Face mask detection using deep convolutional neural network and multi-stage image processing

Article *in* Image and Vision Computing · March 2023

DOI: 10.1016/j.imavis.2023.104657

8 authors, including:

Muhammad Umer
Islamia University of Bahawalpur
106 PUBLICATIONS **2,930** CITATIONS
SEE PROFILE

Reemah Alhebshi
King Abdulaziz University
17 PUBLICATIONS **150** CITATIONS
SEE PROFILE

Shtwai Alsubai
Prince Sattam Bin Abdulaziz University
166 PUBLICATIONS **1,512** CITATIONS
SEE PROFILE

Imran Ashraf
Yeungnam University
372 PUBLICATIONS **6,542** CITATIONS
SEE PROFILE

Face Mask Detection using Deep Convolutional Neural Network and Multi-stage Image processing

Muhammad Umer[a], Saima Sadiq[b], Reemah M. Alhebshi[c], Shtwai Alsubai[d], Abdullah Al Hejaili[e], Ala' Abdulmajid Eshmawi[f], Michele NAPPI[g,*] and Imran Ashraf[h,*]

[a]Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan
[b]Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan
[c]Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
[d]Department of Computer Science, College of Computer Engineering and Sciences in Al-Kharj, Prince Sattam bin Abdulaziz University, P.O. Box 151, Al-Kharj 11942, Saudi Arabia.
[e]Faculty of Computers & Information Technology, Computer Science Department, University of Tabuk, Tabuk 71491, Saudi Arabia
[f]Department of Cybersecurity, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudia Arabia
[g]Department of Computer Science, University of Salerno, Fisciano, Italy
[h]Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea

ARTICLE INFO

Keywords:
Real-time surveillance
biometrics
face mask detection
feature extraction
region of interest extraction

ABSTRACT

Face mask detection has several applications including real-time surveillance, biometrics, etc. Face mask detection is also useful for surveillance of the public to ensure face mask wearing in public places. Ensuring that people are wearing a face mask is not possible with monitoring staff; instead, automatic systems are a much better choice for face mask detection and monitoring to help manage public behaviour and contribute to restricting the outbreak of COVID-19. Despite the availability of several such systems, the lack of a real image dataset is a big hurdle to validating state-of-the-art face mask detection systems. In addition, using the simulated datasets lack the analysis needed for real-world scenarios. This study builds a new dataset namely RILFD by taking real pictures using a camera and annotating them with two labels (with mask, without mask) which are publicly available for future research. In addition, this study investigates various machine learning models and off-the-shelf deep learning models YOLOv3 and Faster R-CNN for the detection of face masks. The customized CNN models in combination with the 4 steps of image processing are proposed for face mask detection. The proposed approach outperforms other models and proved its robustness with a 97.5% of accuracy score in face mask detection on the RILFD dataset and two publicly available datasets (MAFA and MOXA).

1. Introduction

Face mask detection can be used for several applications in real-life scenarios like monitoring people from a distance, real-time biometrics, etc. Mostly criminals cover the mouth area of the face. Occulted face detection problem has been resolved by researchers previously by the shape of the head and shoulder of the person [1]. The situations where many people need to be checked remotely for face mask-wearing add complexity to this task. Moreover, with the spread of the novel coronavirus disease (COVID-2019) necessitated face mask wearing and several other restrictions. With devastating outcomes of COVID-19, rapid spread and lack of appropriate medicine and medical experts, the world health organization (WHO) declared it a pandemic and recommended several precautionary measures that also include the use of face masks [2, 3]. Putting on a mask is the key method to fight against the fatal circumstances of COVID-19. The fact that the use of face masks limits the COVID-19 spread has gained acceptance in the general population. This condition put pressure globally on the general population to maintain social distancing and follow defensive measures to avoid the transmission of this contagious virus. Even though a large part of the population in many countries has been vaccinated, with its evolving variants COVID-19 is still penetrating and spreading. So, the consistent use of face masks is a very significant practice to curb its spread. Moreover, that will help to avoid contamination and keep people safe from contracting the germs of this disease.

Face masks are enforced for people in different countries and entry into public offices is not allowed without wearing a face mask. However, due to a large number of people visiting public offices, public places like airports and train stations, and shopping malls, manual inspection is almost impossible. Recently, research related to the automatic detection and identification of face masks is getting attention. Researchers have started devising automatic detection systems that can help in monitoring and surveillance applications during COVID-19 [4]. The detection process of face masks includes two tasks: identification and classification because it also includes identification of faces and then deciding whether they are wearing masks or not. The first task is an extensively explored research area in the field of com-

*Corresponding author
umersabir1996@gmail.com (M. Umer); s.kamrran@gmail.com (S. Sadiq); ralhebshi@kau.edu.sa (R.M. Alhebshi); Sa.alsubai@psau.edu.sa (S. Alsubai); a.alhejaili@ut.edu.sa (A.A. Hejaili); aaeshmawi@uj.edu.sa (A.A. Eshmawi); mnappi@unisa.it (M. NAPPI); imranashraf@ynu.ac.kr (I. Ashraf)
ORCID(s): 0000-0002-6015-9326 (M. Umer); 0000-0002-2611-3738 (S. Sadiq)

puter vision as many face detection techniques have been developed by the research community [5, 6, 7]. A lot of work has been performed last year for the detection of face masks on different datasets [8, 9]

Two problems with the existing research are the lack of a publicly available real image dataset and the challenge of various face mask colors and wearing styles. Two publicly available datasets have been used for previous research; MAFA [10] and Wider Face [11]. Studies also used the simulated datasets where the face mask is simulated over non-face mask images which makes the models inappropriate for real-world scenarios. A variety of face masks and different positions on the face make the detection more challenging. This study is an effort in a similar direction and aims at improving the detection process of face masks. Toward this problem, we follow an image processing pipeline and make the following key contributions

- A customized image dataset is built for research on face mask detection. The images are real, gathered using a color camera, and separate images are taken for mask and no mask faces. The dataset is manually labelled to provide high annotation accuracy. The dataset is made publicly available for future research.
- For the detection of face masks, various customized off-the-shelf deep learning models are utilized. For this purpose, YOLO v3 and Faster residual convolutional neural network (R-CNN) are leveraged and fine-tuned to obtain a better performance.
- A custom-designed CNN is proposed for face mask detection. Experiments are performed using the collected dataset and performance is evaluated in terms of precision, recall, accuracy, and F1-score. The validation of the proposed CNN and YOLO v3 and Faster R-CNN is carried out with two publicly available datasets including MAFA and MOXA.
- Experiments also involve using logistic regression (LR), extra tree classifier (ETC), random forest (RF), stochastic gradient descent (SGD), and support vector machine (SVM).

