Journal of Theoretical and Applied Information Technology
15th May 2025. Vol.103. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MULTI-CNN MODEL TO EVALUATE THE PERFORMANCE OF FACE DETECTION AND RECOGNITION WITH FACIAL FEATURE DETECTION AND RECOGNITION

[1]G. SUJATHA, [2]MATTA SWATHI, [3]BHAGYA PRASAD BUGGE, [4]SHAIK JOHNY BASHA, [5]ALLURI SWATHI, [6]BALA PRASANTHI PAVULURI, [7]M SITHA RAM, [8]SURYA PRASADA RAO BORRA

[1]Department of CSE, Shri Vishnu Engineering College for Women, Bhimavaram -India.
[2]Department of CSE, LakiReddy Bali Reddy College of Engineering, Mylavaram-India
[3]Department of ECE, SRKR Engineering College, Bhimavaram – India.
[5]Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram – India.
[4,7]Department of CSE, School of SEAS, SRM University-Mangalagiri - India.
[6]Department of ECE, R.V.R. & J.C. College of Engineering, Guntur - India.
[8]Department of ECE, PVP Siddhartha Institute of Technology, Vijayawada - India.
Corresponding Author: suryaborra1679@pvpsiddhartha.ac.in

ABSTRACT

Face Recognition is one of the most advanced and drastically growing research areas because it helps identify people globally in various ethical and unethical applications. Face recognition needs face detection that can be compared with a list of available faces to predict the correct person. Face detection has become popular, easy, and fast since it follows the Viola-Jones FD method. Face comparison is obtained by comparing the internal and external information from the face images, like different features, face structure, key points, and patch-by-patch comparison. Earlier face recognition methods used separate algorithms for feature extraction from the face images, like color, shape, texture, histogram, and local and global binary pattern, to compare pairs of images where they provide more complexity regarding computation, cost, and time. After the evolution of artificial intelligence models, recent research has focused on using machine and deep learning algorithms for face detection and recognition. However, the accuracy of face recognition models needs to be improved under various conditions. Thus, this paper used a two-stage face comparison model to enhance face recognition efficiency. A consequence of three CNN models called CNN-1, CNN-2, and CNN-3 are used to detect the faces, detect the facial features, and recognize the faces, respectively. The CNN models are implemented in Python, and the results are verified by experimenting with multiple benchmark face datasets. The output accuracy obtained from the face detection and recognition is compared with the facial feature detection and recognition to choose the best to identify the criminals. From the comparison, both FDR and FFDR obtained 99.68% accuracy equally

Keywords: *CNN, Deep Learning Algorithm, Face Detection, Facial Feature Detection, Face Recognition.*

1. INTRODUCTION

Artificial intelligence is continuously used in developing applications exponentially due to the wide availability of advanced monitoring, research, and visualization technologies, as well as the rapid growth in computational performance [1]. Many research investigations and commercial organizations have shown that computers can interact with people naturally by employing cameras to observe people and microphones to listen to them. Figuring out these inputs and uniquely responding to them is considered to be very essential. Face detection has been considered the foremost application in image processing for a decade. It is a significant part of face recognition. While considering its performance over human faces, there are various challenging aspects in real-time scenarios since there are several sudden changes in expression, posture, hair color, lighting, camera quality, etc [2]. Object detection is a computerized approach that can be utilized along with computer vision and image-processing techniques to detect occurrences in an object [3-4].

Determining the presence of an image is the crucial objective of face-detection approaches. Many research and investigations have been carried out to enhance the face-detection approach, and the significant milestone of such works is Viola-Jones since it carried out face-detection outcomes in real-time applications such as surveillance monitoring [5].

Earlier, a typical structure was applied to all the Face Detection and Recognition (FDR) models, shown in Fig.1. All the images in the dataset are arranged in folders based on the names of face IDs. All the faces are preprocessed, feature extracted, and compared with the features of the face images available in the database. Different algorithms extract features like geometrical, GLCM, PCA, LCA, etc, for face comparison. Before face comparison, most FR systems used the predefined Violo Jones model for face detection, which gives 100% accuracy.

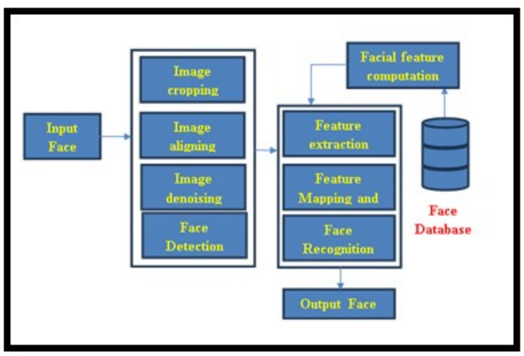

Figure-1. Face Detection Architecture

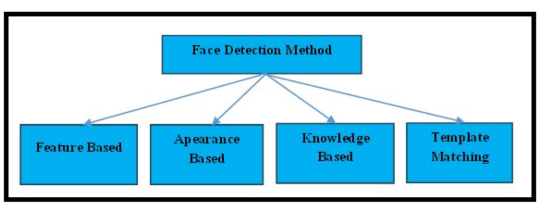

Figure -2. Methods in face detection.

Face detection methods are classified into four types: two or more groups—Yan, Kriegman, and Ahuja present face detection methods. The categories of face detection methods are explained below and shown in Fig.:2

Feature-based methods can place faces with the help of extracting structural features of the face. It should be prepared as an organizer and separate the facial and nonfacial regions. This feature is split into more steps, and they report a success rate of 94% in photos with many faces [6]. **The knowledge-based method is** based on a set of rules and human knowledge to detect faces.

These methods found many problems in building a set of rules. There are so many false positives if the rules are too general. This method alone is insufficient; many faces cannot be found in multiple images. **Template Matching** method uses pre-defined face templates to detect the faces by the correlation between the templates and input images. A face model can typically build the edges using the edge detection method. This approach is simple to implement but inadequate for face detection [7]. Still, the deformable templates have been planned to deal with these problems. **The appearance-based** method mainly depends on a group of delegate training face images to figure out the models [8]. Statistical analysis and machine learning are used to determine appearance-based face images.

