BMJ Open

Acceptance of a computer vision facilitated protocol to measure adherence to face mask use: a single-site, observational cohort study among hospital staff

Peter R Chai ,[1,2,3] Phillip Rupp,[2] Hen-Wei Huang,[2,4,5] Jack Chen,[2,4,5] Clint Vaz,[1] Anjali Sinha,[2] Claas Ehmke,[2] Akhil Thomas,[2] Farah Dadabhoy,[1] Jia Y Liang,[2,4,5] Adam B Landman,[1,6] George Player,[7] Kevin Slattery,[8] Giovanni Traverso[4,5]

ABSTRACT

Objectives Mask adherence continues to be a critical public health measure to prevent transmission of aerosol pathogens, such as SARS-CoV-2. We aimed to develop and deploy a computer vision algorithm to provide real-time feedback of mask wearing among staff in a hospital.
Design Single-site, observational cohort study.
Setting An urban, academic hospital in Boston, Massachusetts, USA.
Participants We enrolled adult hospital staff entering the hospital at a key ingress point.
Interventions Consenting participants entering the hospital were invited to experience the computer vision mask detection system. Key aspects of the detection algorithm and feedback were described to participants, who then completed a quantitative assessment to understand their perceptions and acceptance of interacting with the system to detect their mask adherence.
Outcome measures Primary outcomes were willingness to interact with the mask system, and the degree of comfort participants felt in interacting with a public facing computer vision mask algorithm.
Results One hundred and eleven participants with mean age 40 (SD15.5) were enrolled in the study. Males (47.7%) and females (52.3%) were equally represented, and the majority identified as white (N=54, 49%). Most participants (N=97, 87.3%) reported acceptance of the system and most participants (N=84, 75.7%) were accepting of deployment of the system to reinforce mask adherence in public places. One third of participants (N=36) felt that a public facing computer vision system would be an intrusion into personal privacy.

CONCLUSIONS

Public-facing computer vision software to detect and provide feedback around mask adherence may be acceptable in the hospital setting. Similar systems may be considered for deployment in locations where mask adherence is important.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The YoloV4 object detection model was leveraged as the basis of our computer vision programme.
⇒ A computer vision programme was developed using lower resolution closed circuit television still frames to detect mask wearing.
⇒ An observational cohort survey study at a single site was conducted to understand user experience around using a computer vision programme.
⇒ Study was limited by the single academic hospital setting and by enrolling hospital staff only.
⇒ Surveys were facilitated by study staff, limiting individual-level interactions with the computer vision programme.

COVID-19 has been the adoption of universal face mask recommendations especially in settings where social distancing may not be possible.[1 2] Multiple investigations have demonstrated the efficacy of both surgical masks as well as commercial cloth face masks in preventing spread COIVD-19 with some studies demonstrating a relative risk reduction of up to 70% of acquiring COVID-19 in individuals who wear masks compared with those who do not.[2 3] Despite compelling evidence of the effectiveness of masks, adherence is variable globally.[4 5] Within the USA, the use of face masks is variable, and in addition to mass gathering super-spreader events, the presence of multiple unmasked individuals likely contributes to surges of COVID-19.[6–8]

While universal mask mandates were features of initial surges of COVID-19 cases in the USA, decline of cases and the roll-out of vaccines has relaxed many mask mandates across the country.[4] Yet, the emergence of the delta and omicron variants of COVID-19 has

INTRODUCTION

A primary pillar of public health interventions to prevent the transmission of

resulted in increased disease transmission and altered masking recommendations.[9] In states in the USA where mask adherence has been over 75%, the rate of COVID-19 infection was significantly lower compared with states where there was low or no adherence to masks.[6 10] A cluster randomised controlled trial in Bangladesh also demonstrated that mask wearing was associated with decreased transmission of COVID-19 across communities.[11]

Measuring mask wearing among the general population is difficult especially in the setting of rapidly changing local regulations.[4 12] Most investigations that assess the use of masks during COVID-19 and other infectious disease outbreaks frequently rely on self-report or location-based observations to understand the tendency of respondents to wear masks and comply with local or national rules surrounding mask wearing.[6 8] During COVID-19, the development of computer vision (CV) systems has automated reporting around certain public measures aimed at mitigating COVID-19.[13] Private enterprises have sought to use CV systems to present social distancing data to employees, and some investigations have leveraged closed circuit television (CCTV) footage to provide cross-sectional analyses on the degree of mask wearing in certain settings.[14] Despite this early work, the development of CV algorithms that accurately detect the presence of face masks remains underdeveloped. In addition, it is unclear how individuals would accept feedback and perceive the deployment of these systems in places where mask wearing is especially important. In this investigation, we describe formative work to develop a face mask CV algorithm, initial training and validation in a hospital setting and key user feedback surrounding potential privacy implications of deploying face mask recognition systems in healthcare.

