Ingénierie des Systèmes d'Information
Vol. 28, No. 4, August, 2023, pp. 877-885
Journal homepage: http://iieta.org/journals/isi

Automated Human Recognition in Surveillance Systems: An Ensemble Learning Approach for Enhanced Face Recognition

Bhavana Kanawade[1*], Jyoti Surve[1], Shraddha R. Khonde[2], Shilpa P. Khedkar[2], Jayshree R. Pansare[2], Bhavini Patil[1], Sharayu Pisal[1], Anushka Deshpande[1]

[1] Department of Information Technology, International Institute of Information Technology, S.P. Pune University, Pune 411057, India
[2] Department of Computer Engineering, M. E. S. College of Engineering, Savitribai Phule Pune University, Pune 411007, Maharashtra, India

Corresponding Author Email: bhavanak@isquareit.edu.in

https://doi.org/10.18280/isi.280409

Received: 1 May 2023
Revised: 26 July 2023
Accepted: 17 August 2023
Available online: 31 August 2023

Keywords:
face recognition surveillance system, face recognition & verification, ensemble learning, FaceNet, FaceNet-512, VGGFace, Dlib, ArcFace

ABSTRACT

In the realm of surveillance, closed-circuit television (CCTV) cameras serve as a vigilant watch over unfamiliar entities. However, the unpredictability of such entities necessitates continuous human monitoring, an endeavor prone to error and demanding of significant resources. The automation of this process through face recognition could alleviate these burdens, provided the system delivers high precision and rapid judgment capabilities. This study presents a novel solution to these challenges: an automated human recognition and verification surveillance system, founded on a max-voting ensemble method. This innovative approach amalgamates five influential feature extraction models: VGGFace, FaceNet, FaceNet-512, Dlib, and Arcface, with a support vector machine deployed for classification. The proposed system was subjected to rigorous testing on the AT&T, faces94, Grimace, Georgia Tech, and FaceScrub datasets, demonstrating an impressive accuracy of 100% on the AT&T, faces94, and Grimace datasets, and 99.3% and 98% on the Georgia Tech and FaceScrub datasets, respectively. The system's performance was further enhanced through a re-verification technique, which facilitated swift and precise prediction of unknown entities in real time. This study thus contributes a significant advancement to the field of automated surveillance, offering a potent tool for efficient, accurate human recognition.

1. INTRODUCTION

Closed-Circuit Television (CCTV) has been widely adopted as an efficient security monitoring system, enabling the continuous surveillance of properties and institutions. The ability to monitor unknown individuals in various settings, such as homes and offices, is a key benefit. However, the necessity for constant human monitoring of CCTV footage, coupled with the delay in producing actionable results, means that the system may fail to promptly identify unknown individuals. This methodology is not only time-consuming, but may also compromise security. An urgent need is thus identified for an automated facial recognition system that can instantly notify administrators upon detection of unknown individuals, thereby facilitating immediate intervention and reducing potential future losses.

This paper introduces an automated facial surveillance system, built upon a max-voting ensemble learning method. The ensemble model integrates five established face recognition models: VGGFace [1], FaceNet [2], FaceNet-512 [3], Dlib [4], and ArcFace [5]. These models are employed for feature extraction, whilst the Support Vector Machine (SVM) is utilized for classification purposes. The Haar Cascade method is employed to detect the face within the input image, and five pre-trained models are subsequently applied to calculate the embeddings of the detected face, producing five face embeddings. The SVM classification model is then utilized to predict the class label for the corresponding embedding. Following this process, five class labels are generated, and a max-voting ensemble approach is applied: the class label receiving the maximum votes is deemed the final predicted class label.

A significant challenge within face recognition and verification systems is the detection of unknown individuals. Traditional approaches involve matching the face of an unknown individual with each class image, a method that becomes increasingly time-consuming with larger datasets. To address this challenge, and to enhance the accuracy of the system, a re-verification method is proposed, also utilizing the max-voting ensemble technique. A validation folder, containing 10 images per class, is maintained for re-verification purposes. Each model undergoes the following voting procedure to verify the identity of an individual:

(1) The input face image is matched with 10 images from the validation folder of the predicted class.

(2) If the input face image does not match 6 or more of these 10 images, the individual is voted as 'unknown'; otherwise, the individual is voted as 'known'.

The max-voting ensemble approach is then applied. If the majority of the models vote 'unknown', the individual is classified as such. If classified as 'known', the predicted class label is displayed as output. This re-verification method

enhances the speed and accuracy of the system by confirming the identity of the individual. Developed using Keras [6] and TensorFlow [7] with Python as the programming language, the system offers the following key contributions:

(1) A facial surveillance system, based on ensemble learning, incorporating five of the most popular state-of-the-art real-time face recognition models to ensure optimal accuracy, with SVM as the classifier.

(2) A re-verification method using the max-voting ensemble technique to confirm the identity of the individual, providing fast and accurate results in real-time.

