2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)

Deep Neural Architecture for Face mask Detection on Simulated Masked Face Dataset against Covid-19 Pandemic

1st Alok Negi
Department of Computer Science and Engineering National Institute of Technology, Uttarakhand Srinagar (Garhwal), 246174, India alokitn@gmail.com

2rd Krishan Kumar
Department of Computer Science and Engineering National Institute of Technology, Uttarakhand Srinagar (Garhwal), 246174, India kkberwal@nituk.ac.in

3nd Prachi Chauhan
College of Technology
G.B. Pant University of Agriculture and Technology
Pantnagar, Uttarakhand, 263153, India
prachi3apr@gmail.com

4th R.S. Rajput
Department of Mathematics, Statistics and Computer Science
G.B. Pant University of Agriculture and Technology
Pantnagar,Uttarakhand, 263153, India
rajpoot.rs@gmail.com

2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) | 978-1-7281-8529-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCIS51004.2021.9397196

***Abstract*—The dangerous COVID-19 (SARS-CoV-2) is rising steadily and globally, with more than 72,851,747 confirmed cases observed to WHO including 1,643,339 deaths till 17 December 2020. The country's economy is now almost fully halted, people are stuck up and investment becomes deteriorating. So, this is turning to worry of the government for a development and health. Health organizations are often desperate for evolving decision-making innovations to overcome this viral virus and encourage people to receive rapid and effective responses in real-time. Thus, it is important to create auto-mechanisms as a preventive shield to ensure healthy humanity against SARS-CoV-2. Advanced analytics methods and other strategies could also empower researchers, learners and the pharmaceutical industry to acknowledge the hazardous COVID-19 and speed it up care procedures by efficiently testing vast volumes of research data. The prevention method consequence is being used to effectively manage, calculate, forecast and monitor current infected people and future potential cases. Therefore, we proposed CNN and VGG16 based deep learning models to incorporate and enforce AI-based precautionary measures to detect the face mask on Simulated Masked Face Dataset (SMFD). This technique is capable of recognizing masked and unmasked faces to help monitor safety breaches, facilitate the use of face masks, and maintain a secure working atmosphere.**

***Index Terms*—CNN, Deep Learning, Data Augmentation, Face Mask Detection, Simulated Masked Face Dataset (SMFD), VGG 16, WHO**

I. INTRODUCTION

The novel COVID-19 first recorded in the Chinese province of Wuhan in December 2019 exploded rapidly all over the world and soon became a worldwide problem. It seems to have a profound effect on everyday life, public health and the global economy. COVID-19 is indeed a rare virus correlated with a reasonably common family of viruses with extreme acute respiratory syndrome (SARS) and likely some form of flu or cold.COVID-19 refers to the beta-coronavirus genera, based on both its phylogenetic resemblances and genomic characteristics. Human beta-coronaviruses (MERS-CoV, SARS-CoV, SARS-CoV-2) have some common features, but they also have differences in some genomic and phenotypic compositions that may influence their pathogenesis. COVID-19 composed of a single-stranded (positive-sense) RNA aligned with the nucleoprotein [1] in a capsid consisting of a protein matrix. The traditional CoV incorporates a minimum of six ORFs in its genome. However both functional proteins and additional proteins are perceived as CoV sgRNAs.

When an infected individual coughs, communicates or sneezes then the entire virus is transmitted via the respiratory duct produced between persons in direct contact with one another. Such particles may be inhaled when they approach towards the face, land in the ears, eyes, mouth, nose. Commonly reported symptoms include fever, dry cough, and fatigue, but there is currently no recommended medication to cure COVID-19, and thus no treatment is available. Through time, no antibiotics have been effective against respiratory [2] infections such as COVID-19. Also, it is apparent that preventive measures should be taken to avoid the transmission of these viruses to the human body. The authorities of most nations often undertake other protective measures to protect their citizens. All required steps are also the maintenance of social distance, isolation, sensitization of the skin, lockdown and in particular, the use of masks when stepping outside. Leung et al. [3] demonstrated why and how to use a face mask as a protective measure towards COVID-19.