The rest of the paper is structured as follows. Section 2 describes several articles related to the current study. Section 3 provides a summary of the collected dataset and the proposed model. In addition, the steps performed on the dataset and the basic introductions of deep learning models are provided. Results are discussed in Section 4 and the paper is concluded in Section 5.

2. Related Work

This section provides an overview of literature related to face mask detection in terms of datasets and techniques used for the detection process. Researchers have targeted various computer vision tasks like face detection, face tracking, face retrieval, and face occlusion detection [1]. Under face occlusion detection, researchers performed various steps like head detection, facial feature detection, mask detection [12], skin color-based detector [13], and face parts detection [14] using PCA [15], SVM [16] and Markov model [17]. The major challenge of these studies was low-resolution images of surveillance tools.

2.1. Face Mask Dataset

Amid the COVID-19 pandemic, people across the world start wearing face masks to reduce the spread of the coronavirus. Before it, face masks were used only to avoid pollution, respiratory disease, seasonal flu and cold, etc. Detection of face masks was not as important as it has become after the COVID-19 outbreak. So, very few datasets are publicly available for research purposes. With the spread of the COVID-19 pandemic, face mask datasets have been created by researchers. Authors in [4] developed three face mask datasets including a masked face detection dataset (MFDD), a simulated masked face recognition dataset (SMFRD), and a real-world masked face recognition dataset (RMFRD). MFDD dataset is a modified form of the dataset by [18]. Most images included in this dataset are taken from the internet and are not suitable for a real-world setting. This dataset has been by researchers for face mask detection. RMFRD includes 5000 images of masked people. This dataset includes the labelling of images with respect to the properly placed face mask on the face. While SMFRD includes simulated images of masked faces where masks are added artificially to the images. They made a subset of their dataset available to the public.

The masked faces dataset named MAFA [10] has been proposed as a face detection dataset. Images for this dataset are also collected from the internet and the dataset contains face masks or other occlusions on the face. This dataset is used by the researchers in identifying face masks by combining them with any other dataset containing images without masks [18]. MOXA dataset is created by [19] and contains 3000 images with two classes 'mask' and 'no mask'. Face mask datasets do not consider masks in a proper position. Datasets available at Kaggle called face mask detection (FMD) [20] consists of images labelled with three classes; 'with mask', 'without a mask', and 'incorrect mask'. The main problem related to this dataset is that most of the faces are of small size which makes the image preprocessing task difficult and the detection of face masks very challenging.

Another dataset [21] available at Kaggle includes 6024 images. The dataset contains images of persons from different backgrounds regarding age, region, and ethnicity. Labelling of the dataset involves 20 classes which include 'with face mask', without face mask', 'incorrectly placed' covered with other accessories like a scarf, hats, face shields, and many more. A subset of this dataset called the medical mask dataset (MMD) is relevant to the face mask detection problems during the COVID-19 pandemic. Contrary to the datasets discussed earlier, this study does not collect datasets from internet sources. We have developed our own

customized dataset named a real image-based labelled face mask dataset (RILFD) by taking pictures of people with and without a face mask and labelling them manually.

2.2. Face Mask Detection Techniques

Numerous solutions for face mask detection from face images have been presented in the literature during the last year. Researcher in [22] uses MobileNet for face mask detection. A similar task has been performed by the researcher in [23]. The model [24] based on context attention and feature pyramid has been proposed by researchers as a simple and one-stage model for face detection. The proposed model is tested on datasets like MAFA and wider faces containing with mask and without mask faces. The model was not very efficient in distinguishing faces with masks from faces with masks. In [25], authors classified masked faces from those without masked faces by applying the image resolution enhancement approach to process low-resolution images. The proposed model SRCNet is trained and tested on a dataset containing 671 without masks and 3030 with mask images. Validation of this approach is not possible as the dataset is not publicly available.

A few recent works on face mask detection adopt both customized, as well as, pre-trained deep learning models [26, 27, 28, 29]. For example, [26] uses transfer learning by fine-tuning the ResNet50 model for face mask detection in partially occluded situations. The performance of the model is further improved by using the bounding box affine transformation which helps in localizing the mask area. The study reports a 98.2% accuracy using the oversampled MAFA dataset. Similarly, a single-stage face mask detector called RetinaFaceMask is presented in [27] which utilizes the context attention module for obtaining discriminated features. For experiments, the MAFA dataset is reannotated to formulate MAFA-FMD that contains no mask, and correct and incorrect mask-wearing classes. Transfer learning and context attention module are used to increase the mask detection accuracy of the proposed approach. Experimental results demonstrate a mean average precision (mAP) of 94.8% on the AIZOO dataset while the mAP using MAFA-FMD is only 68.3%. Along the same direction, the use of transfer learning is adopted in [28] where an ensemble learning model is presented that combines a deep neural network, MobileNetV2, and single-shot multi-box detector for face mask detection. MobileNetV2 is a lightweight architecture and shows robust results for real-time operations. Experiments performed using a self-collected dataset show a 92.64% accuracy score. A dual-stage deep learning system is introduced in [29] that can discriminate between masked and unmasked people from the CCTV cameras. The proposed system follows face detection, region of interest extraction, batching, and face mask classification steps to make the final decision. For binary classification, the study shows promising results using NASNetMobile, DenseNet121, and MobileNetV2 models.

Nair et al. [30] detected face masks from videos. The authors discussed that the identification of face masks from videos takes a longer time compared to face detection. The proposed framework is based on Viola Jone's algorithm for face and eye detection. At first, the model detects eyes and then faces; if a face is found it means that the face is without a mask. If after detecting the eyes face is not found, it means the person is wearing a mask. Bu et al. [31] utilized a CNN-based model for face detection. The authors applied three neural network models. First CNN is a shallow network with 5 layers and predicts face masks for each window with non-maximum suppression. Second CNN is a deep model with 7 layers that perform detection after resizing window size and setting a specific threshold value. The third CNN consists of 7 layers and performs prediction by resizing the input window.

Some commercial solutions are also available in the market for face mask detection [32] that provide monitoring of crowds by video streams provided by the camera. Another open-source tool for mask detection is Baidu Paddle Hub, a face mask detector [33].

3. Materials and Methods

This section discusses the proposed framework, dataset collection and steps followed for the proposed framework. The flow of the proposed framework is shown in Figure 1. After the data collection containing mask and no mask classes, images are preprocessed using four stages of image processing. As variations in complexion and lighting effects make face mask detection a more challenging task, image processing steps are used to normalize the input images. After that, a customized deep learning model is employed to detect if a face mask is present or not. Additionally, two off-the-shelf fine-tuned deep learning models are also used.

