A Framework for Real-Time Recognition and Detection of Human Faces in Cctv Images

D. Gayathry[1], R. Latha[2]
[1]Research Scholar, Department of Computer Science, St. Peter's Institute of Higher Education and Research, Chennai - 600 054, Tamil Nadu, INDIA
gaya200689@gmail.com
[2]Professor & Head, Department of Computer Science & Applications, St. Peter's Institute of Higher Education and Research, Chennai - 600 054,INDIA
latharamavel@gmail.com

The goal of this work is to use deep learning and machine learning to develop a real-time framework for the identification and recognition of human faces in closed-circuit television (CCTV) images. A typical CCTV system requires constant human monitoring, which is costly and insufficient. The automated facial recognition technology in CCTV footage can help a lot of businesses, including law enforcement, by lowering expenses and requiring less human intervention to identify suspects, missing people, and anyone entering restricted areas. However, there are other issues with image-based recognition, such as scaling, rotation, and fluctuations in light intensity or busy backdrops. This paper suggests using a variety of face recognition and feature extraction techniques to build a human face recognition system based on CCTV photos.
Face detection, location, extraction from the captured photos, recognition, and image pre-processing are the steps that make up the proposed system. CCTV is used to obtain images. Convolutional neural networks (CNN) and principal component analysis (PCA) are the two feature extraction technologies used. A comparative analysis is conducted between the K-nearest neighbor (KNN), decision tree, random forest, and CNN algorithms. Recognition is achieved by applying these techniques to a dataset of more than 40K real-time images taken under different settings (e.g., rotation, light intensity, and scaling for simulation and performance evaluation). Ultimately, over 90% accuracy was attained, and the processing time needed to identify faces was minimal.
Keywords: KNN, CNN, Deep Learning,CCTV, Machine learning,

1. Introduction

Organizations nowadays must deal with serious security issues; in order to accomplish the necessary security, they require a number of people with specialized training. But people make blunders that compromise safety. Nowadays, closed-circuit television, or CCTV, is employed in many different contexts in daily life. An integrated intelligent control system has replaced basic passive monitoring with the development of video surveillance. Face detection and its novel uses in banking, safe access management, and other areas. The significance of biometric technologies, such as fingerprints, palm prints, and faces, has increased recently. Biometrics is now a commercially feasible technology thanks to developments in microelectronics and vision systems. One crucial component of biometrics is facial recognition. Human basics are matched to contemporary data in biometrics. An effective algorithm is used to extract and apply the face features, and certain modifications are done to enhance the current algorithm model. Computer-generated facial recognition can be used for many practical purposes, such as security systems, crime identification, and authentication. Generally, a facial recognition system consists of face detection phases, in which the input image's face is identified, and image processing procedures that prepare the face image for simple identification.

In this modern age, face recognition has become a necessity as the individual's identification increases daily with globalization. Since the last two decades, face recognition has received much attention because of its various applications, invaluable image analysis, and understanding domains. Face recognition is also becoming important in other fields like image processing, animation security human- computer interface , and medicine . Face recognition is natural, noninvasive, and easy to use. The face recognition system has a wide choice of applications in public safety, entertainment, attendance management, and financial payment. While today's facial recognition systems work researcherll in relatively controlled environments, they suffer from significant problems when used in existing surveillance systems due to image resolution, background clutter, lighting variations, and face and expression posture.

Face recognition systems consist of three steps, such as preprocessing of the image, feature extraction, and classification technique for recognition . Features extracted from the face, such as the mouth, nose, eyebrows, etc., are geometric features. The detected and processed face is compared to a database of known faces to determine who the person is. The surveillance system needs people to monitor it. Human monitoring involves reliability issues, scalability issues, and the inability to identify everyone.

Facial occlusions, such as beards and accessories (glasses, hats, and masks), involve evaluating facial recognition systems, making the subject diverse and challenging to function in a non-simulated environment. Another essential factor to consider is the different terminologies of the same distinct: macro and micro terminologies find their place on someone's face because of changes in an emotional state, and because of the many expressions of this type, effective recognition becomes difficult. A perfe ct face recognition system should be able to tolerate changes in lighting, expressions, poses, and occlusions and can scale for many users who need to capture the feresearcherst images simultaneously.