2. RELATED WORK

This part focuses on a comprehensive literature review that includes many elements of current research areas. Face recognition techniques are the subject of the survey. Wang et al. [8] proposed a novel method where the VGGFACE algorithm was fine-tuned. The system was tested on a dataset that was captured in the real time from real-world surveillance videos and the vgg face algorithm was fine-tuned. The presented approach was 92.1 percent accurate, while the original VGG face was only 83.6 percent accurate. Mattmann and Zhang [9] rebuilt the VGGFace deep learning facial recognition network. On a 4 celeb, 128 celeb, and 2,622 celeb use case, the proposed method provided validation accuracy of 97 percent, 68 percent, and 78 percent, respectively, and approximately 78 percent validation accuracy after optimization using warm-up strategy and learning rate linear scaling on the large 2,622 celeb dataset. The usage of triplet-loss was eliminated in this technique. Oliveira et al. [10] built FaceBank, a dataset of 27,002 authentic images collected from the databases of Brazil's largest public bank, and presented an architecture for cross domain face matching, comparing selfies and IDs. The VGG-Face and OpenFace CNN models were used to normalize the selfies and IDs before extraction and normalization of their deep feature vectors. PmSVM, Linear SVM, Voting RF and RF results were compared to see which one performed better in categorizing a pair of face data (selfie and ID) as genuine or impostor. When tested, OpenFace scored around 92 percent in the LFW benchmark, whereas VGG-Face scored 98 percent. Moustafa et al. [11] developed a system which comprises an image preprocessing phase followed by feature extraction using a pre-trained VGG-Face CNN to get highly discriminative descriptors. After that, a dimension reduction utilizing an efficient MDCA fusion decreased the input feature space greatly and increased the system's recognition rate. The suggested approach has 81.5 percent accuracy for the tough FGNET dataset which consists of face images of people of ages zero to forty-five and 96.5 percent for the MORPH (album-II) dataset Sepas-Moghaddam et al. [12] suggested a method in which the face region of all raw light pictures used to generate the sub-aperture array was clipped and the features retrieved with the pre-trained VGGFace descriptor model. The VGG Very Deep 16 CNN was retrained for the disparity and depth maps retrieved from the light field multi view sub aperture array to fine tune the system. The VGGFace descriptor collected characteristics from three different sorts of data inputs and after completion of all the models, they were concatenated and fed to an SVM classifier. The accuracy of the suggested method is 98.1 percent.

The pre-trained VGGFace model's performance for face verification was inadequate for the LFW and FRGC datasets, according to Lu et al. [13]. They proposed combining non-CNN characteristics with picture representations learnt by CNNs to solve this challenge. When evaluated on the LFW dataset, for the images of size 3232, 6464 and 140140, the proposed method gave an accuracy of 86.85 percent, 92.35 percent and 97.45 percent respectively, whereas, on the FRGC dataset, for images of sizes 3232, 6464 and 211201, it gave an accuracy of 83.32 percent, 90.11 percent and 96 percent respectively.

Astawa et al. [14] changed the pre-trained VGGFace's last three layers, or classification portion. The model produced extremely accurate results with minimal loss. Furthermore, the picture data for training comes from three different places. The best image source, based on the three image sources, is the digital camera, which has an accuracy of 94.69 percent, a loss of 10.41 percent, and a validation accuracy of 99.84 percent. Kumar et al. [15] demonstrated a transfer learning-based system for face identification and verification that needs minimum re-training. For transfer learning, authors have employed the AT&T face database, Essex 94, Essex 95, Essex 96, Essex Grimace, and Georgia Tech databases. Results stated that overall accuracy of suggested model is 96.5% on AT&T dataset, 99.09% Essex94 dataset, 97.43% on Essex95 dataset, 95.7% on Essex96 dataset, 99.25% on grimace dataset, and 96.61% on Georgia Tech dataset. Jose et al. [16] described the development of a smart face recognition surveillance system with several cameras on the Jetson TX2 utilizing FaceNet and the MTCNN algorithm. Using installations of several cameras, this portable system uses the camera position or ID, as well as the timestamp, to track the subject or suspect and reports his existence in the database. This independent system finds the individual who was already assigned to track in the dataset, and an embedding being generated was recognized successfully with a 97 percent accuracy. Manna et al. [17] developed a facial recognition method that makes finding criminals quicker and faster, saving the time of the administration and the police. Face recognition from video is accomplished using FaceNet, a pre-trained model (FN). This model has the benefit of distinguishing between the blurred image & side face that other models cannot. After training with a specific dataset, the FaceNet model has the highest accuracy of any of these models. Recognizing this fact, the dataset was collected and FaceNet was applied to it, resulting in an accuracy of 90%. Anitha et al. [18] presented a system in which faces are detected by using MTCNN i.e., Multi Task Cascaded Neural Network algorithm and individuals are identified by using FaceNet. This methodology is designed to provide a high level of security while reducing manual errors. It updates and prepares an attendance sheet after the facial recognition procedure and sends the report to the appropriate departments and staff members via mail. FaceNet employs the triple loss function and enhances the network with an embedded layer extraction feature. FaceNet is used in the suggested system due to its excellent accuracy. Nyein and Oo [19] presented a methodology for improving the accuracy of multi face identification by using SVM and FaceNet. FaceNet is used to extract features, and SVM is used for the classification of the given training data using the FaceNet derived features. For training and testing, they used private face data sets of roughly 80 people. According to the results, the suggested method is capable of multi-face recognition with a 99.6% accuracy. On the same data set, it outperforms the VGG16 model. Arsenovic et al. [20] studied a novel face

identification attendance system which is based on deep learning. The goal of this research was to apply cutting-edge deep learning algorithms to facial recognition problems. This model is composed of many essential phases that were created using the most cutting-edge methods available at the time: (a) face detection using CNN cascade & (b) face embeddings generation using CNN. The model was trained using the suggested augmentation strategy and a small number of photos per employee. This resulted in the initial dataset being expanded and the overall accuracy being improved. The system has 95.02 percent accuracy on a small collection of original employees facial images in a real time context.

In a notable work, Suguna et al. [21] developed a face recognition model that leverages the combination of FaceNet and Support Vector Machine (SVM) for the extraction and categorization of facial embedding features, respectively. This model incorporates an MTCCN mechanism for the detection of 5-point landmarks, and it utilizes a linear SVM for the classification and recognition of faces. Impressively, the model demonstrated an accuracy of 99.85% in recognising faces that were oriented straight or slightly rotated.

A different approach was taken by Sanchez-Moreno et al. [22], who proposed a real-time face recognition system capable of functioning in an unconstrained environment. This system incorporated a one-stage Deep Neural Network (DNN) methodology that combines FaceNet and classifiers such as Random Forests (RF), SVM, and K-Nearest Neighbors (KNN). The system effectively employed a YOLOv3-based YOLO-Face detector for facial detection, and combined FaceNet with SVM for classification tasks, achieving a remarkable 99.7% accuracy.

