Array 19 (2023) 100294

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

An effective stacked autoencoder based depth separable convolutional neural network model for face mask detection

Sundaravadivazhagan Balasubaramanian *, Robin Cyriac, Sahana Roshan, Kulandaivel Maruthamuthu Paramasivam, Boby Chellanthara Jose

University of Technology and Applied Sciences - Al Mussanah, Department of Information Technology, Al Muladdah, 314, South Al Batinah, Oman

ARTICLE INFO

Keywords:
COVID-19
Stacked auto encoder
Principal component analysis
Depth-wise separable convolutional neural network
Machine learning
Deep learning

ABSTRACT

The COVID-19 pandemic has been infecting the entire world over the past years. To prevent the spread of COVID-19, people have acclimatised to the new normal, which includes working from home, communicating online, and maintaining personal cleanliness. There are numerous tools required to prepare to compact transmissions in the future. One of these elements for protecting individuals from fatal virus transmission is the mask. Studies have indicated that wearing a mask may help to reduce the risk of viral transmission of all kinds. It causes many public places to take efforts to ensure that its guests wear adequate face masks and keep a safe distance from one another. Screening systems need to be installed at the doors of businesses, schools, government buildings, private offices, and/or other important areas. A variety of face detection models have been designed using various algorithms and techniques. Most of the articles in the previously published research have not worked on dimensionality reduction in conjunction with depth-wise separable neural networks. The necessity of determining the identities of people who do not cover their faces when they are in public is the driving factor for the development of this methodology. This research work proposes a deep learning technique to determine if a person is wearing mask or not and identifies whether it is properly worn or not. Stacked Auto Encoder (SAE) technique is implemented by stacking the following components: Principal Component Analysis (PCA) and Depth-wise Separable Convolutional Neural Network (DWSC-NN). PCA is used to reduce the irrelevant features in the images and resulted high true positive rate in the detection of mask. We achieved an accuracy score of 94.16% and an F1 score of 96.009% by the application of the method described in this research.

1. Introduction

Because of the tremendous advancements in science and technology, it has progressed to the point where they can perform tasks that were thought to be impossible a few decades ago. The development of technologies such as Artificial Intelligence (AI) and Machine Learning (ML) has made people's lives easier and provided solutions to a lot of difficult problems in a wide range of fields. Algorithms used in modern computer vision are growing ever closer to being able to execute tasks of visual perception on par with humans. From the classification of pictures to the analysis of videos, computer vision has proven to be a game-changing component of today's technology. Currently, technology is a saviour for fighting viral infections and preparing ourself for any further outbreaks in the future. The concept of "work from home" has largely helped replace the traditional workday schedules in most industries and become ingrained in everyday life as a result of technological advancements. However, it is challenging for several different industries to adapt to this new norm.

People are still reluctant to go back to work even though the virus will eventually be eradicated, and such sectors are becoming increasingly eager to begin employing people face-to-face. 65% of workers today report feeling anxious about returning to the workplace [1]. Since the beginning of the viral pandemic, researchers from all over the world have been looking for answers and strategies that will stop the spread. The current virus travels through the respiratory system of the patient to the lung cells, where it then causes direct harm. The most effective strategy to stop the transmission of the virus is to keep distance from other people and always wear a mask when in a busy or public place.

To begin, masks are not obligatory for everybody, but researchers and medical professionals are becoming more and more inclined to propose that everyone use masks. According to the findings of several studies [2], wearing a face mask reduces the likelihood of viral

* Corresponding author.
E-mail address: sundaravadi@act.edu.om (S. Balasubaramanian).

https://doi.org/10.1016/j.array.2023.100294
Received 24 January 2023; Received in revised form 8 May 2023; Accepted 14 May 2023
Available online 5 June 2023

transmission and provides the user with a sense of safety. Face masks are recommended for use to prevent the virus from spreading. This recommendation comes from the World Health Organization as well as other health organisations for most viral spreads that can hurt the respiratory system. Although every government tries to mandate the use of face masks in public spaces, it might be difficult to manually identify those who are not complying with this requirement in crowded environments.

Researchers have been working to develop automated systems that can identify people using face masks in public locations and enforce their use. It is impossible to manually monitor this policy on a large scale and keep track of any infractions because of the large volume of violations. The use of computer vision offers an alternative that is more desirable. Integrating object detection, object tracking, and image classification resulted in the creation of a sophisticated system that is capable of recognising face masks in both still photographs and moving images. This was accomplished by combining the above mentioned four concepts.

The processing of digital photographs by a computer is referred to as "digital image processing," and it is accomplished using a powerful digital computer. It is also possible to claim that it is the utilisation of computer algorithms to improve the image and extract some significant data. The processing of images also involves significantly completing the following steps [3]:

- Importing the photograph by means of image capture software.
- Analysing and altering the photograph.
- Producing output, the result of which may be a modified image, or a report based on the evaluation of that photograph.

In recent years, there has been a lot of focus placed on Deep Learning (DL), particularly in areas such as computer vision, computational linguistics, object classification, and other elements of information processing. Many previous studies on object identification have made use of models that are based on convolutional neural networks. There has been a boom in the use of CNNs, for several applications, including voice synthesis, picture identification, image thresholding, and object tracking, among others. The disciplines are well-suited to the data extraction capabilities of CNN, which are quite good. CNNs are quickly becoming the method of choice for a rising number of research projects, which are looking to improve their ability to collect visual data and produce more accurate classifications than they could with previous classification approaches. There are several deep neural networks that are not ideal for mobile-based facial picture categorization because of resource constraints, the length of time and amount of money required for their evaluation phase, and the possibility of false positives.

The problem can be summed up as follows: when performing a mask detection job, the classification model is required to categorise the facial image based on the input of a face image. This research is successful in developing the model with fewer learnable parameters and a smaller total number of learnable parameters by employing deep separable convolution layers rather than conventional convolutional layers.

This study proposes a model for categorising facial photos based on MobileNet. To handle this challenge, the model utilises the Depth Wise Separable Convolution (DWSC) technique. DSC is widely used in the process of identifying issues that arise in image processing. In the beginning, it was presented in the article published by Laurent Sifre and Stéphane Mallat [3]. DWSC is a quantized variation of the conventional convolution.

Depth-wise and 1-point convolutions are two categories that are frequently used to categorise convolutions. Instead of using all input channels as in standard convolution, the DWSC layer applies one filtering to a single pulse. The DWSC outputs are then aggregated using a single matrix pointwise convolution. DWSC reduces both the number of parameters that may be learned and the complexity of the test and train simulations.

In the previous study, PCA along with DWSC is not experimented for face mask detection. PCA is an important technique for face mask detection using deep learning because it helps to reduce the dimensionality of the input data while retaining the most important features. In face mask detection, the input data is typically a set of images containing faces with and without masks. These images can be high-dimensional, with each pixel representing a separate feature. However, not all these features are equally important for the task of face mask detection. Some features may be redundant or irrelevant, while others may be critical for distinguishing between masked and unmasked faces.

The primary objective of this research study is to develop a deep learning model that is capable of recognising individuals who are not wearing face masks. The proposed model makes use of MobileNetV2's transfer learning in conjunction with surveillance cameras in order to identify individuals in public places who are not wearing masks. Image augmentation techniques are utilised to broaden the scope of the training data in order to optimise the functionality of the recommended model. The following is a list of the primary contributions that the proposed effort will make:

- Conducted a literature review to gain a better knowledge of the harm caused by respiratory viral transmissions as well as the latest research works on deep learning models for face mask detection.
- Constructed a facemask detector to assist in the precise real-time detection of a mask using picture and video streams.
- PCA is used to reduce the irrelevant features extracted from the images.
- The proposed model makes use of well-known depth wise separable convolutional neural networks techniques to build the classifier, collect images of people wearing masks, and differentiate between classes of face masks and non-facial masks. This activity is accomplished by utilising Open-CV, Keras, and Python.
- When compared to other models, the model that was proposed has a smaller memory footprint and a shorter computation time.
- The article offers suggestions for new lines of inquiry based on the findings, with the goal of developing AI algorithms that are reliable and effective, and which are capable of recognising faces in their natural environments.

