Vol. 4 Special Issue. 1
International Journal of Data Science and Advanced Analytics (ISSN: 2563-4429)

Social Distance Monitoring and Face Mask Detection System

Chin Xin Yee [1], Amad Arshad [2] and Hairul Aysa Binti Abdul Halim Sithiq [3]
[1] Asia Pacific University of Technology and Innovation
[2] Asia Pacific University of Technology and Innovation
[3] Asia Pacific University of Technology and Innovation
[1]cindychin0724@gmail.com, [2] amad.arshad@staffemail.apu.edu.my, [3] hairul-aysa@staffemail.apu.edu.my

***Abstract*—The Covid-19 pandemic has caused a devastating effect toward society, economy and even education system. To prevent the Covid-19 outspread, lock-down seems to be one of the most popular solutions being carried out by governments including Malaysia. In the long run, schools will be reopening anytime when the pandemic is under control. However, the attention towards control measures in school should be paid to prevent the pandemic resurgence. Manual monitoring of the SOP is impractical towards a large population with only limited resources and task forces. To mitigate the issues, the social distancing monitoring and face mask detection system – EyeSpy has been proposed. The proposed system aims to assist teachers in monitoring SOP compliances performed by students in school, thus lowering the risk for Covid-19 outspread. This paper also provides a comparative study of different face detections, CNN architecture on facemask classifiers. Apart from that, this research employed Rapid Application Development as the system development methodology for the development of EyeSpy. Other than that, interview and questionnaire are used as requirement gathering methods to study the feasibility of the project. A system has been developed to act as a proof of concept to underline the implementation of EyeSpy mobile application in restraining the spreading of Covid-19 in schools.**

***Index Terms*— Covid-19, Facemask Classifier, Social Distancing**

1. Project Background

Malaysia has declared multiple lockdown orders to mitigate Covid-19 transmission. All institutions including government primary and secondary schools were shut down during most of the lockdown orders. Even though moving everything online did reduce the number of infections but some researchers have different opinions on this policy. According to the research from reference [9], authors have argued that the prompted of strict-at-home orders will greatly impact people's livelihoods, and when economic lockdown over time, individuals will be started to weigh between the risk of infection of going out to work and the economic cost or burden of staying at home. Eventually, people will all go out for work, the school will be set to reopen, factory is allowed to resume operation even the pandemic is not over. However, wearing the facemask anytime to prevent the transmission of the virus through the air and maintaining appropriate distance with others to reduce close contact will be the only way to combat this pandemic in any situation.

When the time for students to attend physical class has come, manual monitoring will not be effective for limited number of teachers to manage large population of students moving around. The monitoring of SOP is even less effective with possibility of human errors and teacher burnout due to overburden workload including teaching and monitoring student SOP at the same time. Reference [19] reflects that there are challenges found during observing SOP compliance management within the context of secondary school teachers. Two of the SOPs that will be focused on this project will be face masks and social distancing. There are various challenges faced by teachers, such as students' attitude, breaktime control, overburden of teacher's duty on SOP and students' habit to accompany each other. First of all, it will also become a burden for teachers that needed to observe the SOP compliance by the student. Mainly because the teacher will be the one that meets pupils frequently, eventually all the responsibility has fallen under teachers to advise the students to adhere to social distancing and wear the face mask anytime. Most of the teachers have mentioned that the workload has been increased where they needed to oversee the entrance throughout the day of teaching. In addition, students' attitude that does not want to obey the SOP was also a problem faced by teachers. Some students always ignore the SOP during breaktime, sit closely and share drinks with friends. Furthermore, teachers do mention that students will be more difficult to control when they are outside of the classroom, some of them are not wearing mask. By having all the situation above, the rise in number for students tested covid positive is expected. Hence, the researcher has realized the need of an automated system that helps to monitor student's SOP in school. If student behavior is not monitored in school, then school will not be a safe and good environment anymore for students to study. This project aims to use computer vision to assist schools and teacher in enforcing SOP compliance, which can directly or indirectly help in better decision making. The information will be from an online questionnaire and survey.