Table-1. Appearance Based Methods

Method	Description
Eigenface	It is used for Face Recognition and efficiently represents faces using Principal Component Analysis.
Distribution	Distribution-based algorithms can be used to define the subspace representing facial patterns. Here, a trained classifier correctly identifies instances of the target pattern class from the background image patterns.
Neural-Networks	Most detection problems are solved successfully in the neural networks, including some detection problems like object detection, emotion detection, face detection, etc.
Support Vector Machine	It is a linear classifier that increases the margin between the decision hyperplane and the examples in the training set. Osuna et al. applied the first face detection classifier.
Sparse Network of Winnows	These windows have two target nodes, one for face patterns and the other for non-face patterns. This is less time-consuming and more efficient.
Naive Bayes Classifiers	This classifier calculated the probability of a face appearing in the image by counting the frequency occurrence in a pattern series over the training images. Naïve Bayes Classifiers captured the position of the faces and the local appearance of joint statistics.
Hidden Markov Model	HMMs are commonly used with other methods to build detection algorithms. The model is the facial features, usually described as strips of pixels.

It is classified into some face detection methods, as shown in Table-1. One of the **Information Theoretical Approaches** is Markov Random Fields (MRF), which can be used for face patterns and correlated features. This Markov process increases the discrimination between

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classes using Kullback-Leibler divergence. Therefore, this approach can be used in Face Detection. This Inductive Learning has been used to detect faces, and Algorithms like Quinlan's C4.5 or Mitchell's FIND-S have also been used for this purpose. Several methods with high accuracy were proposed for face detection using the abovementioned methods. These methods follow a similar procedure for face detection, like Viola Jones, Matlab, and Neural Networks. Viola Jones is an OpenCV used for face detection, which follows some predefined steps to increase face detection accuracy.

Compared with earlier research methods, this paper contributes the following to increase the overall efficiency of the face recognition model with the highest accuracy.

1. This presents the essential information regarding face detection and recognition and facial feature detection and recognition to help understand the overall research work.
2. The layer performance and sample architecture explain the CNN model in detail.
3. The CNN model is implemented in Python. It has been experimented with different face datasets and demonstrated face detection, facial feature detection, and face recognition through its outputs.
4. The proposed CNN model's performance is evaluated by comparing its experimental results with similar models discussed earlier.

This research addresses these limitations by proposing a two-stage face comparison model using three specialized convolutional neural networks (CNNs) to individually handle face detection, feature extraction, and face recognition. By validating the model with multiple benchmark datasets and achieving a high accuracy of 99.68% in both face detection-based and facial feature-based recognition, this study provides a justified and timely contribution to the ongoing efforts in improving face recognition systems for critical real-world applications.

2. LITERATURE SURVEY

The basic information about various methods proposed earlier for face detection and recognition processes is discussed here. It helps to understand the issues and challenges of the earlier methods, which can be rectified in the proposed model. For the past decade, Face detection has been a field where steady and progressive achievements are getting done. Zaferious et al. (2015) surveyed the various facial detection methods and the growth of the future methodologies required for facial detection. The authors classified the face detection methodologies into two broad categories, starting with the Seminal Viola-Jones Face detector. The authors classified facial detection into general schemes like rigid templates by boosting-based methods, deep learning models, and deformable models by face parts. Yang et al. (2016) have addressed the difference between the real-time requirement and face detection performance. To enhance face detection performance, the authors have introduced the WIDER FACE datasets, which are ten times more than the available datasets. This dataset has data about the scale, pose, occlusion, face bounding, and event categories. Jiang and Learned-Miller (2017) have used this WIDER FACE dataset to fasten the existing R-CNN base framework and two more benchmarks, namely IARPA Janus Benchmark and Face Detection Dataset and Benchmark. The Faster R-CNN resulted in incredible performance with the WIDER FAE dataset. The authors suggest that the performance can be further enhanced by taking the unique patterns in the human faces. In a comparative analysis, Face detection based on various Facial features like nose, eyebrows, mouth, hairline, and skin color with the help of edge detectors need a relationship identifier. This identifier needs to relate the features from the facial or non-facial images, which can be achieved with the help of Machine learning and Statistical analysis. Some of the models are Neural Networks, Adaboost learning, SVM, and HMM, which utilize the standard databases like MIT, PIE, The Yale Face database, FERET, AR, Indian face databases, and SCface-surveillance cameras database, which relates the features by Kumar et al.(2019). The

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

authors suggest that in the future, Facial detection and Improper Illumination and Occlusion can help with critical things like Facial recognition and Facial expression recognition for payment, healthcare, security, criminal identification, etc.

Zhang et al. (2020) and H. Ge et al. (2021) have proposed the Multi-task Convolutional Neural Network (MTCNN) for face detection. The authors considered the face image, improper illumination, low pixel, off-face, and face key point locations for face detection in a challenging situation. These key points of face detection in a critical scenario can be tested with the MTCNN, yielding better results with improved accuracy. The speed of MTCNN's face detection becomes inspiring for the real-time need. It yields better results than the existing yolov3, which uses the WIDER FACE datasets. For face detection, a lot of facial features are required and identified. Various research works were done to make an effective face detection using different Facial Features. N.R. Borkan and S. Kuwelkar (2017) and Hasan et al. (2021) conducted a detailed survey about the implemented face detection methodologies and facial features. This survey was aimed to provide the future scope of Face detection and the Facial Features involved. The authors listed out the various methods like the Viola-Jones algorithm, Local Binary Pattern, Adaboost, and various image-based methods based on Neural Networks like Artificial Neural Networks, Convolutional Neural Networks, Rotation Invariant Neural Networks, Fast Neural Networks, Decision-Based Neural Networks, Fuzzy Neural Network, etc. The abovementioned methodologies considered various facial features, such as existing and newer ones, required at the specific period. They are the eyes, nose, ears, face outline, motion-based information, color information, gray information, and edge. Various methodologies have been used to analyze these techniques; among them, neural networks yielded comparatively better accuracy and results when compared with other existing methods. The authors concluded that Face detection with different facial features concerning the requirement can be developed. In such cases, it does not mean that the specific Neural Network is best for all other purposes; it must be designed based on the demand and facial features. That may be the combination, or the features decided. An upgraded version of the face detector was developed from its existing model, yolo, as the YOLOv5 object detector or YOLO5Face. This particular model was developed for embedded or mobile devices in accurate time detection by F. Majeed et al. (2021) and Qi et al. (2023). The authors have used the WiderFace dataset and achieved state-of-the-art performance in terms of results. H. Wu et al. (2021) and Saleem et al. (2023) have developed a Face Recognition method with the Facial features of Euclidean distance between the eyes, lip-to-lip distance, and nose structure. The authors have analyzed the performance of facial landmarks and facets of crime detection, human tracking, and face verification. In this work, the listed features were changed to the array and region of interest format to recognize the person. These proposed methodologies can be used in industrial areas, identifying and tracking the human in a mass of people and the trespassers. This can be achieved by enabling facial biometric identification in the massive working personnel. It is further extended where a high-security alert is needed.