The following outline has been chosen for this research article: The techniques of image processing and deep learning are utilised in the discussion of the current state of the art (SOTA) of face mask identification in section 2. In addition to this, it addresses the problems and difficulties associated with face mask detection approaches. In section 3, you will find a full presentation of the research methods. It describes the anticipated study endeavour and includes an in-depth architectural breakdown of the architecture. Section 4 provides an in-depth presentation of the datasets that were utilised for this body of research work, as well as the experimental setup, trial findings, and related examinations. The research is ended in section 5. It comprises a summary as well as the potential improvements that can be made in the future because of the research work.

2. Literature review

William Wells, a professor of engineering at Harvard, and his wife, Mildred Wells, a physician, began using more modern experimental methods to investigate airborne transmission in the 1930s [4]. Infectious particles and droplets containing germs are discharged when a person has a cold or illness [5]. The resolution limit for viruses is extremely tiny, much smaller than the one-to two-micron range and is in the five-to-ten-micron region. There are several examples of respiratory viruses involved; regrettably, this happens to them as well [6]. During a few pandemic pathogenicity outbreaks, numerous attempts to prevent or reduce airborne infections were made during this time. This contrasts from all other types of airborne diseases, which are distributed by air pollutants and may move up to 1 m in the air for lengthy periods of time,

in that airborne infections must be trapped within droplets to transmit and spread.

It was previously thought that COVID-19 was transmitted through the respiratory system; however, research then showed that airborne transmission was the primary mode of infection [7]. To control airborne infections effectively, a wide range of strategies needed to be employed. This is necessary since diverse organisations hold varying opinions regarding the efficacy of certain strategies. Both air circulation and droplet dispersion have been the subject of a significant amount of research [8], with the goal of gaining a deeper comprehension of how to achieve maximum droplet dispersion. The term "respiratory system" refers to the mechanism within the body that is responsible for breathing. When utilised near the source of an illness, active protective techniques such as respiratory area cleaners and air purifiers can assist in reducing the risk of exposure [9].

Examples of respirators that protect the user from infection while still offering occupational safety include face masks and N95 respirators [10]. There is a greater need for respirators even though face masks are less expensive and simpler to use, and more testing of their effectiveness has further improved those requirements [11]. Their initial goals were wound dressings, nontemporal surgical exposure to airborne microbes [12], other significant issues regarding airborne infectious illnesses like TB, particularly among the public, and traditional treatments including the use of masks.

Some studies try to prove that face masks have been demonstrated to be of little benefit [13], even if there is no evidence that they reduce the prevalence of respiratory infections and some of the assumptions about their efficacy are up for debate. But new studies have shown that using a face mask can help prevent the flu, even in otherwise healthy individuals. Although face masks were more useful but had limited efficacy, a systematic review of randomised controlled trials evaluating the effectiveness of respirators and face masks against respiratory viruses found that nurses were better protected. Respirators worn by healthcare workers also seemed to be effective, but the results only indicated temporary efficacy.

Early on, the researchers largely concentrated on a few techniques, such as the Gray-scale facial picture algorithms [14]. While AdaBoost, one of the top classifiers at the time, processed the initial information of the face models, a small number of others concentrated on pattern detection techniques. Real-time face detection was made possible later with the development of the Viola-Jones detector. But it had a problem: the dull, low light made it unable to function effectively, and the low light source made it unable to categorise properly. Additional hypotheses have evolved to elucidate these notions, describing the same views in further depth. It has been revealed without doubt that many viruses can spread by airborne pathways.

In the recent past, there has been a trend toward the development of object recognition algorithms that utilise deep learning models rather than shallow models [15]. This trend has resulted in deep learning models being theorised to be superior to shallow models in terms of their ability to perform complex tasks. A good illustration of this would be the development of a model or system that operates in real time and is able to determine whether or not individuals in public locations are wearing masks.

Real-time deep learning was utilised by Shaik and Ahlam [16] to classify and distinguish emotions, while VGG-16 was used to categorise seven faces. This strategy functions well during the current Covid-19 lock-down phase, which is intended to stop the further spread of the disease. In addition, Ejaz et al. [17] made use of main component analysis to differentiate between people who had their faces masked and those who had their faces exposed.

Using CNN, face mask detection was implemented by Li et al. [18]. It could identify whether an individual was wearing a mask or not to track and enforce compliance to directives given by competent authorities. The authors came up with the HGL approach to categorise head postures by making use of masks for faces. This method involved the examination of the colour pattern in pictures and line portraits.

The condition identification approach was utilised in the construction of Qin and Li's [19] face mask detection model. In the paper, the issue was dissected into four distinct parts: pre-processing the picture, clipping the facial portions, performing the super-resolution procedure, and forecasting the final condition. The utilisation of super-resolution as a means of enhancing the functionality of images of lower quality was the key contribution made by this research.

For facial detection in Ref. [20], the authors make use of the Darknet-53 (YOLOv3 algorithm). Machine learning and artificial intelligence are the two primary components of deep learning, which is largely a blend of the two. In general, it has been demonstrated to be more adaptive and to develop more exact models than ML that is inspired by the operation of brain cells [21].

A method of detection that relies on mobile phones was developed by the authors of [22]. From the Gray-Level Co-Occurrence Matrix (GLCM) of face mask micro images, it was able to obtain three different aspects. The KNN algorithm was utilised to carry out a three-result recognition analysis, which ended up having an accuracy rating of 82.87% overall. The system made use of a gray level co-occurrence matrix to determine the presence of face masks in micro photos. However, given that the model was only compatible with mobile phones, it was clear that it could not be used in all situations.

The authors of the study [23] suggested that a pre-trained MobileNet equipped with a global pooling block ought to be utilised for the purposes of facial recognition and detection. A shaded image is used as the starting point for the creation of a multi-dimensional component map by the pre-configured MobileNet. Because it makes use of an overall pooling block, the model that was proposed does not suffer from the issue of over-fitting.

Facial expressions and face mask detection are related in the sense that they both involve analyzing features of the face. Facial expressions are typically used to infer emotions or mood from the face, while face mask detection is used to determine whether a person is wearing a mask over their face or not. In recent studies [25,26], the authors used directional gradients techniques to recognise the face and expressions of face. Yassine et al. [27] published a review paper on the recent works in face mask detection. The paper described the various parameters that have been used to evaluate the face mask detection.

To hasten the development of automated face mask recognition and social distance measuring on public spaces, the authors in [28] have used a massive dataset made up of 10,000 pictures that have been divided into two categories: those wearing face masks and those who are not. The paper provided a complete pipeline for doing outside real-time face mask identification and social distance measuring.

The following Table 1 depicts the comparison of some notable works in the proposed problem domain based on different parameter. The mentioned articles in Table 1 are the most relevant to the problem

Table 1
Comparison of existing work on face mask detection.

Ref.	Techniques	Datasets	Achievements	Accuracy (%)
[18]	HGL	MAFA	It able to identify the face mask on the side-by-side position.	Front face – 93% Side face – 87%
[23]	MobileNet with a global pooling block for face mask detection	Kaggle	It added global pooling block to perform a flatten of the feature vector.	99
[28]	Improved feature extraction and region proposal generation layers on the YOLO V3 architecture	10,000 real-time images	Improved YOLO-v3 outperformed well than the existing baseline techniques	95

domain.

In the previous studies, the YOLO-v3, MobileNet and Inception were used along with various techniques for detecting the mask on the face. The proposed work aims to reduce the dimensions of the datasets using PCA technique. It is worth noting that there exists a paucity of research on the application of a stacked autoencoder-based deep learning model with principal component analysis for face mask detection. The objective of the proposed study is to construct a deep learning framework utilising MobileNET-v2 architecture in conjunction with a stacked auto encoder. The images undergo pre-processing by PCA, which serves to decrease dimensionality and reduce time complexity.

