**INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING ARTIFICIAL INTELLIGENCE AND TECHNOLOGY (AIT), 2020 ICRTAC – AIT 2020**

**Face Recognition With Mask Using MTCNN And FaceNet**

Abhishek Sunil Tiwari, Prajul Gupta, Aanya Jain,

Hari Vilas Panjwani, Malathi G

School Of Computer Science and Engineering

Vellore Institute of Technology, Chennai Campus- 600127

**Abstract**

Face Recognition is a method of identifying or verifying the identity of an individual using their face but what if this recognition method could be extended further to suit the needs of the current scenario. Given this COVID pandemic, this paper fits best by recognising the people wearing masks. The research has been done by creating our own dataset using images from our friends and relatives followed by doing Image augmentation by performing operations like rotating by some angle, changing brightness and contrast, zooming in and out etc. Then face with the mask is extracted from the given image with the help of MTCNN to get a bounding box, width, the height of the face, then segmentation has been done by reducing the height by a factor of 2. FaceNet pretrained model has been used to represent the faces on a 128-dimensional unit hyper-sphere and get the embeddings for further classification. Many different algorithms like Linear Discriminant Analysis, SVM, Ridge Classifier, K-Neighbors Classifier, Logistic Regression, Naive Bayes, XGBoost, Ada Boost, Random Forest Classifier, Decision Tree classifier have been used for experimentation. After testing this good accuracy was obtained as can be seen in the Result section of this paper. The scope of this paper is quite vast as it covers many practical applications in real-scenario like detecting the presence of a particular person from an image or even from video by capturing faces frame by frame.

**Keywords:** Face Recognition with mask, Facenet, MTCNN

# 

# **Introduction:**

Face Recognition systems are becoming an integral part of authentication at organisations nowadays. It is widely adopted due to its non-invasive and contactless nature. These systems are capable of identifying and verifying a person uniquely and are modern substitutes to traditional biometric systems based on finger impression. Other applications include an automatic grouping of images, video surveillance and also important for better human-computer interaction.

After going through so many recognised and established works of researchers, there was a need to design something that could accommodate the current situation and be able to present a solution to the need of the hour. So this paper provides that much-needed solution thereby helping to predict the identity of the person with the mask.

But the pre-existing system of recognising a person with face won’t work as keeping health guidelines it is important to wear the mask and at the same time. Facial recognition is also needed for authentication purposes. So we have proposed facial recognition systems even for people with a mask based on features of eyes and forehead and used the MTCNN(Multi-task Cascaded Convolutional Neural Network) for getting the pixel coordinate values of the left eye, right eye, bounding box. Further, they are passed into the FaceNet model to generate embeddings and lastly passing these values to the models.

This research paper consists of seven sections including the introduction section. Section 2 explains the various previous approaches and models used for face recognition. Section 3 describes the dataset. Section 4 proposes our novel architecture, Section 5 explains in detail about our proposed work, Section 6 discusses outcomes of several experiments conducted and results obtained, Finally, the last section concludes the research paper best architecture among all the experiments conducted.

