Date of publication MAR-29, 2023, date of current version NOV-11, 2022
www.computingonline.net/ computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.22.1.2880

An Automated Face-mask Detection System using YOLOv5 for Preventing Spread of COVID-19

MD ASIFUZZAMAN JISHAN[1], MD SHAHABUB ALAM[1], IMRAN RASHID MAZUMDER[1], KHAN RAQIB MAHMUD[2], ABUL KALAM AL AZAD[2]

[1]Faculty of Statistics, Technische Universität Dortmund, Vogelpothsweg 87, Dortmund, 44227, Germany (e-mail: md-asifuzzaman.jishan@tu-dortmund.de, md-shahabub.alam@tu-dortmund.de, imran.mazumder@tu-dortmund.de)
[2]Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, 688 Beribadh Road, Dhaka, 1207, Bangladesh (e-mails: raqib.mahmud@ulab.edu.bd, abul.azad@ulab.edu.bd)

Corresponding author: Md Asifuzzaman Jishan (e-mail: md-asifuzzaman.jishan@tu-dortmund.de).

ABSTRACT Object detection systems based on deep learning have been immensely successful in complex object detection tasks images and have shown potential in a wide range of real-life applications including the COVID-19 pandemic. One of the key challenges in containing and mitigating the infection among the population is to ensure and enforce the proper use of face masks. The objective of this paper is to detect the proper use of facial masks among the urban population in a megacity. In this study, we trained and validated a new dataset to detect images such as 'with mask', 'without mask', and 'mask not in position' using YOLOv5. The dataset is comprised of 6550 images with the three classes. The dataset attained a commendable performance accuracy of 95% on mAP. This study can be implemented for automated scanning for monitoring the proper use of face masks in different settings of public spaces.

KEYWORDS Mask detection; Object detection; Image classification; YOLOv5; COVID-19; Pandemic.

I. INTRODUCTION

THE Coronavirus disease (COVID-19) was first identified in Wuhan, China, in 2019, and eventually widespread globally to turn out to be the 5th recorded pandemic since the pandemic in 1918. Almost two years after COVID-19 was first identified, there were more than two hundred million identified cases and over 4.6 million deaths. In the early days, following the implementation of special efforts to absorb the virus, the situation in Wuhan became more delicate. China used to record dozens of new cases per day at the start of the outbreak, but that number had dropped to a few dozen by the end of March. On the other hand, cases in Europe were rapidly increasing day by day, with Italy recording over 250 deaths within a day. As a result, the WHO determined on March 13th that Europe had become the pandemic's epicenter. The United States also proclaimed a state of emergency on the same day.

To combat the pandemic, severe precautions have been implemented in various locations around the world. Social distancing, Face-mask usage, and tour regulations, as well as recommendations on proper handwashing practices, were reintroduced [1]. However, because those tactics were only projected to slow the virus's spread, scientists realized that a vaccine would be required to overcome the pandemic. Social distancing techniques were maintained by staying at home, averting crowded areas, the use of no-touch greetings, and bodily distancing people from others [2]. Some governments were mandating and restricting social distancing during the outbreak. On the other hand, the use of a face mask alone was not sufficient to provide appropriate safety against the virus but proper use of the mask began to be practiced worldwide as the virus spread over droplets [3].

One of the main challenges in the process of overcoming the virus spread was detecting the patients and understanding the nature of the virus where machine learning and artificial intelligence have been very useful tools and means to fight against the virus [4]. The machine learning algorithms also had a huge impact on contact tracing of the infected, predicting and forecasting the infection rate as well as drug development research for the Covid-19 pandemic [5]. But it was never easy to implement models that use human or object detection, though some systems were already there, new systems still had to be built. Automatic detection of humans in thermal imaging using YOLO was already there [6]. High-quality detection on generic and specific object datasets using Cascade R-CNN was already available [7]. The detection of facemask and human faces is superior and faster using YOLO, which is useful in detecting Social Distancing violations [8]. But after the pandemic happened, since 2020, research work on mask detection based on machine learning algorithms has skyrocketed. The YOLO-ASC approach was proposed which allows the detection of objects more accurately even without a background in real-time [9]. YOLO could also be improved in terms of predicting the absolute distance of objects using only information from a monocular camera with an inference speed of 45 frames per second [10]. Detecting the human faces was also achievable using the Viola-Jones algorithm. The distance between individuals' was measured to check if people are maintaining social distance [11]. Edge-AI algorithm was proposed to build technology-oriented solutions for detecting mask usage in moving human objects, which could be applied not only in pandemic times but also to general problems in society and healthcare systems [12]. Not only detecting mask usage or social distancing, but Cascade VGGCOV19-NET architecture also improved the already existing automatic detection of COVID-19 cases from X-ray images [13]. Six different versions of the YOLO object detection algorithm (YOLOv3, YOLOv3-tiny, YOLOv4, YOLOv4-tiny, YOLOv5x, and YOLOv5s) were already evaluated for real-time bunch detection and counting in grapes [14]. YOLO was also used to detect the hottest temperatures in the regions of interest (ROIs) of the human face in thermographic images, to identify the febrile states in humans [15].

But one of the main challenges of working with machine learning algorithms is, that to train the machine, a huge amount of data is needed, in this case, it can be footage of masked and unmasked people. Creating a dataset was the main challenge of our project. We gathered thousands of real-life data to create our dataset. We emphasized more on data quality more than machine learning models as they (models) already exist, and the quality of the dataset determines the final accuracy. The uniqueness of our project is, that we did not just classify masked and unmasked but also the proper usage of face-mask; that means our system can also detect if a person is wearing the mask properly. We have used YOLOv5 and YOLOv5n, YOLOv5s, and YOLOv5m pre-trained models.

The organization of the paper is the following: This paper starts with the introduction section which briefly portrays the history of Covid-19, its severity of it, and the challenges the world is going through because of it followed by an outline of our study and its significance. It also talks about the solutions and steps that were taken to overcome the situation and how machine learning and artificial intelligence had an impact on it, which we also used for this project. Next, it has the literature review which includes relevant information regarding past research work on this matter. After the Literature review section, it tells about the dataset and methodology followed by results and conclusions which summarize the whole project. Finally, the references used for this research work can be found in the last part.

II. LITERATURE REVIEW

In the field of image processing and computer vision, datasets always play a vital role. Because accuracy and loss mostly depend on the dataset, model, fine-tuning, and so on. But, if someone wants to implement a new dataset that is related to a new perspective, then definitely data is more important. Moreover, it is also important how one can use and implement the dataset for machine-level processing. Furthermore, an appropriate and efficient model is also important alongside the dataset, because the model will learn from the data and show the accuracy, loss, and the final inference result based on the dataset.

In image classification, it is more important that how can a model identify objects from the dataset. There are different types of image classification models available, but among them, YOLO is mostly popular. In the paper [16], the authors implement a YOLOv5 model for measuring social distancing and social restrictions. Because social distancing understandably plays a crucial role in slowing down the infection rate. Moreover, for their implementation, they used the Common Objects in Context (COCO) dataset along with 200,000 images and 80 classes. In the paper [17], the authors propose the corrected annotations of the Singapore Maritime Dataset (SMD) dataset with DNN algorithms and introduce a new dataset name is SMD-Plus. To extend, they also implemented the YOLOv5 model along with the SMD-Plus dataset for object detection and classifications.