Face recognition is vital in identifying humans using biometric techniques such as facial features. This face recognition is needed to ensure security and protection systems like biometric access to offices or households. To ensure safety using face recognition, Adnan et al. (2020) and A.P. Rajan and A.R. Mathew (2019) have utilized HOG, SURF, LBP, HAAR, and PCA methods to extract the features. The authors adopted the KNN algorithm to analyze the datasets of 100 numbers available. Among the various proposed models, HOG achieves a hit rate of 85%, according to the authors. Upadhyay and Kotak (2020) proposed a model to identify emotions from images by extracting features of the faces, such as mouth, eyes, etc. The authors introduced a facial emotion recognition system that mimics the functions of the human brain to analyze facial expressions. In this work, the author used a geometry-based approach, Template-based approach, Appearance-based approach, and Colour segmentation technique.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Among these techniques, the geometry-based approach stands out from the others with an accuracy of 88.9%. In this work, the authors used both the Frontal and non-frontal images.

From the above discussion, it is noticed that earlier studies have many limitations. The earlier methods provided solutions for solving FDR problems for only one sample and trained the models. Solving large amounts of data, like implementing deep learning models from scratch is not sufficient. Also, the approaches cannot detect the faces with occlusion, mask, poseses, and other conditions on the face images. Most FDR algorithms and methods can provide the best accuracy only for a fixed dataset, not for unknown or conditioned images. This paper tried to provide a better solution for a generic FDR method using a deep learning algorithm, CNN.

Previous studies have employed the Viola-Jones algorithm for face detection due to its speed and simplicity, while others have utilized various machine learning and deep learning techniques for face comparison and recognition (Turk & Pentland, 1991; Ahonen et al., 2006; Parkhi et al., 2015). However, most of these methods either focus on isolated components—such as feature extraction or recognition—or struggle to maintain performance consistency across datasets. Building on this existing body of work, the present research addresses these gaps by proposing a two-stage face comparison model using three CNNs (CNN-1, CNN-2, and CNN-3) that handle face detection, feature extraction, and recognition separately but cohesively. This architecture not only simplifies the computational process but also achieves a high and balanced accuracy of 99.68% across both face detection-based and feature detection-based recognition approaches, thus contributing a novel and practical solution to the ongoing challenges identified in the literature.

3. PROPOSED METHOD

This paper proposes an SNN-based neuromorphic system to detect objects from occluded images. This paper proposed a deep learning Convolution Neural Network model for face detection and recognition. These two methods are related tasks in computer vision and digital image processing applications. The main objective of this paper is to apply facial feature detection-based face detection and recognition to identify a criminal through an image or video. This paper provides the output as a bounding box for face detection and name on face detection. Identifying the region of the face and drawing a bounding box is the output of the face detection process. The bounding box contains the face. Many face detection methods, from traditional Haar-like methods to advanced deep learning methods like R-CNN, are used. It also aims to identify and verify the facial features of the face image, which provides uniqueness in face detection and recognition accuracy. The detected facial features are embedded to compare and verify with each face image available in the face database. This paper uses an efficient deep learning algorithm, the Convolution Neural Network model, for face detection, facial feature detection, and face recognition. The proposed CNN model is trained with a general face image dataset with different conditions, like different poses, angles, and changes in the face due to age. Initially, the CNN model is trained with face images and implemented as CNN-1, and then it is trained with facial feature detection over face images and implemented as CNN-2. Finally, the CNN-3 is created by training with the output obtained from CNN-1 and CNN-2 and compared with the whole set of face images for face recognition.

The entire work process is divided into three portions: face detection, facial feature detection, and face recognition. Face detection locates the face in an image or frame taken from a video. Face recognition helps to identify and verify a person's face by comparing the input face image with the entire database of images. Face detection and recognition are used in several applications, such as surveillance monitoring, security systems, authentication and authorization, and human-computer interaction. The evolving advancements of deep learning algorithms are increasing the strength and accuracy of face detection and recognition systems. The accuracy of FDR is improved here by using trained benchmark face images that are already labeled. Hence, the face recognition process can compare the features of the faces obtained from CNN-1 and CNN-2. This

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

paper also includes distance calculation among the trained, training, and testing images given as input to CNN-1, CNN-2, and CNN-3, which helps to perform various quality-based images, such as low and high qualities. The proposed CNN deep learning model is explained below.

