



OPEN

Artificial intelligence application versus physical therapist for squat evaluation: a randomized controlled trial

Alessandro Luna¹, Lorenzo Casertano², Jean Timmerberg^{1,3}, Margaret O'Neil¹, Jason Machowsky⁴, Cheng-Shiun Leu⁵, Jianghui Lin⁵, Zhiqian Fang^{5,6}, William Douglas² & Sunil Agrawal⁷✉

Artificial intelligence technology is becoming more prevalent in health care as a tool to improve practice patterns and patient outcomes. This study assessed ability of a commercialized artificial intelligence (AI) mobile application to identify and improve bodyweight squat form in adult participants when compared to a physical therapist (PT). Participants randomized to AI group (n = 15) performed 3 squat sets: 10 unassisted control squats, 10 squats with performance feedback from AI, and 10 additional unassisted test squats. Participants randomized to PT group (n = 15) also performed 3 identical sets, but instead received performance feedback from PT. AI group intervention did not differ from PT group (log ratio of two odds ratios = -0.462, 95% confidence interval (CI) (-1.394, 0.471), p = 0.332). AI ability to identify a correct squat generated sensitivity 0.840 (95% CI (0.753, 0.901)), specificity 0.276 (95% CI (0.191, 0.382)), PPV 0.549 (95% CI (0.423, 0.669)), NPV 0.623 (95% CI (0.436, 0.780)), and accuracy 0.565 (95% CI (0.477, 0.649)). There was no statistically significant association between group allocation and improved squat performance. Current AI had satisfactory ability to identify correct squat form and limited ability to identify incorrect squat form, which reduced diagnostic capabilities.

Trial Registration NCT04624594, 12/11/2020, retrospectively registered.

Artificial intelligence technology (AI) is a general term to describe computers exhibiting human-like intelligence and reason¹. AI is becoming more prevalent in health care as a tool to improve practice patterns and patient outcomes. AI holds promise in that it can be used to create programs to replicate complex cognitive tasks to assist clinicians in patient management. Examples include analysis of imaging data for diagnosis of cancer and heart disease; extraction of unstructured data from electronic medical records for evaluation; and creation of socially assistive robot exercise coaching for older adults¹⁻³. These examples demonstrate that AI can be integrated into software systems such as electronic medical records and hardware systems such as robots or devices.

In this study, AI is used to examine bodyweight exercise performance. Benefits of exercise impact multiple areas of health and disease including dementia risk, cardiovascular and musculoskeletal disorders, and obesity⁴⁻⁷. The AI exercise mobile application (app) used for this research is built with patent pending motion tracking technology which monitors and provides real-time audiovisual feedback on a person's exercise performance. The technology relies on mobile phone video capture capability and does not require any additional equipment.

This app benefits from independent operation. Users do not need another person to control the app, and the user does not need to wear additional sensors. Other applications may provide audiovisual instructions, but do not provide corrective feedback when exercises are performed incorrectly⁸. Additionally, a previous trial demonstrated the effectiveness of the AI app in treating lower back pain⁹. This trial tested how the app identifies and

¹Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. ²NewYork-Presbyterian/Columbia University Irving Medical Center, New York, NY, USA. ³Columbia University Irving Medical Center, New York, NY, USA. ⁴Hospital for Special Surgery, New York, NY, USA. ⁵Department of Biostatistics, Columbia University, New York, NY, USA. ⁶Department of Psychiatry, University of Hong Kong, Pok Fu Lam, Hong Kong. ⁷Departments of Mechanical Engineering and Rehabilitation Medicine, Columbia University, 500 W. 120th Street #510, New York, NY 10027, USA. ✉email: sunil.agrawal@columbia.edu

improves bodyweight squats when compared to feedback from a physical therapist since this particular exercise is considered a foundational and compound movement used in activities of daily living^{10,11}.

Methods

Study design. In this randomized, blinded, controlled trial, 30 participants were randomly assigned to either the AI (n = 15) or PT (n = 15) group. The local institutional review board approved the study, and all participants provided written informed consent. Trial number NCT04624594 retrospectively registered on 12/11/2020 (See Supplementary Files 1 and 2).