As disease is becoming a massive disaster, artificial intelligence approaches and methods should also be used to support decision made by the healthcare system and the social system in their efforts to sustain every step of the decline and its consequences: identification, security, reaction, recovery and accelerated study. By use of modern AI techniques combined with face imaging may be advantageous for the effective diagnosis of disease and also helpful in addressing the problem of the absence of medical professionals in remote communities. The study outlined in this paper integrates the relevant improvements to the domain of public protection and the biomedical sector in order to avoid a global pandemic disease.

The proposed work implements a scratch CNN model and VGG16 [4] with transfer learning for the face mask detection from SMFD dataset using Python script and utilizes the concept of deep learning with Tensorflow on Google Colab. The aspects of our study are listed:

- To illustrate the technological viability of the Deep Learning approach to accelerate the deployment of the Global Pandemic Covid-19 protective steps.
- Formulate a revolutionary paradigm associated with a face mask identification to recognize and prevent Covid-19 via Deep Learning.
- Exploration of CNN and VGG16 advanced Deep Convolution Neural Network architectures for the features extraction and classification.
- To provide qualitative and quantitative analysis using accuracy, log-loss value, precision, recall and f1 score.

II. RELATED WORK

As the world started to adopt precautionary measures against the Coronavirus, various Face Mask Detection systems [5]–[10] implementations came out.

Wang et al. [11] proposed three different categories of masked face datasets along with the Masked Face Detection Dataset (MFDD) built-in data consists 24,771 masked facial images. The Real-world Masked Face Recognition Dataset (RMFRD) contains 5,000 pictures of 525 person wearing masks and 90,000 images of same 525 participants without masks and Simulated Masked Face Recognition Dataset (SMFRD) covering 500,000 facial images of 10,000 participants. Out of these, from best of researchers' expertise, RMFRD is globally the leading masked face dataset in the modern world. All such datasets are freely accessible to academia and industry and based on that different applications on masked face images can be implemented. The multigranular masked face image recognition proposed model achieves 95 per cent accuracy, increasing the outcome developed by the industrial sector.

Hariri et al. [12] suggested a technique of deep learning as well as a quantization-based procedure for identification of masked faces. The proposed algorithm may also be widened to better application areas like violence video retrieval and video surveillance. From the authors' point of view, the first task was to remove the masked face region. After that, implement the Pre-Trained Deep Convolutional Neural Nets (CNNs) to retrieve the finest features from the specified area (mostly eyes and forehead area). At last, the Bag-of-features approach was implemented to the feature vectors of the last hidden layer in order to measure them and also to obtain a modest interpretation as compared to the fully - connected layer of traditional CNN. In the end, MLP was added for the classification task. Real-World-Masked-Face-Dataset experimental output demonstrated highest recognition rate.

Roy et al. [13] presented an innovative deep network, abstracted from the Spatial Transformer Networks, that continuously predicts the prevalence score associated with an input block and produces sparse representation of localization of pathological objects. In addition, authors implemented a new framework for efficient frame score accumulation at a multimedia level, predicated on uninorms. After that, reference deep state-of-the-art methods for predicting COVID-19 scanning genetic variants at pixel-level segmentations. Experiments of the proposed system show satisfactory responses on all of the tasks under consideration, preparing the direction for new DL work for the necessary solutions of COVID-19 using LUS data.

Rustam et al. [14] proposed a ML prediction framework for forecasting future COVID-19 epidemic risk. The module analyzes the dataset actually contains the actual day-to-day past results, and uses machine learning methodologies to predict the next few days. The research findings proved that in regard of the complexity and extent of the dataset, ES performed better with respect to current forecasting environment. In some degree, LR and LASSO also work well enough for forecasting in estimate the death rate and validate events. Death levels will rise in near future, and recovery rates will be greatly reduced, as per the effects of these two approaches. Due to the extreme ups and downs in database calculated points, SVM provides bad performance across all situations.