2. Domain Research

There are vast majority of publications regarding facemask detection relies on the clear input from face detection. While in this paper, researcher would like to emphasis on the combination of social distancing and facemask detection system which would brough difficulty for detection in terms of complex environment and background. There are many researchers have proposed many different techniques in the field of facemask detection and social distancing detection. Each of the techniques will have its own advantages and disadvantages, it is crucial to understand the options that available in the market and choose the techniques wisely to make this project achieve its target goal.

2.1. Face Detection System

When comes to the focus of this project, the initial requirement to perform facemask detection is to get facial detail as an input. The requirement for the face classifier of this project will be complex since it required the face classifier to be able to deal with different facial orientation or under complex background. In reference [15], researchers have taken multiple faces detection and different face orientations into consideration. Researchers have adopted pre-defined training weights of VGG-16 architecture for feature extraction and prediction. The final propose model have consisted of 17 convolutional layer and 5 max pooling layers to ensure that the size of feature map is halved to reduce the parameter, and achieved 93% accuracy to detect faces in multiple angles. Other than having transfer learning as the method, reference [11] have proposed two models for extracting features from non-frontal faces of 30º to 90º. Active Appearance Model (AAM) is utilized to interpret faces. Then by training the model to fit the new images and improve the efficiency of fitting by applying nonlinear parameter update method which could be able to obtain the appearance parameter for face feature vector.

2.2. Facemask detection system

Transfer learning is being employed by many researchers in training facemask classifier. In reference [16], researchers have proposed a transfer learning method adopting ResNet50 for facemask detection in a classroom. Researcher added an average pooling layer followed by a fully connected layer containing 1000 nodes and end with a SoftMax function. The other group of researchers from reference [5], have proposed in the used of MobileNetV2, researchers were using Data Augmentation techniques to create a new set of training data to improve the accuracy of model. From the study of reference [14], researchers have used methodology based on transfer learning and adopted Simulated Masked Face Dataset (SMFD), researcher have fine-tuned the last layer of InceptionV3 by removing it and adding 5 layers to the network. Other focus of this project will be improving the model capability to deal with the complex dataset. While reference [1] have proposed another system on face mask detection system aimed at tackling real-world complex problems. Researchers have comprised of eight object detection models and specifically fine-tune them for face mask detection. The models which include MaskRCNN, YOLO-v3 (Darknet-53), YOLO-v3-tiny, YOLOv3-SPP (Spatial Pyramid Pooling), RetinaNet, Faster-RCNN, YOLO-v5x, and YOLO-v5x6-TTA (Test Time Augmentation).

2.3. Social Distancing Detection System

According to reference [4] proposed system, researchers have used Euclidean distance to directly compute the distance between two individuals without performing any view calibration. Their work did not take the distance between pedestrian and camera into account. Other than directly computing with centroid distance using Euclidean distance formula, DBSCAN algorithm is another famous approach when it came to social distancing detection. Based on a set of centroids, DBSCAN groups together those centroids that are closed to each other as one cluster based on distance measurement. The DBSCAN algorithm is usually used in segmenting an object from an image, when it came to calculating distance based on pixel distance in an image, the result is notas good as expected, since the DBSCAN algorithm does not consider the camera angle and it is clustered based on the 2-D pixel information. Hence, an algorithm that put camera calibration into consideration, is aimed to determine the specific camera parameter, which is a crucial step in computer vision application specifically when the metric information of the pixel is required to increase the accuracy of measurement [8].

3. System Development Methodology

3.1 Rapid Application Development

System development methodology chosen to guide the process of the development of the project is Rapid Application Development (RAD). Unlike other software development, the first phase of the RAD model asked for a broad requirement. Developers, team members, and the client will discuss the goal and expectation for this project [17]. Then the second phase, the project will start with the User Design phase. By taking user feedbacks and then building several prototypes of the android mobile application to mimic the final version of the product and collect user feedbacks for further improvement. Then in the third phase - Rapid Construction Phase will be focused on refining all the previous prototypes into a working model. After creating several prototypes for client to review, the developer then started to leverage multiple AWS services as backend, facemask detection model and social distancing detection model. Lastly, phase 4 is Cutover, the last phase will be including the finalization of features, functions, interface, or everything related to the project. The interface between independent module needs to be tested properly, this should be done in the cutover phase. Hence, the final product with all the changes being updated is going to launch at this phase.

