Real-Time Multiface Mask Automatic Detection System in Classroom Learning using YOLOv4 Deep Learning

Arif Fadllullah
Department of Computer Engineering,
Universitas Borneo Tarakan
Tarakan, Indonesia
arif.fadl@borneo.ac.id

Rahmatuz Zulfia
Department of Nursing,
Universitas Borneo Tarakan
Tarakan, Indonesia
rahmatuzzulfia@borneo.ac.id

Tegar Palyus Fiqar
Department of Mathematics and Information Technology,
Institut Teknologi Kalimantan
Balikpapan, Indonesia
tegar@lecturer.itk.ac.id

Awang Pradana
Department of Computer Engineering,
Universitas Borneo Tarakan
Tarakan, Indonesia
awang.pradana@borneo.ac.id

Abstract—During the Covid-19 pandemic, students were required to wear masks in classroom learning. However, students often do not use masks, so they are prone to transmission of Covid-19. For this reason, this study proposes the development of a real-time multiface mask automatic detection system in classroom learning using YOLOv4 deep learning. Experimental results on 22 samples of students who collected real-time/live video data every 3 minutes for 20 scenarios proved that the proposed system was successful in detecting objects wearing masks (WM) and not wearing masks (NWM) with the average percentage of *precision* was 95.63% for WM and 97.33% for NWM, the average percentage of *recall* was 61.61% for WM and 60.23% for NWM, and the average percentage of *F-measure* was 74.55% for WM and 74.00% for NWM. This results indicate an effective, valid and accurate proposed system for monitoring the use of masks in classroom learning easily and automatically.

Keywords—Multiface Mask, Covid-19, Real-Time Image, YOLOv4, Deep Learning

I. INTRODUCTION

In Indonesia, the education sector has implemented lockdown on almost all schools and universities for more than a year due to the Covid-19 pandemic. This has led to the teaching and learning process being carried out online with students learning from home, in order to ensure that learning does not stop and to minimize the risk of Covid-19 transmission in schools and universities. Various variations of online learning have been developed by teachers and lecturers with all their limitations in terms of ability, facilities, and infrastructure. There are two variations of online learning used, namely: 1) Asynchronous learning, which is carried out through WhatsApp group media, Google Classroom, Moodle, and other online learning applications, and 2) Synchronous learning, which uses Google Meet, Zoom Cloud Meetings, Cisco Webex, and others [1].

However, these efforts have not been enough to overcome the learning loss that has occurred nationally. The results of the Kemdikbudristek survey in the 2020/2021 Academic Year emphasize that prolonged distance learning risks increasing learning loss, especially in students' cognitive and character development [1]. In addition, not all regions in Indonesia have internet access, while learning from home requires an internet connection and good quota during the pandemic [2]. Based on data from the Indonesian Internet Service Providers Association (APJII), Indonesia's internet penetration rate has experienced significant growth in recent years, with 73.7% of the population having internet access in 2020 [3]. This translates to approximately 196.71 million internet users. Despite this progress, there remains a significant disparity in internet access across the country. While internet access is relatively widespread in Java and Sumatra, the situation in other regions, particularly Kalimantan, is less favorable. Only 7.97% of Kalimantan's territory has adequate internet connectivity, hindering access to online resources and services for a significant portion of the population. This shows that children in rural/border areas, especially in North Kalimantan who do not have an internet connection, also experience severe limitations in receiving education services during school lockdowns. In fact, a UNICEF survey shows that 70% of students are frustrated, anxious, and stressed due to school/university lockdowns which are certainly not effective if distance learning is applied continuously [2].

Therefore, in 2022, the government is targeting to reopen schools and universities in full, considering that vaccination is being intensified by the government. It was recorded that as of July 5, 2021, 32,301,268 people in Indonesia had received their first vaccination and 14,035,934 people in Indonesia had completed their second vaccination. The government is targeting a national vaccination target of 181,554,465 people in Indonesia. Vaccination is still ongoing in Indonesia to this day, although not as massive as when Covid-19 occurred [4][5]. School-age children are prioritized to participate in mass vaccination activities throughout Indonesia, including in North Kalimantan. As a result, several local governments, institutions, and universities have succeeded in reopening schools and universities that have been in lockdown for a long time. Students can return to face-to-face learning and experience face-to-face schooling, as well as use campus facilities, especially laboratory practicum which has been difficult to access due to lockdown. However, all of this is done with strict Covid-19 protocols, namely a) wearing masks; b) washing hands; c) maintaining distance; d) avoiding crowds; e) and limiting mobility and interaction) [6][7].