Participants. Research population was academic institution affiliates, age 20–35, without any preexisting medical condition that precluded them from participating in bodyweight exercise for 10 min. Participants could withdraw at any time and were not paid to be in the study. Person-to-person recruitment and flyers were used in October 2019. Participants were randomly assigned to either the AI or PT group in a 1:1 ratio using the random choice selection function in Excel. Participants were also assigned a unique identifier number to sign up for a time slot. Both the AI and PT groups had 7 female and 8 male participants.

Squat definition. Based on pre-existing squat literature descriptions and published squat best practices^{12–18}, the PT and three independent evaluators collectively agreed on this study's official squat definition:

Individual starts in a standing position with feet flat on the floor, knees and hips in a neutral, extended anatomical position, spine in an upright position with preservation of its natural curves, and hands held in front of body. Squat movement begins with descent phase initiated by “sitting back” as hips, knees, and ankles flex simultaneously. Individual should descend until hip joint becomes level with knee joint, without letting the knees extend past toes. Ascent is achieved through simultaneous extension of the hips, knees, and ankles, continuing until the subject has returned to starting position.

Intervention. On assigned day and time, each participant reported to campus research space individually for 15 min to maintain anonymity. The first 5 min were reserved for a presentation of key information about the trial. Participants had the opportunity to discuss the information before volunteering their written informed consent.

In the subsequent 10 min, participants were observed by the AI (operating from an iPhone X) and PT from the sagittal right plane 3 m away. A video camera simultaneously recorded the exercises at this same position. Additional supervising researcher was always present for added safety. For standardization, all participants were instructed to keep hands in front of their body, squat until knees flexed to 90 degrees, and maintain a cadence of 2 s for the descent phase and two seconds for the ascent phase.

The PT providing feedback had more than a decade of training and practice in neurorehabilitation and extensive experience training athletes. The PT was provided a standardized list of corrections based on the common AI evaluations, and was also free to provide any necessary feedback not contained in the list. The common AI evaluations included: body leaning too far to the front, not squatting deep enough (<90°), squatting too deep (>90°), knees extending past toes, neck extended too far upwards, neck flexed too far downwards, motion was too fast, and motion was too slow.

All participants performed 10 bodyweight squat “control” repetitions without feedback followed by one minute of rest. Those in the AI group then performed 10 more “practice” repetitions with real-time audiovisual feedback from the app followed by 1 min of rest. The AI's design provided one piece of feedback, if necessary, with a voice statement and on-screen video per repetition (e.g. when participant performed squat repetition with their neck flexed downward, AI suggested keeping their head up with on-screen instruction). Those in the PT group (n = 15) also performed 10 “practice” repetitions with one piece of feedback per repetition, if necessary, from the PT. Participants in both groups then performed 10 “test” repetitions without feedback.

Outcomes. Filmed repetitions from the “control” and “test” sets were scored by three independent evaluators as correct or incorrect. They were blinded to participant group and squat set. If repetition was incorrect, a one sentence justification was provided; they used the aforementioned standardized correction list as a guide and were free to make other form corrections they deemed necessary. If at least two panel evaluators scored a squat as correct, the “majority score” was correct, and vice versa for incorrect squats. The outcome of the intervention for a given participant was considered a “success” if participant had more correct than incorrect squats after intervention.

Statistical analysis. Graphical presentation and descriptive statistics were generated to show the frequencies and percentages of correct and incorrect squats in control and test sets as scored by AI and each evaluator. In addition, AI and evaluators also provided one piece of feedback per incorrect squat. Evaluators identified 8 additional corrections in addition to the standardized list: arms not out in front of chest, asymmetrical weight-bearing, incomplete extension on ascent, early lumbar spine motion, body leans too far back, trunk folds or torso bends, knee dominant movement, and torso-initiated movement. Graphical presentation and descriptive statistics were presented to report the feedback provided by AI and evaluators for incorrect squats.