A unique "two-tier authentication" concept was introduced by Mehta et al. [23]. This concept was designed to enhance system accuracy and incorporate a time allowance mechanism for students. The system, which exhibits an overall accuracy of 93.33%, utilizes FaceNet for the creation of face embeddings.

Further, a smart classroom attendance management system was proposed by Seelam et al. [24], which merges computer vision and deep learning methodologies. Built on a Raspberry Pi, the system utilized a facial detection method for tracking attendance, followed by facial identification. The system achieved a facial recognition accuracy of 98% when the dataset was split into training and testing sets in an 80:20 ratio.

Subsequently, a real-time attendance system based on facial identification was developed by Kuang and Baul [25], using pre-trained deep neural networks. The system employed a pre-trained Haar Cascade model to recognize faces in webcam video and generated 128-dimensional face embeddings using FaceNet. The system demonstrated a face recognition accuracy of about 95% in a consistent image acquisition scenario for a class of 28 students.

The exploration of data augmentation techniques was the focus of D'Silva et al. [26], who developed a facial recognition attendance system based on deep learning. This study established the superiority of training a model with an enriched dataset. Two FaceNet models, the 2017 and 2018 variants, were trained and tested, with the 2018 model demonstrating superior performance.

Sikarwar et al. [27] proposed an automatic and interactive biometric verification application, which also offers attendance record management and visualization features. The FaceNet was employed to create a corresponding 512-dimensional embedding for the face identified by MTCNN. The accuracy on the Faces94 dataset without image enhancement was 98.19 percent.

Chanda et al. [28] introduced a novel hybrid approach for handling face identification tasks in the one-shot learning framework. This approach combined features extracted using the ResNet architecture from Dlib with a Siamese-Network classifier. The hybrid network performed admirably, especially on 5-way and 50-way one-shot tasks, achieving an accuracy of more than 90% and 84% respectively.

In an innovative application, Xia and Li [29] proposed a facial identification model for cinema and television. All faces in a scene were used as input for generating the feature vector of the face and as a training set for a given scene. The system achieved a high recognition rate of 95.7% for white individuals, with somewhat lower rates for black and yellow individuals, but overall, the accuracy was greater than 85%.

Marsi et al. [30] implemented a facial recognition system on an Odroid XU-4 platform, which can recognize faces up to 4 m at a speed of 1.8 seconds per frame. Despite training the classifier on very few samples, a confidence assessment allowed false recognitions to be minimized and true positives to be maximized.

In a recent study, Su et al. [31] proposed a dual-channel image-based method. The original VGG-like model, the ResNet called by Dlib, MobileFaceNet, and the new VGG-cut model were compared on a RISC-V SoC to emulate an embedded environment. For a model of size 99.72MB, the Dlib algorithm exhibited a training accuracy of 99.2% and a testing accuracy of 99.1%.

Chinapas et al. [32] introduced a system of personal verification based on an ID card and a face image, employing facial detection and facial comparison by utilizing three widely acknowledged methods: Dlib, Facenet, and ArcFace. The testing results demonstrated that the system with ArcFace performed the best, with an accuracy of 99.06 percent for face detection and 96.09 percent for face comparison, since it straightens the facial picture and compares important facial traits better than the other approaches.
Son et al. [33] presented the architecture for developing an automated attendance system that uses CCTV Camera. They compared the facial feature representations of models, Arcface and Facenet, and opted to employ ArcFace as a feature extractor due to its superior performance. On the test dataset, they achieved an accuracy of 91.3 percent. Guo and Nie [34] proposed a face recognition system that is primarily dependent on RetinaFace for face detection and alignment, and it employs a lightweight system with good detection on small faces and good real-time performance for complicated surveillance situations. Deep residual neural network is paired with ArcFace loss for feature extraction purpose. The face recognition system achieved good real-time performance, accuracy, and resilience, as per the testing results. Jiao et al. [35] proposed a Dynamic Additive Angular Margin Loss Function for Deep Facial Recognition (Dyn-arcFace) in which the fixed additive angular margin is transformed into a dynamic one. The overfitting produced by the fixed additive angular margin is decreased by Dyn-arcFace. The experimental findings demonstrated that the suggested loss function outperformed ArcFace on various benchmarks, confirming the efficacy and resilience of the proposed strategy. Based on ArcFace, Li et al. [36] presented a unique additive margin loss function called Li-ArcFace for deep face recognition. It shows good performance and convergence when using embedding feature learning which is low-

dimensional. On various face verification datasets, it obtained state of the art results.

3. ALGORITHMS USED FOR ENSEMBLE

3.1 VGGFace

VGG-Face [1] is an implementation of the extremely deep ConvNet architecture VGG-16 created at Oxford University's Visual Geometry Group (VGG). The database utilized is made up of up to a thousand instances of each subject and is trained on a collection of 2.6 million face photos and 2622 unique identities. It takes input of size 224×224.

3.2 FaceNet

FaceNet [2] model requires 160×160 RGB images to represent facial images as 128-dimensional vectors. The face embedding is achieved using a batch input layer, a deep CNN, and L2 normalization which during training is followed by the triplet loss.

3.3 FaceNet-512

David Sandberg [3] released an expanded version of Facenet, which generates 512 dimensions. On the LFW data set, he achieved 99.60 percent accuracy.

3.4 Dlib

Dlib [4] is based on a ResNet-34 model [37]. The usual ResNet structure was changed by removing a few layers and rebuilding a neural-network with 29 convolution layers. It requires 150×150×3 input to represent facial pictures as 128 dimensional vectors. The model was then put to the test on the LFW (labeled faces in the wild) data set, which is widely used as a benchmark in face recognition research and achieved 99.38% accuracy. It was trained from the ground up on a dataset of around 3 million faces. The dataset was created by combining many datasets. The FaceScrub dataset [38], the VGG dataset [1] and a huge number of photos from the internet were used. The collection has 7485 unique individual identities. There was no overlap with the Labeled Faces in the Wild (LFW) dataset.

