Face Recognition and Detection using CNN and Cascade Detection

Jyothi Kiran G C[1], Ramachandra A C[2], and Raja K B[3]

[1]BMS College of Engineering, VTU, Bengaluru, India, *jyothirkk1@gmail.com
[2] NMIT, VTU, Bengaluru, India, ramachandra.ac@nmit.ac.in
[3]University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India,

Abstract:

Because of their great accuracy and non-intrusive character, physiological biometrics are applied extensively in many different fields technological world. With an emphasis on persons of Asian ancestry, we show in this work a complete system developed using the capabilities of Cascade Detector and Convolutional Neural Networks (CNNs) for accurate and efficient face detection and recognition. With Cascade Detector, the face detection component finds important facial characteristics including the lips, nose, and eyes. We simultaneously use three CNN models—AlexNet, SqueezeNet, and GoogleNet—for RGB image facial recognition. Furthermore, used for grayscale images is a custom-built CNN model, hence increasing the system's adaptability in several imaging contexts. The models are trained on a varied dataset comprising people of Asian race, therefore allowing the system to shine in identifying facial traits. Computation of evaluation metrics including Accuracy, False Acceptance Rate (FAR) and False Rejection Rate (FRR) across several datasets helped to provide a thorough study of model performance. Among the color datasets, the Indian female dataset always showed the best accuracy, ranging from 65% to 95%; the Southeast Asian dataset showed rather less accuracy, ranging from 40% to 60%. Grayscale datasets, our custom-built CNN model attained remarkable accuracy of 92.68%; the Indian male dataset came within the range of 55% to 75% accuracy.

Keywords: Cascade Detector, Asian race, AlexNet, SqueezeNet, Demographic, Accuracy, FAR, FRR

I. Introduction

Face recognition, taken more broadly, is the process of recognizing or authenticating an individual using a digital photograph of their face. It can be difficult to recognize a human face image when partially obscured in unconstrained face recognition. Face recognition takes this a step further by identifying or verifying a person's identity based on their facial features. Recent years have seen a significant improvement in the use of face recognition systems because face biometrics are non-intrusive and human imaging equipment is widely available. Pre-processing, feature extraction, categorization, and image collection are the key components of a general face recognition system. Despite achieving human-level performance in person identification, face recognition systems still struggle with some facial images. One of the primary issues in this field is identifying identical twins, also known as monozygotic twins. Biometrics such as the face, fingerprint, and iris are used to identify identical twins. However, it is quite challenging to identify them based solely on their facial image because their faces are remarkably similar in look.

Once faces are detected, recognition involves comparing those features to a database to match or verify a person's identity. Face recognition has repeatedly shown its importance over the last years and so not only it is a widely research area of image analysis, pattern recognition in more precisely biometrics, but also it has become an important part of our everyday lives since it was introduced as one of the identification methods to be used in e- passports.

Face detection and recognition are foundational tasks in image processing with wide-ranging applications. Traditional methods like Haar cascades and HOG+SVM paved the way, In deep learning, face detection and recognition have become highly accurate and can handle complex challenges like varying lighting, facial expressions, and occlusions. Recent deep learning-based approaches, such as CNNs and YOLO, have revolutionized the field by offering much higher accuracy and robustness. The combination of these techniques allows for real-time, accurate face detection and recognition in varied conditions, enabling their deployment in everything from security systems to mobile applications. As the field evolves, new techniques continue to emerge, pushing the boundaries of what's possible in facial recognition technology.

II. Literature Survey

The process of matching NIR to VIS facial images is known as near infrared-visible (NIR-VIS) heterogeneous [1] face recognition. By creating VIS images from NIR photos, current heterogeneous approaches attempt to expand VIS face recognition techniques to the NIR spectrum. However, compared to VIS face images, NIR face images are always incomplete because of the self-occlusion and sensing gap, which cause certain visible lighting contents to be lost. High-resolution heterogeneous face synthesis is modeled by RanHe et al. as a complementing combination of two elements: a pose correction component and a texture in painting component. VIS image textures are synthesized and painted from NIR image textures by the in painting component. The corrective component creates paired NIR and VIS textures by mapping every stance in NIR images to a frontal pose in VIS images.