3.2 Unit Testing

Unit testing is being carried out on the basic components of the system to ensure all the components are working as intended [13]. Unit testing of this project consist of two main components which encompasses model performance testing and unit testing of EyeSpy mobile application. The model performance testing will be focused on checking if the machine learning models are performed accurately, and one of the most important tests that was included is invariance test. Invariance test is which the test used to check how much the model could handle the variance of data without affecting the accuracy of the model [13], in this case will be testing on how well the model could handle complex and extreme data. On the other hand, the mobile application will undergo testing of each unit that make up the whole application. The unit test on mobile application is to validate that each of the components in the software code performs as expected. For example, the sign-up page unit testing will make sure the user entered correct information before signing up an account.

3.3 User Acceptance Testing

User Acceptance Testing will usually be done on the last stage of the software development life cycle and is conducted by end-users of the products to verify the product. This project was constructed during the time of pandemic which make it impossible to meet the end-users physically. Hence the user acceptance testing was conducted via Google Meet. Due to some hardware limitations, the connection to a real-time CCTV stream is not able to carry out in this project. Hence, the component of getting video frame from real-time CCTV stream will be replace with running different CCTV footage on different angles in Amazon Sagemaker and showing the result of detection in the EyeSpy mobile application to the end-user. All of the end-users are satisfied with the product's result and some feedbacks is given on the mobile application itself such as the font size could be larger by considering there are some aged teachers teaching in school.

4. Implementation

4.1 AWS Architecture

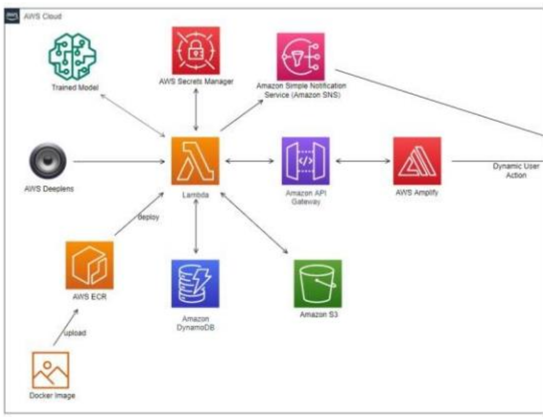

Figure 1. EyeSpy AWS Architecture

The Amazon AWS Services that are used in the EyeSpy application would be AWS Deeplens, AWS ECR, AWS Sagemaker, AWS lambda, Amazon API Gateway, Amazon S3, Amazon DynamoDB, AWS Secrets Manager, Amazon Simple Notification, AWS Amplify. AWS Deeplens is proposed to act as the CCTV in this system. The AWS Deeplens will be placed approximately 2-meter height and the angle of CCTV should be approximately 15^{o} which considered as the maximum angle to capture facial details [10]. Teachers as the user will be able to use mobile application to view the people that violating SOP which has been detected by the system. The logic of detection and reporting is being done through the AWS backend services. The docker image that containing the script for detection will be upload to AWS ECR and run the detection through AWS Lambda Container Image. The pretrained model such as facemask classification model will be deployed in AWS Sagemaker endpoint and ready to be invoked by lambda function. When a person is detected for not wearing facemask or not keeping a safe distance with others, Lambda will perform to save the image and details inside Amazon S3 and DynamoDB respectively. When a picture being save to Amazon S3, the new image entry will trigger Lambda function to notify all users through Amazon SNS.

On the user side, user will be able to access the information in the mobile app such as evidence that students violating SOP compliance that stored in S3 and DynamoDB. Other than that, user will be able to save those evidence and view it anytime. The logic is being done through Amazon Amplify which is a framework to build fullstack IOS, android, web or react native application. By setting up Amplify API in android studio, it will be able to make request to the Amazon API gateway and invoke the lambda function that programmed in a way that return what the user request for.

4.2 Model Development

The Social Distancing Monitoring and Facemask detection system is formed by having 4 computer vision models, which are face detection, facemask classifier, social distancing detection and tracking algorithm.