# **Related Works :**

Face Recognition is a very popular field among researchers, over time several approaches have been proposed. Principal Component Analysis (PCA)[1] was used for facial recognition purposes, it basically reduces large dimensional data space to small dimensional feature space, by computing a subspace vector also known as Eigenfaces. It does so by representing 2-D images into 1-D vectors of pixels also known as Eigenspace vectors. Linear Discriminant Analysis (LDA) method for face recognition purposes popularly termed as Fisherface [3] method. For Facial recognition method as described by Belhumeur [4], uses both LDA and.PCA, the reason for using LDA is as it does dimensionality reduction with help of linear projection and preserving linear separability at the same time.Although for detecting masked faces[5] used LLE-CNN it only detects whether a face is masked or not, and that with 76.4% accuracy and also doesn’t do face recognition. When these above-mentioned approaches were not satisfactory in a changing environment. In recent years, the combination between GPU and convolutional networks has brought breakthroughs in benchmark evaluations, such as FDDB [6,7]. And further major recent benchmarks in facial recognition domain developed are as follows MTCNN [8] rained on the WiderFace [9] become a popular face detection framework due its proposal of adopting three stages carefully engineered deep convolutional networks that particularly predicts in landmark and face location in coarse to fine manner, exploiting the inherent correlation between them to improve their performance.Facenet[[10]](https://www.zotero.org/google-docs/?KfJzT8) which gives better representational efficiency due to its novel algorithm in which they do training with triplets aligned roughly with matching/non-matching face patches generated using a method of online triplet mining, which helps in directly optimizing embedding itself.Deepface[[11]](https://www.zotero.org/google-docs/?YSlZyf) which performed affine transformation in piecewise form then obtained face representation from a deep neural network consisting of nine layers. This network involves close to 120 million parameters without weight sharing of many locally connected layers.

Since we all know post-COVID masks will be a common thing at work places. And that time another challenge of detecting a person in spite of wearing a mask, and existing approaches wouldn’t be enough to overcome this challenge.M. S. Ejaz and M. R. Islam[13] have tried faces with masks, in which they used MTCNN[8] for face detection,FaceNet[10] for face embeddings and SVM[13] for Classification. M. S. Ejaz and M. R. Islam have done experimentation on standard datasets which have mixed images. We have compared the same with our approach in which we have used LDA[14] for classification, giving decent accuracy of 94% which outperforms all existing approaches in the domain of face recognition of masked faces. We have also made our own dataset for training purposes. Our architecture is novel because we only take the upper half of face for embedding purposes, so that forehead, eyes features are taken into consideration rather than mask type, color which may rather act as noise for actual recognition.

# **Dataset:**

Although it was hard to find the labelled dataset of images with people wearing masks, we have created our own dataset by taking mask images from our friends and relatives for training the model. The dataset has been divided to Training set and Testing set.Both has images of 49 people wearing face masks.The images looked quite alike so to create variation and a diversified dataset,images have been augmented by changing angle,height,width.Finally a new dataset has been generated with better chances of training the model.The dataset consists of 2 subparts,training set and testing set.The training has 20 images per person with masks on their faces with different variations.The other is the testing set with 10 images per person with masks to actually test that the person wearing mask belongs to any one of the labels in training set.

# **Proposed Architecture :**

In the current paper we developed a system which detects the face of the person with a mask. In our system architecture the first stage is the augmentation of an image dataset using Image-DataGenerator class in python. After the initial stage the augmented image is passed through the MTCNN model for face detection purpose and extracted face image will be the output of that model. Then FaceNet is used for converting the detected image into an embedded vector. After that we have just preprocessed our dataset and splitted them into train and test images. For the classification, LDA(Linear Discriminant Analysis) is used. The overview of current system is demonstrated in figure 1:

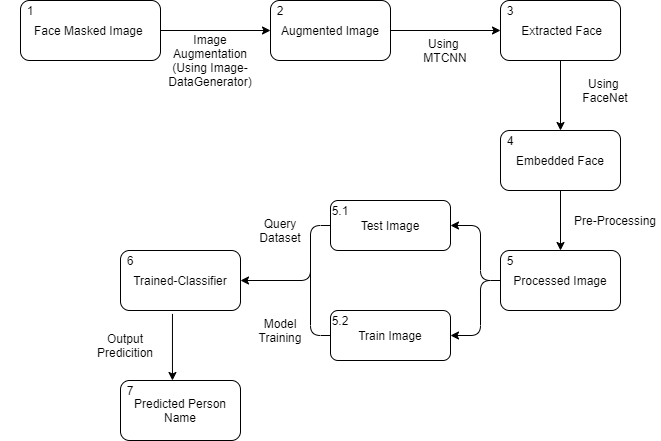


Fig 1: Architecture Diagram

The System consists of 3 important Models, they are:

1. **MTCNN:**

Multi-task Cascaded Convolutional Networks (MTCNN) is the combination of several convolutional networks strung together which gives out several pieces of information.This network detects individual faces and it locate facial landmarks like eyes and forehead in our case and gets their pixel coordinate values. The MTCNN model consists of 3 different networks i.e P-Net(Pnet network layers takes image as an input and output 2 things: 1. probability of face being in bounding box 2. coordinates of bounding box),R-Net(Output of Pnet acts as an input for the Rnet and it gives out more accurate coordinates of the bounding box),O-Net(outputs of R-Net act as its input and outputs 3 sets of data: 1.Bounding Box coordinates, 2.the probability of the face being in the box 3. the coordinates of the facial landmarks.).

1. **FaceNet:**

FaceNet uses deep convolutional neural network(CNN) along with triplet loss for better accuracy. FaceNet has unique architecture and because of it tasks like face recognition, verification as well as clustering can easily be performed. It also provides unified embedding for all tasks. According to the face similarity it maps each face into an euclidean space. Let’s consider an image dataset, faceNet will place a person image more closer to its own image as compared to images of any other person. It learns from the mapping of the images and creates embedding rather than using any bottleneck layer approach and these embedding can be further used for tasks like recognition, verification etc.

Another feature of FaceNet is its triplet loss function. We need 3 images namely anchor,positive and negative, for the calculation of triplet loss.



Equation 1: Triplet Loss Formula

Where f(a): embedding of an anchor image

f(p): embedding of an positive image

f(n): embedding of an negative image

N: margin between positive and negative pairs(threshold value)

1. **LDA:**

Linear Discriminant Analysis(LDA) is a dimensionality reduction technique in the preprocessing step which is used for classification. As the name implies dimension reduction technique project dataset onto lower- dimensional space by retaining as much information as possible. LDA is quite similar to PCA but in LDA we chose that axis which maximizes the separation between the multiple classes. The working principle of LDA is to use the information of both x,y axes and then to create a new axis and project that data onto the new axis such that variance can be minimized and the distance between the means of two classes could be maximized.

LDA consists of statistical properties of a dataset like mean variance, standard deviation etc calculated for each class. For a single variable it is mean and variance whereas for multivariate Gaussian it is mean and the covariance matrix.

**5. Proposed Work:**

This work is an improved version of a pre-existing recognition system which took full face as input. But as we know post - COVID era people will be wearing masks and as for recognising person masks we have come with architecture which would take features of eyes and forehead features and will generate encoding using Facenet model architecture.

***5.1 MTCNN(Multi-task Cascaded Convolutional Neural Network):***

The MTCNN works by extracting face from given filename and then converting to RGB, getting the pixel coordinates which include centre coordinates, detected face height and width, obtained and then we have done height/2 for a portion of the face above mask.It is resized to (160,160). After resizing, store the resized image into trainX,testX variable and store class name in trainY,testY and finally store all variables in a .npz file. So up till now we have extracted the face upper part only without a mask consisting of eyes and forehead.

***5.2 FaceNet Model***

Now creating face embedding comes into play and we had done this part using FaceNet. So in face embedding features are extracted from the face and are represented in the form of a 128 element vector. Later on, with the help of these vectors distinguishing between 2 persons can be done. Suppose two-element vectors of the same person will be similar whereas for different people element vectors will vary a lot. The work of FaceNet in our project is simple, it will generate the embedding for a given face in the form of an element vector and will store them with the name of the person whose face is embedded.After MTCNN is done,load the detected faces .npz file using load function in NumPy library and again store values in train and test.Loading the pre-trained Facenet keras model is the easiest step as it helps to generate embeddings for face.Now scale the pixel values and then standardize them so that they can meet the input expectations of the FaceNet model.Expand the dimensions and then use the “facenet\_keras.h5” model for prediction and extraction of the resultant embedding/element vector.The function will receive the model as well as detected face as input and will return the embedded vector.New variable created after getting embeddings each embedding of (128X1) Now the most important steps i.e. Face Classification comes. Under this we will fit a different classification model using pycaret on our dataset. So Firstly we loaded face-embeddings.npz file and then we developed our classification model. But Before developing our classification model we required to perform some data pre-processing.

