Copyright © 2024, Massachusetts Medical Society. All rights reserved.

This is an Author Accepted Manuscript, which is the version after external peer review and before publication in the Journal. The publisher's version of record, which includes all NEJM AI editing and enhancements, will be available at https://ai.nejm.org/doi/full/10.1056/AIoa2400741.

The ADVENT Study: A multi-centre, prospective, double-blinded, pilot study evaluating the sensitivity of artificial intelligence driven automatic detection of proximal deep vein thrombosis.

Short title: ADVENT: A pilot study to evaluate Automatic-DVT diagnostic software.

Authors:

Nicola Curry,[1,2] Elisa Allen,[3] Laura Silsby,[3] Steve Goodacre,[4] Christopher Deane,[1,3] Alison Deary, [3] Ashley Foster,[3] James Griffiths,[3] Rupa Sharma,[3] Helen Thomas,[3] Sven Mischewitz,[5] Fouad Al-Noor,[6] and the ADVENT Investigators.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Affiliations:

1 - Oxford University Hospitals NHS Foundation Trust, Nuffield Orthopedic Centre, Windmill Road, Oxford, UK

2 – Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

3 - NHS Blood and Transplant Clinical Trials Unit, Fox Den Road, Stoke Gifford, Bristol BS34 8RR, UK

4 – Population Health, School of Medicine and Population Health, University of Sheffield, Sheffield, UK.

5 - ThinkSono GmbH, Potsdam, Germany.

6 - ThinkSono Ltd, London, UK.

Corresponding author

Name: Nicola Curry

Address: Oxford University Hospitals NHS Foundation Trust, Nuffield Orthopaedic Centre, Oxford. UK.

Email: nicola.curry@ouh.nhs.uk

Tel: 00-44-1865-225316

Word count: 2999

Abstract word count: 298

ABSTRACT

Background: Ultrasound is one of the most widely requested forms of diagnostic imaging. The costs for diagnosing deep vein thrombosis (DVT) in the UK are £175 million, annually. In at least 80% of cases, DVT is excluded. As healthcare becomes increasingly stretched, resource utilisation needs to be optimised. This prospective, double blind, test accuracy study was designed to test whether an artificial intelligence (AI)-guided software device (AutoDVT) could support non-radiology specialists to diagnose proximal DVT.

Methods: Eleven regional hospital DVT diagnostic clinics enrolled adult patients, age 18 or older, who were referred for investigation of symptoms suggestive of DVT, including a compression ultrasound. Prior to the clinical compression ultrasound, an AutoDVT scan was completed. This was a two-point AI-guided compression ultrasound scan. The main primary outcome was sensitivity of AutoDVT within a diagnostic algorithm for the detection of proximal DVT by non-radiology trained staff. Other outcomes included specificity and positive/negative predictive value (PPV/NPV) of AutoDVT.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Results: 414 participants were enrolled. Proximal DVT was detected in 10.5% of those analysed. AutoDVT resulted in 68% (95% confidence interval [CI]: 49 - 83%) sensitivity and 80% (95% CI: 74 –85%) specificity for detection of proximal DVT. NPV for AutoDVT was 95% (95% CI: 92 – 98%) with a PPV of 28% (95% CI: 19 – 40%). 63/294 results (21%, 95% CI: 17 – 27%) were discrepant compared to compression ultrasound.

Conclusion: Though AI-guided ultrasound use can detect proximal DVT, test accuracy was not sufficient for this device to be used safely. Further optimisation of the software is required prior to use in clinical practice by non-radiology trained healthcare professionals.

Trial Registration: ISRCTN (11069056).

INTRODUCTION

Venous thromboembolism (VTE) is a leading cause of death and disability[1] with an estimated incidence of 1.2 – 2.7 per 1,000.[2] Deep vein thrombosis (DVT) accounts for two-thirds of cases.[1] DVT is difficult to diagnose clinically; For every one patient confirmed to have a DVT up to nine more individuals will seek consultation only to have DVT excluded.[3,4] Predictive algorithms based on D-dimer testing and a clinical risk score (e.g., Wells' DVT score) can streamline which individuals require an ultrasound scan (USS)[5-7] but recent data show that 14 USS are conducted for every confirmed proximal DVT.[4]

Guidelines recommend that confirmation of DVT occurs within four hours, or 24 hours if interim anticoagulation is administered.[8] Compression USS is the routine imaging modality of choice,[9] with a sensitivity and specificity of 94% and 97% respectively when performed by radiologists.[10] Diagnostic delays for USS are common, largely driven by the scarcity of radiographers and radiologists.[11] Previous studies aimed at reducing the reliance of DVT-diagnosis on these individuals has involved training of nurses, [12] general practitioners, [13] or the use of point-of-care USS in emergency departments.[14] These approaches have been broadly successful, with diagnostic accuracy ranging from 91.7% (95% confidence interval [CI]: 85.0 - 95.6%) to 95.8% (95% CI: 94.7 - 97.0%). [13] Nevertheless, these approaches continue to rely on skilled healthcare professionals.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Artificial intelligence (AI)-guided diagnostic tools are increasingly used to support efficient delivery of resource-limited healthcare.[15,16] AutoDVT is an AI software that, when coupled to a handheld ultrasound device, is designed to guide non-radiology specialists to perform a 2-point compression USS and make a diagnosis — in real time — of a DVT. [17] This study was a diagnostic test accuracy study comparing AutoDVT with standard UK DVT diagnostic pathways and evaluating the sensitivity of AutoDVT for diagnosis of proximal DVT.