3. Proposed methodology

Within the scope of this study, an attempt is being made to investigate a variety of significant facets of face mask identification. The objective of face detection is to search for and determine the identities of all faces that may be seen in an image or video. If there are numerous faces, each one will have a bounding box surrounding it. This will allow other people to locate each individual face if there are multiple faces.

The modelling of human faces is difficult due to the large number of variables that might change, including facial expression, direction, illumination, and partial occlusions caused by objects such as sunglasses, scarves, and masks, amongst other things. The outcome of the detection offers the face location parameters, which may be required in a variety of different shapes, such as a rectangle containing the middle section of the face, eye centres, or characteristics such as the corners of the mouth and nose, the eyebrows, the nostrils, etc.

It is common practise to identify things based on the unique qualities that they possess. A human face can be distinguished from a wide variety of other things due to the fact that it possesses a number of distinctive traits. It does this by isolating structural features of the face, such as the eyes, nose, and mouth, and then utilising those features to determine the identity of a face. A statistical classifier is helpful in most situations for distinguishing between areas of the face and areas that do not contain faces. Additionally, human faces have distinct textures that can be used to differentiate them from the textures of other objects. In addition, one can identify elements on a face by looking at the boundaries of the characteristics that make up that item. In the following section, it will utilise OpenCV to design a feature-based approach, and then it will use NumPy to evaluate that approach.

Image-based systems frequently utilise statistical analysis and machine learning approaches to detect the significant characteristics of photographs containing faces and pictures containing objects other than faces. When it comes time to detect faces, the learned characteristics are use in the form of distribution models or discriminant functions. This strategy implements several different methodologies, such as neural networks, HMM (Hidden Markov Model), SVM (Support Vector Machine), and AdaBoost learning, among others. In the following section, we will investigate how to use MTCNN, which stands for Multi-Task Cascaded Convolutional Neural Network and is an image-based approach of face recognition.

In this research work, a new method of facial identification has been developed, and it makes use of complex depth-wise separable convolutional neural networks in conjunction with principal component analysis.

3.1. Stacked autoencoder based depth separable convolutional neural network model for face mask detection

A sophisticated framework successfully developed has been demonstrated in this article for the purpose of identifying masks in the face of individuals. The learning architecture generates results by classifying the input picture according to whether it has a mask. These findings are output. A warning is displayed if an individual is not wearing a mask or is not correctly applying it to their face. If the user puts on a mask and applies it correctly, the remaining procedures can continue. Because of this, everyone will be safe from the respiratory droplet transmissions so long as they wear their masks appropriately. In this way, the utilisation of this technology will contribute to the reduction in the proliferation of any droplet-based transmissions. If the system recognises the user's face while they are not wearing a mask or when the mask is not worn in the correct manner, it will notify the user and sends an alert to the governance system. The workflow flowchart for the complicated construction is depicted in Fig. 1.

The proposed work starts out by having the datasets imported into it. Image processing scrubs the unprocessed data and sends it on to subsequent steps where it is augmented with further information. The datasets are first partitioned into train and test data before the augmentation of the data takes place. The well-known ratio of 80:20 is utilised for both the test data and the train data. The data from the train is modified before being fed into the suggested deep learning model known as DWSC-PCA-SAE (Depth-wise Separable Convolutional Neural Network using Principal Component Analysis and Stacked Autoencoder). The model is put through its work, and the outcomes are assessed in further sections.

3.2. Image pre-processing

Before moving on to the testing and training of the model, it is necessary to complete this crucial phase, which assists in the processing of the data. As soon as it reaches this point, it will initiate the process of turning all of the images that are contained within the dataset files into arrays. It will develop the deep learning module by utilising these arrays as its building blocks and all the relevant tools are going to be imported from the corresponding modules. Then, following the completion of the data processing and the delivery of the conclusions to user in the form of labels and then it will proceed to construct the variables and objects.

The device's built-in camera's video input will be recognised by the Haar cascade classifier. Before proceeding to the next stages, pre-processing is necessary for video and image data captured by the system's camera. The RGB colour image contains a lot of extraneous data that is not required for mask detection, thus during pre-processing the image is changed to a grayscale image. To maintain the input picture consistency across the model, it then scaled the image to 224×224 pixels. The acquired pictures are then normalised, resulting in pixel values that range from 0 to 1. The normalization helps the learning system understand and pick up the necessary information from the photos more quickly.

3.3. Data augmentation

The training of the DWSC-PCA-SAE model requires an enormous quantity of data in order to be carried out effectively. This is because there is not currently adequate data for training the model that has been suggested. This is due to the requirement of providing the model with an acceptable quantity of data while it is being trained. The method of data augmentation is going to be utilised to get to the bottom of this issue. In this process, the picture is transformed in a variety of ways, such as by rotating it, zooming it, moving it, severing it, and flipping it, to obtain several different copies of the same image. Specifically, the image is sheared, moved, rotated, and flipped. In the model that has been described, the process of data augmentation includes the utilisation of image augmentation as one of its components. Creating a portrait called "image data creation" is a function that is designed for the goal of improving images. Test and training sets of data are both returned by this function. The parameters for data augmentation are outlined in Table 2, which may be found here.

3.4. Dimensionality reduction

This method not only ensures that an adequate quantity of

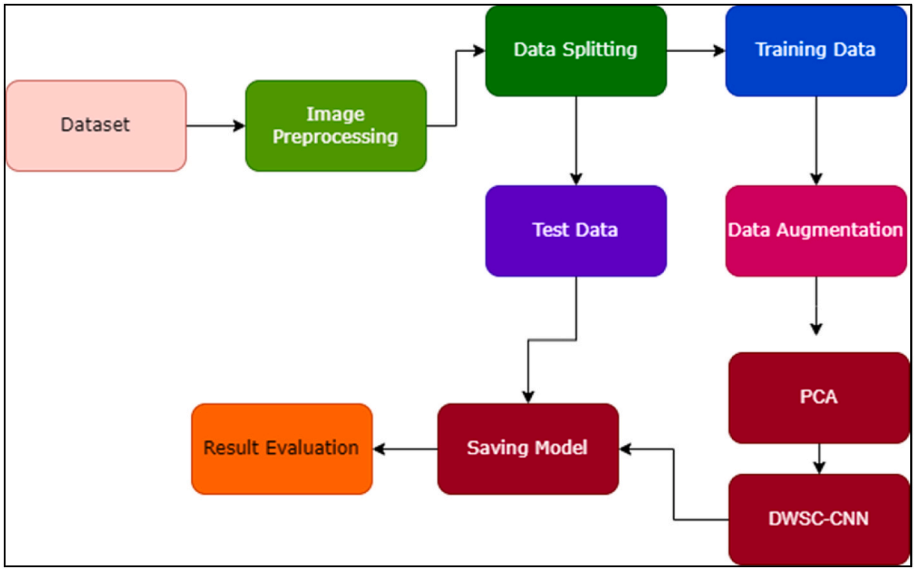


Fig. 1. Proposed workflow.

Table 2
Parameters for data augmentation.

Technique	Parameter
Rotation	-30^{o} to 30^{o}
Gaussian Blur (in %)	25, 50, 75, 100
Sharpness (in %)	50, 100, 150, 200
Emboss (in %)	50, 100, 150, 200
Shearing X and Y axis	10^{o}
Gaussian Noise	0.1 to 0.9
Jitter	−4 to 4
Flips	Horizontal and Vertical

information is preserved for deep learning, the extraction of picture features, and the categorization of images, but it also shortens the amount of time that is required for training and testing. In its most basic form, data dimensionality reduction is the process of mapping data from their original, high-dimensional space to their final, low-dimensional space destination. Depending on whether or not labels are used in the process, this procedure can be further broken down into nonlinear or linear or supervised or unsupervised dimensionality reduction. The principal component analysis (PCA) and the stacked autoencoder are the two dimensionality reduction components that are utilised in this research work.