3.1. Real Image-based Labeled Face Mask Dataset

A manually labelled and structured face mask dataset is a useful and serviceable addition to the research community. Although in this regard, some datasets are now available for experiments, these datasets have different limitations. For example, some datasets have small-sized images while others contain internet-based images. Similarly, simulated face masks have been placed on faces in the datasets. This study presents a labelled dataset called

Stage 1: It involves applying filters on the original data. Filter size ([0,-1,0],[-1,6,-1],[0,-1,0]) has bee applied on images for this purpose as shown in Figure 2c.

Stage 2: At the third stage, images are transformed from blue, green, and red (BGT) to luma components, red and blue projections (YUV). This step keeps the Y a luma component at full resolution and reduces the U and V resolutions. Because luminance is more significant as compared to colo, the reduction of U and V also helps in reducing the complexity of the training model. Figure 2d presents the conversion of BGR to YUV.

Stage 3: At the last stage, images are normalized by converting back to the BGR. This step smoothens the images and histogram normalization is applied as shown in Figure 2a.

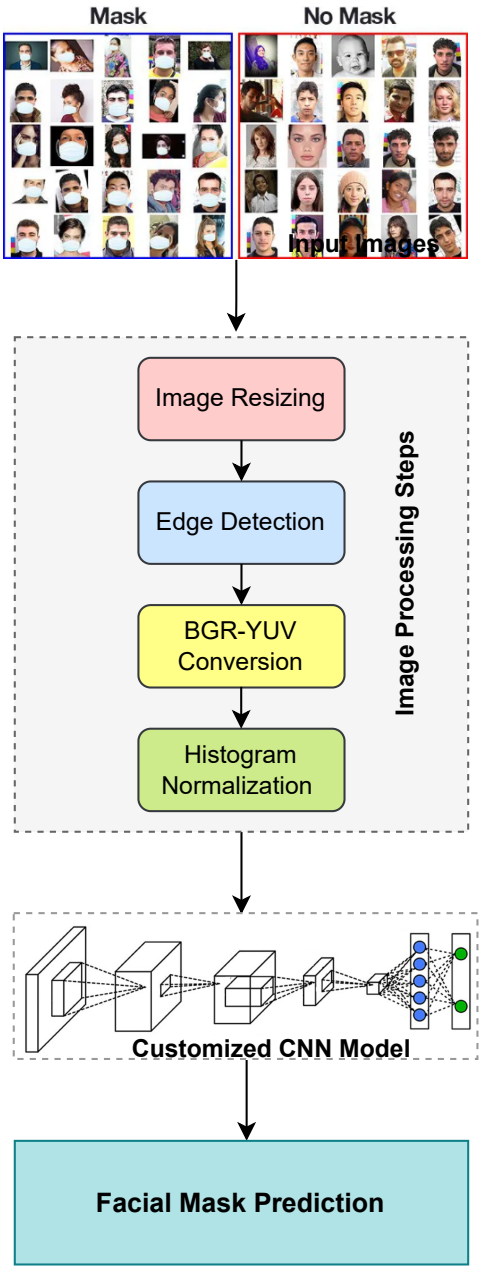

Figure 1: Workflow diagram of the proposed framework for face mask detection.

Stage 4: This stage involves contrast conversion and image resizing. It consists of reducing the size of the input image. The dimensions of images in the dataset are 11903x13096 pixels. At this stage, the size of the image is reduced to $120 \times 120 \times 3$ as shown in Figure 2b. It helps to reduce the computational time of the detection.

3.2. Proposed Model for Face Mask Detection

Face mask detection models have made notable progress over the last year, mainly during the COVID-19 pandemic [34]. Face detection has shown remarkable improvements with the development of deep learning models like convolutional neural networks (CNN). Deep CNN models have been employed by researchers for a variety of tasks like numeric data analysis [35]and text data analysis [36]. Predominantly, face detection techniques are based on CNN [37] and show excellent performance by efficiently dealing with many challenges like pose, low-resolution images, illumination, and other types of noises. Although face mask detection has been studied for different occlusions it is still a less explored research area. This paper utilizes a CNN-based deep learning model extensively used for object detection tasks.

CNN performs a significant role in computer vision tasks, due to its exceptional feature extraction ability and low computational cost [56]. Binary classification of the images is one of its extremely helpful applications. CNN convolves with the feature maps or images to obtain high-level features by utilizing various kernels. Though, the main question that always stays is how to construct a better CNN structural design. Another most employed and acknowledged CNN is the inception model [57]. A residual network is a deeper neural network model that can learn the mapping from the prior layer. For object detection using mobile devices, a low computational cost-based model, MobileNet was proposed [58]. The depth and channel-wise convolution reduce the computational cost in the model.

CNN consists of convolutional layers, pooling layers, and fully connected layers. Each layer performs different tasks. In the CNN model, Kernel or filter is a number matrix that convolves over the image and transforms it according to the values of the filter. After each step of convolution, image size is reduced and this process can be performed in a limited time. Let $I(x, y)$ be the input image, $f(x, y)$ be the kernel used for convolution, the output $y(i, j)$ can be defined as

$$y(i,j) = (I,f)(x,y) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} I(x-u, y-v) f(u,v) \quad (1)$$

As the kernel moves on the image, its impact on the centre is smaller as compared to the other parts of the image. Therefore image can be padded with a border. Padding should fulfil the following requirements

$$P = \frac{(f-1)}{2} \quad (2)$$

where f represents filter dimensions and p represents padding.