Majority scores were used as the gold standard to calculate sensitivity (AI ability to identify a correct squat), specificity (AI ability to identify an incorrect squat), positive predictive value (PPV), negative predictive value (NPV), and accuracy of the scores determined by AI. Furthermore, each evaluator was also considered as the

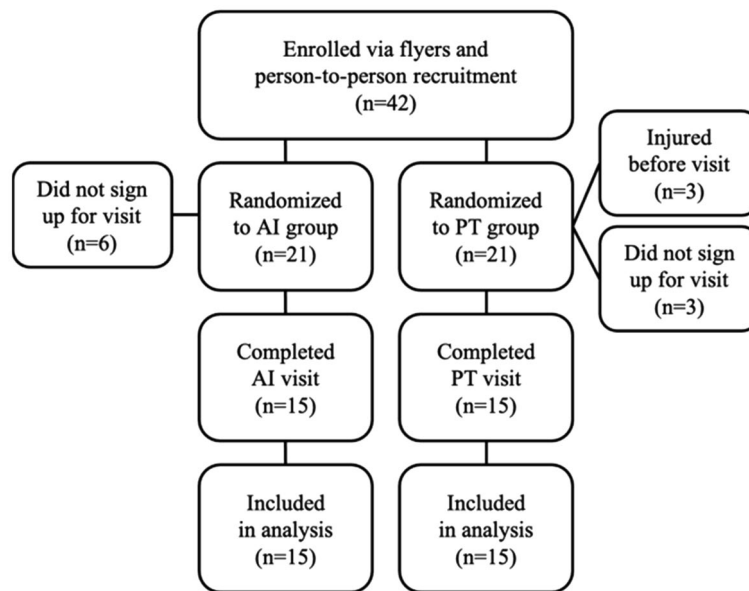


Figure 1. Participant flowchart. 42 people were eligible, but 6 people did not sign up for a time slot and 3 people were injured prior to participation.

gold standard to examine the performance of AI. Generalized Estimating Equations (GEE) were used to account for within subject correlation due to repeated measures when calculating the 95% confidence intervals for the above operating characteristics.

A generalized linear model was used to compare the change over time of the probability of doing correct squats between the AI and PT interventions. GEE with exchangeable correlation structure was employed to account for within subject correlations¹⁹. The statistical model included the intercept, post-intervention indicator (vs. pre-intervention), AI group indicator (vs. PT group), and the interaction between the two indicators. Gender was also included in the model to adjust for potential confounding. The regression coefficient corresponding to the interaction term represented the log ratio of two odds ratios and allowed comparison of AI and PT group intervention effect. The odds of “success” (i.e., more correct than incorrect squats after intervention) were compared between AI and PT group via logistic regression analysis. In addition, Light’s Kappa was used to evaluate inter-rater reliability of the three evaluators²⁰. Findings were declared statistically significant if $p \leq 0.05$. Analyses were performed in RStudio²¹.

Sample size calculation. The study was designed to detect a 40-percentage point difference (i.e., 65% vs. 25%, which we considered as a clinically important difference) with 80% power for a two-sided 0.05-level test. We conducted the sample size calculation assuming that each participant would complete 20 squats, having a within subject correlation (i.e., intra-subject correlation ICC) no greater than 0.7. The above calculation led to a recruitment of 15 subjects per group (30 in total) and required data with 300 observations per group (600 in total).

Ethics approval. Columbia University institutional review board approved study protocol AAAS7301 on October 15, 2019, which was performed in accordance with the standards defined in the 1964 Declaration of Helsinki.

Informed consent. Informed consent was obtained from all individual participants included in this study.

Results

Descriptive statistics. Data for 30 participants were collected (see Fig. 1). Correct and incorrect squats were tabulated for AI and the three evaluators (see Fig. 2). Of the 600 squat repetitions performed in the control and test sets, 307 (51.2%) received a majority score of correct and 293 (48.8%) a majority score of incorrect. The three evaluators completely agreed for 294 (49%) repetitions: 155 (25.8%) squats scored as correct and 139 (23.2%) squats scored as incorrect. The most common feedback provided by 2/3 evaluator majority was “squat too shallow (<90 degrees)” (11.3%). The most common feedback provided by AI was “neck extends too far upwards” (15%) (see Fig. 3).

