Indonesian Journal of Electrical Engineering and Computer Science
Vol. 34, No. 1, April 2024, pp. 302~311
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v34.i1.pp302-311 ❒ 302

Masked facial recognition using ensemble convolutional neural network and grey-level co-occurrence matrix

Om Pradyumana Gupta[1], Arun Prakash Agrawal[1], Om Pal[2]
[1]Department of Computer Science and Engineering, Sharda University, Greater Noida, India
[2]Department of Computer Science, University of Delhi, New Delhi, India

Article Info

Article history:

Received Oct 7, 2023
Revised Dec 26, 2023
Accepted Jan 3, 2024

Keywords:

CNN
DenseNet-121
GLCM ANN
LR
MRMR
SVM
VGG-16

ABSTRACT

The COVID-19 pandemic proved how face masks became necessary to stop the spread of infection. Due to this, effective identification of people wearing face mask became challenging. Masked facial recognition has significantly increased in accuracy because of developments in convolutional neural networks (CNNs). Small size of the dataset of masked facial images has been a problem in earlier research. As would be expected, this results in poorer accuracy when the model tries to identify faces. In this study, a novel model is proposed with textural feature extraction using grey-level co-occurrence matrix (GLCM) and an ensemble of two pre-trained CNNs DenseNet-121 and VGG-16. Using the minimum redundancy and maximum relevance, the model has improved accuracy by choosing the most important features of the image. The model was trained using in-house dataset that included 38,290 photos of 2,500 people with approximately equal distribution of properly masked, partially masked, and unmasked images. In this, we evaluated the performance of the model on different classifiers multi-class logistic regression (LR) and support vector machine (SVM) with one-vs-rest (OvR) classification and artificial neural network (ANN) and applied a soft voting scheme. The model achieved the highest accuracy of 98.56% at a learning rate of 0.001 on the ANN classifier.

Corresponding Author:

Om Pradyumana Gupta
Department of Computer Science and Engineering, Sharda University
Greater Noida, India
Email: op.gupta@nic.in

1. INTRODUCTION

With the recent emergence of COVID-19 pandemic across the world, detection, identification, and recognition of a human’s face have become important and critical challenges. The pandemic has enforced face masks as a critical element to avoid the spread of infection amongst humans. The use of face masks has reduced the overall accuracy of the existing face detection and recognition systems which has become a threat to other security applications. This has brought the emerging need for enhancements in the existing systems and to explore new solutions to tackle the problem of masked face recognition. Deep learning has been extensively used to enhance facial recognition and address the shortcomings of traditional techniques [1]. Convolutional neural networks (CNNs) are key components of deep learning-based implementations. Rather than manually selecting the features, deep learning CNNs may automatically identify higher-level features from data. Small CNNs were trained to learn DeepID [2], which had a 97.45% recognition accuracy. Each CNN receives the facial image's component parts such as the mouth, nose, and eyes independently so that it may be trained. Then, the learnt features are integrated to create a possible model. The other deep learning-based face recognition algorithms, such as deep face identification [3] were able to identify faces with an accuracy of 98.95% in the labeled faces in the wild (LFW) dataset [4] and 97.3% in the YouTube faces (YTF) dataset [5].

Previous deep learning-based techniques to facial recognition have been shown to have several drawbacks, including their dependency on vast volumes of data and powerful computing facilities. Deep transfer learning can be used to get around this deep learning issue [6]. In the field of machine learning, transfer learning is the process of using a previously trained model as the basis for a new assignment. In transfer learning, the model uses the knowledge it learned from its first training to enhance its performance and learning on a new, related task. Applying transfer learning can be especially helpful in situations where the new work has different data distribution from the original job or has limited annotated data. Additionally, transfer learning can improve the accuracy and generalization of the mode while saving time and processing resources by reusing the pre-trained model [7]. These state-of-the-art methods' (SOTA) use of fundamental datasets, where a limited number of parameters are present to evaluate their model, is a restriction. The suggested model makes use of the ideas of transfer learning and ensemble learning. Ensemble learning averages the weights of several deep learning models to improve recognition accuracy. Local binary pattern (LBP) was put out in various studies [8]–[11]. To extract features from the face that relate to texture, utilize the LBP operator. To further extract features for training, network parameter optimization, and achieving a classification result after the layer is fully connected (FC), ten CNNs with five different network topologies are used. A majority vote is used to determine the outcome of face recognition when employing the parallel ensemble learning method. CNNs apply filters to the input data to extract specific features. These filters slide across the input, producing feature maps that capture patterns and relationships within the data [12]–[14].