4.2.1 Face Detection

Due to the requirement for analyzing social distancing might need some distances between person and the camera, it makes face detection and facemask detection a problem since the face will be small and blur in the frame. Then the developers have tested multiple famous face classifiers available and found that DSFD have outshined many face classifiers including Haar classifier, DNN and MTCNN under various extreme condition which include dim lighting images, crowd images, blurry faces, different face orientations and etc.
According to the observation, the Haar classifier has shown to have the lowest accuracy, Haar are not able to detect frontal face and profile face together in a single system, thus as expected the classifier did not do well on faces in different orientations which makes it comparatively limited [6]. Hence, Haar becomes the last choice when comes to face detection in this proposed system. DNN face detector are the algorithm provided in OpenCV deep neural network module, it is a caffe model which is built based on SSD and uses ResNet-10 as the architecture backbone [3], it works better than Haar algorithm, it can give an optimum detection when the image lighting situation is good and less people, when the environment is dim or crowded, the accuracy will decrease drastically. One of the biggest issues for DNN classifier, it could not detect any of the faces in crowd.

Figure 2. Output of DSFD in crowded situation

However, by having a strong face detection is still not enough to apply in video surveillance devices. Hence developer has added some process in Figure 2. The process will start with people detection and performed face detection on the detected people area, which could ease the process to find 'faces' and save computation powers since the classifier will not require to detect every single pixel in the frame.

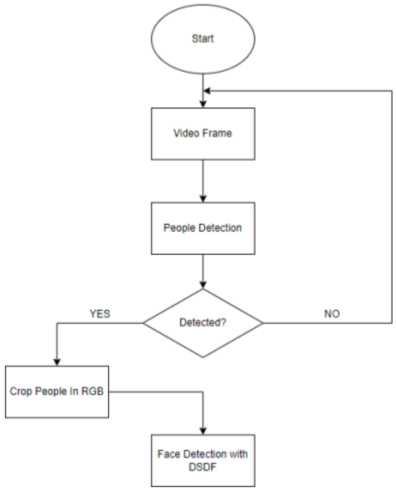

Figure 3. Process of face detection

4.2.2 Facemask Classifier

Facemask classifier in this system will be required to adapt to the quality faces, blurry faces, faces in different orientations, low resolution faces, changes in illumination direction and etc. Facemask classifier models in Machine Learning and Deep Learning have achieved promising result in those use case of facemask detector in shopping centre that wait for the person well-centred their face and started the detection. However, the existing methods need to be more robust in complex environment. In order to fulfil the requirement of the system that can classify the face under extreme conditions, the dataset being used for facemask classifier training is crucial, the dataset should be complex enough to deal with the problem mentioned. For this purpose, the developers prepared a large-scale dataset that consists of two classes, i.e., 'mask' and 'nomask'. The dataset consists of around 44000 face images of people. The dataset is collected from multiple sources, which are MAFA public dataset, Thang Pham face classifier dataset from Github, and the developer's self-collected face images from social media. Three different sources of the dataset are shown in Figure 3, it is expected to have different orientations of faces being detected, hence the dataset prepared encompasses faces with various orientations and different types of complexity as well.

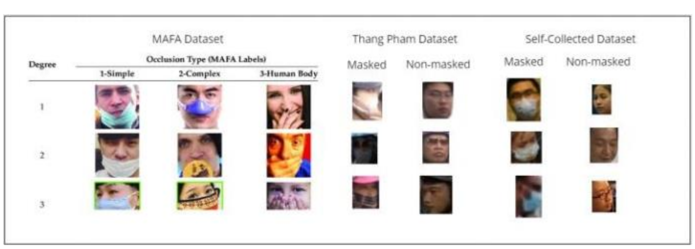

Figure 4. Dataset from different sources

Developer has developed several different CNN architectures to explore the possibility of these model to plug-in with the proposed architecture so that highly accurate result can be achieved parallel with social distancing detection. The experiment is conducted with three popular baseline models which are ResNet50, ResNet50V2, Inception Resnet, and MobileNetV2. Before the training starts, the developer first initializes the network with ImageNet weights, then followed by replacing the last SoftMax activation function with the sigmoid activation function. Loss function is also being changed to Binary Crossentropy loss. All the networks were trained without freezing any layers. During the training progress, Stochastic Gradient Descent is used with a learning rate of 0.0001. Other than that data augmentation techniques were implemented to form new and different examples to train the datasets. It is observed that the proposed technique achieves high accuracy (98.98%) when implemented ResNet50V2. The outstanding performance of the proposed model is highly suitable for video surveillance devices.

