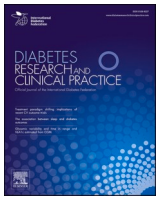
diabetesresearchandclinicalpractice 205 (2023) 110951

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.journals.elsevier.com/diabetes-research-and-clinical-practice



Artificial intelligence for automated detection of diabetic foot ulcers: A real-world proof-of-concept clinical evaluation

Bill Cassidy [a,*], Moi Hoon Yap [a], Joseph M. Pappachan [b], Naseer Ahmad [c], Samantha Haycocks [d], Claire O'Shea [e], Cornelious J. Fernandez [f], Elias Chacko [g], Koshy Jacob [h], Neil D. Reeves [i]

[a] Department of Computing Mathematics, Manchester Metropolitan University, John Dalton Building, Manchester M1 5GD, UK
[b] Lancashire Teaching Hospitals NHS Foundation Trust, Preston PR2 9HT, UK
[c] Manchester University NHS Foundation Trust, Manchester M13 9WL, UK
[d] Salford Royal NHS Foundation Trust, Stott Lane, Salford M6 8HD, UK
[e] Te Whatu Ora Health New Zealand Waikato, Pembroke Street, Hamilton 3240, New Zealand
[f] Department of Endocrinology and Metabolism, Pilgrim Hospital, United Lincolnshire Hospitals NHS Trust, Boston LN2 5QY, UK
[g] Jersey General Hospital, The Parade, St Helier, JE1 3QS Jersey, UK
[h] Eastbourne District General Hospital, Kings Drive, Eastbourne BN21 2UD, UK
[i] Faculty of Science & Engineering, Manchester Metropolitan University, John Dalton Building, Manchester M1 5GD, UK

A B S T R A C T

Objective: Conduct a multicenter proof-of-concept clinical evaluation to assess the accuracy of an artificial intelligence system on a smartphone for automated detection of diabetic foot ulcers.

Methods: The evaluation was undertaken with patients with diabetes ($n = 81$) from September 2020 to January 2021. A total of 203 foot photographs were collected using a smartphone, analysed using the artificial intelligence system, and compared against expert clinician judgement, with 162 images showing at least one ulcer, and 41 showing no ulcer. Sensitivity and specificity of the system against clinician decisions was determined and inter- and intra-rater reliability analysed.

Results: Predictions/decisions made by the system showed excellent sensitivity (0.9157) and high specificity (0.8857). Merging of intersecting predictions improved specificity to 0.9243. High levels of inter- and intra-rater reliability for clinician agreement on the ability of the artificial intelligence system to detect diabetic foot ulcers was also demonstrated ($K\alpha > 0.8000$ for all studies, between and within raters).

Conclusions: We demonstrate highly accurate automated diabetic foot ulcer detection using an artificial intelligence system with a low-end smartphone. This is the first key stage in the creation of a fully automated diabetic foot ulcer detection and monitoring system, with these findings underpinning medical device development.

1. Introduction

Diabetic foot ulcers (DFUs) and associated amputations are a global health and economic burden. In the United Kingdom, they account for 10% of the diabetes health services budget [14]. A reduction of DFU cases by one-third would result in a gross annual saving of more than £250 GBP million in the UK [16]. In the United States, the cost related to DFU is estimated to be approximately $9–13 USD billion in addition to the cost associated with diabetes [15]. Patients with diabetes have a lifetime risk of up to 34% of developing a DFU, with more than half of all cases leading to infection [4]. DFU is considered to be a clinical marker for increased risk of amputation and mortality [19,11].

Smartphone healthcare apps and associated research has seen notable growth in recent years, with increased engagement with self-monitoring health apps [3,23]. Artificial intelligence (AI) has been used in recent studies for real-time screening of diabetic retinopathy [3] (*accuracy* = 94.7%), diabetes prediction using lifestyle data [28] (*accuracy* = 82.1%), and screening for pre-diabetes in children and adolescents [29] (*accuracy* = 90.13%). AI applications are also being developed for other clinical areas such as endometriosis screening [30] and screening for genetic syndromes in children [31]. These studies showed promising results, indicating the potential importance of AI in clinical settings.

There have also been numerous advances in the use of AI for automated and semi-automated DFU screening, detection, and monitoring. Brown et al. [6] developed the MyFootCare smartphone app, which

* Corresponding author.
E-mail addresses: bill.cassidy@stu.mmu.ac.uk (B. Cassidy), m.yap@mmu.ac.uk (M. Hoon Yap), pappachan.joseph@lthtr.nhs.uk (J.M. Pappachan), naseer.ahmad@mft.nhs.uk (N. Ahmad), sam.haycocks@hotmail.co.uk (S. Haycocks), claire.o'shea@waikatodhb.health.nz (C. O'Shea), cornelius.fernandez@ulh.nhs.uk (C.J. Fernandez), e.chacko@health.gov.je (E. Chacko), k.jacob1@nhs.net (K. Jacob), n.reeves@mmu.ac.uk (N.D. Reeves).

https://doi.org/10.1016/j.diabres.2023.110951
Received 18 April 2023; Received in revised form 2 October 2023; Accepted 11 October 2023
Available online 15 October 2023

attempted to motivate self-care using aspects such as personal goals. Three patients participated in the study, with delineation of DFU wound region completed using a semi-automated procedure where the user was required to manually indicate DFU location and surrounding skin within the app.

Yap et al. [26] developed a smartphone app used to standardise capture of DFU photographs to improve follow-up analysis. The app used basic image processing techniques to display a ghost image of the initial foot photo which could be used to align subsequent follow-up photographs for standardisation purposes [1]. However, this system was used only for image capture and was not designed for automated DFU recognition.

