2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

An Intelligent Augmented Lifelike Avatar App for Virtual Physical Examination of Suspected Strokes

Kevin Yao, Kelvin K. Wong, Xiaohui Yu, John Volpi, and Stephen T. C. Wong*

Abstract— **An intelligent-augmented lifelike avatar mobile app (iLAMA) that integrates computer vision and sensor readings to automate and streamline the NIH Stroke Scale (NIHSS) physical examination is presented. The user interface design is optimized for elderly patients while the app showcases an animated lifelike 3D model of a friendly physician who walks the user through the exam. The standardized NIHSS examination included in iLAMA consists of five core tasks. The first two tasks involve rolling the eyes to the left and then to the right, and then smiling as wide as the user can. The app determines facial landmarks and analyzes the palsy of the face. The next task is to extend the arm and hold the phone at the shoulder level, and the smart phone gyroscope is used to detect acceleration to determine possible weakness in the arm. Next, the app tracks the location of the hand keypoints and determines possible ataxia based on the precision and accuracy of the locations of the touches. Finally, the app determines the user's forward acceleration in walking and possible imbalances using the accelerometer. The app then sends analyzed results of these tasks to the neurologist or stroke specialist for review and decisions.**

Clinical Relevance— **The physical examination of a stroke patient is a time consuming and repetitive process, and there is a lack of infrastructure and resource to monitor patient in post-stroke recovery after they leave the hospital for home or rehabilitation facilities. iLAMA app aims to automate a subset of the NIHSS physical examinations in measuring motor function recovery and also allows individual patients to track their performance over time. It will be an essential component in monitoring rehabilitation recovery and therapy effectiveness after hospitalization and can easily scaled to lo help millions of patients at a fraction of the cost.**

I. Introduction

Artificial Intelligence (AI) powered telemedicine—including virtual diagnostics, checkups, and physical exams through apps—have become widespread, especially during the COVID-19 pandemic. Among telemedicine apps being developed, apps that examine for signs of cerebrovascular accident or stroke are of particular research interest, as strokes are the second leading cause of death and the leading cause of serious long-term disability worldwide [1]. Stroke survivors are often left with any combination of speech, vision, or motor deficits, which can be improved or remedied through physical therapy. Progress tracking is needed in determining the effectiveness of physical therapy and to guide timely clinical decision making. The National Institutes of Health Stroke Scale or NIH Stroke Scale (NIHSS) is a standard-of-care tool used by healthcare providers to objectively examine and quantify a patient's cognitive impairment and motor skills after a suspected stroke [2]. The NIHSS exam consists of eleven tasks that test for facial or arm palsy, ataxia, balance, aphasia, and consciousness. Doctors or certified stroke nurses ask the patient to talk, look left and right, smile, raise their arms, touch repeatedly between their nose and the doctor's finger, and slide their legs, among other tasks. It is clinically important to continue such examinations after the patient leaves the hospital, but currently there is no infrastructure nor resources to hire certified personnel in post-stroke rehabilitation [3]. To fill this opportunity, we developed a virtual NIHSS physical exam app that integrates computer vision and sensor readings to automate and streamline part of the NIHSS exam. To our knowledge, such an intelligence-augmented app for physical exams has not been previously developed. Current developments on AI stroke detection apps focus on applying computer vision to support the initial detection of stroke, often through analysis using the F.A.S.T. test (facial palsy, arm weakness, and speech difficulty) [4-9]. These apps are meant to support ER doctors in emergency situations of possible stroke for pre-hospital self-diagnosis. However, they are much less likely to be used due to the intensity of the situation as ER doctors are rushed to provide care and limited in time to test experimental apps. In contrast, iLAMA app is meant to be used in post-stroke patient in order to automate the repetitive and time-consuming tasks of performing the NIHSS physical exam to monitor patient status, especially during lengthy rehabilitation months. The low-stress and repetitive nature of the work would make doctors more lenient towards relying on automation. We also envision that people with a high risk of repeated stroke could use iLAMA app regularly to objectively detect a gain/loss of motor functions based on their own physical exam history. The physical exam history can also serve as objective evidence to be considered by emergency room physicians, especially in determining whether a motor deficiency was newly developed from the stroke or present previously.

II. Methods

App Design Considerations

The target audience of the proposed app is post-stroke patient population, which is primarily patients 60 years or older. We made specific design choices, such as incorporating a talking 3D avatar of a physician to facilitate better patient-doctor interaction and to help the elderly visually understand on how to perform our tasks. In addition, we made text and buttons bigger, and every page can be reached in a maximum of three clicks. The five core NIHSS physical exam tasks are performed in order: 1. Roll your eyes

K. Yao, K. Wong, X. Xu, J. Volpi, and S. T. C. Wong (e-mail: stwong@houstonmethodist.org) are with the T.T. & W.F. Chao Center for BRAIN and Houston Methodist Cancer Center, Houston Methodist Hospital, Houston, TX 77030 USA.

