Journal of Intelligent Medicine and Healthcare
DOI: 10.32604/jimh.2022.033058

Article

Tech Science Press

CNN-LSTM Face Mask Recognition Approach to Curb Airborne Diseases COVID-19 as a Case

Shangwe Charmant Nicolas*

School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China
*Corresponding Author: Shangwe Charmant Nicolas. Email: charmantshangwe@yahoo.fr

Received: 06 June 2022; Accepted: 12 August 2022

Abstract: The COVID-19 outbreak has taken a toll on humankind and the world's health to a breaking point, causing millions of deaths and cases worldwide. Several preventive measures were put in place to counter the escalation of COVID-19. Usage of face masks has proved effective in mitigating various airborne diseases, hence immensely advocated by the WHO (World Health Organization). A compound CNN-LSTM network is developed and employed for the recognition of masked and none masked personnel in this paper. 3833 RGB images, including 1915 masked and 1918 unmasked images sampled from the Real-World Masked Face Dataset (RMFD) and the Simulated Masked Face Dataset (SMFD), plus several personally taken images using a webcam are utilized to train the suggested compound CNN-LSTM model. The CNN-LSTM approach proved effective with 99% accuracy in detecting masked individuals.

Keywords: COVID-19; face mask; airborne diseases; SARS-CoV-2; CNN; LSTM

1 Introduction

With an official death count of 171, COVID-19 was publicized in January 2020 as a public health threat with fever to cold, dire breathing ability, loss of taste and smell sense in some cases as the main symptoms [1], affecting the respiratory system as well as kidneys and the Liver [2]. The COVID-19 pandemic has claimed a huge number of lives, 6,152,095 deaths of 489,779,062 confirmed cases as of April 2022 reported to the World Health Organization globally.

To limit the further escalation of COVID-19, extreme measures had to be taken; such as local lockdowns, nationwide lockdowns, social distancing (standing 6ft apart), mask mandates, and hand hygiene have been implemented as primary Covid counter stratagems [3]. Several studies were done to evaluate the efficacy of facemasks in the fight against COVID-19. In Beijing ménages, a study investigated the depletion of secondary spread of SARS-CoV-2 by facemask employment [4]. It descried that face masks were 79% helpful in fending off transmission if used by all ménage members before the occurrence of symptoms though the study did not investigate the sorts of masks. Leffler et al. [5] investigated the correlation between facemask use and the diffusion of SARS-CoV-2 using a multiple regression approach. Their study noticed that SARS-CoV-2 diffusion was multiple times higher in countries a facemask decree had not been put in place equivalent results were noticed in other studies in some countries [6].

Be it well noted that facemask usage as a pandemic prevention tool is not a 20th or 21st-century invention. Wu Lien Teh recognized the cloth mask, which he designed immensely used in 20th-century all over the world as the chief mean of personal protection [7]. Although Wu designed it, he stated that the airborne diffusion of plague was notorious in the 13th century. A century later, face coverings started being employed for airborne disease protection [8].

2 Related Works

All the above and more studies supported putting in place social distancing and face mask mandates. Researchers and developers have hopped on the opportunity to create systems to assist the compliance of the set rules during the fight against COVID-19.

Using a Deep CNN model created from a convolutional neural network (CNN) structure with a slight kernel value and two fronts that operate in a variety of scene settings with no prior familiarity, Jarraya et al. [9] proposed a crowd monitoring system for social distancing in public places of high and low concentrations. Sonn et al. [10] illustrated the role of intelligent cities in countering the coronavirus diffusion in South Korea. Hastening the city's communication survey together with patient displacement, transaction history, cell phone use, and locating by a time-space cartographer. Rahman et al. [11] designed a model based on CNN, capable of detecting an unmasked individual with a 98% accuracy and informing law enforcement to deter the diffusion of COVID-19.

Ge et al. [12] suggested detecting masked individuals using an LLE-CNN network, manifesting them with high dimensional descriptors after fusing pre-trained CNNs to extract relevant facial regions. Using the Viola-Jones algorithm, Fan et al. [13] developed a deep learning-based feather-light facemask detection system to satisfy embedded systems' low computational necessities, introducing the SL-FMDet that operated commendably because of its little hardware obligations Using the nearest neighbor (NN) classifier distance for face recognition, Ejaz et al. [14] developed PCA for face portion, masked and unmasked facial image detection, and for working out Eigenface simultaneously.

The COVID-19 pandemic has drastically boosted research hunger in the domain of MFR as a whole that protracted existing methods like OFR and attained outstanding accuracy results by a significant margin. Above all, deep learning methodologies have gradually been advanced to counter MFR trials, as illustrated in Fig. 1 below [15].

3 Methods and Materials

This study is carried out in 5 essential stages. Starting with data collection, data pre-processing, and data division into training, validation, and testing sets are the subsequent steps. Model development is the next step, based on a combination of CNN and LSTM architecture, and model evaluation is the last step, used to assess how well the produced model performs in terms of accuracy, recall, precision, and F1-Score ranking. A fair representation of all classes in the testing set was taken into account throughout the experiment, and the Test set was deliberately set aside for the model evaluation stage. The entire work is summarized in Fig. 2 below, which shows the process in a left to right direction.

