An Automatic Student Attendance Monitoring System Using an Integrated HAAR Cascade with CNN for Face Recognition with Mask

Kosuri Naresh Babu[1]* , Suneetha Manne[2]

[1] Department of Information Technology, Geethanjali College of Engineering and Technology, Hyderabad 501301, India
[2] Department of IT, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada 520007, India

Corresponding Author Email: naresh.kosuri@gmail.com

ABSTRACT

In the olden day's many organizations including private and government finds it difficult to mark the attendance manually. A few decades back with the research on biometrics and image processing many smart applications like face recognizers and scanners came into existence but all these apps suffer from single face scanning problem but from the past 5 years many object detection algorithms help us to classify many objects or faces at a time based on multi facial points using boundary boxes to segment the regions. Many research works are carried out for the recognition of faces without masks. With the help of detection algorithms, the proposed algorithm tries to recognize the face of the students with or without masks to mark the attendance in this pandemic situation by designing HAAR integrated with LBP and CNN to find the multiple persons based on the facial points associated with the upper nose, eyes and other regions to extract the features.

1. INTRODUCTION

HAAR is popular for detecting the edges and lines in an image based on their intensity levels. It identifies the variation in the intensity levels to classify whether it is an important pixel or not. To detect the components of the image, a sliding window is defined to extract intensity values by constructing the five rectangular components, which cover all the regions of the images in all the possible directions. The features of the image are computed using the Eq. (1).

$$\text{Feature\_value} = \left(\frac{\sum_i^k \text{dark\_pixels}}{\text{number of dark\_pixels}} - \frac{\sum_i^k \text{light\_pixels}}{\text{number of light\_pixels}} \right) \quad (1)$$

where, $\sum_i^k dark\_pixels$ represents the sum of all the pixel values with 1.
$\sum_i^k light\_pixels$ represents the sum of all the pixel values with 0.
k represents the corresponding number of pixels.

The computation of the feature value can be described in Figure 1 and Eq. (1), Eq. (2), Eq. (3) by assuming some random values for both image intensity and HAAR values.

In General Face Detection→LBP, HAAR cascade and HoG, and Face Recognition→Improved Inception CNN.

Figure 1. Example for edge detection using HAAR

$$\text{Dark Pixels Calculations} = \frac{1+0.4+0.8+1+0.4+0.5+0.7+0.8}{8} = 0.7 \quad (2)$$

$$\text{Light Pixels calculations} = \frac{0.6+0.9+0.2+0.3+0.3+0.4+0.2+0.7}{8} = 0.45 \quad (3)$$

$$\text{Feature\_value} = 0.7\text{-}0.45 = 0.25 \quad (4)$$

Since the value is closer to 0, it concludes that there is no edge in the given image. Depending on the Feature_value the algorithm determines there is an edge if the value is closer to 1. The five rectangular components of HAAR are described in Figure 2.

Figure 2. Rectangular components of HAAR for feature extraction

2. LITERATURE REVIEW

Patil et al. [1] used the Viola-Jones technique for detecting the face and LDA with machine learning algorithms to automatically recognize the face. This work is divided into five sections which can be explained briefly as follows on the

self-designed dataset containing images of students in various scenarios and conditions: the first section deals with pre-processing using histogram equalization to convert the color images into grayscale images and are stored in the database for future operations. The second section deals with face detection using the Viola-Jones Method in which the features are transformed through a sliding window of fixed size to extract the features using HAAR combined with Adaboost to classify the features into two groups based on the threshold values. The third section deals with feature extraction using LDA, intending to maximize the discriminate analysis in between inter and intraclass variations by reducing the dimensions by computing the maximum values among the Eigenvectors of the scattered matrix. Finally, recognition is performed based on the KNN classifier. The objective of the work is the maximization of activation function by using the Radial Bias Function parameter in the SVM algorithm to design the best hyperplane.