The overall contributions of the research paper can be summarized as follows:
(i) A machine learning-based framework for detecting and recognizing faces in CCTV images with various clutter backgrounds and occlusion
(ii) A dataset of 40K images with different environ- mental conditions, clutter backgrounds, and occlusion
(iii) Performance comparison of classical machine learning and deep learning algorithms for faces recognition in CCTV images the rest of the paper is organized as follows: Section 2 briefly introduces the related works. Section 3 explains the methodology, and the results are discussed in Section 4. Finally, researcher conclude the paper in Section 5.

2. Related Work

In this section, Researcher briefly introduce the related works about face detection and recognition using classical approaches and deep learning.

2.1 Face Detection Algorithms

2.1.1 Geometric Methods for Face Detection.

Early on in the history of computer vision, scientists experimented with a variety of methods to extract features from images and use geometric requirements to understand the features' supplies. A portion of this was caused by extremely little processing power. The earliest computers' ability to do computer vision was made feasible by the reduction of information from functionality extraction.

2.1.2 Template-Based Face Detection .

The majority of face detection techniques rely on models, directly encoding facial images based on pixel intensity. Most often, probabilistic models are utilized in conjunction with neural networks or other processes to characterize these facial picture datasets. These models' parameters can be manually or automatically changed based on sample photos.

2.1.3 Simple Templates.

These algorithms display misleading results if you are using a skin-based technique and there are areas of the image (such as hands and arms) where another skin color is present. Numerous scholars attempted to get around issue by integrating the results of skin color matching using straightforward models. These models range from correlation models for the regions of skin color and skin color (such as lips, hands, or eyes) to certain ovals connected to the image of the entrance's edge. However, these methods can also improve speed and strengthen the color-based detectors' resilience.

2.2 Face Recognition algorithms.

A technique that is currently being considered in artificial intelligence and machine learning is face recognition. In a lot of social security applications, it is crucial. The issue of facial recognition is now being researched through a wide range of studies and techniques. Combining genetic algorithms (GA), particle swarm optimization (PSO), and cat swarm optimization (CSO) was suggested by the researcher. Many others have been motivated to combine SVM, random transformation (RT), and higher-order spectral (HOS) techniques in a similar manner by this hybrid approach.

2.2.1 Iterative Closest Point-Based Alignment.

The aim of the alignment method is to convert the point cloud iteratively by identifying the translation and rotation parameters based on the nearest iterative point. When both point clouds are aligned, the mean square error of the clouds decreases. Thus, the lowest distance between point clouds is achieved by translating and rotating one point cloud relative to the others, as well as by measuring the distance between each point in the original point clouds every second and figuring out the average of all the distances. The alignment strategy based on the closest iterative point has a major drawback in that it requires the convergence trajectory to be initially aligned. Another drawback of this method is its high computational cost.

2.2.2 Simulated Annealing-Based Alignment.

It is a local research algorithm based on a stochastic process. In contrast to simulated annealing, researcheren hill-climbing has the ability to compute an even worse solution throughout the iteration process. It is more likely that you will find a solution because local minima do not impose constraints on simulated annealing. To define the transformation matrix utilized for an alignment between two 3D faces, simulated annealing requires six parameters, three for each translation and one for rotation reference to a 3D coordinate system. Using this method, facial image alignment occurs in three stages: initial level alignment, approximate level alignment, and final level alignment. First, the two-sided mass's center is being positioned. This method provides to minimize an approximation measure that combines the mean square error corresponding point of two faces that will be compared with the consensus of multiple estimators M (MSAC). Then, using the mean of a simulated annealing search technique that employs the measurement of the interpenetration of surfaces (SIM) as an estimation criterion, an accurate alignment is obtained. The drawback of simulated annealing alignment is that it requires additional calculation time, around the same as alignment based on the closest

2.2.3 Average-Based Face Model.

The medium-based face model is the foundation for this alignment. First, either manually or electronically, the reference points are on the face. To create a face model, the average of the crucial coordinates, followed by the Procrustes inspection and the converted milestone, are then mediated once more. Using an alignment on the closest iterative point, the probe face picture is aligned with the average model in this manner. The poor precision index and partial loss of spatial material during the development of the medium face model are noteworthy research weaknesses of the alignment based on the medium face model.