However, when students returned to school, some of them were seen ignoring Covid-19 protocols. This also happens to some teachers and lecturers. They do not limit interaction, do not keep their distance, still gather, and secretly take off their masks in the classroom, which puts them at risk of contracting Covid-19. This is evidenced by several schools and universities in North Kalimantan that have re-locked down after several students, students, and lecturers were exposed to Covid-19. Extra attention is needed for students, students, teachers, and lecturers to always comply with Covid-19 protocols, especially in the use of masks. Direct reprimands are very difficult to do, especially when students secretly take off their masks when the lecturer is not looking or when the lecturer is focused on delivering material. If they are caught not wearing a mask and are reprimanded, they put the mask back on, but then they take it off again shortly afterwards. This condition makes the classroom an unsafe, uncomfortable, and unhealthy place for students, because it has the potential to become a hotbed for the rapid transmission of Covid-19. Therefore, it is necessary to develop automatic mask detection technology so that students who open their masks can be immediately detected.

Among the studies conducted, one research on real-time face detection for people counting in indoor areas using CCTV cameras employed the YOLO method as a CNN architecture. This method successfully captured objects up to 10 meters away with angles of 75° and 60° at heights of 1.5 meters, 2 meters, and 2.5 meters [8]. Another study investigated mask detection using a combination of Viola-Jones and CNN (Convolutional Neural Network) methods, achieving an accuracy rate of 84.23%. However, the images used as research samples were static/still images and not real-time [9]. Additionally, research on mask detection for the visually impaired using Raspberry Pi hardware and Keras and Tensor Flow CNN algorithms for video/real-time images achieved an accuracy of 98%. Nevertheless, video streaming using Raspberry Pi in this study only obtained an FPS of 0.33, could only perform mask detection at a distance of less than 3 meters, and did not include testing based on the angle of the mask object [10].

Mask detection was successfully performed for real-time images using the Haar Cascade method (Viola-Jones architecture) [11][12] and the single shot multibox detector and mobilenetv2 algorithms [13]. The highest accuracy rates achieved in each study exceeded 88%. However, these three studies were unable to detect masks on multiple faces in the same image [13] and were not robust under low light conditions [11]. Subsequently, real-time mask detection research employed the CNN YOLOv3 algorithm and tested image capture at angles of 180°, 90°, and 45°. The highest accuracy rates were obtained at distances ranging from 2 to 5 meters, with accuracy levels between 71% and 99%. For light condition testing, accuracy rates reached 80-99% [14].

Based on the issue of mask usage in classrooms and several similar studies related to mask detectors that have been conducted, this research proposes the development of an automatic real-time multi-face mask detection system in classrooms to enforce Covid-19 protocols. The research will use CCTV cameras/similar to capture real-time images and the captured images are then processed using the deep learning method based on YOLOv4 [15][16], until the automatic detection results are obtained whether the object is using a mask or not. If not wearing a mask, the system will display a notification in the form of a sound/alarm. The detection method uses YOLOv4 which is robust to lighting conditions, object angles, and the attributes used, as well as faster detection speed and more accurate results compared to YOLOv3 [15][17].

It is expected that the proposed research design will produce effective, valid, and accurate real-time camera-based mask detection with dynamic, fast, and lightweight performance. Including the ease of which it can automatically monitor mask usage. Teachers or lecturers do not need to monitor students who are not wearing masks at all times, because the monitoring process is carried out by the proposed system. In other words, this research is also built to support the implementation of strict Covid-19 health protocols on face-to-face learning limited to classrooms.