Halder et al. [2] designed a smart framework based on deep learning algorithms to mark attendance for the students/faculty of a university directly into the database using the time of entry and time of exit. In the processing step, the details of the image are obtained in the form of a tensor with its image notation in the form of gray pixel values and class labels associated with each image. Convolution layer with 64 units and a sliding window kernel size of 3 is designed.

To recognize the students, the model has implemented a 2 layered dense neural network with softmax activation function and it efficiently solved the problem of gradient vanishing problem by adding dropout layers. The input to the system comes from the live video images, so the motion-triggered system identifies the local as well as global changes in the intensity levels from one frame to another frame. The system handles the dynamic variants in intensity levels by applying the subtraction operation on the background. The comparison of the real frames for the variation of intensity levels is performed concerning the time factors. By applying the hyperparameters concept, it studied the three different scenarios and timings.

Reddy et al. [3] explored a novel approach based on the features and their matching patterns using artificial intelligence. The system is implemented in American, US schools to provide security to the students while their parents want to take their wards home after the completion of school hours. At the security gates, attached cameras capture the live stream of parents and allow them to enter the campus if and only if their facial features match the database images.

Ray et al. [4] designed a portal for marking the attendance automatically by capturing the facial features of the image. A high-resolution webcam is attached in the class and through the server, the student's movements are continuously monitored. The frames captured are compared against the student's images stored in the database. A multi-layer perceptron has been implemented with many normalization layers. This model has maintained a high precision rate even when the number of students is more in classrooms like seminar halls, auditoriums, and other crowded rooms.

Patil et al. [5] experimented with an image tagging mechanism to mark the attendance, which is proposed to overcome the drawbacks associated with biometrics and the facial scanning process. In the tagging process, every record is assigned with the class label, generally, the name of the person. In this system, a LBP is used to train the algorithm

for the recognition and detection of a face from the images stored in the local server. The main advantage of this model is a single click on record can mark attendance for all the students simultaneously and it automatically sends an email notification to the parents about their ward absence. In this model, the pre-processing is performed on two major operations: one is related to resizing to maintain constant pixel resolution and another is related with stage by stage histogram calculation to produce grayscale images. In the feature extraction phase, for all student's faces detected, the algorithm generates separate blocks, and on each block, histogram combined with linear binary patterns to generate a new single histogram, and the image is further processed using training results based on the patterns matched.

Bah and Ming [6] constructed a framework known as "ARRAY", for recognizing the employee faces to mark attendance. In this research, an improved version of LBP is proposed, in which pre-processing is performed using the contrast adjustment by varying the alpha and beta values to detect the important features. Using the hyper turning method it conducted a test between three types of filters and selected bilateral filter. This filter is combined with a histogram equalizer to perform feature engineering. The HAAR classifier in this work performs a blending task to train the system. This blending task helps to improve the quality of the image by combining more than one image of a particular employee in different environments and to predict the employee accurately in different backgrounds and low lightening conditions. The entire face is divided into many regions and for every pixel in the region; LBP is computed to replace the pixels with binary values based on the threshold levels. Finally, these values are converted into a feature vector and histogram bins are formed to find the more concentrated regions to project the robust facial features.

Khan et al. [7] designed an API for smartphones to mark the attendance using Object Detection algorithms. The new parameter considered in this research is, the marking of the attendance is performed during entry as well as exit points to ensure that the student has stayed for the complete class. In this model, YOLO V3 is implemented by replacing the detection of objects with the detection of faces along with several students present in the class. The recognition of face and generating the spreadsheet with attendance marked automatically is taken care by AZURE. Here, to implement a convolution neural network, pixels are arranged as tensors, because they efficiently handle the features related to edges and colors and they arrange the features from highest to lowest based on the details of the project. In this architecture, the formed regions are manipulated by the Fast RCNN to identify the region of interest by drawing boundary boxes at a single step and the classification of images is performed region by region using the DARKNET, which is popular for its parallel computations in GPU environments.

The proposed algorithm tries to recognize the face of the students with or without masks to mark the attendance in this pandemic situation by designing HAAR integrated with LBP and CNN to find the multiple persons based on the facial points associated with the upper nose, eyes and other regions to extract the features.