Pre-processing is the first stage of face recognition. When images are captured, transformed, converted, or compressed, they may experience a number of degradations. These degradations can occur in real-time video surveillance systems or images obtained from cameras. For example, low-resolution, noisy, and blurry photos have an impact on face recognition. These problems could make the facial recognition technique very difficult to use and reduce its effectiveness. Pre-processing is therefore a crucial stage in any face recognition system. Numerous statistical, convolutional, and color normalizing techniques are employed as preprocessing instruments. An additional significant issue with face identification via security cameras is the large number of captured images of an individual, which makes it expensive to run a face recognition algorithm on each one in terms of processing time and energy usage. An investigator presented

The most used method for processing images is PCA. Using the Euclidean distance method, image processing techniques were utilized to extract 16 face attributes with a ratio of area, angle, and distance, resulting in a performance of 75%. This served as the foundation for the initial proposal of the eigenface face recognition technique. Principal component analysis (PCA), an algorithm, is created as a result of this process. From that point on, PCA attracted a lot of interest and emerged as the most successful method for facial recognition. To achieve the best results, the PCA algorithm has undergone numerous changes. For feature extraction and face identification, the researchers employed PCA and Kernel-PCA, respectively. They investigate the non-linear kernel function to enhance PCA.

A CNN-based facial representation called Deep Hidden ID Entity feature (Deep ID) is proposed in. Deep ID gathers features from a group of small CNNs that are used for network fusion, in contrast to Deep Face, which gathers characteristics from a single giant CNN. In a similar vein, Researcherb Face, a face recognition pipeline, is presented in and use CNN to acquire the face representation. Over the past ten years, one of the most popular methods in computer vision has been the convolutional neural network (CNN), with uses in face recognition, object identification, and image categorization. Different approaches use the nearest neighbor (NN) classifier and its derivatives, such as PCA-based eigenfaces and LDA-based Fisher faces. Support vector machines and other supervised classifiers are used in face recognition systems.

3.Proposed Framework for Face Detection and Recognition in CCTV Images

As seen in Figure 1, the suggested approach consists of four important steps: (i) image acquisition, (ii) image enhancement, (iii) face detection, and (iv) face recognition. They used a variety of machine learning methods, such as convolutional neural networks (CNN), random forests, decision trees, and K-nearest neighbor (KNN), to achieve recognition.

3.1 Image Acquisition.

In this phase, researcher acquire an image.
Processing is not possible, hence the first stage in the workflow sequence is image restoration from the source camera, which is often a hardware source. The preprocessed input we use is the images that our CCTV continuously reads.

3.1.1 Camera Interfacing.

The Hikvision DS-2CD2T85FWD-15/18 IP camera is utilized for image acquisition. It has an 8-megapixel camera that records video at a resolution of $1248 * 720$ at 15 frames per second. First, the camera will take a picture, which a software program like MATLAB will then be used to preserve and access. The CCTV camera specifications utilized to acquire images are displayed in Table 2. The faces of individuals it will recognize are included in the face database. Because facial recognition uses algorithms for classification, every image in the collection has a label. Pictures of over 41,320 photos featuring 90 individuals. As a result, these classes' (people's) labels range from 1 to 90. It implies that there are several photos on each label. The description of the dataset is shown below. So, label 1 has 775 images approximately and same as others displayed in the figure (classes on the *x*-axis and number of images on the *y*-axis). Figure 2 shows the sample images in the dataset.

3.2 Preprocessing.

Preprocessing the image after it has been acquired gets it ready for handling more. Edge detection methods and grayscale conversion are the two primary preprocessing procedures.

3.2.1 Grayscale Conversion.

Researchers obtain the RGB image (R for red, G for green, and B for blue) from the camera. One red pixel paired with blue and green pixels makes up an RGB pixel. Since one pixel in an RGB image is equal to eight bits, the computation became more complex. Since every pixel in a grayscale image is a scalar, the image will have eight bits. Thus, the formula for converting RGB to grayscale is

$$\text{Grayscale} = 0.3 * \text{R} + 0.59 * \text{G} + 0.11 * \text{B}. \tag{1}$$

Here R, G, and B represent red, green, and blue pixels, respectively.