The paper [18] utilized a dataset that is a combination of WIDER-FACE and MAFA for medical face mask detection. That dataset belongs to 7,959 images along with two classes, a covered face, and an uncovered face. For that research, they used the YOLOv4 framework and gained 99.98% accuracy during the training and testing. Furthermore, in the paper [19] the authors showed a real-time face mask detection system using a Tiny-YOLOv4 model. They implemented the Kaggle face mask dataset for their investigation. For the face mask detection based on transfer learning and modeled with PP-YOLO [20], the author used Pascal VOC standard dataset which contains 7959 images and 16,635 annotations. They achieved the mask detection mAP of that model is 86.69%. These papers report the benchmark performances for this task and we carry our investigation by drawing inspiration from these works.

A device called a face mask detector was installed at Politeknik Negeri Batam for real-time face mask detection using YOLO V4 deep learning algorithm [21]. In this work [22], the approach used to detect face masks achieved high accuracy. It had a 96% classification and detection accuracy.

In the paper [23], they replaced Mask-R-CNN with You only look once (YOLO) to increase the real-time processing speed without losing the accuracy. YOLO is also a more efficient model. The paper [24] shows that, on the self-made verification set, the algorithm's mAP is 86.92 percent, and the preset process may be accomplished in the actual identification process by operating the robot. Two state-of-the-art object detection models, YOLOv3 and faster R-CNN were used to achieve the accuracy in the paper [25]. It proposed a technique that drew bounding boxes around the faces, based on the presence of the mask, and kept a record of the ratio of mask usage.

Two different datasets containing masked and unmasked images were combined in one dataset in the research [26]. To achieve better accuracy, the best number of anchor boxes was calculated by mean IoU. The highest average precision percentage was 81%. In the paper [27], researchers proposed a new system using a YOLOv4 object detector that integrated tiny YOLO v4 with spatial pyramid pooling (SPP) module and an additional YOLO detection layer. The dataset had more than 50,000 images. The paper [28] was also based on YOLO-v4, but with an improved CSPDarkNet53 into the trunk feature extraction network to reduce the computing cost of the network and improve the learning ability of the model. The model in [29] used the Keras, Tensor-flow, and OpenCV methods. They used two datasets, with and without a facemask. A Raspberry- PI camera was used which captured the live streaming video and converted them into images, which were later used as the data. The paper [30] talks about Darknet which is also used for object detection which is an open source neural network framework written in the C and CUDA programming languages. The Jetson Nano is a System On Module (SoM) and developer kit from the Nvidia Jetson family. Finally the system used in the paper [31], can distinguish the detected faces with three labels as with a mask, without a mask, and mask worn incorrectly, similarly to our work, and achieve a mean Average Precision (mAP) of 87.94%. A lot of work had already been done in the improvement of object detection using YOLO and other systems. However, there are limitations in every system. Our work focused more on the proper usage of face masks with a better detection rate.

III. DATASET

Nowadays data is an important part of our life. From the morning until late at night we use and depend on lots of types of data. For doing research it is more important to collect and choose relevant data. In general, for doing research the priority of data is almost 60% but in the industry field that rate is almost 90% [32-33]. Apart from that, we can understand how important data is in our real-life problems and find out their solutions. We can collect data from various sources and implement them to solve real-life problems.

At present, in the backdrop of the COVID-19 outbreak, the topmost priority in our life is to bring about appropriate

Figure 1. Samples of the proposed dataset for unmasked faces

Figure 2. Samples of the proposed dataset for masked faces

lifestyle changes to mitigate the transmission and severity of the infection. Wearing a mask in public places has become an imperative civil practice so much so that judicial interventions are imposed where the practice of wearing a mask appropriately is not followed. Many public places including shopping malls, public transportation, offices, educational premises, and so on do not allow admittance without wearing a mask. So, the monitoring of face mask practices on public premises is an integral civil practice now. In our paper, we have introduced a new dataset named Covid Face-Mask Monitoring Dataset [34]. Our main focus is to detect whether people wearing face masks or not in the streets and crowded public domains. Moreover, it is noted that a few people do not wear masks properly which is as bad as not wearing masks at all and contribute to the spreading of infection is harmful to other. So, we posit to detect the proper wearing of face masks as well. Moreover, in the perspective of a densely populated country such as a large number of the population try to avoid the covid rules and regulations and they have no intention to use facial masks [34]. This malpractice regarding aversion to using face-masks is commonly noticed in the walking street, bus stops, street tea stalls, foot-over bridges, and at similar public places in the country.

Our proposed dataset contains 6,550 images and those images were collected from the walking street, bus stop, street tea stall, foot-over bridge, and so on. We used personal cell phones and DSLRs for collecting frames and adding them to our final dataset. We have also planned to collect images from similar public places using an action camera or

Table 1. Resizing and Formation of Dataset

No of frames	File Format	Image Dimension	Class Size
6550	JPG	1080 × 720	3

CCTV surveillance camera. But, in Bangladesh, most public places lack CCTV coverage. CCTV surveillance cameras are mostly used in educational institutions, shopping complexes, hospitals, and offices where using a mask is mandatory. But the scope of our investigation is to incorporate multiple perspectives and vantage points of monitoring wearing face masks. in our dataset, there are three classes which are 1. Mask, 2. No Mask, and 3. Mask not in position.

We represented the sample images of our dataset in Figures 1, 2, and 3. In Figure 1, we represented the unmasked faces of people from the dataset as we previously described that in our dataset three types of classes exist. Furthermore, in Figure 2 we represented peoples images who are wearing masks properly. In the end, in Figure 3 we represented those people who are wearing the mask but not correctly and that is harmful to others people.

Figure 3. Samples of the proposed dataset for masked but not weared in correct position of faces

We classified the full dataset into three classes and we resized the whole dataset in a fixed image dimension format. The dimension of the images is 1080 × 720 pixels. We selected 5,750 images for training purposes and 800 images for validation purposes. All the frame formats are set as JPG, and the data formats, image dimensions, their resolutions, and bit depth are presented in Table 1. We used LabelImg tools for the labeling of the whole self-made dataset. During the dataset labeling, we implemented class 0 for mask, class 1 for no mask, and class 2 for mask not in position.

IV. METHOD

You Only Look Once (YOLO) is one of the fastest and most popular algorithms for object detection. YOLO is more popular than others because it has remarkable learning capability, accuracy, and training speed. It has been used in various types of applications such as detection of the traffic light, object detection, mask detection, car plate detection, and so on. YOLO mainly works depending on residual blocks, bounding box, and intersection over union [9-10].