Sensitivity, specificity, PPV, NPV, and accuracy. Operating characteristics of AI were reported with their 95% confidence intervals (see Table 1). AI was also compared with instances where panel of evaluators had 3/3 complete agreement instead of 2/3 majority agreement: sensitivity was 0.865 (95% CI (0.751, 0.931)), speci-

Squats in control and test sets as scored by AI and each evaluator

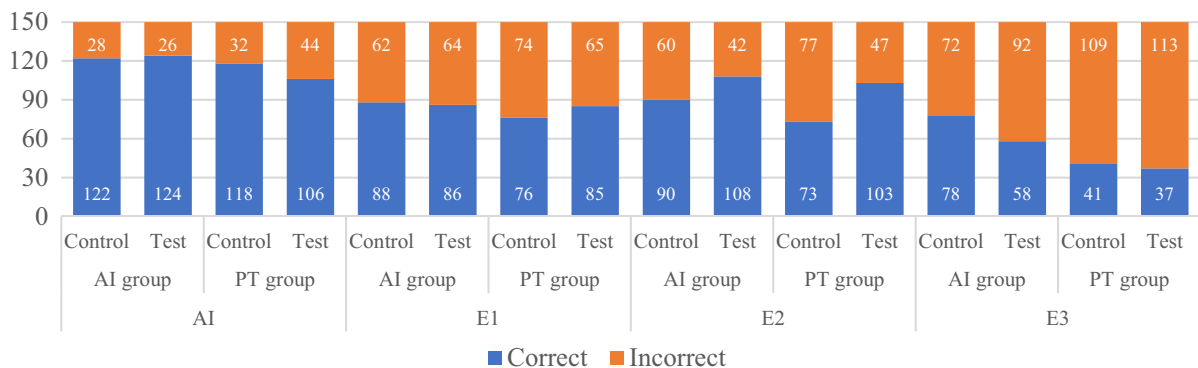


Figure 2. Correct and incorrect squats as scored by AI and evaluators (E1 = Evaluator 1, E2 = Evaluator 2, E3 = Evaluator 3). “Control” refers to the first set of 10 unassisted squat repetitions. “Test” refers to the third and last set of 10 unassisted squat repetitions performed by participants after receiving feedback in the second set.

ficity = 0.281 (95% CI (0.179, 0.411)), PPV = 0.573 (95% CI (0.390, 0.737)), NPV = 0.650 (95% CI (0.370, 0.855)), and accuracy = 0.588 (95% CI (0.466, 0.701)).

Comparison of AI versus PT intervention effects. Findings from the GEE analysis suggested that AI group intervention effect, in terms of change over time in the probability of a correct squat pre- versus post-intervention, did not differ from the PT group (log ratio of two odds ratios = -0.462 , 95% CI ($-1.394, 0.471$), $p = 0.332$). Proportion of participants with more correct squats after the intervention was greater in the PT group, but was not statistically significant at the 0.05 level (PT vs. AI: 60% vs. 27%, odds ratio = 4.125, 95% CI (0.883, 19.273), $p = 0.072$).

Inter-rater reliability. Light’s Kappa (weighted average of Cohen’s Kappa for each evaluator pair) for inter-rater reliability of the three evaluators scoring 600 squat repetitions was 0.337. Cohen’s Kappa for evaluator 1 and 2, 1 and 3, and 2 and 3 were 0.320, 0.266, and 0.319, respectively. In the subset of squats determined to be incorrect by 2/3 panel majority, Light’s Kappa for inter-rater agreement on the feedback provided for these incorrect squats was 0.407 (Supplementary Material).

Discussion

This trial was an independent university medical center evaluation of commercialized private sector technology to study the ability of an AI exercise mobile application to identify and improve bodyweight squat form in 30 adult participants. The GEE analysis revealed no statistically significant difference between AI and PT group on squat performance. While not statistically significant at $p < 0.05$, trends of these analyses suggested that PT intervention may have had favorable effects on squat improvement; PT group had above 4 times greater odds of having more correct squats when compare to AI group. Lack of statistical power may have been an issue for such an effect not attaining the 0.05 level of statistical significance.