The intra-class differences are greatly increased by the presence of position variations [2], age changes, and various attributes in faces seen in the wild. Despite significant advancements in face recognition, only a small number of current studies are able to jointly learn local and multi-scale representations. A novel model known as Local and Multi-Scale Convolutional Neural Networks (LS-CNN) is put forth by Qiangchang et al. Multi-scale characteristics must be learned since similar discriminative face regions may appear at different scales. Introduce a novel backbone network called the harmonic Multi-Scale Network (HSNet) to achieve this goal. It uses two harmonic viewpoints to extract rich multi-scale features: concatenating multi-scale feature maps from separate layers and using multiple kernel sizes in a single layer. When the look of the face varies dramatically on a global scale, it is crucial to identify related local patches. At the same time, distinct facial regions possess varying capacities for discrimination. A spatial attention is suggested in order to weigh various local patches adaptively and capture important local commonalities. Third, different convolutional kernels in different channels are capable of detecting characteristics of varying relevance.

With the use of recognizers based on convolutional neural networks (CNNs), face recognition has advanced [3]. Current recognizers usually show strong ability to identify faces that are not obscured, but they frequently experience a decline in accuracy when attempting to directly identify faces that are obscured. The primary reason of this is the lack of identification and visual clues brought on by occlusions. However, generative adversarial networks (GANs) are especially well-suited for tasks involving the reconstruction of visually convincing occlusions by face in paintings. Inspired by these findings, Shiming et al. suggest using identity diversity in paintings to make it easier to recognize occluded faces. By differentiating variability within the same identity class, the third participant to compete with the generator is an optimal pre-trained CNN recognizer that is integrated with a GAN. A CNN-LSTM architecture and WTPCA-L1 features are used by AYYAD et al. to propose a robust facial recognition [4] model named DeepWTPCA-L1. First, face features are extracted using the WTPCA-L1 technique, which is made up of the PCA-L1 algorithm and the three-level decomposition of the discrete wavelet transform. The suggested CNN-LSTM architecture then uses the retrieved features as inputs. Several facial recognition datasets have been utilized to assess the suggested method's robustness. Furthermore, the suggested approach is trained on noisy photos by adding Gaussian and Salt & Pepper noise to each dataset's facial images.

The facial feature-based recognition job has become more challenging due to the high degree of face similarity between twins [5]. To address the issue, Shokoufeh et al. provide a Distinctive Landmark-based Face Recognition (DLFR) method. The features are suggested based on the most notable landmark area of the face as well as the quantity of important points derived via a modified scale-invariant feature transform technique. A support vector machine classifier is used to classify the weighted features. Extensive tests on 440 identical twin and non-twin face photos demonstrate the DLFR system's superior performance. Given that aging has a substantial impact on facial appearance, Cross-Age Face Recognition (CAFR) [6] continues to be one of the most difficult issues in the field of face recognition. The lack of adequate special datasets over a broad age range is another drawback. Separating aging-related fluctuations from face features and obtaining stable person-specific traits is the key to solving this issue. In particular, Yangjian et al. presented the Age Adversarial Convolutional Neural Network (AA-CNN), a revolutionary end-to-end CNN technique with a parallel network topology. The features recovered by AA-CNN are invariant to age variation through adversarial training in the Age Discrimination Network (ADN), and they maintain their identity discriminativeness through joint training in the Identity Recognition Network (IRN).

Research in the field of face recognition [7] has made significant strides in the last ten years, especially when it comes to uncontrolled situations (facial recognition in the wild). The enormous popularity and efficacy of deep convolutional neural networks, as well as the accessibility of larger unconstrained datasets, have contributed to this progress. Nonetheless, a number of face recognition issues still exist in the context of homogeneous (same domain) and heterogeneous (different domain) face recognition at very low resolution. The very low resolution face recognition problem is addressed by LUIS et al., who also offer a thorough examination of the methods' design, efficacy, and efficiency for a real-time surveillance application. Face recognition is recognized as the fundamental biometric security system and is a significant area of research in computer vision applications [8]. An Optimal Face Recognition Network (OPFaceNet) is created by Guru Kumar Lokku et al. to identify face photos that are obstructed and noisy. LBP, FLBP, and NRLBP feature patterns that are exposed to noise are extracted. The Convolutional Neural Network (CNN) classifier that is suggested is given the average of all three patterns. The Fitness Sorted Rider Optimization Algorithm (FS-ROA) optimization is the primary contribution to the CNN model. The convolutional layer, pooling layer, fully connected layer, number of hidden layers, and kind of pooling are among the CNN hyperparameters that are optimized by this technique. A type of biometric [9] security called facial recognition is commonly utilized in many different businesses to recognize and verify a person's identity by looking at their face. By collecting and training millions of face photos, face recognition datasets are helping to achieve state-of-the-art accuracy in the current deep learning era. Low-quality face photos and unintentional inaccurate annotations by annotators make it difficult to annotate such a large-scale face recognition collection. When a deep learning model is trained with these kinds of uncertainties, it overfits noisy, uncertain samples and loses its discriminative power. In order to overcome these problems, Abhijeet et al. suggest a straightforward yet powerful uncertainty learning network that effectively lessens over-fitting brought on by ambiguous face images. More precisely, at the decision layer, the FC module assigns a weight to every sample in the mini-batch.