3.1 Spiking Neural Network

Hubel and Wiesel initiated the individual network structure that can effectively reduce the complexity of the neural network feedback in the study of neurons with local sensitivity and directions in the cortex of cats, which was found in the 1960s. CNN has become a vast research topic in many scientific fields, mainly in the pattern classification fields, and it avoids the complex preprocessing of the image and directly inputs the original image so that it can be used in various fields. It is non-fully connected with a multilayer neural network. This layer contains convolutional layers, a sampling layer, and a hidden layer composed of these two layers in the whole network model. The convolution operation of the CNN model is illustrated in Fig.:3

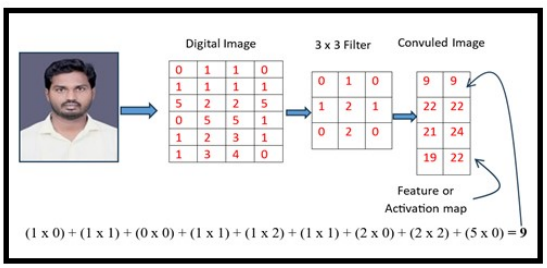

Fig.:3. Convolution Operation of CNN

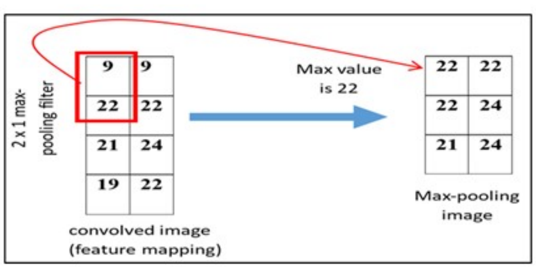

Fig.4. Max-Pooling Operation of CNN

Generally, the basic structure of CNN consists of two layers: one is the convolution layer, and the next one is the extraction layer (Figure-4). In the previous layer, the input of each neuron is connected to the local acceptance domain, and then the regional features are extracted. When the local features are extracted, the positional relationship with other features is determined; each layer has a plurality of feature maps, and this map has multiple neurons. The sampling layer is a feature map layer (Figure-4), and each computing layer of the network has a plurality of feature maps. This map is a plane where the weights of all neurons are equal. The mapped features are flattened using the flattened layer (Fig.5).

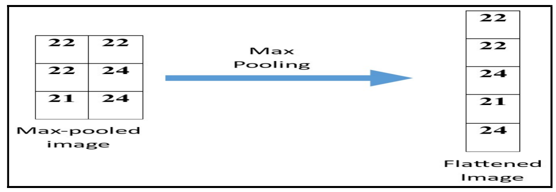

Fig.5. Flattening Operation of CNN

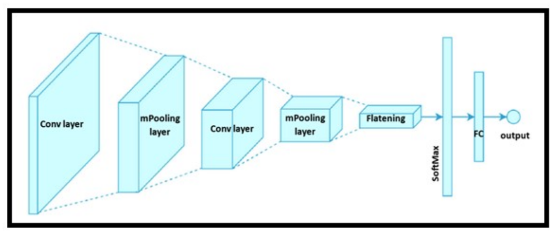

Figure-6. CNN Architecture.

The number of free parameters is reduced because the neurons share their weight on the map. A computational layer for local averaging and quadratic extraction follows CNN. The individual structure reduces the feature solution twice with the feature extraction. The proposed model is experimented on the Windows system, training a CNN model on TensorFlow. Fig.6 shows the network structure. The network design comprises 11 layers, including four convolution layers, four sampling layers, two fully connected layers, and one output layer. Each input layer contains the training parameters (connection weights).

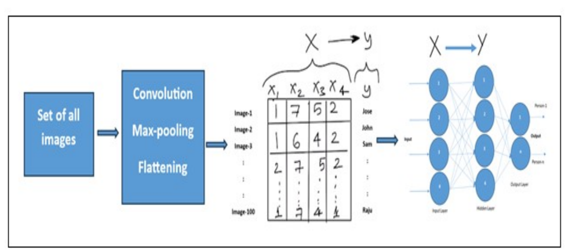

Figure-7. CNN Workflow Diagram

Journal of Theoretical and Applied Information Technology
15th May 2025. Vol.103. No.9
© Little Lion Scientific

The size of the input images is 112*112, normalized by the face detection process TensorFlow is an open-source software library that uses data flow graphs for numerical computing. It transfers complex data structures to artificial neural networks for analysis. It can be used in many fields, especially deep learning, such as speech and image recognition. During training, each image is fed into CNN as a sequence of rows in a matrix.

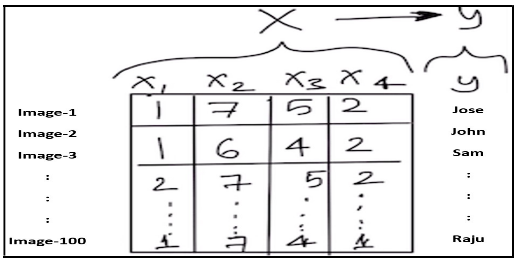

Fig.8. Training Process

Regarding face recognition, the name of the face indicates the person of the face. Different person names, like Jose, Kanee, Raju, Sam, etc., are the labels of the face images that will be classified as recognition output. Fig.7 summarizes the entire CNN workflow diagram. It shows that, initially, the features are extracted from the input images to convert them into numerical format. Then, the artificial neurons in the CNN model learn the number against the class label.

Algorithm: CNN Model Implementation Steps.

Step 1: Import the input image and convert it from RGB to Grayscale.

Step 2: Manipulate the image concerning cropping, resizing, blurring, enhancement, and sharpening.

Step 3: Find and detect the contour and segment the image objects because it helps to classify accurately.

Step 4: The haar-like features method is called the Viola-Jones method to identify and detect facial features from the faces.

Step 5: Using the eyes, nose, mouth, and other facial features, it is easy to recognize the face among several face images in a picture.