3.4.1. Principal component analysis (PCA)

PCA is among the best approaches for dimensionality reduction in the area of unsupervised dimensionality reduction. Simply decomposing the data's eigenvalues is all that is required to reach the objectives of data compression as well as the removal of redundancy. An image is another type of data that has a high degree of correlation. When it comes to the extraction of characteristics from pictures, PCA performs significantly better. The image matrix is subjected to a variety of processes, after which it is turned into an eigen subspace with less dimensions. After that, the covariance matrix is computed using the lower-dimensional matrix, as a starting point for computation. The covariance matrix is a representation of the relative variation that exists between individual pixels in an image. After that, Eigen vectors are computed using the covariance matrix as a starting point. The Eigen vectors that have the highest values are the ones that are taken into consideration to be the principal components.

In PCA, every new feature is a linear combination of previous features, and it brings the total number of features from the original '*x*' features down to a smaller set of '*y*' features. These linear combinations have as their goal the decoupling of the new y features from one another while simultaneously increasing the sample variance. The process involves in the PCA is given in the algorithm 1.

In order to perform face recognition, it must first train a dataset and then apply PCA to the trained dataset. In the beginning, it will need to transform the train dataset into a face vector. Every two-dimensional train picture (x, y) is transformed into a one-dimensional face vector (x·y by 1). After all the pictures have been converted, they are loaded into the stack-autoencoder process. In this body of work, the total number of 'n' images is taken. Here, the value of 'n' is determined by the total number of pictures generated after the augmented data for training, and 'P' is a variable that loads each image. The mean value of augmented data is evaluated by using the following equation (1),

$$\overline{M} = \frac{1}{n}\sum_{i=0}^{n-1} Pix(n) \tag{1}$$

where $\overline{M}$ is the mean value of images taken, *Pix(n)* represents the pixels of nth image.

Image normalization is the important process to generate normalize data. Normalization involves taking each image in the collection and subtracting the value of image that appears on average within those images. The final result of the normalization process is stored in the variable denoted by δ_n given in equation (2). After that, a new matrix 'T' is produced according to equation (3).

$$\Delta_n = Pix(n) - \overline{M} \tag{2}$$

$$T \in \{\delta_1, \delta_2, \delta_3, \ldots.., \delta_n\} \tag{3}$$

Following the process of normalization, the 'mask feature' in the image that has been normalised. Later it computes the covariance matrix of the normalised vector by using equation (4). Because of the high computational and memory costs associated with explicitly determining the covariance matrix, the I covariance method is rarely used in practical implementations, particularly when dealing with high dimensional data (large 'n'). This is primarily since it is inefficient. The computational cost of traditional covariance matrix is xy^2 where '*x*' and '*y*' is the number of row vectors and the number of elements in each row respectively. This research work uses covariance-free approach. The computational cost of covariance-free approach is $2 \bullet (xy)$.

$$Cov(T) = T^X(Tr) = \left\{ \begin{matrix} T^X_{x*y} \\ T_{y*x} \end{matrix} \right\}_{x*x} \tag{4}$$

Eigen value is calculated by using equation (5).

$$|T - \lambda| = 0 \tag{5}$$

where T is the matrix and λ is the eigen value.

Cov(T) is a big square matrix. Calculating the eigenvalue and eigenvector helps reduce the size of the mask/non-mask space vector, which improves both the efficiency and accuracy of the calculation. To achieve recognition, it is necessary to compute a weight, which is then compared with the weight of the test image. The following equation (6) helps to calculate the weight.

$$\omega_n = \lambda^X(n - \overline{M}) \tag{6}$$

where ω is the weight of image, λ is the eigen vector, X is the matrix and $\overline{M}$ is the mean value.

Algorithm 1. Dimensionality Reduction

1: Load the training data.
2: Compute mean value using equation (1)
3: Compute data normalization using equation (2)
4: Find covariance-free matrix using equation (4)
5: Find Eigen vectors and values using equation (5)
6: For value in eigenvalue's:
7: Find the greatest eigenvalue.
8: Find the weight of image using equation (6)

3.4.2. Stacked autoencoder (SAE)

Encoding and decoding are the two parts that make up the Auto-Encoder (AE) model, which is a representation of an unsupervised neural network. During the encoding stage, the implicit properties of the input data are learned, and through the stage of decoding, the objective is to reproduce the correct data (input) by making use of the new features that have been learned. Because the neural network model can understand new features more correctly and carry out the job of feature extraction, the capacity to represent features in the data that is processed by AE is significantly improved. Only data that are equivalent to the training data can be compressed since the data that are created by AE are correlated with one another. A particular encoder is trained with the use of input from a particular class in order to accomplish the objective of autonomous learning from data samples. The purpose of the learning process in AE, which falls under the domain of unsupervised learning, is to restore input without the need of labelling. One may consider it to be composed of three layers: input layer, hidden layer, and output layer. Both the output layer and the input layer have the same data size in their respective scales. Fig. 2 represents the architecture of autoencoder.

The encoder's output of the hidden layer feature, also known as the coding feature, may be thought of as a characterisation of the data D that was entered into the encoder. At the same time, the feature of the hidden layer is the feature that is acquired by the dimensionality reduction of the encoder. In this case, the data of the hidden layer H have a smaller dimensionality than the data of the input layer D and the data of the output layer D_o; this is shown by the fact that |D| is greater than |H| and |D_o| and that |D| is equal to |D_o|. First, calculate H using the mapping matrix H = x(D), which will take you from the input layer D to the hidden layer H. Next, calculate D_o using the mapping matrix $D_o = y$(H), which will take you from the hidden layer H to the output layer D_o. The transformation may be represented by equation (7).

$$x : \alpha \to \delta$$

$$y : \delta \to \alpha$$

$$x, y = \underset{x,y}{\operatorname{argmin}} ||D - y[x(D)]||^2 \tag{7}$$

The magnitude of the hidden space, denoted by δ, is denoted by the letter α, which stands for the embedding input space (which is also known as the output space). The input space D is the element of α and the characteristic space H is the element of δ are both sent to the self-encoder, and it is its job to solve the mapping (x, y) that exists between the two spaces in order to minimise the reconstruction error of the input feature.

There is a possibility that a single autoencoder will be unable to bring down the dimensionality of the input features. As a result, for situations like these, we make use of stacked autoencoders. As its name indicates, stacked autoencoders are a collection of encoders that are arranged in a vertical stack. The stacked autoencoder works based on the following principle.

i. Train the first autoencoder based on the input data, and then acquire the feature vector that was learnt.
ii. The feature vector from the previous layer is then utilised as the input for the layer that comes after it, and this process is continued until the training is finished.
iii. After all of the hidden layers have been trained, the back-propagation algorithm (BP) is utilised to perform fine-tuning. This is accomplished by minimising the cost function and updating the weights using labelled training sets.

3.5. Development of deep learning architecture

The architecture is developed as a less complex and effective

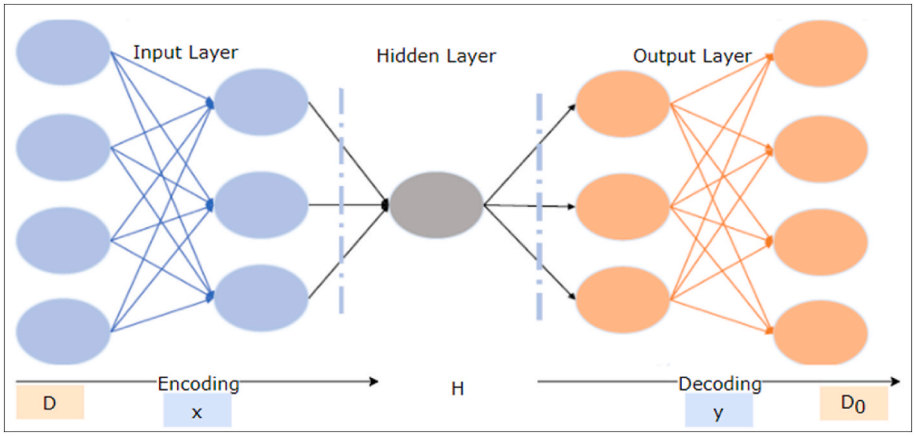


Fig. 2. Architecture of autoencoder.