Sigmoid is used as a non-linear activation function as shown in Equation 3

$$sigmoid(x) = \frac{1}{1+e^{-x}} \quad (3)$$

the pooling layer subsamples the convolutional output and it can be a max or average according to the requirement. Pooling can be computed as follows

$$X_{ij}^{[l]} = \frac{1}{MN}\sum_{m}^{M}\sum_{n}^{N} X_{iM+m,jN+n}^{[l-1]} \quad (4)$$

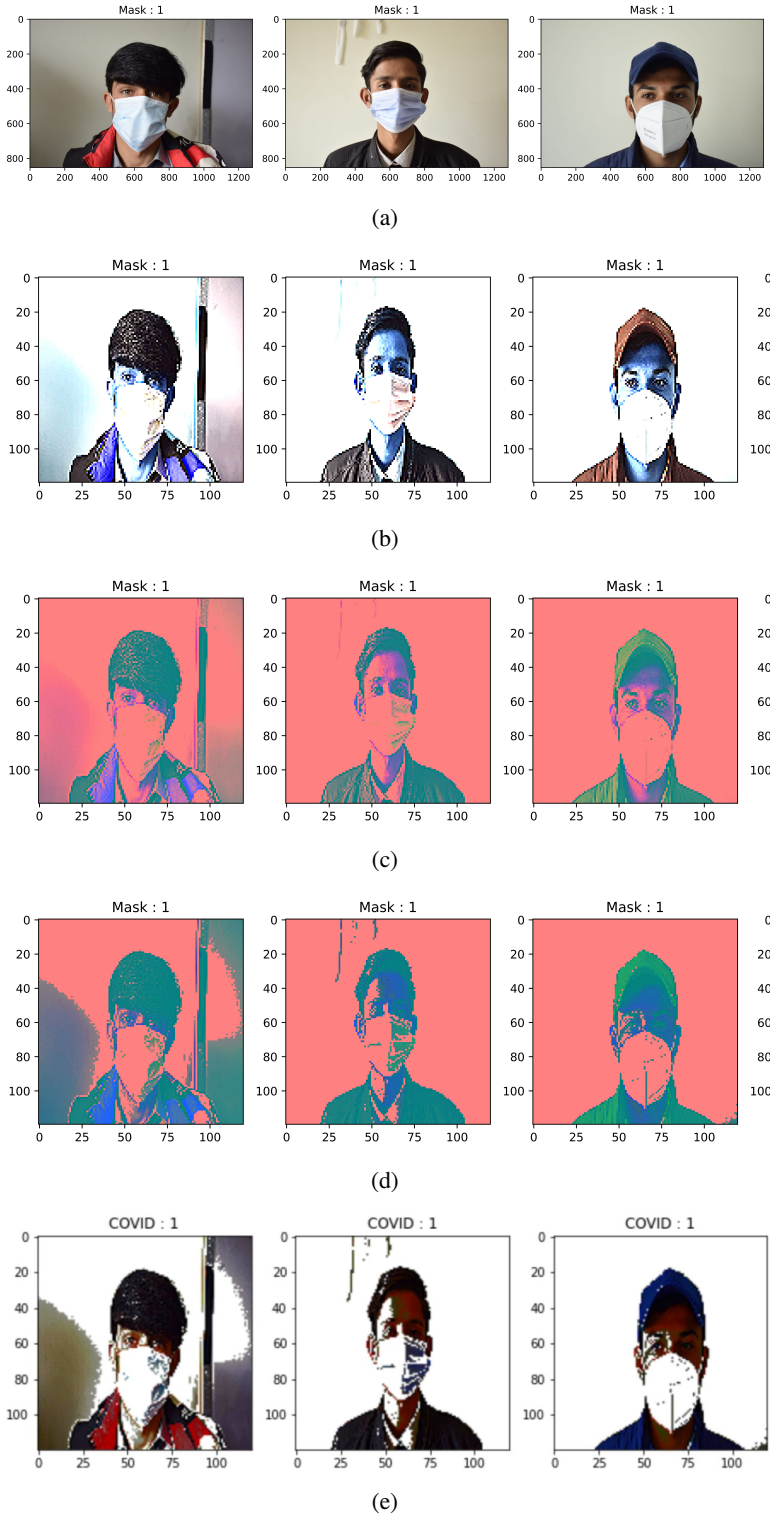

Figure 2: Image preprocessing steps followed in the experiments, **(a)** Original Image, **(b)** Kernel is applied for edge detection, **(c)** BGR image is converted to YUV to get Y_0, and **(d)** Histogram Equalization of YUV image, and **(e)** YUV image is converted to BGR image.

where i, and j present the output map positions, while M and N show the sample sizes.

For optimal performance, several hyperparameters of CNN are tuned. The details of the hyperparameters used in the customized CNN are presented in Table 2. Machine learning and deep learning models used for comparison are discussed in Table 1.

Table 1
Machine Learning and Deep Learning Models used for comparison.

Ref.	Model	Description
L. Breiman (2001) [38]	RF	RF is a supervised machine learning model that is based on trees and makes a final decision from the results of several decision trees. The model is trained on a randomly selected dataset by bootstrapping [39]. The root node of decision trees is generated on measures like information gain and the Gini index. RF has shown remarkable results on classification tasks like EEG data [40], and cardiotocograph data classification [41].
Nick and Kathleen (2007) [42]	LR	LR is based on a statistical approach and maps input features into target variables using the Sigmoid function. This function contains an S-shaped curve that specifies the target value [43]. LR has been used for different tasks like comment classification [44].
Sharaff and Gupta (2019) [45]	ETC	The ETC is also based on multiple trees similar to RF. It makes trees from the original data, unlike RF which uses bootstrap data for training [46]. The root node is selected using Gini Index. It works on numeric input and selects an optimal value for a cut point which reduces the complexity of the model. ETC shows better results for complex problems and produces multilinear approximation.
Gardner (2014) [47]	SGD	SGD is an optimization machine learning algorithm and is mostly used to determine best-fit parameters between actual and predicted results for the model. SGD is a simple but powerful model and has been widely used in classification problems [48]. It is faster than gradient descent and shows speedy results.
Scholkopf et al. (1996) [49]	SVM	SVM is a supervised machine learning algorithm that uses a support vector to make hyperplanes [50]. It works efficiently in solving multi-dimensional problems. It has been extensively used in classification, outlier detection, and regression. It shows improved results in solving many problems like image steganography detection [51] and plant disease classification [52].
Redmon and Farhadi (2018) [53]	YOLO V3	YOLO (You Only Look Once) is a deep neural network model with the capability of object detection in real time. It is a very fast and accurate algorithm and it has been used in the detection of traffic signals, parking, animals, and people. It predicts the boundary (coordinates) using fully-connected layers next to convolutional layers. The algorithm works by dividing images into N number of parts (grids) having $S \times S$ dimensions. Identification and localization of the specific objects have been performed with the help of each grid of the image. This method simplifies the problem and easy to learn the model. YOLO v3, an improved version of YOLO [53] is a deep neural model and can identify objects from images or videos using learned features in real-time objection detection. It works on three scales by reducing the dimensions through downsampling of input images by 32 then 16 and then 8. As it is a variant of Darknet [54] (having 53 layers), it is staked with 53 more layers to perform object detection and has a total of 106 layers. The first detection of an object takes place at 82 layers after downsampling the images till the 81st layer has a stride of 32.
Ren et al. (2015) [55]	Faster R-CNN	Faster R-CNN is an improved form of Fast R-CNN that applies a regional proposal network (RPN) as a region proposal along with CNN as a detector model. RPN extracts convolutional features and passes them to the detection model. It is a fully convolutional model and computes object bounds at each place and position. RPN generates high-quality region proposals that are used by R-CNN. They both act as a single network having two modules. The first is a deep network that suggests regions and the second is the fast R-CNN that uses suggested regions for detection.