METHODS
Development of a CV model for face mask detection
The system comprises an FLIR Chameleon Camera (FLIR Systems, Wilsonville, Oregon, USA) for capturing video, a Nvidia Jetson Nano (NVidia, Santa Clara, California, USA) for data processing, an object detection and mask classification model, a person tracking system, and a high definition television for displaying real-time results (figure 1).

For person detection and face mask classification, we used the YoloV4 model, a state-of-the-art object detection model that has been adapted to various detection tasks.[15] Furthermore, the model can be optimised to make real-time predictions on edge computing devices.[16] YoloV4 detects bounding boxes of the desired classes and classifies them. Since neural networks are based on supervised learning techniques, YOLOv4 can only differentiate between predefined classes—in our case only two (mask=positive, no mask=negative). The original model was trained on 80 different classes, none of which are equal to the classes mentioned. Therefore, we used our own dataset to retrain the model.[17] To do so, we employed

a method called transfer learning where only the classification layers of the neural networks are retrained, which are usually the last (fully connected) layers in the neural network.[18] This allows us to use the convolutional layers as a predefined preprocessor for the given object detection task. Because the original model was trained on ImageNet, which is significantly larger than the dataset we use (15 million images compared with 2400 images), we assume that the convolutional layers are suited for bounding box regression tasks. Nevertheless, the classification layer of the neural network needed to be optimised again. YOLOv4 has approximately 64 million parameters of which 21 million define the classification layers. Transfer learning is achieved through only updating the weights in the classification layer in the backward pass of the backpropagation algorithm.

As the dataset, we used five CCTV videos from our hospital recording two different entries. We obtained these videos through approval from our hospital Office of General Counsel and Security and obtained approval by the Mass General Brigham Institutional Review Board. These CCTV videos depicted staff entering the hospital prior to COVID-19 and during the pandemic to demonstrate the presence and absence of mask wearing. The videos have a resolution of 1280×720 pixels and were recorded at approximately 15 frames per second. The videos used came from three different cameras, recording two locations. One member of our study team annotated videos through visual inspection to label individuals who were wearing masks (positive condition) compared with not wearing masks (negative condition).

We split the dataset in a training and test set, the first to calibrate the model and the latter to evaluate the model performance (online supplemental material 1). We randomly selected 70% of the images from the video from all four footages to be the training set, and the remaining 30% assigned to the dataset. This resulted in approximately 1500 images in the training and 900 images in the test set. To further ensure robustness and minimise overfitting of the model, we employed a method called early-stopping where we split the training set into two datasets, one to calibrate the models parameters, the other to track the models performance while training.[19] We used the third dataset (validation set) in the training process only. Again, we split the training set with a ratio of 70:30 into a new training set and a validation set (online supplemental material 1). We use the validation set to evaluate the accuracy every epoch. If we see an increase of the validation accuracy over 10 epochs, we assume that the model is starting to overfit and we stop the training process. This resulted in 160 training epochs.