II. RESEARCH METHODS

A. YOLOv4 (*You Only Look Once* v4)

YOLOv4 is a single-stage object detection algorithm for real-time images, representing an advanced development of YOLO [15]. Single-stage means that the neural network predicts the image bounding box into its class directly from the entire image in just one evaluation [18]. This makes the YOLO algorithm and its derivatives fast enough to detect real-time images/live video. YOLOv4 employs the CSPDarknet53 CNN backbone architecture [15] consisting of 162 layers. The input layer is a still image or a frame from a moving image that is resized according to the configuration initialization. Next, the input layer is processed into the CSPDarknet53 architecture. This architecture features additional implementations such as CSP (Cross Stage Partial), Cross Mini-Batch Normalization [19], Self Adversarial Training, FPN (Feature Pyramid Network), SPP (Spatial Pyramid Pooling), BoS (Bag of Specials), and so on, which enhance the accuracy performance and detection efficiency of YOLOv4. The Bag of Specials (BoS) approach is additionally received where 'specials' allude to systems that improve network execution while expanding the expense of observation at a lower cost [20]. Furthermore, performance enhancement is also supported by the output generated from the convolution process, which is combined with the convolution results (feature results) of the previous output layer.

B. Design and Creation of the Proposed System

The proposed system in this research aims to develop an automatic mask usage detection system in classrooms as part of enforcing Covid-19 protocol rules. To ensure the system's suitability for real-world conditions, a preliminary analysis is necessary to identify the specific tasks required to address the research problem. The preliminary analysis revealed the following needs: a) The proposed system should automatically detect mask usage in classroom learning spaces; and b) The proposed system should display notifications in the form of messages, marker images, or beep sounds if students are detected by CCTV/similar cameras as not wearing masks[21].

Based on the initial requirements, we designed the hardware system to be built as shown in Figure 1. The figure illustrates three hardware components: a camera sensor, a laptop/PC, and a tripod (serving as the camera sensor's stand). Figure 1 also depicts the estimated camera mounting height, as well as the distance, range, and angle of the camera's image capture relative to the object. Furthermore, the minimum specifications for the computer and camera are provided in Table 1.

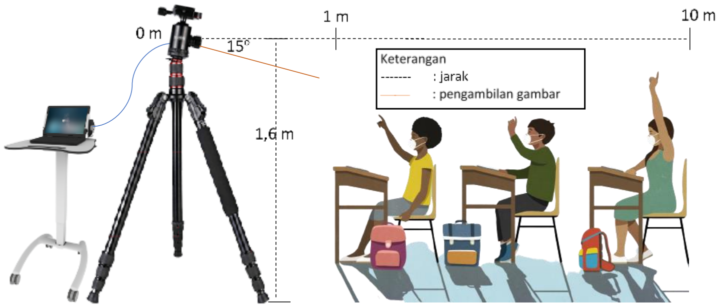

Figure 1. The Proposed Hardware Design

Tabel 1.
Minimum Computer and Camera Specifications

Component	Minimum Specification
Computer	
CPU	Core I5 8th 1,8 Ghz
VGA	NVDIA Geforce 4 GB
OS	Windows 11
CUDA	11.2
OpenCV	4.2
YOLO	Versi 4
Camera Stream Webcam or Similar	
Quad	10 MP Autofocus
Resolution	Full HD 720p/25fps