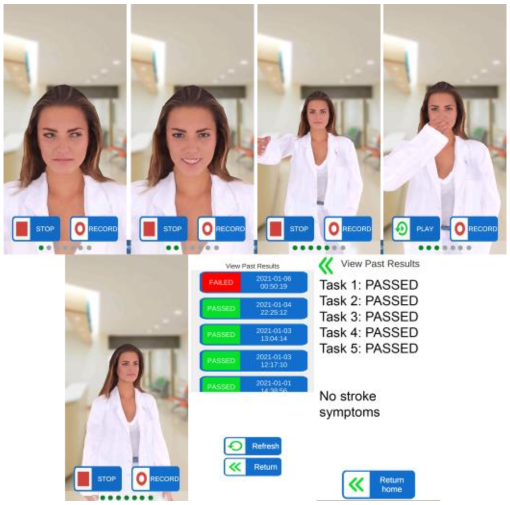


Figure 1. Sample images from app. Left to right: the five NIHSS tasks performed by the lifelike 3D avatar, view medical history, and view details.

to the right and then to the left (NIHSS test 2), 2. Smile as wide as you can and show your teeth (NIHSS test 4), 3. Raise your arm and hold it at shoulder level for ten seconds, repeating for each arm (NIHSS test 5), 4. Touch your nose three times, repeating for each arm (custom NIHSS test 7), and 5. Walk forward in a straight line (custom NIHSS test 7). To analyze the data, which consists of facial video from front-camera (tasks 1, 2, 4) and sensor (tasks 3, 5) data, we applied computer vision algorithms from Google's Mediapipe to recognize for body parts detection and use the supervision by registration (SBR) method [10] to extract 68 facial landmarks from the facial videos. We also archive the results of previous virtual examinations. The user or doctor is able to access those results anytime and view patient history or track patient recovery through the scores.

An overview of the tasks and the user interface is shown in Figure 1. At each task, the user has the option to press the play button, which will play audio instructions and an animation of the task for the user to perform and mirror. Text instructions at each step is available using the (i) button. The user has the option of viewing all of their past results, which indicates if they failed any of the five tasks when they took that test. Each task may have multiple sub-scores ranging from 0 to 100, where 100 represents a perfect score and 0 represents a completely failed task. A task is "PASSED" if the user scores a zero on the NIHSS scale, otherwise the NIHSS score is shown. The algorithm calculates an overall risk based on the NIHSS scale.

Task 1: Eyes

For the first task, the app collects a MP4 video of the user first looking directly at the camera, and then rolling his/her eyes to the left (L) and right (R). This task is analogous to task 2 on the NIHSS scale, which accesses the severity of gaze palsy. To analyze the user's video, we first detect the facial landmarks using a face landmark detector based on supervision by registration (SBR) [10] and use the eye keypoints for analysis. To find the position of the pupil, we process the eye to a pure black-and-white image and approximate the centroid of the black iris to be the pupil location. We then calculate scores for the R eye looking in the middle (RM), L eye looking in the middle (LM), R eye looking L (RL), R eye looking R (RR), L eye looking L (LL), and L eye looking R (LR). A high LL score is given if the pupil in the L eye is able to move far left relative to the width of the eyes. A similar method is used to assign scores for the other metrics. Our algorithm is then able to diagnose partial gaze paresis (PGP) if an eye experiences an inability to move far left or right (but not both), total gaze paresis (TGP) if an eye is unable to move from the center, and forced deviation (FD) if an eye only looks in one direction but cannot move. A PGP is ranked 1 on the NIHSS while TGP and FD are ranked 2.