Although latest work and research paper offer significant advances in facial recognition, in this research, the authors concentrate on creating a "Masked face detection and recognition system" to address the issues of low accuracy in current face recognition systems due to the use of face masks. To the best of our knowledge, this is the first paper in masked facial recognition with ensemble CNN learning. The proposed model is extremely critical in cases where security is of utmost concern since the model provides very high accuracy but has high computational processing as well. It is a perfect solution for airports, banks, hospitals and law institutions. The method that is proposed in this paper first recognizes the masked face by detecting it in the uploaded image and then maps it to the training image datasets. The model is novel having textural feature extraction with grey-level co-occurrence matrix (GLCM) and an ensemble of two pre-trained CNNs-DenseNet-121 and VGG-16. Using the minimum redundancy and maximum relevance (mRMR), the model has improved accuracy by choosing the most important features of the image. The model was trained using in-house dataset that included 38,290 photos of 2,500 people with approximately equal distribution of properly masked, partially masked, and unmasked images. In this, we evaluated the performance of the model on different classifiers-multi-class logistic regression (LR) and support vector machine (SVM) with one-vs-rest (OvR) classification and artificial neural network (ANN) - and applied a soft voting scheme, ensuring that the weaknesses of individual classifiers are minimized. With the soft voting scheme, the model achieved the highest accuracy of 98.56% at a learning rate of 0.001 on the ANN classifier. The main contributions of this paper are, feature extraction with the ensemble CNN and GLCM, extraction of best features with mRMR, and classification with voting across SVM, ANN, and LR.

The paper is structured as follows. The methodology is detailed in section 2 in which pre-processing layer, ensemble CNN, GLCM, mRMR, classification layer, voting schemes, and experimental setup are detailed. In section 3, describes the experimental result analysis, comparing proposed model with other models. Conclusion and future works is described in section 4.

2. METHOD

The pre-trained DenseNet-121 and visual geometry group (VGG)-16 CNN as well as the GLCM for texture extraction have been used by the authors for face detection and identification. Due to the variety of human faces, including position, illumination, and emotion, face identification is a difficult process. Deep learning has recently demonstrated considerable potential for face recognition, with models like VGG-16 and DenseNet-121 obtaining cutting-edge outcomes. In Figure 1, we describe the methodology used to identify people using an ensemble CNN and GLCM for texture classification.

VGG-16 architecture and DenseNet-121 architecture make up the overall CNN. In this ensemble CNN, we have replaced the final classification layer with a 1,024 features output from the Dense layer of each CNN as input to the mRMR. In a similar manner, 80 of the 144 retrieved characteristics for GLCM were chosen. The 160 most crucial features are then extracted from by the mRMR from each CNN output (1,024 features) and GLCM output (144 features) to accurately classify the faces. This feature selection is done to save computational time and minimize the impact of irrelevant features that may lead to a slower classification, affecting the performance of the model. The 400 total features were concatenated after the feature selection process. To fit the classification model, we have used 3 classifiers - ANN, multi-class LR and SVM with OvR methodology. Probabilities from each classifier were fed to a soft voting scheme, selecting the mean class probabilities of individual classifiers. The model was further evaluated on the test dataset. To build an efficient model, a 3×3 kernel size with a stride of 1 and 64 filters (VGG-16) and 128 (DenseNet-121) were used. The small number for kernel size and filters was used to avoid the problem of overfitting. Max pool layer has a 2×2 kernel size with a stride of 2, reducing the feature maps size and extracting salient features. A low dropout layer rate of 0.2 helped preserving the model's performance without the problem of overfitting. Leaky rectified linear unit (ReLU) activation function is preferred over the ReLU activation function as it fixes the dying ReLU problem. The model uses Adam optimizer and has a learning rate 0.001 which helps to prevent the model from converging too quickly and missing important features.

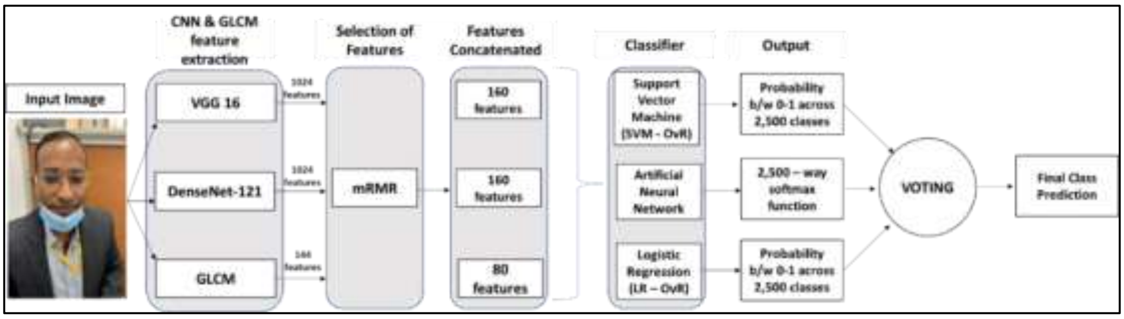

Figure 1. High level diagram of the proposed model

2.1 Dataset

The pre-trained, VGG-16 and DenseNet-121, CNNs on ImageNet dataset have very high accuracies. In this paper, an in-house dataset was developed of images of 2,500 people:38,290 facial images, of which 17,596 have proper masks, 11,263 have partial masks (improperly worn), and 9,431 have no masks. With a right mix of different type of images and with the large number of pictures, the dataset was ideal avoiding the problems of overfitting and biased learning. An 80:20 ratio was finalized for training and testing data. Number of images by gender and age is presented in Table 1:

Table 1. Dataset represented by gender and age-group

Age-group	Total	Male	Female
18-25	10,338	5,479	4,859
26-40	8,041	3,699	4,342
41-60	14,550	8,002	6,548
60+	5,361	3,109	2,252

2.2. Pre-processing layer

The pre-processing layer includes following steps: i) removal of noise from the input face image using various techniques such as salt and pepper technique as shown in Figure 2. This technique helps in detection of the noise as well as smoothening of the noise using non-local switching filter method. The smoothening of the image is done using gaussian filtering techniques to improve the visual quality of the image. The below block diagram represents the process involved in removal of noise. ii) the size of the image dataset is augmented by performing operations such as image rotation, flipping of the image.

2.3. Ensemble CNN

CNNs are capable of accurately classifying images on their own. However, the combined efforts of several CNNs can produce better outcomes than those of any one CNN alone [15]–[22]. Trained on ImageNet dataset, Two CNNs in the current model are loaded with weights. VGG-16 and DenseNet-121, two pre-trained CNNs created in Python using Tensorflow and Keras, were employed for masked face classification. In this ensemble CNN, we removed the classification layer and added a Dense layer that outputs 1024 features as shown in Figure 3.

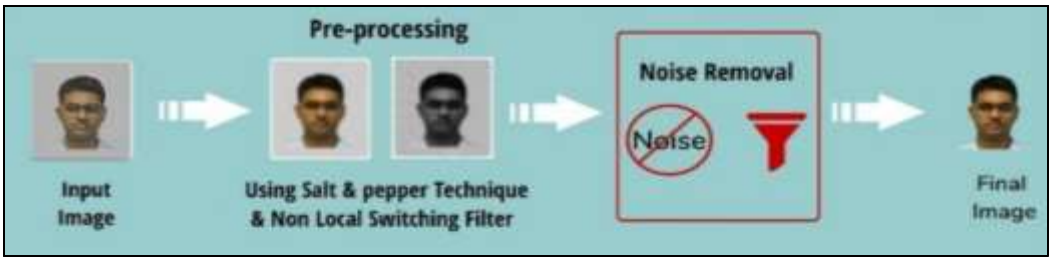

Figure 2. Pre-processing using salt and pepper technique

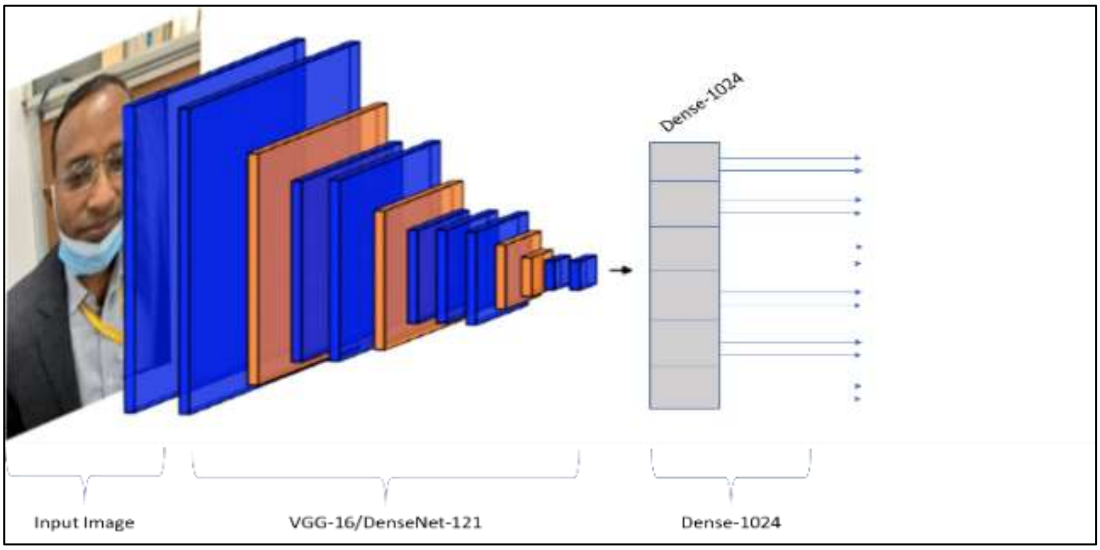

Figure 3. Architecture of modified VGG-16 and DenseNet-121 CNN

Due to their ability to learn from one another's failures, ensemble models are usually more accurate than individual models. The VGG-16 and DenseNet-121 models each have unique advantages and disadvantages. Face detection under challenging situations, such as dim lighting or occlusions, is improved by VGG-16. DenseNet-121 performs better at recognizing faces in a variety of positions and expressions and, generally, has higher accuracy because of the large number of neural networks but is computationally expensive to train. An ensemble can attain greater accuracy than each model alone by combining the benefits of these two models.