3.2.2. Canny Edge Detection. The Canny filter recognizes sharp color shifts in images to identify edges. This is being used by researchers to improve the image edges. Researchers can identify facial expressions with more precision the more benefits are enhanced. Gaussian and Sobel filters make up the filter. Initially, grayscale images are subjected to a Gaussian filter with a preset value of σ in order to facilitate edge detection.

$$G = \frac{1}{(2\pi\sigma^2)} e^{-(x^2+y^2)/2\sigma^2}. \tag{2}$$

In the second step, the Sobel filter is applied for finding the edges in the images. The filter used for finding the horizontal edges is

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}. \tag{3}$$

For horizontal edges, the filter is

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \tag{4}$$

The horizontal and vertical edges are calculated in order to find all the edges in the filter.

$$A = x = \sqrt{G_x^2 + G_y^2}. \tag{5}$$

The third and last step of the

canny edge detector, the hysteresis threshold, is applied to images containing the edges. The threshold is expressed as

$$H = \frac{1}{1 + e^{-x}}. \quad (6)$$

First, the lowest and maximum thresholds are chosen. A value of one is assigned to the pixel if its value above the threshold, and 0 is set for all values below the threshold. Another scenario is when the value stays constant and is equal to the threshold. To create the final improved image, the edges are lastly added to the original image. As a result, facial feature extraction and detection become simple, improving system efficiency as a whole.

3.3 Face Detection

After obtaining the image from the camera, the following stage involves using the Viola-Jones algorithm to identify faces and non-face regions in the images. After that, the facial region is removed for additional processing.

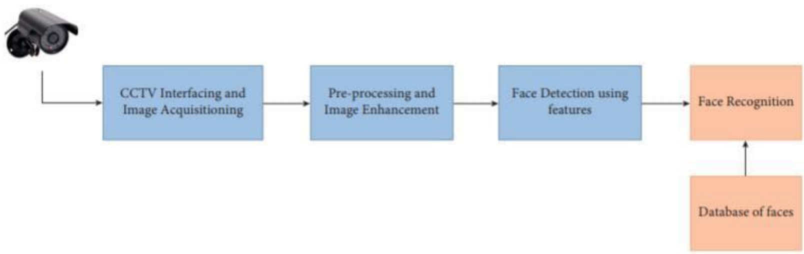

FIGURE 1: Process flow of the proposed system.

TABLE 2: Camera properties.

DS-2CD2T85FWD-15/18
Up to 8-megapixel high resolution Digital noise reduction
Day and night vision
Max. resolution 3840 × 2160

3.3.1Face Detection Using Viola–Jones Algorithm.

The first algorithm to offer competitive object detection rates in real-time is the Viola-Jones algorithm. Because it can process two frames per second, it offers resilience and high detection rates, making it simple for real-time applications. Following this, the image is recognized using several categorization algorithms. The following are some of the primary steps:

1. Integral image
2. Ada boost training
3. Cascading classifiers

3.3.2. ROI Extraction and Resizing.

The face detected by the Viola–Jones technique is extracted and resized as a 40 X 40 image, then used by various feature extraction techniques to find the features.

3.3.3 Features Extraction from Detected Face Images.

Researcher have used the principal component analysis (PCA) technique to extract features of the face in order to detect the face in later steps.

3.3.4 PCA-Based Facial Feature Extraction

PCA is a technique used to reduce the dimensions of the images in our dataset. It finds the characteristics of images, the difference and variance in pixels in one column from the other . PCA has the following steps as shown in Figure 3:

(1) Mean of each column

In this step, researcher have calculated the mean value of each column. The sum of the means of the columns are expressed as

$$\gamma_i = \sum_{i=1}^{n} \frac{a_{1i} + a_{2i} + a_{3i} + \cdots + a_{mi}}{m}. \qquad (7)$$

Here, c_i is the mean of i-th column.

(2) Covariance matrix

The second step is calculating the covariance of the matrix. The variance of the pixels is calculated as

FIGURE 2: Sample of face images used for recognition.