For other hyperparameters, we used the initial training parameters as used to train the original YoloV4 model. We defined the overlap threshold to be 50%. The overlap threshold determines how much of the area of the predicted bounding and the ground truth bound must overlap to be classified as a positive.
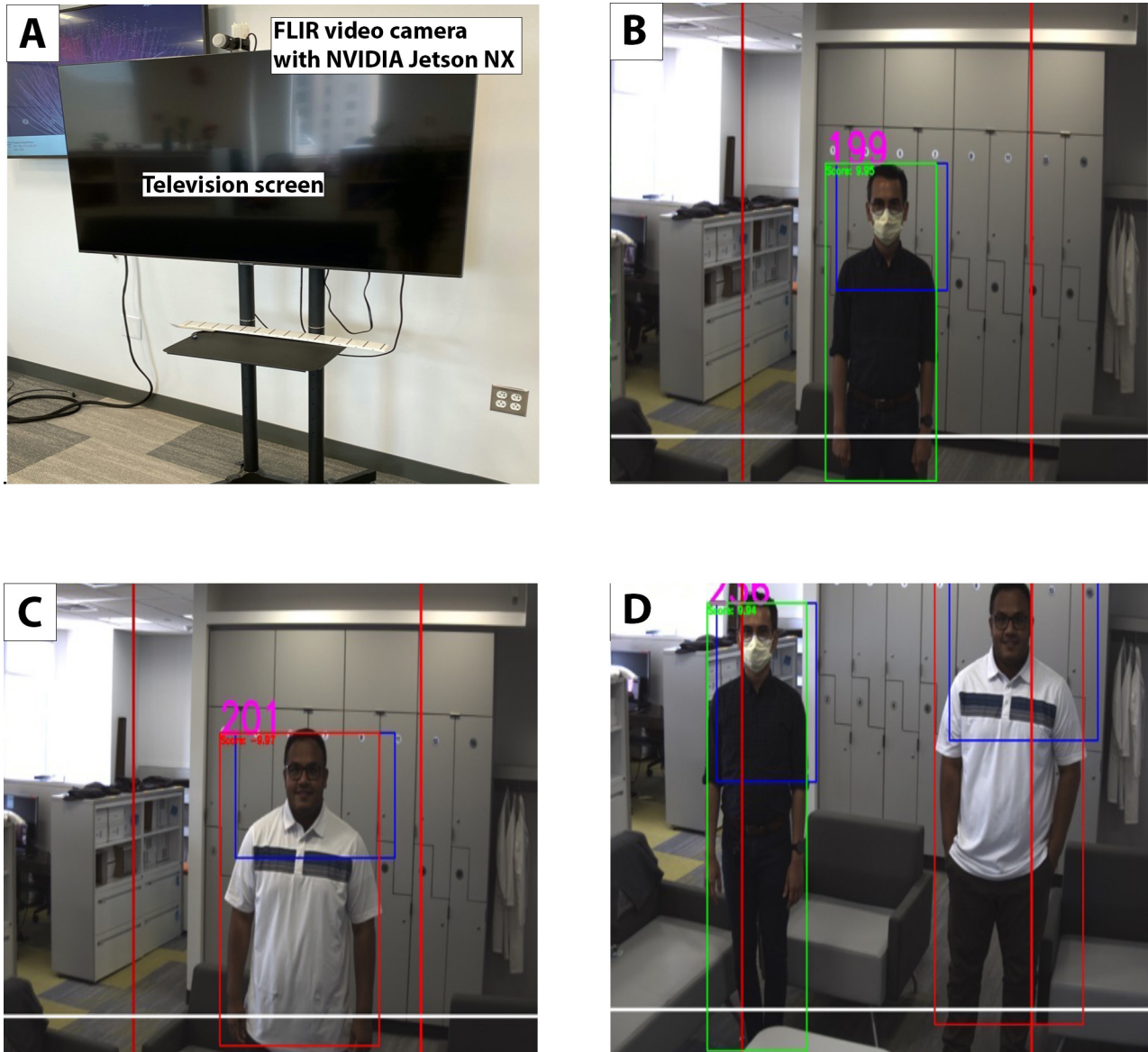
Figure 1 Sample images of the computer vision mask system the system (A) is able to detect both individual positive detection of mask adherence (B), mask non-adherence (C), and simultaneous adherence and non-adherence (D). Positive detection results in a blue bounded box around the individual wearing a mask. Non-adherence is indicated by a pink bounding box around the individual who is not wearing a mask. Both individuals in the figure have consented for their images to be published in this manuscript. Individuals in the image are non-patients.

We evaluated the model on three, metrics, namely precision, recall and F1 score which are defined as

$$Precision = \frac{TP}{TP+FP},$$
$$Recall = \frac{TP}{TP+FN},$$
$$F1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Where *TP*, *FP* and *FN* mean true positive, false positive and false negative respectively. A face is classified as a true positive if the predicted bounding box and the ground truth bounding box do overlap by at least 50% in our case and the predicted class corresponds to the ground truth. A false positive is given, when the predicted bounding box and face bounding box do not overlap with at least 50%, but the predicted class is correct. F1

score is a metric to evaluate the balance between precision and recall of the model. We enabled individual tracking through a tracking-by-detection scheme. In each frame, the positions of all individuals are determined using the YOLOv4 object detector. Between frames, the intersection-over-union method is applied to associate the positions of detected persons. This method achieves state-of-the-art tracking accuracy without requiring additional image information and operates in real time.[20] Accurate tracking of individuals allows for accurate counting of the different classes of mask wearing during clinical trials.

We also evaluated this method on a subset of the CCTV footage to show that this tracking method is applicable to our setup. We adapted this method slightly, namely

instead of taking the intersection over union (IOU) as an evaluation metric, we calculated the Euclidean distances of the centroids of the bounding boxes over time.[21] A new identification number (ID) is assigned, if a new bounding box is detected. The ID will be kept for 60 additional frames if the YOLOv4 network is not able to detect a bounding box correctly. If a bounding box is detected close to where the tracking was lost, the ID will be assigned to the new location.