3. PROPOSED METHODOLOGY

The main objective of the proposed system is to mark the attendance of the students by comparing their images in the

database with the captured videos with masked faces. To achieve this objective, it has divided the entire process into two phases: one deal with face detection and another deal with face recognition.

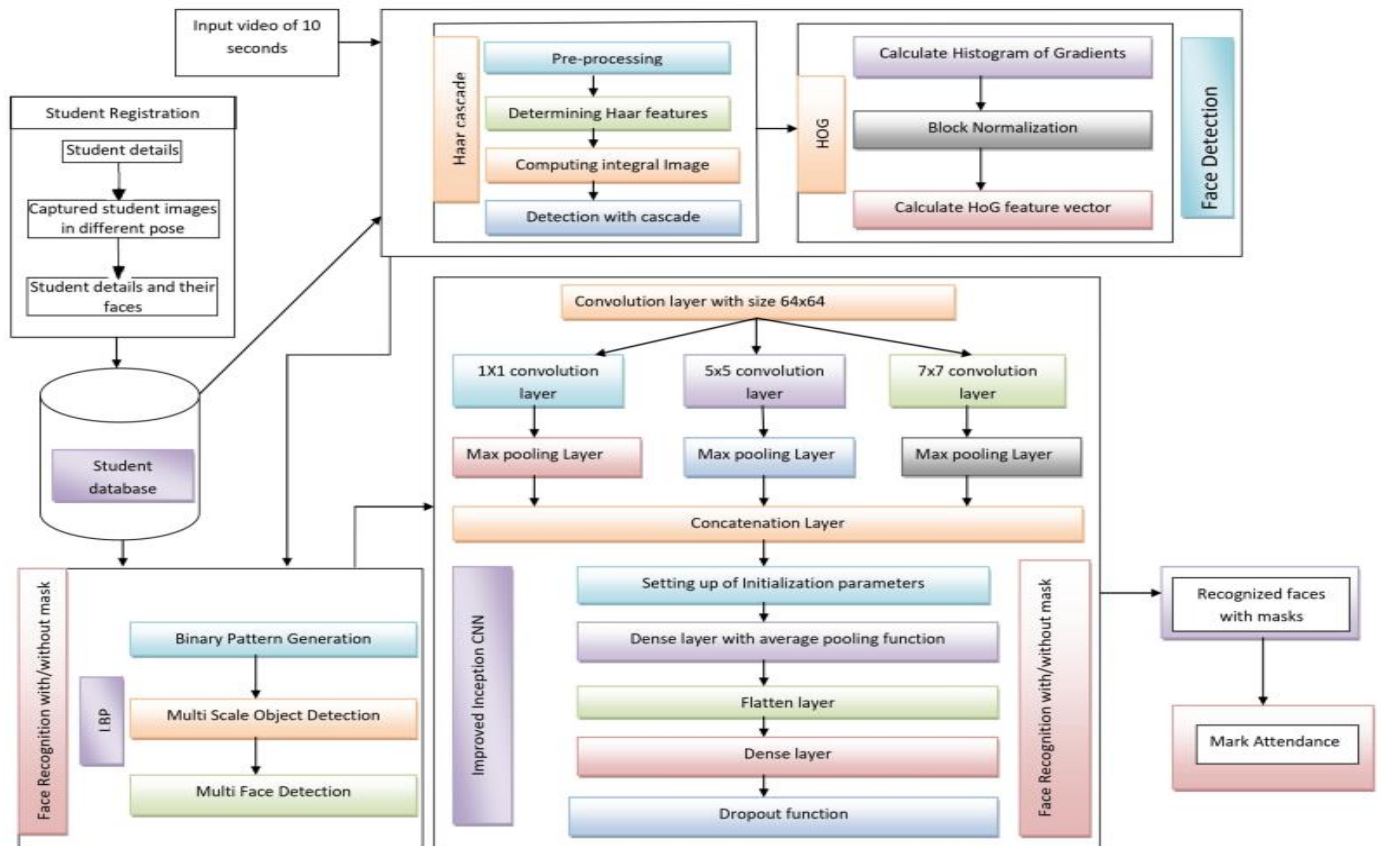The overall process of marking attendance can be illustrated in Figure 3.