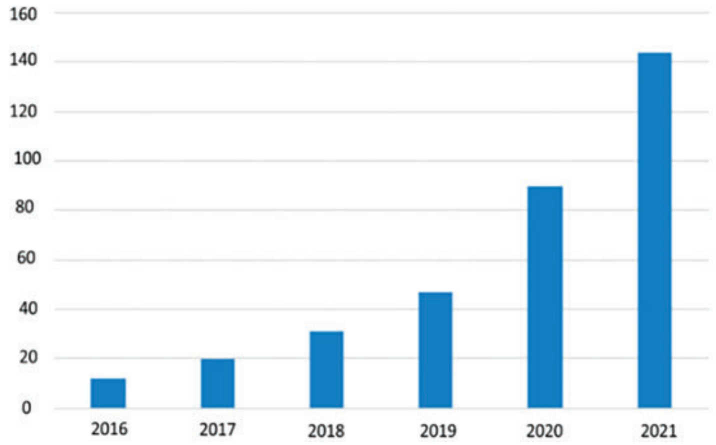


Figure 1: Demonstrates a hike in research endeavors in the MFR field since the beginning of the pandemic

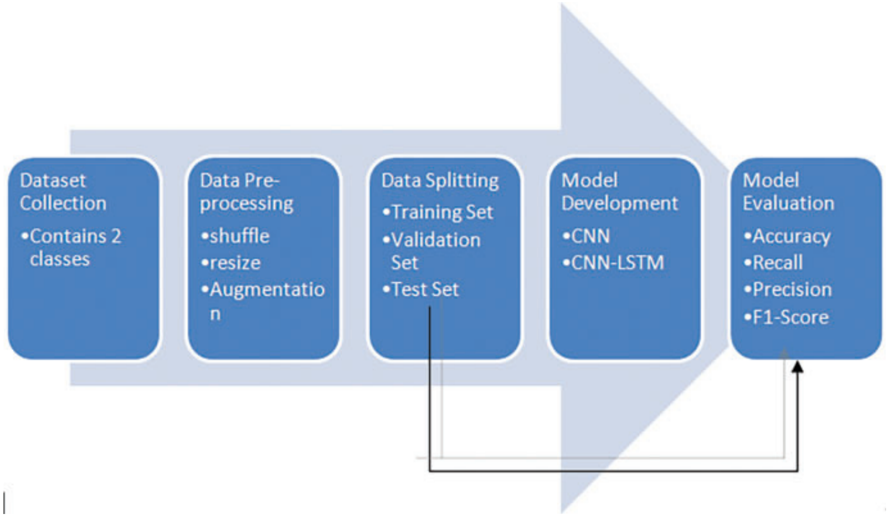


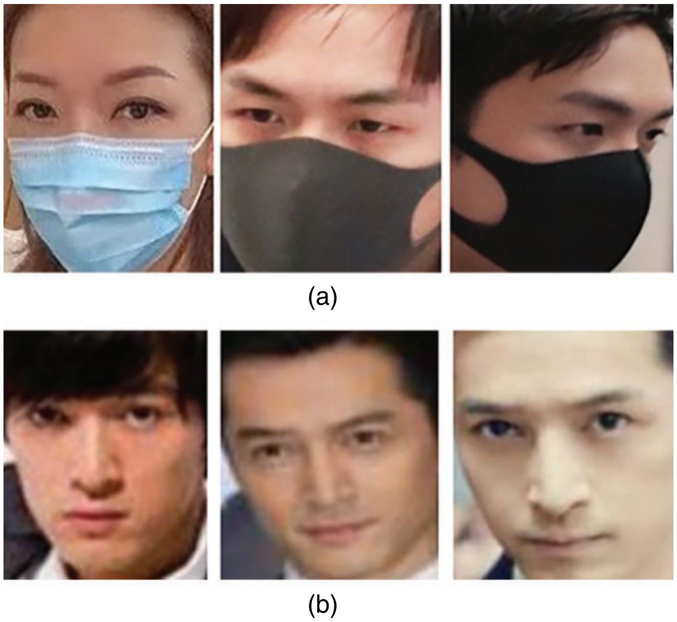
Figure 2: The overall study process

4 Dataset Collection and Description

Using publicly available datasets, the dataset employed in this research endeavor was generated 3833 RGB images in total, including 1915 masked and 1918 unmasked images sampled from a combination of the RMF Dataset (publicly available on GitHub) and the SMF Dataset (publicly available on Kaggle), plus several personally taken images using a webcam, dataset sample is showcased in Fig. 3. The images were split and used for the model's validation, training, and testing purposes throughout this endeavor.

4.1 Data Preprocessing and Augmentation

Due to the fact that elements of the dataset are not of the equivalent size, it is essential for resize to come into play. Employing Keras' Image Data Generator method to resize all images to 128 × 128 size. Images are later converted to NumPy arrays and later normalized in the range [0–255] to make the model converge faster. Data augmentation is done on the dataset by rescaling, rotation, width shift, height shift, shear, zoom, vertical, and horizontal flip by applying the corresponding Keras method.