METHODS

Study design and patients

ADVENT was a multicentre, prospective, single arm, double-blind, pilot, test accuracy study conducted at 11 UK hospital DVT diagnostic services. The study protocol is available (Supplement 1). Patients were eligible if they were referred to a DVT diagnostic service with symptoms suggestive of a DVT, 18 years or older, and the DVT diagnostic algorithm indicated the need for an USS (Supplement Fig. 1). Exclusion criteria were pregnancy at 12+ weeks, D-dimer could not be performed (e.g., prior anticoagulation), or the patient previously had a

radiologically confirmed DVT in the symptomatic leg. If a participant was found to have an isolated distal DVT during the diagnostic process, they were retrospectively excluded, as the study design was for proximal DVT alone. The study was approved by the East of Scotland Research Ethics Service (REC reference: 21/ES/0070) and followed the Standards for Reporting of Diagnostic Accuracy (STARD) guideline.[18]

An independent data monitoring committee reviewed interim data and monitored patient safety. A Trial Steering Committee provided oversight. The trial is registered with ISRCTN 11069056.

Participant flow

Eligibility for study entry was confirmed at the point of patient referral to the DVT service. At some sites participants were screened prior to arrival, otherwise screening was conducted by research personnel in the DVT clinic. Fully informed, written consent was taken prior to study procedures. Baseline demographic data were collected at study entry. Recruited participants had a Wells score calculated and a D-dimer test (according to the DVT-diagnostic algorithm. The Auto-DVT scan was completed prior to the clinical USS (Supplement Fig. 2A/2B). The result of the AutoDVT scan was not disclosed to participants, research personnel, or remote reviewers and was not used to direct clinical care. No longer than 72 hours was permitted between the AutoDVT and clinical USS.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Once research procedures were completed, participants underwent a compression USS. Six sites used above knee scanning, with the first scan at presentation ('initial scan') and a second scan 6 – 8 days later described as day 7 (D7) where clinically indicated. Five sites used full leg scanning. The management and treatment of participants was led solely by routine clinical practices and clinical USS reports. Clinical outcome data were obtained by the research team after the participant had completed their clinical evaluation.

Description of AutoDVT software

All research personnel who were to conduct AutoDVT scans received an in-person training session, lasting an average of 60 minutes. The AutoDVT software was designed using machine learning to guide the user to conduct a 'two-point' compression ultrasound (e.g., assessment of common femoral and popliteal veins) to diagnose the presence/absence of DVT (Figure 1). A video showing how the software can be used is found here: AutoDVT training video. Prior software development is published.[17]

Briefly, a compatible handheld ultrasound device (Clarius L7 HD) is linked to the AutoDVT software application. The software requests study participant ID, symptomatic leg (right/left), and the research personnel ID prior to scan start. The researcher is directed to place the ultrasound probe on the groin region and the software detects the site of the probe, directs the operator to move the probe, and confirms correct positioning. The operator is requested to compress the vein several times, and once the software confirms receipt of sufficient data, the operator is directed to perform a similar operation in the popliteal fossa. Once the scan is completed by AutoDVT, the results are not shared, rather the software declares 'scan complete,' or if more data are required 'repeat exam.'

Outcomes

The main primary outcome was sensitivity of AutoDVT within a diagnostic algorithm for the detection of proximal DVT by non-radiology trained staff. The reference standard was a compression ultrasound completed by a trained radiographer or radiologist with real-time diagnosis using local procedures. The desired sensitivity of AutoDVT for the primary endpoint was set at 90% or above, in line with the reference standard.[10] Sensitivity was prioritised due to the dangers of missing clots.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

A further test accuracy outcome was assessed; The sensitivity of AutoDVT for the detection of a proximal DVT by non-radiology trained staff with remote retrospective analysis of the AutoDVT scans by a qualified clinician with more than one year's experience in diagnosing DVT by USS (e.g., sonographer or radiologist). This outcome was intended to provide insight into whether an inaccuracy in the primary outcome was due to AutoDVT captured image quality or the AI interpretation of the images. For this outcome the data stored from each AutoDVT scan (without the AutoDVT diagnosis) were reviewed retrospectively by five remote clinicians. Each clinician rated each AutoDVT scan using the American College of Emergency Physicians (ACEP) image quality 1 to 5 scale.[19] A rating of 3 or more, by three or more specialists, was chosen to mean the image data met or exceeded the minimal criteria for a diagnosis ('diagnostic quality scan'). If the reviewer scored a scan as ACEP 3+, they were asked to make a diagnosis (DVT, no DVT, or DVT indeterminate). Here DVT indeterminate describes a situation where the image quality is adequate (ACEP 3+), but a diagnosis could not be made (e.g., full compression could not be seen on all compression points). The diagnosis made by one randomly selected remote specialist was compared with the results from the participant's standard compression USS (AutoDVT with

review). This random selection was also implemented using Bootstrap methods, producing 'Bootstrap AutoDVT with review' results.