Step 6: The x, y, height, and width coordinates are obtained to draw a bounding box on the face to show the exact face region in the image.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This paper used an Intel Pentium Core-i5 processor with 12GB RAM and 1TB HDD. The processor's speed is 2.34GHz, installed with NVIDIA GTX, a 64-bit Windows 10 operating system. This paper takes the input data from two different datasets to perform face recognition and feature extraction. The ORL-face dataset is used to evaluate the model's performance on facial recognition. The ORL_face dataset includes 400 facial images from 40 subjects from different locations, times, and poses. But all the images are captured with dark homogenous backgrounds with the pixel size and grey level of 92 x112 and 256, respectively. And to perform facial feature extraction process the input image samples are taken from the pin face recognition dataset. It includes 17534 images of 105 celebrities captured at different times, poses, and locations. As mentioned, the main goal of this research is to identify the face from a video stream or images. For this, a face recognition algorithm is applied and evaluated. The face recognition process is entirely based on matching the facial features of the input samples with listed samples. The pre-trained CNN model is used in this paper to recognize and classify facial images.

4.1 CNN Based Face Recognition system

In this research work, a deep learning-based CNN model is applied to detect and recognize faces similar to those in the input images.

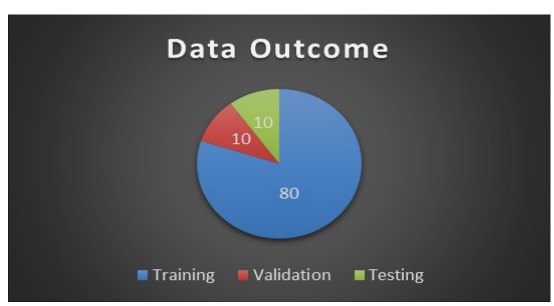

Fig.9. Classified Data Sample

Table-1. Layer Configuration of CNN

Model: "sequential"		
Layer (type)	Output shape	Param #
conv2d (Conv2D)	(None, 106, 86, 36)	1800
max_pooling2d (MaxPooling2D)	(None, 53, 43,36)	0
conv2d _1 (Conv2D)	(None, 49, 39, 54)	48654
max_pooling2d_1 (MaxPooling2	(None, 24, 19, 54)	0
flatten (Flatten)	(None, 24624)	0
dense (Dense)	(None, 2024)	49841000
dropout (Dropout)	(None, 2024)	0
dense_1 (Dense)	(None, 1024)	2073600
dropout_1 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 512)	524800
dropout_2 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 20)	10260
Total params: 52, 500, 114 Trainable params: 52, 500, 114 Non-trainable params: 0		

To accurately detect the input images, the overall input samples are classified into three phases: training, testing, and validation The classified input image is shown in Fig.9.

The analysis result depicts that due to a smaller number of input samples, only a smaller percentage of input samples are used for validation. The input facial image samples are classified into three phases with a ratio of 8:1:1.

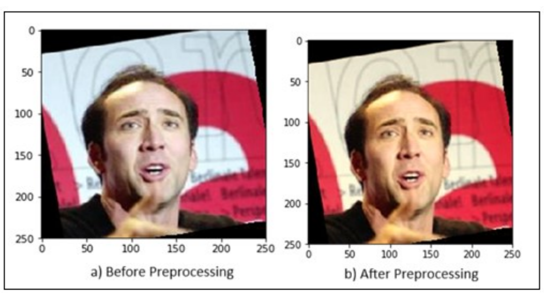

Figure -10. Pre-Processing Output

Table-1 shows the total number of parameters used in the proposed model to recognize the input facial images. The proposed CNN model contains 1 convolutional, 2 pooling, and 3 fully connected layers. The input convolutional layer uses 36 filters with a kernel size 7 and ReLU as an activation function. The pooling includes 54 filters with a size of 5. The Adam optimizer is used to optimize the accuracy result of the model.

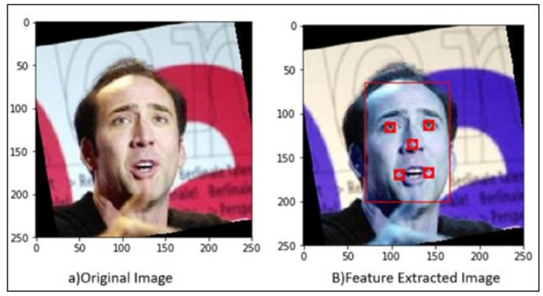

Figure -11. Feature Extraction Output

Once the data is classified into three phases, the pre-processing step is performed to reduce the over fitting problems. This process is also applied to minimize the noise of images in the input dataset. Fig.10 depicts the sample image before and after pre-processing. It is clear from the image pre-processing sample result that the image's dimensionality is increased. It will increase the model accuracy and maintain the same size as the input sample. Once the input images are pre-processed, similar facial images among the input samples are extracted using facial features such as eyes, nose and mouth. The sample and feature-extracted images using key points are shown in Fig.11.

The Fig.12 shows the accuracy of the result of the proposed CNN model in recognizing the input facial images.

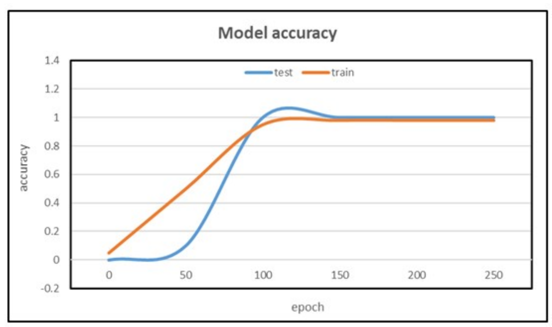

Figure-12. Accuracy Result

Generally, the accuracy of the model is evaluated based on the training sample result. The accuracy result is evaluated by experimenting with the model with different epochs from 0 to 250. It is clear from the analysis that the more increases, the more the accuracy rate increases. That is, from 0 to 50 epochs, the model achieves 15%, and while it reaches more than 50 epochs, it achieves more than 90% accuracy. Finally, the model achieves 93.75% accuracy when experimented with 250 epochs.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fig.13 shows the model's training and testing loss rate for recognizing the input facial images. The analysis results emphasize that the loss rate of the model decreases as the number of epochs increases. That is, as the number of epochs increases from 0 to 250 epochs, the loss rate decreases from 3.0 to 0.32.