Convolutional Neural Network (CNN) model utilising libraries such as TensorFlow with Keras and the OpenCV to determine whether the test subjects are donning a face mask to protect themselves. The 'work's various aspects are all explained here. From the provided samples, the deep learning architecture begins to distinguish and classify several significant non-linear properties. To estimate the training pattern issues that are posed to it other than examples fed using the process depicted in Fig. 3, this informed architecture is employed. The pseudocode of the proposed DWSC-PCA-SAE algorithm is given below.

Algorithm 2. DWSC-PCA-SAE

- Input: Image files
- Output: Mask or No-Mask

1: Load the dataset
2: Load OpenCV
3: Find the face in the datasets
4: Generate data augmentation
5: Perform PCA to reduce dimension
6: Build DWSC-CNN using Stacked Autoencoder
7: Load MobilNetv2
8: Train model
9: Test model
10: Classify the output

3.5.1. *Depth-wise separable convolutional neural network*

Fig. 3 provides a visual representation of the multi-layered characteristics, as well as a comparison between traditional and depth-wise separable features. When combined, the depth-wise (dw) and point-wise (pw) convoluted structures produce what is known as a "Depth-wise separable" convoluted structure. The standard convolutional structure is outperformed by the Depth-wise separable convolutional structure, which offers a function that is equivalent to the previous method but operates at a much quicker pace. Because the frames may be separated from one another in terms of depth, the offered approach does not have a pooling layer in between them. To minimise the impact of the spatial dimension, a few of the depth-wise layers each include a stride of two. In this instance, the collection of output channels is also included into the point-wise layer that comes after it.

3.5.2. *DSC-PCA-SAE architecture*

The CNN principles, which are essential for seeing patterns in pictures, are used in the learning model. The neural network must be able to view data from both classes. An input layer, many hidden layers, and an output layer make up a network. There are several levels of convolution in hidden layers. Numerous dense neural networks employ the characteristics collected from CNNs for categorization. There are three pairs of convection levels in each of the 32 formations, followed by the highest pooling level. 100 kernels with a 3x3 window size and a maximum pooling level of 2x2 window size make up the convolution layer. The highest value inside the 2x2 frame will be chosen by this layer after combining the results of the preceding convolution level.

The number of parameters is decreased because of the local level of performance being decreased. The computations for the network become easier consequently. The resolution 'evel's output will be flattened and transformed into a 1-D array. Consequently, there are two thick layers and a dropout layer. The dropout layer eliminates the exclusion level drive to prevent network overfitting. Each neuron in the dense layer picks up nonlinear features as they are learned. The first dense layer of 50 knots will contain the flat outcome. Finally, there is another thick layer with two nodes that has two classes: one with a mask and the other without.

For unlabelled input, the autoencoder consists of two core models: an encoder and a decoder. The input feature map is encoded using an encoder, and then the feature map is decoded and rebuilt using a decoder module. For a more comprehensive modelling of the features in this study, it squeezed the output feature vector from the MobileNetV2 model using the encoder function of the stacked autoencoder. Global average pooling is used to decrease the output dimensions of the MobileNetV2 model, which are $7 \times 7 \times 1280$, to 1280 dimensions. The suggested encoding technique is then given access to the output of the global average pooling to further extract more representative components for the last stage classification. The features vectors 1280 dimensions are first converted to 640 and then to 320 dimensions. The autoencoder complexity is mostly decreased by feature encoding utilising their halves.

In this research work, it employed stacked auto-encoding layers to reduce the size of the MobileNetV2 output feature vector while maintaining an abstract representation of all feature mappings. To use ReLU and softmax in a stacked autoencoder, it typically applies the ReLU activation function to the hidden layers of the encoder and decoder, and apply the softmax activation function to the output layer of the decoder when the task is multi-class classification. The weights, which include a bias term and an ReLU activation function (given in equation (8)) are multiplied with the data in the stacked autoencoder encoding module.

$$\mathrm{ReLU} : f(n) = \max(0, n) \tag{8}$$

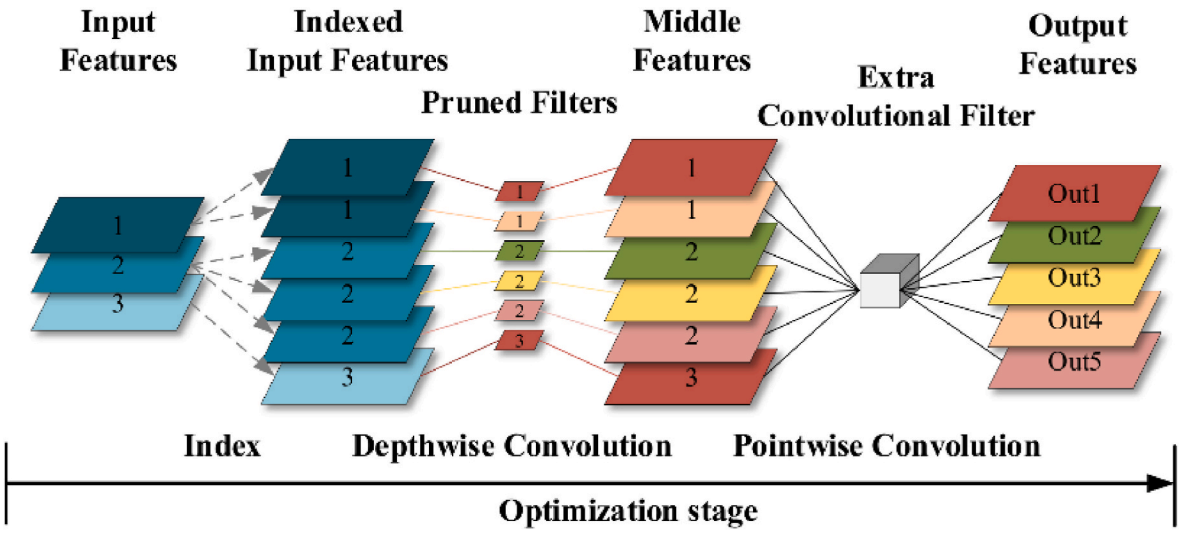


Fig. 3. Architecture of depth-wise convolutional neural network.

where 'n' represents the positive value return by the function and false positive return '0'. The value of rectified linear unit is between '0' to infinity.

The first encoding layer in the suggested stacked encoded layers employs MobileNetV2's output feature vector, while the next layer stacks feature from the first layer. To understand the encoded characteristics before the classification layer, two fully linked (Dense) layers receive the output of the encoding layers. In the output layer, softmax is used as the activation function and it is given in equation (9).

$$softmax(n_j) = \frac{\exp(n_j)}{\sum_k \exp(n_k)} \tag{9}$$

where k is the number of classes. This means that for each class j, we take the exponent of the input n_j and divide it by the sum of the exponentials of all the inputs. This ensures that the outputs add up to one, and can be interpreted as probabilities of each class.

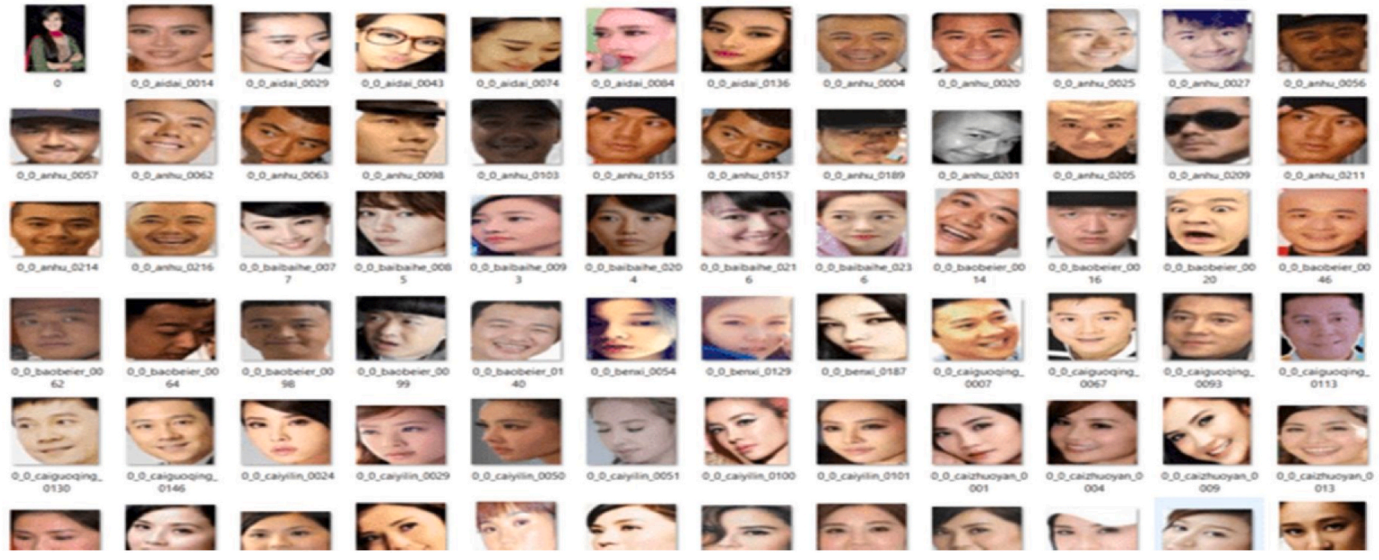
The overall proposed model can be expressed mathematically in equation (10).

