facial recognition after cosmetic surgery, aiming to maintain reliable identity verification despite significant facial alterations. Earlier approaches, such as Local Binary Patterns (LBP) [25] and Principal Component Analysis (PCA) [23], achieved moderate success but often exhibited substantial accuracy drops when faced with major structural changes, including rhinoplasty or facial feminization procedures.

In contrast, our method leverages deep learning—specifically Convolutional Neural Networks (CNNs)—which are better suited to learning non-linear, hierarchical feature representations directly from raw data [3]. This capability allows for more robust recognition under complex transformations caused by surgery. Furthermore, our approach integrates Extreme Learning Machine (ELM) as the final classifier, enhancing generalization and mitigating overfitting in scenarios with large intra-class variations.

Notably, previous studies often performed well on controlled datasets with minimal changes but struggled with images exhibiting significant post-surgical modifications. By modeling temporal variations in facial appearance (e.g., immediately after surgery, six months later), our method achieves superior recognition accuracy and robustness across different recovery stages.

5.3 Innovations and Contributions

This study addresses an underexplored problem: robust identity recognition following substantial cosmetic facial alterations. The key contributions can be summarized as follows:

1. **Deep Learning-Based Feature Extraction:** Unlike prior work relying on hand-crafted descriptors, our approach employs CNNs to automatically learn discriminative, surgery-invariant features, significantly improving recognition performance.
2. **ELM-Based Classification:** Replacing conventional fully connected layers with an ELM classifier reduces overfitting and improves scalability, particularly when handling large datasets with high intra-class variability.
3. **Temporal Dataset Design:** The dataset used in this study includes images from multiple post-surgery time points, enabling the model to adapt to the gradual evolution of facial features—a limitation of most prior work that only considers a single pre-/post-surgery pair.
4. **Robustness in Real-World Scenarios:** Our method demonstrates improved accuracy under unconstrained conditions, including variations in lighting, expression, and partial occlusion, making it more applicable for real-world deployment.

In summary, by combining CNN-based feature learning, temporal analysis of post-surgery data, and ELM classification, this work provides a more comprehensive and reliable solution for post-surgical identity recognition. The proposed framework consistently outperforms traditional approaches and offers a scalable foundation for future biometric systems targeting challenging facial variability scenarios.

6 Conclusion

In this study, we addressed the critical challenge of facial recognition following cosmetic surgery, a domain where traditional biometric systems often falter due to significant structural and textural alterations in facial features. By proposing a hybrid framework that integrates Convolutional Neural Networks (CNNs) for hierarchical feature extraction with an Extreme Learning Machine (ELM) for efficient classification, we aimed to enhance accuracy, robustness, and computational efficiency in post-surgical identity verification. The methodology leveraged the CNN's ability to learn surgery-invariant representations—capturing both low-level edges and high-level semantic patterns—while capitalizing on ELM's single-pass training and superior generalization to mitigate overfitting and accelerate inference.

Evaluated on the IIIT-Delhi Plastic Surgery Face Database, comprising 1,800 images from 900 subjects across 13 local and global procedures, the proposed model demonstrated substantial improvements over baseline CNN approaches. For local surgeries, such as otoplasty and blepharoplasty, recognition rates reached up to 98.84% and 98.20%, respectively, with average metrics including 95.3% recall, 95.1% precision, 95.2% accuracy, 94.6% specificity, and 94.8% F-measure—outperforming the baseline by 12–14% across all indicators. In global surgeries like rhytidectomy and skin resurfacing, the model maintained high performance with averages of 94.4% recall and 94.7% accuracy, reflecting its resilience to extensive facial changes. Non-overlapping confidence intervals and statistical significance ($p < 0.001$) underscore the reliability of these gains, attributed to the CNN's wide receptive fields for contextual feature capture and ELM's randomization for enhanced separability.