3.3. Evaluation Metrics

To evaluate the effectiveness of the deep learning models, this study utilizes accuracy, precision, recall, and F1-score, These metrics are computed on four basic terms, i.e., true positive (TP), false positive (FP), true negative (TN), and false negative (FN). TP refers to the person wearing a face mask and the model predicts it as 'with mask' while FP is those persons who do not wear the mask but are predicted as 'with mask'. Similarly, TN refers to those persons who do not wear a mask and are predicted as 'without mask' and FN are the ones wearing a mask but predicted as 'without mask'. Table 3 presents the formulae for the evaluation metrics.

4. Results and Discussions

For face mask detection, deep learning models used in this study are customized CNN, YOLO V3, and Faster R-CNN. In the first step, the face mask dataset RILFD has been created by taking a picture from the camera containing the 'with mask' or 'without mask' classes. For efficient training of the deep learning model, four-stage image processing has been performed. Results are compared with and without applying image processing. Experiments have been performed using two publicly available datasets called MAFA and MFDD. The data split ratio for experiments is 0.7 to 0.3 for training and testing, respectively. Training is done us-

Table 2
Detail of the hyperparameters used in the customized CNN model.

Name	Description
Convolution	Filters=$(3\times3, @16)$, Strides=(1×1)
Convolution	Filters=$(3\times3, @128)$,Strides=(1×1)
Max pooling	Pool_ size=(2×2),Strides=(2×2)
Convolution	Filters=$(2\times2, @256)$,Strides=(1×1)
Average pooling	Pool_ size=(3×3), Strides=(1×1)
Layer	Flatten()
Fully connected	Dense (120 neurons)
Fully connected	Dense (60 neurons)
Fully connected	Dense (10 neurons)
Sigmoid	Sigmoid (2-class)

Table 3
Performance Measures used for Evaluation in this study

Evaluation Metric	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$2\times\frac{precision\times recall}{precision+recall}$

Table 4
Face mask detection using deep learning models on RILFD dataset without image preprocessing steps.

Models	Accuracy	Precision	Recall	F1-score
YOLO v3	72.84	78.12	81.46	79.79
Faster R-CNN	83.17	84.01	84.41	84.21
Customized CNN	88.52	89.77	91.34	90.55

ing a Tesla K80 Tensor Processing Unit (TPU) available at Google Colab. It provides 180 TFlops, 16 GB of random access memory (RAM), and 128 GB of disk space. Training took 1.5 hours to run 12 epochs on the dataset for two classes.

4.1. Results Without Image Preprocessing Steps

Initial experiments are carried out without using the fours steps-based image processing pipeline to analyze the performance of all the models. In this regard, the images are fed into the models without performing any of the steps like color transformation, applying the filter, etc. and results are given in Table 4. It can be seen that the performance of the models is not very good. Despite that, the performance of the customized CNN is better than other models with an 88.52% accuracy. Similarly, its precision, recall, and F1-scores are much better than YOLO v3 and Faster R-CNN. Analyzing the dataset, it is revealed that it is difficult to differentiate face masks from other stuff. For example, skin color and facial hair may be similar to the color of the mask which makes mask detection very difficult. Therefore, four stages of image preprocessing techniques have been applied before training the classifiers.

Table 5
Face mask detection using deep learning models on RILFD dataset with image preprocessing techniques.

Models	Accuracy	Precision	Recall	F1-score
YOLO v3	89.44	87.67	91.23	89.45
Faster R-CNN	92.65	91.82	94.24	92.23
Customized CNN	97.25	96.20	97.34	96.77

4.2. Results Using Image Preprocessing

The second set of experiments involves using the image preprocessing steps as described previously. Experiment results are given in Table 5 indicating better performance than without the use of image preprocessing. It can be observed that all deep learning models show significantly better performance with image preprocessing. The Faster R-CNN shows better results with 92.65% accuracy, 91.82% precision, 94.24% recall, and 92.23% F1-score as compared to the YOLO V3 algorithm. Overall customized CNN achieved the highest performance with 97.25% accuracy, 96.20% precision, 97.34% recall, and 96.77% F1-score.

YOLO V3 algorithm is a simple and single-shot algorithm with less inference time. But in comparison, Faster R-CNN has achieved better results. There is a trade-off between performance efficiency and speed. Figure 3 shows the accuracy curve for the best-performing customized CNN model. It shows that the proposed model smoothly improves the training and testing accuracy which proves the robustness of the proposed approach. Hence, face mask detection using CNN is superior to other off-the-shelf deep learning models used in the experiment as shown in figure 4.

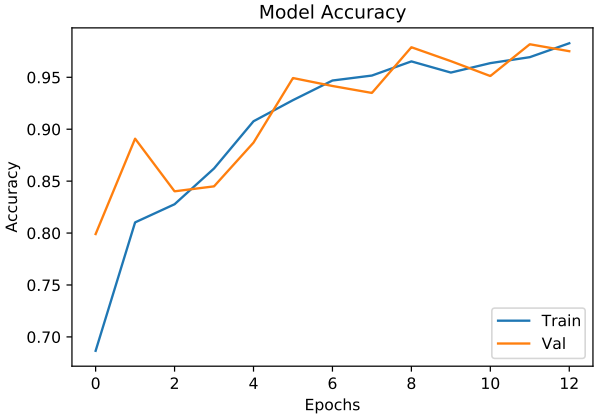

Figure 3: Accuracy Curve of the customized CNN model.

4.3. Results of K-Fold Cross-Validation

For validating the performance of deep learning models and analyzing the overfitting, this study applied 10-fold cross-validation. The result of the 10-fold cross-validation results using a customized CNN model with image preprocessing steps is presented in Table 6. Results prove that the performance of the customized CNN is better with 96.13%

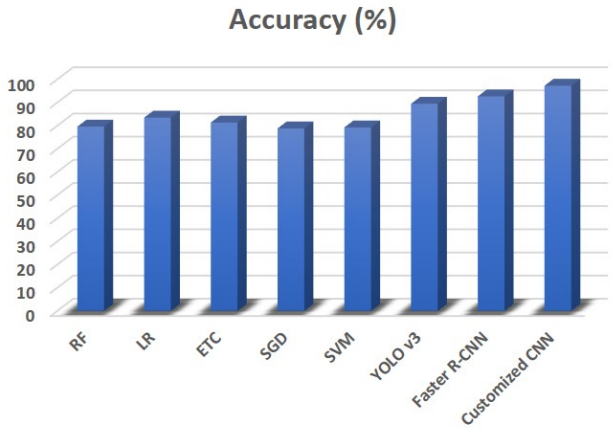

Figure 4: Accuracy comparison of customized CNN with other models.