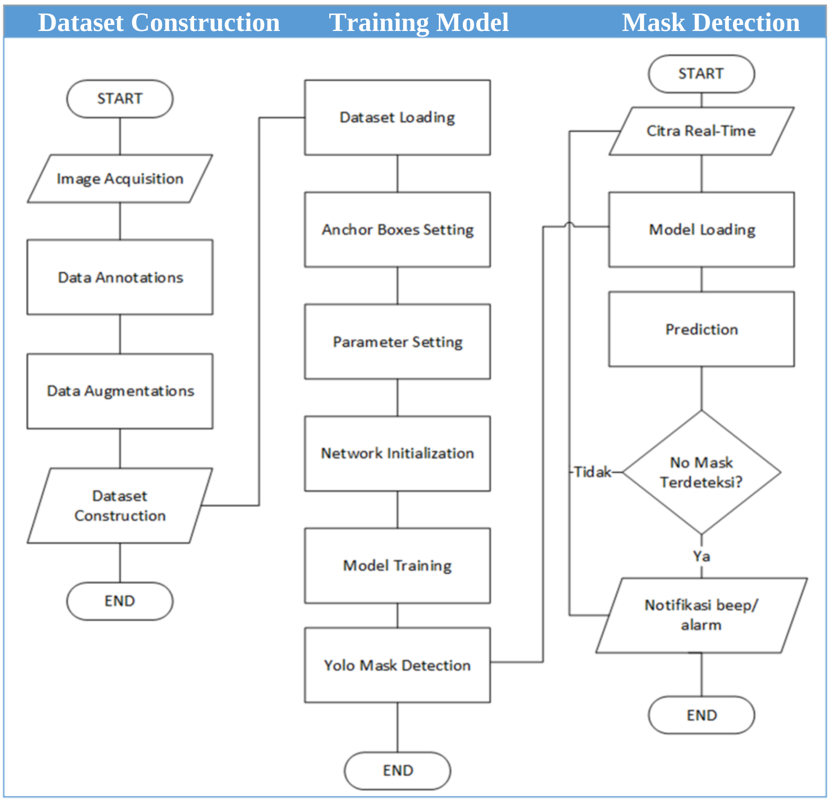

Figure 2. The Proposed Software Design

For the software design of the system, particularly in the image classification section, this research proposes the use of the YOLOv4 deep learning method, which employs the CSPDarknet53 CNN backbone architecture. The proposed system flowchart can be seen in Figure 2. The proposed system design consists of three main parts: dataset construction, model training, and mask detection.

Dataset construction is a technique for creating a ready-to-use data collection for training into a model. To create a ready-to-use training dataset, the first step is annotation, which involves labeling and categorizing images containing faces with and without masks. The labeling format used in data annotation is <size> and <object>, where the <size> format contains <width><height>, which are values based on the length and width of the image, and <object> contains <name>, which is the type of class or category of the object in the image, and <bndbox> contains <ymin><ymax><xmin><xmas>, which are the bounding box prediction values for each object detected in each image [14].

After the annotation process is complete, the next step is to introduce variations to the dataset, such as rotation, cropping, and flip, through the data augmentation process. Data augmentation is a technique for artificially increasing the size of a training dataset by creating variations of existing images. This is done to improve the generalization ability of the model, which means that the model will be better able to perform well on unseen data. The results of this process are then built into a dataset construction that is ready to be used as training data.

Next, the anchor boxes setting is performed, which will be used to detect multiple objects in the image with different sizes and with their center points located in the same cell. After that, parameter settings are performed, such as *batch*, *max_batches*, *subdivisions*, *steps*, *width*, *height*, *classes* to adjust the training scenario configuration. The *max_batches* value is the iteration limit for stopping the training process. The *batch* value is used to determine the number of images to be processed by the network weight before it changes. The smaller the *batch* value, the faster the training process, but the lower the accuracy.

Next, subdivision is responsible for processing the batch value into smaller parts, often called mini-batches. The division is done by taking the subdivision value. For example, if the batch value is 64 and the subdivision value is 16, then the training process is performed for 4 (64/16) images for each mini-batch. This process will take place 16 times until the training process on one batch is complete. Then, the system will switch to the next batch, which is also valued at 64. The classes value is the class value to be predicted, in this research it is 2 (Masked and Unmasked).

The next step is to train the configured model using the YOLOv4 method. The training process is where the neural network is trained to learn a pattern to recognize objects based on the parameter settings made in the previous steps. This training process will then produce a weight file. Then, the final step is to detect mask usage on real-time images. The testing process is done by inputting the image to be detected, then performing image detection analysis based on the weight file. At this stage, the system performs the detection and prediction process on the objects in the tested image. If the system finds a human object without a mask, a notification appears in the form of a message/alarm/sound.

C. Image Dataset Collection Technique

In this research, dataset collection is carried out in a secondary manner so that the training data used to model the weight file as an object prediction tool is very complete and rich in image libraries. In this case, the dataset that will be trained to obtain the weight file model is from Kaggle:

https://www.kaggle.com/datasets/andrewmvd/face-mask-detection.

This data is in the form of still images containing human objects wearing or not wearing masks, which have been annotated. Then, for testing data, the image data will be tested to see if there are students wearing masks and those who are not wearing masks. The source of this image testing data is moving images (videos) that are taken directly using a camera and then compared with the weight file. The location that will be used as the testing data collection room is the Department of Computer Engineering classroom, Faculty of Engineering, Universitas Borneo Tarakan.