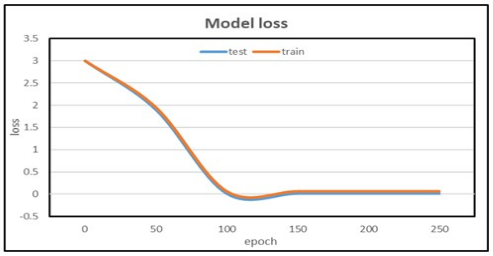

Figure-13. Training Loss Rate

4.2 CNN based Facial Feature Extraction and Recognition

Fig.14 shows the randomly selected images from the input dataset to perform the face recognition process. The face recognition is performed based on the face landmarks and face embedding. These randomly selected images are classified into two phases: training and testing. That is, 80% of the data are used for training and 20% for testing, as shown in Fig.15. Four images are selected from the randomly selected image, and a pre-processing step is performed to improve further model efficiency, accuracy, computational speed, noise reduction, dimensionality reduction, and resource allocation. For this, 100 input train and tested images are taken and processed. The pre-processed input image is depicted in Fig.14. The pr-processed input images are processed and saved into separate folders. The different pre-processed images under each category are shown in Fig.14.

Now, the popular deep learning-based face detection algorithm multi-task cascaded convolutional network (MTCNN) is applied to detect face and facial landmarks or features like eye, mouth, neck, and nose. The CNN model creates a bounding box around the region of the face, which is the initial step. The boxed images performed shows the selected two raw images and their feature-extracted images; after adjusting the bounding box using key points, features are extracted by the CNN model are shown in Fig.15 and Fig.16.

The embedding distance between the facial images is identified after extracting the facial features from the selected samples. The image of the same or different people is identified based on the embedding distance value. For this, two images are randomly selected, and the embedding distance between those two images is evaluated. Fig. 17 (a) (b) shows the resultant image of the embedding distance evaluated using Euclidean distance calculation on the same and different facial images

Fig.17 shows that the model has achieved a 0.6794 embedding value when comparing original face images 1 and 2. Fig.17(b) shows that when compared to original images 1 and 2, the model achieved a 1.3845 embedding distance value. It is clear from the result that the images compared in Fig.17(a) are both. Meanwhile, in Fig.17 (b), the images are dissimilar. The similarity between the images is predicted based on the lowest embedding value. Each image's actual and predicted result in the input dataset is evaluated based on this result.

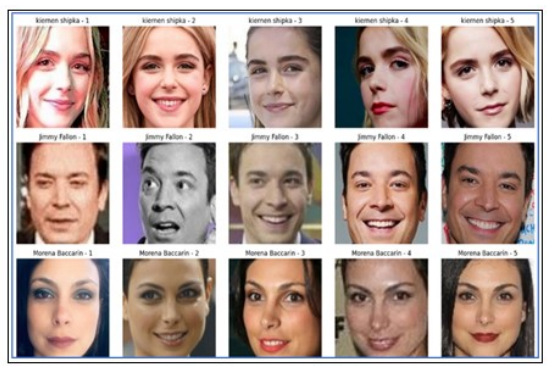

Figure 14. Preprocessing Image.

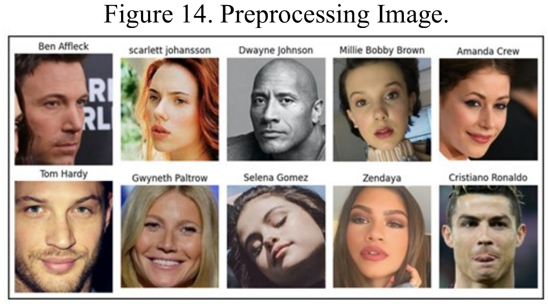

Figure 15. Input sample Images

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

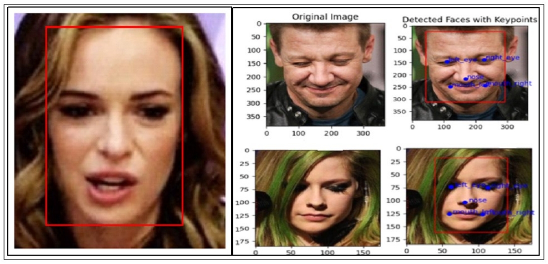

Figure-16. Bounding Box Using CNN and Feature Extracted Image

Figure-17. Predicted A Similar Image and Predicted Dissimilar Image.

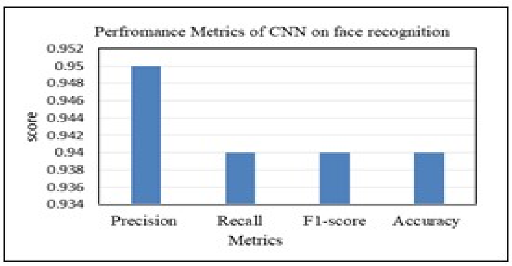

Figure-18. Performance Metrics Result

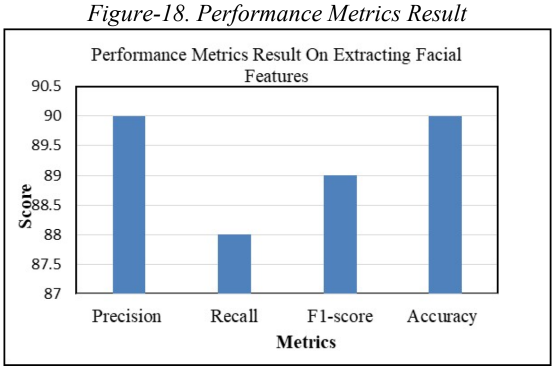

Figure-19. Performance Metrics Result on Extracting Facial Features

Some images with actual and predicted resultant values are shown in Fig.18, Fig.19, Fig.20 and Fig.21 represent the accuracy and loss rate of the model in predicting the similarity between the images in the input data samples.