Table 6
Ten-fold cross-validation results using customized CNN model with image preprocessing techniques.

Fold Number	Accuracy	Precision	Recall	F1-score
1st-Fold	94.50	95.60	95.10	95.10
2nd-Fold	96.20	95.70	95.20	95.60
3rd-Fold	96.10	94.30	95.30	95.40
4th-Fold	94.80	94.70	95.90	95.50
5th-Fold	95.40	97.10	96.80	96.30
6th-Fold	97.60	97.60	96.70	96.20
7th-Fold	94.40	94.70	96.60	97.10
8th-Fold	94.40	95.40	98.50	97.70
9th-Fold	98.20	94.40	94.80	97.80
10th-Fold	99.70	98.50	98.70	98.90
Average	**96.13**	**95.80**	**96.72**	**96.56**

accuracy, 95.80% precision, 96.72% recall, and 96.56% F1-score. These values are averaged ten folds and show very little variance.

4.4. Performance of Machine Learning Models

In addition to off-the-shelf deep learning models and the customized CNN model, this study also implements several machine learning models including RF, LR, ETC, SGD, and SVM. These models have been selected in view of their reported performance for similar tasks. The performance of the customized CNN model is compared with these machine learning models.

These models are tested on the RILFD dataset containing face images with face masks and without face masks. Table 7 presents the results of machine learning models indicating that LR achieves the highest accuracy score of 83.45% with image preprocessing steps. On average, the performance of the machine learning models is much better when used with image preprocessing steps. Despite the best performance of LR among machine learning models with 83.45% accuracy, its performance is still inferior to the customized CNN models that achieve 97.25% accuracy for face mask detection on the collected dataset.

Table 7
Comparative analysis of machine learning models for face mask detection.

Model	Accuracy (%)	
	With preprocessing	Without preprocessing
SGD	78.80	73.47
RF	79.60	71.77
ETC	81.35	78.78
SVM	79.24	74.11
LR	83.45	77.34
Customized CNN	97.25	88.52

Furthermore, if we compare the suggested model's performance with that of earlier research in the literature, it can be shown that some studies have been conducted to identify the face from masked face images or videos, and some studies are conducted to identify masks. Due to the COVID-19 epidemic, several public health organizations throughout the world have mandated the wearing of face masks to prevent the spread of the virus. Checking masks on faces is of great importance to reduce the spread of contagious diseases. Authors in [59] applied traditional machine learning and deep learning models on masked faces to perform recognition and mainly focused on the biometric system to improve the performance of the models. However, we have used four stages of image processing to improve the performance of the models.

4.5. Validating Performance of Deep Learning Models With MAFA and MOXA Datasets

To further corroborate the performance of the proposed customized CNN model and other deep learning models, additional experiments have been performed using three well-known datasets including MAFA, MOXA, and RMFRD. These datasets have been selected due to their frequent use for face mask detection studies. These experiments also intend to prove the robustness and generalizability of the proposed customized CNN model. Experimental results prove the efficiency of the proposed approach as the customized CNN achieves the best results on these datasets as compared to YOLO v3 and Faster R-CNN. It obtains a 95.74% accuracy score and 94.29% recall on the MAFA dataset and 94.37% accuracy score and 95.28% recall using the MOXA dataset. Furthermore, the performance of customized CNN is much better on RMFRD dataset where it obtains a 99.63% accuracy and a 99.69% recall which shows its superior performance.

4.6. Performance Comparison with Existing Studies

The performance of the customized CNN is also compared with existing state-of-the-art studies [26, 27]. For comparison, only those studies are selected that use a publicly available dataset containing real images, as often the accuracy with simulated face masks is high as reported in [22]. Table 9 shows the comparison with the selected studies indicating that the proposed approach outperforms existing models in terms of accuracy and precision. A few studies re-

Table 8
Performance comparison of models using MAFA, MOXA and RMFRD datasets.

Model	Dataset	Accuracy	Recall
YOLO v3	MAFA	90.47	88.24
Faster R-CNN	MAFA	91.65	90.87
Customized CNN	MAFA	95.74	94.29
YOLO v3	MOXA	86.34	88.85
Faster R-CNN	MOXA	88.34	87.31
Customized CNN	MOXA	94.37	95.28
YOLO v3	RMFRD	98.55	98.74
Faster R-CNN	RMFRD	98.97	99.31
Customized CNN	RMFRD	99.63	99.69

Table 9
Comparison with existing state-of-the-art studies.

Reference	Dataset	Accuracy
Sethi et al. (2021) [26]	MAFA+oversampling	98.20%
Current study	MAFA+oversampling	99.10%
Fan et al. (2021) [27]	MAFA	68.30%
Current study	MAFA	94.11%

port high accuracy for face mask detection like [29] reports an accuracy of 99.40%. However, this approach follows a face detection, region of interest, and face mask detection procedure and the reported accuracy does not consider the error in face detection, so accuracy is overrated. Also, the study points out that partially occluded faces can not be detected which further highlights its limitations.

5. Conclusions

Wearing a face mask is one of the basic restrictions imposed by governments to curb the spread of COVID-19. However, ensuring that people wear face masks becomes a laborious and time-consuming task in public places. This study presents a CNN-based approach to provide a solution for automatic face mask detection. For this purpose, RILFD an image dataset containing high-definition images is collected from a real-world environment and annotated manually. A four-stage image preprocessing is utilized to obtain high performance for face mask detection. Experiments are performed using off-the-shelf YOLO v3 and Faster R-CNN and machine learning models including LR, RF, SGD, ETC, and SVM. Results prove that the use of image preprocessing yields better results, both for machine learning and deep learning models. The highest accuracy of 97.25% is obtained using the customized CNN which is substantially better than machine learning and YOLO v3 and Faster R-CNN models. The proposed CNN model is simple and less complex and has shown less training time when trained on the RILFD dataset. However, YOLO v3 and Faster RCNN have shown poor performance and took more time to predict face masks. Results from cross-validation and experiments on two publicly available datasets MAFA and MOXA confirm the superior performance of the proposed customized CNN. This study performs experiments on a self-collected dataset collected in a real-world environment, however, the influence of dark and low-light environments is not considered which can affect the robustness of the proposed approach.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The datasets generated during and/or analyzed during the current study can be downloaded from the following link.
https://github.com/MUmerSabir/MDPIDiagnostic

CRediT authorship contribution statement

Muhammad Umer: Writing - Original draft preparation, Methodology, Software. **Saima Sadiq:** Writing - Original draft preparation, Conceptualization of this study. **Reemah M. Alhebshi:** Final manuscript review, Project Supervision. **Shtwai Alsubai:** Final manuscript review. **Abdullah Al Hejaili:** Visuallization, Software. **Ala' Abdulmajid Eshmawi:** Project Supervision, Methodology. **Michele NAPPI:** Funding Acquisition. **Imran Ashraf:** Final manuscript review.