Pre-specified secondary outcomes included: Specificity of AutoDVT (desired specificity was set at 60%);[20] positive and negative predictive values of AutoDVT; numbers of discrepant results when comparing AutoDVT to standard ultrasound; diagnostic image quality of AutoDVT (using ACEP scale);[19] rates of imaging failure for AutoDVT; inter-operator variability of AutoDVT results (10% initial scans); interobserver agreement between five reviewers for each AutoDVT scan; numbers of eligible patients not enrolled; feasibility of recording AutoDVT scans; duration of AutoDVT scans; number of indeterminate and repeat scans; feedback from research personnel as to the ease of use of AutoDVT using the Post-Study System Usability Questionnaire (PSSUQ, version 3) score. [21,22] A safety outcome was included: The number of radiologically confirmed VTE events and/or deaths related to VTE during the 3-month follow up period.

Statistical Analysis

The sample size was specified at 30 participants with confirmed proximal DVT by clinical USS and who also completed the AutoDVT scan. As approximately 1 in 10 patients with symptoms will test positive for thrombosis, ten times as many participants as the required sample size needed to be scanned in total.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

A summary of the statistical methods is in the Supplementary Statistical Analysis, and the full statistical analysis plan is available in Supplement 2. All analyses were carried out using the software Base SAS version 9.4M7 (SAS Institute, Cary, North Carolina, USA).

RESULTS

ADVENT enrolled 414 patients from 11 diagnostic DVT services, between 23 December 2021 and 6 February 2023. In total, 1,302 patients were screened (Figure 2). Of the 879 eligible patients, 47% were recruited. Reasons for non-enrolment are detailed in Supplementary Table 2. The percentage of consented participants out of those eligible ranged from 10 – 100% across sites. Of the 414 enrolled, 120 were excluded from the outcome analysis for 'AutoDVT alone' and 191 were excluded from the outcome analysis for 'AutoDVT with review,' leaving a total of 294 and 223 participants for final analysis, respectively. Reasons for exclusion are listed in Figure 2. Baseline

characteristics of the enrolled and analysed (AutoDVT alone outcome) cohorts are in Table 1. The median age was 64 years, 54% male. 3% had a prior history of DVT. Median D-dimer value and Wells' score was 1028 µg/L and 2, respectively.

AutoDVT alone resulted in 68% (95% CI: 49 - 83%) sensitivity and 80% (95% CI: 74 – 85%) specificity for detection of proximal DVT. Of the 31 DVT-positive clinical scans, ten were diagnosed as DVT-negative by AutoDVT (Figure 3). Of these, one participant had a normal D-dimer where a DVT could have been missed. Four of the ten missed AutoDVT scans were DVT in the distal popliteal/proximal calf region, the remaining six were more proximal. Sensitivity and specificity were 85% (95% CI: 65 – 96%) and 73% (95% CI: 66 – 79%), when diagnosis was via a remote retrospective analysis of AutoDVT images ('AutoDVT + review') (Table 2). Diagnostic classification of AutoDVT scans given by five independent radiology trained clinicians is summarised in Table 3. There were 34 incomplete pairs of initial diagnoses excluded from the estimates of AutoDVT alone or AutoDVT + review results, mostly due to lack of an AutoDVT diagnosis (Supplementary Table 1). Other secondary endpoints relating to AutoDVT performance are in table 4. The negative predictive values (NPV) and positive predictive values (PPV) for AutoDVT were 95% (95% CI: 92 – 98%) and 28% (95% CI: 19 – 40%), and for AutoDVT + review: 97% (95% CI: 93 – 99%) and 29% (95% CI: 19 -40%), respectively.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

