# Optimizing User Recognition with Deep CNN and Cropped Facial Inputs

Jansi Rani S[1], Bhuvanesh P[2], and Dinesh Kumar B[2,] Hari Krishna P A[2]
[1] *Assistant Professor (Sl.Gr) – Sri Ramakrishna Engineering College, Coimbatore, India*
[2] *UG Students - Sri Ramakrishna Engineering College, Coimbatore, India*

***Abstract* - Systems for facial recognition have become widely used in fields including attendance tracking, security, and authentication. However, different lighting conditions, background noise, and facial occlusions make it difficult to achieve great accuracy in real-world circumstances. To improve recognition accuracy and model efficiency, this research proposes a deep learning-based method that makes use of a Convolutional Neural Network (CNN) trained just on cropped face areas. Prior to feature extraction, faces are cropped to improve attention on discriminative facial traits and remove irrelevant background data. After separating and aligning face areas using a face detection pipeline, the system sends the information to a deep CNN that has been trained on a labelled dataset for classification. Results from experiments show that, especially in hardware-constrained contexts, the suggested approach greatly enhances recognition performance when compared to models trained on full-frame images. This study offers insights into creating lightweight yet precise recognition algorithms and emphasizes the significance of region-focused preprocessing in deep facial recognition.**

***Keywords:* CNN, Cropped Face Images, Face Detection, Haar Cascade, Recognition.**

## I. INTRODUCTION

One of the most dependable and non-intrusive biometric methods for user identification and verification is facial recognition technology, which has quickly gained popularity. The need for precise and effective facial recognition systems has never been greater due to the expanding applications in smart access control, mobile security, surveillance, and attendance systems. Conventional facial recognition pipelines frequently use full-frame photos, which contain occlusions, background noise, and other visual information that is not necessary

Convolutional Neural Networks (CNNs), a recent development in deep learning, have demonstrated impressive performance in extracting intricate hierarchical characteristics from image data. However, the caliber and applicability of the input data have a significant impact on CNN performance in facial recognition. The model can learn more robust and discriminative facial features by excluding unnecessary input and concentrating only on selected face regions. Results from experiments show that, especially in hardware-constrained contexts, the suggested approach greatly enhances recognition performance when compared to models trained on full-frame images. This study offers insights into creating lightweight yet precise recognition algorithms and emphasizes the significance of region-focused preprocessing in deep facial recognition.

However, different lighting conditions, background noise, and facial occlusions make it difficult to achieve great accuracy in real-world circumstances. To improve recognition accuracy and model efficiency, this research proposes a deep learning-based method that makes use of a Convolutional Neural Network (CNN) trained just on cropped face areas. Prior to feature extraction, faces are cropped to improve attention on discriminative facial traits and remove irrelevant background data. After separating and aligning face areas using Haar-Cascade, the system sends the information to a deep CNN that has been trained on a labelled dataset for classification. particularly in areas with little hardware.

## II. LITERATURE REVIEW

Ridha Ilyas, Bendjillali, Beladgham [1], presented a cutting-edge facial recognition system that improves detection and classification accuracy by combining conventional and deep learning techniques. The three main components of the suggested pipeline are the deep CNN architectures (ResNet50 and VGG16) for feature extraction and classification, the Viola-Jones method for face detection, and Adaptive Histogram Equalization (AHE) for picture enhancement. The Extended Yale B and CMU PIE databases served as two benchmark datasets on which their approach was thoroughly evaluated. The authors highlighted how AHE improved facial image quality, which improved feature learning and resilience across a range of illumination scenarios. ResNet50 outperformed the other examined designs, with 98.38% accuracy on the CMU PIE dataset and 97.23% accuracy on the Extended Yale B dataset.

Randa Nachet; Tarik Boudghene Stambouli [2], proposed an improved facial recognition system that combines an optimized Convolutional Neural Network (CNN) for recognition and classification with a Multi-Task Convolutional Neural Network (MTCNN) for face detection. The CNN's capacity to extract highly discriminative characteristics from facial photos and then classify them using a softmax function is the fundamental strength of their methodology. The 400 photos from 40 different persons in the ORL face database were used to train and assess the system. Their model outperformed a few traditional and cutting-edge recognition techniques, achieving a recognition accuracy of 97.50% through intensive testing and hyperparameter optimization. This paper emphasizes the usefulness of MTCNN for preprocessing tasks such as face alignment and cropping, which considerably improves the CNN's downstream performance.

Jinhui Zhang [3], created an online face recognition system that overcomes the shortcomings of conventional machine learning techniques in managing changes in angle, lighting, and facial expression. The system offers strong performance in real-world video and picture input settings by integrating the FaceNet model for feature extraction and recognition with a Multi-Task Convolutional Neural Network (MTCNN) for face detection. Users can upload video or image files to the backend system, which uses a Browser/Server (B/S) architecture Higher accuracy, enhanced adaptability, and real-time responsiveness are guaranteed when CNN-based deep learning is integrated with a full-stack web application. This work shows how deep learning models can be successfully integrated into cloud-accessible services for wider deployment, in addition to improving the dependability of face recognition systems in real-world settings.