The AI had satisfactory ability to identify correct squat form as evidenced by its sensitivity values (see Table 1) which are comparable with each individual evaluator and the collective panel²². Conventional motion-tracking systems use multiple high-speed cameras with ground force plates or wearable inertial measurement units²³. When comparing the present AI data to previous studies validating these conventional systems, the AI sensitivity matched or exceeded these systems for squat movements^{23–28}. When compared with existing systems, a clear benefit of the AI mobile application is the absence of complex machinery or expensive wearable sensors for functioning. However, the AI specificity to identify incorrect squats was insufficient and attributed to the low accuracy of 0.565 (95% CI 0.524–0.605).

Descriptive statistics indicated possible factors contributing to the low accuracy. AI only identified squats that were deemed too shallow (< 90 degrees) 12 times while the panel majority provided this feedback 68 times. Additionally, the most common AI feedback for incorrect squats was “neck extended too far upwards”, a correction provided 90 times by the AI and only 2 times by the panel. These under-corrections and over-corrections may be sources of diagnostic error that explain the equivocal PPV, NPV, and low specificity despite satisfactory sensitivity. Another source of diagnostic discrepancy was the panel’s ability to identify eight additional corrections beyond the standardized list. These include “incomplete extension on ascent” and “asymmetrical weightbearing” (see Fig. 3). The AI may be limited currently in its capacity to identify subtle changes in three-dimensional space in comparison to the evaluators.

Inter-rater reliability merits further explanation. All three evaluators provided input and established the working definition for a correct squat; received the same standardized list of corrections; performed their analyses blinded to set number and group allocation; and attained at least a decade of experience in physical therapy and exercise instruction, yet the homogeneity of evaluators was not reflected in the heterogeneity of IRR calculations. Although the three evaluators completely agreed for 294 (49%) repetitions, Light’s Kappa was 0.337 and is interpreted as minimal agreement²⁹. The operating characteristics (i.e., sensitivity, specificity, PPV, NPV, and