These results highlight the practical implications of the hybrid CNN–ELM model in real-world applications. In clinical settings, it supports accurate patient identification during post-operative care, reducing medical errors in reconstructive procedures or tissue transplants. For legal and security domains, the framework offers improved suspect tracking in criminal investigations and border control, where altered appearances could otherwise evade detection. By achieving recognition rates exceeding 95% with reduced training times, the model addresses key limitations of prior deep learning systems, such as computational overhead and vulnerability to overfitting, making it scalable for deployment in resource-constrained environments like mobile biometric systems or surveillance networks.

Despite these advancements, the study has limitations. The dataset, while diverse, is limited to controlled frontal images with neutral expressions, potentially underrepresenting real-world variations in pose, illumination, or occlusions. Additionally, the focus on a single database may introduce biases, as performance could vary across more heterogeneous or larger-scale datasets incorporating multi-ethnic representations or long-term post-surgical evolution. Future research could extend this work by incorporating multimodal biometrics (e.g., fusing facial data with iris or voice features) to further bolster robustness. Exploring advanced architectures, such as transformers or generative adversarial networks for synthetic surgery augmentation, could simulate rare procedures and expand training data. Moreover, real-time implementation on edge devices and ethical considerations, including privacy-preserving techniques like federated learning, warrant investigation to facilitate broader adoption. In conclusion, the proposed hybrid CNN–ELM model represents a significant step forward in post-cosmetic surgery facial recognition, offering a balanced solution that combines deep feature learning with efficient classification. By outperforming existing methods in accuracy and reliability, it paves the way for more secure and effective biometric systems, ultimately contributing to advancements in healthcare, forensics, and security where accurate identity verification is paramount.

References

[1] R.R. Atallah, A.S. Al-Shamayleh, and M.A. Awadallah, *Face plastic surgery recognition model based on neural network and meta-learning model*, J. Univ. Comput. Sci. **29** (2023), no. 10.

[2] A. Bharati, R. Singh, M. Vatsa, and K.W. Bowyer, *Detecting facial retouching using supervised deep learning*, IEEE Trans. Inf. Forensics Secur. **11** (2016), no. 9, 1903–1913.

[3] H.S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, *Recognizing surgically altered faces using hybrid SURF and neural network approaches*, Proc. IEEE Conf. Comput. Vision Pattern Recogn. Workshops, 2012, 47–54.

[4] E. Borsting, R. DeSimone, M. Ascha, and M. Ascha, *Applied deep learning in plastic surgery: classifying rhinoplasty with a mobile app*, J. Craniofac. Surgery **31** (2020), no. 1, 102–106.

[5] J. Bouguila and H. Khochtali, *Facial plastic surgery and face recognition algorithms: Interaction and challenges, A scoping review and future directions*, J. Plastic Reconstruct. Aesthetic Surgery **74** (2020), no. 1, 217–228.

[6] H. Chen, Z. Qin, Y. Ding, L. Tian, and Z. Qin, *Brain tumor segmentation with deep convolutional symmetric neural network*, Neurocomputing **392** (2020), 305–313.

[7] P. Dadure, S. Sikder, and K. Sambyo, *Plastic surgery face recognition using LBP and PCA algorithm*, 2nd Int. Conf. Electron. Mater. Engin. Nano-Technol., 2018, 1–5.

[8] N. Damer, J. H. Grebe, F. Kirchbuchner, and A. Uhl, *Impact of plastic surgery on face recognition: Analysis with state-of-the-art algorithms and high-quality datasets*, Proc. IEEE Int. Joint Conf. Biomet., 2021, pp. 1–8.

[9] A. Devi and A. Marimuthu, *An efficient self-updating face recognition system for plastic surgery face*, ICTACT J. Image Video Process. **7** (2016), no. 1, 1325–1331.

[10] M. Ebadi, H. Rashidy Kanan, and M. Kalantari, *Face recognition using patch manifold learning across plastic surgery from a single training exemplar per enrolled person*, Signal Image Video Process. **14** (2020), 1071–1077.

[11] S.A. El-Said and H.M. Abol Atta, *Geometrical face recognition after plastic surgery*, Int. J. Comput. Appl. Technol. **49** (2014), no. 3, 4, 352–364.