Second, compared to individual models, ensemble models are less likely to overfit the training data. This occurs as a result of the ensemble's models' training on various subsets of the data. This aids in avoiding overfitting by preventing the models from picking up on the noise in the training set. Thirdly, compared to individual models, ensemble models are more resistant to noise and outliers. This is because it is less likely that a single outlier will have an impact on each model in the ensemble. In real-world applications, where the input data may be erratic or include outliers, ensemble models are more dependable. Additionally, ensemble models are more computationally effective than training two models separately. This is possible due to the parallel training of the models and the ability to combine the predictions of the several models. Because of this, ensemble models are an excellent option for applications with limited computational resources.

2.4. Grey level co-occurrence matrix

A GLCM is a concise way to indicate how frequently a specific pair of pixels appears in a picture along a specific direction and spaced a specific amount apart from one another [23]. Its function is to make it possible to compute textural attributes (contrast and homogeneity) of the image. GLCM has various advantages that complement those of CNN. The lighting condition does not impact the number of features extracted from the image. Focusing on the statistical relationships between the image pixel's grey levels, GLCM extracts such features as correlation, contrast, homogeneity, entropy, and energy of the image. These features help the model to be more accurate, especially when the face is occluded with mask. A co-occurrence matrix with columns j and rows i for a scale image with 256 different shades of grey, P_{ij} will be of size 256×256. The value (i, j) in the matrix means how many times a pair of pixels (i, j) appears in the original horizontally i.e i and j should be adjacent to each other, as shown in Figure 4.

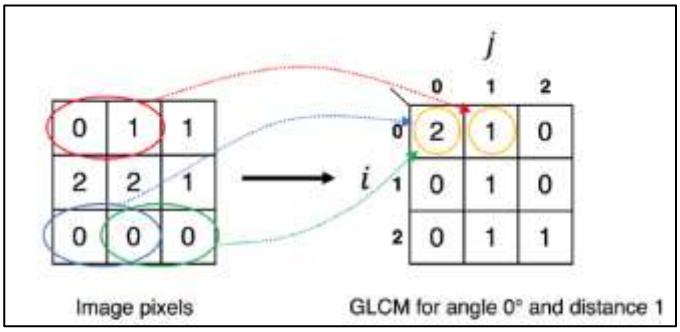

Figure 4. Computation of GLCM matrix in horizontal direction [15]

To maximize the information gained from each input image, six attributes were computed contract, dissimilarity, homogeneity, angular second moment (ASM), energy, correlation. These six attributes were calculated taking not only the horizontal direction but in eight different directions with three different distances. The resultant, 6×8×3 features (total 144) were the output of the GLCM, critical for the extraction of the textural features.

$$\text{Contrast}=\sum_{i,j=0}^{N_{levels}-1} P_{i,j}(i-j)^2 \quad (1)$$

$$\text{Dissimilarity}=\sum_{i,j=0}^{N_{levels}-1} P_{i,j}|i-j| \quad (2)$$

$$\text{Homogeneity}=\sum_{i,j=0}^{N_{levels}-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (3)$$

$$\text{ASM}=\sum_{i,j=0}^{N_{levels}-1} {P_{i,j}}^2 \quad (4)$$

$$\text{Energy}=\sqrt{ASM} \quad (5)$$

$$\text{Correlation}=\sum_{i,j=0}^{N_{levels}-1} P_{i,j}\left[\frac{(1-\mu_i)(j-\mu_i)}{\sigma_i.\sigma_{\sigma j}}\right] \quad (6)$$

2.5. Minimum redundancy and maximum relevance

When utilizing conventional (non-CNN) classifiers, the minimum redundancy maximum relevance approach seeks to choose the characteristics that are most crucial to classification [24], [25]. As there are fewer characteristics to examine and less opportunities for model misunderstanding, features that do not add value or were similar to other features are removed that has improved the computational performance and the accuracy of the model considerably. The mRMR works by extracting, at each iteration, the feature that is most relevant and most irrelevant. Those having the greatest F-test statistic are chosen in the initial iteration. As illustrated in (7), F-test correlation quotient (FCQ), the criteria for selection on following iterations, is computed.

$$Score\ of\ X_i = \frac{F(Y,X_i)}{\{\frac{1}{|S|}\sum_{X_s \in S} \rho(X_s,X_i)\}} \quad (7)$$

X_i represents the feature at the i[th] iteration which will be selected, $F(Y,X_i)$ is the feature's F-test statistic in relation to its matching class label. The set of previously chosen features is represented by Y, S and the $\rho(X_s,X_i)$ Pearson correlation coefficient of each feature against every X_s

2.6. Classification layer

For each image, once the ensemble CNN and GLCM output the 1024×2 and 144 features respectively, from every CNN, the most prominent 160 features and, from GLCM, 80 features are selected. Lastly, a set of 400 features is created for each input of the dataset by concatenating the generated features. The output was passed through an ANN.