Feedback in control and test sets as provided by AI and evaluators

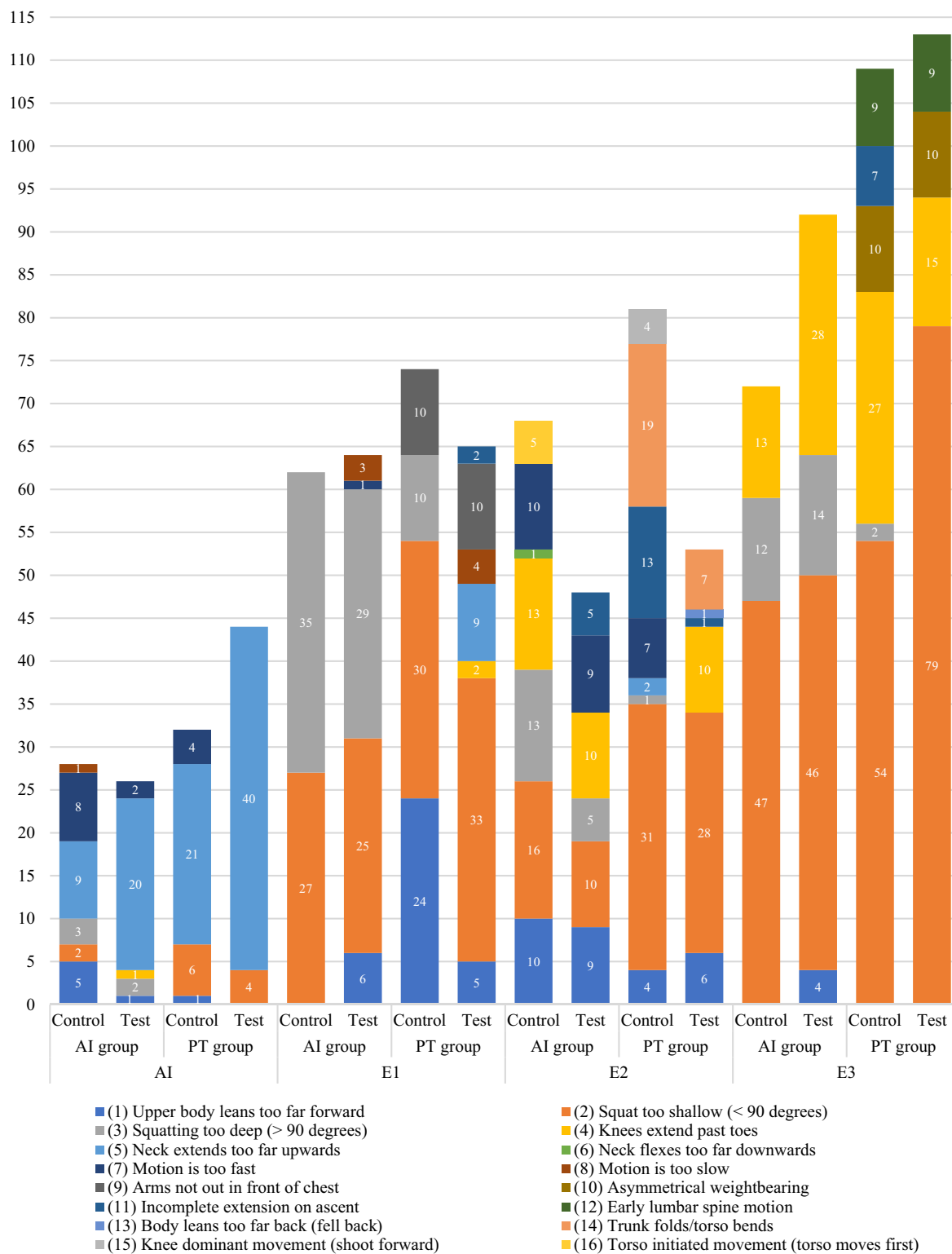


Figure 3. Feedback for incorrect squats as provided by AI and evaluators (E1, E2, E3).

accuracy) of comparing AI with instances where the panel of evaluators had 3/3 complete agreement fall within the confidence intervals of the original test characteristics calculated with 2/3 majority agreement. Such findings suggest that the consensus derived from panel majority maintains consistency with more stringent criteria. The test characteristics for AI versus each individual evaluator are comparable with the panel majority as well.

	AI vs. majority (control + test set)		AI vs. majority (control set)		AI vs. majority (test set)		AI vs. E1 (control + test set)		AI vs. E2 (control + test set)		AI vs. E3 (control + test set)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Sensitivity	0.840	0.753–0.901	0.871	0.770–0.931	0.812	0.697–0.891	0.812	0.727–0.875	0.826	0.746–0.885	0.846	0.748–0.910
Specificity	0.276	0.191–0.382	0.268	0.178–0.382	0.286	0.181–0.421	0.253	0.179–0.344	0.288	0.192–0.407	0.251	0.177–0.343
PPV	0.549	0.423–0.669	0.533	0.389–0.672	0.565	0.426–0.695	0.579	0.464–0.685	0.657	0.531–0.765	0.385	0.264–0.522
NPV	0.623	0.436–0.780	0.683	0.482–0.834	0.571	0.356–0.763	0.515	0.372–0.656	0.500	0.323–0.677	0.746	0.550–0.876
Accuracy	0.565	0.477–0.649	0.563	0.458–0.663	0.567	0.467–0.662	0.565	0.488–0.639	0.623	0.537–0.703	0.463	0.380–0.549

Table 1. Operating characteristics and 95% confidence intervals for AI versus evaluators. “AI vs. Majority” compares AI with panel majority of the 3 evaluators: E1, E2, and E3. “Control” refers to the first set of 10 unassisted squat repetitions. “Test” refers to the third and last set of 10 unassisted squat repetitions performed by participants after receiving feedback in the second set.

One risk of the AI ability to detect incorrect squats is patient exercise safety. While this study population was comprised of healthy individuals, patients with specific musculoskeletal rehabilitation requirements may be more vulnerable to errors in exercise form, and thus more likely to experience insufficient improvements or injury due to improper movement. Of note, there were no adverse events in the PT or AI group and the AI intervention was well tolerated by study participants. A distinct advantage of this evolving technology is the potential for more equitable dissemination of safe exercise coaching. In metropolitan areas, gym memberships can cost 20 to 100 USD per month, which does not always include costs associated with hiring personal trainers or enrolling in group fitness classes³⁰. For individuals who cannot afford or do not have access to these facilities and resources, or for those who prefer to exercise at home and in outdoor spaces with minimal equipment, the on-demand mobile app format is appealing as the AI technology advances.