In total, there were 392 initial and 52 D7 clinical ultrasound scans conducted on the 414 enrolled participants, compared with 344 presentation and eight D7 scans using the AutoDVT device. 5% of initial and 0% of D7 AutoDVT scans were indeterminate (e.g., AutoDVT device could not collect adequate ultrasound cine loops from each defined anatomical region and/or the user had to abort the scan). 81% of initial AutoDVT scans had to be repeated due to a failure in the AutoDVT scanning procedure a median of three times while 75% of D7 AutoDVT scans had to be repeated a median of 1.5 times. 24% and 15% of the remote reviews were indeterminate (either the scan was given an ACEP score of 1 or 2, or ACEP 3+ but the reviewer could not make a diagnosis) for initial and D7 AutoDVT scans, respectively. The average AutoDVT scan was completed in just under four minutes. The number of AutoDVT imaging failures at the initial scan was small, n = 6 (6/344, 1.7%), and all six were due to the user aborting the test. There were no AutoDVT imaging failures at D7. 95% of AutoDVT scans were able to be recorded, when requested (n = 335/352). 6% of AutoDVT initial scans were completed by two independent research personnel (n = 19 scans). There was good agreement between the two operators (kappa statistic 0.73) (Supplementary Table 3). Twenty-eight AutoDVT users were included in the study and provided answers to the

PSSUQ, v.3 questionnaire. The median (IQR) global PSSUQ score was 3 (2-3), meeting the pre-specified definition of the software being considered 'user-friendly' (defined by a median <4). Full data can be found in Supplementary Table 4. Sensitivity (86%), NPV (98%), and other diagnostic accuracy measures of AutoDVT + review using Bootstrap methods are presented in Supplementary Table 5.

Five participants (1%) developed seven reported additional VTE events within 3-months follow-up (Table 5). In two of these instances, AutoDVT reported the presence of a proximal DVT when none was detected on the clinical USS. Two deaths, from non-VTE related causes occurred during follow up. One participant died of complications from a subdural haematoma after a fall. No cause of death was given for the second participant.

DISCUSSION

In this multicentre, prospective, double blind, pilot, test accuracy study, AI-guided diagnosis by non-radiology specialists, when used alone, achieved 68% sensitivity and 80% specificity for detection of proximal DVT. The pre-specified desired sensitivity of 90% or more was not achieved.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

The AutoDVT software was designed to be used here as a limited two-point compression scan for operator ease, concentrating on the compressibility of the common femoral and popliteal veins. Using AutoDVT alone, ten of the 31 positive clinical scans were diagnosed as DVT negative. This lack of sensitivity is not acceptable in a clinical diagnostic setting as missed DVTs could place patients at risk of DVT extension or, more concerningly, a life-threatening pulmonary embolism. Of the ten misdiagnosed as DVT negative, four were in the distal popliteal veins, but six were in the larger proximal veins. The clinical scan operators at participating sites reported, after study completion, to perform compression approximately every 10 – 15 cm along the length of the deep vein system for the reference-standard USS. This difference, which has been highlighted in clinical settings as a limitation to diagnostic accuracy of 2-point compression,[23] is a limitation of AutoDVT as conducted in this study.

AutoDVT alone incorrectly diagnosed 53 participants as having a proximal DVT (53/263 = 20%). Of these, 44 clinical scans were normal, and the remainder detailed: calf haematoma (n=1); Baker's cyst (n=3); short segment calf thrombophlebitis (n=1); extensive long saphenous vein thrombophlebitis (n=4, one abutting the saphenofemoral junction). Of these 53 false positives, two participants went on to develop VTE — one two months later having been found initially to have extensive (>40cm) thrombophlebitis, treated with intermediate

dose low molecular weight heparin (LMWH) for six weeks. The other participant had a full leg clinical scan that was negative and was diagnosed 22 days later with PE (Table 5).

When evaluating the AutoDVT + review data, sensitivity was higher at 85% but with reduced specificity at 73%, suggesting reviewers had a lower threshold for diagnosing DVT but an overall similar accuracy. Information from 71 participants was excluded as the expert reviewer deemed the image 'DVT indeterminate.' Within that number, five of the 31 DVT-positive USS cases were removed from evaluation (Tables 2A, 2B). Of these five, three AutoDVT scans that correctly confirmed a DVT and two that misdiagnosed a DVT as negative were removed. Four of the false negative AutoDVT scans were redefined as being positive after remote radiology review. Overall 24% (407/1720) of all the remote reviews of initial AutoDVT scans were deemed indeterminate.

The variability of reporting by the five reviewers (Table 3) is noteworthy, with a kappa statistic of 0.13 (95% CI: 0.07 – 0.18) for image quality, highlighting poor interobserver agreement. A parallel study using the same AutoDVT software reported an identical 75% figure, similarly showing that three-quarters of the captured AutoDVT scans were of sufficient image quality to be evaluated by remote radiologists.[24] This is an important finding, as poor image quality will itself limit AI accuracy. Optimising AutoDVT could involve measures such as restricting determination to good quality images only or improving the scanning and image capture technique performed by operators. These results suggest that some of the real-time feedback that radiologists and radiographers experience whilst scanning (e.g., the physical resistance to compression, the recognition of a poor image and USS probe adjustment) should not be underestimated and greater AI guidance for optimal probe placement could be developed. Notably, these findings highlight a marked difference between AI-assisted reporting for X-ray or CT-scan images, where the AI software uses the exact same data as a clinician for diagnosis (e.g., with no physical/real-time intra-scan feedback). Real-world, large numbers of scans to provide iterative training of the software, using images captured by non-radiology trained operators, could help address this limitation. In addition, in-depth analysis of the conditions which lead to any suboptimal AutoDVT results, using a combination of expert radiology opinion as well as statistical modelling, will provide important information to guide algorithm refinement.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