4.2.3 Social Distancing Detection

YOLOv3 is being employed in this project for people detection due to its ability of fast detection with high accuracy. After detection, the centroid information of individuals is used to compute the horizontal distance and vertical distance between each other. After getting the horizontal and vertical distance, the individual's height was compared, and identified which individual is nearer to the camera, the width and height of the individual is being calibrated. After all, the camera factor is added to the calibration process. Distance between all individuals is compared using horizontal distance and vertical distance that are being calibrated at first. The developer has visualized how horizontal distance and vertical distance are calculated and used in indicating if the individuals are close to each other in Figure 5.

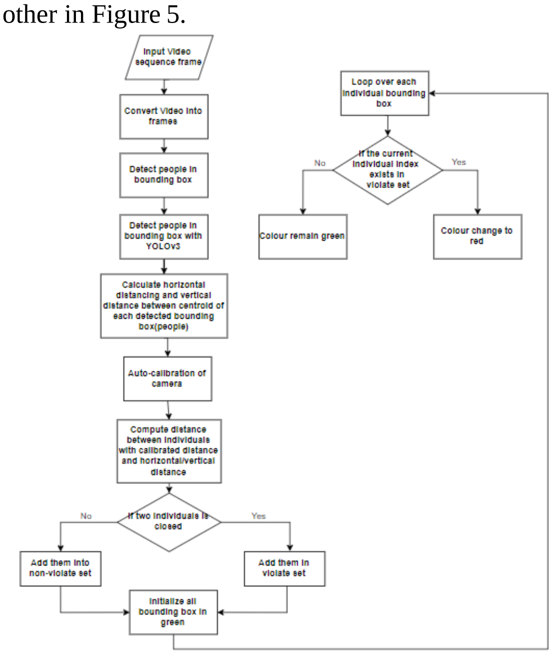

Figure 5. Process of social distancing detection

4.2.4 Tracking Algorithm

A simple tracking algorithm is being implemented in this project. The frame that extracted from video are one of many still images, it was composed to form the complete video. Each of the frames is independent on each other, hence the tracking algorithm is required to help in analysing the sequence of video frames and stitch the person's previous location with the current location [7]. The algorithm works by first getting all detected people location in the initial frames and assigned a unique ID to it. In the following frames, comparison will be made between the person in the previous location and location in current frames to make sure the objects are belong to the same objects. In this case, the closer the location is, the greater the probability for tracking the same object. Students that are violating the SOP will only be captured once by the system using the tracking algorithm. This is to prevent the duplication of capturing the same person which is considered as repeated information. A CCTV has a frame rate of 25 FPS. If an unmasked student appears on the CCTV for over five minutes, this equals to 7500 frames captured (25 fps × 60 seconds × 5 minutes = 7500). It is ineffective and a waste of resources to capture the same person over 7000 times. Therefore, the tracking algorithm is needed, if the same person appears to violate the SOP for over 50 frames, they will be captured and uploaded to the mobile application. 50 frames are chosen in this case to ensure that the system correctly classifies students when they violate the SOP.