***5.3 Data Preprocessing***

Under Data preprocessing do normalization by using Normalizer class which scaled our values to unit length.This Normalization is performed both on training dataset as well as testing dataset. After Normalization, label encoding is done using the LabelEncoder class. Label Encoding converts our string target variables i.e. name of the person into some integers.This leads to our end of data preprocessing. Now comes the final part which is developing a model and fitting that model to our train dataset and then predicting the accuracy score.

***5.4 Models using Pycaret***

Using Pycaret which is an open-source machine learning library which is simple and easy to use.It allows us to compare different machine learning models and fine-tune them very easily.Initialise the models by creating them and tune them to give best results,as pycaret itself does grid search to find the best parameters in case of getting maximum accuracy. Now just compare all the classification models to check which model did its job best by doing the most accurate prediction of the person wearing the mask just by inputting a masked-image of them.

# **6. Result and Discussion**

As of now, our proposed architecture has outperformed all existing approaches with very good accuracy in spite of the mask on the face. We have experimented with several algorithms and some of our key observations are recorded in Table1 and are as follows:Tuned LDA Model:

Table 1: Tuned LDA Model Accuracy Table

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No.** | **Accuracy** | **AUC** | **Recall** | **Prec.** | **F1** | **Kappa** | **MCC** |
| **1** | 0.967 | 0 | 0.9651 | 0.9799 | 0.9656 | 0.9662 | 0.9666 |
| **2** | 0.967 | 0 | 0.969 | 0.9789 | 0.966 | 0.9662 | 0.9666 |
| **3** | 0.967 | 0 | 0.969 | 0.9771 | 0.9653 | 0.9662 | 0.9666 |
| **4** | 0.9451 | 0 | 0.9496 | 0.967 | 0.9462 | 0.9437 | 0.9443 |
| **5** | 0.956 | 0 | 0.9535 | 0.956 | 0.9491 | 0.955 | 0.9557 |
| **6** | 0.956 | 0 | 0.9574 | 0.9698 | 0.9535 | 0.955 | 0.9554 |
| **7** | 0.9111 | 0 | 0.9147 | 0.9185 | 0.9037 | 0.909 | 0.91 |
| **8** | 0.9444 | 0 | 0.9419 | 0.9593 | 0.9415 | 0.9431 | 0.9436 |
| **9** | 0.9333 | 0 | 0.9341 | 0.9426 | 0.9333 | 0.9317 | 0.9319 |
| **10** | 0.9444 | 0 | 0.9496 | 0.963 | 0.9444 | 0.9431 | 0.9437 |
| **Mean** | 0.9492 | 0 | 0.9504 | 0.9612 | 0.9469 | 0.9479 | 0.9484 |
| **SD** | 0.0168 | 0 | 0.0161 | 0.018 | 0.0178 | 0.0172 | 0.017 |

LDA works on probability of the inputs to each class and then the one with the highest probability is the output class.So the model gave the highest accuracy in this case and predicted the person who is actually wearing the mask.