Applications that employ selfie photos are particularly fond of beautification and augmented reality effects [10]. However, they have the power to alter or distort biometric traits, which can seriously impair the ability to identify a person or even recognize their face.

Consequently, to examine how these filters affect automated face detection and identification accuracy. The social media image filters under study alter the lighting, contrast, or obscure certain facial features. Some of these filters have been found to have negative effects on face detection and identity recognition, particularly when they obscure the eye or, to a lesser extent, the nose. Pontus et al. create a technique to reverse the applied modification using a modified version of the U-NET segmentation network in order to counteract such an effect. It has been noted that this technique improves the accuracy of face identification and recognition. The Covid-19 virus [11] and its variants have spread over the world, creating new requirements and issues that have a significant impact on our daily life. Although masks are the best way to stop the virus from spreading, wearing them has led to a number of security flaws. Additionally, this pandemic reduces the effectiveness of many standard biometric verification methods, including facial security checks, gated community entry management, and facial attendance. Therefore, this study's primary goal is to find people who either don't wear masks or wear them improperly and then utilize a masked face dataset to confirm their identification. Based on an ensemble of fine-tuned lightweight deep Convolutional Neural Networks (CNN), BUSRA et al. created a revolutionary real-time masked detection service and facial recognition mobile application. Using 1849 face samples from 12 people, the suggested model obtains a validation accuracy of 90.40 percent.

In many different fields, Face Recognition (FR) has been a popular biometric [12] method for identification verification. Even though FR has advanced significantly in recent years, there are still issues that need to be resolved, such as the recognition of faces with significant posture variation. JUAN P et al. report a collection of convolutional neural networks (CNNs) that we have constructed to represent face patches specialized in a certain face pose orientation range, based on the concept of biological face patches identified in and utilized by human and macaque monkey brains. Genetic algorithms (GAs) are used in neuroevolution to specify the structure of three CNNs. For small, medium, and large face rotations, an evolutionary algorithm evolves each CNN to a certain face pose orientation range. One convolution neural network (CNN) plus a series of long short-term memory (LSTM) models make up the sequential recurrent convolution network (SRCN), which was proposed by Chih-lyang et al. [13]. The CNN is used to obtain the spoken command or face emotion feature vector. A (pre-trained) CNN then provides a succession of input sub-images or spectrograms that correlate to spoken instruction and face emotion, respectively, which are reflected in a series of LSTM models with a shared weight. In short, two SRCNs are developed: one for wireless speech command recognition (SRCN-WSCR) and another for dynamic face emotion recognition (SRCN-DFER). With an average generalized recognition rate of 98% and 96.7%, the suggested method not only successfully tackles the recognition of dynamic mapping off ace mood and speech command. Inspired by the hybrid design of EdgeNeXt, Anjith et al. present EdgeFace, a lightweight [14] and effective face recognition network. EdgeFace delivers exceptional facial recognition performance tailored for edge devices by skillfully combining the advantages of both CNN and Transformer models with a low rank linear layer. The suggested EdgeFace network is appropriate for deployment on edge devices since it not only maintains low computing costs and small storage, but it also achieves excellent face recognition accuracy. Due to the absence of features, face masks present difficulties for face recognition [15]. To increase recognition, current methods mostly rely on painting and reconstruction; however, reconstructed images frequently lose their identity and have low facial likeness. This occurs because of the newly reconstructed traits being either newly developed or derived from other people. By enhancing only the lower portion of masked face photos rather than creating the complete face, SUSANTA et al. provide a way to increase the SSIM value, face recognition accuracy, and identity preservation. It first learns the boundary between the visible and invisible portions of a face by analysing many masked face photos to identify the occluded area. After that, it generates two datasets with the faces' upper and bottom portions.

III. Proposed Methodology

Figures 1 shows the block diagram of Face Recognition and Detection Model based on CNN and Cascade Detector.