The negative predictive value (NPV) of both AutoDVT alone and AutoDVT + review pathways are, at face value, high at 95% and 97%, respectively. However, NPV values depend on the underlying disease prevalence[25] and, with a DVT prevalence of 10% as in this study, NPV values are likely to be high even with relatively low sensitivity

and should not be overinterpreted. Importantly, NPV values for the reference-standard compression USS are 99%.[10]

Strengths of this study include its multicentre nature and the double-blind study design, which supports the reliability of the AutoDVT results presented. This is one of the only diagnostic test accuracy studies conducted in a clinical DVT diagnostic setting and demonstrates that AutoDVT can be used by a range of non-radiology trained staff. Speed of scanning was short at four minutes, whereas a clinical USS takes 15–20 minutes on average, even in experienced hands.[26] Furthermore, it demonstrates that remote review of the recorded AutoDVT scans can be used, which could support diagnosis. Inter-operator agreement analysis was not fully completed due to staff shortages, which is a limitation of this study. The handheld ultrasound used was the Clarius L7, and although unlikely that the operability will differ with another ultrasound machine, this has not been tested.

In conclusion, test accuracy was not sufficient for this device to be used safely in clinical practice when compared to the reference standard for diagnosis of proximal DVT. Skilled radiographers and radiologists incorporate real-time, iterative feedback during USS procedures to optimise image capture and provide an accurate result. The data from this study show that capture of better DVT image quality is vital to further optimisation of AutoDVT software.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Contributors

NC and FAN conceived the study. NC, FAN, SG, CD, HT initiated the study design and CD, CF, AD, AF, JG and JL helped with implementation and coordinated the project. FAN is the grant holder. EA and HT led and supervised statistical analysis. All authors contributed to refinement of the study protocol and approved the final version. The site principal investigators supervised patient recruitment. EA, LS, HT, RS, were responsible for data curation, formal analysis, and data visualization. NC, EA, FAN, SM wrote the original draft, which was then reviewed and approved by all co-authors. All authors had access to the raw data. EA, HT, FAN, SM, NC verified the data.

Additional Contributors

We thank the patients who supported the study, principal investigators and other research staff at the study sites, NHS Blood & Transplant Clinical Trials Unit (including Claire Foley, Mark Jones, Jonathan Last and Joanne Lucas) and Oxford University Hospitals NHS Foundation Trust Joint Research Office and Sponsors Office, the Trial Steering Committee (Dr Daniel Horner (Chair), Dr Catherine Bagot, Rupert Mann, Jonathan Clarke, Katie Flight) and the Data Monitoring Committee (Professor Cheng Hock Toh (Chair), Professor Gavin Murphy, Professor Paul White).

Principal Investigators and researchers at participating sites

We thank the principal investigators and all researchers involved with the ADVENT study at all participating sites: Oxford University Hospitals NHS Foundation Trust (Nuffield Orthopaedic Centre, John Radcliffe Hospital: PIs – Jamie Burbage, Dr. Dan Lasserson), Buckinghamshire Healthcare NHS Trust (PI: Dr. Alice Ngumo), Cardiff and Vale University Health Board (PI: Prof. Raza Alikhan), King's College Hospital NHS Foundation Trust (PI: Dr. Lara Roberts, King's College Hospital, Princess Royal University Hospital), Maidstone and Tunbridge Wells Hospitals NHS Trust (PI: Dr. Andrew Ross-Parker, Maidstone Hospital, Tunbridge Wells Hospital), Sheffield Teaching Hospitals NHS Foundation Trust (PI: Prof. Steve Goodacre), Stockport NHS Foundation Trust (PI: Dr. Chaminda Jayawarna), and University Hospitals of Leicester NHS Trust (PI: Dr. Mauro Culasso).

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Declaration of Interests

NC, EA, LS, SG, CD, AD, AF, HT, JG, RS have no relevant declaration of interests.

SM and FA-N co-founded the ThinkSono company and developed the software tested in this study.

Funding

Wellcome Trust, Wellcome Innovator Award 220505/Z/20/Z.

Role of the Funding Support

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of this manuscript.