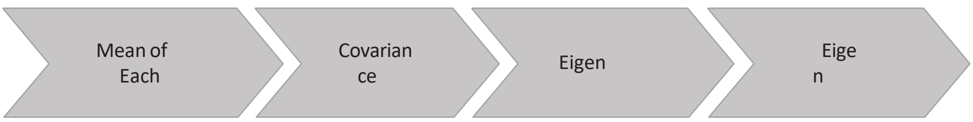

FIGURE 3: PCA steps for feature extraction.

$$cov(X_i, X_j) = \frac{1}{n} \sum_{k=1}^{m} (X_i^k - \gamma_i)(X_j^k - \gamma_i). \qquad (8)$$

In the above equation, *i* is the number of columns in the original image matrix, *j* is the second column in the image, and *k* is the number of rows. The fol- lowing equation shows the result.

$$\begin{bmatrix} cov(X_1, X_1) & cov(X_1, X_2) & \dots & cov(X_1, X_n) \\ cov(X_2, X_1) & cov(X_2, X_2) & \dots & cov(X_2, X_n) \\ \vdots & \vdots & \vdots & \\ cov(X_n, X_1) & cov(X_n, X_2) & \dots & cov(X_n, X_n) \end{bmatrix}. \qquad (9)$$

Eigenvalues
After the covariance matrix is calculated, the eigenvalues of the covariance matrix can be calculated by

$$\text{covariance} - c\boldsymbol{I}_n = 0. \quad (10)$$

Eigenvectors
Using the eigenvalues calculated in the previous step, researcher can find the eigenvectors from the following

$$\text{covariance} - c_i\boldsymbol{I}_i * X_i = 0. \quad (11)$$

Eigenvalues are the features of an extracted face. These values will be used for recognition.

3.5 Face Recognition Using Machine Learning Algorithms

3.5.1 Random Forest. This machine learning method addresses regression and classification issues. It employs ensemble learning, a method that combines numerous classifiers to solve complex problems. A random forest algorithm is made up of several decision trees. The "forest" generated by the random forest technique is trained via bootstrap aggregation or bagging. By combining them, a meta-algorithm called bagging improves accuracy.

3.5.2. Decision Tree. The decision tree is a nonparametric supervised learning method for both regression and classification. In order to build a model that predicts the value of a target variable, the goal is to understand fundamental decision rules from data features. It has a tree structure that resembles a flowchart, with each internal node denoting an attribute test, each branch designating the result, and each leaf node—or terminal node—carrying a class label.

3.5.3. K-Nearest Neighbor. Researchers have employed our features, which are 5, 10, and 15 eigenvectors. These vectors are used to generate the dataset, and the fresh face image will go through each PCA phase. The closest image in the dataset will be our prediction once the researcher has calculated its distance using the features of other images in the dataset. Because the Manhattan distance formula is more exact, researchers have been using it to determine distance. The formula for Manhattan distance is

$$D(Z, B) = \sum_{x=1}^{n} |z_x - b_x|. \quad (12)$$

Here, z is for the dataset, and b is for the test image. Then researcher will check which instance in the dataset has the minimum distance with the test image, which will be our prediction.

3.6 Face Recognition Using Convolutional Neural Network.

Convolutional neural networks consist of convolutional layers, pooling layers, and, at the end, a fully connected layer. A CNN has a much different architecture than a simple neural network. It has an input layer, a convolutional layer, a max-pooling layer, and at the end, a fully connected neural network as shown in Figure 4.

Researcher have used Adam optimizer for training in optimizing researcherights.

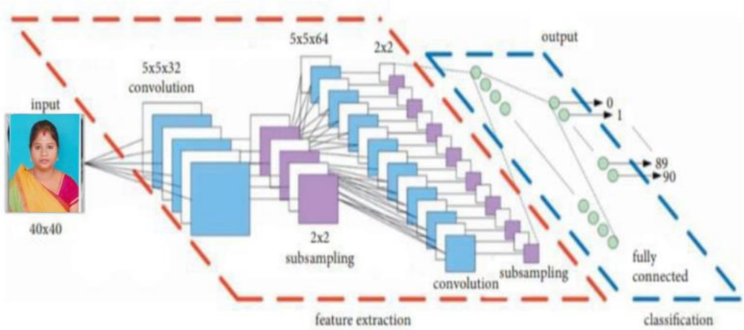

FIGURE 4: The architecture of CNN.