There are many improvements from the YOLO to YOLOv5. In YOLO, it was mainly responsible for object

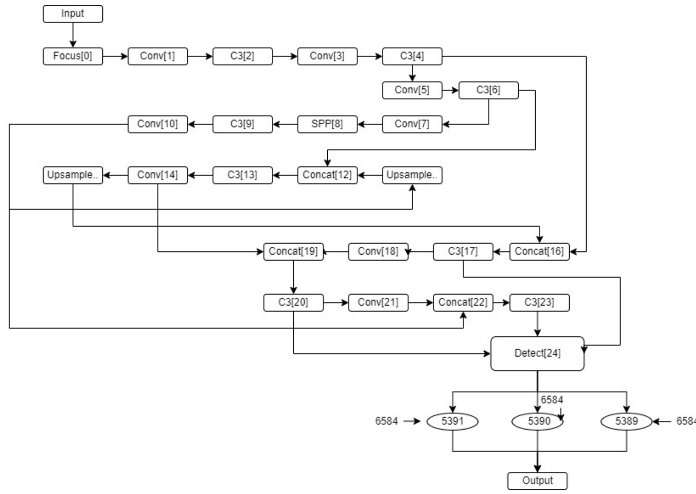


Figure 4. Model architechture of YOLOv5

detections and loss of confidence. But, in YOLOv2 K-Means with anchor added. Moreover, YOLOv2 also added a fully convolutional neural network system along with two stages of training. On the other hand, in the YOLOv3 multi-scale detection using FPN was added. To extend, the YOLOv4 also added new features like SPP, the activation function of MISH, data enhancements using Mosaics, loss function, and so on. Finally, YOLOv5 came with flexible control of the model, data enhancement, and application of the Hardswish activation function [16].

We implemented YOLOv5 for our proposed dataset which was used for the facial mask detection system. YOLOv5 family has three different architecture blocks, and they are backbone, neck, and head. YOLOv5 backbone mainly employs CSPDarknet as the backbone of extraction features from the image dataset using a partial network. In addition, if we give concern on the YOLOv5 neck, then we see that uses for the PANet to generate pyramids feature network [9]. To extend, it performs aggregation on the feature and then passes it towards the head for the predictions. And YOLOv5 head has layers that can generate predictions from the anchor boxes for the detection objects. For the activation function and optimization techniques, YOLOv5 mainly uses Leaky ReLU and Sigmoid activation functions. Moreover, as the optimization, it uses SGD and ADAM optimization techniques [16]. We represented the YOLOv5 model architecture in Figure 4.

In YOLOv5, it has five different pre-trained checkpoints for object detection. Firstly, we explore the case of the YOLOv5n (Nano) pre-trained model. In general, it always uses an image size of 640 × 640, and its mAP is 28.0 on the scale of 0.5:0.95 and 45.7 on the scale of 0.5. That pre-trained model CPU (b1) based speed is 45 ms, 6.3 ms V100 (b1), 0.6 ms V100 b32. It has 1.9 million parameters and 4.5 billion FLOPs @640 (B) [16]. Secondly, the YOLOv5s (Small) pre-trained model always uses an image size of 640 × 640, and its mAP is 37.4 on the scale of 0.5:0.95 and 56.8 on the scale of 0.5. That pre-trained model CPU (b1) based speed is 98 ms, 6.4 ms V100 (b1), 0.9 ms V100 b32. It has 7.2 million parameters and 16.5 billion FLOPs @640 (B) [17].

Likewise, in the YOLOv5m (Medium) pre-trained model, it always uses an image size of 640 × 640, and its mAP is 45.4 on the scale of 0.5:0.95 and 64.1 on the scale of 0.5. This pre-trained model CPU (b1) based speed is 224 ms, 8.2 ms V100 (b1), 1.7 ms V100 b32. It has 21.2 million parameters and 49.0 billion FLOPs @640 (B). Furthermore, the YOLOv5l (Large) pre-trained model always uses an image size of 640 × 640, and its mAP is 49.0 on the scale of 0.5:0.95 and 67.3 on the scale of 0.5. That pre-trained model CPU (b1) based speed is 430 ms, 10.1 ms V100 (b1), 2.7 ms V100 b32. It has 46.5 million parameters and 109.1 billion FLOPs @640 (B). Finally, the YOLOv5x (Extra Large) pre-trained model uses an image size of 640 × 640, and its mAP is 50.7 on the scale of 0.5:0.95 and 68.9 on the scale of 0.5. That pre-trained model CPU (b1) is based on the speed of 766 ms, 12.1 ms V100 (b1), and 4.8 ms V100 b32. It has 86.7 million parameters and 205.7 billion FLOPs @640 (B) [16-17]. But, for our proposed dataset of facial mask recognition tasks, we implemented YOLOv5 nano, small and medium pre-trained checkpoints.

V. RESULT AND DISCUSSION

For the mask detection, we implemented the YOLOv5 model along with YOLOv5n, YOLOv5s, and YOLOv5m pre-trained models. During the training period, the proposed model has a batch size of 8, a learning rate of 0.001, and trained 75 epochs for each pre-trained model. Precision, recall, F-score, and mAP are mainly used for measuring the exact accuracy of detecting the object. Precision mainly focused on the accurate prediction and recall measures that the positive classes detect easily. Precision and recall are calculated by the below formulas. Here, Pos denoted Positive, Neg denoted Negative, T denoted True and F denoted False.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1)$$

$$Recall = \frac{T\_Positive}{T\_Positive + F\_Negative} \quad (2)$$

Accuracy is mainly used as another metric for measuring the performance of the classification. Here, T denoted True and F denoted False also. The equation of accuracy and error is illustrated below:

$$Accuracy = \frac{T\_Pos + T\_Neg}{T\_Pos + T\_Neg + F\_Pos + F\_Neg} \quad (3)$$

$$Error = 1 - Accuracy \quad (4)$$

The mean average of precision (mAP) is the average of the AP calculated for all the classes. The equation of mAP is given below:

$$Mean\ Avg\ Precision\ (mAP) = \frac{\sum_{q=0}^{Q} AveP(q)}{Q} \quad (5)$$

In the statistical analysis of the binary classification, the F-score or F-measure is a proportion of a test's precision. It is determined by precision and recall of the test, where the precision is the quantity of genuine positive outcomes partitioned by the quantity of every single positive outcome, including those not recognized accurately, and the recall is the quantity of genuine positive outcomes separated by the quantity of all examples that ought to have been distinguished as sure. Precision is otherwise called positive predicted value, and recall is also known as sensitivity in diagnostic binary classification. The equation of F-score is the given below:

$$F = 2 \times \frac{P \times R}{P + R} \quad (6)$$

A. YOLOV5N PRE-TRAINED MODEL IMPLEMENTATIONS OUTCOME

Our achieved outcomes are remarkable compared with the data and image quality. Because from Bangladesh's perspective it is very tough to collect data properly due to the limitations of camera facilities. Moreover, there are existing different types of images such as noisy images, light capacity, sunlight, shadow, and others. For that reason achieving good results is comparatively hard.