Therefore, we recorded 3 min of one of the five videos and manually counted the number of people and compared it to the total number of assigned IDs. We would like to note that the mentioned video also contains images from the training set for the YOLOv4 since these images were selected randomly. This should not play a major role because we verified the YOLOv4 mask detector on a separate test set. We tuned the ID tracked such that the tracking can only be lost in the areas where the subjects can actually leave the room. We tuned the ID tracker to lose tracking only in areas where subjects can leave the room. These areas were determined empirically.

We conducted preliminary testing of our CV mask detection system among the study team. To provide live user feedback, we decided to annotate live image with a person bounding box, face bounding box, the mask detection result, mask detection confidence and a counter for past mask detection results. We used a colour scheme of red, yellow and green to delineate the presence or absence face masks on individuals using the CV system. A red bounding box indicates that no mask is worn, and a green one that the person is wearing a face mask (figure 1).

Clinical deployment
We conducted a pilot observational study to understand the CV mask detection system user experience among employees. The study took place at a large, academic, urban tertiary care hospital in Boston, Massachusetts, USA. We developed a quantitative assessment that collected information around the perception of privacy with the system, willingness to engage with the system and comfort interacting with the mask system. Questions were developed by key members of the study team (PC, FD, GT and H-WH) and tested among the study team for clarity and completeness. Questions included basic demographics surrounding age, gender, race/ethnicity, baseline attitudes and awareness of mask use, and finally attitudes towards their experience with the CV mask system. We also asked participants to rate their willingness to have facial recognition software in public places to encourage mask wearing. We also asked participants to describe whether they considered these systems a violation of personal privacy and their comfort with interacting with video camera systems in public places. All questions pertaining to the CV mask system and perceptions of these software were presented with a five point Likert scale. Finally, we asked how likely participants would be to recommend the system to a colleague using the 10-point

Net Promoter Score, a common metric to understand acceptance and usability of novel systems.[22]

A trained research assistant recruited hospital staff entering one of the main hospital entrances to engage with the mask system over the course of 5 days from 26 pril 2020 to 30 April 2020. We placed the CV mask detection system in a separate location after individuals had completed their attestation and switched into the hospital issued mask. Participants were then invited to stand in front of the CV mask detection system which measured whether the participant was wearing a mask. The research assistant introduced the system to the participant by explaining that the CV mask system was attempting to detect the presence of mask wearing among individuals in the hospital and may have applicability outside of the hospital setting as well to annotate and eventually provide feedback on mask adherence to users in real time. We also described the colours of the bounding boxes (eg, green for mask detected, red for no mask detected), how the system was developed and how the system detects an individual and measures mask wearing. Participants were allowed to interact with the CV mask detection system for as long as they wanted to experience detection of mask wearing compared with detection of non-mask wearing. The research assistant conducted a ground truth assessment by physically observing the presence, appropriate positioning or absence of the mask on the participant. Finally, participants completed the quantitative assessment with the assistance of the study team.

Data analysis
For the study, we conducted basic descriptive statistics to define mean age, gender, proportions of hospital staff members enrolled and ethnicity. We additionally calculated percent accuracy of the mask system detecting mask adherence compared with ground truth observation by a study staff member. We calculated mean scores for response categories and reported percentages of participants who rated the mask system through each category. Finally, we used an ordinal logistic regression to examine the relationship between participant's experience with the CV mask system and their attitudes towards CV systems in other public places and comfort with having video systems in public places.

Patient and public involvement
Patients and the public were not involved in the design, conduct, reporting or dissemination plans of our research. The office of the general counsel of Massachusetts General Brigham and the COVID-19 Critical Response Team were involved in the conduct of our research.

RESULTS
We evaluated the retrained YOLOv4 detection system on the test set, which resulted in an area under the receiver operating characteristic curve (AUC) with a value of 0.81. To assess performance, we tested our model with
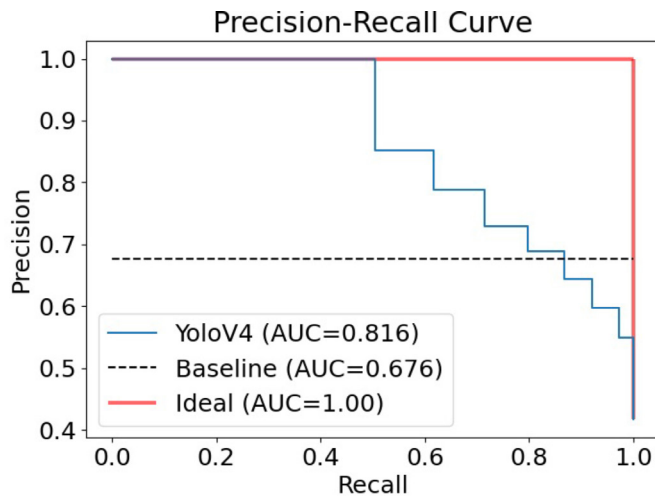
Precision-Recall Curve