(a)

(b)

Figure 3: Dataset samples (a) Masked (b) Unmasked

5 Proposed Method

A fusion of Convolutional Neural Network (CNN) and LSTM (Long Short Memory) for facemask detection is proposed in this paper and described below.

5.1 Convolutional Neural Network

Many researchers have admired the use of Convolutional Neural Network (CNN) algorithm designs for Segmentation, Image Classification, and Object Recognition. Generally, a typical CNN structural design comprises three main components (layers), namely: Convolutional, pooling, and fully connected layers, in that order. To acquire deep features, which is the architecture's primary goal, the inputs are scanned by the initial layer (Convolutional) through a filter. To condense the computational burden, the pooling layer offers the selection of more significant features.

Lastly, the fully connected layer flattens the inputs and calculates the probabilities of the labels.

A tensor of feature maps is defined by a batch of kernels [16] included in the stated above Convolutional layer. By the use of strides, the mentioned kernels convolve the input completely turning the output volume dimensions to integers in the process [17].

Zero-padding [18] is vital to pad the input volume with zeros to preserve the input volume dimension with low-level features since for the striding process the input volume dimensions reduce after the convolution layer. The convolutional layer maneuver is mathematically illustrated below as:

$$G\,(\mathrm{i},\mathrm{j}) = (\mathrm{E} * \mathrm{K})\,(\mathrm{i},\mathrm{j}) = \Sigma\Sigma\ \mathrm{E}\,(\mathrm{i}+\mathrm{t},\mathrm{j}+\mathrm{p})\,\mathrm{K}\,(\mathrm{t},\mathrm{p}) \tag{1}$$

where E stands for the input matrix, K denotes a size t × p 2D filter, and 2D feature map output represented by G. E ∗ K denotes the convolution operation. The Rectified Linear Unit (ReLU) layer is employed to aggregate the nonlinearity in the feature maps (mages are naturally non-linear). By maintaining the threshold of the input at zero, ReLU works out the activation, illustrated graphically in Fig. 4. Given mathematically as:

$$g(x) = \max(0, x) \tag{2}$$

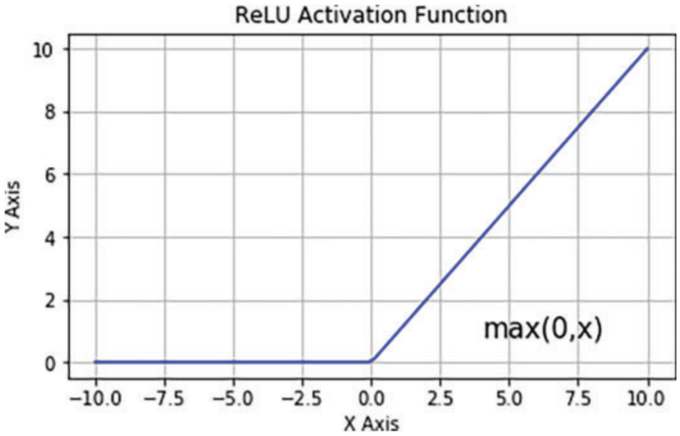


Figure 4: ReLU activation function graph

The pooling layer [19] comes next to diminish the number of parameters by the dint of down sampling the input dimension. Being one of the notorious methods, Max pooling yields the topmost value in the input region. Based of acquired features from the above-mentioned layers namely: pooling and convolutional layers, the fully connected layers [20] is later employed as a classifier to make a judgment. A whole CNN structural design can be built up with the above-discussed layers, graphically as Fig. 5.

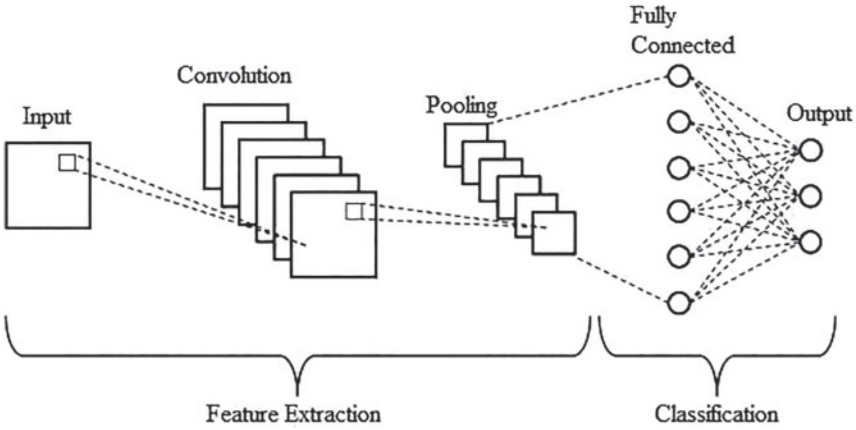


Figure 5: Graph representation of a CNN structural design

5.2 Long Short-Term Memory

Usually called "LSTMs"–Long Short Term Memory–an advance of RNN, Hochreiter et al. (1997) [21] instigated fit for learning long-term dependencies. They later gained notoriety due to the refinement work of various people [22,23]. Their immense popularity is based on their outstanding performance-proven solving various problems.