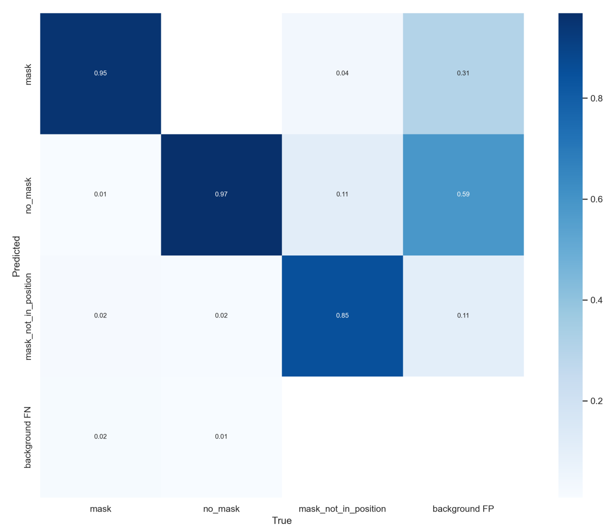
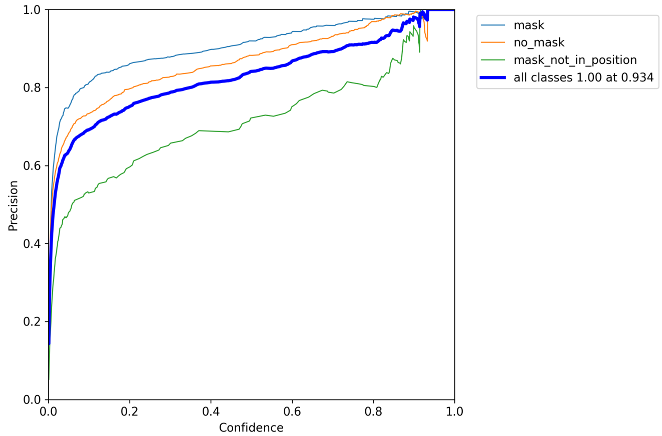
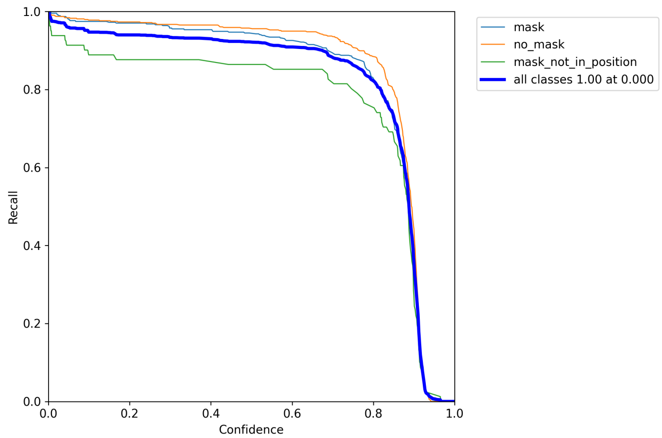


Figure 5. Confusion matrix for YOLOv5n using proposed dataset

In this section, we represented the outcomes of the YOLOv5n pre-trained model using our proposed dataset for mask detections. In Figure 5, we showed the confusion matrix where the result is 0.95 for mask detection, 0.97 for no mask detection, and 0.85 for mask not in the correct position detection. This is the benchmark score for the proposed dataset.



(a) Precision

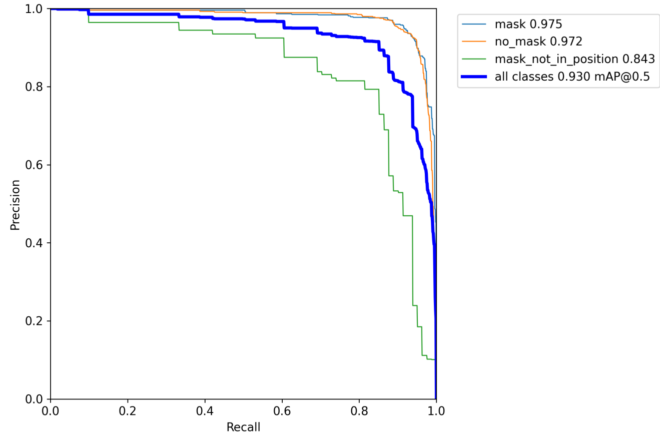


(b) Recall

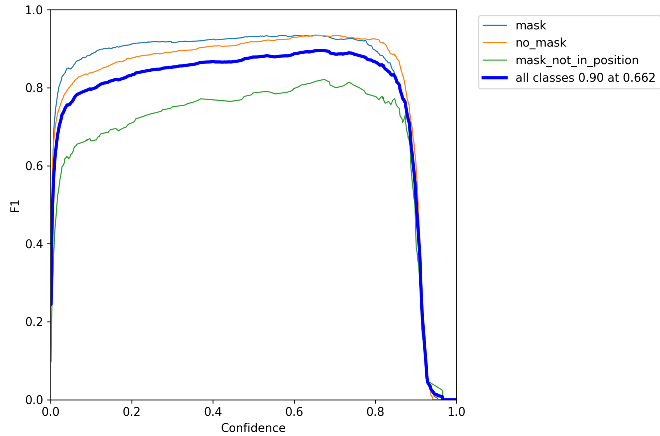
Figure 6. Precision and recall curve for YOLOv5n

After that, our concern is the precision and recall score for the proposed dataset. Precision scores are mainly dependent on accurate prediction, and recall scores can measure easily detect positive classes. We trained the whole dataset using batch size 8 and ran 75 epochs. During the training, for the YOLOv5n pre-trained model the best precision score is 0.89078 which is the benchmark score of our proposed dataset. To extend, for the same pre-trained model the best recall score is 0.9019 which is also a benchmark score for the same dataset. Precision and recall curves are represented in Figure 6. Figure 6(a), represented the precision vs confidence score and Figure 6(b) represented the recall vs confidence score.

After the precision and recall score, precision vs recall and F1 scores are more important for the proposed dataset. Precision vs recall and F1 curves are represented in Figure 7. In Figure 7(a), the precision vs recall score is represented,



(a) Precision vs Recall



(b) F1 curve

Figure 7. Precision vs Recall and F1 curve for YOLOv5n

and Figure 7(b) represented the F1 score curve for the YOLOv5n pre-trained model.

B. YOLOV5S PRE-TRAINED MODEL IMPLEMENTATIONS OUTCOME

In this section, we represented the outcomes of the YOLOv5s pre-trained model using our proposed dataset for mask detections. In Figure 8, we showed the confusion matrix where the result is 0.97 for mask detection, 0.97 for no mask detection, and 0.81 for mask not in the correct position detection. This is the benchmark score for the proposed dataset.

After that, our concern is the precision and recall score for the proposed dataset. Precision scores are mainly dependent on accurate prediction, and recall scores can measure easily detect positive classes. We trained the whole dataset using batch size 8 and ran 75 epochs. During the training, for the YOLOv5s pre-trained model the best precision score is 0.90517 which is the benchmark score of our proposed dataset. To extend, for the same pre-trained model the

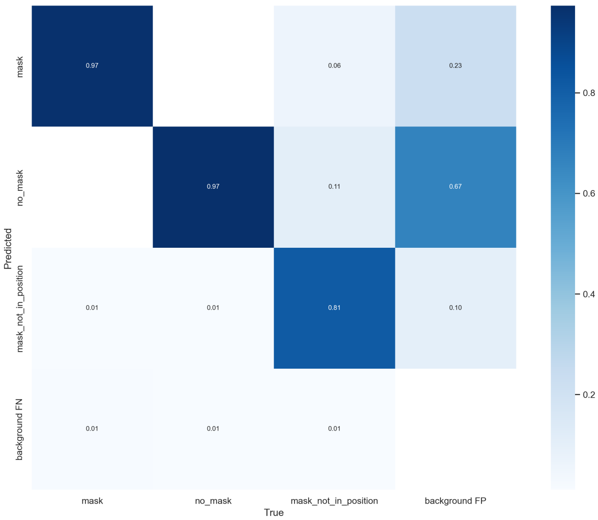


Figure 8. Confusion matrix for YOLOv5s using proposed dataset

best recall score is 0.87601 which is also a benchmark score for the same dataset. Precision and recall curves are represented in Figure 9. Figure 9(a), represented the precision vs confidence score and Figure 9(b) represented the recall vs confidence score.