Abbreviations

The following abbreviations are used in this manuscript:

Acronyms	Definition
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
COVID-2019	CoronaVirus disease 2019
ETC	Extra Tree Classifier
FMD	Face mask detection
LR	Logistic Regression
MAFA	Masked Face
MFDD	Masked Face Detection Dataset
MMD	Medical Mask Dataset
R-CNN	Residual Convolutional Neural Network
RF	Random Forest
RILFD	Real Image-based Labelled Face Mask Dataset
RMFRD	Real-world Masked Face Recognition Dataset
SGD	Stochastic Gradient Descent
SMFRD	Simulated Masked Face Recognition
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
WHO	World Health Organization
YOLO	You Only Look Once

References

[1] Tao Zhang, Jingjing Li, Wenjing Jia, Jun Sun, and Huihua Yang. Fast and robust occluded face detection in atm surveillance. *Pattern Recognition Letters*, 107:33–40, 2018.

[2] Nancy HL Leung, Daniel KW Chu, Eunice YC Shiu, Kwok-Hung Chan, James J McDevitt, Benien JP Hau, Hui-Ling Yen, Yuguo Li,

Dennis KM Ip, JS Peiris, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature medicine*, 26(5):676–680, 2020.

[3] Shuo Feng, Chen Shen, Nan Xia, Wei Song, Mengzhen Fan, and Benjamin J Cowling. Rational use of face masks in the covid-19 pandemic. *The Lancet Respiratory Medicine*, 8(5):434–436, 2020.

[4] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al. Masked face recognition dataset and application, 2020.

[5] Stefanos Zafeiriou, Cha Zhang, and Zhengyou Zhang. A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding*, 138:1–24, 2015.

[6] Xin Li, Shenqi Lai, and Xueming Qian. Dbcface: Towards pure convolutional neural network face detection. *IEEE Transactions on Circuits and Systems for Video Technology*, PP:1–1, 05 2021. doi: 10.1109/TCSVT.2021.3082635.

[7] Ashu Kumar, Amandeep Kaur, and Munish Kumar. Face detection techniques: a review. *Artificial Intelligence Review*, 52(2):927–948, 2019.

[8] Y Wang. Which mask are you wearing. *Face Mask Type Detection with TensorFlow and Raspberry Pi. Medium*, 1, 2020.

[9] G Jignesh Chowdary, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal. Face mask detection using transfer learning of inceptionv3. In *International Conference on Big Data Analytics*, pages 81–90. Springer, 2020.

[10] Shiming Ge, Jia Li, Qiting Ye, and Zhao Luo. Detecting masked faces in the wild with lle-cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2682–2690, 2017.

[11] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5525–5533, 2016.

[12] Che-Yen Wen, Shih-Hsuan Chiu, Yi-Ren Tseng, and Chuan-Pin Lu. The mask detection technology for occluded face analysis in the surveillance system. *Journal of Forensic Science*, 50(3):1–9, 2005.

[13] Praveen Kakumanu, Sokratis Makrogiannis, and Nikolaos Bourbakis. A survey of skin-color modeling and detection methods. *Pattern recognition*, 40(3):1106–1122, 2007.

[14] Chung J Kuo, Tsang-Gang Lin, Ruey-Song Huang, and Souheil F Odeh. Facial model estimation from stereo/mono image sequence. *IEEE transactions on multimedia*, 5(1):8–23, 2003.

[15] Zhaoqing Pan, Peng Jin, Jianjun Lei, Yun Zhang, Xingming Sun, and Sam Kwong. Fast reference frame selection based on content similarity for low complexity hevc encoder. *Journal of Visual Communication and Image Representation*, 40:516–524, 2016.

[16] Daniel MM Da Costa, Sarajane M Peres, Clodoaldo AM Lima, and Pollyana Mustaro. Face recognition using support vector machine and multiscale directional image representation methods: A comparative study. In *2015 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2015.

[17] Huy Tho Ho and Rama Chellappa. Pose-invariant face recognition using markov random fields. *IEEE transactions on image processing*, 22(4):1573–1584, 2012.

[18] Daniell Chiang. Detect faces and determine whether people are wearing mask. *Face Mask Detection*, 15, 2020.

[19] Biparnak Roy, Subhadip Nandy, Debojit Ghosh, Debarghya Dutta, Pritam Biswas, and Tamodip Das. Moxa: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks. *Transactions of the Indian National Academy of Engineering*, 5(3):509–518, 2020.

[20] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N Taha, and Nour Eldeen M Khalifa. Fighting against covid-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection. *Sustainable cities and society*, 65:102600, 2021.

[21] Wobot Intelligence. Face mask detection dataset, 2021.

[22] Isunuri B Venkateswarlu, Jagadeesh Kakarla, and Shree Prakash. Face mask detection using mobilenet and global pooling block. In *2020 IEEE 4th Conference on Information & Communication Technology (CICT)*, pages 1–5. IEEE, 2020.

[23] Prateek Khandelwal, Anuj Khandelwal, Snigdha Agarwal, Deep Thomas, Naveen Xavier, and Arun Raghuraman. Using computer vision to enhance safety of workforce in manufacturing in a post covid world, 2020.

[24] Mingjie Jiang, Xinqi Fan, and Hong Yan. Retinamask: A face mask detector, 2020.

[25] Bosheng Qin and Dongxiao Li. Identifying facemask-wearing condition using image super-resolution with classification network to prevent covid-19. *Sensors*, 20(18):5236, 2020.

[26] Shilpa Sethi, Mamta Kathuria, and Trilok Kaushik. Face mask detection using deep learning: An approach to reduce risk of coronavirus spread. *Journal of biomedical informatics*, 120:103848, 2021.

[27] Xinqi Fan and Mingjie Jiang. Retinafacemask: A single stage face mask detector for assisting control of the covid-19 pandemic. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 832–837. IEEE, 2021.

