
Brief Communications

Automatic detection of hand hygiene using computer vision technology

Amit Singh,^{1,*} Albert Haque,^{2,*} Alexandre Alahi,³ Serena Yeung,⁴ Michelle Guo,² Jill R Glassman,⁵ William Beninati,⁶ Terry Platchek,^{1,5} Li Fei-Fei,^{2,†} and Arnold Milstein^{5,†}

¹Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA, ²Department of Computer Science, Stanford University, Stanford, California, USA, ³Department of Civil Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, ⁴Department of Biomedical Data Science, Stanford University, Stanford, California, USA, ⁵Clinical Excellence Research Center, Stanford University School of Medicine, Stanford, California, USA, and ⁶Intermountain TeleHealth Services, Murray, Utah, USA

*These authors contributed equally.

†co-last authors.

Corresponding Author: Amit Singh, MD, 300 Pasteur Drive MC 5776, Stanford, CA 94305-5776, USA; atsingh@stanford.edu

Received 27 February 2020; Revised 14 April 2020; Editorial Decision 12 May 2020; Accepted 21 May 2020

ABSTRACT

Objective: Hand hygiene is essential for preventing hospital-acquired infections but is difficult to accurately track. The gold-standard (human auditors) is insufficient for assessing true overall compliance. Computer vision technology has the ability to perform more accurate appraisals. Our primary objective was to evaluate if a computer vision algorithm could accurately observe hand hygiene dispenser use in images captured by depth sensors.

Materials and Methods: Sixteen depth sensors were installed on one hospital unit. Images were collected continuously from March to August 2017. Utilizing a convolutional neural network, a machine learning algorithm was trained to detect hand hygiene dispenser use in the images. The algorithm's accuracy was then compared with simultaneous in-person observations of hand hygiene dispenser usage. Concordance rate between human observation and algorithm's assessment was calculated. Ground truth was established by blinded annotation of the entire image set. Sensitivity and specificity were calculated for both human and machine-level observation.

Results: A concordance rate of 96.8% was observed between human and algorithm ($\kappa = 0.85$). Concordance among the 3 independent auditors to establish ground truth was 95.4% (Fleiss's $\kappa = 0.87$). Sensitivity and specificity of the machine learning algorithm were 92.1% and 98.3%, respectively. Human observations showed sensitivity and specificity of 85.2% and 99.4%, respectively.

Conclusions: A computer vision algorithm was equivalent to human observation in detecting hand hygiene dispenser use. Computer vision monitoring has the potential to provide a more complete appraisal of hand hygiene activity in hospitals than the current gold-standard given its ability for continuous coverage of a unit in space and time.

Key words: computer vision, hand hygiene, healthcare acquired infections, patient safety, machine learning, artificial intelligence, depth sensing

INTRODUCTION

Hospital-acquired infections are a seemingly recalcitrant challenge. According to data from the Centers for Disease Control and Prevention, in the United States about 1 in 31 hospitalized patients has at least 1 healthcare-associated infection daily.¹ Hand hygiene is critical in preventing hospital acquired infections yet is very difficult to monitor and consistently perform.² For decades, in-person human observation has been the gold-standard for monitoring hand hygiene.² However, it is widely known that this method is subjective, expensive, and discontinuous. Consequently, only a small fraction of hand hygiene compliance is observed.²

Technologies such as radiofrequency identification have been tested in hospitals to help monitor hand hygiene but do not correlate with human observation.^{3,4} Moreover, objects and humans must be manually tagged and remain physically close to a base station to enable detection, thereby disrupting workflow. Video cameras can provide a more accurate picture of hand hygiene activity but raise privacy concerns and are costly to review.⁵⁻⁷

Computer vision technology may provide an innovative solution to perform more accurate and privacy-safe appraisals of hand hygiene. Depth sensors capture 3-dimensional silhouettes of humans and objects based on their distance from the sensor. As they are not actual cameras and do not utilize color photo or video, they preserve the privacy of individuals in the image. Identification of diverse visual images by computer vision has already entered the healthcare setting by assessing pathologic images in diabetic retinopathy, skin cancer, and breast cancer metastases.⁸⁻¹¹ In these studies, machine learning was applied to static images with clear pathologic definitions and criteria for diagnosis. However, hand hygiene dispenser usage is dynamic, fast, and related to physical human movements. Moreover, understanding human activities is a challenging task, even for more mature artificial intelligence disciplines such as robotics.¹²⁻¹⁵ Computer vision has recently been evaluated in simple hand hygiene appraisals, but only in simulated environments.^{16,17}

We present an application of computer vision in observing hand hygiene dispenser use. The primary objective was to determine if a computer vision algorithm could accurately identify hand hygiene

dispenser usage from images collected from depth sensors placed in the unit with equivalent accuracy compared with simultaneous human observation.