$$y = f(W_2 \circ f(W_1 \circ P_x + b_1) + b_2) \tag{10}$$

where:

- y is the binary classification output indicating whether the input image contains a mask or not.
- P is the PCA projection matrix of size (HWC) x k. H-Height, W-Width and C-number of channels.
- x is the PCA pre-processed image.
- W_1 and W_2 are the weight matrices of the DSCNN model.

(a) With Mask



(b) Without Mask

Fig. 4. Dataset bifurcation in categories (a) with-mask, (b) without-mask.

- b_1 and b_2 are the bias vectors of the DSCNN model.
- f is the activation function.
- + denotes element-wise addition.
- ° denotes element-wise multiplication.

4. Results and disscussions

4.1. Dataset description

For training the proposed deep learning model, it has gathered a variety of photos. CNN plays a significant role in the learning strategy's framework. The dataset consists entirely of facial photos. It comprises 658 pictures in the category without masks and 670 pictures in the category with masks. Ninety percent of the photos in each class are used for training, while the remaining ten percent of each dataset are utilised for testing. The numerous objects under each category—with mask and without mask—are shown in Fig. 4.

The outcomes of the experiment are thoroughly explained in this section. The Keras framework with TensorFlow as the backend is used for all experiments, which runs on a GPU with 64 GB of RAM. The datasets utilised for each model's assessment, evaluation metrics, are all included in this part. The further sections provide a brief description of each of these sections.

4.2. Evaluation metrics

The comparison of the effectiveness of proposed algorithm makes use of a confusion matrix. By combining the values of True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP), this matrix is used to generate various metrics. The performance metrics used to assess models using the confusion matrix are listed below.

Accuracy is a measure of a model's estimated value corresponds to its real or actual value, which corresponds to the proportion of all samples that are properly categorised. The model's accuracy is determined using the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

Precision reveals the proportion of relevant occurrences among the chosen examples that are truly affirmative. To determine accuracy, apply the formula below:

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

The percentage of genuine positives that are accurately detected is determined by recall or true positive rate (TPR). The following equation is used to calculate recall:

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

The harmonic mean of accuracy and memory, which sums the weighted average of precision and recall, is how the F1 score is often understood. The F1-score is calculated using the following formula:

$$F - Score = \frac{Precision * Recall}{Precision + Recall} \tag{14}$$

The proposed work is experimented using Windows 10 computer, an Intel Core i5 processor, and Nvidia GTX 1080 GPU for the tests. Programming was carried out using Python 3.8. The pre-trained model is using 224x3 STIF frames. The proposed approach is tested using this dataset. Datasets for training and testing have been produced. The scaling factor was first set at 0.001, and after that, it decreases by a factor of 0.9 every 10 epochs. The Adam optimizer uses a momentum value of 0.999. The training process is repeated up till 100 epochs have transpired. The layout of the recommended MobileNetV2 parameter settings for efficient face picture classification (mask/no-mask) is used. Models for categorization were tested using the 10-fold cross-validation method. Table 3 represents the results obtained with the training model on Adam optimizer.

4.3. Results

Table 4 depicts the comparative analysis of proposed DSC-SAE-MN with AlexNet model. Proposed model outperformed well than AlexNet model. It shown a 1.5% improvement in-terms of accuracy. Figs. 5 and 6 illustrates the comparison of training accuracy with augmented and without augmented data. It shows that a decent improvement in the data augmentation model. Training loss is an important parameter that is used to evaluate the loss of a model. Figs. 7 and 8 represented the training loss of the proposed DSC-SAE-MN model. The proposed model is compared with Alexnet [24] and the result is given in Fig. 9.

A classification model can be evaluated in accordance with the Receiver Operating Characteristic Curve (ROC) (Fig. 10), which states that this evaluation can be done by comparing the model's true positive rate against its false positive rate at a variety of threshold values. Because lower values on the x-axis indicate a lesser number of false positives and a higher number of true negatives, the fact that the curve is located in the top-left corner is seen as a sign of excellent performance. On the other hand, a higher significance level on the y-axis indicates a lower number of false negatives and a higher number of real positives.

The time complexity of using PCA with DWSC for detecting masks in the face will depend on several factors, including the size of the input images, the number of principal components selected, and the architecture of the DWSC. PCA involves computing the eigenvectors and eigenvalues of the covariance matrix of the input data. The time complexity of this step is typically $O(n^3)$, where n is the number of features in the input data. However, there are faster algorithms such as the Power iteration method that can reduce the time complexity to $O(k * n^2)$, where k is the number of principal components selected.

The time complexity of the DWSC architecture will depend on the number of layers and the size of the filters used in each layer. Typically, the time complexity of a single convolutional layer is $O(k * n^{2h})$, where k is the number of filters, n is the size of the input feature map, and h is the size of the filter. However, depth-wise separable convolutions are more efficient than standard convolutions, as they require fewer parameters and computations. However, in general, DS-CNNs are known for their efficiency, and PCA can help to further reduce the dimensionality of the input data, potentially improving the overall speed and performance of the system. The proposed DWSC-SAE technique was tested on MobileNet and AlexNet architectures, and the corresponding time consumption was 17.49 ms and 53.49 ms, respectively.

5. Conclusion

The Mobile Net-based Depth-wise Separable Convolution Neural Network using Principal Component Analysis and Stacked Autoencoder (DWSC-PCA-SAE) is proposed as a method for detecting masks in face pictures within the scope of this research. On several datasets, it

Table 3
Training model on adam optimizer.

Epochs	Loss	Validation Loss	Accuracy	Validation Accuracy
10	29.38	27.28	86.12	91.31
20	24.08	26.98	88.02	91.81
30	18.78	23.38	89.72	92.31
40	11.48	20.88	91.32	92.81
50	8.58	18.38	92.72	93.31
60	5.68	15.88	94.62	93.81
70	3.78	13.38	96.12	94.31
80	2.88	10.88	97.02	94.81
90	1.98	8.38	98.92	95.31
100	1.08	5.88	99.12	95.31

Table 4
Comparative Analysis of proposed method with existing methods.

	Precision	Recall	F-Score	Accuracy
AlexNet	0.966587	0.935335	0.950704	0.927959
Proposed	0.976134	0.944573	0.960094	0.941681

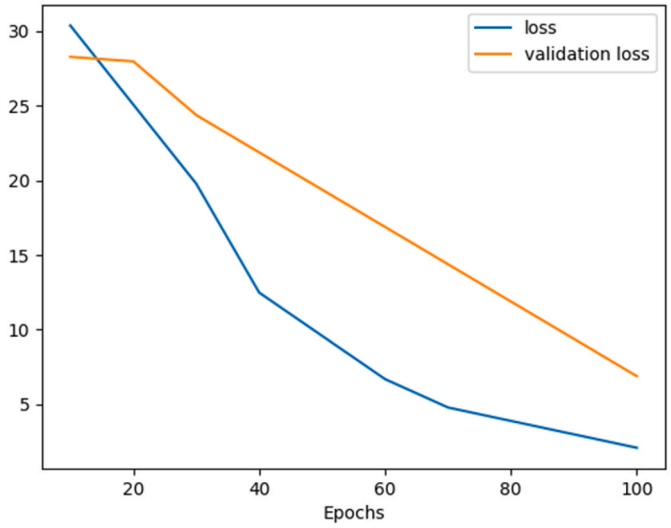


Fig. 5. Results of training loss and validation loss – without data augmentation.