Figure 3. The overall process for face recognition with and without a mask

3.1 Face detection using HAAR-CASCADE+HOG

Histogram of Oriented Gradient is treated as "Feature Descriptors", which can extract the necessary and detailed information from the images based on their structure during the process of feature engineering. In this mechanism, the overall image is partitioned into regions by defining the local portions and generating the histograms for them [8].

The gradients represent the changes in the magnitude concerning x and y directions and their computations are stored separately in two different matrices. The Gradient change concerning x-direction can be computed as shown in Figure 4.

Figure 4 has assumed, the part of the nose can be represented in pixel notation as shown in matrix format. Since it is a change with respect to the horizontal direction, the right value is subtracted from the left side value, and the highlighted pixel is updated accordingly. Like this, the entire matrix is updated with respect to both directions. The orientation for each pixel is computed using the Pythagoras theorem and finally, a histogram is created based on the bins computed for the continuous values obtained from the pixels.

Even though HAAR is the oldest technology, but still it remains in the top position because of its high computational speed in handling both images and videos when compared to many of the latest detection algorithms. To work with the image convolutions, a sliding window known as a kernel of multiple scales is operated from top to bottom to construct a new matrix with fewer dimensions to simplify the process.

Using the constructed data points, it generates two class labels namely "yes", for facial points detected and "no", for non-facial points detected. These values play a crucial role in the detection of facial parameters based on nose, eyes, and cheeks.
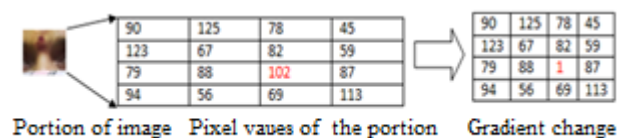
Figure 4. A sample gradient change value for a pixel with respect to X-direction

Table 1. Multi-scale detection parameters for cascade

S.No	Parameter Name	Description
1	Cascade	It reads the facial parameters obtained from HAAR using the YML file
2	Image	The detected points are represented as pixel notation using matrix
3	Objects	It consists of five rectangles in the form of vectors, each rectangle represents the detected points of a particular portion like eyes, nose
4	Scale Factor	It represents the amount of size to be reduced at every iteration
5	Minimum Neighbors	It represents the number of neighborhood cells information to retain

The integration of HAAR with the cascade mechanism divides the entire problem into smaller subsets and the solutions are computed stage by stage to detect the objects in the real-time environment. The multi-scale values are computed using the parameters as discussed in Table 1.

3.2 Face recognition using LBP matching and CNN

The main aim of facial recognition is to compare the captured images with the training images in the dataset by performing 1* N comparisons. The name itself says that the pixel values are filled with binary values based on the threshold values of the neighborhood. Similar to HAAR, LBP also deals with four types of parameters as shown in Table 2.

Table 2. LBP parameters details for feature extraction

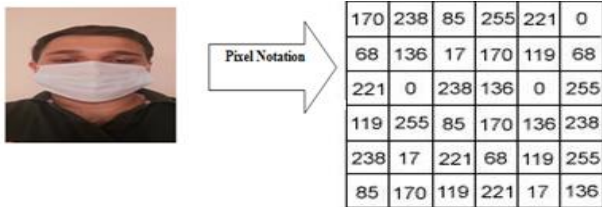| S.No	Parameter Name	Description
1	Radius	The radius values that define the central pixel value
2	Neighbors	It represents the number of nearest pixel points to be considered
3	Grid_X	The number of pixels to be considered in the horizontal direction
4	Grid_Y	The number of pixels to be considered in the vertical direction

Figure 5. Pixel notation for the image

Initially, the facial images are represented as a two-dimensional pixel notation as shown in Figure 5.