2.6.1. Artificial neural network

ANNs refer to stacking neurons in various 1D layers, unlike that in deep neural networks (DNNs) which employ 2D layers. They are also referred to as feed-forward neural networks or multi-layer perceptron. They have been utilized successfully in the categorization of medical images, including the classification of cancerous skin lesions and lung nodules in computed tomography (CT) scans. As shown in Figure 5, the ANN was developed for the current study in Python's Keras package utilizing an input layer that was equal in length to the number of features (400), with five hidden Dense layers that each had 550 neurons, providing output to a 2,500-way SoftMax function for exact matches.

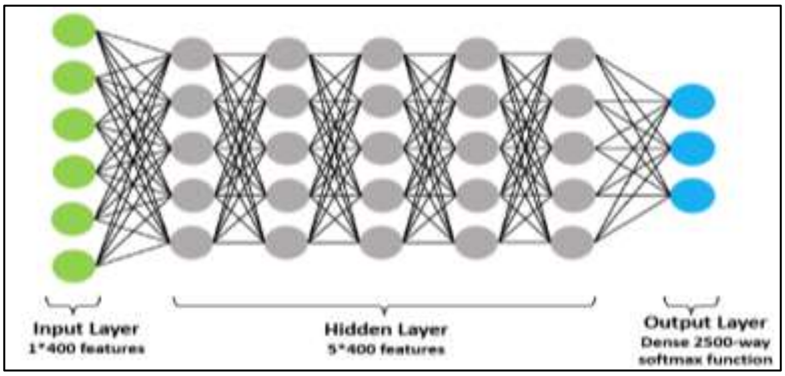

Figure 5. High level architecture of ANN

2.6.2. Logistic regression and support vector machine

LR and SVM are primarily used for binary classification. The idea of binary classification was applied to multi-class classification (2,500 in this study). We have used the OvR multi-class classification. The scikit-learn library implements the OvR strategy.

2.7. Voting scheme

Preserving the output of each classifier, a soft voting scheme was applied to reach the final output. With only three classifiers, hard voting was not reasonable as it could significantly impact the accuracy of the model. Figure 5 describes the two voting schemes, soft voting averages the probabilities of the predictions of the three classifiers, thereby minimizing the individual weaknesses and bias of each classifier in case of mutually exclusive predictions of the three classifiers.

$$k = mean(w_1 \times p_1(x) + w_2 \times p_2(x) + \ldots + w_n \times p_n(x)) \tag{8}$$

where:
k is t
he final output (final class label)
x is the input
w_i is the weight assigned to individual classifiers
p_i is the class probabilities of the individual classifier
In our model, w_i is same for all classifiers.

2.8. Experimental setup

As part of the experimental setup, python libraries such as Imutils, numpy, dlib, random, and OpenCV2, are used for face detection, pre-processing and face cropping. Other python libraries such as Keras, Tensorflow, Zipp, markdown, Scikit learn, and matplotlib, are used for training of the model. The combination of two different models namely, VGG-16 CNN model and DenseNet-121 model is used for the feature extraction from the masked face using deep learning CNN network. We have trained the model, having a batch size of 32, with the Adam optimizer. Table 2 shows the experimental setup details.

Table 2. Experimental setup details

Resource	Details
Processor	M1 Max Apple MacBook Pro CPU
Random access memory	32GB
Graphics processing unit	M1 Max
Language	Python

3. RESULTS AND DISCUSSION

Validation metrics as described in Table 3 abbreviations for true positive (TP), true negative (TN), false positive (FP), false negative (FN) are mentioned. Tables 4 and 5 displays the test image classification results for each dataset, and Figure 6 compares the classifier performances graphically on different configurations. The data in Tables 4 and 5 shows the comparison with other models including VGG-16, DenseNet-121, an ensemble of VGG-16 and DenseNet. In this paper, proposed model's performance is compared with that of the mentioned models on different classifiers: SoftMax, ANN, SVM, and LR. Although parameter adjustments were made, the performance was judged with the parameters at par for every model. The best accuracy of VGG-16, DenseNet, an ensemble of VGG-16 and DenseNet, and the proposed model across different classifiers is 92.07%, 93.72%, 93.70%, and 98.56%, respectively. Since creating a confusion matrix of 2,500×2,500 is difficult to demonstrate, a confusion matrix with the age-group classification, as shown in Figure 7, is proposed. For the age groups 18-25, 26-40, 41-60, and 60+, TP samples were identified by the proposed model across the age groups at 99.13%, 99.25%, 99.41%, and 98.21% respectively; TN samples were identified by the proposed model at 93.63%, 92.15%, 93.07%, and 95.29% respectively; FP samples were identified by the proposed model at 0.34%, 0.48%, 0.30%, and 1.46% respectively; FN samples were identified by the proposed model at 0.53%, 0.27%, 0.22%, and 0.34% respectively. The proposed model, when compared with the other models, showed the highest F1-score of 95.21% at the learning rate of 0.001.