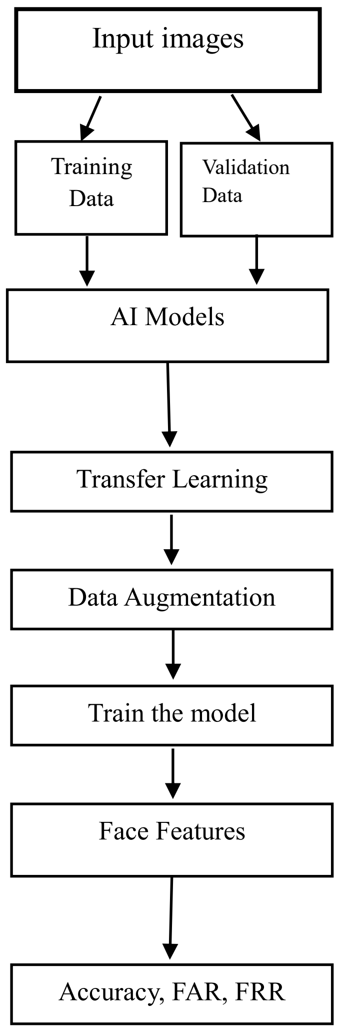

Figure 1: Block diagram of FRD-CNN-CD model

A. Datasets

Development and evaluation of fair, reliable facial recognition algorithms depend critically on datasets. Datasets corresponding to several South Asian Ethnicities—Japanese, Korean, Thai, Indian Male and Female individuals—are thus developed. One may essentially categorize these datasets as Gray Scale pictures and RGB images. Along with Southeast Asian companies including Korean, Japanese, and Thai individuals, the RGB pictures files feature Indian Male and Female as well. The Grayscale images comprise a dataset of photos of Japanese people with 10 classes consisting of 213 images that was well as a reduced ORL, face dataset that has 100 images distributed over 10 classes. Figure 2 to 6 shows the datasets as they stand.

Figure 2: Sample of ORL_face dataset

Figure 3: Sample of JAFFE dataset

The created datasets are matching numerous South Asian ethnicities: Japanese, Korean, Thai, Indian male and female individuals. These sets can be basically classified as RGB images and Gray Scale images. Along with corporations from Southeast Asia including Thai, Japanese, and Korean, the RGB photo files also include Indian Male and Female. The Grayscale pictures comprise a dataset of photos of Japanese individuals with 10 classes consisting of 213 images that was well as a reduced ORL, face dataset of 100 images dispersed over 10 class.

Figure 4: Sample of Indian male dataset

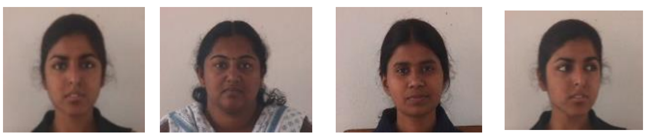

Figure 5: Sample of Indian female dataset

The produced datasets reflect many South Asian ethnicities: Japanese, Korean, Thai, Indian male and female people. One may mostly categorize these sets as Gray Scale images and RGB images. Along with businesses from Southeast Asia including Thai, Japanese, and Korean, the RGB picture files also feature Indian Male and Female. The Grayscale images consist of a dataset of photos of Japanese people with 10 classes totalling 213 images that was well as a reduced ORL, face dataset of 100 images scattered over 10 class.

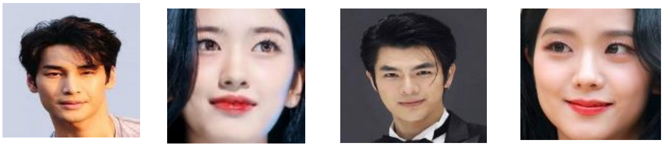

Figure 6. Sample of South Asian Dataset

B. Face Recognition

CNNs are specialized deep neural networks used mostly for processing visual input, such as photos and videos—in this case, facial images. CNNs usually have alternating convolutional and pooling layers—also known as feature extractors—followed by fully connected layers—for classification or regression—shown in Figure 7. Sliding the learnable filters across the input images and running convolution operations helps convolutional layers extract features from input data. The following convolutions are used to provide non-linearity are activation functions such as ReLU. By summarizing areas, so lowering spatial dimensions, and so controlling overfitting, pooling layers down sample feature maps. By choosing the highest activations inside the local region, max pooling helps shrink the spatial dimensions of the feature map while still preserving the most salient features. Usually, fully connected layers are employed for classification problems at the end of the network. Usually, soft max activation is used to derive probability for several classes. Equation 1 presents a broad statement of the convolutional layer.