**After comparing all models, the results are recorded in Table2.**

Table 2: Modal Accuracy Comparison Table

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Model** | **Accuracy** | **AUC** | **Recall** | **Prec** | **F1** | **Kappa** | **MCC** | **TT(Sec)** |
| **1** | Linear Discriminant Analysis | 0.9481 | 0 | 0.9492 | 0.9586 | 0.9451 | 0.9468 | 0.9473 | 0.051 |
| **2** | SVM - Linear Kernel | 0.8499 | 0 | 0.8465 | 0.8742 | 0.8373 | 0.8462 | 0.848 | 0.2092 |
| **3** | Extra Trees Classifier | 0.8388 | 0 | 0.8345 | 0.8645 | 0.8282 | 0.8349 | 0.8364 | 0.3894 |
| **4** | Ridge Classifier | 0.8334 | 0 | 0.8283 | 0.8479 | 0.8151 | 0.8293 | 0.8311 | 0.0123 |
| **5** | K Neighbors Classifier | 0.8123 | 0 | 0.8109 | 0.8257 | 0.7957 | 0.8077 | 0.8094 | 0.0141 |
| **6** | Logistic Regression | 0.7958 | 0 | 0.7907 | 0.8091 | 0.7764 | 0.7907 | 0.7928 | 0.207 |
| **7** | Naive Bayes | 0.7958 | 0 | 0.7965 | 0.8295 | 0.7828 | 0.7908 | 0.7929 | 0.0115 |

Many papers are published on face recognition and detection but ours is a different and unique approach for detection with a mask. After successful experimentation on several algorithms and with various approaches we have found that Linear Discriminant Analysis is performing considerably well in comparison to others.The top 5 models Linear Discriminant Analysis,SVM-Linear Kernel,Extra Trees Classifier,Ridge Classifier,K Neighbors Classifier gave accuracies of: 94.04,84.99,83.38,83.33,81.23 percent respectively.

# **7. Conclusion:**

After experimenting with several approaches as discussed above, We can finally conclude that use of LDA is better choice for classification of embedding and also it is much better to take above half face specifically focusing on eye and forehead features and which helps our model to outperform all existing approaches that detect masked face images.and getting much better accuracy of 94% and add on to that our work is independent of the mask shape ,size, color, design etc and also our model will predict more accurately irrespective of the person is having a mustache or beard. Future scope of our system is to work with more datasets and try for getting better accuracy.

# **8. References:**

[1] Wendy S. Yambor Bruce A. Draper J. Ross Beveridge, “Analyzing PCA based Face Recognition Algorithms: Eigenvector Selection and Distance Measures”, July 1, 2000.

[2] Wendy S. Yambor, “Analysis of PCA Based and Fisher Discriminant- Based Image Recognition Algorithms”, M.S. Thesis, July 2000

[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711– 720.

[4] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg, ―Face Recognition by Elastic Bunch Graph Matching‖, ―Intelligent Biometric Techniques in Fingerprint and Face Recognition‖, eds, L.C . Jain et al., publ. CRC Press, ISBN 0-8493- 2055-0, Chapter 11, pp. 355-396, (1999) [7] T. Ojala, M. Pietik¨ainen, and D. Harwood

[5] S. Ge, J. Li, Q. Ye and Z. Luo, "Detecting Masked Faces in the Wild with LLE-CNNs," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 426-434, doi: 10.1109/CVPR.2017.53.

[6]. Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.

[7]. M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Proceedings of European Conference on Computer Vision, Springer, Zurich, Switzerland, 2014, pp. 818–833.

[8] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503, 2016.

[9]. S Yang, P. Luo, C. Chen and X. Tang. WIDER FACE: A Face Detection Benchmark. In CVPR, 2016.

[[10] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” *2015 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR*, pp. 815–823, Jun. 2015, doi: 10.1109/CVPR.2015.7298682.](https://www.zotero.org/google-docs/?q1HcHk)

[[11] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015, Accessed: Jul. 22, 2020.](https://www.zotero.org/google-docs/?q1HcHk)

[12] Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7\_12.

[13] M. S. Ejaz and M. R. Islam, "Masked Face Recognition Using Convolutional Neural Network," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/STI47673.2019.9068044.

[14] S. J. D. Prince and J. H. Elder, "Probabilistic Linear Discriminant Analysis for Inferences About Identity," 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, 2007, pp. 1-8, doi: 10.1109/ICCV.2007.4409052.