D. Real-Time Image Testing Data Collection with Masked and Unmasked Objects

The testing data collection technique is carried out through direct observation of the real-time testing video/images that will be used. The testing sample used is real-time moving images in the classroom taken through CCTV/similar cameras. These real-time moving images will contain around 22 objects of students from the Computer Engineering Department who are randomly arranged wearing or not wearing masks. The shooting process will be set for 3 minutes and taken 20 times. This process is carried out live and in real-time, meaning that when the CCTV camera takes pictures of students, then at that time it is immediately processed until the system is able to directly detect student objects wearing masks or not wearing masks, including the system being able to activate notification messages/ alarms/sounds when there are students who are not wearing masks.

E. System Performance Evaluation Technique

The system detection performance evaluation technique is carried out on the testing data. On the testing data, evaluation is carried out by comparing the performance of the proposed system in detecting masks against the actual conditions (classification) to obtain the values of *precision*, *recall*, and *F-measure* [22] based on the identification results from: a) TP (True Positive), which is the number of real objects that are correctly detected by the system; b) FP (False Positive), which is the number of noise objects that are correctly detected by the system; and c) FN (False Negative), which is the number of real objects that are not correctly detected by the system [23].

Precision is the number of correct predictions compared to the total number of objects detected by the system, with the formula (1):

$$precision = \frac{TP}{TP + FP} \qquad (1)$$

Recall is the number of correct predictions compared to the total actual results, with the formula (2):

$$recall = \frac{TP}{TP + FN} \qquad (2)$$

F-measure is the harmonic mean of *precision* and *recall*, where the best score is close to 1 and the worst is close to 0. *F-measure* formula (3):

$$F - measure = \frac{2 \times precision \times recall}{(precision + recall)} \qquad (3)$$

In this research, *F-measure* is used to assess the classification of image objects of students wearing masks (WM) and those not wearing masks (NWM), rather than accuracy. This is because accuracy is well-suited for use when the distribution of positive and negative classifications is balanced (the number of masked and unmasked objects is balanced), while *F-measure* is better suited for use when the distribution of masked and unmasked object classifications is unbalanced and random, and can change over time.

III. RESULTS AND DISCUSSION

Figure 3 shows the implementation of the hardware design for the proposed system. The camera sensor is mounted on a tripod and placed on a table with a sensor height of

1.6 m. The camera sensor is then connected to a laptop/PC using a USB type C cable and is directed at a specific angle to capture images of students in front of it. Figure 3 shows that the system successfully captures image objects of students wearing and not wearing masks in the classroom of the Department of Computer Engineering, Faculty of Engineering, Universitas Borneo Tarakan.

Figure 3 Hardware Design Implementation for the Proposed System

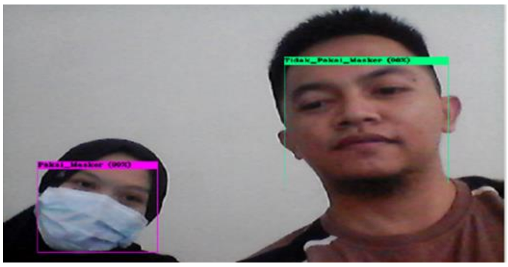

Figure 4 The Initial System Testing with Mask Detection

Figure 5 The proposed system's testing successfully detected WM (wearing masks) and NWM (not wearing masks/unmasked) student objects in the classroom

The testing outcomes of the proposed system are presented in Figures 4 and 5. Figure 4 showcases the system's effectiveness in identifying mask usage on two image objects. It is observed that for two closely positioned image objects, the proposed system accurately captures and classifies mask usage on the image objects. Additionally, Figure 5 highlights the system's ability to detect mask usage on image objects with multiple faces (22 student samples). The system successfully identified the majority of student objects wearing masks and conversely, successfully identified the majority of student objects not wearing masks. Nonetheless, there were still cases of object misidentification, where the system incorrectly classified objects as wearing or not wearing masks.