Limitations

As expected in this healthy adult population, some participants may not have been entirely naive to the squat movement, which could have limited the ability of this intervention to demonstrate clinical improvement in squat outcomes. As the recruitment of participants was limited to academic institution affiliates, ages 20 to 35, and without any preexisting medical condition, further studies will be necessary to generalize the findings to patient populations or individuals with specific physical rehabilitation requirements. The AI and PT only evaluated squats in the sagittal plane and adding multiple views could change the accuracy of either evaluator. The low Light’s Kappa for feedback provided in the subset of incorrect squats could have been due to evaluators’ subjective interpretations of each participant’s anatomical variance; individual evaluators could have also focused on different aspects of the squat as more important at a single point in time.

Conclusions

While there was no statistically significant association between group allocation and improved squat performance, the current iteration of AI has satisfactory ability to identify correct squat form and is well-tolerated in a healthy adult population. However, the AI has limited ability to identify incorrect squat form, which reduces its diagnostic capabilities. Specific improvements could include enhanced recognition of squat depth and spine biomechanics via anatomical subtleties in three-dimensional spatial detection. Future research studies should consider expanding population demographics to include various levels of squat familiarity for identification and improvement of squat form.

Data availability

The datasets generated and analyzed during the current study are not publicly available to maintain privacy of participants; relevant de-identified data and statistical analyses are included in the manuscript.

Received: 1 June 2021; Accepted: 18 August 2021

Published online: 13 September 2021

References

- Bini, S. A. Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care? *J. Arthroplasty* **33**(8), 2358–2361 (2018).
- Jiang, F. *et al.* Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2**(4), 230–243. <https://doi.org/10.1136/svn-2017-000101> (2017).
- Hamet, P. & Tremblay, J. Artificial intelligence in medicine. *Metabolism* **69**, S36–S40 (2017).
- Fasola, J. & Mataric, M. A socially assistive robot exercise coach for the elderly. *J. Hum. Robot Interact.* <https://doi.org/10.5898/JHRI.2.2.Fasola> (2013).
- Hamilton, M. T., Hamilton, D. G. & Zderic, T. W. The necessity of active muscle metabolism for healthy aging: Muscular activity throughout the entire day. *Progr. Mol. Biol. Transl. Sci.* **155**, 53–68 (2018).
- Gopinath, B., Kifley, A., Flood, V. M. & Mitchell, P. Physical activity as a determinant of successful aging over ten years. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-28526-3> (2018).
- Jakovljevic, D. G. Physical activity and cardiovascular aging: Physiological and molecular insights. *Exp. Gerontol.* **109**, 67–74 (2018).