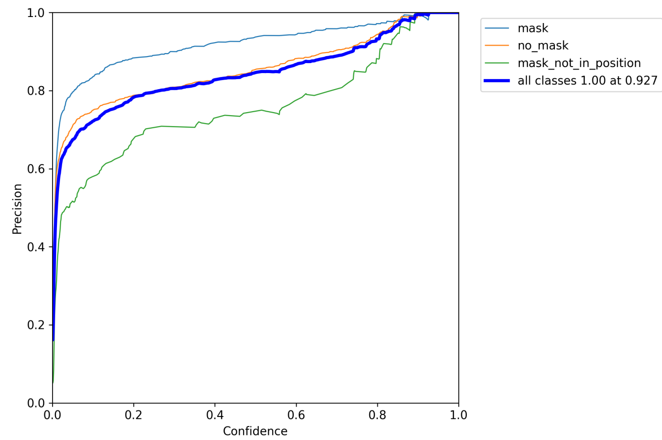
After the precision and recall score, precision vs recall and F1 scores are more important for the proposed dataset. Precision vs recall and F1 curves are represented in Figure 10. In Figure 10(a), the precision vs recall score is represented, and Figure 10(b) represented the F1 score curve for YOLOv5s pre-trained model.

C. YOLOV5M PRE-TRAINED MODEL IMPLEMENTATIONS OUTCOME

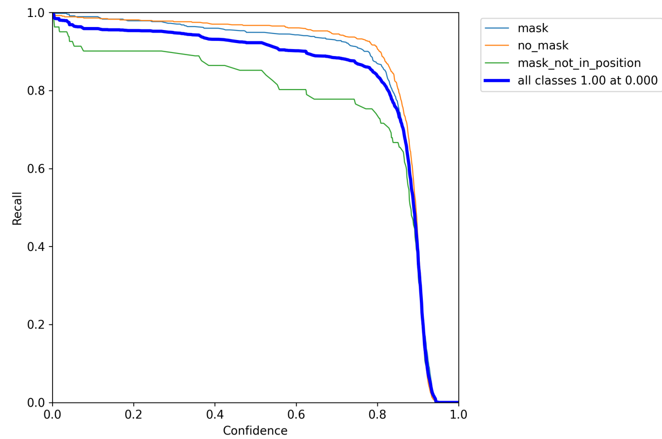
In this section, we represented the outcomes of the YOLOv5m pre-trained model using our proposed dataset for mask detections. In Figure 11, we showed the confusion matrix where the result is 0.96 for mask detection, 0.95 for no mask detection, and 0.84 for mask not in the correct position detection. This is the benchmark score for the proposed dataset.

After that, our concern is the precision and recall score for the proposed dataset. Precision scores are mainly dependent on accurate prediction, and recall scores can measure easily detect positive classes. We trained the whole dataset using batch size 8 and ran 75 epochs. During the training, for the YOLOv5m pre-trained model the best precision score is 0.85522 which is the benchmark score of our proposed dataset. To extend, for the same pre-trained model the best recall score is 0.8947 which is also a benchmark score for the same dataset. Precision and recall curves are represented in Figure 12. Figure 12(a), represented the precision vs confidence score and Figure 12(b) represented the recall vs confidence score.

After the precision and recall score, precision vs recall



(a) Precision



(b) Recall

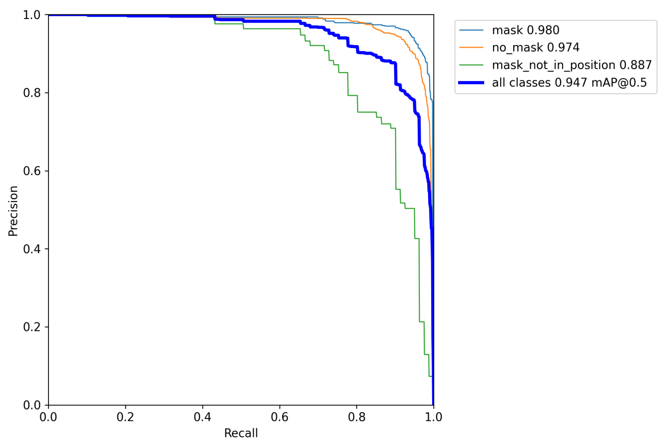
Figure 9. Precision and recall curve for YOLOv5s

and F1 scores are more important for the proposed dataset. Precision vs recall and F1 curves are represented in Figure 13. In Figure 13(a), the precision vs recall score is represented, and Figure 13(b) represented the F1 score curve for YOLOv5m pre-trained model.

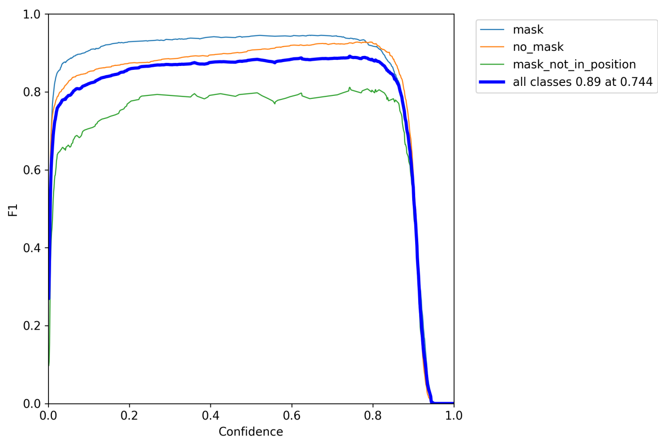
D. THE MEAN AVERAGE OF PRECISION SCORE (MAP) IN THE SCALE OF 0.5

Many classification and object detection algorithms, such as Faster R-CNN, MobileNet SSD, and YOLO, use Mean Average Precision (mAP) to evaluate their models and publish their research outcomes. The Mean Average Precision (mAP) for object detection is the average of the AP calculated for all the classes. It is also important to note that in some studies AP and mAP were used interchangeably.