Karlupia, Namrata & Mahajan [4], used a face recognition framework that uses a Genetic Algorithm (GA) for hyperparameter tuning to improve the performance of Convolutional Neural Networks (CNNs). The authors used GA to automatically find the best settings for factors like filter sizes, number of filters, and number of hidden layers because they understood the drawbacks of human tuning. By iteratively improving the CNN architecture and reducing the objective function, this evolutionary method methodically investigates the parameter space. The study emphasizes the value of adaptive optimization in increasing model precision and decreasing human intervention in deep learning-based facial recognition systems by including GA for network construction.

## III. DATASET

This study used public datasets with real-time user data to efficiently train, test, and evaluate the facial recognition models. For early research and model comparison, the Kaggle Face Dataset—which includes photos of 20 distinct people with varying backgrounds, poses, and expressions—was utilized, offering a consistent standard by which to measure accuracy and consistency. Furthermore, the system's embedded camera interface was used to gather real-time user data, taking samples in real-world settings like low light levels, live authentication circumstances, and changes in look (e.g., with or without spectacles). To ensure dependable performance in terms of speed, accuracy, and robustness in real-world applications, the system was assessed under both organized and unpredictable conditions using a mixed-data approach.

## IV. METHODOLOGY

Haar-Cascade is used for both face detection and cropping in the suggested system. The cropped face images are then sent to a Convolutional Neural Network (CNN) model for facial recognition. The primary objective is to effectively recognize faces in photos, crop them to highlight the face area, and then use CNN for recognition to identify people based on their facial traits.

1. Image Acquisition

The system begins by acquiring an input image, either from a camera (real-time capture) or a pre-existing file. This image may contain various objects and backgrounds, with one or more faces potentially present.

2. Face Detection using Haar-Cascade

The Haar-Cascade classifier handles face detection, the first and most important stage in the pipeline. For computational efficiency, the input image is transformed to grayscale since Haar-Cascade works best on grayscale images. Face detection in images is accomplished by using the Haar-Cascade classifier. This classifier is a trained model based on Haar features, which can identify face characteristics (such the mouth, nose, and eyes) based on how they appear in various lighting conditions and orientations. The bounding box's dimensions (x, y, width, and height) surrounding identified faces are returned by the Haar-Cascade. The bounding boxes are saved for later processing if more than one face is found.

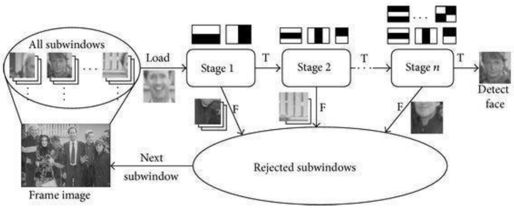


Figure 1 – Haar-Cascade Face Detection

3. Cropping the face

Following face detection, each face is cropped from the original image using the coordinates of the bounding box: To guarantee compatibility with the CNN input layer, the cropped face images are enlarged to a consistent size (e.g., 224x224 pixels). If more than one face is found, each one is cropped separately, and the procedure is then repeated for each face found.

4. Face Recognition using CNN

Once the face images are preprocessed, they are fed into a Convolutional Neural Network (CNN) for facial recognition. The CNN model consists of several convolutional layers, pooling layers, and fully connected layers. The convolutional layers extract hierarchical features from the face images, such as textures, shapes, and distinctive facial components. The pooling layers reduce the dimensionality of the feature maps while retaining essential information.

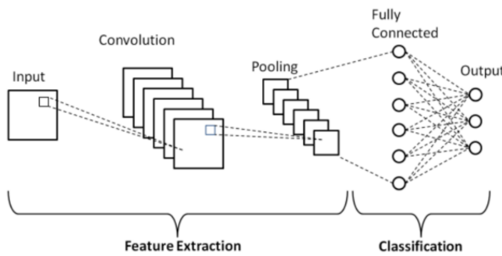


Figure 2 – CNN Architecture

The fully connected layers are used for classification, mapping the extracted features to the final output classes (i.e., the identities of the individuals). The final layer of the CNN is a softmax classifier, which outputs a probability distribution over the possible classes. Each class corresponds to a specific individual, and the CNN predicts the individual with the highest probability.

5. Training the model

Labeled face picture datasets are used to train the CNN model. The folder name act as the label name. The model is trained using a labeled dataset of face photos, each of which is linked to a class (the person's identification). Supervised learning is used to train the model. After passing the images through the network, the network uses an optimization algorithm (like Adam or SGD) to modify its weights to minimize the classification loss, which is usually categorical cross-entropy. Accuracy, precision, recall, and F1-score are used to gauge the performance of the trained model on a different validation set.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Basic CNN | 49.55 | 52.51 | 49.55 | 49.04 |
| Deep CNN | 53.12 | 54.72 | 53.50 | 52.17 |
| CNN + Face Cropped (Haar) | 71.42 | 70.37 | 68.77 | 67.44 |

Table 1 – Evaluation metrics Comparison

Depending on the data used for training and testing, facial recognition algorithms can perform very differently. We employed two different kinds of datasets—the Kaggle face dataset and real-time user data—to make sure the suggested solution is reliable, flexible, and accurate in a variety of situations. This method addresses both controlled and real-world difficulties, enabling a thorough assessment of the system's capabilities.