MATERIALS AND ETHODS

Image acquisition

Sixteen privacy-safe depth sensors were installed in an acute care unit of Lucile Packard Children's Hospital Stanford (LPCH) (Palo Alto, California). As the sensors are not video cameras, faces of people and colors of clothing are not discernable and therefore are privacy safe (see [Figure 1](#)). Sensors were mounted on the ceiling above wall-mounted hand hygiene dispensers outside patient rooms. These hand hygiene dispensers are triggered by motion and dispense hand sanitizer when a hand is placed underneath them.

Imaging data were collected at LPCH from March to August of 2017. Sensors detect motion and continuously collect image data when movement is detected. Depending on how quickly the subjects in the images were moving, the images could be static or could be short (2-3 seconds) video clips. This project was reviewed by the Stanford University Institutional Review Board and designated as nonhuman subjects research.

Development, training, and testing of the algorithm

A machine learning algorithm was developed to analyze the images obtained by the sensors to identify when hand hygiene dispensers were used/not used in the images. The algorithm was implemented in Python 3.6 (Python Software Foundation, Wilmington, DE) using the deep learning library Pytorch 1.0 (<https://pytorch.org/>). The algorithm is a densely connected convolutional neural network trained on 111 080 images (training set) (see [Supplementary Appendix](#)).^{18,19} The research team annotated whether the hand hygiene dispenser was used in each image. This annotation was done by 9 computer scientists and 4 physicians. The computer scientists were trained by physicians to detect hand hygiene dispenser use. For the purposes of our study, appropriate use was defined as use of the hand hygiene dispenser upon entry or exit of a patient room as dem-

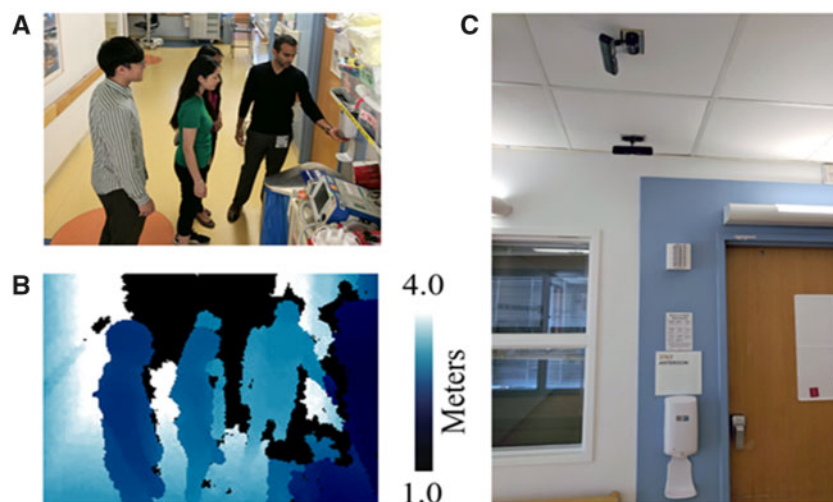


Figure 1. Depth sensor placement and image comparison. Privacy-safe depth image compared with color photos. (A) Example standard color photo and (B) corresponding privacy-safe depth image captured by the sensor. The depth images are artificially colored: the darker color indicates pixels closer to the sensor. Black pixels indicate that the pixel is out of range. (C) Two depth sensors attached to the ceiling of a hospital ward. Subjects in the photo consented to have their photograph taken for illustrative purposes.