Figure 2 Precision recall curve on test set of hospital CCTV dataset the plot shows the accumulated averaged precision and recall on the evaluation of the test sorted with varying confidence thresholds. AUC integral under the blue curve. The baseline is the number of positive samples over the number of all samples. The red line shows an ideal classifier, which always classifies correctly regardless of the confidence threshold. AUC, area under the curve; CCTV, closed circuit television.

decreasing confidence thresholds—starting at 1—with step size 0.1. The IoU threshold was set to 0.5. We set the IoU higher than the initial YoloV4 training IoU value because we track individual faces across multiple frames. Lowering the IoU would lead to more variance in the centre position over time of a specific detected face bounding box which makes tracking more difficult. Increasing the IoU results in more stable centre coordinates across frames which enables more robust tracking. A perfect classifier has an AUC of 1 while a random classifier has AUC equal to the ratio of positive samples to all samples of the test set. The maximum F1 score is 0.74 where the precision and recall at this maximum are 0.64 and 0.87. The baseline F1 score at the mentioned recall is 0.66. Figure 2 shows the precision-recall curve evaluated on the Brigham CCTV test set. This curve shows the trade-off between precision and recall with varying confidence thresholds. The dashed black line shows the prediction baseline which is the ratio of positive samples in the test set. The baseline shows the precision and recall if the algorithm would only predict the positive class. It is the number of positive samples over all samples. The red line shows an ideal classifier which always classifies the given sample as a true positive regardless of the confidence threshold. The better the model, the closer the blue line is to the red line. Thus, our model showed a strong correlation between detected mask wearing and the ground truth.

To count the number of individuals present during the live CCTV stream, we employed a person tracker to uniquely track the subjects in the video. The number of assigned IDs is then compared with the ground truth number of counted subjects. A video of 3 min was used

Table 1

Characteristic	Participants N=111 (%)
Age, mean (SD)	40.49 (15.57)
Sex	
Male	53 (47.7)
Female	58 (52.3)
Race/ethnicity*	
White	54 (49.1)
Asian or Asian American	21 (19.1)
Black or African American	17 (15.5)
Latino or Hispanic	10 (9.1)
Middle Eastern	4 (3.6)
Mixed race	4 (3.6)
Other	3 (2.7)

*Participants were able to report more than one category.

to validate the proposed tracker. In the video, 42 subjects appear while the tracker assigned 43 unique IDs. This demonstrates that the proposed tracker is suitable for the counting subjects in a video.

Clinical study

During the study period, we approached 184 eligible individuals and we were able to consent and enrolled 111 participants. Mean age was 40.49 (SD 15.57), and 52.3% (N=58) participants were female (table 1). A total of 49.1% (N=54) identified as white, 19.1% (N=21) were Asian or Asian American, 15.5% (N=17) were black or African American, 9.1% (N=10) were Latino or Hispanic, 3.6% (N=4) were Middle Eastern and 3.6% (N=4) identified as mixed race. All participants were observed to be wearing a face mask by the study team, and the CV system accurately detected the presence of mask adherence 100% of the time.

When asked about baseline attitudes towards camera and CV systems, 43.2% (N=48) participants reported they felt either comfortable or very comfortable with camera systems in public spaces (table 2). Only 16 participants (14.4%) reported concern around public use of these systems. A majority of participants (N=84, 75.7%) reported that they were in favour of the deployment of facial recognition software similar to that described in this investigation to reinforce mask adherence in public places. In addition, almost all participants (N=104, 93.7%) reported support for universal face mask mandates.

After participants interacted with the CV mask system, we asked about their overall experience. Most participants (N=97, 87.3%) reported a positive experience interacting with the system in the hospital. In addition, 83.7% (N=92) of participants reported they would recommend the system after their experience. Given concerns regarding potential privacy issues associated with automated facial recognition systems, we asked if participants considered the system an intrusion into their personal