$$Y_i = f\left(\sum_{j=1}^{N}(X * W_{ij}) + bi\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

In a Convolutional Neural Network, each feature map Y_i, which denotes the output of a convolutional layer, is generated by applying an activation function f to the convolution operation ($*$) between the input data X and the associated filters or kernels W_{ij}. The filters are designed to capture various patterns or features present in the input data. Each feature map Y_i is linked to a bias term bi, which enhances flexibility and assists in the representation of complex relationships. The subscript i indicates the index of the feature map, which varies from 1 to N, with N signifying the total count of filters present in the convolutional layer. CNNs utilize this process to extract hierarchical representations of input data, facilitating tasks including image recognition and object detection, among others. Equation for Maxpooling layer is given by Equation 2.

$$Y_{i,j,k} = max_{p,q}\left(X_{i\cdot s+p,j\cdot s+q,k}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

In the pooling layer of a convolutional neural network (CNN), the output at position (i, j, k), represented as $Y_{i,j,k}$ is calculated by executing pooling operations on the input feature map $X_{i\cdot s+p, j\cdot s+q, k}$. The variable s denotes the stride, which defines the step size utilized during the pooling operation on the input feature map. The indices p and q traverse the pooling window, during which a maximum operation is generally executed to identify the highest value contained within the window. This procedure efficiently minimizes the spatial dimensions of the input feature maps while maintaining critical features. Pooling layers enhance the network's capacity to abstract and summarize information, facilitating tasks like feature extraction and dimensionality reduction in CNN architectures.

The proposed methodology utilizes three distinct pre-trained convolutional neural networks (CNNs): GoogleNet, AlexNet, and SqueezeNet, for the classification of RGB images, alongside a custom-built model designated for the classification of grayscale images. Each of these networks comprises the previously mentioned layers in varying quantities, which contribute to their distinct architecture and computational capabilities.

Transfer learning in convolutional neural networks (CNNs) involves utilizing a pre-trained model, developed on a substantial dataset for a specific task, as the foundational model for a new, yet related task. The pre-trained CNN models serve as feature extractors, with modifications applied to the last few layers of the pre-trained model. This adaptation enables the models to align their high-level representations with the specific requirements of the new task. The convolutional layers that have acquired hierarchical features from the original task are preserved. The fully connected layers are adjusted to create Facial Feature Learners through the alteration of bias and weight parameters. The output layer has been adjusted to create the face classifier layer, which categorizes facial images into their corresponding identities as shown in figure 7.

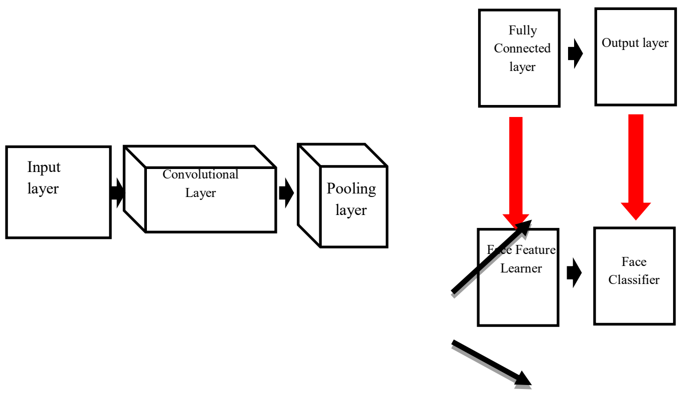

Figure 7: Modification in GoogleNet due to Transfer learning

C. Face Detection

The Cascade Object Detector operates on the principles of the Viola-Jones object detection framework, specifically for the identification of facial features such as the eyes, nose, and mouth. The Cascade Object Detector is implemented in MATLAB via the Computer Vision Toolbox™, which includes the function for training a cascade object detector utilizing both positive and negative training samples. Vision. The Cascade Object Detector functions to identify frontal faces, profile faces, noses, eyes, and the upper body within images by utilizing a pre-trained cascade classifier. The detector is configured by default to identify faces within an image. However, it can detect additional features such as the nose, mouth, eyes, or the upper part of the body, as specified by the input string Classification Model. The process of training a cascade object detector utilizing the train Cascade Object Detector function results in the storage of the trained detector in an XML file format.