LSTMs are unequivocally aimed to sidestep the long-term dependency problem. Retention of information for extensive periods of time is their inherent behavior conduct, a task they do with ease.

To work out exploding and vanishing gradient glitches [24], memory blocks are suggested by the LSTM in spite of traditional RNN units hence an upgrade of the RNN. An LSTM network is able to recall and link former information to the extant information [25].

LSTM augmented a cell state to save long-term states distinguishing them from RNN. The design of an LSTM network is illustrated in Fig. 6. The LSTM associates the output, forget, and input gates. Where n*t* denotes the current input, Zt represents the new and later Z$t-1$ represents the previous cell state, the current and previous outputs are denoted by *Ot* and Ot_{-1} in that order. The LSTM's input gate mathematical principle is presented next form.

$$\mathrm{dt} = \varphi\,(\mathrm{Mi}\ \cdot [\mathrm{O}t - 1,\ \mathrm{nt}] + \mathrm{pi}) \tag{3}$$

$$\mathrm{Zt} = \tanh\,(\mathrm{Mi}\ \cdot [\mathrm{O}t - 1,\ \mathrm{nt}] + \mathrm{pi}) \tag{4}$$

$$\mathrm{Zt} = \mathrm{yt}\ \mathrm{Zt} - 1 + \mathrm{dt}\ \mathrm{Zt} \tag{5}$$

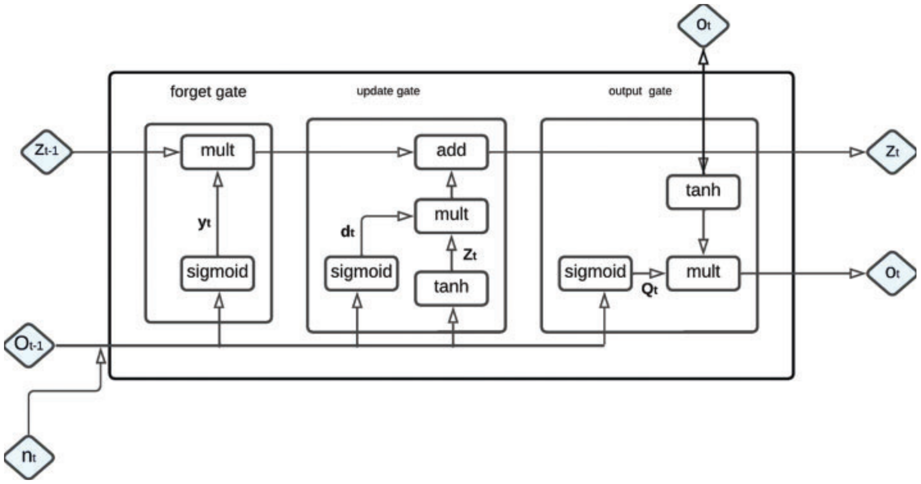


Figure 6: Long short term memory graph representation

To elect the quota of information to be added, Ot_{-1} and n*t* are delivered via the sigmoid layer. Afterward, novel information is obtained at (4) once Ot_{-1} and n*t* are delivered via the tanh layer. In (5) Zt () and Z$t-1$ (long-term memory information) are joined into Zt with d*t*, Zt representing the sigmoid, and the tanh outputs respectively, with Mi and pi representing weight matrices, LSTM's input gate bias in that order. Then by a dot product and a sigmoid layer, the forget gate of LSTM permits the discerning information passage. With a given likelihood, the judgment to forget the former cell's associated information is carried out with (6), the weight matrix, the offset, and the sigmoid function represented by Mf, sf, φ respectively.

$$yt = \varphi\,(\mathrm{Mf}\ \cdot [\mathrm{Zt} - 1,\ \mathrm{nt}] + \mathrm{sf}) \tag{6}$$

After (7) and (8), the essential states to be carried on by the Ot_{-1} and nt inputs are defined by the output gate of the LSTM. The concluding output is found to multiply with the state decision vectors delivering novel information through the tanh layer.

$$\mathrm{Qt} = \varphi\,(\mathrm{Mo}\ \cdot [\mathrm{Q}t - 1,\ \mathrm{nt}] + \mathrm{ro}) \tag{7}$$

$$Ot = \mathrm{Qt}\tanh(\mathrm{Zt}) \tag{8}$$

With $\mathrm{M_o}$ denoting the output gate's weighted matrices and $\mathrm{r_o}$ representing LSTM's output gate bias.

5.3 Fused CNN-LSTM Network

This study developed a combination of CNN and LSTM networks to differentiate masked individuals from unmasked personnel. In this developed architecture, CNN was employed for feature extraction purposes from images and LSTM was employed for the classification role. Table 1 below illustrates the proposed architecture.