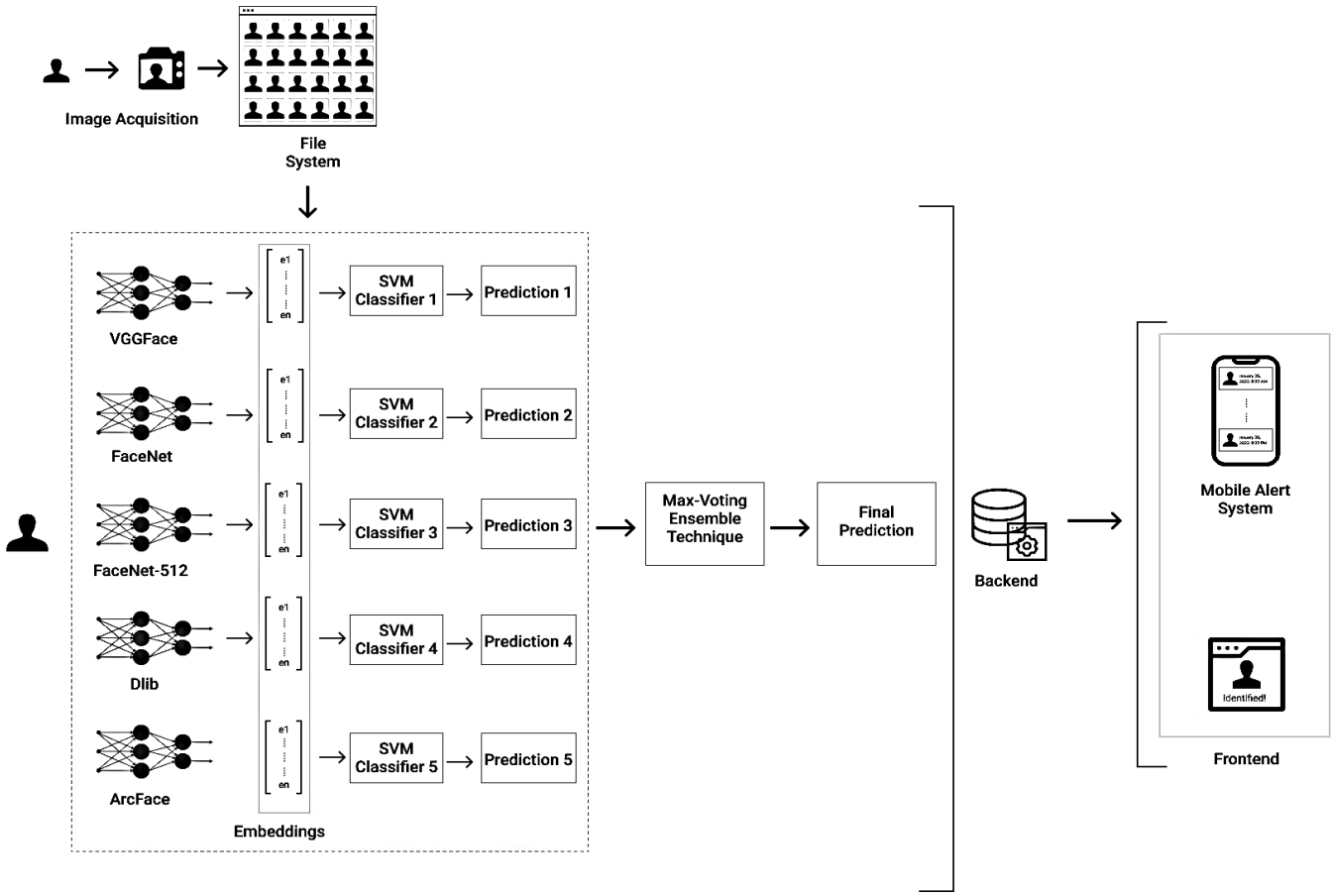


Figure 1. Architecture diagram of face recognition surveillance system using ensemble learning

3.5 ArcFace

ArcFace [5] model requires (112, 112, 3) shaped inputs and generates 512-dimensional vector representations. On the LFW data set, the original research scored 99.83 percent accuracy, whereas Keras re-implementation scored 99.40 percent accuracy.

4. PROPOSED METHOD

The Figure 1 depicts the proposed system's detailed design.

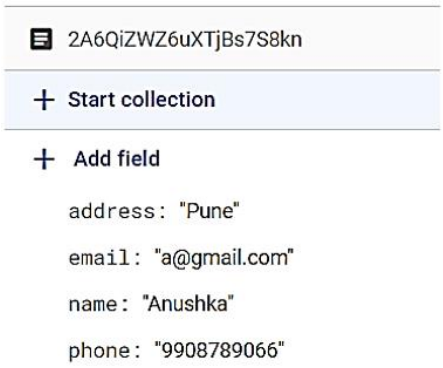
4.1 Image acquisition

This section consists of two modules:

4.1.1 Registration

Before capturing the images, users are required to get themselves registered by filling the registration form as shown in Figure 2 (a). The entered information is stored on Google Firebase's Firestore as shown in Figure 2 (b) and the directories of the user's name are created under the train, test and validation directories in the file system.

(a) Registration form



(b) Google firebase's firestore

Figure 2. Registration form and firebase information view

4.1.2 Face detection and preprocessing

For detecting the face of the individual Haar Cascade classifier is used. After detecting the face, the face image is resized to 400×400. The captured images are stored in the train, test and validation directories under the user's name in the file system. 80-20 rule is followed while storing images in the train and test directories and additionally, 10 images are captured and stored for validation purposes. The number of validation images can be varied as per the convenience of the admin.

4.2 Feature extraction

4.2.1 Data preparation

All the images in the captured dataset are resized as per the input size of each pre-trained model. The processed training and testing data and the corresponding labels are stored together in.npz file format. The following Table 1 shows the standard input size for each pretrained model.

4.2.2 Generation of embeddings

The training and testing data saved in.npz format is loaded and the corresponding pretrained model in .h5 format (downloaded from the internet) is loaded. The embeddings of training and testing data are calculated through the corresponding model and are saved in.npz file format.