In Figure 14, we represented the mean Average Precision (mAP) score on a scale of 0.5. For our proposed dataset, we got 3 types of results for three pre-trained models. For the YOLOv5n pre-trained model, we achieved a 0.92943 score which is the benchmark result for our proposed



(a) Precision vs Recall



(b) F1 curve

Figure 10. Precision vs Recall and F1 curve for YOLOv5s

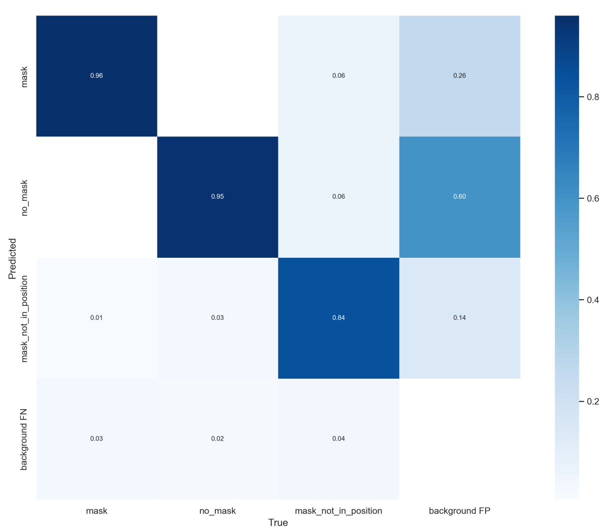
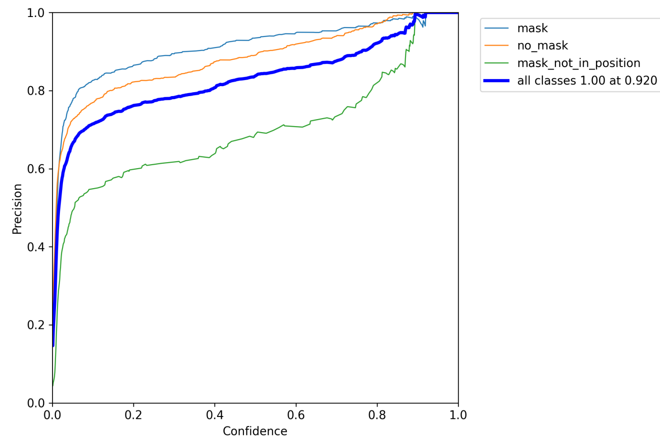
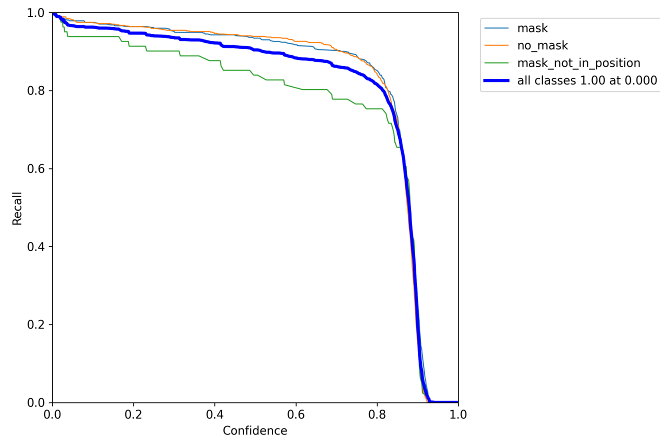


Figure 11. Confusion matrix for YOLOv5m using proposed dataset



(a) Precision



(b) Recall

Figure 12. Precision and recall curve for YOLOv5m

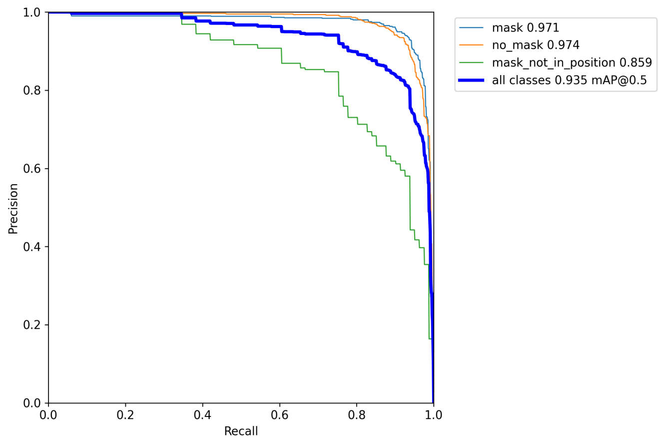
dataset. Moreover, when we worked on the YOLOv5s pre-trained model at that time we see that the pre-trained model achieved a 0.9469 score which is the benchmark result for our proposed dataset. To extend, for the YOLOv5m pre-trained model that score was 0.93473.

E. THE MEAN AVERAGE OF PRECISION SCORE (MAP) IN THE SCALE OF 0.5:0.95

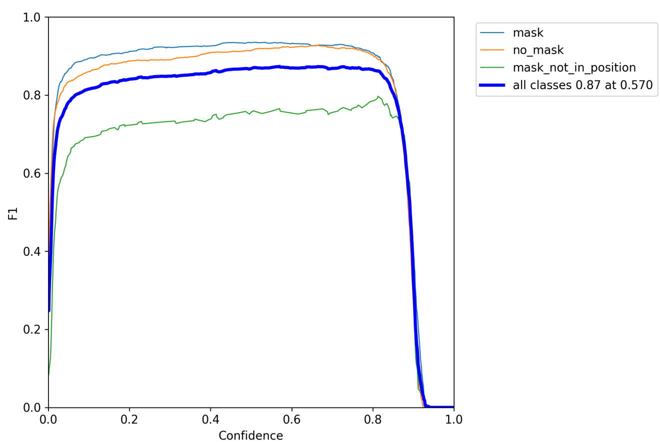
In Figure 15, we represented the mean Average Precision (mAP) score on a scale of 0.5:0.95. For our proposed dataset, we got results for three pre-trained models which are YOLOv5n, YOLOv5s, and YOLOv5m. For the YOLOv5n pre-trained model, we achieved a 0.71373 score on the scale of mAP 0.5:0.95 which is the benchmark result for our proposed dataset. Moreover, during the training period of the YOLOv5s pre-trained model we see that model achieved a 0.72532 score on the scale of mAP 0.5:0.95 which is the benchmark result for our proposed dataset. To extend, for the YOLOv5m pre-trained model that score was 0.71753.

Table 2. Models Comparison of Covid Face-Mask Monitoring Dataset

Model	Precision	Recall	mAP_0.5	mAP_0.5:0.95
YOLOv5n	0.89078	0.9019	0.92943	0.71373
YOLOv5s	0.90517	0.87601	0.9469	0.72532
YOLOv5m	0.85522	0.8947	0.93473	0.71753



(a) Precision vs Recall



(b) F1 curve

Figure 13. Precision vs Recall and F1 curve for YOLOv5m

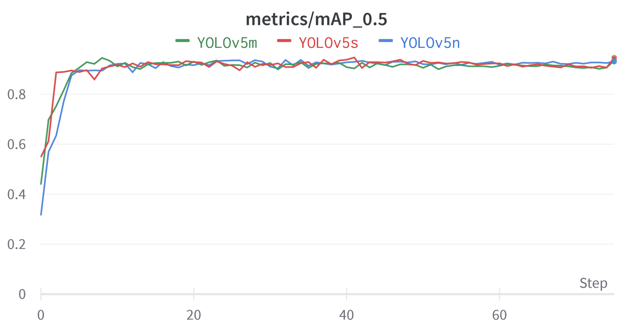


Figure 14. Mean average of precision score (mAP) in the scale of 0.5 for proposed dataset