8. Dahlberg, L. E., Dell'Isola, A., Lohmander, L. S. & Nero, H. Improving osteoarthritis care by digital means—Effects of a digital self-management program after 24- or 48-weeks of treatment. *PLoS ONE* **15**(3), e0229783. <https://doi.org/10.1371/journal.pone.0229783> (2020).
9. Toelle, T. R., Utpadel-Fischler, D. A., Haas, K.-K. & Priebe, J. A. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: A randomized controlled trial. *NPJ Digit. Med.* **2**(1), 34 (2019).
10. Kritz, M., Cronin, J. & Hume, P. The bodyweight squat: A movement screen for the squat pattern. *Strength Cond. J.* **31**(1), 76–85 (2009).
11. O'Reilly, M. A., Whelan, D. F., Wassrd, T. E., Delahunt, E. & Caulfield, B. M. Technology in strength and conditioning. *J. Strength Cond. Res.* **31**, 2303–2312 (2017).
12. Myer, G. D., Ford, K. R. & Hewett, T. E. Rationale and clinical techniques for anterior cruciate ligament injury prevention among female athletes. *J. Athl. Train.* **39**(4), 352–364 (2004).
13. Escamilla, R. F. Knee biomechanics of the dynamic squat exercise. *Med. Sci. Sports Exerc.* **33**(1), 127–141 (2001).
14. Heijne, A. *et al.* Strain on the anterior cruciate ligament during closed kinetic chain exercises. *Med. Sci. Sports Exerc.* **36**(6), 935–941 (2004).
15. Schoenfeld, B. J. Squatting kinematics and kinetics and their application to exercise performance. *J. Strength Cond. Res.* **24**(12), 3497–3506 (2010).
16. Myer, G. D. *et al.* The back squat. *Strength Cond. J.* **36**(6), 4–27 (2014).
17. Swinton, P. A., Stewart, A. D., Lloyd, R., Keogh, J. W. L. & Agouris, I. A biomechanical comparison of the traditional squat, powerlifting squat, and box squat. *J. Strength Cond. Res.* **26**(7), 1805–1816 (2012).
18. Brocki, K. C. & Bohlin, G. Executive functions in children aged 6 to 13: A dimensional and developmental study. *Dev. Neuropsychol.* **26**(2), 571–593 (2004).
19. Shults, J. *et al.* A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Stat. Med.* **28**(18), 2338–2355. <https://doi.org/10.1002/sim.3622> (2009).
20. Hallgren, K. A. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant. Methods Psychol.* **8**(1), 23–34 (2012).
21. RStudio. RStudio (2020). <https://rstudio.com/>. Accessed 16 Jan 2020.
22. Power, M., Fell, G. & Wright, M. Principles for high-quality, high-value testing. *Evid. Based Med.* **18**(1), 5–10. <https://doi.org/10.1136/eb-2012-100645> (2013).
23. Renggli, D. *et al.* Wearable inertial measurement units for assessing gait in real-world environments. *Front. Physiol.* <https://doi.org/10.3389/fphys.2020.00090/full> (2020).
24. Giggins, O. M., Sweeney, K. T. & Caulfield, B. Rehabilitation exercise assessment using inertial sensors: A cross-sectional analytical study. *J. Neuroeng. Rehabil.* **11**(1), 158. <https://doi.org/10.1186/1743-0003-11-158> (2014).
25. O'Reilly, M., *et al.* Evaluating squat performance with a single inertial measurement unit. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. *IEEE*, 1–6 (2015). <http://ieeexplore.ieee.org/document/7299380/>.
26. Whelan, D. F., O'Reilly, M. A., Ward, T. E., Delahunt, E. & Caulfield, B. Technology in rehabilitation: Evaluating the single leg squat exercise with wearable inertial measurement units. *Methods Inf. Med.* **56**(2), 88–94 (2017).
27. Whelan, D., O'Reilly, M., Ward, T., Delahunt, E. & Caulfield, B. Evaluating performance of the single leg squat exercise with a single inertial measurement unit. In *Proc. 3rd 2015 Workshop on ICTs for Improving Patients Rehabilitation Research Techniques—REHAB'15*, 144–147 (ACM Press, 2015). <http://dl.acm.org/citation.cfm?doid=2838944.2838979>. Accessed 28 Aug 2020.
28. Whelan, D., O'Reilly, M., Ward, T., Delahunt, E. & Caulfield, B. Evaluating performance of the lunge exercise with multiple and individual inertial measurement units. In *Proc. 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM. <https://doi.org/10.4108/eai.16-5-2016.2263319> (2016).
29. McHugh, M. L. Interrater reliability: The kappa statistic. *Biochem. Med.* **22**(3), 276–282 (2012).
30. Zagnit, E. A., Rajan, S. & Basch, C. H. Prevalence and pricing of chain gyms in New York City. *Int. J. Health Promot. Educ.* **54**(1), 50–57. <https://doi.org/10.1080/14635240.2015.1069717> (2016).

Acknowledgements

Thank you to Kaia Health for providing the Motion Coach application (December 2019 version) used in this study. We are also grateful for the volunteers who participated in this research.

Author contributions

A.L. conceived and designed this study with S.A. and W.D. L.C. provided the PT group feedback. J.T., M.O'N. and J.M. evaluated the recorded exercises. C.-S.L., J.L. and Z.F. performed the statistical analyses. A.L. drafted the first manuscript. All authors reviewed the results and provided feedback on subsequent versions to generate the final manuscript.

Funding

UHF/NMF Diverse Scholars Program, CTSA UL1TR001873.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97343-y>.

Correspondence and requests for materials should be addressed to S.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021