REFERENCES

1. ISTH Steering Committee for World Thrombosis Day. Thrombosis: a major contributor to global disease burden. *Thromb Haemos*. 2014; **112(5):** 843-52.
2. Wendelboe AM, Raskob GE. Global Burden of Thrombosis: Epidemiologic Aspects. *Circ Res* 2016; **118(9):** 1340-7.
3. Büller HR, Ten Cate-Hoek AJ, Hoes AW, et al. AMUSE (Amsterdam Maastricht Utrecht Study on thromboEmbolism) Investigators. Safely ruling out deep venous thrombosis in primary care. *Ann Intern Med* 2009; **150:** 229-35.
4. Willan J, Agarwal G, Varghese T, Cairns K, Baker P, Curry N. Diagnosis of deep vein thrombosis using an adjusted clinical probability scoring: supportive evidence for a change in clinical practices. *J Thromb Haemost* 2023; **21(8):** 2300-4.
5. Wells PS, Anderson DR, Rodger M, et al. Evaluation of D-dimer in the diagnosis of suspected deep-vein thrombosis. *N Engl J Med* 2003; **349(13):** 1227-35.
6. Stubbs MJ, Mouyis M, Thomas M. Deep vein thrombosis. *BMJ* 2018; **360:** k351.
7. Kearon C, de Wit K, Parpia S, et al, on behalf of the Designer D-Dimer Deep vein thrombosis (4D) Study Investigators. Diagnosis of deep vein thrombosis with D-dimer adjusted to clinical probability: prospective diagnostic management study. *BMJ* 2022; **376:** e067378.
8. NICE Clinical Guideline 158 updated 02 August 2023. Venous thromboembolic diseases: diagnosis, management and thrombophilia testing. Found here: https://www.nice.org.uk/guidance/NG158 Accessed 15th December 2023.
9. Needleman L, Cronan JJ, Lilly MP, et al. Ultrasound for Lower Extremity Deep Venous Thrombosis: Multidisciplinary Recommendations from the Society of Radiologists in Ultrasound Consensus Conference. *Circulation* 2018; **137(14):** 1505-15.
10. Goodacre S, Sampson F, Thomas S, van Beek E, Sutton A. Systematic review and meta-analysis of the diagnostic accuracy of ultrasonography for deep vein thrombosis. *BMC Med Imaging* 2005; **5:** 6.
11. All Party Parliamentary Thrombosis Group Annual Report, 2018. Accessed 8th March 2024: http://apptg.org.uk/wp-content/uploads/APPTG-annual-survey-report-2018-compressed.pdf
12. Mumoli N, Vitale J, Cocciolo M, et al. Accuracy of nurse-performed compression ultrasonography in the diagnosis of proximal symptomatic deep vein thrombosis: a prospective cohort study. *J Thromb Haemost* 2014; **12(4):** 430-5.
13. Mumoli N, Vitale J, Giorgi-Pierfranceschi M, et al; PRACTICUS Study Investigators. General Practitioner-Performed Compression Ultrasonography for Diagnosis of Deep Vein Thrombosis of the Leg: A Multicenter, Prospective Cohort Study. *Ann Fam Med* 2017; **15(6):** 535-9.
14. Javaudin F, Seon J, Le Bastard Q, et al: Winfocus-France Study Group; Le Conte P. Evaluation of short training session for venous limited compression ultrasonography: prospective multicenter study. *Ultrasound J* 2020; **12(1):** 5.
15. Tang X. The role of artificial intelligence in medical imaging research. *BJR Open* 2019; **2(1):** 20190031.
16. Oren O, Kebebew E, Ioannidis JPA. Curbing unnecessary and wasted diagnostic imaging. JAMA 2019; **321:** 245–6.
17. Kainz B, Heinrich MP, Makropoulos A, et al. Non-invasive diagnosis of deep vein thrombosis from ultrasound imaging with machine learning. *NPJ Digit Med* 2021; **4:** 137.
18. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351:** h5527.
19. ACEP, 2018: Found here: https://www.acep.org/siteassets/uploads/uploaded-files/acep/clinical-and-practice-management/policy-statements/information-papers/emergency-ultrasound-standard-reporting-guidelines---2018.pdf Accessed 14th December 2023.
20. Stein PD, Hull RD, Patel KC, Olson RE, Ghali WA, Brant R, Biel RK, Bharadia V, Kalra NK. D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review. Ann Intern Med. 2004; **140(8):** 589-602.
21. PSSUQ. Found here: https://uiuxtrend.com/pssuq-post-study-system-usability-questionnaire/ Accessed 14th December 2023.
22. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int J Hum-Comput Int* 2002; **14:** 463–88.
23. Bhatt M, Braun C, Patel P, et al. Diagnosis of deep vein thrombosis of the lower extremity: A systematic review and meta-analysis of test accuracy. *Blood Adv* 2020; **4:** 1250–64.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

24. Oppenheimer J, Mandegaran R, Staabs F, et al. Remote Expert DVT Triaging of Novice-User Compression Sonography with AI-Guidance. *Ann Vasc Surg* 2024; **99:** 272-9.
25. Rouvière O, Souchon R, Melodelima C. Pitfalls in interpreting positive and negative predictive values: Application to prostate multiparametric magnetic resonance imaging. *Diag Inter Imaging* 2018; **99(9):** 515-8.
26. Sanyal R, Kraft B, Alexander LF, Ismail A, Lockhart ME, Robbin ML. Scanner-Based Protocol-Driven Ultrasound: An Effective Method to Improve Efficiency in an Ultrasound Department. *Am J Roentgenol* 2016; **206(4):** 792-6.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Figure Legends.