To extract the features, first a sliding window of "k" size can be defined, which in general is called "Kernel Size". For simplicity, let us assume the system has 3*3 grayscale pixel values and a threshold value of 60 is computed based on the number of neighbors and radius parameters. The binary values replacement takes places as follows: if the value of the pixel is greater than the threshold value i.e., center pixel value then it is marked as "1" otherwise it is marked as "0". This process can be illustrated as shown in Figure 6.

Now, the binary pattern should be converted to decimal value by concatenating the binary values in the row-wise fashion i.e., 10101110, which is equivalent to 174. This computed value is considered as "central pixel value". The Grid coordinates X and Y define the number of cells to be considered to form the histograms to extract the features by combining small histograms created to generate a new image but containing all the characteristics of the original image so that it ensures no loss of important information from the image.

From the dataset, the algorithm has to find the matching image to mark the attendance. This is done by computing the "Euclidean distance" between the histograms of the input image and generated histograms. Pseudocode for Training and Testing Process is described in Table 3. The efficiency of this algorithm is measured by the parameter known as "confidence". Face recognition along with confidence value during the initial stage of training are sown in Figure 7.
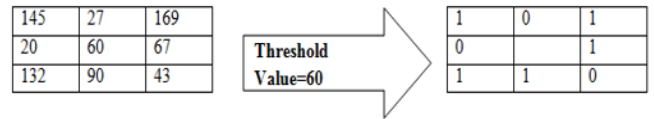
Figure 6. Conversion of gray scale values to binary patterns

Figure 7. Face recognition along with confidence value during the initial stage of training

Table 3. Pseudocode for training and testing process

Pseudocode for Training Process
1. input_img←Load the images from the disk
2. for i in input_img:
a. path,id←get_imgpath(),get_imgid()
b. face_parameters←path
c. extract_parameters←multiscale()
d. face_parameters←extract_parameters+face_parameters
e. train_img←face_parameters.append(id)
f. face_reg←LBHP()
3. return face_reg

Pseudocode for Testing Process
1. input_video←Load the video from the disk/capture directly
2. face_par←multiscale()
3. imgx_train,imgx_test,imgy_train,imgy_test←get_data_image()
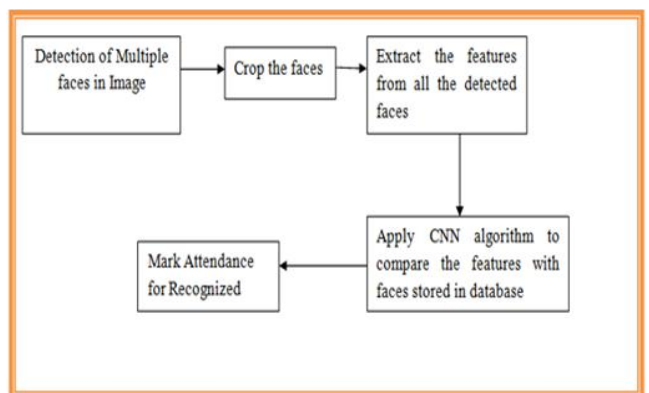4. predict(yml)
5. print confidence label.

Figure 8. Step by step process in marking attendance using CNN

The main advantage of using CNN in facial recognition is, it reduces the parameters to be trained and makes fewer computations during the learning and adaptability phases because of the automatic processing nature of the layers in the network. A three-layered architecture, which are both convoluted and pooled is designed to recognize the multiple faces in a single capture from different sources. The step by step process is presented in Figure 8.

In the previous approaches, the popular process for recognizing the faces is "PCA", where the images are converted into Eigen vectors and then computations are performed to recognize the face. But in this process, most of the time has to be spent on the training phase. Still, it cannot recognize the unknown faces. The system needs a mechanism, in which even an unknown face can be recognized so that the training time becomes less and computational complexity reduces.