Table 1: Architecture of the proposed CNN-LSTM network

Layers	Type	Kernel size	Strides	Kernel	Rate	Input size
1	Conv2D	3 × 3	1	64	-	128 × 128 × 3
2	MaxPooling	2 × 2	2		-	126 × 126 × 64
3	BatchNormalization	-	-		-	63 × 63 × 64
4	Conv2D	3 × 3	1	64	-	63 × 63 × 64
5	MaxPooling	2 × 2	2		-	61 × 61 × 64
6	BatchNormalization	-	-		-	30 × 30 × 64
7	Conv2D	3 × 3	1	128	-	30 × 30 × 64
8	MaxPooling	2 × 2	2		-	28 × 28 × 64
9	BatchNormalization	-	-		-	14 × 14 × 128
10	Conv2D	3 × 3	1	128	-	14 × 14 × 128
11	MaxPooling	2 × 2	2		-	12 × 12 × 128
12	BatchNormalization	-	-	-	-	6 × 6 × 256
13	Conv2D	3 × 3	1	256	-	6 × 6 × 256
14	MaxPooling	2 × 2	2		-	4 × 4 × 256
15	BatchNormalization	-	-	-	-	2 × 2 × 256
16	Reshape				-	2 × 2 × 256
17	LSTM				-	4 × 256
18	BatchNormalization	-	-	-	-	4 × 256
19	Flatten	-	-	-	-	4 × 256
20	Dense	-	-	128	-	1024
21	Dropout	-	-	-	0.4	128
22	Dense	-	-	64	-	128

(Continued)

Table 1: Continued

Layers	Type	Kernel size	Strides	Kernel	Rate	Input size
23	Dropout	-	-	-	0.4	64
24	Output	-	-	1		64

The input shape requires a size of 128 × 128 pixels. For the training cost not to be augmented, Convolution layers were kept to 5 layers max. Input Images of size 128 × 128 × 3 go through the initial five convolution layers, with 64, 64,128,128, and 256 filters, respectively. A (3 × 3) kernel and in the convolution layer, the ReLU function is employed as the activation function plus a stride of 1. Further, a pooling layer (MaxPooling) of stride 2 and kernel size (2 ∗ 2) follows each of the convolution layers, and finally, a Batch Normalization layer follows. For time information to be extracted, the function map is passed to the LSTM layer in the following section of the architecture. A dense block entailing 2 fully connected layers is found in the last phase after the flattening of the model, a 0.4 drop rate is set in the Dropout layers of the dense block inhibiting overfitting past each respective fully connected layer.

Due to the fact that the model gives us a score, applying a threshold will provide a prediction (by comparison) which is 0.5 by default. The threshold was tuned to 0.946058 based on the ROC curve in this study, shown in Fig. 7. The notion of using the ROC curve for tuning the threshold is to spot that threshold that proffers the upper-left corner of the curve. In mathematical terms, that threshold p satisfies the Eq. (9):

$$TPR\,(p) = 1 - FPR\,(p) \tag{9}$$

It is equivalent to finding the value of p for which the True Positive Rate is equal to the True Negative Rate (1-FPR). In simple terms, we want to find the threshold that satisfies the following:

$$\mathrm{p}{*} = \mathrm{argp\ min}\ \ |\mathrm{TPR\ (p) + FPR\ (p) - 1}| \tag{10}$$

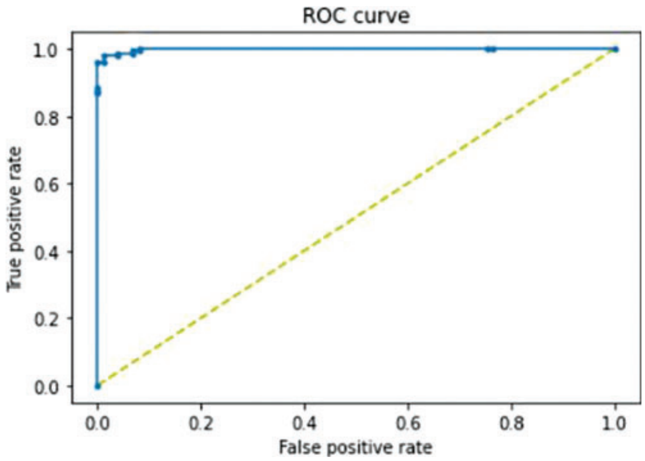


Figure 7: ROC curve of the proposed model with AUC = 0.998

Throughout the training period of the proposed network due to its practical choice of hyperparameters [26], the Adam optimizer [27] was employed. Furthermore, various batch sizes were inspected to achieve a minor error proportion possible (256, 128, 64, plus 32) and was finally set to 64. To inhibit the saturation of the model and to maintain a nominal error rate, numerous learning rates were tested and later set to 0.0001. Epochs are set to 200 to monitor the sturdiness of the model and relate the test scores with comparable conditions.