3.6. 1. Adam Optimizer

$$v_t = \beta_1 * v(t-1) - (1-\beta_1) * g_t,$$

$$s_t = \beta_2 * s(t-1) - (1-\beta_2) * g_t^2,$$

$$\Delta\omega_t = -\eta \frac{v_t}{\sqrt{s_t + \epsilon}} * g_t, \tag{13}$$

$$\omega_{t+1} = \omega_t + \Delta\omega_t,$$

where η: learning rate (0.001), g_t: gradient at time t, v_t: exponential average of the gradient, s_t: exponential average of the square of Gradient, and $\beta_{1,2}$: hyperparameters.

4.Results and Discussion

When Researcher apply PCA, Researcher get eigenvectors; these eigenvectors are our features. Researcher have used different features, such as Investigator have used 5, 10, and 15 eigenvectors.

4.1 K-Nearest Neighbour (KNN) Algorithm Results.

Table 3 displays the outcomes of simulating various values of k. With five eigenvectors, Figure 5 displays the obtained findings with a maximum accuracy of 94.7%. The accuracy reduced when the researcher increased the value of K. Researchers obtain about 95% accuracy for K 1 with Manhattan distance and 89% accuracy with Euclidean distance.

Ten coefficient PCA characteristics are displayed in Figure 6. The researcher achieved a maximum accuracy of 93.7% using Manhattan distance using 10 eigenvectors, and 87.6% with Euclidean distance. Then, when K increased in value, the accuracy dropped. Researchers have also found that the Manhattan distance outperforms the Euclidean distance in this instance. Furthermore, because the beginning eigenvectors exhibit maximal feature relevance, the accuracy likewise drops as the eigenvectors advance.

PCA features with fifteen coefficients are displayed in Figure 7. The same thing applies here: accuracy drops as features rise. Likewise, applying the value of k.

4.2. Decision Tree Result. For the decision tree, the results obtained for different features are given below, both in tabular form Table 4 and graphical form Figure 8.

4.3. CNN Results. As in CNN, researcher must train our dataset. Investigator have trained our data in 5000 steps and obtained 95.7% accuracy with only 30 images for testing and 30 for training.

4.3.1. With 50% Training and Testing Data. Researcher have obtained a maximum of 95.67% accuracy with 50% data of training and testing. Researcher trained it in 4000 steps. In some steps, the training steps, the accuracy increased, and at some points, it decreased, but at the end, researcher have obtained a maximum accuracy of 95.67%, accuracy as shown in Figure 10.

4.3.2. With 90% Training and 10% Testing Data. Now researcher have obtained 95% accuracy in this section, maybe because testing data is much less than training. And researcher have obtained this accuracy in 300 steps, as shown in graph Figure 11.

4.3.3. With 80% Training and 20% Testing Data. Now researcher have obtained 97.5% accuracy in this section, which may be because testing data is much less than training data. And researcher have trained data in 5000 steps, as shown in the graph Figure 12.

Table 3: Results for KNN.

No. of features	Training data	Numerical methods	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
5	90	Euclidean	89.0115%	89.7889%	79.0841%	76.5861%	75.278%
		Manhattan	94.7623%	90.0457%	88.6113%	86.9326%	86.0086%
	80	Euclidean	87.8664%	80.2338%	77.7842%	75.2137%	73.5403%
		Manhattan	93.7989%	89.0839%	87.6401%	85.9456%	84.8975%
10	90	Euclidean	88.3589%	79.7717%	77.8163%	76.1214%	75.0927%
		Manhattan	93.7989%	89.4494%	88.4072%	87.1582%	77.0927%
	80	Euclidean	86.8185%	77.905%	75.9567%	74.377%	73.4407%
		Manhattan	93.6811%	88.3288%	87.335%	86.0392%	85.3475%
15	90	Euclidean	86.383%	76.6452%	74.7327%	73.293%	72.5651%
		Manhattan	93.9484%	88.2299%	87.3221%	86.1644%	85.618%
	80	Euclidean	84.5773%	74.3589%	72.5164%	71.1934%	70.4926%
		Manhattan	92.8172%	86.6614%	85.9425%	84.7646%	84.1484%

Performance Comparison of K Nearest Neighbor with cross validation folds 10 with 5 Eigen vectors

Figure 5: Comparison of KNN results for 5 eigenvalues.