Loey et al. [15] presented a complex model that uses deep and classic learning algorithms to detect face masks. The developed scheme composed in two phases. The first phase was for the extraction of features using Resnet50. Whereas the second phase was used to detect face masks through traditional machine learning methods. Support Vector Machine (SVM), Decision Trees and Ensemble Algorithms have been chosen as conventional machine learning towards analysis. The proposed finding confirmed that the SVM classifier obtained the maximum feasible accuracy with minimum time spent throughout the training. The SVM classifier in RMFD obtained 99.64 percent test accuracy. In SMFD, 99.49 per cent was achieved, whereas in LFW, 100 per cent of the test accuracy was achieved.

III. PROPOSED WORK

In this work, we implemented a face mask detection mechanism using CNN and VGG16 model for identifying those among the crowd, who has not worn a mask. The proposed work uses the concept of data augumentation, dropout, normalization and transfer learning. In hospitals, malls, transport hubs, restaurants and some other community meetings where surveillance is required, this technique can be used.

For CNN implementation, Using 16, 16, 32, 32, 64, 64, 96, 96, 128 and 128 filters respectively with a size of 3 x 3, ten convolution layers are implemented and Relu is used as an activation function calculated by $f(x) = \max(0,x)$. The model used five layers of maxpooling with stride 2 followed by the flattening layer. Then four dense layers are built using the relu activation function in which there are 512, 128 and 64 hidden nodes respectively in the first, second and third dense. The fourth dense layer with two hidden nodes using the softmax activation function is used for the final output. Figure 1a demonstrates the layered architecture for this classification. Fully connected layers are omitted from the original VGG16 architectures for proposed VGG16 implementation and additional dense layers are applied to the top of the model. Three fully connected layers are replaced with two dense layers in this work (128 and 2 hidden nodes respectively). For the final output, the second dense layer is used using the softmax activation function. In order to prevent overfitting, pre-trained weights and transfer learning are used with dropout, augmentation and normalisation. Figure 1b demonstrates the layered architecture for this classification.

IV. RESULT AND ANALYSIS

Both the model are trained on Google Colab using python script and Tensorflow for just 25 epochs. Adam optimizer and Batch size 32 are used for it. For CNN, there are total 2,181,778 trainable parameter and there is no non-trainable parameter. For VGG16, There are total 14,780,610 parameter out of which 65,922 are trainable and remaining 14,714,688 non-trainable parameter.

A. Dataset Description

In this work, Simulated Masked Face Dataset (SMFD)[18] is used for training, validation and testing set which is available in the public domain. 1315 images (658 mask and 657 without mask) are included in the training set while 142 images are included in the validation set (71 mask and 71 without mask). Testing set contains 194 images (97 mask and 97 without mask). The dataset distribution is plotted in Figure 2.

B. Kernel Density Plot

The distribution of data over a continuous interval or time span is visualized by a density plot as shown in Figure 3a and Figure 3b. The density plot peaks help to demonstrate where values over the interval are concentrated. A benefit of density plots over histograms is that the distribution form is better calculated since the number of bins used does not affect them.

C. Data Preprocessing and Augmentation

Input images also come in different sizes and resolutions for this dataset, so they have been resized to 224×224×3 to decrease the scale. By generating new examples by arbitrarily transforming existing ones, data augumentation enriches training data. This way, we artificially lift the size of the training set, eliminating overfitting. For the proposed work, zoom range (0.2), rescale (1./255), horizontal flip (true) and shear range(0.2) are used as augmentation parameter using ImageDataGenerator. The random transformation Using data augmentation are shown in Figure 4.