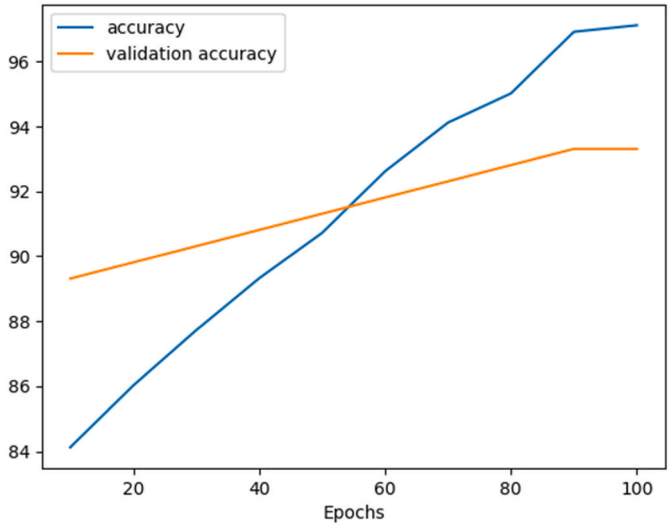


Fig. 6. Results of accuracy and validation accuracy – without data augmentation.

evaluates the results of the research in comparison to the original convolutional filters. According to the findings of the trials, the proposed system performed much better than the existing conventional convolutions. In addition, past work on a motivated baseline approach is compared with the strategy that has been offered here. According to the results, which include an accuracy score of 94.1%, a precision score of 97.6%, a recall score of 94.4%, and an F-score of 96.01%. The DWSC-PCA-SAE algorithm generates the greatest overall performance across a range of evaluation criteria. This model needs additional processing to create visualisations, and due to the limitations of the dataset, it is unable to differentiate between appropriate and inappropriate mask application. In the future, one of our goals is to provide face mask identification datasets that include a variety of various mask wearing

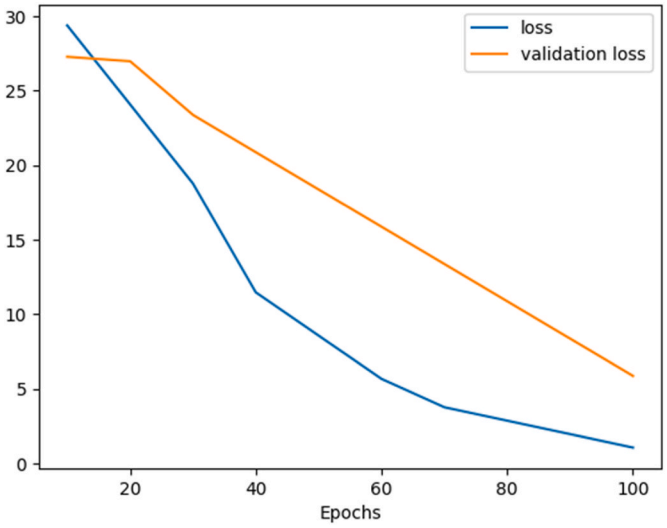


Fig. 7. Results of Training loss and validation loss.

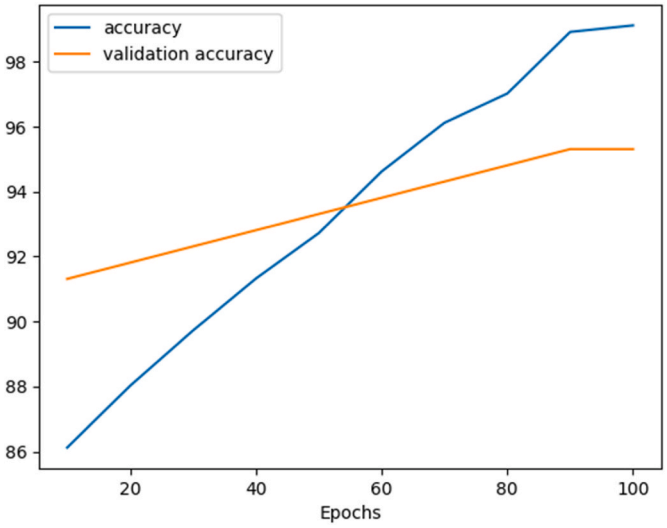


Fig. 8. Results of accuracy and validation accuracy.

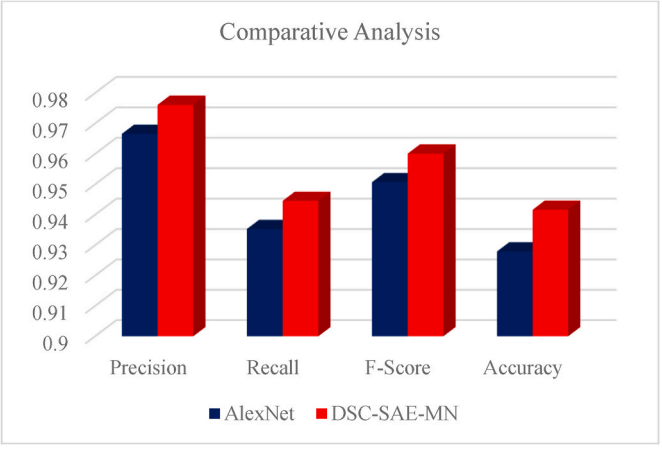


Fig. 9. Comparative analysis of DWSC-SAE-MN with AlexNet.

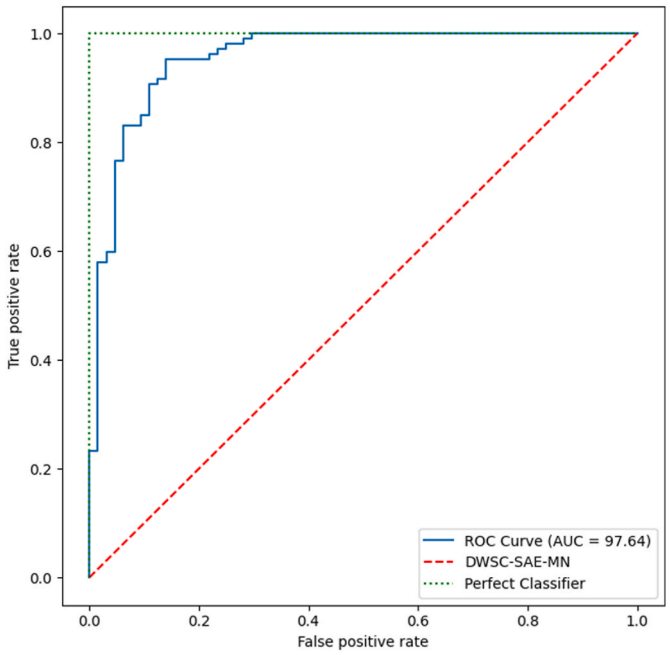


Fig. 10. Comparison of true positive rate with false positive rate.

states. Alternatively, in the future, the work needs to use zero-shot learning to have the design recognise incorrect mask wearing states.

Authorship statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *Array-Open Access-Elsevier*.

Authorship contributions

Category 1

Sundaravadivazhagan Balasubaramanian, Robin Cryic: Conception and design of study, Robin Cyriac, Sahana Roshan, Boby Chellanthara Jose,: Acquisition of data, Sundarvadivazhagan Balasubaramanian, Kulandaivel Maruthamuthu Paramasivam: analysis and/or interpretation of data.