Table 3. Validation metrics

Validation metric	Equation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
F1-Score	$\frac{2TP}{2TP+FP+FN}$

Table 4. Accuracy of different configurations of features and classifiers

Metric	VGG-16	DenseNet-121	VGG-16+DenseNet-121				
	Softmax	Softmax	ANN			SVM	LR
	Learning rates		0.01	0.001	0.05		
Accuracy	0.9207	0.9372	0.9218	0.9349	0.9162	0.9370	0.9131
Precision	0.9007	0.9142	0.9244	0.9012	0.9196	0.9519	0.9110
Sensitivity	0.9282	0.9135	0.9226	0.9144	0.9223	0.9661	0.9150
Specificity	0.9153	0.9175	0.9131	0.9254	0.9214	0.9276	0.9244
F1-score	0.9057	0.9426	0.9346	0.9238	0.9115	0.9017	0.9251

Table 5. Accuracy of different configurations of features and classifiers with soft voting

Metric	VGG-16	DenseNet-121	VGG-16+DenseNet-121+GLCM							
	Softmax	Softmax	Soft voting (ANN*+LR+SVM)			ANN*			SVM	LR
	Learning Rates		0.01	0.001	0.05	0.01	0.001	0.05		
Accuracy	0.9207	0.9372	0.9610	0.9856	0.9594	0.9501	0.9726	0.9442	0.9664	0.9219
Precision	0.9007	0.9142	0.9656	0.9760	0.9599	0.9546	0.9607	0.9537	0.9551	0.9104
Sensitivity	0.9282	0.9135	0.9623	0.9597	0.9444	0.9600	0.9533	0.9366	0.9683	0.9108
Specificity	0.9153	0.9175	0.9218	0.9823	0.9013	0.9128	0.9800	0.9253	0.9402	0.9154
F1-score	0.9057	0.9426	0.9427	0.9521	0.9311	0.9287	0.9391	0.9222	0.9121	0.9201