Task 2: Mouth

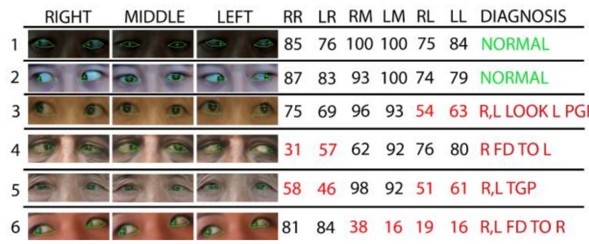
Task 2 involves smiling wide and showing teeth, with the jaw clenched. This is analogous to the examination of lower face palsy in task 4 of the NIHSS examination. The same face landmark detector [10] is used to detect the mouth keypoints. We detect that the mouth is opened when the ratio of the mouth height to the mouth width is between 0.15 and 0.4. To quantify the symmetry of the smile, we reflect one side of the mouth landmark points across the line connecting the center of the eyes to the bottom jaw. We then calculate the distance between this reflected point and the corresponding point on the other side of the face. The distance is then normalized by the width of the mouth (to account for differences in magnification), and all of these normalized distances are summed together. The score for the entire video is the average of these summed distances (ASD). A higher ASD represents greater asymmetry. Two thresholds are set to classify the palsy as minor (score of 1), partial (score of 2), or complete (score of 3). Note that this threshold value can easily change, and that the exact score is not of clinical significance: rather, the relationship between previous scores and the current score can indicate if a change from baseline has occurred.

Task 3: Hold Arms at Shoulder Level

Task 3 consists of holding each arm at shoulder level for ten seconds. This is analogous to tasks 5 and 6 on the NIHSS exam. Unity has a gyroscope module that allows the smart phone to collect the rotation speed (ω_x, ω_y, ω_z) and acceleration with the effect of gravity removed (a_x, a_y, a_z) at any time in the x, y, and z directions. We calculate the magnitude of rotation and acceleration as $\sqrt{\omega_x^2+\omega_y^2+\omega_z^2}$ and $\sqrt{a_x^2+a_y^2+a_z^2}$ respectively. Two scores are then calculated, the gyro score and the acceleration score. A score of 100 is assigned when the magnitudes of rotation or acceleration is close to 0, and a score of 0 is calibrated to the magnitudes when the hand is dropped in free fall. We then analyzed the lowest of these two scores. We set three thresholds to classify if the patient is normal (score 0), has minor weakness (score 1), has weakness but still has effort against gravity (score 2), or no weakness against gravity (score 3).

Task 4: Touch Nose Three Times

Task 4 consists of touching the nose three times with each arm, which mimics the standard NIHSS test 7 of touching the examiners finger and then touching your nose repeatedly. We first used Google's Mediapipe library to detect the hand keypoints [11]. We then calculated the position of each touch by determining the positions where the velocity of the index finger keypoint was close to zero. The nose keypoint was determined using the facial landmarks. We then calculated the center score, where high scores are given to touches close to



	RIGHT	MIDDLE	LEFT	RR	LR	RM	LM	RL	LL	DIAGNOSIS
1				85	76	100	100	75	84	NORMAL
2				87	83	93	100	74	79	NORMAL
3				75	69	96	93	54	63	R,L LOOK L PGP
4				31	57	62	92	76	80	R FD TO L
5				58	46	98	92	51	61	R,L TGP
6				81	84	38	16	19	16	R,L FD TO R

Figure 2. Table showing the results of the algorithm on various cases of gaze palsy. The first two letters (R, L, or both R, L) indicate which eye is affected. If the patient is experiencing a partial gaze paresis (PGP), we specify look R or look L depending on which direction the patient is having trouble moving their eyes to. If an eye is experiencing total gaze paresis (TGP), no further specification of direction is needed. If the patient is experiencing forced deviation (FD), we specify if the FD is to the R or L.

Figure 3. Task 2 example case where the user made a symmetrical smile and an asymmetrical smile. The average of the summed distances (ASD) score is significantly different.

the nose keypoints (quantifies accuracy), and the closeness score by calculating the variance between the three touches (quantifies precision).

Task 5: Walk in straight line

In task 5, the user is asked to walk in a straight line for 10 seconds. This is also a part of task 7 in the NIHSS exam, which tests for the ataxia of the lower body. While the user walks, we collect acceleration and rotation data from the gyroscope. To analyze the sensor data, we first detect if there is an initial acceleration in the first 2 seconds of recording that indicates the start of walking by determining if the moving average acceleration passes a certain threshold. Our reasoning behind using a moving average is because it averages the noise from the gait, allowing us to distinguish larger-scale trends in user acceleration. We also monitor the number of steps using the sensormotion package in python. If initial acceleration is detected and we detect at least five steps, we run a moving average analysis of the x acceleration of the phone to determine swaying in the walk that is convoluted with noise from natural gait. If either initial acceleration is not detected or if less than five steps were detected, the user is deemed to be unable to walk, and scored a 2 on the NIHSS scale. Otherwise, we set two thresholds based on the sum of squares (SS) moving-average acceleration to classify the sway as mild (score=1) or severe (score=2).