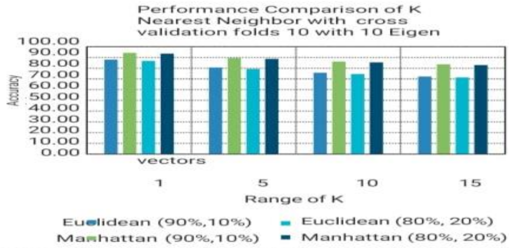

FIGURE 6: Comparison of KNN results for 10 eigenvalues.

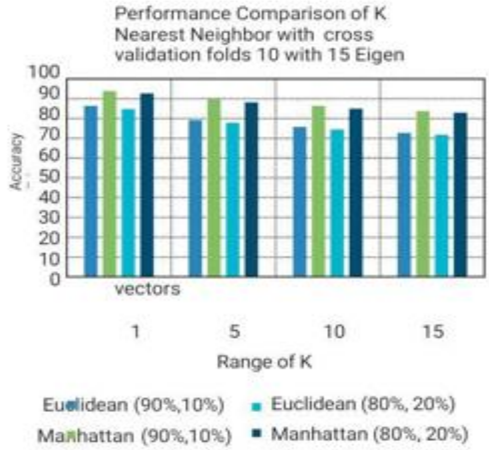

FIGURE 7: Comparison of KNN results for 15 eigenvalues.

TABLE 4: Results for decision tree.

No. of features	Training data	Testing data	Accuracy
5	90%	10%	70.34%
	80%	20%	68.75%
10	90%	10%	68.88%
	80%	20%	68.39%
15	90%	10%	68.64%
	80%	20%	68.28%

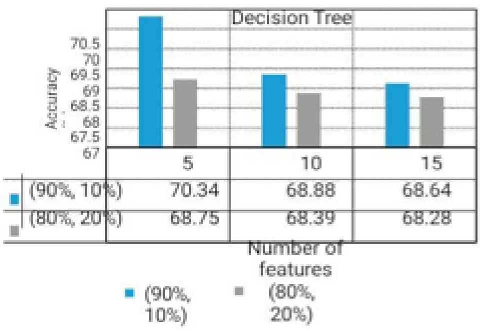

FIGURE 8: Comparison of decision tree results.

Performance Comparison of K Nearest Neighbor with cross validation folds 10 with 15 Eigen vectors

TABlE 5: Results for random forest.

No. of features	Training data	Testing data	Accuracy
5	90%	10%	93.20%
	80%	20%	92.65%
10	90%	10%	91.38%
	80%	20%	90.71%
5	90%	10%	89.95%
	80%	20%	88.60%

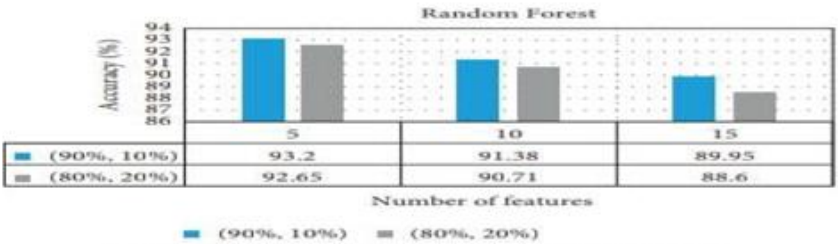

FIGURE 9: Comparison of random forest results.

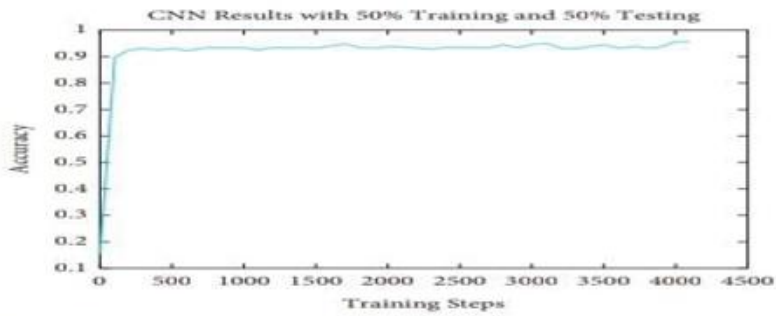

FIGURE 10: Results of 50% training and 50% testing data using CNN.