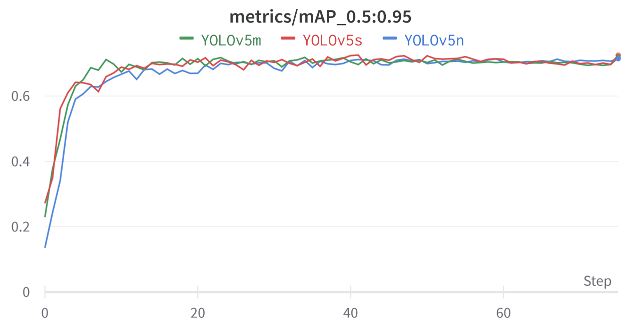


Figure 15. Mean average of precision score (mAP) in the scale of 0.5:0.95 for proposed dataset

Figure 16. Output of the model during testing

F. DISCUSSION

We represented a dataset named Covid Face-Mask Monitoring Dataset for detecting the mask, no mask, and mask not wearing in the correct position and that dataset is based on the Bangladesh perspective images. We achieved remarkable precision, recall, mAP, and F-score results for our proposed dataset using three pre-trained models. We represented the confusion matrix, precision and recall curve, precision vs recall curve, and F-score for the YOLOv5n pre-trained model in Figures 5, 6, and 7 respectively. To extend, we also represented the confusion matrix, precision and recall curve, precision vs recall curve, and F-score for the YOLOv5s pre-trained model in Figures 8, 9, and 10 respectively. Moreover, we represented the confusion matrix, precision and recall curve, precision vs recall curve, and F-score for the YOLOv5m pre-trained model in Figures 11, 12, and 13 respectively. After that, we showed a mAP 0.5 score which is the benchmark score for three pre-trained models using our proposed dataset in Figure 14. In addition, we represented the mAP 0.5:0.95 score for that dataset using three pre-trained models in Figure 15. Finally, we showed a comparison result in Table 2 and the test results in Figures 16 and 17.

Figure 17. Output of the model during testing

VI. CONCLUSION

In this work, we have presented a comprehensive study for the detection of proper face mask usage based on a novel dataset, as it is proven to be a vital way of preventing COVID-19 infection from person to person. For object detection, we have used YOLOv5 and YOLOv5n, YOLOv5s, and YOLOv5m pre-trained models to achieve higher accuracy and better performance. To train and validate our detection system, we prepared a dataset consisting of thousands of masked and unmasked face images. We did not just focus on the presence of masks but also on the proper use and positioning of them, and we were able to show that YOLOv5 is an effective and efficient model for the use case. Our proposed solution may be implemented in a myriad of circumstances where monitoring the proper usage of face-mask in a crowded setup is not only limited to CCTV observations.

References

[1] Chiu, N. C., Chi, H., Tai, Y. L., Peng, C. C., Tseng, C. Y., Chen, C. C., Tan, B. F., & Lin, C. Y., "Impact of Wearing Masks, Hand Hygiene, and Social Distancing on Influenza, Enterovirus, and All-Cause Pneumonia During the Coronavirus Pandemic: Retrospective National Epidemiological Surveillance Study," *Journal of medical Internet research*, 2020, 22(8), e21257. https://doi.org/10.2196/21257.

[2] Kebede Y, Yitayih Y, Birhanu Z, Mekonen S, Ambelu A., "Knowledge, perceptions and preventive practices towards COVID-19 early in the outbreak among Jimma university medical center visitors, Southwest Ethiopia," *PLoS ONE*, 2020, 15(5): e0233744. https://doi.org/10.1371/journal.pone.0233744.

[3] Martinelli, L., Kopilaš, V., Vidmar, M., Heavin, C., Machado, H., Todorović, Z., Buzas, N., Pot, M., Prainsack, B., & Gajović, S., "Face Masks During the COVID-19 Pandemic: A Simple Protection Tool With Many Meanings," *Frontiers in public health*, 2021, 8, 606635. https://doi.org/10.3389/fpubh.2020.606635.

[4] Kushwaha, S., Bahl, S., Bagha, A. K., Parmar, K. S., Javaid, M., Haleem, A., & Singh, R. P., "Significant applications of machine learning for COVID-19 pandemic," *Journal of Industrial Integration and Management*, 2020, 5(04), 453-479.

[5] Lalmuanawma, S., Hussain, J., & Chhakchhuak, L., "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review," *Chaos, Solitons & Fractals*, 2020, 139, 110059.

[6] Ivašić-Kos, M., Krišto, M., & Pobar, M., "Human detection in thermal imaging using YOLO," *In Proceedings of the 2019 5th International Conference on Computer and Technology Applications*, 2019, pp. 20-24.

[7] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High Quality Object Detection and Instance Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, vol. 43, no. 5, pp. 1483-1498, 1 May 2021, doi: 10.1109/TPAMI.2019.2956516.

[8] K. Bhambani, T. Jain and K. A. Sultanpure, "Real-time Face Mask and Social Distancing Violation Detection System using YOLO," *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)*, 2020, pp. 1-6, doi: 10.1109/B-HTC50970.2020.9297902.

[9] P. Hurtik, V. Molek and P. Vlasanek, "YOLO-ASC: You Only Look Once And See Contours," *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1-7, doi: 10.1109/IJCNN48605.2020.9207223.

[10] Vajgl, M.; Hurtik, P.; Nejezchleba, T., "Dist-YOLO: Fast Object Detection with Distance Estimation," *Appl. Sci.*, 2022, 12, 1354. https://doi.org/10.3390/app12031354.

[11] Hardan, F., & R. J. Almusawi, A., "Developing an Automated Vision System for Maintaing Social Distancing to Cure the Pandemic," *Al-Khwarizmi Engineering Journal (Alkej)*, 2022, 18(1), 38–50. https://doi.org/10.22153/kej.2022.03.002.

[12] Sengupta, K., Srivastava, P.R., "HRNET: AI-on-Edge for Mask Detection and Social Distancing Calculation," *SN COMPUT. SCI.*, 2022, 3, 157 (2022). https://doi.org/10.1007/s42979-022-01023-1

[13] Karacı A., "VGGCOV19-NET: automatic detection of COVID-19 cases from X-ray images using modified VGG19 CNN architecture and YOLO algorithm," *Neural Comput Appl.*, 2022, Jan 24:1-22. doi: 10.1007/s00521-022-06918-x. Epub ahead of print. PMID: 35095212; PMCID: PMC8785935.

[14] Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F., "Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms," *Agronomy*, 2022, 12, 319. https://doi.org/10.3390/agronomy12020319.

[15] da Silva, J.R.; de Almeida, G.M.; Cuadros, M.A.d.S.L.; Campos, H.L.M.; Nunes, R.B.; Simão, J.; Muniz, P.R., "Recognition of Human Face Regions under Adverse Conditions—Face Masks and Glasses—In Thermographic Sanitary Barriers through Learning Transfer from an Object Detector," *Machines*, 2022, 10, 43. https://doi.org/10.3390/machines10010043.

[16] imam al amin and F. Arby, "Implementation of YOLO-v5 for a Real Time Social Distancing Detection," *JAIC*, 2022, vol. 6, no. 1, pp. 01-06.

[17] Kim, J.-H.; Kim, N.; Park, Y.W.; Won, C.S., "Object Detection and Classification Based on YOLO-V5 with Improved Maritime Dataset," *J. Mar. Sci. Eng*, 2022, 10, 377. https://doi.org/10.3390/jmse10030377.