In the proposed system, it has used a pre-defined model known as "Inception", which has considered the base database as "ImageNet", which is intended to provide approximately 21 classes of different groups with nearly 15 million images stored in the form of annotations. This has its great impact on the object detection i.e., drawing a boundary box around the recognized object in the design. The major advantage of using this pre-trained model is, it gets the ability to learn both low level and high level features and the annotated features in the images reduces the burden of training on the designed systems to train from the scratch. The main intention of the proposed system is to design the recognition system with constant time complexity even the number of layers are increased.

Table 4. Pseudo code for face recognition using modified Inception CNN model

Begin:
1. Define a 1×1 pooling layer with 64×64 pixel size to reduce the dimensionality
2. Design a concatenation layer with 5×5 and 7×7 with filter size and 32×32 image size
3. Define a max pooling with a stride size of 3
4. Design another concatenation layer with above three layers as input layer
5. Design a convolution layer by initializing the bias parameter, activation function as swish
6. Set up initialization parameters for Inception Integrated with ReLu architecture
7. Apply average pooling function as dense layer
8. Define flatten function with stride size as 1
9. Define another dense convolution layer with 32 neurons
10. Apply Dropout function to normalize the units
11. Go to step 7 till the size becomes 8×8
End

The working on the ImageNet, is done using the Inception Model, which is designed using stacked layers of neurons. The system to solve the vanishing problem, in the proposed system the concept of batch normalization is implemented, in which connections are arranged with skip connections involved in it. The Inception model combines different size filters as single concatenation layer. The concatenation layer combines the data only on the dimensions mentioned in the system. Here, the proposed system along with the specified filters, also applies to the pooling layer. Pseudo code for Face Recognition using Modified Inception CNN model is described in Table 4.

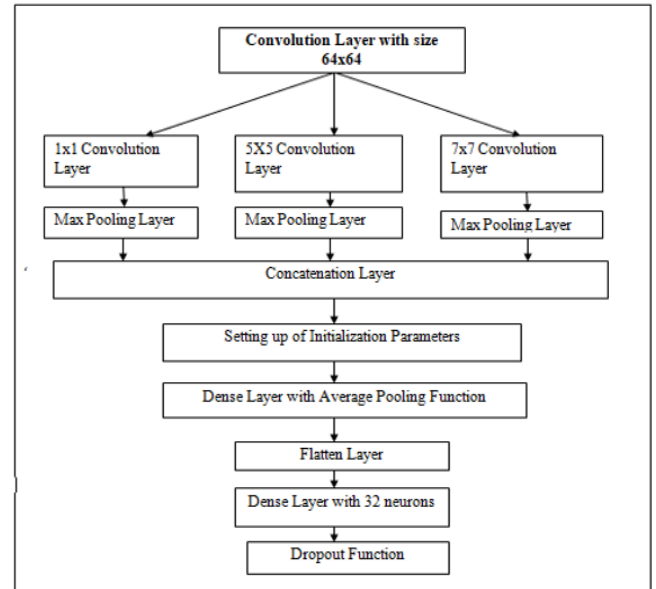The flowchart for designing of neural network for face recognition is illustrated in Figure 9.

Figure 9. Designing of neural network for face recognition

The model implements down sampling technique at each iteration to identify the patterns associated with the stacked layers in the spatial hierarchy to summarize the data present in the current dimensionality. The pooling operator that is implemented in down sampling is "Average Pooling", which is explained in Figure 10.
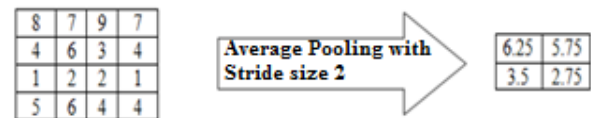

Figure 10. Average pooling operation in neural networks

4. RESULTS AND DISCUSSION

The proposed method implemented the HAAR integrated with HOG and LBP [9] to identify the students with and without masks in the classroom to mark the attendance. Many researchers have developed automation systems for attendance marking as discussed, few works in the literature survey. These works are compared as shown in Table 5.