The Viola-Jones approach makes use of a cascade classifier structure, AdaBoost training, and Haar-like features. The difference in the sum of the pixel intensities in neighboring rectangular parts defines Haar-like features, which are calculated over rectangular regions of an image. The 24x24 pixel windows that are shifted on the image where we wish to detect faces have their features extracted. An technique called AdaBoost builds a strong classifier by combining the predictions of several weak classifiers.

A collection of weak classifiers is present in each of the several steps that make up the cascade structure. A portion of the image moves on to the next level of detection if it passes each of the weak classifiers in that stage. If not, it is denied right away because it does not include the item that is being detected. Equation 3 serves as the basis for the decision at each stage j.

$$H(x) = \begin{cases} 1, & if\ \Sigma_{k=1}^{k_j} \cdot \alpha_j^k h_j^k(x) \geq \theta_j \\ 0, & otherwise \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots \ (3)$$

k_j Is the number of weak classifiers in stage j, α_j^k is the weight of the k^{th} weak classifier in stage j, $h_j^k(x)$ is the output of the k^{th} weak classifier in stage j and θ_j is the threshold for stage j.

First, the dataset is loaded and divided into training and validation sets. Pre-trained CNN models such as Google Net, Squeeze Net, and Alex Net are used for RGB photos, however a specially designed CNN model is used for grayscale images. The pre-trained model architectures are modified using transfer learning approaches to meet the demands of the task. To improve model generalization, transformations including random reflections, translations, and scaling are used to the training and validation datasets. The altered model is then tested against the validation set after being trained with the augmented training set. Using samples from the validation set, the trained model is tested with an emphasis on identifying face features such as the mouth, eyes, and nose. Model performance is evaluated by computing evaluation measures such as False Acceptance Rate (FAR), False Rejection Rate (FRR), and overall accuracy. In order to give information about the model's performance and accuracy in classifying facial features, the detected facial features are finally shown together with the predicted class label and matching prediction score.

IV. Results and Discussion

The Indian_Male, Indian_Female, and Southeast Asian face datasets are used in the experiment. The datasets are divided into training and validation sets in four distinct ratios: 8:2, 7:3, 6:4, and 3:7. Pixel ranges in data augmentation are varied in the following ranges for each dataset split ratio: [-30,30], [-50,30], [-30,50], and [-50,50]. All the chosen datasets—three RGB datasets for the pre-trained models and a grayscale dataset for the custom-built model—have FAR, FRR, and Accuracy for each combination of dataset split ratio and pixel range. Testing the model on people with noses, mouths, and eyes reveals that the suggested model performs better when it comes to nose recognition.

GoogleNet for Nose Detection

Nose detection is used in conjunction with the GoogleNet model. Tables 1 for the Indian female dataset, Table 2 for the Indian male dataset, and Table 3 for the Southeast Asian dataset display the accuracy values for the same. For all pixel ranges, it is found that the overall accuracy is poor for the 3:7 split ratio and high for the 8:2 ratio. Additionally, it is noted that the suggested model performs less accurately on the Southest Asian dataset.

Table 1: GoogleNet model on indian_female face dataset

Pixel_Range	Training Dataset:Testing Dataset			
	08:02	07:03	06:04	03:07
	Accuracy%			
[-30,30]	93.18	96.97	97.73	72.16
[-50,30]	95.45	87.88	94.32	61.96
[-30,50]	88.64	96.97	75.00	63.32
[-50,50]	97.97	88.88	82.92	58.52

Table 2: GoogleNet model on indian_male face dataset

Pixel_Range	Training Dataset:Testing Dataset			
	08:02	07:03	06:04	03:07
	Accuracy%			
[-30,30]	95.00	88.89	75.90	48.08
[-50,30]	85.00	90.48	78.31	37.18
[-30,50]	87.50	85.71	90.36	47.44

[-50,50]	87.50	80.93	81.93	35.26

Table 3: GoogleNet model on southeast_asian face dataset

Pixel_Range	**Training Dataset**		**Dataset:Testing**	
	08:02	07:03	06:04	03:07
	Accuracy%			
[-30,30]	43.75	33.33	43.75	31.90
[-50,30]	37.50	39.58	43.75	27.59
[-30,50]	34.38	39.58	42.19	24.19
[-50,50]	37.50	41.67	28.12	27.56

For four randomly chosen samples from the Indian male, Southeast Asian, and Indian female datasets, Figures 08, 09, and 10 show the results of the FDR system employing GoogleNet with nose, mouth, and eye detection, respectively. The projected labels and confidence ratings are presented.