Figure 1.

ThinkSono software showing Mobile App, Clarius handheld ultrasound device (left and ThinkSono cloud based dashboard for qualified clinician review (right)

Figure 2.

CONSORT Diagrams.

2A. AutoDVT alone.

Figure 2A legend. For AutoDVT, there were 414 participants enrolled; 383 of whom had an initial USS and AutoDVT scan pair (Scan completion (able to diagnose) box). Of all enrolled participants, 45 did not complete the trial (Follow-Up box) but 30 were able to be included in the primary outcome analysis. So, 15 of these 45 participants were 'Excluded from analysis' and 30 were 'Analysed.' Of all enrolled participants, 294 were analysed and 120 were excluded from analysis (Analysis box).

2B. AutoDVT + review.

Figure 2B legend. For AutoDVT with review, there were 414 participants enrolled (the same participants as for AutoDVT alone); 380 of whom had an initial USS and AutoDVT with review pair (Scan completion (able to diagnose) box). Of all enrolled participants, 45 did not complete the trial (Follow-Up box) but 23 were able to be included in the primary outcome analysis. So, 22 of these 45 participants were 'Excluded from analysis' and 23 were 'Analysed.' Of all enrolled participants, 223 were analysed and 191 were excluded from analysis (Analysis box).

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Figure 2 Key:

DVT – D7 – day 7 (those participants scanned at presentation with a proximal scan only may require a second follow up scan at 1 week); deep vein thrombosis; USS – ultrasound scan; bug fix 1 – the partnering company to the ADVENT study (ThinkSono Ltd) updated the AutoDVT software as the study was recruiting and, as the update was likely to affect the primary outcome, it was decided that participants scanned prior to the software update would be discarded for primary outcome calculation

Figure 3.

DVT scan details of the 31 positive ultrasound scans and AutoDVT results alone, and outcomes after remote radiology review.

Figure 3 Key:

DVT – deep vein thrombosis; n/a – not available; SFV – superficial femoral vein; USS – ultrasound scan

Table 1.

Baseline characteristics of enrolled and participants analysed, AutoDVT alone

	Participants enrolled (N=414)	Participants analysed (N=294)
Characteristic		
Sex, male; n/N (%)	223/414 (54)	156/294 (53)
Age (years)	64 (53-74)	65 (53-75)
Height (cm)	170 (162-179)	170 (160-178)
Weight (kg)	87 (74-101)	85 (73-100)
Wells' score	2 (1-2)	2 (1-2)
D-dimer result (µg/L, FEU)	1028 (566-2197)	1052 (595-2270)
Leg with symptoms, left; n/N (%)	210/414 (51)	149/294 (51)
Previous DVT, yes; n/N (%)	11/414 (3)	7/294 (2)
Underlying malignancy, yes; n/N (%)	10/414 (2)	7/294 (2)
Currently pregnant, yes; n/N (%)	2/414 (0)	0/294 (0)

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Data are number/total number (%) for categorical variables and median (IQR) for continuous variables.

Key: DVT – deep vein thrombosis; FEU – fibrinogen equivalent units; IQR – interquartile range

Characteristic (number of participants with missing data in Enrolled): Age (0); sex (0); height (2); weight (1); Wells' score (13); D-dimer (79); leg with symptoms (1); previous DVT (2); underlying malignancy (2); currently pregnant (1).

Characteristic (number of participants with missing data in Analysed): Age (0); sex (0); height (1); weight (0); Wells' score (7); D-dimer (56); leg with symptoms (0); previous DVT (1); underlying malignancy (1); currently pregnant (0).

Table 2.

2A. Relationship between initial AutoDVT scan and standard compression ultrasound scan result

AutoDVT Result	Standard Compression Ultrasound Result		Total
	DVT yes	DVT no	
DVT yes	21	53	74
DVT no	10	210	220
Total	31	263	294

2B. Relationship between initial 'AutoDVT scan + review' † and standard compression ultrasound scan result

AutoDVT + Review Result	Standard Compression Ultrasound Result		Total
	DVT yes	DVT no	
DVT yes	22	54	76
DVT no	4	143	147
Total	26	197	223

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Data are number of participants.

Key: † – randomly selected diagnosis, with equal probability, from 5 reviews; DVT – deep vein thrombosis

Table 3.