Category 2

Drafting the manuscript: Kulandaivel Maruthamuthu Paramasivam, Boby Chellanthara Jose, revising the manuscript critically for important intellectual content: Sundarvadivazhagan Balasubaramanian, Robin Cyriac.

Category 3

Approval of the version of the manuscript to be published (the names of all authors must be listed):

Sundarvadivazhagan Balasubaramanian, Robin Cyriac, Sahana Roshan, Kulandaivel Maruthamuthu Paramasivam, Boby Chellanthara Jose.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sundaravadivazhagan Balasubaramanian reports financial support was provided by University of Technology and Applied Sciences - Al Musannah. Sundaravadivazhagan Balasubaramanian reports financial support was provided by The Research Council-(TRC)-Block Funding-Research Grant-Oman.

Data availability

The data that has been used is confidential.

Acknowledgements

All persons who have made substantial contributions to the work reported in the manuscript have agreed to published the document in *Array-Open Access-Elsevier Journal.*

References

[1] Woods A, Daily News B, Jun. Britain faces an anxiety crisis as people return to work. 2020. https://bdaily.co.uk/articles/2020/06/22/britainfaces-an-anxiety-crisis-as-people-return-to-work.

[2] Howard J, Huang A, Li Z, Tufekci Z, Zdimal V, van der Westhuizen H, von Delft A, Price A, Fridman L, Tang L, Tang V, Watson GL, Bax CE, Shaikh R, Questier F, Hernandez D, Chu LF, Ramirez CM, Rimoin AW. Face masks against COVID-19: an evidence review. Preprints 2020:2020040203. https://doi:10.20944/preprints202004.0203.v1.

[3] Laurent Sifre, Mallat Stéphane. Rigid-motion scattering for texture classification. Int J Comput Vis 2014. https://doi.org/10.48550/arXiv.1403.1687.

[4] Jimenez JL, Marr LC, Randall K, Ewing ET, Tufekci Z, Greenhalgh T, Tellier R, Tang JW, Li Y, Morawska L, Mesiano-Crookston J, Fisman D, Hegarty O, Dancer SJ, Bluyssen PM, Buonanno G, Loomans MGLC, Bahnfleth WP, Yao M, Sekhar C, Wargocki P, Melikov AK, Prather KA. What were the historical reasons for the resistance to recognizing airborne transmission during the COVID-19 pandemic? Indoor Air 2022 Aug;32(8):e13070. https://doi.org/10.1111/ina.13070. PMID: 36040283; PMCID: PMC9538841.

[5] Ting Daniel Shu Wei, Lawrence Carin, Dzau Victor, "Tien Y Wong. Digital technolog" and COVID-19. Nat Med 2020;26(4):459–61.

[6] Leung Nancy HL, Chu Daniel KW, Shiu Eunice YC, Chan Kwok-Hung, McDevitt James J, Hau Benien JP, Yen Hui-L"ng, et al. Respiratory virus shedding in exhaled breath and efficacy"of face masks. Nat Med 2020;26(5):676–80.

[7] Ather B, Mirza TM, Edemekong PF. Airborne precautions. [Updated 2022 aug 29]. In: StatPearls [internet]. Treasure island (FL). StatPearls Publishing; 2022 Jan. Available from: https://www.ncbi.nlm.nih.gov/books/NBK531468/.

[8] Bae Seongman, Kim Min-Chul, Kim Ji Yeun, Cha Hye-Hee, Lim Joon Seo, Jung Jiwon, Kim Min-"ae, et al. Effectiveness of surgical and cotton masks in blocking SARS–CoV-2: a controlled comparison"in 4 patients. Ann Intern Med 2020;173(1):W22–3.

[9] Tang JW, Li Y, Eames I, Chan PK, Ridgway GL. Factors involved in the aerosol transmission of infection and control of ventilation in healthcare premises. J Hosp Infect 2006 Oct;64(2):100–14.

[10] Paton R, Tolhurst N, Perisa M, Dempsey K, Tallon J. What mask to use? Aust Nurs Midwifery J 2014 Nov;22(5):31.

[11] Emma P, Martin C. David Grass, Isaac Henrion, Warren S. Warren and Eric Westman. Low-cost measurement of face mask efficacy for filtering expelled droplets during speech. Sci Adv 2 Sep 2020;6(36). https://DOI:10.1126/sciadv.abd3083.

[12] Interim Infection Prevention and Control Recommendations for Healthcare Personnel During the Coronavirus Disease 2019 (COVID-19) Pandemic. Internet source. 2022. https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control-recommendations.html. Update on September 23.

[13] Javaid Mohd, Abid Haleem, Raju Vaishya, Bahl Shashi, Suman Rajiv, Vaish A'hishek. Industry 4.0 technologies and their applications in fighting COV"D-19 pandemic. Diabetes Metabol Syndr: Clin Res Rev 2020;14(4):419–22.

[14] Loey M, Manogaran G, Taha MHN, Khalifa NEM. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement 2021;167:108288.

[15] 21. Shashi Y. Deep learning based safe social distancing and face mask detection in public areas for covid-19 safety guidelines adherence Int J Res Appl Sci Eng Technol 2020;8(7):1368–75. https://doi.org/10.22214/ijraset.2020.30560.

[16] Hussain Shaik Asif, Ahlam Salim Abdallah Al Balushi. A real time face emotion classification and recognition using deep learning model. In: Journal of physics: conference series, vol. 1432. Bristol: IOP Publishing; 2020, 012087.

[17] Ejaz MS, Islam MR, Sifatullah M, Sarker A. Implementation of principal component analysis on masked and non-masked face recognit^{io}n. In: 2019 1St international conference on advances in science, engineering and robotics technology (ICASERT). IEEE; 2019. p. 1–5. https://doi.org/10.1109/ICASERT.2019.8934543.

[18] Li S, Ning X, Yu L, Zhang L, Dong X, Shi Y, He W. Multi-angle head pose classification when wearing the mask for face recognition under the covid-19 coronavirus epidemic. In: 2020 International conference on high performance big data and intelligent systems (HPBD&IS). IEEE; 2020. p. 1–5. https://doi.org/10.1109/HPBDIS49115.2020.9130585.

[19] Qin B, Li D. Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19. Sensors 2020;20(18):5236. https://doi.org/10.3390/s20185236.
[20] Li C, Wang R, Li J, Fei L. Face detection based on yolov3. In: Recent trends in intelligent computing, communication and devices. Springer; 2020. p. 277–84. https://doi.org/10.1007/978-981-13-9406-5 34.
[21] Ochin S. Deep challenges associated with deep learning. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE; 2019. p. 72–5. https://doi.org/10.1109/COMITCon.2019.8862453.
[22] Chen Y, Hu Menghan, Hua C, Zhai G, Zhang J, Li Q, Yang SX. Face mask assistant: detection of face mask service stage based on mobile phone. IEEE Sensor J 2021;21(9):11084–93. https://doi.org/10.1109/JSEN.2021.3061178.
[23] Venkateswarlu IB, Kakarla J, Prakash S. Face mask detection using mobilenet and global pooling block. In: 2020 IEEE 4Th conference on information & communication technology (CICT). IEEE; 2020. p. 1–5. https://doi.org/10.1109/CICT51604.2020.9312083.
[24] Krizhevsky Alex, Sutskever Ilya, Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks.". Adv Neural Inf Process Syst 2012.
[25] Ayeche F, Alti AHDG, HDGG. An extensible feature extraction descriptor for effective face and facial expressions recognition. Pattern Anal Appl 2021;24:1095–110. https://doi.org/10.1007/s10044-021-00972-2.
[26] Ayeche Farid, Alti Adel. Local directional gradients extension for recognising face and facial expressions. Int J Intell Syst Technol Appl 2023;20(No. 6). https://doi.org/10.1504/IJISTA.2022.128525.
[27] Himeur Y, Al-Maadeed S, Varlamis I, Al-Maadeed N, Abualsaud K, Mohamed A. Face mask detection in smart cities using deep and transfer learning: lessons learned from the COVID-19 pandemic. Systems 2023;11:107. https://doi.org/10.3390/systems11020107. 2023.