The proposed method is compared with other works in terms of accuracy and is plotted as shown in Figure 11, in which the x-axis represents different models and the y-axis represents the percentage of accuracy [10, 11]. From the figure, it is evident that the proposed algorithm performs better than the remaining works even in the case of face masks also [12-14].

The analysis shows that HAAR integrated with LBP is efficient in handling the extraction of facial expression based on the upper nose and eyes whereas most of the other works try to find the patterns based on the eyes, nose, cheeks, and lips [15].

Table 5. Comparative study on past works and proposed mechanism

| S.No	Author	Algorithm Implemented	Accuracy	Limitations
1	Vidya Patil	LDA+ML	95%	Very few images are considered
2	Rohit Halder	CNN	96%	More time is taken to identify the best parameters
3	K P Naveen Reddy	LBP	76%	Mechanism related to pre-processing and extractions are not taken care
4	Rahul Ray	MLP	88.4%	Matching with local server data takes a lot of time to achieve the task
5	Vishal Patil	Image Tagging	70%	Cannot work with unknown images of unknown persons
6	Serign Modou	HAAR with blending approach	89.06%	Construction of k*k regions makes the system difficult to build binary vector
7	Sikandar Khan	YOLO V3+AZURE	94.5%	Cannot recognize the facial features in low light conditions
8	**Proposed Algorithm**	**HOG+HAAR+LBP+CNN**	**97.74%**	
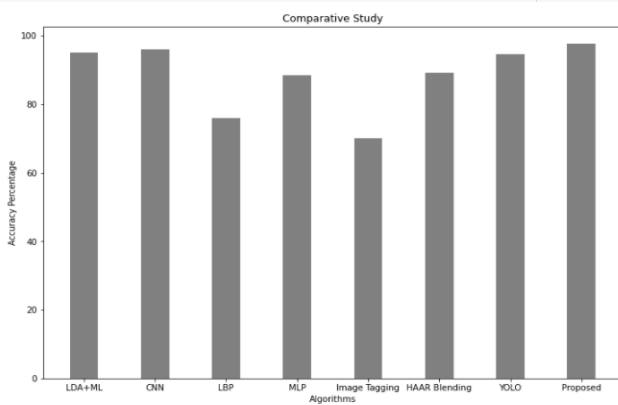
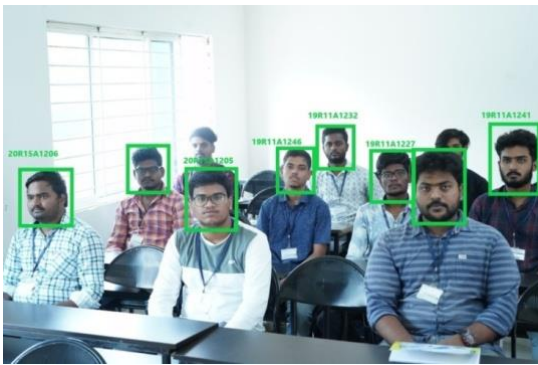**Figure 11.** Analysis of different algorithms on face recognition

Figure 12. Attendance monitoring system without mask

Figure 13. Multiple face recognition system using integrated HAAR cascade

In Figure 12, the attendance is marked automatically by the proposed research algorithm. The images are collected from the real time data and it is evident that out of 10 people, the system is able to identify 8 people. Finally, the proposed system has successfully identified multiple images as shown in Figure 13.

5. CONCLUSIONS AND FUTURE SCOPE

To date research works carried out on facial recognitions are based on the full face without any obstacles but the proposed research is intended to find the facial recognition with mask during this pandemic situation where the mask is a must for 24/7. So, the proposed system designed an algorithm to mark the attendance in a classroom containing students by automatically detecting all of them simultaneously. In the designed system, when a mask is covered over the face, it becomes difficult to identify the facial features because it hides most of the important regions like cheeks, mouth, and part of the nose. But, still, the designed model combines the HOG with HAAR cascade to extract more information from few features available. The LBP plays a vital role in finding the students with accurate confidence levels by defining the right parameters and by choosing the correct neighbors.