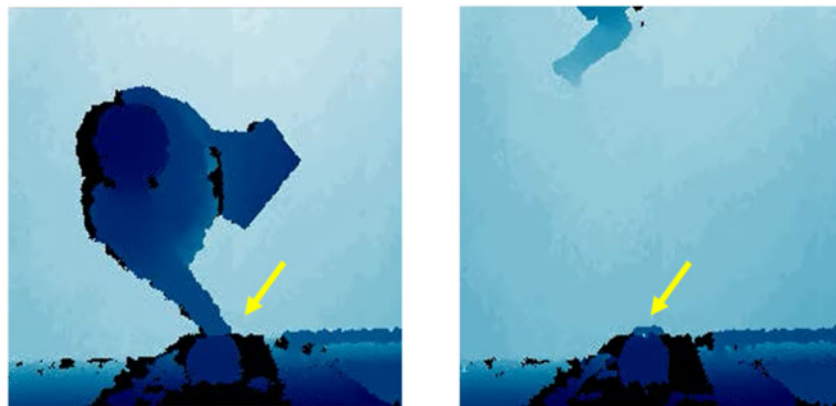


Figure 2. Positive and negative examples of depth sensor data. Examples of hand hygiene dispenser use (left) and nonuse (right) in image dataset. Yellow arrows indicate the location of hand hygiene dispensers in the images.

onstrated by a human with their hand underneath the dispenser indicating its usage. Nonuse was defined as a healthcare professional entering or exiting a room without putting a hand underneath the hand hygiene dispenser (see [Figure 2](#)). If a healthcare professional used the dispenser but did not enter the room, this was not classified as an event of interest. Annotation ambiguities were infrequent and resolved by discussion among the annotators. It is important to note that compliance rates were not calculated as this was not the focus of the study; the focus was on if the physical movement captured was consistent with dispenser usage or not.

Once all the images in the training set were annotated, 5-fold cross validation was used to train the algorithm to detect hand hygiene dispenser use or nonuse.^{17,20,21} Five-fold cross-validation is an evaluation framework in machine learning whereby an algorithm is trained on a subset of the data (80% of the training set, 88 864 images) and then applied to the remaining data (20%, 22 216 images) to test its performance. The process is then repeated 4 more times (each time on a different 80% subset) in order to continuously refine the algorithm. Results from each round of cross-validation training and validation can be found in the [Supplementary Appendix](#).

Algorithm comparison to human observation to establish concordance

Because human observation remains the current gold-standard for the auditing of hand hygiene, human observation and labeling were used in 2 separate tasks. In this first human-centered task, trained observers individually audited hand hygiene dispenser use in-person at various dispenser locations on the unit of study for 2 hours. Results of human audits were compared with the algorithm's assessment of images from the sensors covering the exact same dispensers during the same time period to calculate concordance between human observation and the algorithm's determination. A total of 718 images demonstrating dispenser use or nonuse were obtained during the human observation period. This set of images served as the primary test set for our study.

Audits included observing for 2 events of interest as noted previously: hand hygiene dispenser use or nonuse upon room entry. For the purposes of our study, *concordance* referred to the ability of the algorithm to provide the same assessment of an event as a human observer recorded it. The level of agreement between the human audits and the algorithm's assessments was calculated using Cohen's

kappa. Cohen's kappa adjusts the concordance rate downward based on the rate of agreement expected by chance.

Determination of human and algorithm accuracy

In the absence of high-definition video to serve as a source of ground truth to compare human audit results with, the second human-centered task was the manual annotation of the entire test image set by 3 of the authors (A.S., W.B., and T.P.).^{17,22} This step involved human labeling and observation, which was separate from the concordance exercise noted previously. In this portion of the study, the 3 referenced authors independently annotated the set for events of dispenser use and nonuse. This process was "blinded" in that the results of the prior human observation were not accessible to the annotators. Interrater reliability was computed between the 3 authors using Fleiss's kappa.²³ Fleiss's kappa is the analog to Cohen's kappa when there are more than 2 raters, in that it adjusts the raw proportion of agreements for the proportion expected by chance. Any differences in labeling were resolved by adjudication by the 3 authors. The final result served as the source of ground truth to which both human and machine labels were compared via calculation of sensitivity and specificity of both the human observers and the algorithm.

RESULTS

A concordance rate of 96.8% was observed between human and machine for overall labeling of events with a kappa of 0.85 (95% confidence interval [CI], 0.77-0.92). Concordance among the 3 independent auditors to establish ground truth was 95.4%, with a Fleiss's kappa of 0.87 (95% CI, 0.83-0.91).

The machine learning algorithm showed a sensitivity and specificity of 92.1% (95% CI, 84.3%-96.7%) and 98.3% (95% CI, 96.9%-99.1%) in correctly identifying hand hygiene dispenser use and nonuse. Comparatively, human observations had a sensitivity and specificity of 85.2% (95% CI, 76.1%-91.1%) and 99.4% (95% CI, 98.4%-99.8%), respectively. The results of the labeling of dispenser use or nonuse by machine, human, and ground truth annotation are shown in [Table 1](#).