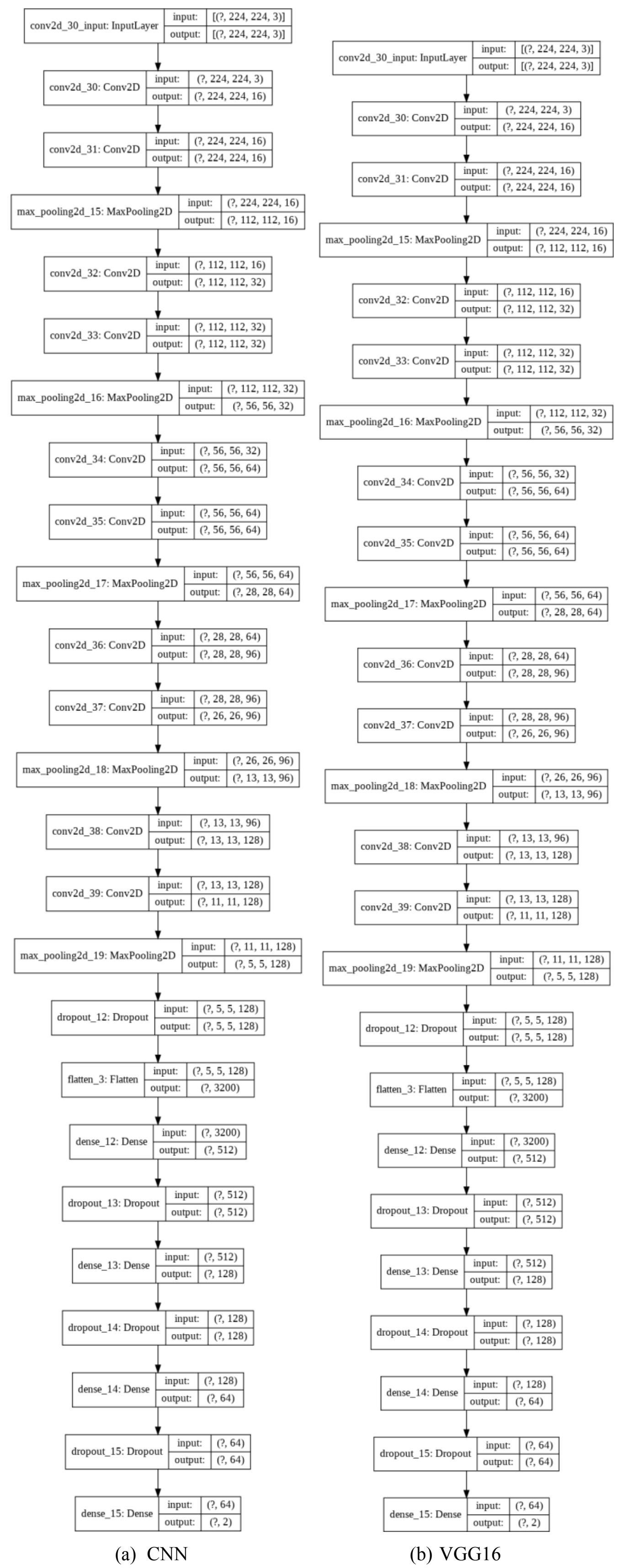
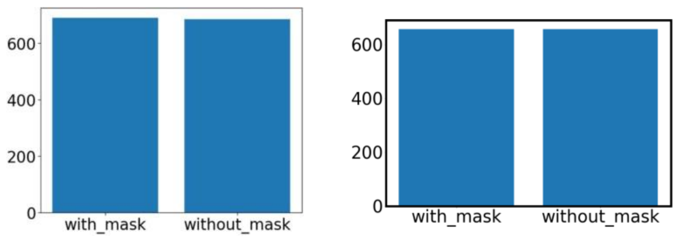
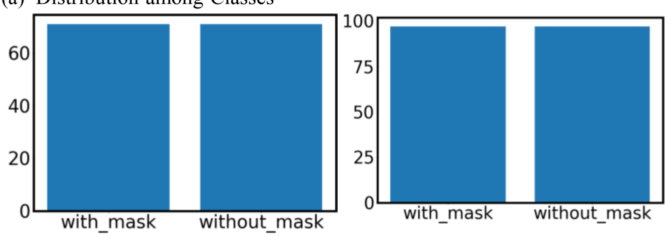