6 Performance Evaluation Criterion

To gauge the performance of the proposed approach, the highlighted metrics below were utilized. AP represents the correctly predicted unmasked individuals, FP is the masked individuals that were wrongfully classified as unmasked by the proposed model, and AN represents the masked cases that were suitably classified. At the same time, the FN denotes the unmasked cases that were misclassified as masked.

$$\text{Accuracy} = (\text{AP} + \text{AN})/(\text{AN} + \text{FP} + \text{AP} + \text{FN}) \tag{11}$$

$$\text{Recall} = \text{AP}/(\text{AP} + \text{FN}) \tag{12}$$

$$\text{Precision} = \text{AN}/(\text{AN} + \text{FP}) \tag{13}$$

$$\text{F1 score} = (\text{AP} \times 2)/(2 \times \text{AP} + \text{FP} + \text{FN}) \tag{14}$$

7 Evaluation Results and Discussion

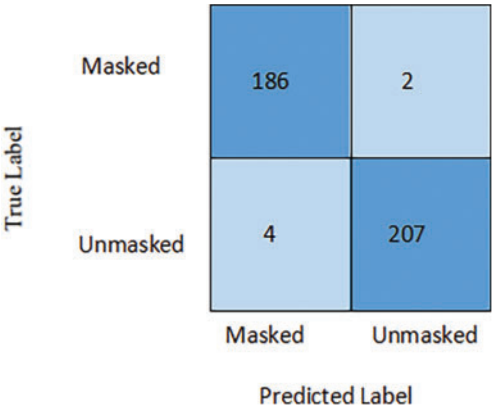
The training, testing, and validation of the suggested architecture are carried out using Cuda 10.1 and TensorFlow 2.8.0 on an 84 GB RAM and Nvidia GeForce RTX 3080 Ti GPU computer. Images are equally distributed according to their respective classes. Considering 3833 total images, they were split into training, testing, and validation set. With 3036 images in the Training set (1519 unmasked, 1517 masked images), 399 images in the Testing set (211 unmasked, 188 masked), and 398 images in the Validation set (210 masked and 188 unmasked images Moreover, the ROC-AUC, accuracy, recall, and precision are computed during the validating and testing stages to examine the performance of the model.

7.1 Results Evaluation

Fig. 8 portrays the proposed CNN-LSTM architecture's test phase confusion matrix. Though the architecture classifies all images almost perfectly, there is a total of 6 misclassified images. Two were misclassified as unmasked, and four images were misclassified as masked. It is found that the suggested CNN-LSTM provides an outstanding performance based on harmonious actual-negative, actual-positive results and fewer false negative and false-positive results. Consequently, the proposed architecture can proficiently detect masked individuals.

Additionally, Fig. 9 demonstrates the graphical performance assessment of the bid CNN-LSTM classifier by loss and accuracy in the training stage and validation phase. The training and validation accuracy is 98.2% and 97%, in that order. Correspondingly, the training loss and validation loss is 0.18 and 0.05, in turn, for the bid architecture.

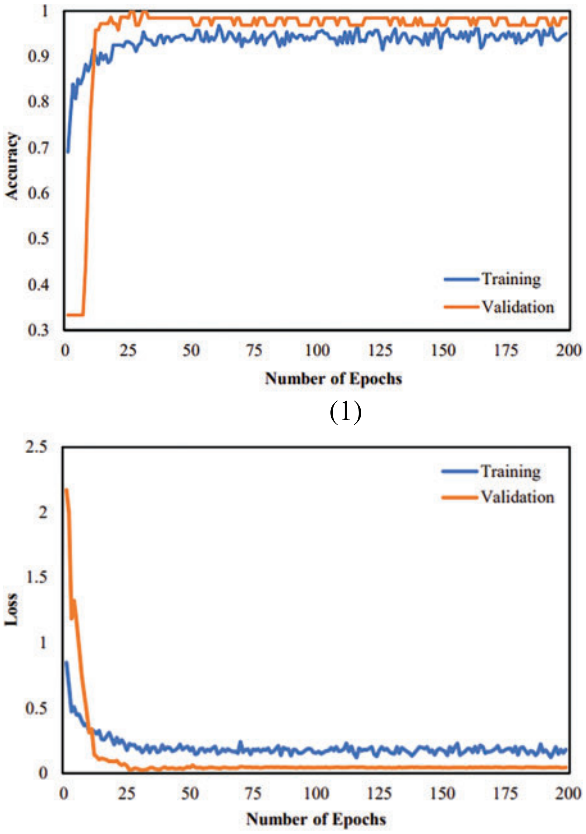
Additionally, Table 2 plus Fig. 10 illustrates each class's assessment metrics of the fused CNN-LSTM approach. Accuracy is found to be 99% for masked cases and 98% for unmasked cases. The Precision is at 99% and 98% for masked, unmasked cases, in that order. The Recall is achieved at 98% for both cases. The f1-score is found to be 98% for the two classes.



		Masked	Unmasked
True Label	Masked	186	2
	Unmasked	4	207

Predicted Label

Figure 8: Confusion matrix of the detailed face mask detection system



(1)

(2)

Figure 9: Assessment metrics of the bid LSTM-CNN Face mask detection architecture (1) Accuracy (2) Loss

From the experimental point of view, it is eminent that the bid CNN-LSTM architecture achieved an overall 99% accuracy, 99% Precision, and 98% Recall, respectively, for the Masked cases. This endeavor's principal point is attaining outstanding results in masked individuals' detection. Hereafter, results disclose that the bid CNN-LSTM design provides reliable performance on that task.