[Table 1](#) illustrates the results of labeling by both human observers and the machine algorithm as compared with the ground truth established by remote labeling by the 3 authors (A.S., W.B., and T.P.).

Table 1. Results of machine and human labeling

	Human observers	Machine	Ground truth
Dispenser used	79	92	88
Dispenser not used	639	626	630
Sensitivity, %	85.2 (95% CI, 76.1-91.1)	92.1 (95% CI, 84.3-96.7)	—
Specificity, %	99.4 (95% CI, 98.4-99.8)	98.3 (95% CI, 96.9-99.1)	—

CI: confidence interval.

DISCUSSION

Our results demonstrate that a computer vision algorithm is concordant with human audits in detection of hand hygiene dispenser usage or nonusage in a hospital setting. Its specificity was high and achieved a lower false-negative rate compared with human observation. Human auditing of hand hygiene is often done monthly for only a few hours at a time. Moreover, it is not often possible to cover an entire hospital unit covertly. Our study demonstrates that automated detection of hand hygiene dispenser use is a trustworthy assessment when compared with human observation as a gold-standard. Of note, during our observation period we noted more events of dispenser nonuse rather than use. However, given our experiment was not focused on calculating compliance rates, but rather to discern if a machine learning algorithm could correctly classify movements, both events (use and nonuse) were of equal value in capturing.

Real-time video surveillance and feedback has been studied in prior experiments with impressive improvement in hand hygiene compliance.⁷ However, it requires manual human review of real-time video data and as noted previously, video data may raise staff privacy concerns. Consequently, the ability of a machine learning algorithm as used in our study to take sensor data and in real-time proactively intervene with an output represents the next logical step in harnessing this technology. Our current work is focused on developing this real-time feedback loop. As a case example, a visual cue such as a red light surrounding a patient's bed or entry/exit threshold could alert a clinician if hand hygiene is not performed upon entering a patient room. This could further influence vital behavior that is not sustained at 100% anywhere in the world.² Extending beyond hand hygiene, one could combine depth sensor analysis of patient movements inside a hospital room with routine vital sign alarm data. The algorithm can fuse this data to learn when to silence inappropriate alarms, liberating bedside staff from "alarm fatigue," a known patient safety concern.²⁴

There are limitations to our study. First, annotation errors may have mislabeled events. We tried to mitigate this with multiple reviewers and discussions to resolve discrepancies. Second, depth sensors lack fine-grained detail. Color video is the optimal choice but raises privacy concerns for patients and staff. Depth sensors at LPCH only recorded data when triggered by motion, generating in certain cases only a few images versus full video, depending on how fast or slowly the subject moved. Third, our study was limited to one location and may not generalize to other hospitals or clinics without more detailed training of the algorithm on location-specific images. Fourth, our human observation time was only for a few hours in one day. However, this is similar to current practice for covert human observations and may not prove to be an accurate representation of the true number of hand hygiene events in a 24/7 care delivery system. Additionally, as this was a proof-of-concept study, we only had sensors installed on one unit to minimize disruption in

patient care. Installation of sensors throughout an entire hospital would take mobilization and coordination of many resources in units in which active patient care is taking place, possibly limiting active installation this type of technology in an existing hospital structure. Last, while the cost of performing the experiment on one unit may be relatively small for a hospital budget, it could prove more expensive to perform in an entire hospital building. At the time of our study, the cost of the sensors used and their associated hardware used was approximately \$50 000 USD. Additional expenses included time and labor for installation by the hospital facilities staff and engineers, which may vary substantially by institution. However, when compared with the expense of covert observers to provide a similar level of data as continuous and autonomous sensors, it is likely less expensive in the long term.

CONCLUSION

We demonstrate a novel application of a passive, economical, privacy-safe, computer vision-based algorithm for observing hand hygiene dispenser use. The algorithm's assessment is equal to the current gold-standard of human observation with improved sensitivity. Its capacity of continuous observation and feedback to clinicians may prove useful in efforts to mitigate a seemingly recurrent source of healthcare-induced harm.