5. Evidence of outcome

5.1 Mobile application Snapshot

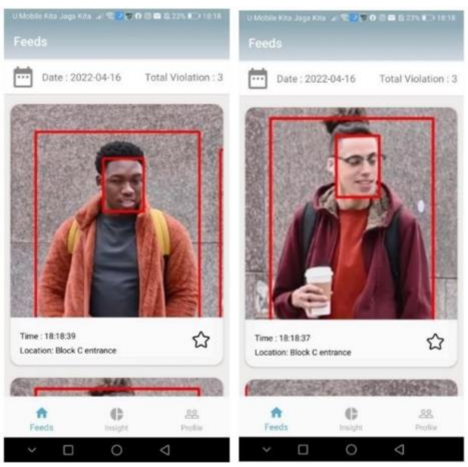

Figure 6. Feeds page that recorded students violate SOP

Figure 7. Analysis page that shows the insight of the violation cases

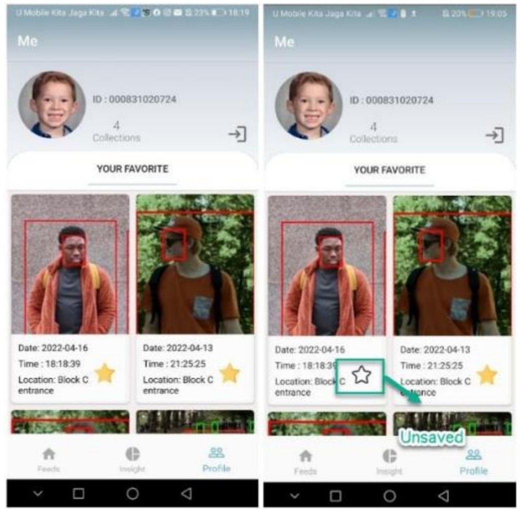

Figure 8. Profile page that shows saved records

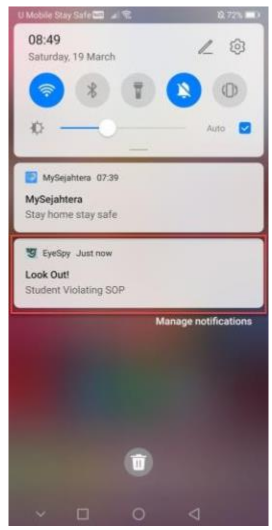

Figure 9. Notification of violation cases

5.2 Facemask classifier output

Figure 10. Facemask classifier output on extreme dataset

The example testing outputs given in Figure 10 have shown the ability and advantages of the model, the labels 'Non-masked' and 'Masked' are the result of classification using the trained model, it is capable to deal with the extreme scenario where the faces might be blurred, motion-blurred, dim, small in size, or different orientations.

5.3 Video Detection output.

The example testing outputs is given in this link.

6. Future Enhancement

One of the future enhancements for the proposed system would be to connect with local authorities and make the real-time CCTV connection possible. Due to the on-going pandemic and hardware limitation, the connection to real-time CCTV stream is not able to carry out in this project. This could help developers some spaces to finetune the detection model in real-time environment. Furthermore, instead of just knowing where and when those violation cases occur, the system could then step into identifying accurately who are those persons that are violating rule through recognition features. By having full grasp of people who violated SOP, the analysis page for the mobile application could be more advanced and in-depth to checking each student's violation records. Lastly, some improvement will required to be done in tracking algorithm part such as exploration on different advance tracking algorithm could have more robust performance in dealing with occlusion, multiple spatial scales and background distractions.

7. Conclusion

In order to complete this research paper, the researcher invest many time in doing findings and underwent various research studies to gather enough knowledge and information for the development of deep learning model and mobile application in this project. This project aimed to develop a system that utilized Deep Learning and could be used in detecting students that did not wear their facemask or maintain a safe distance from others, which can lessen the teacher's burden to make the teacher more focused on their teaching and at the same time have finger on pulse know the information about students that attempt to break the SOP. Facemask classification and social distancing are two major ways that should be practiced to reduce the spread of Covid-19. One of the major strengths of this system is it does not divide the facemask detection and social distancing into 2 products that are popularly been used in most of the existing systems in the market. Most of the existing system does not combine the two distinct features into a single detection system is due to the requirement for analysing social distancing might need some distances between person and camera. When there is a certain distance between the person itself with CCTV, it makes facemask detection a problem since the face will be small and blur in the frame. However, in this system, developer has successfully used her own architecture on the face detection algorithm and trained a facemask detection model that is capable capable to deal with small and complex face datasets and its accuracy has reached around 98% on the validation dataset.