Table 1. Input image size for pretrained models

Model	Input Size
VGGFace	224 × 224
Facenet	160 ×160
Facenet-512	160 × 160
Arcface	112 × 112
Dlib	150 × 150

4.3 Classification

After loading the embeddings in NPZ format (refer section 4.2.2) of the corresponding model, the performance of the model is evaluated by using a support vector machine as a classifier for each model. All the five generated fitted models are saved in .sav format.

4.4 Real time working

In real time when the person comes in front of the camera, Haar Cascade is used to detect the face of the person. After detection of face, all five pre-trained models in .h5 format are loaded. For each model following procedure is followed:

(1) Calculate the embeddings of the detected face.

(2) Predict class label by using the classification model in .sav format (refer section 4.3) for the corresponding embedding.

Then the max-voting ensemble approach is applied on the five predicted class labels, where the class label with the maximum votes is the final predicted class label.

Re-verification is performed to confirm the identity of the individual.

4.4.1 Reverification

For re-verification purposes, we have maintained a validation folder which contains 10 images per class.

Once prediction of the class label (refer section 4.4) is done using max-voting ensemble technique, for each model following procedure is followed:

(1) Find the embeddings of the input face image for each model.

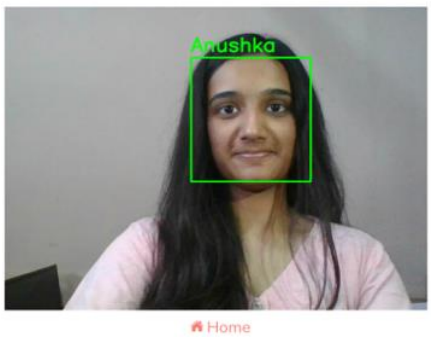
(2) Using cosine similarity [39], the embeddings of the input face image are compared to the 10 validation images embeddings of the predicted class label from the validation folder. While comparing the embeddings, if the estimated cosine distance exceeds the threshold value, the input face image is considered as 'unknown' in relation to the validation image.

(3) If 6 or more out of 10 predictions are 'unknown', the corresponding model's final vote is 'unknown'.

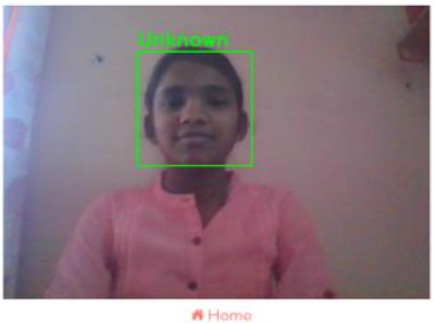
Then the max-voting ensemble approach is applied on the final vote of each model. If the majority of the models out of five vote 'unknown', the person is categorized as 'unknown'. Otherwise, if the person is categorized as 'known', then the predicted class label is shown as output. If the person is predicted as unknown, then the notification alert is fired to the admin along with the face image of an unknown person.

Figure 3 (a) depicts identification of a known person and Figure 3 (b) depicts identification of an unknown person.

The Figure 4 (a) shows the alert notification that is received by the admin when an unknown person is encountered along with the face image of the unknown person. Figure 4 (b) depicts the app, which allows the admin to browse a log of unknown people.

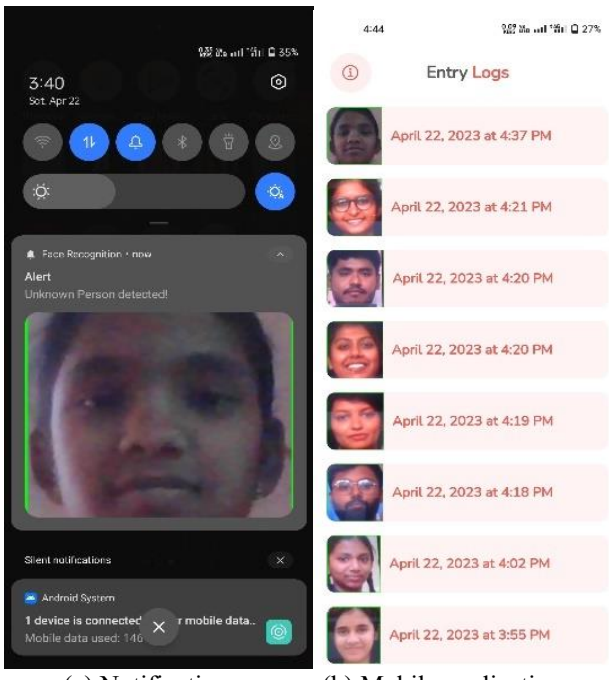


(a) Known individual



(b) Unknown individual

Figure 3. Face recognition output



(a) Notification (b) Mobile application

Figure 4. Screenshots of the mobile application

5. DATASETS

We have used five Datasets to test the performance of the system. For each dataset, 80-20 rule is followed while storing images in the train and test directories and 10 images are stored for validation purposes. For validation purposes, the 10 images at varying angles are considered.

5.1 FaceScrub

The FaceScrub dataset [38] was developed using a technique where faces were detected from the images fetched from the Internet followed by the discard of the images that did not classify to the person in question. It consists of 106,863 images of 530 celebrities with approximately 200 images per person.

5.2 AT&T

AT&T dataset [40] is organized into 40 classes, each of which has ten face images. The images were taken at various times of day, with varying lighting, face emotions, and facial features. All of the images were taken with the person against a dark, uniform background.

5.3 Faces94

Faces94 dataset [41] contains images of 153 individuals and there are 20 images per individual at a resolution of 180 by 200 pixels. The participants were told to speak while a series of twenty photographs were captured while sitting at roughly the same distance from the camera.

5.4 Grimace

Grimace dataset [42] contains images of 18 individuals and there are 20 images per individual at a resolution of 180 by 200 pixels. The subject was moving his/her head while capturing the images. The interval between frames in the series was around 0.5 seconds.