Table 2: Performance assessment of fused CNN-LSTM

Class	Accuracy	Precision	Recall	F1-score
Masked	0.99	0.99	0.98	0.98
Unmasked	0.98	0.98	0.98	0.98

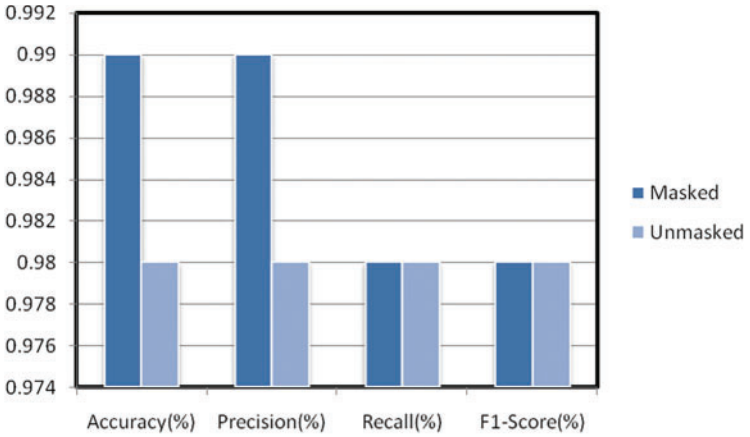


Figure 10: Graphical bar chart representation of the assessment metrics results for each class

7.2 Discussions

Analyzing the results demonstrates that a fusion of CNN and LSTM has a noteworthy ability to detect masked individuals. Based on automatically extracted features, the bid system could distinguish masked class from unmasked class with high accuracy from the input images. In Table 3 below, a comparative study of approaches with the bid CNN-LSTM approach is presented. Based on Table 3, it is proven that a number of the proposed approaches [28–30] achieved a moderately lower accuracy of 97.3%, 96.5%, and 96.2%, in that order. The soberly highest accuracy of 98.7% and 99.82% are found at [30]. The result of the bid system is superior in terms of performance compared to some existing systems. Most present systems achieved accuracy in the 0 range of 80% to 96%, slightly less than our proposed system.

Table 3: Comparative table

Author	Architecture	Dataset	Size	Accuracy (%)
Rahman et al. [11]	CNN	Publicly available on Kaggle and GitHub	858 masked, 651 unmasked	98.7

(Continued)

Table 3: Continued

Author	Architecture	Dataset	Size	Accuracy (%)
Sidik et al. [28]	CNN based model	Face mask detection by Larxel on Kaggle and partially selected dataset from MAFA [12]	4423 images and 250 images [12]	97.3
Saravanan et al. [29]	Pre-trained Vgg16 based model	Kaggle sourced (Dataset 1 & Dataset 2)	1484 images in 2 classes (Dataset 1) & 7200 images in 2 classes (Dataset 2)	96.5 (Dataset 1) & 91 (Dataset 2)
Shamrat et al. [30]	CNN based model (Average pooling approach)	RMFD, SMFD, personally collected images	1340 images in 2 classes	98.67
Shamrat et al. [30]	CNN based model (Max pooling approach)	RMFD, SMFD, personally collected images	1340 images in 2 classes	96.23
Shamrat et al. [30]	MobileNetV2	RMFD, SMFD, Personally collected images	1340 images in 2 classes	99.82
Proposed system	CNN-LSTM	Webcam images, RMFD, SMFD	3833 images (1915 masked and 1918 unmasked)	99

8 Conclusion and Future Scope

This study combined CNN and LSTM networks building a collaborative network (CNN-LSTM) for face mask detection. The primary objective was to suggest a highly accurate attuned model such that mask identification will be fast and simple to curb the spread of COVID19 and other airborne diseases in general. The proposed model provided outstanding performance based on experimental results. Though able to correctly classify images with various occlusion and face angles in almost all cases, it slipped in some. It is quite a dare that must be addressed to increase the performance of the employed model for masks in community areas.

Funding Statement: The author received no specific funding for this study.

Conflicts of Interest: The author declare that he has no conflicts of interest to report regarding the present study.

References

[1] J. Elliott, M. Whitaker, B. Bodinier, O. Eales, S. Riley *et al.,* “Predictive symptoms for COVID-19 in the community: REACT-1 study of over 1 million people,” *PLoS Medicine*, vol. 8, no. 9, pp. e1003777, 2021.

[2] W. C. Culp, “Coronavirus disease 2019: In-home isolation room construction,” *A & A Practice*, vol. 14, no. 6, pp. e01218, 2020.

[3] T. D. Hollingsworth, D. Klinkenberg, H. Heesterbeek and R. M. Anderson, “Mitigation strategies for pandemic influenza a: Balancing conflicting policy objectives,” *PLoS Computational Biology*, vol. 7, no. 2, pp. e1001076, 2011.

[4] Y. Wang, H. Tian, L. Zhang, M. Zhang, D. Guo *et al.,* “Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: A cohort study in Beijing, China,” *BMJ Global Health*, vol. 5, no. 5, pp. e002794, 2020.

[5] C. T. Leffler, E. Ing, J. D. Lykins, M. C. Hogan, C. A. McKeown *et al.,* “Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks,” *The American Journal of Tropical Medicine and Hygiene*, vol. 103, no. 6, pp. 2400–2411, 2020.