Fig. 1: Proposed Layered Architecture for CNN and VGG16



(a) Distribution among Classes

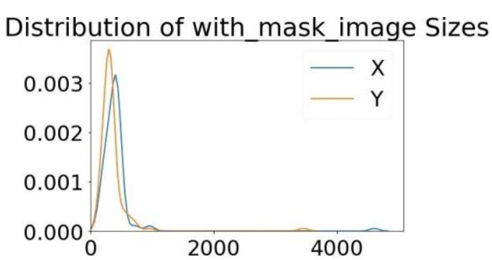
(b) Training Set



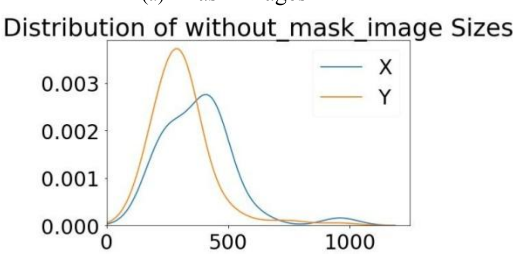
(c) Validation Set

(d) Test Set

Fig. 2: Distribution of Dataset



(a) Mask Images



(b) Without Mask Images

Fig. 3: Kernel Density Plot

D. Performance Matrics

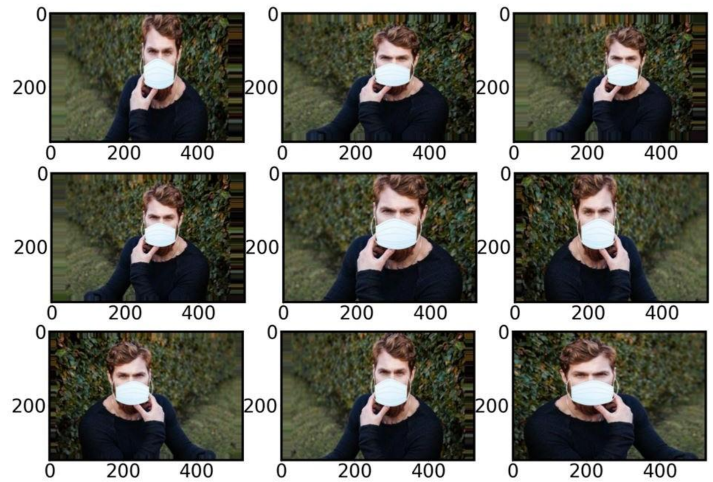
Accuracy, Categorical cross entropy, Precision, Recall, F1 score and Confusion matrix are used as a performance matrics and mathematics behind them are given by the Equation 1, 2, 3, 4 and 5.

$$Accuracy = (TP + TN)/\ (FN + TP + TN + FP) \quad (1)$$

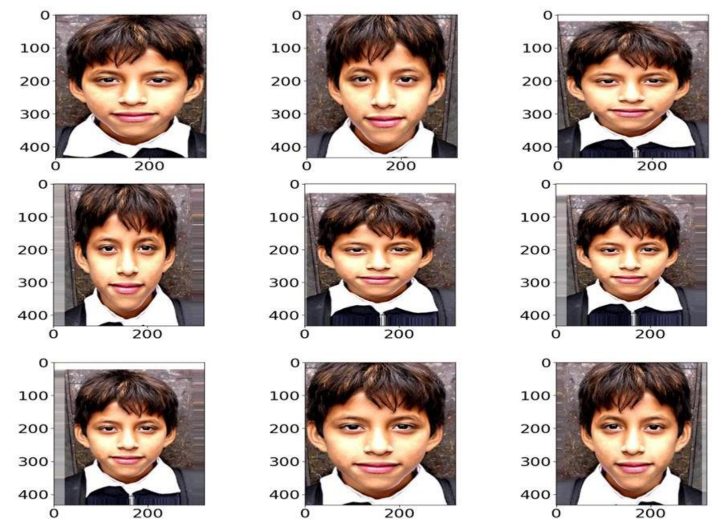
Categorical cross entropy as shown in equation 2 is used as a metric for this work. A perfect classifier gets the logloss of 0.

$$logloss = \frac{-1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} * log(p)_{ij} \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$