5.5 Georgia Tech

The Georgia Tech face database [43] comprises photos of 50 persons collected at Georgia Institute of Technology's Center for Signal and Image Processing between June 1, 1999 and November 15, 1999. Each person in the database is represented by fifteen 640×480 pixel color JPEG pictures with a cluttered backdrop. The faces in these images are 150×150 pixels on average. The images depict frontal and/or angled faces with a variety of facial expressions, lighting, and size. To establish the position of the face in the picture, each image is manually labeled.

6. RESULTS

This research work is performed on five datasets-AT&T, Faces94, FaceScrub, Georgia Tech, & Grimace. The performance metric used in this research work is "accuracy" to compare results on different datasets. The accuracy Eq. (1) is calculated from TP, TN, FP, FN where TP: True Positive, TN: True Negative, FP: False Positive, and FN: False Negative.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Table 2 shows accuracy comparison of various methods

across various datasets. Table 2 illustrates that, when compared to other methods, the suggested approach produced improved results.

Each model is tested on each of the 5 datasets. After that the proposed technique is also evaluated on all five datasets. Figure 5 depicts the comparison of accuracy of various algorithms across different datasets. Arcface showed a consistent accuracy of >97.4% whereas, Dlib showed a lot of variation in accuracy with the best accuracy on the Grimace dataset (100%) and accuracy of 73.14% on the FaceScrub dataset. On the other hand, Facenet gave an accuracy of 100% on all datasets except the FaceScrub dataset (98.27%) while Facenet-512 showed a similar trend with 100% on all datasets except for the FaceScrub dataset (85.45%). VGGFace gave an accuracy of 100% on AT&T, Faces94 and Grimace dataset, 93.95% on the FaceScrub dataset and 99.33% on the Georgia.

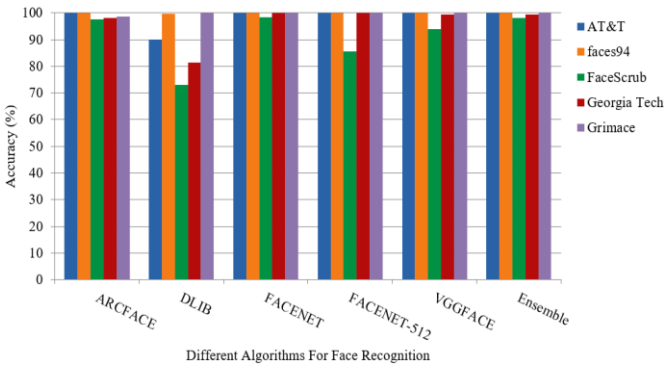


Figure 5. Accuracy comaprision across different datasets for various algorithms

Table 2. Accuracy comparison with other existing approaches

Method	Accuracy	Datasets
Moustafa et al. [11]	81.5% 96.5%	FGNET MORPH (album-II)
Astawa et al. [14]	99.84%	KomNet
Kumar et al. [15]	96.5% 99.09% 97.43% 95.7% 99.25% 96.61%	AT&T Essex94 Essex95 Essex96 Grimace Georgia Tech
Sikarwar et al. [27]	98.19% 95.32%	Faces94 Grimace
Proposed Method	100% 100% 98% 99.3% 100%	AT&T Faces94 FaceScrub Georgia Tech Grimace

Table 3. Accuracy (%) comparison with other approaches

Method	Accuracy (Faces94)	Accuracy (Grimace)
Kumar et al. [15]	99.09%	99.25%
Sikarwar et al. [27]	98.19%	95.32%
Proposed Method	100%	100%

Tech dataset. Finally, the proposed method outperformed all the other algorithms and showed an accuracy of 100% on the AT&T, Faces94 and Grimace datasets, 99.3% on the Georgia Tech dataset and 98% on the FaceScrub dataset. The proposed system employs ensemble learning technique of 5 models. As it is challenging to produce better outcomes with a single model, the combination of 5 models helped to improve the prediction accuracy.

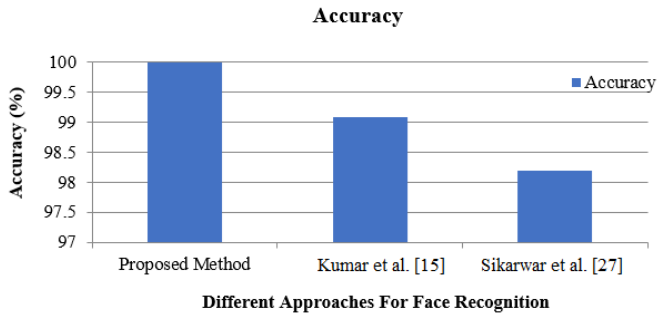


Figure 6. Comparison of % accuracy gained using different approaches (Faces94 dataset)

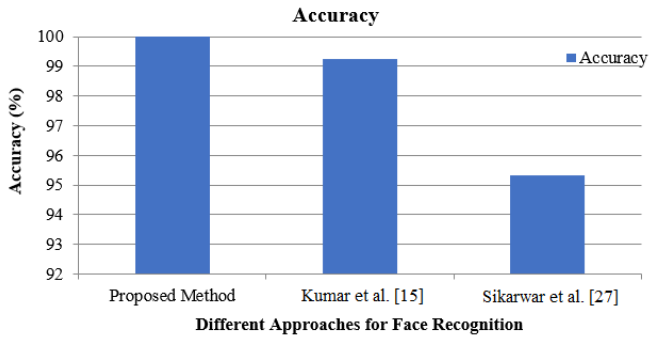


Figure 7. Comparison of % accuracy gained using different approaches (Grimace dataset)

Figures 6 and 7 show the accuracy gained using the various approaches listed in Table 3 on the Faces94 and Grimace datasets. On the Faces94 dataset, the suggested method utilising the ensemble approach achieved 100% accuracy, while Kumar et al. [15] achieved 99.09% and Sikarwar et al. [27] achieved 98.19%. On the Grimace dataset, the suggested method employing the ensemble approach achieved 100% accuracy, while Kumar et al. [15] achieved 99.25% and Sikarwar et al. [27] achieved 95.32%. This demonstrates that the suggested system using the ensemble approach outperformed the other current approaches listed in Table 3. The maximum-voting ensemble increased classification performance, yielding the best accuracy.