[28] Preeti Nagrath, Rachna Jain, Agam Madan, Rohan Arora, Piyush Kataria, and Jude Hemanth. Ssdmnv2: A real time dnn-based face mask detection system using single shot multibox detector and mobilenetv2. *Sustainable cities and society*, 66:102692, 2021.

[29] Amit Chavda, Jason Dsouza, Sumeet Badgujar, and Ankit Damani. Multi-stage cnn architecture for face mask detection. In *2021 6th International Conference for Convergence in Technology (I2CT)*, pages 1–8. IEEE, 2021.

[30] Aishwarya Radhakrishnan Nair and Amol D Potgantwar. Masked face detection using the viola jones algorithm: A progressive approach for less time consumption. *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, 6(4):4–14, 2018.

[31] Wei Bu, Jiangjian Xiao, Chuanhong Zhou, Minmin Yang, and Chengbin Peng. A cascade framework for masked face detection. In *2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, pages 458–462. IEEE, 2017.

[32] Ziwei Song, Kristie Nguyen, Tien Nguyen, Catherine Cho, and Jerry Gao. Camera-based security check for face mask detection using deep learning. In *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 96–106. IEEE, 2021.

[33] Xu Tang, Daniel K Du, Zeqiang He, and Jingtuo Liu. Pyramidbox: A context-assisted single shot face detector. In *Proceedings of the European conference on computer vision (ECCV)*, pages 797–813, 2018.

[34] Sohaib Asif, Yi Wenhui, Yi Tao, Si Jinhai, and Kamran Amjad. Real time face mask detection system using transfer learning with machine learning method in the era of covid-19 pandemic. In *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 70–75. IEEE, 2021.

[35] Mubariz Manzoor, Muhammad Umer, Saima Sadiq, Abid Ishaq, Saleem Ullah, Hamza Ahmad Madni, and Carmen Bisogni. Rfcnn: traffic accident severity prediction based on decision level fusion of machine and deep learning model. *IEEE Access*, 9:128359–128371, 2021.

[36] Musarat Karim, Malik Muhammad Saad Missen, Muhammad Umer, Saima Sadiq, Abdullah Mohamed, and Imran Ashraf. Citation context analysis using combined feature embedding and deep convolutional neural network model. *Applied Sciences*, 12(6):3203, 2022.

[37] Puranjay Mohan, Aditya Jyoti Paul, and Abhay Chirania. A tiny cnn architecture for medical face mask detection for resource-constrained endpoints. In *Innovations in Electrical and Electronic Engineering*, pages 657–670. Springer, 2021.

[38] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[39] Mostafa El Habib Daho and Mohammed Amine Chikh. Combining bootstrapping samples, random subspaces and random forests to build classifiers. *Journal of Medical Imaging and Health Informatics*, 5(3): 539–544, 2015.

[40] Damodar Reddy Edla, Kunal Mangalorekar, Gauri Dhavalikar, and Shubham Dodia. Classification of eeg data for human mental state

analysis using random forest classifier. *Procedia computer science*, 132:1523–1532, 2018.

[41] MM Imran Molla, Julakha Jahan Jui, Bifta Sama Bari, Mamunur Rashid, and Md Jahid Hasan. Cardiotocogram data classification using random forest based machine learning algorithm. In *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019*, pages 357–369. Springer, 2021.

[42] Todd G Nick and Kathleen M Campbell. Logistic regression. *Topics in biostatistics*, pages 273–301, 2007.

[43] Ernest Yeboah Boateng and Daniel A Abaye. A review of the logistic regression model with emphasis on medical research. *Journal of data analysis and information processing*, 7(4):190–207, 2019.

[44] Mujahed A Saif, Alexander N Medvedev, Maxim A Medvedev, and Todorka Atanasova. Classification of online toxic comments using the logistic regression and neural networks models. In *AIP conference proceedings*, volume 2048, page 060011. AIP Publishing LLC, 2018.

[45] Aakanksha Sharaff and Harshil Gupta. Extra-tree classifier with metaheuristics approach for email classification. In *Advances in computer communication and computational sciences*, pages 189–197. Springer, 2019.

[46] Nur Heri Cahyana, Yuli Fauziah, and Agus Sasmito Aribowo. The comparison of tree-based ensemble machine learning for classifying public datasets. In *RSF Conference Series: Engineering and Technology*, volume 1, pages 407–413, 2021.

[47] William A Gardner. Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal processing*, 6(2):113–133, 1984.

[48] EM Dogo, OJ Afolabi, NI Nwulu, B Twala, and CO Aigbavboa. A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In *2018 international conference on computational techniques, electronics and mechanical systems (CTEMS)*, pages 92–99. IEEE, 2018.

[49] Bernhard Schölkopf, Chris Burges, and Vladimir Vapnik. Incorporating invariances in support vector learning machines. In *International Conference on Artificial Neural Networks*, pages 47–52. Springer, 1996.

[50] Madiha Khalid, Imran Ashraf, Arif Mehmood, Saleem Ullah, Maqsood Ahmad, and Gyu Sang Choi. Gbsvm: sentiment classification from unstructured reviews using ensemble classifier. *Applied Sciences*, 10(8):2788, 2020.

[51] Wenyuan Liu and Jian Wang. Research on image steganography information detection based on support vector machine. In *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pages 631–635. IEEE, 2021.

[52] KR Aravind, P Raja, KV Mukesh, R Aniirudh, R Ashiwin, and Cezary Szczepanski. Disease classification in maize crop using bag of features and multiclass support vector machine. In *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, pages 1191–1196. IEEE, 2018.

[53] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement, 2018.

[54] Mehul Mahrishi, Sudha Morwal, Abdul Wahab Muzaffar, Surbhi Bhatia, Pankaj Dadheech, and Mohammad Khalid Imam Rahmani. Video index point detection and extraction framework using custom yolov4 darknet object detection model. *IEEE Access*, 9:143378–143391, 2021.

[55] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[56] Rahul Chauhan, Kamal Kumar Ghanshala, and RC Joshi. Convolutional neural network (cnn) for image detection and recognition. In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pages 278–282. IEEE, 2018.

[57] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[58] Wei Wang, Yutao Li, Ting Zou, Xin Wang, Jieyu You, and Yanhong Luo. A novel image classification approach via dense-mobilenet models. *Mobile Information Systems*, 2020, 2020.

[59] Lucia Cimmino, Michele Nappi, Fabio Narducci, and Chiara Pero. M2fred: Mobile masked face recognition through periocular dynamics analysis. *IEEE Access*, 10:94388–94402, 2022.