Table 2.
Evaluation Outcomes Based on *Precision*, *Recall*, and *F-measure*

Scenario	***Precision* (%)**		***Recall* (%)**		***F-measure***		**Message Notification/ Alarm NWM**
	WM	**NWM**	**WM**	**NWM**	**WM**	**NWM**	
1	90,91	100,00	62,50	50,00	74,07	66,67	ON
2	100,00	100,00	53,85	60,00	70,00	75,00	ON
3	90,00	100,00	60,00	66,67	72,00	80,00	ON
4	100,00	100,00	63,64	57,14	77,78	72,73	ON
5	100,00	100,00	46,15	62,50	63,16	76,92	ON
6	90,91	100,00	62,50	50,00	74,07	66,67	ON
7	100,00	100,00	64,29	75,00	78,26	85,71	ON
8	91,67	100,00	61,11	40,00	73,33	57,14	ON
9	100,00	100,00	57,14	57,14	72,73	72,73	ON
10	100,00	100,00	69,23	62,50	81,82	76,92	ON
11	100,00	80,00	44,44	57,14	61,54	66,67	ON
12	72,73	100,00	61,54	55,56	66,67	71,43	ON
13	88,89	83,33	66,67	62,50	76,19	71,43	ON
14	87,50	83,33	63,64	62,50	73,68	71,43	ON
15	100,00	100,00	72,73	55,56	84,21	71,43	ON
16	100,00	100,00	69,23	55,56	81,82	71,43	ON
17	100,00	100,00	71,43	71,43	83,33	83,33	ON
18	100,00	100,00	66,67	66,67	80,00	80,00	ON
19	100,00	100,00	57,14	70,00	72,73	82,35	ON
20	100,00	100,00	58,33	66,67	73,68	80,00	ON
Mean	**95,63**	**97,33**	**61,61**	**60,23**	**74,55**	**74,00**	**(20/20 *100) = 100,00**

To validate the proposed system's ability to accurately capture and classify images of students wearing masks (WM) or not (NWM) in real-time within the classroom, the *precision*, *recall*, and *F-measure* values were calculated. Table 2 presents the testing results of the proposed system based on *precision*, *recall*, and *F-measure*. According to the *precision* calculations, the proposed system achieved an average *precision* score of 95.63% for WM and 97.33% for NWM. The highest *precision* scores for WM and NWM were 100%, while the lowest *precision* scores were 72.73% for WM and 80% for NWM. The *precision* results indicate that the proposed system, which utilizes deep learning CNN based on YOLOv4, has a prediction accuracy close to perfect for the student image objects successfully detected by the system. This is evidenced by the fact that most scenarios, both WM and NWM, have the highest *precision* score of 100%.

This signifies that in most scenarios, when the proposed system accurately detects student image objects, it can also effectively differentiate between student objects wearing masks (WM) and those not wearing masks (NWM) from the successfully detected objects. However, there remain *precision* scores falling below <100% in certain scenarios. This implies that when the system successfully detects WM objects, there are instances where NWM objects are misclassified as WM by the system, and vice versa. This suggests that there are cases where the system successfully detects the object but incorrectly determines whether it falls into the WM or NWM classification. Detection errors can arise due to several factors, including: a) The mask's color closely resembles the student's skin tone or the color of the student's hijab, leading to the misclassification of the student object as NWM; b) The NWM student's facial area is obscured by hands or other objects during camera sensor capture, causing the system to misidentify the student object as WM; and c) Students cover their faces with hijabs without wearing masks, sometimes leading to misclassification as WM objects by the system.

The *recall* calculations reveal that the proposed system achieved an average *recall* score of 61.61% for WM and 60.23% for NWM. The highest *recall* scores for WM and NWM were 72.73% and 75%, respectively, while the lowest *recall* scores were 44.44% for WM and 40% for NWM. This signifies that the proposed system utilizing YOLOv4 deep learning is reasonably capable of accurately detecting WM and NWM student image objects from the entire dataset. However, it also indicates that there are WM and NWM student image objects that were not successfully captured by the system. The system may fail to capture objects due to several factors, including: a) The student image objects are situated at the maximum distance or beyond the camera's range; b) Both WM and NWM student objects are in a lowered and sideways position, as illustrated in Figure 6.

Figure 6 Evaluation Outcomes of the Proposed System for Scenario-13

Figure 6 illustrates that some NWM and WM students cannot be detected by the system due to their images being captured while they are lowering their heads or facing sideways. Additionally, students sitting at the back of the classroom cannot be detected or misidentified by the system due to their significant distance from the camera sensor position and the insufficient lighting, which causes color blurring in the facial area below the nose, leading the system to assume that these students are wearing masks when they are not.

Based on the *F-measure* calculations, the proposed system achieved an average *F-measure* score of 74.55% for WM and 74.00% for NWM. The highest *F-measure* scores for WM and NWM were 84.21% and 85.71%, respectively, while the lowest *F-measure* scores were 61.54% for WM and 57.14% for NWM. The average success rate of message/alarm/sound notifications when NWM students were successfully detected by the system across all scenarios was 100%. This demonstrates that the system successfully classified or distinguished between real-time image objects or live video of WM and NWM students due to the average *F-measure* score > 70% approaching 1 (100%) and the sound notification/message display becoming ON (active) when there are students who are not wearing masks.