7. CONCLUSIONS

In a real-time context, implementing a facial surveillance system is a difficult challenge. Furthermore, the performance of a single face recognition model is insufficient because it may result in incorrect predictions if the face is not stable in real time. To address these issues, this study proposes a surveillance system based on the max-voting ensemble technique, which integrates predictions from five well-known face recognition algorithms: VGGFace, Facenet, Facenet-512, Dlib, and Arcface. Several public datasets were used in the tests, including AT&T, Faces94, Grimace, FaceScrub and Georgia Tech. On the AT&T, Faces94, and Grimace datasets,

the proposed technique was 100 percent accurate. On the other hand, on the FaceScrub dataset it had a 98 percent accuracy and on the Georgia Tech dataset it had a 99.3 percent accuracy. The ensemble method, along with the re-verification methodology, makes the prediction of the unknown exceedingly fast and precise in real time, improving the system's performance. The proposed system can be used in:

(1) Assisting law enforcement by identifying suspects and locating missing persons by comparing images or videos with known databases.

(2) Border control and immigration to verify traveler identities, swiftly identifying forged documents to prevent fraud and unauthorized entry.

(3) Access control for buildings and smart homes, replacing traditional methods, such as key cards or passwords, ensuring secure entry only for authorized individuals.

While the proposed system enhances the accuracy of face recognition and verification, the simultaneous utilization of five models can introduce some performance constraints, resulting in slower processing times. However, by providing ample computing resources, such as higher-capacity RAM and a more powerful graphics card, the system's performance can be significantly improved. These upgrades would enable smoother execution and expedite the overall process, enhancing the efficiency and effectiveness of the system.

REFERENCES

[1] Parkhi, O., Vedaldi, A., Zisserman, A. (2015). Deep face recognition. In BMVC 2015-Proceedings of the British Machine Vision Conference 2015. British Machine Vision Association.

[2] Schroff, F., Kalenichenko, D., Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 815-823. https://doi.org/10.1109/CVPR.2015.7298682

[3] Davidsandberg. https://github.com/davidsandberg/facenet, access on April 16, 2018.

[4] High Quality Face Recognition with Deep Metric Learning. (2017). http://blog.dlib.net/2017/02/high-quality-face-recognition-with-deep.html

[5] Deng, J., Guo, J., Xue, N., Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Beach, CA, USA, pp. 4690-4699 https://doi.org/10.48550/arXiv.1801.07698

[6] Chollet, F. (2015). Keras. https://github.com/fchollet/keras.

[7] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Zheng, X. (2016). {TensorFlow}: A System for {Large-Scale} Machine Learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16), Savannah, GA, USA, pp. 265-283. https://doi.org/10.48550/arXiv.1605.08695

[8] Wang, Y., Bao, T., Ding, C., Zhu, M. (2017). Face recognition in real-world surveillance videos with deep learning method. In 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, pp. 239-243. https://doi.org/10.1109/ICIVC.2017.7984553

[9] Mattmann, C.A., Zhang, Z. (2019). Deep facial recognition using tensorflow. In 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS), Denver, CO, USA, pp. 45-51. https://doi.org/10.1109/DLS49591.2019.00011

[10] Oliveira, J.S., Souza, G.B., Rocha, A.R., Deus, F.E., Marana, A.N. (2020). Cross-domain deep face matching for real banking security systems. In 2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG), Buenos Aires, Argentina, pp. 21-28. https://doi.org/10.1109/ICEDEG48599.2020.9096783

[11] Moustafa, A.A., Elnakib, A., Areed, N.F. (2020). Age-invariant face recognition based on deep features analysis. Signal, Image and Video Processing, 14(5): 1027-1034. https://doi.org/10.1007/s11760-020-01635-1

[12] Sepas-Moghaddam, A., Correia, P.L., Nasrollahi, K., Moeslund, T.B., Pereira, F. (2018). Light field-based face recognition via a fused deep representation. In 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), Aalborg, Denmark, pp. 1-6. https://doi.org/10.1109/MLSP.2018.8516966

[13] Lu, Z., Jiang, X., Kot, A. (2017). Enhance deep learning performance in face recognition. In 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, pp. 244-248. https://doi.org/10.1109/ICIVC.2017.7984554

[14] Astawa, I.N.G.A., Radhitya, M.L., Ardana, I.W.R., Dwiyanto, F.A. (2021). Face Images Classification using VGG-CNN. Knowledge Engineering and Data Science, 4(1): 49-54. https://doi.org/10.17977/um018v4i12021p49-54

[15] Kumar, S., Athul, A. R., Sethi, A., Bombay, I.I.T. (2022). Face recognition and verification using transfer learning. Transfer Learning, 1: 1. https://doi.org/10.13140/RG.2.2.26851.99367

[16] Jose, E., Greeshma, M., Haridas, M.T., Supriya, M.H. (2019). Face recognition-based surveillance system using facenet and mtcnn on jetson tx2. In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, pp. 608-613. https://doi.org/10.1109/ICACCS.2019.8728466

[17] Manna, S., Ghildiyal, S., Bhimani, K. (2020). Face recognition from video using deep learning. In 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 1101-1106. https://doi.org/10.1109/ICCES48766.2020.9137927

[18] Anitha, G., Devi, P.S., Sri, J.V., Priyanka, D. (2020). Face recognition-based attendance system using MTCNN and Facenet. Zeichen Journal, 6(8): 189-195.