(a) Mask Images



(b) Without Mask Images

Fig. 4: Random Transformation Using Data Augmentation

$$Recall = TP/\ (FN + TP) \quad (4)$$

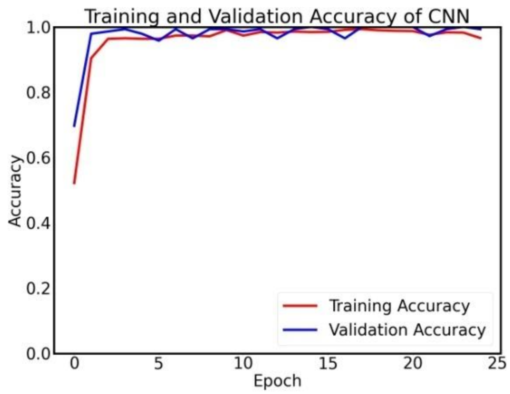
$$f1\ Score = 2\ x\ (Precision\ Recall) / (Precision + Recall) \quad (5)$$

The accuracy curve and loss curve for CNN are shown by Figure 5a and 5b. For VGG16, Figure 5c and 5d shows accuracy and loss curve. The accuracy and logloss recorded by the CNN and VGG16 model are shown in Table I.

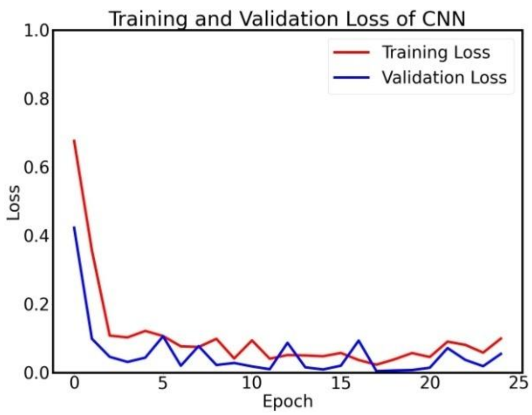
TABLE I: Accuracy (In Percent) and Logloss Score

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss	Testing Accuracy	Testing Loss
CNN	96.35	0.14	99.30	0.05	97.42	0.10
VGG16	99.47	0.02	98.59	0.02	98.97	0.02

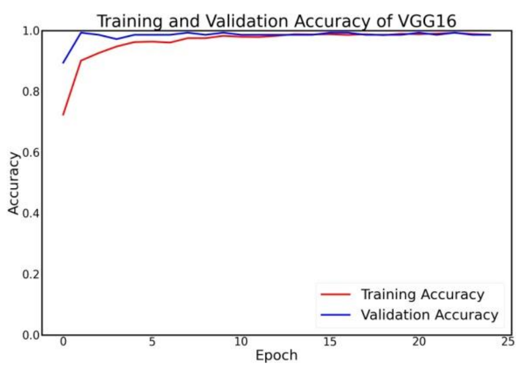
Confusion Matrix are shown in Figure 6 for validation and testing set while Classification report are given by Table II and