Nowadays, wearing face masks is made compulsory. People have to cover their faces in public areas, supermarkets, offices, stores, railway stations, shopping malls, schools, airports, etc. Future work focuses on face mask detection over multi-face images and marking attendance automatically in an excel sheet.

REFERENCES

[1] Patil, V., Narayan, A., Ausekar, V., Dinesh, A. (2020). Automatic students attendance marking system using image processing and machine learning. In 2020 International Conference on Smart Electronics and Communication (ICOSEC), IEEE, pp. 542-546. https://doi.org/10.1109/ICOSEC49089.2020.9215305

[2] Halder, R., Chatterjee, R., Sanyal, D.K., Mallick, P.K. (2020). Deep learning-based smart attendance monitoring system. In Proceedings of the Global AI Congress 2019, Springer Singapore, pp. 101-115. https://doi.org/10.1007/978-981-15-2188-1_9

[3] Reddy, K.N., Alekhya, T., Sushma Manjula, T., Krishnappa, R. (2019). AI-based attendance monitoring system. International Journal of Innovative Technology

and Exploring Engineering, 9(2S): 592-597. http://dx.doi.org/10.35940/ijitee.B1057.1292S19

[4] Ray, R., Khan, F., Sharma, H., Kumar, G. (2020). Class Attendance Portal (CAP) using Face recognition. In International Journal for Research in Applied Science and Engineering Technology IJRASET. https://doi.org/10.22214/ijraset.2020.30745

[5] Patil, V., Kapadia, K., Khokrale, A., Jain, P. (2020). Intelligent college attendance system using image tagging. In Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST). https://dx.doi.org/10.2139/ssrn.3568112

[6] Bah, S.M., Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. Array, 5: 100014. https://doi.org/10.1016/j.array.2019.100014

[7] Khan, S., Akram, A., Usman, N. (2020). Real time automatic attendance system for face recognition using face API and OpenCV.Wireless Personal Communications, 113: 469-480. https://doi.org/10.1007/s11277-020-07224-2

[8] Reddy, A.M., Krishna, V.V., Sumalatha, L., Niranjan, S.K. (2017). Facial recognition based on straight angle fuzzy texture unit matrix. In 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), IEEE, pp. 366-372. https://doi.org/10.1109/ICBDACI.2017.8070865

[9] Reddy, A.M., SubbaReddy, K., Krishna, V.V. (2015). Classification of child and adulthood using GLCM based on diagonal LBP. In 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), IEEE, pp. 857-861. https://doi.org/10.1109/ICATCCT.2015.7457003

[10] Katamaneni, M., Mayuri, A.V.R. (2022). A comprehensive survey on COVID-19 detection and classification using chest-x-ray images. Traitement du Signal, 39(4): 1407-1419. https://doi.org/10.18280/ts.390435

[11] Viola, P., Jones, M.J. (2004). Robust real-time face detection. International Journal of Computer Vision, 57: 137-154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

[12] Farfade, S.S., Saberian, M.J., Li, L.J. (2015). Multi-view face detection using deep convolutional neural networks. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp. 643-650. https://doi.org/10.1145/2671188.2749408

[13] Anwarul, S., Dahiya, S. (2020). A comprehensive review on face recognition methods and factors affecting facial recognition accuracy. Proceedings of ICRIC 2019: Recent Innovations in Computing, 495-514. https://doi.org/10.1007/978-3-030-29407-6_36

[14] Wang, Z., Huang, B., Wang, G., Yi, P., Jiang, K. (2023). Masked face recognition dataset and application. IEEE Transactions on Biometrics, Behavior, and Identity Science. https://doi.org/10.1109/TBIOM.2023.3242085

[15] Bah, S.M., Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. Array, 5: 100014. https://doi.org/10.1016/j.array.2019.100014