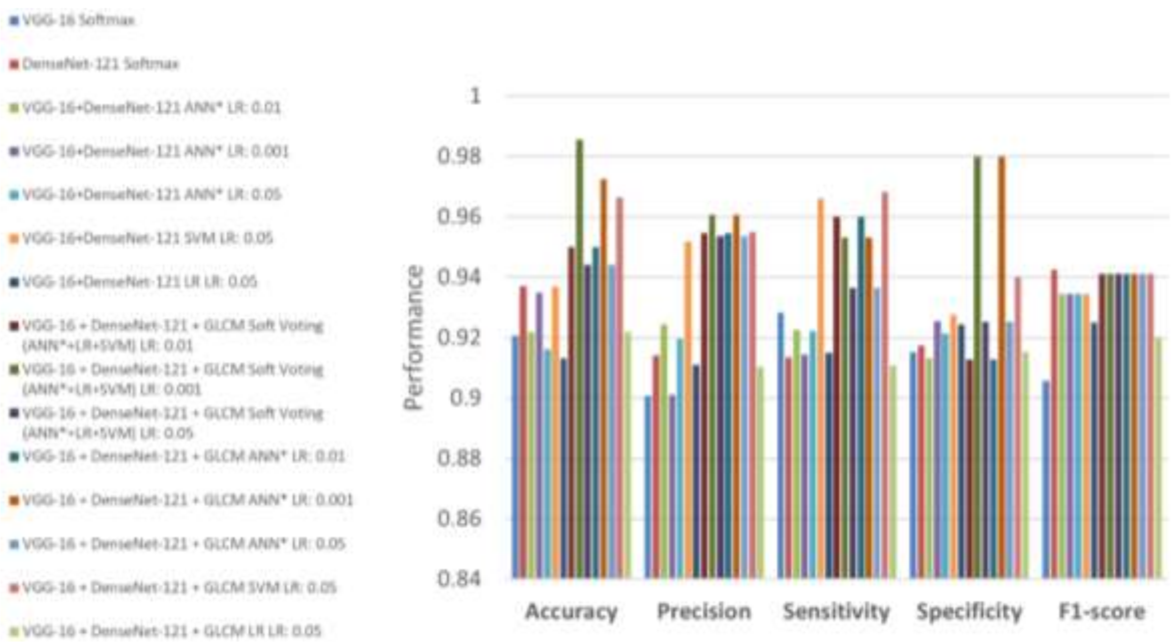

Figure 6. Performance of different configurations across classifiers

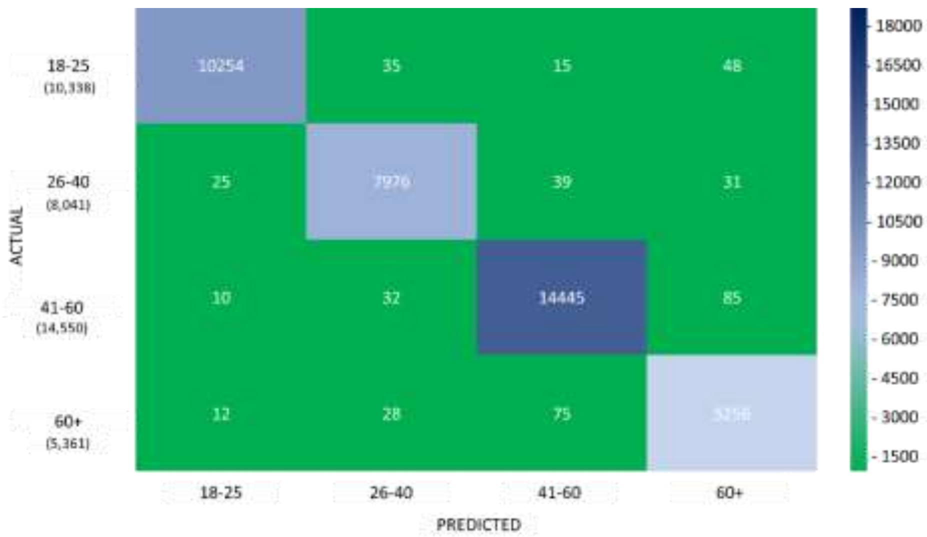

Figure 7. Confusion matrix of the proposed model

4. CONCLUSION

This work investigated various deep learning methods for images and clarified the advantages of merging CNNs to enhance classification performance. It was discovered that choosing mRMR features preserves pertinent picture data while speeding up calculation for to fit features. The features that were gathered from DenseNet-121, VGG16, and GLCM were integrated at 98.56% using the soft voting scheme that performed best with the ANN classifier at a learning rate of 0.001 to reach the dataset's highest classification accuracy. The ANN classifier was initiated at the learning rate of 0.01, however, overfitting was observed as we increased the learning rate from 0.001. Overfitting of data could also be reduced if the size of the dataset be increased, which could've resulted in better generalization for different configurations. Another limitation of the model was the low accuracy in correctly classifying images of people wearing mask of different types – fabric, pattern, and size. For One-vs-Rest (OvR) LR and SVM, it was observed that it was computationally expensive to train the model. A high degree of overfitting was also observed. Therefore, when developing new or enhanced masked face recognition classification models, it is advised to emphasize training with larger and diversified datasets. Another recommendation is to leverage Intersection over union

(IoU) for performance measurement of trains of object detection. DarkNet53_YOLOv3 may also be chosen as a better model with fast detection.

REFERENCES

[1] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," *Signal, Image and Video Processing*, vol. 16, no. 3. Research Square Platform LLC, pp. 605–612, 2022. doi: 10.1007/s11760-021-02050-w.

[2] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2014, pp. 1891–1898. doi: 10.1109/CVPR.2014.244.

[3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Procedings of the British Machine Vision Conference 2015*, British Machine Vision Association, 2015, pp. 41.1-41.12. doi: 10.5244/C.29.41.

[4] M.-W. Huang, Z. Wang, and Z.-L. Ying, "A new method for facial expression recognition based on sparse representation plus LBP," in *2010 3rd International Congress on Image and Signal Processing*, IEEE, Oct. 2010, pp. 1750–1754. doi: 10.1109/CISP.2010.5647898.

[5] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *CVPR 2011*, IEEE, Jun. 2011, pp. 529–534. doi: 10.1109/CVPR.2011.5995566.

[6] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Artificial Neural Networks and Machine Learning – ICANN 2018*, Springer International Publishing, 2018, pp. 270–279. doi: 10.1007/978-3-030-01424-7_27.

[7] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2020, doi: 10.1109/tai.2021.3054609.

[8] J. Tang, Q. Su, B. Su, S. Fong, W. Cao, and X. Gong, "Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105622, Dec. 2020, doi: 10.1016/j.cmpb.2020.105622.

[9] K. S. Reddy, V. V. Kumar, and B. E. Reddy, "Face recognition based on texture features using local ternary patterns," *International Journal of Image, Graphics and Signal Processing*, vol. 7, no. 10, pp. 37–46, Sep. 2015, doi: 10.5815/ijigsp.2015.10.05.

[10] K. Alhanaee, M. Alhammadi, N. Almenhali, and M. Shatnawi, "Face recognition smart attendance system using deep transfer learning," *Procedia Computer Science*, vol. 192, pp. 4093–4102, 2021, doi: 10.1016/j.procs.2021.09.184.

[11] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009, doi: 10.1016/j.imavis.2008.08.005.

[12] Y. Chen *et al.*, "Face mask assistant: detection of face mask service stage based on mobile phone," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 11084–11093, May 2021, doi: 10.1109/JSEN.2021.3061178.