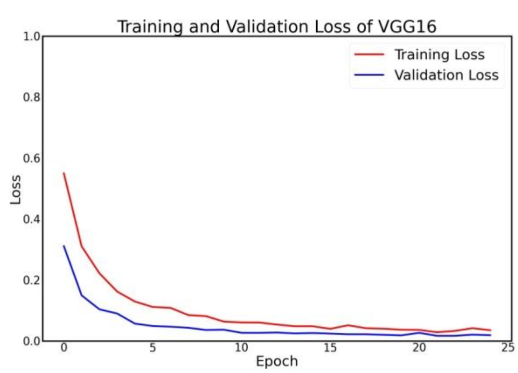
(a) Accuracy Curve CNN



(b) Loss Curve CNN



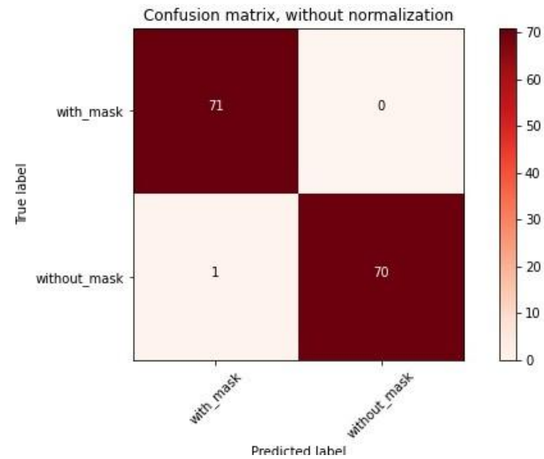
(c) Accuracy Curve VGG16



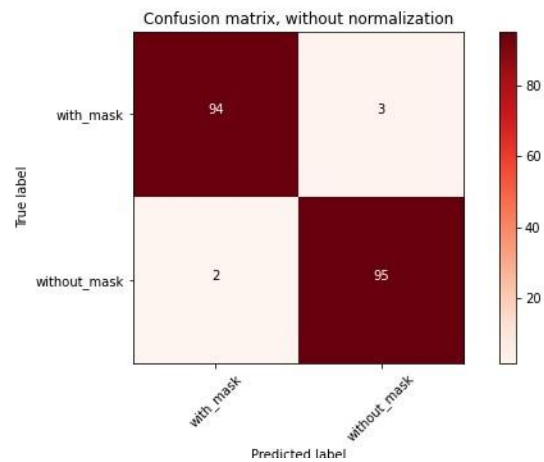
(d) Loss Curve VGG16

Fig. 5: Accuracy and Loss Curve

III. Using CNN, the value of True Negatives, False Positives, False Negatives, True Positives are recorded 71, 0, 1, 70 respectively for the validation while 94, 3, 2, 95 respectively for the testing set. Using VGG16, the value of True Negatives, False Positives, False Negatives, True Positives are recorded 70, 1, 1, 70 respectively for the validation while 97, 0, 2, 95 respectively for the testing set.



(a) Validation Set



(b) Test Set

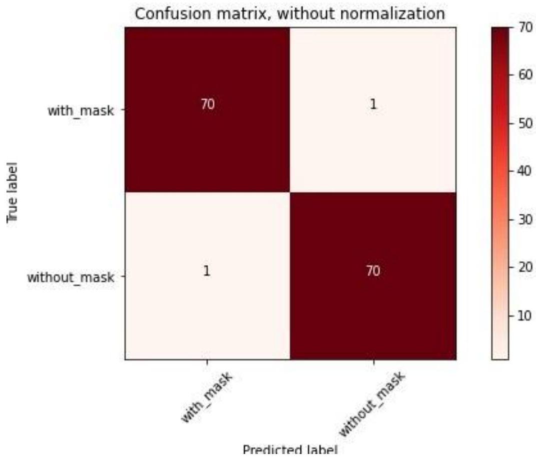
Fig. 6: Confusion matrix for CNN

TABLE II: Classification Report for Validation(In Percent)

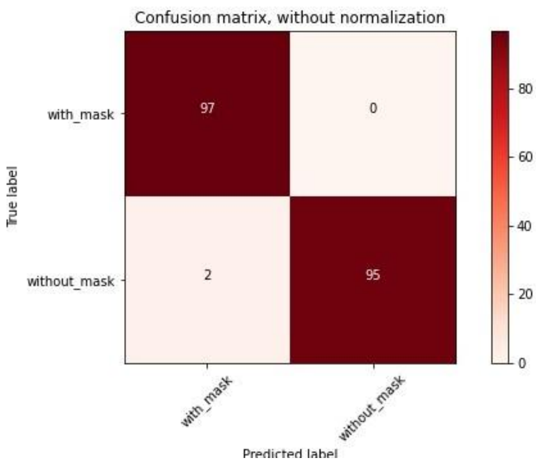
Model	Precision	Recall	f1 score
CNN	100.00	98.59	99.29
VGG16	98.59	98.59	98.59

TABLE III: Classification Report for Testing(In Percent)