Beta testing of iLAMA app

To ensure the proper functioning of the app, we conducted comprehensive beta testing of the app by extended friends and families (normal volunteers) by open invitation through email. This is a critical process in making sure the proper functioning of the app at different lighting conditions for people of different color and ethnicities before the app can be evaluated in post-stroke patients. We asked for electronic consent from beta testers upfront in the app for facial video and sensor data to be recorded and send to our server for analysis. In total, we tested on 140 videos for all tasks.

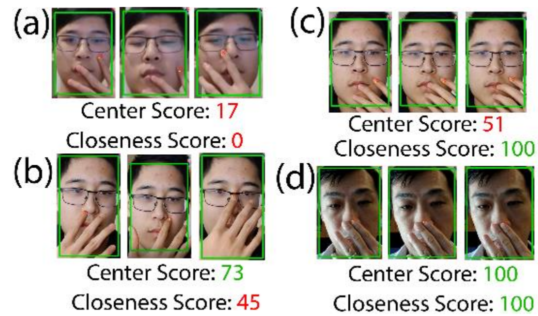


Figure 4. Results from task 4. (a-d) Scores from touching the nose with no precision or accuracy, some accuracy but no precision, precision but no accuracy, and accurately and precisely, respectively. The dot on the index finger indicates its predicted location.

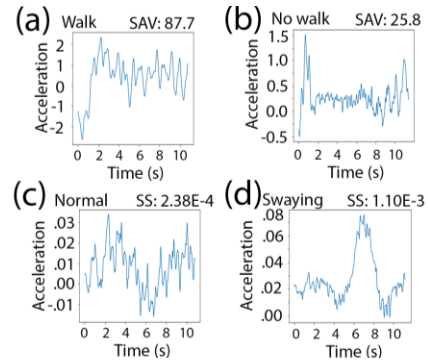


Figure 5. Results for task 5. (*a*) and (*b*) show acceleration in the y direction as a function of time for a user who walks normally and a user who shakes the phone but does not walk, respectively. (*c*) and (*d*) show acceleration in the x direction as a function of time for a user who walks normally and a user who sways to the side.

III. Results

Our results for task 1 are shown in Figure 2. We arbitrarily set the threshold score as 65 to differentiate between a good score (>65) and bad score (<65). The first two rows in Figure 2 are healthy controls. In the case of the third row (taken from video data on a L and R eye PGP patient taken from Ref. [12]), the RL and LL <65, which tells us that both the R and L eyes experience paresis looking L. The video data for the patient in the fourth row of Figure 2 was found on google images, presenting with an R eye FD to the L, which was correctly predicted. The image data for the fifth row was taken from Ref. [13] as an example of a patient with TGP. Finally, the data for the sixth row was taken from google as an example of a patient with FD to the R. Both of these were correctly predicted by our algorithm. The scores can differentiate the extent and type of the patient's gaze palsy.

Figure 3 shows a test of the facial paralysis algorithm in task 2. The outputs of the reflected landmarks are shown in red, and the original outputs of the landmark detector are shown in white. The numbers below the images show the ASD score. Note that the ASD score for the face with the asymmetrical smile is over three times as high as the ASD score for the face with the symmetrical smile.

Figure 4 shows the results of our algorithm for task 4. Each analysis was done on the video image of a phone camera, and each image in Figure 4 (a-d) is a single frame that our algorithm predicted was a part of one of the three nose touches. Figure 4 (a-d) shows an example of an imprecise and inaccurate test, an imprecise but relatively accurate test, a precise but inaccurate test, and a precise and accurate test. The center score and closeness score match with what is expected, showing that our methods are effective.

Figure 5 shows the results for task 5. Figures 5 (a) and (b) show moving average filtered plots of the y axis acceleration from a user who was walking normally and a user who shook the phone to mimic walking but did not walk, respectively. The sum of squares (SS) for the first two seconds for the user who walked (87.7) is three times larger than that of the user who pretended to walk by shaking the phone (25.8), indicating that our algorithm is capable of detecting the initial acceleration of walking. Figures 5 (c) and (d) show moving average filtered plots for users who walked in a straight line and who swayed while walking, respectively. The SS for Figure 5 (d) is five times larger than (c), indicating that this method of scoring the patient is able to distinguish if the gait has a net sway in the horizontal direction.