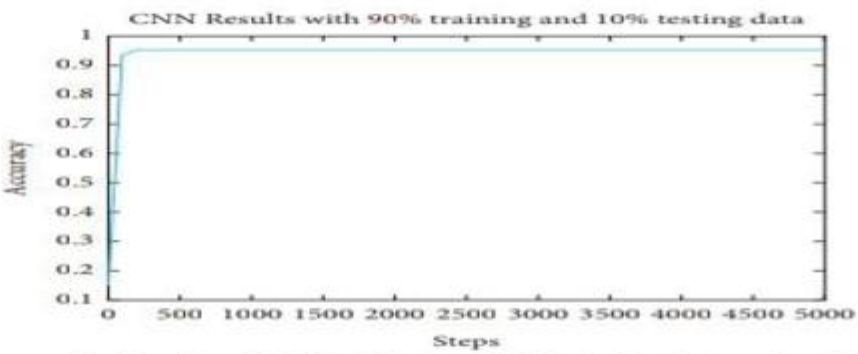

FIGURE 11: Results of 90% training and 10% testing data using CNN.

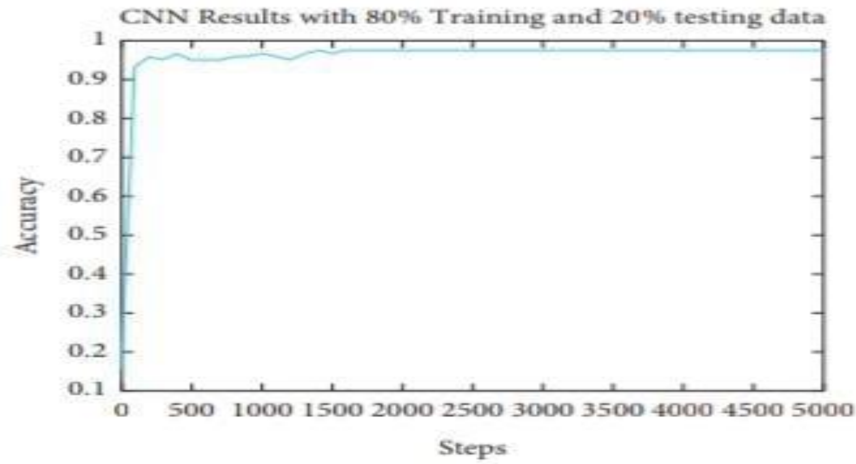

FIGURE 12: Results of 80% training and 20% testing data using CNN.

5. Conclusion

In this work, a framework for automatic face identification based on CCTV photos is developed using several machine learning methods. In order to get the best recognition accuracy, the work's goal is to gather over 40,000 face photos and evaluate how well different algorithms perform. Various algorithms have been employed to achieve superior accuracy for CNN. When using DT, RF, and KNN, CNN is far more dependable than PCA. KNN is a sluggish method; it looks for predictions in every instance in the dataset, whereas CNN quickly recognizes patterns from its model. For PCA, 41,320 photos were utilized for 90 classes, while for CNN, researchers employed 10 classes with 30 images each, and they produced good results.

References

1. F. Schroff, D. Kalenichenko, and J. Philbin,"Facenet: A unifiedembeddingforfacerecognitionandclustering,"in *Proceedings oftheIEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 815-823.
2. Y. Sun, D. Liang,X. Wang, andX.Tang,"Deepid3: Face recogniith very deep neural networks,"*arXiv preprint arXiv:1502.00873*, 2020.
3. Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface:Closingthegaptohuman-levelperformancein face verification," in*ProceedingsoftheIEEEconference on computer vision and pattern recognition*, 2020, pp. 1701- 1708.
4. P. Viola and M. J. Jones,"Robust real-time face detection,"*International journal of*
5. *computer vision,*vol. 57, no. 2, pp. 137-154, 2021