AUTHOR CONTRIBUTIONS

All authors contributed to the study concept and design. Data acquisition, analysis, and interpretation of data were completed by AS, AH, SY, MG, and JRG. Drafting of the initial article was completed by AS and AH. Critical revision of the article was completed by all authors. Statistical analysis was performed by AS, AH, and JRG. The clinical study portion was supervised by AS and TP, with technical oversight provided by AH, SY, MG, AA, and LF-F.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

The authors thank the patients, families, and staff of the study unit for their cooperation and assistance in the study. The authors additionally thank the facilities and engineering staff of Lucile Packard Children's Hospital Stanford, who assisted with installation of the sensors. The authors also thank Mr. Zelun Luo, MS; Mr. Sanyam Mehra, MS; Ms. Alisha Rege, MS; and Ms. Lily Li with their assistance in annotation of the image dataset. Finally, the authors thank the hospital administrative staff and executive board for their approval, encouragement, and support in pursuing the study.

CONFLICT OF INTEREST STATEMENT

AA serves as the chief executive officer of VisioSafe Inc and provided the depth sensors used in this study. The remaining authors have no potential conflicts of interest, relevant financial interests, activities, relationships, or affiliations to report.

REFERENCES

- Centers for Disease Control and Prevention. Healthcare associated infections data. <https://www.cdc.gov/hai/data/index.html> Accessed April 2020.
- World Health Organization. *WHO Guidelines on Hand Hygiene in Health Care: First Global Patient Safety Challenge: Clean Care is Safer Care*. Geneva, Switzerland: World Health Organization and Patient Safety; 2009.
- Pineles LL, Morgan DJ, Limper HM, et al. Accuracy of a radiofrequency identification (RFID) badge system to monitor hand hygiene behavior during routine clinical activities. *Am J Infect Control* 2014; 42 (2): 144–7.
- Morgan DJ, Pineles L, Shardell M, et al. Automated hand hygiene count devices may better measure compliance than human observation. *Am J Infect Control* 2012; 40 (10): 955–9.
- Sharma S, Khandelwal V, Mishra G. Video surveillance of hand hygiene: a better tool for monitoring and ensuring hand hygiene adherence. *Indian J Crit Care Med* 2019; 23 (5): 224–6.
- Vaidotas M, Yokota PKO, Marra AR, et al. Measuring hand hygiene compliance rates at hospital entrances. *Am J Infect Control* 2015; 43 (7): 694–6.
- Armellino D, Hussain E, Schilling ME, et al. Using high-technology to enforce low-technology safety measures: the use of third-party remote video auditing and real-time feedback in healthcare. *Clin Infect Dis* 2012; 54 (1): 1–7.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316 (22): 2402–10.
- Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019; 25 (1): 24–9.
- Esteva A, Kuprel B, Novoa RA, et al. Corrigendum: dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 546 (7660): 686.
- Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; 318 (22): 2199–210.
- Koppula H, Saxena A. Learning spatio-temporal structure from RGB-D videos for human activity detection and anticipation. In: *International Conference on Machine Learning*; 2013: 792–800.
- Sung, J, Ponce C, Selman B, Saxena A. Unstructured human activity detection from RGBD images. In: 2012 IEEE International Conference on Robotics and Automation; 2012: 842–9.
- Xia L, Gori I, Aggarwal JK, et al. Robot-centric activity recognition from first-person RGB-D videos. In: 2015 IEEE Winter Conference on Applications of Computer Vision; 2015: 357–64.
- Rehder E, Wirth F, Lauer M, et al. Pedestrian prediction by planning using deep neural networks. In: 2018 IEEE International Conference on Robotics and Automation (ICRA); 2018: 5903–8.
- Chen J, Cremer JF, Zarei K, et al. Using computer vision and depth sensing to measure healthcare worker-patient contacts and personal protective equipment adherence within hospital rooms. *Open Forum Infect Dis* 2016; 3 (1): ofv200.
- Awwad S, Tarvade S, Piccardi M, et al. The use of privacy-protected computer vision to measure the quality of healthcare worker hand hygiene. *Int J Qual Health Care* 2019; 31 (1): 36–42.
- Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. <http://doi.org/10.1109/cvpr.2017.243>
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521 (7553): 436–44.
- Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Vol. 2; 1995: 1137–43. <https://dl.acm.org/doi/10.5555/1643031.1643047>
- Haque A, Guo M, Alahi A, et al. Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. *Mach Learn Healthcare* 2017; 68: 75–87.
- Yeung S, Rinaldo F, Jopling J, et al. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. *NPJ Digit Med* 2019; 2 (1): 11.
- Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. London, United Kingdom: Wiley; 2013.
- The Joint Commission. *National Patient Safety Goal on Alarm Management*. http://www.jointcommission.org/assets/1/18/JCP0713_Announce_New_NSPG.pdf Accessed April 14, 2020.