Diagnostic classification of AutoDVT scans by five independent remote radiology-trained clinicians

	Independent reviewer				
	1	**2**	**3**	**4**	**5**
Technically unsatisfactory AutoDVT scan*	2 (1)	50 (15)	62 (19)	25 (7)	1 (0)
DVT yes	103 (31)	24 (7)	115 (34)	81 (24)	50 (15)
DVT no	221 (66)	158 (47)	125 (37)	147 (44)	238 (71)
DVT indeterminate	9 (3)	103 (31)	33 (10)	82 (24)	46 (14)
TOTAL	335	335	335	335	335

Key – * - defined by an ACEP score < 3; DVT – deep vein thrombosis

Data are N (%). The data show the distribution of diagnoses over repeated reviews

Agreement in diagnosis between individual pairs of reviewers (excluding the technically unsatisfactory category), summarised by a kappa statistic, is 1v2=0.44; 1v3=0.44; 1v4=0.52; 1v5=0.56; 2v3=0.34; 2v4=0.47; 2v5=0.48; 3v4=0.47; 3v5=0.35; 4v5=0.49, where 1v2 denotes the kappa statistic for reviewers 1 and 2. The closer the kappa to 1, the greater the agreement. The highest agreement is between reviewers 1 and 5, and the poorest agreement is between reviewers 2 and 3.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Table 4.

Secondary outcomes

	AutoDVT	AutoDVT + review
Diagnostic ability: AutoDVT compared to standard ultrasound		
Specificity, % (95% CI)	80 (74-85)	73 (66-79)†
Positive predictive value, % (95% CI)	28 (19-40)	29 (19-40)†
Negative predictive value, % (95% CI)	95 (92-98)	97 (93-99)†
Discrepant results, n/N (%, 95% CI)	63/294 (21, 17-27)	58/223 (26, 20-32)†
AutoDVT software performance		
ACEP score, median (IQR), N		4 (3-4), 335†
Minimal image quality met*, n/N (%, 95% CI)		330/335 (99, 97-100)
Interobserver image quality agreement, weighted kappa statistic (95% CI)		0.13 (0.07-0.18)
Interobserver DVT diagnosis, kappa statistic (95% CI)**		0.46 (0.37 – 0.55)
Interoperator DVT diagnosis agreement, n/N (%)	17/19 (89)	
Interoperator DVT diagnosis agreement, kappa statistic (95% CI)	0.73 (0.38-1)	
Feasibility of scan being recorded, n/N (%)	335/352 (95)	
Scan duration in seconds, median (IQR) (Range, N)	234 (160-377) (74-1393, 335)	

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Key – † – randomly selected radiologist review; * - definition of minimal image quality met = an ACEP score of 3 or more given by three or more of the five reviewers; ** - 99 scans (29.6%) were excluded for having one or more technically unsatisfactory reviews (e.g. ACEP score <3); ACEP – American College of Emergency Physicians; CI – confidence intervals; DVT – deep vein thrombosis; IQR – interquartile range

There is no universally agreed kappa value that indicates good agreement, the following values have been proposed [Altman, 1991]: value of kappa (strength of agreement) = <0.20 (poor); 0.21-0.40 (fair); 0.41-0.60 (moderate); 0.61-0.80 (good); 0.81-1.00 (very good).

Table 5.

Reported VTE events during study follow up.

Participant number	Study entry date	Initial clinical USS report	Initial AutoDVT report	SAE date	SAE VTE event	Description
A023079	18/06/2022	No proximal DVT	AutoDVT: negative AutoDVT + review: 5/5 negative	18/06/2022	Segmental PE	SOB at presentation, CTPA confirmed PE
A023150 (2 VTE events)	04/11/2022	Mid-thigh STP No proximal DVT	AutoDVT: positive AutoDVT + review: 5/5 positive	16/01/2023	Distal/mid femoral and popliteal DVT	Re-referred for clinical evaluation of worsening leg swelling
				18/01/2023	Incidental PE on CT CAP	Asymptomatic PE
A309009	03/11/2022	No proximal DVT or STP on whole leg scanning	AutoDVT: positive AutoDVT + review: 4/5 indeterminate 1/5 positive	25/11/2022	Segmental PE	Pleuritic chest pain, pain in left calf resolved.
A309012	14/11/2022	Common femoral DVT	AutoDVT: positive AutoDVT + review: 5/5 positive	14/11/2022	Bilateral PE	SOB at time of entry into study, referred for CTPA
A204027 (2 VTE events)	13/01/2023	Small below knee left DVT, no anticoagulation given	AutoDVT: negative AutoDVT + review: 5/5 negative	20/01/2023	More extensive distal left DVT and new distal right DVT found on clinical scan at d7	Clinical decision to anticoagulate at the d7 scan
				17/02/2023	Subsegmental PE	Symptomatic chest pain, CTPA showed PE whilst compliant on treatment anticoagulation

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Key – AutoDVT – report based on the results from the AutoDVT alone; AutoDVT + review – report based on AutoDVT scan being retrospectively reviewed by 5 independent experienced radiology clinicians; CT CAP – computerised tomography scan of chest, abdomen and pelvis; CTPA – CT pulmonary angiography; d7 – day 7; DVT – deep vein thrombosis; PE – pulmonary embolus; SOB – shortness of breath; VTE – venous thromboembolism.