Model	Precision	Recall	f1 score
CNN	96.94	97.94	97.44
VGG16	100.00	97.94	98.96



(a) Validation Set



(b) Test Set

Fig. 7: Confusion matrix for VGG16

E. Comparison with Related Works

We compared the classification accuracy of our proposed system as shown in Table I with other approaches and obtained the higher and nearer accuracy among some methods. The same SMFD dataset were used in the work presented in[[15]. The author achieved 96 percent and 95.64 percent accuracy for validation and testing using decision trees classifier. The SVM classifier achieved 100 percent and 99.49 percent while ensemble classifier achieved 94 percent and 98.7 percent accuracy for validation and testing. Our Model achieved 96.35 percent, 96.35 percent and 97.42 percent accuracy for training, validation and testing using CNN while 99.47 percent, 98.59 percent and 98.97 perecent using VGG16.

Although our proposed work recorded better results in just 25 epochs with the SMFD dataset but no of epochs with more standard dataset can be increased for better analysis.

V. CONCLUSION

The pandemic of the COVID-19 coronavirus is causing a global health crisis. Studies have shown that wearing a face mask decreases the risk of viral transmission substantially as well as offers a sense of security. In this work, we proposed two deep neural models for the detection of the facemask using SMFD dataset. CNN model achieved training, validation and testing accuracy 96.35 percent, 96.35 percent and 97.42 percent respectively while another model VGG16 achieved
99.47 percent, 98.59 percent and 98.97 perecent respectively. In restaurants, bus terminals, some other social occasions and public places, the proposed work can be used where monitoring is needed. In future, this work can be extend for face mask detection from more standard dataset and video surveillance by using advanced computer vision approach.

REFERENCES

[1] D. Hunt, “Pathogenesis of tissue injury in the brain in patients with systemic lupus erythematosus,” in *Systemic Lupus Erythematosus*. Elsevier, pp. 341–348.

[2] Z. Xu *et al.*, “Lancet respir. med,” 2020.

[3] N. H. Leung, D. K. Chu, E. Y. Shiu, K.-H. Chan, J. J. McDevitt, B. J. Hau, H.-L. Yen, Y. Li, D. K. Ip, J. M. Peiris *et al.*, “Respiratory virus shedding in exhaled breath and efficacy of face masks,” *Nature medicine*, vol. 26, no. 5, pp. 676–680, 2020.

[4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[5] O. Cakiroglu, C. Ozer, and B. Gunsel, “Design of a deep face detector by mask r-cnn,” in *2019 27th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2019, pp. 1–4.

[6] T. Meenpal, A. Balakrishnan, and A. Verma, “Facial mask detection using semantic segmentation,” in *2019 4th International Conference on Computing, Communications and Security (ICCCS)*. IEEE, 2019, pp. 1–5.

[7] M. R. Bhuiyan, S. A. Khushbu, and M. S. Islam, “A deep learning based assistive system to classify covid-19 face mask for human safety with yolov3,” in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2020, pp. 1–5.

[8] A. S. Joshi, S. S. Joshi, G. Kanahasabai, R. Kapil, and S. Gupta, “Deep learning framework to detect face masks from video footage,” in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 2020, pp. 435–440.

[9] Y. Wang, B. Luo, J. Shen, and M. Pantic, “Face mask extraction in video sequence,” *International Journal of Computer Vision*, vol. 127, no. 6-7, pp. 625–641, 2019.

[10] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, “Multi-stage cnn architecture for face mask detection,” *arXiv preprint arXiv:2009.07627*, 2020.

[11] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei *et al.*, “Masked face recognition dataset and application,” *arXiv preprint arXiv:2003.09093*, 2020.

[12] W. Hariri, “Efficient masked face recognition method during the covid-19 pandemic,” 2020.

[13] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli *et al.*, “Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound,” *IEEE Transactions on Medical Imaging*, 2020.

[14] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B. On, W. Aslam, and G. S. Choi, “Covid-19 future forecasting using supervised machine learning models,” *IEEE Access*, 2020.

[15] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, “A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic,” *Measurement*, vol. 167, p. 108288, 2020.