[18] Degadwala, S., Vyas, D., Chakraborty, U., Dider, A. R., & Biswas, H., "Yolo-v4 deep learning model for medical face mask detection," *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, (pp. 209-213). IEEE.

[19] K. V. Sathyamurthy, A. R. Shri Rajmohan, A. Ram Tejaswar, K. V and G. Manimala, "Realtime Face Mask Detection Using TINY-YOLO V4," *2021 4th International Conference on Computing and Communications Technologies (ICCCT)*, 2021, pp. 169-174, doi: 10.1109/ICCCT53315.2021.9711838.

[20] Jian, W., & Lang, L., "Face mask detection based on Transfer learning and PP-YOLO," *In 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2021, (pp. 106-109). IEEE.

[21] S. Susanto, F. A. Putra, R. Analia and I. K. L. N. Suciningtyas, "The Face Mask Detection For Preventing the Spread of COVID-19 at Politeknik Negeri Batam," *2020 3rd International Conference on Applied Engineering (ICAE)*, 2020, pp. 1-5, doi: 10.1109/ICAE50557.2020.9350556.

[22] M. R. Bhuiyan, S. A. Khushbu and M. S. Islam, "A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225384.

[23] Liu, R., & Ren, Z., "Application of Yolo on Mask Detection Task," *In 2021 IEEE 13th International Conference on Computer Research and Development (ICCRD)*, 2021, (pp. 130-136). IEEE.

[24] Li, Y., Yan, J., & Hu, B., "Mask Detection Based On Efficient-YOLO" *In 2021 40th Chinese Control Conference (CCC)*, 2021, (pp. 4056-4061). IEEE.

[25] Singh, S., Ahuja, U., Kumar, M., Kumar, K., & Sachdeva, M., "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment," *Multimedia Tools and Applications*, 2021, Spring: 80(13), 19753-19768

[26] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M., "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*, 2021, 65, 102600.

[27] Kumar A, Kalia A, Sharma A, Kaushal M., "A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system," *J Ambient Intell Humaniz Comput*, 2021, 20:1-14. doi: 10.1007/s12652-021-03541-x.

[28] Yu, J., & Zhang, W., "Face mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, 2021, 21(9), 3263.

[29] Balaji, S., Balamurugan, B., Kumar, T. A., Rajmohan, R., & Kumar, P. P., "A brief Survey on AI Based Face Mask Detection System for Public Places," *Irish Interdisciplinary Journal of Science & Research (IIJSR)*, 2021.

[30] Prasetia, D. D., Yuswanto, A., & Wibowo, B., "DESIGN OF MACHINE LEARNING DETECTION MASK USING YOLO AND DARKNET ON NVIDIA JETSON NANO," *TEKNOKOM*, 2022, 5(1), 88-95.

[31] Z. Qin, Z. Guo and Y. Lin, "An Implementation of Face Mask Detection System Based on YOLOv4 Architecture," *14th International Conference on Computer Research and Development (ICCRD)*, 2022, pp. 207-213, doi: 10.1109/ICCRD54409.2022.9730470.

[32] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," *2nd ed., Morgan Kaufmann*, 2006.

[33] Mohammed J. Zaki, Wagner Meira JR, "Data Mining and Analysis," *Fundamental Concepts and Algorithms: Campridge University Press*, 2014.

[34] Masum, M. Islam; Jishan, Md Asifuzzaman; Mahmud, Khan Raqib; Azad, Abul Kalam Al, "Covid Face-Mask Monitoring Dataset," *Harvard Dataverse*, https://doi.org/10.7910/DVN/BKPMHT.

MD ASIFUZZAMAN JISHAN is studying as a student in the Master of Science in Data Science program within the Faculty of Statistics at the Technische Universität Dortmund, Germany. He has completed his Bachelor of Science in Computer Science and Engineering from the Department of Computer Science and Engineering at the University of Liberal Arts Bangladesh (ULAB). He has expertise in C, C++, Java, Python, MATLAB, and R programming languages as well as Tableau data visualization system. He also has working knowledge in different web programming languages: HTML, CSS, JavaScript (JS), Laravel framework, and database system. He has been active in the research with research interest in the area of big data analytics, bioinformatics, image processing, artificial intelligence, machine learning, and neural systems. (Email: md-asifuzzaman.jishan@tu-dortmund.de).

MD SHAHABUB ALAM is a full-stack Software Engineer with experience developing front-end and back-end applications. As well as machine learning models and data analysis and interpretation, he has developed and implemented research tools. In addition to expertise in R, Python, and C#, he is skilled at implementing web applications. He is also interested in the intersection of applied Data Science, Machine Learning, and Software Engineering. He is a strong engineering professional with a B.Sc. in Computer Science and Engineering from Ahsanullah University of Science and Technology. Working in a diverse team, collaborating and sharing ideas, and driving an organization forward are some of his favourite things to do. (Email: md-shahabub.alam@tu-dortmund.de)

IMRAN RASHID MAZUMDER is studying as a student of Master of Science in Data Science at Technische Universität Dortmund (TUD). He has completed his Bachelor of Science in Computer Science and Engineering from the Department of Computer Science and Engineering at the University of Information Technology and Science (UITS). He has expertise in C, C++, Python, and R programming languages as well as Tableau and Power BI. He also has working knowledge as well experience in web programming languages: HTML5, CSS, and database management systems. He is very much passionate about learning Cyber Security as well as Data Analytics. (Email: imran.mazumder@tu-dortmund.de)

KHAN RAQIB MAHMUD currently working as a lecturer within the department of Computer Science and Engineering at the University of Liberal Arts Bangladesh (ULAB). He has completed Bachelor of Science (Honors) and Master of Science in Mathematics from ShahJalal University of Science and Technology, Bangladesh. He received an Erasmus Mundus Scholarship from the Education, Audiovisual and Culture Executive Agency of the European Commission, to pursue a double Masters in Science degree in Computer Simulation for Science and Engineering and Computational Engineering, from Germany and Sweden. He was an MSc thesis student within the Computational Technology Laboratory of the Department of High Performance Computing and Visualization at KTH Royal Institute of Technology, Sweden. His current research interest includes machine learning and-pattern recognition, image processing and computer vision and adaptive dynamic system. (Email: raqib.mahmud@ulab.edu.bd)

ABUL KALAM AL AZAD received his PhD in Applied Mathematics from University of Exeter, United Kingdom, Masters of Science in Theoretical Physics and Bachelor of Science in Physics from University of Dhaka. He is currently an Associate Professor at the Department of Computer Science and Engineering, University of Liberal Arts Bangladesh (ULAB). Previously, he undertook post-doctoral research at Department of Computing and Mathematics, University of Plymouth, United Kingdom, and School of Biological Sciences, University of Bristol, United Kingdom, on a BBSRC fellowship. His research interest includes areas of theoretical and computational neuroscience, connectomics, multitimescale dynamics, self-organized criticality (SOC) and artificial intelligence. He has published a number of papers in peer-reviewed international journals and presented original research articles in numerous international conferences. (Email: abul.azad@ulab.edu.bd)