These findings demonstrate that the successfully designed proposed system can effectively, validly, and accurately detect whether students are wearing masks with dynamic performance and lightweight resources. It facilitates automated mask usage monitoring in classrooms. Teachers and lecturers are relieved from the constant obligation of monitoring and reminding students who fail to wear masks, as the proposed system handles this process automatically. The testing results substantiate the system's readiness to support the implementation of strict Covid-19 health protocols in limited face-to-face learning settings in classrooms. Automated mask detection through CCTV cameras will undoubtedly contribute to enforcing Covid-19 protocols more objectively. The proposed system automatically detects such violations and triggers message/alarm/sound notifications, eliminating the need for educators to constantly monitor and remind negligent students who remove their masks in class.

IV. CONCLUSIONS

This study proposes the development of an automatic multiface mask detection system in classrooms for Covid-19 protocol enforcement. The proposed system utilizes a CCTV camera or similar device to capture real-time images, which are then processed using YOLOv4 deep learning to identify whether the image objects are wearing masks or not. If a mask is not detected, the system will trigger a message/sound/alarm notification. Experimental results on 22 samples of students who collected real-time/live video data every 3 minutes for 20 scenarios proved that the proposed system was successful in detecting objects wearing masks (WM) and not wearing masks (NWM) with the average percentage of precision was 95.63% for WM and 97.33% for NWM, average recall scores of 61.61% for WM and 60.23% for NWM, and average F-measure scores of 74.55% for WM and 74.00% for NWM. These findings indicate that the successfully designed proposed system can effectively, validly, and accurately detect whether students are wearing masks with dynamic performance and lightweight resources. This includes the ease of automatically monitoring mask usage in classrooms.

ACKNOWLEDGEMENTS

The author would like to thank LPPM UBT for providing research grant funding under the RKD (Riset Kompetensi Dasar) scheme from DIPA funds for the year 2022-2023.

REFERENCES

[1] A. Wahyudi, “Learning Loss During Covid-19 Pandemic in Indonesia and The Strategies to Minimize it,” *J. English Educ. Linguist.*, vol. 2, no. 2, hal. 18–25, 2021.

[2] C. F. Pasani dan R. Amelia, “Learning Loss and Education Inequality in Indonesia (Mapping the Potential , Consequences , and the Covid-19 Crisis),” *Review of International Geographical Education*, vol. 11, no. 10. hal. 1171–1181, 2021.

[3] S. I. Mawarni dan W. Syafitri, “DIGITALISASI DAN PENGANGGURAN DI PEDESAAN JAWA TIMUR PADA MASA PANDEMI COVID-19 (STUDI KASUS PADA SEKTOR INFORMAL),” *Contemp. Stud. Econ. Financ. Bank.*, vol. 2, no. 1, hal. 156–167, 2023.

[4] T. Tamara, “Gambaran Vaksinasi COVID-19 di Indonesia pada Juli 2021,” *Medula*, vol. 11, no. 1. hal. 180–183, 2021.

[5] P. Maria, F. Akbar, A. Ullah, dan H. Zarory, “Kunci Kelas Terjadwal Teknik Elektro UIN SUSKA Riau Berbasis RFID dan IoT,” *J. Sist. Cerdas*, vol. 6, no. 1, hal. 19–28, Apr 2023, doi: 10.37396/jsc.v6i1.282.

[6] I. S. Joyosemito dan N. M. Nasir, “Gelombang Kedua Pandemi Menuju Endemi Covid-19: Analisis Kebijakan Vaksinasi Dan Pembatasan Kegiatan Masyarakat Di Indonesia,” *Jurnal Sains Teknologi dalam Pemberdayaan Masyarakat*, vol. 2, no. 1. hal. 55–66, 2021, doi: 10.31599/jstpm.v2i1.718.