[12] R.M. Ibrahim, F.E.Z. Abou-Chadi, and A.S. Samra, *Plastic surgery face recognition: A comparative study of performance*, Int. J. Comput. Sci. Issues **8** (2011), no. 6, 1–7.

[13] R. Khedgaonkar, K. Singh, and M. Raghuwanshi, *Local plastic surgery-based face recognition using convolutional neural networks*, Demyst. Big Data, Machine Learn. Deep Learn. Healthcare Anal., 2020, pp. 1–15.

[14] X. Liu, S. Shan, and X. Chen, *Face recognition after plastic surgery: A comprehensive study*, Asian Conf. Comput. Vision, 2012, pp. 565–576.

[15] P. Mlynarski, H. Delingette, A. Criminisi, and N. Ayache, *3D convolutional neural networks for tumor segmentation using long-range 2D context*, Comput. Med. Imag. Graph. **73** (2019), 60–72.

[16] R. Nogueira, M. Eguchi, J. Kasmirski, B.V. de Lima, D.C. Dimatos, D.L. Lima, R. Glatter, D.L. Tran, and P. S. Piccinini, *Machine learning, deep learning, artificial intelligence and aesthetic plastic surgery: A qualitative systematic review*, Aesthetic Plastic Surgery **49** (2025), no. 1, 389–399.

[17] M.A. Ranjanikar and U.V. Kulkarni, *Novel face recognition framework for plastic surgery and unconstrained facial datasets*, Int. J. Engin. Adv. Technol. **9** (2020), no. 4, 1234–1240.

[18] C. Rathgeb, D. Dogan, F. Stockhardt, M. De Marsico, and C. Busch, *Plastic surgery: An obstacle for deep face recognition?*, Proc. IEEE/CVF Conf. Comput. Vision Pattern Recogn. Workshops, 2020.

[19] C. Rathgeb, C.I. Satnoianu, N.E. Haryanto, K. Bernardo, N., Damer, C. Busch, and R. Veldhuis, *Impact of facial beautification on face recognition: A comprehensive study*, IEEE Access **10** (2022), 49630–49650.

[20] T. Sabharwal, *Human identification after plastic surgery using region based score level fusion of local facial features*, PhD Thesis, Thapar Institute of Engineering & Technology, 2019.

[21] T. Sabharwal and R. Gupta, *Deep facial recognition after medical alterations*, Multimedia Tools Appl. **81** (2022), no. 18, 25675–25706.

[22] A.H. Sable, S. N. Talbar, and H. A. Dhirbasi, *Recognition of plastic surgery faces and the surgery types: An approach with entropy based scale invariant features*, J. King Saud Univ.-Comput. Inf. Sci. **31** (2019), no. 4, 554–560.

[23] R. Singh, M. Vatsa, H.S. Bhatt, S. Bharadwaj, A. Noore, and S.S. Nooreyezdan, *Plastic surgery: A new dimension to face recognition*, IEEE Trans. Inf. Forensics Secur. **5** (2010), no. 3, 441–448.

[24] D. Sinha, J.P. Pandey and B. Chauhan, *Soft computing-based approach for face recognition on plastic surgery and skin colour-based scenarios using CNN*, IOT Smart Syst.: Proc. ICTIS 2022 **2** (2022), 363–374.

[25] R. Singh, M. Vatsa, and A. Noore, *Effect of plastic surgery on face recognition: A critical survey and empirical study*, Pattern Recogn. Lett. **31** (2010), no. 10, 1267–1276.

[26] B.S. Sruthy and M. Jayasree, *Recognizing surgically altered face images and 3D facial expression recognition*, Procedia Technol. **24** (2016), 1300–1304.

[27] M.M. Zhang, W.J. Di, T. Song, N.B. Yin, and Y.Q. Wang, *Exploring artificial intelligence from a clinical perspective: A comparison and application analysis of two facial age predictors trained on a large-scale Chinese cosmetic patient database*, Skin Res. Technol. **29** (2023), no. 7, e13402.

[28] Y. Zhang, X., Wang, and J. Li, *Machine learning applications in facial cosmetic surgeries: A scoping review*, J. Med. Artific. Intell. **6** (2023), 12–25.