IV. Discussion

Our current approach may be limited by the accuracy of the facial or hand landmark models under certain subtle stroke conditions or unfavorable ambient conditions, especially from occlusion, poor lighting, or non-frontal angles. Our ongoing work is thus investigating non-video based analysis of facial palsy, such as through facial sensors or other physical approaches. In addition, our analysis requires that the users perform the tasks according to the NIHSS defined procedures. Small innocuous mistakes may be interpreted as signs of stroke, such as a natural unpredictable imbalance while walking, especially in unsupervised environments such as patient self-diagnosis at home. It is difficult to distinguish these small deviations from instructions from real signs of stroke. Thus, observation of common misinterpretations of instruction in clinical trials is necessary as well as a careful analysis of natural mistakes from stroke signs. Finally, the entire screening process takes about five minutes. This length of time is the tradeoff to get the most comprehensive snapshot of a patient's status at a given time, since the app is not meant to be used in acute situations. However, some users may find finishing the entire test on a regular basis to be too cumbersome, so future work will explore faster testing algorithms, such as combining multiple tasks together to expedite the process.

We have developed an intelligent-augmented mobile health app, iLAMA, that screens patients for suspected stroke based on the standard-of-care NIHSS examination. The iLAMA app features a lifelike, 3D avatar doctor that guides the user through five required NIHSS tasks, including moving the eyes to the left and right, smiling, raising the arms, touching the nose, and walking. We deployed a combination of computer vision algorithms and embedded phone sensor filters to analyze those tasks and score each of them on the NIHSS scale. This intelligent augmented app serves as a blueprint for the next generation of telemedicine apps that perform virtual physical examinations, that is, without the presence of a physician or disease specialist. Such intelligent augmented apps would be useful in continuing the line of care and monitoring after a patient is released from the hospital while still facilitating a healthy and productive patient-doctor interaction with a consistent scoring performance. We are planning a clinical trial to validate the function and design of the iLAMA app at Houston Methodist for post-stroke in patients before further testing it in post-stroke rehabilitation facilities.

Acknowledgment

The authors would like to acknowledge T.T. & W.F. Chao Foundation and John S Dunn Research Foundation for their funding support. We appreciate Dr. Rebecca Danforth for proofreading the manuscript.

References

[1] George MG, Fischer L, Koroshetz W, Bushnell C, Frankel M, Foltz J, et al. CDC grand rounds: public health strategies to prevent and treat strokes. MMWR Morbidity and mortality weekly report. 2017;66(18):479.

[2] National Institutes of Health Stroke Scale (NIHSS). Wiley StatsRef: Statistics Reference Online.

[3] Hartford W, Lear S, Nimmon L. Stroke survivors' experiences of team support along their recovery continuum. BMC health services research. 2019;19(1):1-12.

[4] Gupta A. StrokeSave: A Novel, High-Performance Mobile Application for Stroke Diagnosis using Deep Learning and Computer Vision. arXiv preprint arXiv:190705358. 2019.

[5] Yu M, Cai T, Huang X, Wong K, Volpi J, Wang JZ, et al., editors. Toward Rapid Stroke Diagnosis with Multimodal Deep Learning2020; Cham: Springer International Publishing.

[6] Chien T-Y, Chen C-Y, Jin G-L, editors. A Stroke Detection System Based on Cincinnati Prehospital Stroke Scale. Proceedings of the third International Conference on Medical and Health Informatics 2019; 2019.

[7] Simionescu C, Insuratelu M, Herscovici R. Prehospital Cerebrovascular Accident Detection using Artificial Intelligence Powered Mobile Devices. Procedia Computer Science. 2020;176:2773-82.

[8] García L, Tomás J, Parra L, Lloret J. An m-health application for cerebral stroke detection and monitoring using cloud services. International Journal of Information Management. 2019;45:319-27.

[9] Phienphanich P, Tankongchamruskul N, Akarathanawat W, Chutinet A, Nimnual R, Tantibundhit C, et al., editors. Automatic stroke screening on mobile application: Features of gyroscope and accelerometer for arm factor in FAST. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2019: IEEE.

[10] Dong X, Yu S-I, Weng X, Wei S-E, Yang Y, Sheikh Y, editors. Supervision-by-registration: An unsupervised approach to improve the precision of facial landmark detectors. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018.

[11] Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, Sung G, Chang C-L, et al. MediaPipe Hands: On-device Real-time Hand Tracking. arXiv preprint arXiv:200610214. 2020.

[12] Kim J-S, Jeong S-H, Choi J-Y, Kim H-J. Teaching Video NeuroImages: Palsy of conjugate horizontal gaze and face due to isolated abducens nuclear infarction. Neurology. 2017;89(14):e180-e1.

[13] Lee E, Kim JS, Kim JS, Song HS, Kim SM, Kwon SU. A small dorsal pontine infarction presenting with total gaze palsy including vertical saccades and pursuit. Journal of Clinical Neurology. 2007;3(4):208-11.