Table 2 User response to interacting with the computer vision mask system

	Participant N=111
Baseline attitudes towards having cameras in public places	
Very uncomfortable	5 (4.5%)
Uncomfortable	11 (9.9%)
Neutral	47 (42.3%)
Comfortable	33 (29.7%)
Very comfortable	15 (13.5%)
Use of facial recognition software in public places to encourage use of face masks	
Very unfavourable	13 (11.7%)
Somewhat unfavourable	8 (7.2%)
Somewhat favourable	27 (24.3%)
Very favourable	57 (51.4%)
Don't know	6 (5.4%)
Support of universal face mask mandates	
Not supportive	6 (5.4%)
Supportive	104 (93.7%)
No opinion	1 (0.9%)
Overall experience interacting with the computer vision system	
Very poor	1 (0.9%)
Poor	1 (0.9%)
Neutral	12 (10.8%)
Good	42 (37.8%)
Excellent	55 (49.5%)
Would recommend the system to a colleague	
Not at all likely	2 (1.8%)
Unlikely	4 (3.6%)
Neutral	12 (10.9%)
Likely	62 (56.4%)
Extremely likely	30 (27.3%)
Computer vision systems are an intrusion of privacy	
Strongly disagree	24 (22.0%)
Disagree	28 (25.7%)
Neutral	21 (19.3%)
Agree	25 (22.9%)
Strongly agree	11 (10.1%)

information. In response, only 33% (N=36) reported concerns regarding personal privacy.

Next, we considered the participant experience with the CV mask system and examined whether individuals who had better experiences with the system had at baseline more favourable attitudes towards use of facial recognition software in public places. Based on an ordinal logistic regression, we found that individuals who had increasing positive experiences interacting with our CV mask system had 1.62-times increased odds of also being favourable of facial recognition software in public places (p=0.038). We also found that there was no significant relationship between the positive experiences interacting with the CV mask system and baseline attitudes towards having cameras in public places.

DISCUSSION

Mask wearing has been a significant component of managing airborne pathogens like SARS-CoV-2. Even standard cloth masks have been demonstrated to provide at least 50% reduction in disease transmission compared with no masks at all.[23–25] Several other empirical investigations including a large national randomised controlled trial in Bangladesh have also demonstrated the efficacy of face masks to prevent COVID-19 transmission.[11] In the face of rising and falling background infection rates, municipalities have sought to enact mask mandates to mitigate disease spread. Despite this, mask uptake is variable as is reinforcing adherence.[26] This investigation demonstrates the development of a CV system that can innovatively detect face mask adherence both from a distance using CCTV footage as well as live in-person video streams. In addition, our data demonstrate that a CV system is acceptable by individuals in a hospital system.

Compared with the previously proposed retrained YOLOv4 mask detection network, our model had similar performance compared with Kumar *et al*'s model.[15] This is encouraging especially because previous models used a dataset of over 50 000 images while our model was trained on 1,500 CCTV images. Despite the lower number of datapoints in our training set, our model importantly was demonstrated to effectively differentiated between masked and unmasked individuals in two key scenarios: CCTV footage as well as in person footage. These two scenarios suggest potential future applications of the algorithm as a measure of overall mask adherence in larger public spaces with CCTVs such as hospitals, transportation hubs, sporting events and grocery stores while an in-person mask detection system may be applied at key junctures like the entrance to a hospital, store or stadium.[27 28] In addition, face bounding boxes used in our CCTV dataset were low resolution (30×40 pixels) compared with the resolution of the dataset to train the mentioned model (130×145 pixels). This indicates that our model performs well despite the significantly lower resolution in the training and test set, and therefore, may be applicable as a plugin for commercial CCTV systems, which may be of lower resolution. This would enable broader scalability of our CV mask detection system without the need to upgrade legacy systems into modern high resolution security cameras. Importantly, our model deviates only 5% in F1 score from the mentioned model by Kumar *et al*, but is able to infer face bounding boxes 15 times smaller than the bounding boxes used to train and test the Kumar *et al* model.[15]

Our formative data suggest that individuals in hospitals are receptive to the use of CV systems to provide visual

annotation of mask adherence. We decided to integrate our CV system at an entrance to the hospital as part of the initial screening to enter the building. This strategy effectively detected the presence of masks, a task that was previously delegated to a hospital staff member. In addition, individuals who used our system did so in public and importantly did not consider interaction with the system in public a breach of personal privacy. These data suggest that one potential place to deploy such CV systems that detect mask wearing would be at the entrance to key public buildings like grocery stores, museums, schools and transportation hubs. By using standard off the shelf hardware and software modules to build and programme the system, we anticipate that the cost of assembling and maintaining a system like this may reduce deployment and maintenance costs for individuals who wish to leverage CV for mask detection. For public health entities or individuals who do not wish to purchase additional hardware to deploy such a system, our data additionally demonstrate that a low cost computer can run our CV programme, which can be integrated into existing surveillance systems such as CCTV.