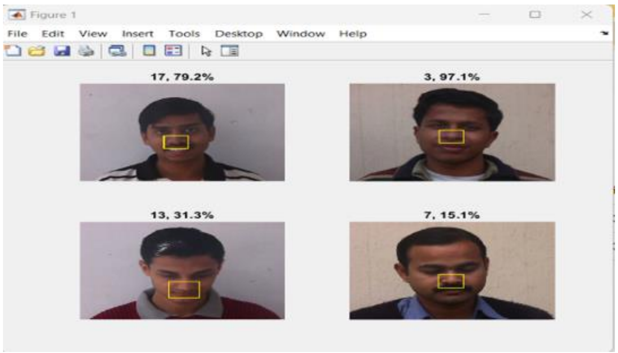

Figure 8. FDR for GoogleNet on indian_male dataset

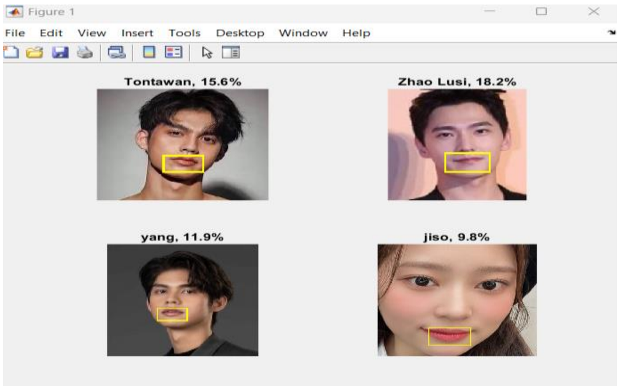

Figure 9. FDR for AlexNet on southeast_asian dataset

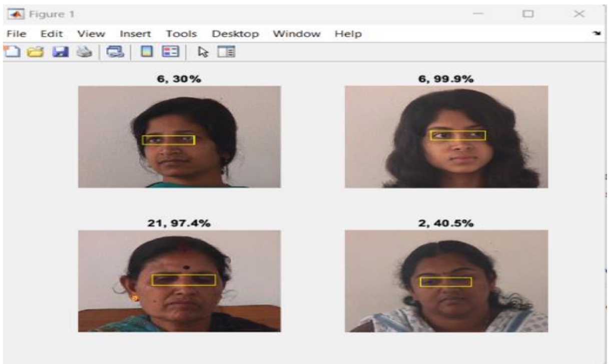

Figure 10. FDR for SqueezeNet on Indian_female dataset

Custom Built model with Face Detection

Eyes detection is used in conjunction with the SqueezeNet model. Table 4 for the Gray Scale dataset displays the accuracy values for the same. For both datasets, it is found that the overall accuracy is poor for the 3:7 split ratio and good for the 6:4 ratio.

Table 4: custom-build model on grayscale face dataset

Dataset	Training Dataset: Testing Dataset			
	08:02	07:03	06:04	03:07
	Accuracy%			
Orl_faces	90.00	63.33	85.00	74.29
JAFFE	92.68	95.31	95.24	73.15

The output of the FDR system utilizing a custom-built model with face detection is shown in Figure 11. The predicted labels and confidence ratings for four randomly chosen samples from the smaller JAFFE dataset are shown.

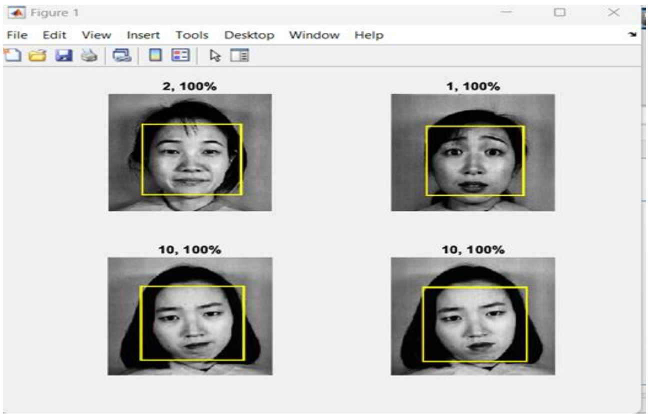

Figure 11. FDR on JAFFE dataset

Table 5: Comparison of Proposed and Existing work

Paper id	Accuracy%
Shivalila et al., [13]	94.23
Fahima et al.,[16]	93.34
Proposed	97.97

Table 5 compares the proposed work with the existing work. It is found that the suggested work provides greater accuracy than the existing work, and it will work with both RGB and grayscale photos.