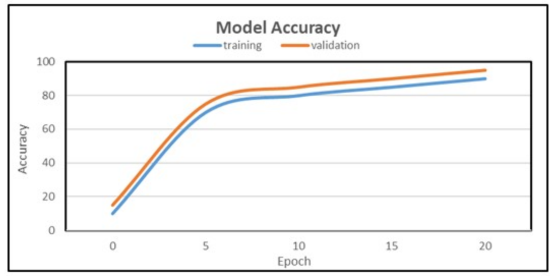

Figure-20. Model Accuracy

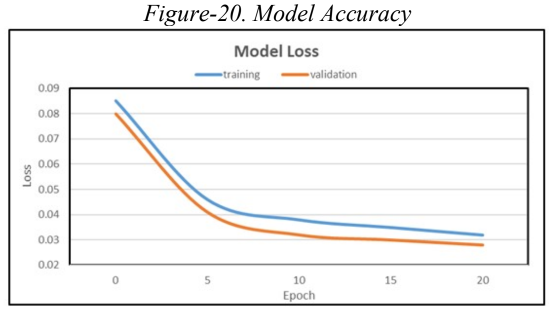

Figure-21. Model Loss

The accuracy result of the model indicates that the proposed model achieved 90% accuracy when performed with 20 epochs. And process the data with a loss rate of 0.035. These results show that the proposed model is more suitable for recognizing facial images.

Once the model's accuracy and loss rate are evaluated, its performance is evaluated using various performance metrics, such as precision, F1-score, Recall, and accuracy. These metrics are evaluated using flowing equations (1-4).

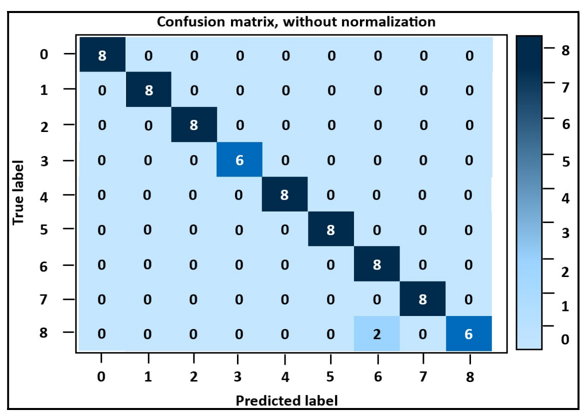

Figure-22. (a). Confusion Matrix Without Normalization

Fig.18 depicts the performance metrics result of the proposed CNN model on recognizing the input of similar facial images. The experimental result shows that the CNN model has achieved 0.95, 0.94, 0.94, and 0.93 in precision, F1-score, recall, and accuracy, respectively. Fig.19. graphically shows the analysis result of the proposed model on recognizing the facial image from facial features, which indicates that the model has achieved precision, recall, F1-score, and accuracy of 0.90, 0.88, 0.89, and 0.90, respectively.

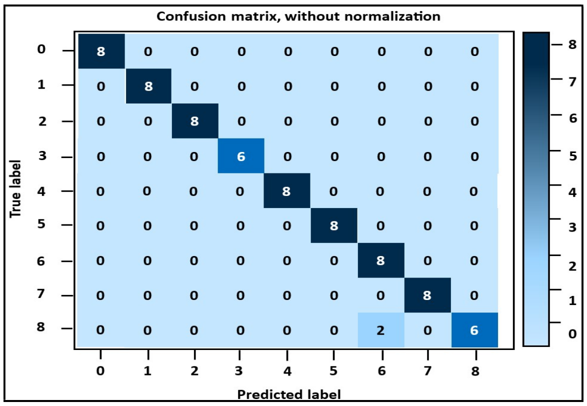

Figure-22 (b). Confusion Matrix with Normalization

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$Accurcay = \frac{TP + FP}{TP + TN + FP + FN}$$

After evaluating the performance metrics result of the model, the confusion matrix results with and without normalization are analysed and shown in Fig.22(a) and 22(b), respectively. The plotted confusion matrix result depicts the actual and predicted result of the proposed model on recognizing the input facial images. From the analysis, before normalization, the model achieved 93.75% accuracy, and after normalization, it achieved 94% accuracy. Fig.23(a) and Fig.23(b) shows the performance comparison result of the proposed and existing model [1] on recognizing the input facial image. The analysis shows that the existing models have achieved 89.28% and 92.7% accuracy, whereas the proposed model has achieved 94% accuracy.

Finally, the performance of the proposed CNN model is evaluated by comparing tis FPS measures is calculated from video input data and compare with the other similar models (Fig.24). FPS measures the number of times your graphical hardware redraws the screen every second. In the above package the CNN has the low range resolution of Frame Per Second. In the existing part compare to all the packages, the CNN give best result for the face detection. The CNN is optimized in terms of speed to achieve an execution time beyond real time (faster than 30 frames per second). CNN resulting high recognition rate because low range of resolution give the high speed of frames.

Figure-23(a). Performance Comparison of FTR.

Unlike earlier face recognition approaches that relied on handcrafted feature extraction techniques such as color histograms, texture descriptors, and local binary patterns—which often increased computational complexity and reduced robustness under varying conditions—this research introduces a two-stage face comparison model utilizing three distinct convolutional neural networks (CNN-1 for face detection, CNN-2 for facial feature detection, and CNN-3 for recognition).

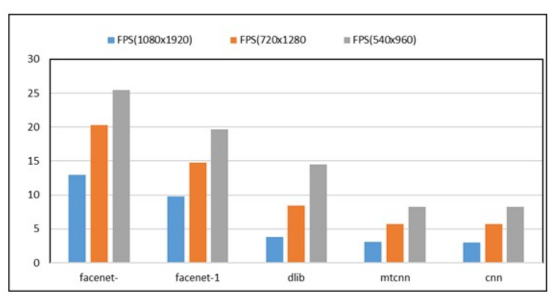

Figure-23(a). Performance Comparison of CNN.