As COVID-19 continues to mutate it may become increasingly virulent and necessitate reinforcement of face mask rules to mitigate disease spread.[9] Given the fatigue associated with COVID-19 leading to lower face mask compliance and the potential willingness to take corrective action in response to posters and guidance, an interactive CV system could serve as a potential measure to improve face mask wearing in public spaces.[4 13] In addition, future airborne viral pandemics may occur where rapidly providing just-in-time feedback around mask adherence may be a key public health effort. Real-time interventions that provide corrective feedback when adherence is poor can use CV data in several ways. First, individual detected non-adherence could trigger a response on a display screen providing prerecorded verbal or visual feedback to encourage mask wearing. Second, mask adherence detected through a CV system could be linked to a second activity like unlocking a door to a hospital room thereby ensuring adherence before entering a specific space. Finally, passive CV data may be leveraged to provide objective measures of other public health interventions like in-person reminders from workers or public health officials to wear masks, or may even be used to direct the placement of free face masks in places where adherence is poor.

Overall, this investigation demonstrates that a low cost CV system is feasible to deploy in a hospital setting to measure mask adherence. Importantly, individuals who experienced our CV system did not consider it an intrusion into their privacy, were accepting of the system itself, and thought it could be used in other public spaces to encourage mask adherence. This suggests that continued development of CV mask detection systems can be a key method to encourage adherence to a key public health measure for the COVID-19 pandemic and future airborne pathogens. This system may additionally be leveraged as a method to better understand population level mask adherence in key locations. Future work should consider further human factors research into how individuals interact with CV mask systems including the time it takes for them to interact with the system, baseline attitudes that may affect perceptions of such systems, the frequency in which individuals may adjust or wear a mask based on feedback and the annotations in the form of bounding boxes or other messaging that serves to reinforce mask adherence. In addition, qualitative methodologies may be helpful in elucidating key facilitators and barriers surrounding how to best implement these systems in healthcare settings and other public institutions where masks may prevent airborne transmission of infectious pathogens.

This investigation had several limitations. First, our limited number of training images and the vantage point of CCTV images available may have impacted the accuracy of our CV model. All individuals in our training set and CCTV images were wearing standard disposable surgical face masks; future work may need to test our CV algorithm on other face coverings. Despite this, our unique perspective from a CCTV demonstrates the feasibility of applying a mask detection model to CCTV systems in the future. Second, lighting conditions were consistent throughout our training set videos, which may have contributed to an accurate mask detection algorithm in the hospital, but may limit generalisability outside of a controlled lighting environment. Third, we conducted our study at an urban, academic medical centre in Massachusetts, USA. Perspectives around baseline support for mask mandates and willingness to wear masks may differ across the USA and decrease the generalisability of this investigation. Fourth, the experience interacting with the CV system was chaperoned by a study staff member. Perspectives and willingness to interact with such a system may differ in the setting of standalone kiosks or other iterations of the system.

Author affiliations

[1]Emergency Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[2]Massachusetts Institute of Technology Koch Institute for Integrative Cancer Research, Cambridge, Massachusetts, USA
[3]The Fenway Institute, Boston, MA, USA
[4]Medicine/Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts, USA
[5]Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[6]Digital Innovation Hub, Brigham and Women's Hospital, Boston, MA, USA
[7]Facilities, Brigham and Women's Hospital, Boston, Massachusetts, USA
[8]Security, Safety and Parking, Brigham and Women's Hospital, Boston, Massachusetts, USA

Twitter Clint Vaz @clint_vaz and Giovanni Traverso @cgtraverso

initial version of the manuscript. All authors provided critical edits and approve of the final version of the manuscript. GT took responsibility for the data as a whole and the content of the manuscript.

Competing interests GT has a financial interest in Teal Bio, a biotechnology company making transparent and reusable respirators. ABL is a member of the Abbot Medical Device Cybersecurity Council. Complete details of all relationships for profit and not for profit for G.T. can be found at the following link: https://www.dropbox.com/sh/szi7vnr4a2ajb56/AABs5N5i0q9AfT1IqIJAET5a?dl=0. No other authors have competing interests to report.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study was designated not human subjects research by the Mass General Brigham Institutional Review Board: Protocol 2020P00247.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