V. Conclusions

We investigated a wide variety of models and datasets in our thorough examination of face recognition models utilizing CNNs and cascade detection. In addition to custom-built CNN models for the grayscale datasets JAFFE and orl_face, our analysis included well-known CNN architectures including SqueezeNet, AlexNet, and GoogleNet that were trained on datasets representing Southeast Asian, Indian male, and Indian female populations. Our research showed that accuracy varied significantly between models and datasets. The Southeast Asian dataset showed somewhat lower accuracy, ranging from 40% to 60%, whereas the Indian female dataset consistently produced the best accuracy, ranging from 65% to 95%, across the color datasets. The accuracy range for the Indian of our method in capturing facial features even in grayscale photos, achieving remarkable accuracies of 92.68% for JAFFE and 90.00% for orl_face for grayscale datasets.

References

[1] R. He, J. Cao, L. Song, Z. Sun and T. Tan, "*Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition*," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 5, pp. 1025-1037, 1 May 2020.

[2] Q. Wang and G. Guo, "LS-CNN: Characterizing Local Patches at Multiple Scales for Face Recognition," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1640-1653, 2020.

[3] S. Ge, C. Li, S. Zhao and D. Zeng, "Occluded Face Recognition in the Wild by Identity-Diversity In painting," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 10, pp. 3387-3397, Oct. 2020.

[4] A. Maafiri, O. Elharrouss, S. Rfifi, S. A. Al-Maadeed and K. Chougdali, "DeepWTPCA-L1: A New Deep Face Recognition Model Based on WTPCA-L1 Norm Features," in IEEE Access, vol. 9, pp. 65091-65100, 2021.

[5] Shokoufeh Mousavia, Mostafa Charmia," Adistinctive landmark-based face recognition system for identical twins by extracting novel weighted features" Computers & Electrical Engineering, ISSN: 0045-7906, Vol: 94, Page: 107326, 2021.

[6] Y. Huang and H. Hu, "A Parallel Architecture of Age Adversarial Convolutional Neural Network for Cross-Age Face Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 1, pp. 148-159, Jan. 2021.

[7] L. S. Luevano, L. Chang, H. Méndez-Vázquez, Y. Martínez-Díaz and M. González-Mendoza, "A Study on the Performance of Unconstrained Very Low Resolution Face Recognition: Analyzing Current Trends and New Research Directions," in IEEE Access, vol. 9, pp. 75470-75493, 2021.

[8] Gurukumar Lokkua, G.Harinatha Reddyb," OPFaceNet: OPtimized Face Recognition Network for noise and occlusion affected face images using Hyper parameters tuned Convolutional Neural Network", Applied Soft Computing, ISSN: 1568-4946, Vol: 117, Page: 108365, 2022.

[9] Abhijeet Boragule; Hamna Akram;" Learning to resolve uncertainties for large-scale face recognition", ISSN: 0167-8655, Vol: 160, Page: 58-65, 2022.

[10] Pontus Hedman; Vasilios Skepetzis; "On the effect of selfie beautification filters on face detection and recognition" ISSN: 0167-8655, Vol: 163, Page: 104-111, 2022.

[11] B. Kocacinar, B. Tas, F. P. Akbulut, C. Catal and D. Mishra, "A Real-Time CNN-Based Lightweight Mobile Masked Face Recognition System," in IEEE Access, vol. 10, pp. 63496-63507, 2022.

[12] J. P. Perez and C. A. Perez, "Face Patches Designed Through Neuroevolution for Face Recognition With Large Pose Variation," in IEEE Access, vol. 11, pp. 72861-72873, 2023.

[13] C. -L. Hwang, Y. -C. Deng and S. -E. Pu, "Human–Robot Collaboration Using Sequential-Recurrent-Convolution-Network-Based Dynamic Face Emotion and Wireless Speech Command Recognitions," in IEEE Access, vol. 11, pp. 37269-37282, 2023.

[14] A. George, C. Ecabert, H. O. Shahreza, K. Kotwal and S. Marcel, "EdgeFace: Efficient Face Recognition Model for Edge Devices," in IEEE Transactions on Biometrics, Behaviour, and Identity Science, vol. 6, no. 2, pp. 158-168, April 2024.

[15] S. Malakar, W. Chiracharit and K. Chamnongthai, "Masked Face Recognition With Generated Occluded Part Using Image Augmentation and CNN Maintaining Face Identity," in IEEE Access, vol. 12, pp. 126356-126375, 2024.