[13] W. Ji and L. Jin, "Face shape classification based on MTCNN and FaceNet," in *2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, IEEE, Nov. 2021, pp. 167–170. doi: 10.1109/ICHCI54629.2021.00042.

[14] Y. Fan, J. C. K. Lam, and V. O. K. Li, "Multi-region ensemble convolutional neural network for facial expression recognition," in *Artificial Neural Networks and Machine Learning – ICANN 2018*, Springer International Publishing, 2018, pp. 84–94. doi: 10.1007/978-3-030-01418-6_9.

[15] D. Kuzinkovas and S. Clement, "The detection of COVID-19 in chest X-rays using ensemble CNN techniques," *Information*, vol. 14, no. 7, p. 370, Jun. 2023, doi: 10.3390/info14070370.

[16] S. Anwarul, T. Choudhury, and S. Dahiya, "A novel hybrid ensemble convolutional neural network for face recognition by optimizing hyperparameters," *Nonlinear Engineering*, vol. 12, no. 1, Jun. 2023, doi: 10.1515/nleng-2022-0290.

[17] A. Shah, B. Ali, M. Habib, J. Frnda, I. Ullah, and M. Shahid Anwar, "An ensemble face recognition mechanism based on three-way decisions," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 4, pp. 196–208, 2023, doi: 10.1016/j.jksuci.2023.03.016.

[18] E. S. Smitha, S. Sendhilkumar, and G. S. Mahalakshmi, "Ensemble convolution neural network for robust video emotion recognition using deep semantics," *Scientific Programming*, vol. 2023, pp. 1–21, May 2023, doi: 10.1155/2023/6859284.

[19] F. Bougourzi, F. Dornaika, N. Barrena, C. Distante, and A. Taleb-Ahmed, "CNN based facial aesthetics analysis through dynamic robust losses and ensemble regression," *Applied Intelligence*, vol. 53, no. 9, pp. 10825–10842, May 2023, doi: 10.1007/s10489-022-03943-0.

[20] J. Tang *et al.*, "The two-stage ensemble learning Model based on aggregated facial features in screening for fetal genetic diseases," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, p. 2377, Jan. 2023, doi: 10.3390/ijerph20032377.

[21] A. Iyer, S. S. Das, R. Teotia, S. Maheshwari, and R. R. Sharma, "CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings," *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 4883–4896, Feb. 2023, doi: 10.1007/s11042-022-12310-7.

[22] A. Banerjee, S. Roy, R. Kundu, P. K. Singh, V. Bhateja, and R. Sarkar, "An ensemble approach for still image-based human action recognition," *Neural Computing and Applications*, vol. 34, no. 21, pp. 19269–19282, 2022, doi: 10.1007/s00521-022-07514-9.

[23] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[24] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[25] Z. Zhao, R. Anand, and M. Wang, "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform," in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, Oct. 2019, pp. 442–452. doi: 10.1109/DSAA.2019.00059.

BIOGRAPHIES OF AUTHORS

Om Pradyumana Gupta is pursuing his Ph.D. in the field of Artificial Intelligence, Computer Vision and Deep Learning from Sharda University, Greater Noida, India. Earlier he has received degree of M.Tech. in Information Technology and MCA. Presently he is working as Scientist-E in National Informatics Centre, Ministry of Electronics and Information Technology, Government of India. He is involved in a number of projects of national level. Gov.in Secure Intranet, Dididhan Dashboard, Electronics Transaction Aggregation and Analysis Layer (eTaal) are among few of them. His area of interest includes artificial intelligence, computer vision and deep learning. He can be contacted at email: op.gupta@nic.in.

Dr. Arun Prakash Agrawal is currently working as a Professor with the Department of Computer Science and Application and Engineering at Sharda University, Greater Noida, India. He did Ph.D. and Masters in Computer Science and Engineering from Guru Gobind Singh Indraprastha University, New Delhi. He was the gold medalist of his Batch. Prior to his current assignment he has served many academic institutions including Amity University, Noida. He has also taught short term courses at Swinburne University of Technology, Melbourne, Australia and Amity Global Business School, Singapore. His research interests include machine learning, AI and soft computing. He can be contacted at email: arun.agrawal@sharda.ac.in.

Dr. Om Pal received B.E. in Computer Science and Engineering from Dr. B. R. Ambedkar University, Agra, MBA(O&M) from IGNOU, MS (Research) in field of Cryptography from IIT Bombay and Ph.D. in field of Cyber Security from Jamia Millia Islamia, New Delhi. Presently he is working as an Associate Professor in Department of Computer Science, University of Delhi (Central University), North Campus, Delhi. His area of interest includes cryptography, information security, cyber security, network security, post-quantum cryptography, quantum computing, blockchain technology, algorithms analysis and design, cyber law, artificial intelligence for cyber security, machine learning for cyber security, theory of automata, operation research and mathematics. He can be contacted at email: opal@cs.du.ac.in.