[19] Nyein, T., Oo, A.N. (2019). University classroom attendance system using facenet and support vector machine. In 2019 International conference on advanced information technologies (ICAIT), Yangon, Myanmar, pp. 171-176. https://doi.org/10.1109/AITC.2019.8921316

[20] Arsenovic, M., Sladojevic, S., Anderla, A., Stefanovic, D. (2017). FaceTime-Deep learning-based face recognition attendance system. In 2017 IEEE 15th International symposium on intelligent systems and informatics (SISY), Subotica, Serbia, pp. 53-58. https://doi.org/10.1109/SISY.2017.8080587

[21] Suguna, G.C., Kavitha, H.S., Sunita, S. (2021). Face recognition system for realtime applications using SVM combined with FaceNet and MTCNN. International Journal of Electrical Engineering and Technology (IJEET), 12: 328-335. https://doi.org/10.34218/IJEET.12.6.2021.031

[22] Sanchez-Moreno, A.S., Olivares-Mercado, J., Hernandez-Suarez, A., Toscano-Medina, K., Sanchez-Perez, G., Benitez-Garcia, G. (2021). Efficient face recognition system for operating in unconstrained environments. Journal of Imaging, 7(9): 161. https://doi.org/10.3390/jimaging7090161

[23] Mehta, R., Satam, S., Ansari, M., Samantaray, S. (2020). Real-time image processing: face recognition based automated attendance system in-built with "Two-Tier Authentication" method. In 2020 International Conference on Data Science and Engineering (ICDSE), Kochi, India, pp. 1-6. https://doi.org/10.1109/ICDSE50459.2020.9310090

[24] Seelam, V., kumar Penugonda, A., Kalyan, B.P., Priya, M.B., Prakash, M.D. (2021). Smart attendance using deep learning and computer vision. Materials Today: Proceedings, 46: 4091-4094. https://doi.org/10.1016/j.matpr.2021.02.625

[25] Kuang, W., Baul, A. (2020). A real-time attendance system using deep-learning face recognition. Kuang, W., & Baul, A. (n.d.). A Real-time Attendance System Using Deep-learning Face Recognition. 2020 ASEE Virtual Annual Conference Content Access Proceedings. https://doi.org/10.18260/1-2--33949

[26] D'Silva, K., Shanbhag, S., Chaudhari, A., Patil, M.P. (2019). Spot me-a smart attendance system based on face recognition. International Research Journal of Engineering and Technology (IRJET), 6(3): 4239.

[27] Sikarwar, A., Chandra, H., Ram, I. (2020). Real-Time Biometric Verification and Management System Using Face Embeddings. In 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, pp. 1-4. https://doi.org/10.1109/INDICON49873.2020.9342551

[28] Chanda, S., GV, A.C., Brun, A., Hast, A., Pal, U., Doermann, D. (2019). Face recognition-A one-shot learning perspective. In 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Sorrento, Italy, pp. 113-119. https://doi.org/10.1109/SITIS.2019.00029

[29] Xia, H., Li, C. (2019). Face recognition and application of film and television actors based on DLIB. In 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, pp. 1-6. https://doi.org/10.1109/CISP-BMEI48845.2019.8965869

[30] Marsi, S., De Bortoli, L., Guzzi, F., Bhattacharya, J., Cicala, F., Carrato, S., Ramponi, G. (2019). A face recognition system using off-the-shelf feature extractors and an ad-hoc classifier. In Applications in Electronics Pervading Industry, Environment and Society: APPLEPIES 2018 6, Pisa, Italy, pp. 145-151. https://doi.org/10.1007/978-3-030-11973-7_18

[31] Su, D., Li, Y., Zhao, Y., Xu, R., Yuan, B., Wu, W. (2020). A face recognition algorithm based on dual-channel images and VGG-cut model. Journal of Physics: Conference Series, 1693(1): 012151. https://doi.org/10.1088/1742-6596/1693/1/012151

[32] Chinapas, A., Polpinit, P., Intiruk, N., Saikaew, K.R. (2019). Personal verification system using ID card and face photo. International Journal of Machine Learning and Computing, 9(4): 407-412. https://doi.org/10.18178/ijmlc.2019.9.4.818

[33] Son, N.T., Anh, B.N., Ban, T.Q., Chi, L.P., Chien, B.D., Hoa, D.X., Hassan Raza Khan, M. (2020). Implementing CCTV-based attendance taking support system using deep face recognition: A case study at FPT polytechnic college. Symmetry, 12(2): 307. https://doi.org/10.3390/sym12020307

[34] Guo, X., Nie, J. (2020). Face recognition system for complex surveillance scenarios. Journal of Physics: Conference Series, 1544(1): 012146. https://doi.org/10.1088/1742-6596/1544/1/012146

[35] Jiao, J., Liu, W., Mo, Y., Jiao, J., Deng, Z., Chen, X. (2021). Dyn-arcFace: dynamic additive angular margin loss for deep face recognition. Multimedia Tools and Applications, 80(17): 25741-25756. https://doi.org/10.1007/s11042-021-10865-5

[36] Li, X., Wang, F., Hu, Q., Leng, C. (2019). Airface: Lightweight and efficient model for face recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea (South). https://doi.org/10.1109/iccvw.2019.00327

[37] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770-778. https://doi.org/10.1109/cvpr.2016.90

[38] Ng, H.W., Winkler, S. (2014). A data-driven approach to cleaning large face datasets. In 2014 IEEE international conference on image processing (ICIP), Paris, France, pp. 343-347. https://doi.org/10.1109/icip.2014.7025068

[39] Nguyen, H.V., Bai, L. (2010). Cosine similarity metric learning for face verification. In Asian Conference on Computer Vision, New Zealand, pp. 709-720. https://doi.org/10.1007/978-3-642-19309-5_55

[40] Samaria, F.S., Harter, A.C. (1994). Parameterisation of a stochastic model for human face identification. In Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, pp. 138-142. https://doi.org/10.1109/acv.1994.341300

[41] Face Recognition Data, University of Essex, UK, Faces94. https://cmp.felk.cvut.cz/~spacelib/faces/faces94.html.

[42] Face Recognition Data, University of Essex, UK, Grimace. https://cmp.felk.cvut.cz/~spacelib/faces/grimace.html.

[43] Nefian, A.V., Hayes, M. (1999). Face recognition using an embedded HMM. In IEEE Conference on Audio and Video-Based Biometric Person Authentication, USA, pp. 19-24.