[6] I. J. Rao, J. J. Vallon and M. L. Brandeau, “Effectiveness of face masks in reducing the spread of COVID-19: A model-based analysis,” *Medical Decision Making*, vol. 41, no. 8, pp. 988–1003, 2021.

[7] L. G. Goh, T. M. Ho and K. H. Phua, “Wisdom and western science: The work of Dr wu lien-teh,” *Asia Pacific Journal of Public Health*, vol. 1, no. 1, pp. 99–109, 1987.

[8] M. P. Ravenel, “A treatise on pneumonic plague,” *American Journal of Public Health*, vol. 17, no. 5, pp. 500–500, 1927.

[9] S. K. Jarraya, M. H. Alotibi and M. S. Ali, “A Deep-CNN crowd counting model for enforcing social distancing during COVID19 pandemic: Application to Saudi Arabia’s public places,” *Computers, Materials & Continua*, vol. 66, no. 2, pp. 1315–1328, 2021.

[10] J. W. Sonn and J. K. Lee, “The smart city as time-space cartographer in COVID-19 control: The South Korean strategy and democratic control of surveillance technology,” *Eurasian Geography and Economics*, vol. 61, no. 4–5, pp. 482–492, 2020.

[11] M. M. Rahman, M. M. H. Manik, M. M. Islam, S. Mahmud, and J. -H. Kim, “An automated system to limit COVID-19 using facial mask detection in smart city network,” in *IEEE Int. IoT Electronics and Mechatronics Conf. (IEMTRONICS)*, Vancouver, BC, Canada, pp. 1–5, 2020.

[12] S. Ge, J. Li, Q. Ye and Z. Luo, “Detecting masked faces in the wild with LLE-CNNs,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 426–434, 2017.

[13] X. Fan, M. Jiang and H. Yan, “A deep learning based light-weight face mask detector with residual context attention and Gaussian heatmap to fight against COVID-19,” *IEEE Access*, vol. 8, pp. 96964–96974, 2021.

[14] M. S. Ejaz, M. R. Islam, M. Sifatullah and A. Sarker, “Implementation of principal component analysis on masked and non-masked face recognition,” in *2019 1st Int. Conf. on Advances in Science, Engineering and Robotics Technology (ICASERT)*, Dhaka, Bangladesh, pp. 1–5, 2019.

[15] A. Alzu’bi, F. Albalas, T. AL-Hadhrami, L. Younis and A. Bashayreh, “Masked face recognition using deep learning: A review,” *Electronics*, vol. 10, no. 21, pp. 2666, 2021.

[16] A. M. Hasan, H. A. Jalab, F. Meziane, H. Kahtan and A. S. Al-Ahmad, “Combining deep and handcrafted image features for MRI brain scan classification,” *IEEE Access*, vol. 7, no. 1, pp. 79959–79967, 2019.

[17] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy *et al.,* “Recent advances in convolutional neural networks, pattern recognition,” *Pattern Recognition*, vol. 77, pp. 354–377, 2021.

[18] H. Kutlu and E. Avci, “A novel method for classifying liver and brain tumors using convolutional neural networks, discrete wavelet transform and long short-term memory networks,” *Sensors*, vol. 19, no. 9, pp. 1992, 2019.

[19] A. S. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.

[20] P. Chang, J. Grinband, B. D. Weinberg, M. Bardis, M. Khy *et al.,* “Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas,” *American Journal of Neuroradiology*, vol. 39, no. 7, pp. 1201–1207, 2018.

[21] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] F. Gers, “Long Short-Term Memory in Recurrent Neural Networks,” Ph. D. Dissertation, École Polytechnique Fédérale de Lausanne, Switzerland, 2001.

[23] F. A. Gers, J. Schmidhuber and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[24] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998.

[25] M. Rathi and A. Sinha, "*Advanced Computational Techniques for Sustainable Computing*," London, United Kingdom: Taylor & Francis Ltd, 2021. [Online]. Available: https://www.routledge.com/Advanced-Computational-Techniques-for-Sustainable-Computing/Rathi-Sinha/p/book/9780367495220.

[26] M. A. Ozdemir, O. K. Cura and A. Akin, "Epileptic EEG classification by using time-frequency images for deep learning," *International Journal of Neural Systems*, vol. 31, no. 8, pp. 2150026, 2021.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. on Learning Representations (ICLR)*, San Diego, SD, USA, pp. 13–26, 2015.

[28] R. P. Sidik and E. C. Djamal, "Face mask detection using convolutional neural network," in *2021 4th Int. Conf. of Computer and Informatics Engineering (IC2IE)*, Depok, Indonesia, pp. 85–89, 2021.

[29] T. M. Saravanan, K. Karthiha, R. Kavinkumar, S. Gokul and J. P. Mishra, "A novel machine learning scheme for face mask detection using pretrained convolutional neural network," *Materials Today: Proceedings*, vol. 58, pp. 150–156, 2021.

[30] F. M. J. M. Shamrat, S. Chakraborty, M. M. Billah, M. A. Jubair, M. S. Islam *et al.,* "Face mask detection using convolutional neural network (CNN) to reduce the spread of covid-19," in *2021 5th Int. Conf. on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, pp. 1231–1237, 2021.