6. Face Detection

The CNN model can identify faces in real time once it has been trained. The trained CNN model is used to process newly discovered and clipped faces. Based on the highest likelihood score, the model then produces a forecast that matches the individual's identification. Depending on the needs of the application, such as access control or attendance logging, the anticipated identity is either shown or recorded in the system.

## V. ADVANTAGES OF PROPOSED SYSTEM

The first benefit of the suggested approach is its ease of use. The method simplifies the process by classifying faces directly rather than extracting and comparing information, which lowers processing time and complexity. As a result, activities run more quickly and require less computing power. Another important advantage is real-time performance. The system is well-suited for real-time applications like access control and attendance logging because Haar-Cascade's facial detection provides rapid and effective processing.

The technology is also very scalable. The model may be retrained to include more faces when new people are added to the dataset, enabling the system to expand without requiring a major redesign. The model's capacity to generalize is further improved by its scalability, which allows it to be adjusted to various contexts and use situations.

## VI. LIMITATIONS AND FUTURE SCOPE

### System Limitations

The proposed system has several limitations that need to be addressed for improved performance and scalability. One of the primary limitations is that when a new user needs to be added to the system, the entire model must be retrained. This could lead to significant downtime and delays, especially as the number of users grows.

The system may struggle with face occlusions, such as those caused by glasses, hats, masks, or other obstructions. These occlusions can block critical facial features, reducing the model's ability to accurately recognize the individual. Additionally, the accuracy of the system is limited by the model's architecture, which could be enhanced by using more advanced models and architectures, such as Vision Transformers (ViTs), which may offer better performance in complex face recognition tasks. While the system performs well with a small dataset, its performance may degrade as the number of users increases.

Larger datasets might present challenges in terms of processing time, memory requirements, and maintaining accurate predictions for all individuals. Another limitation is that, on real-time processing, the face detection may cause delays or incorrectly detect faces under certain conditions, especially in environments with low lighting or fast movement, affecting the overall reliability and responsiveness of the system. Moreover, while the model performs well in controlled conditions, real-time performance may fluctuate depending on the environmental and computational constraints.

### Future Scope

The pipeline's architecture is one crucial area that has to be improved in the future. The system now employs a streamlined method that combines facial detection and classification. The system might become more modular, adaptable, and simpler to update by adding distinct phases for face identification, embedding extraction, and classification. The pipeline's architecture is one crucial area that has to be improved in the future. The system now employs a streamlined method that combines facial detection and classification. The system might become more modular, adaptable, and simpler to update by adding distinct phases for face identification, embedding extraction, and classification.

Improving the system's real-time speed is essential for its use in dynamic settings. Future research could concentrate on streamlining the face detection and recognition pipeline to manage uninterrupted real-time processing. This could entail employing hardware acceleration strategies like GPUs or specialized neural processing units (NPUs) or lowering the computational load by putting more effective methods into practice.

The future acceptance of the system will depend on its ability to handle increasingly diverse user demographics and greater datasets. Techniques for dynamically adding new users to the system without necessitating retraining from scratch may be included in future developments.

## VII. REFERENCES


[1] Ridha Ilyas, Bendjillali & Beladgham, Moh & Merit, Khaled & Miloud, Kamline. (2019). Enhanced Face Recognition System Based on Deep CNN, 2019 *6th International Conference on Image and Signal Processing and their Applications (ISPA)*.

[2] R. Nachet and T. B. Stambouli, "Improved Face Recognition Rate Using Convolutional Neural

Networks," *2022 2nd International Conference on New Technologies of Information and Communication (NTIC)*.

[3] J. Zhang, "Online Face Recognition System Based on Convolutional Neural Network (CNN)," *2022 3rd International Conference on Computer Science and Management Technology (ICCSMT).*

[4] Karlupia, Namrata & Mahajan, Palak & Abrol, Pawanesh & Lehana, Parveen. (2023). "A Genetic Algorithm Based Optimized Convolutional Neural Network for Face Recognition ". *International Journal of Applied Mathematics and Computer Science.*

[5] Kaipeng Zhang et al., "Joint face detection and alignment using multitask cascaded convolutional networks", *IEEE signal processing letters*, vol. 23, no. 10, pp. 1499-1503, 2016.

[6] Ruian Liu et al., "Research on face recognition technology based on an improved LeNet-5 system", *2022 International Seminar on Computer Science and Engineering Technology (SCSET)*, 2022.

[7] "Recognition Model Based on MTCNN And Facenet," 2022 IEEE *2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*.