ORCID iD
Peter R Chai http://orcid.org/0000-0003-0955-4117

REFERENCES

1 Howard J, Huang A, Li Z, *et al*. An evidence review of face masks against COVID-19. *Proc Natl Acad Sci U S A* 2021;118:e2014564118.
2 Wu J, Xu F, Zhou W, *et al*. Risk factors for SARS among persons without known contact with SARS patients, Beijing, China. *Emerg Infect Dis* 2004;10:210–6.
3 Kenyon C. Widespread use of face masks in public may slow the spread of SARS CoV-2: an ecological study. *medRxiv* 2020:2020.03.31.20048652.
4 Petherick A, Goldszmidt R, Andrade EB, *et al*. A worldwide assessment of changes in adherence to COVID-19 protective behaviours and hypothesized pandemic fatigue. *Nat Hum Behav* 2021;5:1145–60.
5 Badillo-Goicoechea E, Chang T-H, Kim E, *et al*. Global trends and predictors of face mask usage during the COVID-19 pandemic. *BMC Public Health* 2021;21:2099.
6 Fischer CB, Adrien N, Silguero JJ, *et al*. Mask adherence and rate of COVID-19 across the United States. *PLoS One* 2021;16:e0249891.
7 Kantor BN, Kantor J. Non-pharmaceutical interventions for pandemic COVID-19: a cross-sectional investigation of US General public beliefs, attitudes, and actions. *Front Med* 2020;7:384.
8 Tamamoto KA, Rousslang ND, Ahn HJ, *et al*. Public compliance with face mask use in Honolulu and regional variation. *Hawaii J Health Soc Welf* 2020;79:268–71.
9 US Centeres for Disease Control. Omicron variant: what you need to know: US centers for disease control, 2022. Available: https://www.cdc.gov/coronavirus/2019-ncov/variants/omicron-variant.html
10 Lyu W, Wehby GL. Community use of face masks and COVID-19: evidence from a natural experiment of state mandates in the US. *Health Aff* 2020;39:1419–25.
11 Abaluck J, Kwong LH, Styczynski A, *et al*. Impact of community masking on COVID-19: a cluster-randomized trial in Bangladesh. *Science* 2022;375:eabi9069.
12 Ford N, Holmer HK, Chou R, *et al*. Mask use in community settings in the context of COVID-19: a systematic review of ecological data. *EClinicalMedicine* 2021;38:101024.
13 Ulhaq A, Born J, Khan A, *et al*. COVID-19 control by computer vision approaches: a survey. *IEEE Access* 2020;8:179437–56.
14 Amazon Web Services. TensorIoT SafetyVisor: Next Generation Worker Safety Platform Built on AWS IoT and ML [cited 2022 February 1]. Available: https://aws.amazon.com/iot/solutions/TensorIoTSafetyVisor/
15 Yu J, Zhang W. Face mask wearing detection algorithm based on improved YOLO-v4. *Sensors* 2021;21. doi:10.3390/s21093263. [Epub ahead of print: 08 May 2021].
16 Jung J. TensorRT MODNet, YOLOv4, YOLOv3, SSD, MTCNN, and GoogLeNet, 2021. Available: https://github.com/jkjung-avt/tensorrt_demos [Accessed cited 2021 December 10].
17 Chai PR, Rupp P, Huang H. Data from: the acceptance of a computer vision facilitated protocol to measure adherence to face masks: a single site, observational cohort study. *Github* 2022 https://github.com/Jhc999/CVMask2022
18 Tan C, Sun F, Kong T, et al. A survey on deep transfer learning. *International conference on artificial neural networks*, 2018.
19 Goodfellow I, Bengio Y, Courville A. *Deep learning: MIT press*, 2016.
20 Bochinski E, Eiselein V, Sikora T. High-Speed tracking-by-detection without using image information. *2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS)*, 2017.
21 Rosebrock A. Simple object tracking with OpenCV, 2018. Available: https://www.pyimagesearch.com/2018/07/23/simple-object-tracking-with-opencv/ [Accessed cited 2021 December 10].
22 Reichfield F. The one number you need to grow. *Harvard Business Review* 2003.
23 Davies A, Thompson K-A, Giri K, *et al*. Testing the efficacy of homemade masks: would they protect in an influenza pandemic? *Disaster Med Public Health Prep* 2013;7:413–8.
24 Eikenberry SE, Mancuso M, Iboi E, *et al*. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect Dis Model* 2020;5:293–308.
25 Mahase E. Covid-19: what is the evidence for cloth masks? *BMJ* 2020;30:m1422.
26 Betsch C, Korn L, Sprengholz P, *et al*. Social and behavioral consequences of mask policies during the COVID-19 pandemic. *Proc Natl Acad Sci U S A* 2020;117:21851–3.
27 Draughon GTS, Sun P, Lynch JP. Implementation of a computer vision framework for tracking and visualizing face mask usage in urban environments. *2020 IEEE International Smart Cities Conference (ISC2)*; 28 Sept.-1 Oct. 2020, 2020.
28 Nayak R, Manohar N. Computer-Vision based face mask detection using CNN. *2021 6th International Conference on Communication and Electronics Systems (ICCES)*; 2021 8-10 July, 2021.