While previous literature has shown the effectiveness of individual deep learning models in face detection or recognition tasks, the proposed method differs by integrating these components

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

into a cohesive, modular pipeline that improves both efficiency and accuracy. The model achieves a high accuracy of 99.68% in both face detection-based recognition (FDR) and facial feature detection-based recognition (FFDR), validated across multiple benchmark datasets, surpassing many traditional methods in consistency and reliability. A key advantage of this approach is its structured design that simplifies the recognition process while ensuring high precision; however, a potential limitation lies in the system's reliance on three separate CNNs, which may increase memory usage and computational load compared to lightweight or single-model solutions. Additionally, while the model performs well on benchmark datasets, further testing in real-world, unconstrained environments is necessary to fully assess its generalization capabilities.

5. CONCLUSION

The problem at the heart of this research was clear: despite the progress in face recognition technology, existing models continue to struggle with achieving consistently high accuracy under real-world conditions, while also facing challenges in computational efficiency and integration. This study addressed that gap by designing a two-stage face comparison model that not only separates but optimally coordinates the tasks of face detection, feature extraction, and recognition through three specialized CNNs. The experimental results speak directly to the problem statement—both the face detection-based recognition (FDR) and facial feature detection-based recognition (FFDR) achieved an equally remarkable accuracy of 99.68%, a figure that surpasses many conventional approaches and affirms the viability of modular, deep learning-based architectures in practical applications. These results are not just numbers; they validate the research hypothesis that a multi-stage CNN approach can deliver both high accuracy and structural efficiency in complex recognition scenarios. By grounding this model in benchmark dataset testing, this research not only proves its effectiveness but also offers a replicable and scalable framework for real-world implementation, particularly in high-stakes domains like criminal identification. Thus, this work not only reinforces the urgency and relevance of improving face recognition systems but also provides a data-driven, technically sound path forward.

REFERENCES:

[1] Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, 1-66.

[2] Juneja, K., & Rana, C. (2021). An extensive study on traditional-to-recent transformation on face recognition system. *Wireless Personal Communications*, *118*(4), 3075-3128.

[3] Amit, Y., Felzenszwalb, P., & Girshick, R. (2021). Object detection. In *Computer Vision: A Reference Guide* (pp. 875-883). Cham: Springer International Publishing.

[4] Murthy, C. B., Hashmi, M. F., Bokde, N. D., & Geem, Z. W. (2020). Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review. *Applied sciences*, *10*(9), 3280.

[5] Hasan, M. K., Ahsan, M. S., Newaz, S. S., & Lee, G. M. (2021). Human face detection techniques: A comprehensive review and future research directions. *Electronics*, *10*(19), 2354.

[6] Hatem, H., Beiji, Z., & Majeed, R. (2015). A survey of feature base methods for human face detection. *International Journal of Control and Automation*, *8*(5), 61-78.

[7] Hassner, T., Masi, I., Kim, J., Choi, J., Harel, S., Natarajan, P., & Medioni, G. (2016). Pooling faces: Template based face recognition with pooled face images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 59-67).

[8] Payal, P., & Goyani, M. M. (2020). A comprehensive study on face recognition: methods and challenges. *The Imaging Science Journal*, *68*(2), 114-127.

[9] Majeed, F., Khan, F. Z., Iqbal, M. J., & Nazir, M. (2021, July). Real-time surveillance system based on facial recognition using YOLOv5. In *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)* (pp. 1-6). IEEE.

[10] Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: past,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

present and future. Computer Vision and Image Understanding, 138, 1-24.

[11] Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). Wider face: A face detection benchmark. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5525-5533).

[12] Jiang, H., & Learned-Miller, E. (2017, May). Face detection with the faster R-CNN. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017) (pp. 650-657). IEEE.

[13] Kumar, A., Kaur, A., & Kumar, M. (2019). Face detection techniques: a review. Artificial Intelligence Review, 52, 927-948.

[14] Zhang, N., Luo, J., & Gao, W. (2020, September). Research on face detection technology based on MTCNN. In 2020 international conference on computer network, electronic and automation (ICCNEA) (pp. 154-158). IEEE.

[15] Ge, H., Dai, Y., Zhu, Z., & Wang, B. (2021). Robust face recognition based on multi-task convolutional neural network. *Math Biosci Eng*, *18*(5), 6638-6651.

[16] Borkar, N. R., & Kuwelkar, S. (2017, July). Real-time implementation of face recognition system. In *2017 international conference on computing methodologies and communication (ICCMC)* (pp. 249-255). IEEE.

[17] Hasan, M. K., Ahsan, M. S., Newaz, S. S., & Lee, G. M. (2021). Human face detection techniques: A comprehensive review and future research directions. Electronics, 10(19), 2354.

[18] Qi, D., Tan, W., Yao, Q., & Liu, J. (2022, October). YOLO5Face: Why reinventing a face detector. In European Conference on Computer Vision (pp. 228-244). Cham: Springer Nature Switzerland.

[19] Wu, H., Cao, Y., Wei, H., & Tian, Z. (2021, February). Face recognition based on Haar like and Euclidean distance. In *Journal of Physics: Conference Series* (Vol. 1813, No. 1, p. 012036). IOP Publishing.

[20] Saleem, S., Shiney, J., Shan, B. P., & Mishra, V. K. (2023). Face recognition using facial features. Materials Today: Proceedings, 80, 3857-3862.

[21] Adnan, S., Ali, F., & Abdulmunem, A. A. (2020, November). Facial feature extraction for face recognition. In Journal of Physics: Conference Series (Vol. 1664, No. 1, p. 012050). IOP Publishing.

[22] Rajan, A. P., & Mathew, A. R. (2019). Evaluation and applying feature extraction techniques for face detection and recognition. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *7*(4), 742-749.

[23] Upadhyay, V., & Kotak, D. (2020, January). A review on different facial feature extraction methods for face emotions recognition system. In 2020 Fourth International Conference on Inventive Systems and Control (ICISC) (pp. 15-19). IEEE.

[24] Serengil, S. I., & Ozpinar, A. (2020, October). Lightface: A hybrid deep face recognition framework. In 2020 innovations in intelligent systems and applications conference (ASYU) (pp. 1-5). IEEE.

[25] Jin, B., Cruz, L., & Gonçalves, N. (2020). Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis. *IEEE Access*, *8*, 123649-123661.