References

[1] Adhikarla, E., & Davison, B. D. (2021). Face Mask Detection on Real-World Webcam Images. *Proceedings of the Conference on Information Technology for Social Good*. https://doi.org/10.1145/3462203.3475903

[2] Adhinata, F. D., Rakhmadani, D. P., & Segara, A. J. T. (2021). YOLO Algorithm for Detecting People in Social Distancing System. *Jurnal Transformatika*, *19*(1),1. https://doi.org/10.26623/transformatika.v19i1.3582

[3] Agarwal, V. (2020). *Face Detection Models: Which to Use and Why?* Medium. https://towardsdatascience.com/face-detection-models-which-to-use-and-why-d263e82c302c

[4] Ahmed, I., Ahmad, M., Rodrigues, J. J. P. C., Jeon, G., & Din, S. (2020). A deep learning-based social distance monitoring framework for COVID-19. *Sustainable Cities and Society*, 102571. https://doi.org/10.1016/j.scs.2020.102571

[5] Arora, M., Garg, S., & A., S. (2021). Face Mask Detection System using Mobilenetv2. *International Journal of Engineering and Advanced Technology*, *10*(4),127–129. https://doi.org/10.35940/ijeat.d2404.0410421

[6] Bardhan, A. (2020). *Face Detection- Comparing MTCNN, ResNet-10 and Haar Classifiers*. Www.linkedin.com. https://www.linkedin.com/pulse/face-detection-comparing-mtcnn-resnet-10-haar-adreeja-bardhan/

[7] Barla, N. (2022). *The Complete Guide to Object Tracking [+V7 Tutorial]*. Www.v7labs.com. https://www.v7labs.com/blog/object-tracking-guide

[8] Cattaneo, C., Mainetti, G., & Sala, R. (2015). The Importance of Camera Calibration and Distortion Correction to Obtain Measurements with Video Surveillance Systems. *Journal of Physics: Conference Series*, *658*, 012009. https://doi.org/10.1088/1742-6596/658/1/012009

[9] Chowdhury, A., Kabir, K. M. A., & Tanimoto, J. (2020). How quarantine and social distancing policy can suppress the outbreak of novel coronavirus in developing or under poverty level countries : a mathematical and statistical analysis. *How Quarantine and Social Distancing Policy Can Suppress the Outbreak of Novel Coronavirus in Developing or under Poverty Level Countries : A Mathematical and Statistical Analysis.* https://doi.org/10.21203/rs.3.rs-20294/v2

[10] Derek Ward. (2013, September 27). *Testing Camera Height vs Image Quality*. IPVM. https://ipvm.com/reports/testing-camera-height

[11] Do, A. T., Ngo, T. D., & Bui, T. D. (2016). Feature Extraction for Non-frontal Faces. *Journal of Automation and Control Engineering*, *4*(2), 171–176. https://doi.org/10.12720/joace.4.2.171-176

[12] Gavrilova, Y. (2020). *Testing Machine Learning Models*. Serokell Software Development Company. https://serokell.io/blog/machine-learning-testing

[13] Hamilton, T. (2019, June 10). *Unit Testing Tutorial: What is, Types, Tools, EXAMPLE*. Guru99.com. https://www.guru99.com/unit-testing-guide.html

[14] Jignesh Chowdary, G., Punn, N. S., Sonbhadra, S. K., & Agarwal, S. (2020). Face Mask Detection Using Transfer Learning of InceptionV3. *Big Data Analytics*, 81–90. https://doi.org/10.1007/978-3-030-66665-1_6

[15] Meenpal, T., Balakrishnan, A., & Verma, A. (2019, October 1). *Facial Mask Detection using Semantic Segmentation*. IEEE Xplore. https://doi.org/10.1109/CCCS.2019.8888092

[16] Nithiyasree, K., & Kavitha, T. (2021). Face Mask Detection in Classroom using Deep Convolutional Neural Network. *Turkish Journal of Computer and Mathematics Education*.

[17] O'Carroll, B. (2020). *What Is Rapid Application Development (RAD)?* Codebots. https://codebots.com/app-development/what-is-rapid-application-development-rad

[18] Ubaidah, A., Mansor, A. N., & Hashim, H. (2021). Challenges of COVID Standard Operating Procedure Compliance in a Secondary Private School. *Creative Education*, *12*(08), 1881–1890. https://doi.org/10.4236/ce.2021.128143