[7] I. B. M. Gandamayu, N. W. S. Antari, dan I. A. S. Strisanti, “The level of community compliance in implementing health protocols to prevent the spread of

COVID-19," *Int. J. Heal. Med. Sci.*, vol. 5, no. 2, hal. 177–182, 2022.
[8] Y. Adinata, K. Gunadi, dan I. Sugiarto, "Aplikasi Deteksi Jumlah Orang pada Area Indoor Untuk Mendukung Pelaksanaan PPKM dengan Metode YOLO," *J. Infra*, vol. 10, no. 1, hal. 135–141, 2022.
[9] F. A. Hermawati dan R. A. Zai, "Sistem Deteksi Pemakaian Masker Menggunakan Metode Viola-Jones dan Convolutional Neural Networks (CNN)," in *Konferensi Nasional Ilmu Komputer (KONIK) 2021*, 2021, hal. 182–187.
[10] M. Aminullah, "Alat Deteksi Masker Dengan Metode Convolutional Neural Network Untuk Tunanetra Pada Era New Normal," in *Seminar Nasional Fortei Regional 7*, 2020, hal. 321–324.
[11] F. L. Ahmad, A. Nugroho, dan A. F. Suni, "Deteksi Pemakai Masker Menggunakan Metode Haar Cascade Sebagai Pencegahaan COVID 19," *Edu Elektr. J.*, vol. 10, no. 1, hal. 13–18, 2021.
[12] G. Aprilian Anarki, K. Auliasari, dan M. Orisa, "PENERAPAN METODE HAAR CASCADE PADA APLIKASI DETEKSI MASKER," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, hal. 179–186, 2021, doi: 10.36040/jati.v5i1.3214.
[13] R. R. Ramdhani, R. I. Adam, dan A. A. Ridha, "Implementasi Deep Learning untuk Deteksi Masker," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2, hal. 384–391, 2021.
[14] L. Agustien, T. Rahman, dan A. W. Hujairi, "Real-time Deteksi Masker Berbasis Deep Learning menggunakan Algoritma CNN YOLOv3," *J. Teknol. Inf. dan Terap.*, vol. 8, no. 2, hal. 129–137, 2021, doi: https://doi.org/10/25047/jtit.v8i2.246.
[15] A. Bochkovskiy, C.-Y. Wang, dan H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv Prepr. arXiv2004.10934*, hal. 1–17, 2020, doi: 10.48550/arXiv.2004.10934.
[16] T. M. Đức dan N. T. Hương, "Improvement of Vehicle detection and classification performance with Region of Interest," *ITEJ (Information Technol. Eng. Journals)*, vol. 8, no. 1, hal. 34–41, Jul 2023, doi: 10.24235/itej.v8i1.116.
[17] C.-Y. Wang, A. Bochkovskiy, dan H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," in *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 2021, hal. 13029–13038, doi: 10.48550/arXiv.2011.08036.
[18] J. Redmon, S. Divvala, R. Girshick, dan A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection."
[19] Z. Yao, Y. Cao, S. Zheng, G. Huang, dan S. Lin, "Cross-Iteration Batch Normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, hal. 12331–12340.
[20] S. Degadwala, D. Vyas, U. Chakraborty, A. R. Dider, dan H. Biswas, "Yolo-v4 Deep Learning Model for Medical Face Mask Detection," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, hal. 209–213, doi: 10.1109/ICAIS50930.2021.9395857.
[21] I. A. Dahlan, D. Ariateja, M. A. Arghanie, M. A. Versantariqh, M. David, dan U. D. Fatmawati, "Sistem Deteksi Senjata Otomatis Menggunakan Deep Learning Berbasis CCTV Cerdas," *J. Sist. Cerdas*, vol. 4, no. 2, hal. 126–141, Agu 2021, doi: 10.37396/jsc.v4i2.172.
[22] T. J. Lakshmi, N. M. R. Karri, N. B. K. Y. Teluguntla, J. Vallabhaneni, dan G. K. Nanduri, "Facemask Detection Using Machine Learning Techniques: A Review," *Webology (ISSN: 1735-188X)*, vol. 19, no. 1, 2022.
[23] D. I. Mulyana, A. C. Aditya, S. Amelia, dan S. Agustiansyah, "Implementasi Framework Mask R-CNN Object Detection API Dalam Mengklasifikasi Jenis Kendaraan Bermotor," *J. Tek. Elektro dan Komputasi*, vol. 4, no. 1, hal. 63–72, 2022, doi: 10.32528/elkom.v4i1.7148.