Wang et al. [24] proposed a system to perform wound area measurement to monitor healing progress. In a later study they used a smartphone app with a capture box to perform automated wound region delineation. However, these studies were limited due to a small sample size of 65 photographs from patients and hand-crafted wound models.

Thermographic technologies have shown promise for providing early indication of DFU development [2,12]. However, from a self-monitoring perspective, thermographic smartphone attachments are prohibitively expensive. Therefore, photography remains an important research field for automated DFU screening, detection and monitoring, as it allows for accurate indication of pathology, whereas thermography is mainly aimed at identifying DFU before they appear. Additionally, monthly temperature monitoring failed to show a significant reduction in ulcer recurrence rates or increased ulcer-free survival [20].

Chan et al. [9] tested a smartphone app capable of measuring DFU wounds in hospital settings with 28 patients. However, this solution was not able to accurately delineate wound regions and required manual intervention to correct, further emphasising the importance of accurate DFU detection as the first key stage in automated DFU monitoring.

Goyal et al. [13] demonstrated high levels of accuracy in DFU detection using fully automated AI techniques in experimental settings. These AI techniques showed high sensitivity and specificity (>0.9) for automatic detection of DFU from 355 test images.

Semi-automated wound monitoring systems have been introduced in recent years using smartphone technologies [25], however, they are not suitable for patient use as they require manual adjustments to prediction results. As evidenced by the literature reviewed here and further highlighted by a recent review [10], most wound detection and monitoring solutions lack rigorous evaluation studies to determine their efficacy in real-world settings.

Automated AI detection of DFU with high accuracy will allow for remote screening and self-monitoring of patients outside hospital/clinic settings, resulting in a reduction in hospital visits. Associated financial impacts are particularly pertinent for patients in developing countries with restricted access to healthcare where the cost of treating the disease can be equivalent to 5.7 years of annual income [8]. Accurate remote screening and monitoring of DFU will aid earlier DFU detection to reduce amputations and improve overall health outcomes.

The aim of this study is to investigate if an AI algorithm running on a cloud platform connected to low-end smartphone devices can be used for screening and accurate automatic detection of DFU in 'real-world' settings.

2. Subjects, materials and methods

A proof-of-concept clinical evaluation for the smartphone app and cloud-based framework took place at two hospitals in the United Kingdom: Salford Royal NHS Foundation Trust, and Lancashire Teaching Hospitals NHS Foundation Trust, between September 2020 and January 2021. Ethical and governance approvals were obtained from the Northern Care Alliance NHS Foundation Trust (REF: S19HRANA37) and Lancashire Teaching Hospitals NHS Foundation Trust (REF: SE-281). Written informed consent was obtained from all participating clinicians and patients.

Main inclusion criteria for patient participants included: aged > 18 years, diagnosed with diabetes type 1 or 2, and had at least one intact foot. The main exclusion criteria for patients was amputation of both feet at or above the ankle. Participating clinicians were clinical specialists in the diabetic foot (podiatrist (n = 3), surgeon (n = 1), and consultant (n = 1)).

2.1. AI algorithm development

The deep learning model used in the evaluation is a single classifier localisation model using the Faster R-CNN and Inception-ResNetV2 architectures for feature extraction and object localisation [13]. Transfer learning from the MS COCO dataset was followed by model training using 1775 DFU photographs and expert labels obtained from Lancashire Teaching Hospitals.

2.2. Data collection

The smartphone app created for this proof-of-concept clinical evaluation was used by clinicians on a low-end smartphone device to automatically detect a DFU if present at various stages of development. A total of 203 foot photographs were captured from 81 patients at two hospitals in the United Kingdom. The app is a non-contact solution that connects to a cloud-based platform to run the artificial intelligence algorithm and complete the automated analysis. The cloud platform hosts an AI algorithm that is capable of identifying the location of DFUs on photographs acquired using the smartphone app. Six clinicians participated in the evaluation, comprising podiatrists, surgeons, and consultants. Each clinician was provided with a low-end Android smartphone (a Nokia 1 Plus or a Motorola E5 Play) with the app pre-installed. Low-end smartphones were used to demonstrate that the efficacy of the system was not affected by the quality of the smartphone device or acquired photograph, since the automated analysis was performed on the cloud platform. The app was used to acquire photographs of patient's feet during scheduled appointments where a DFU was either present or not present to allow for the analysis of all possible detection results. Fig. 1 shows the main data capture screens within the smartphone app and the associated clinical workflow used during the proof-of-concept clinical evaluation.

General guidelines were agreed with clinicians for the acquisition of foot photographs when using the app. These included orienting the foot vertically within the photograph so that the heel was positioned at the bottom of the image and the toes at the top. DFUs were photographed in different states of preparation. Each smartphone device used was purposely configured to use a different image resolution to challenge the AI algorithm: 700 × 933; 700 × 525; 700 × 350, and 500 × 1000 pixels. Flashes and other image processing features were disabled on all devices. Of the DFU cases photographed during the proof-of-concept clinical evaluation, some wounds may have surfaced naturally, while others had been exposed as a result of debridement.

The data that the results of the study were derived from can be divided into three distinct sets:

1. Patient foot photographs, consisting of lesion-free feet and feet exhibiting DFU and other associated pathology at various stages of development.
2. Diagnostic results returned by the AI algorithm for each foot photograph taken using the app.
3. Diagnosis results returned by the AI algorithm for each photograph taken using the app following a post-processing stage to refine predictions.

At the end of the evaluation, two additional podiatrists who did not participate in the initial evaluation were asked to provide answers to four questions for all 203 foot photographs together with the AI predictions:

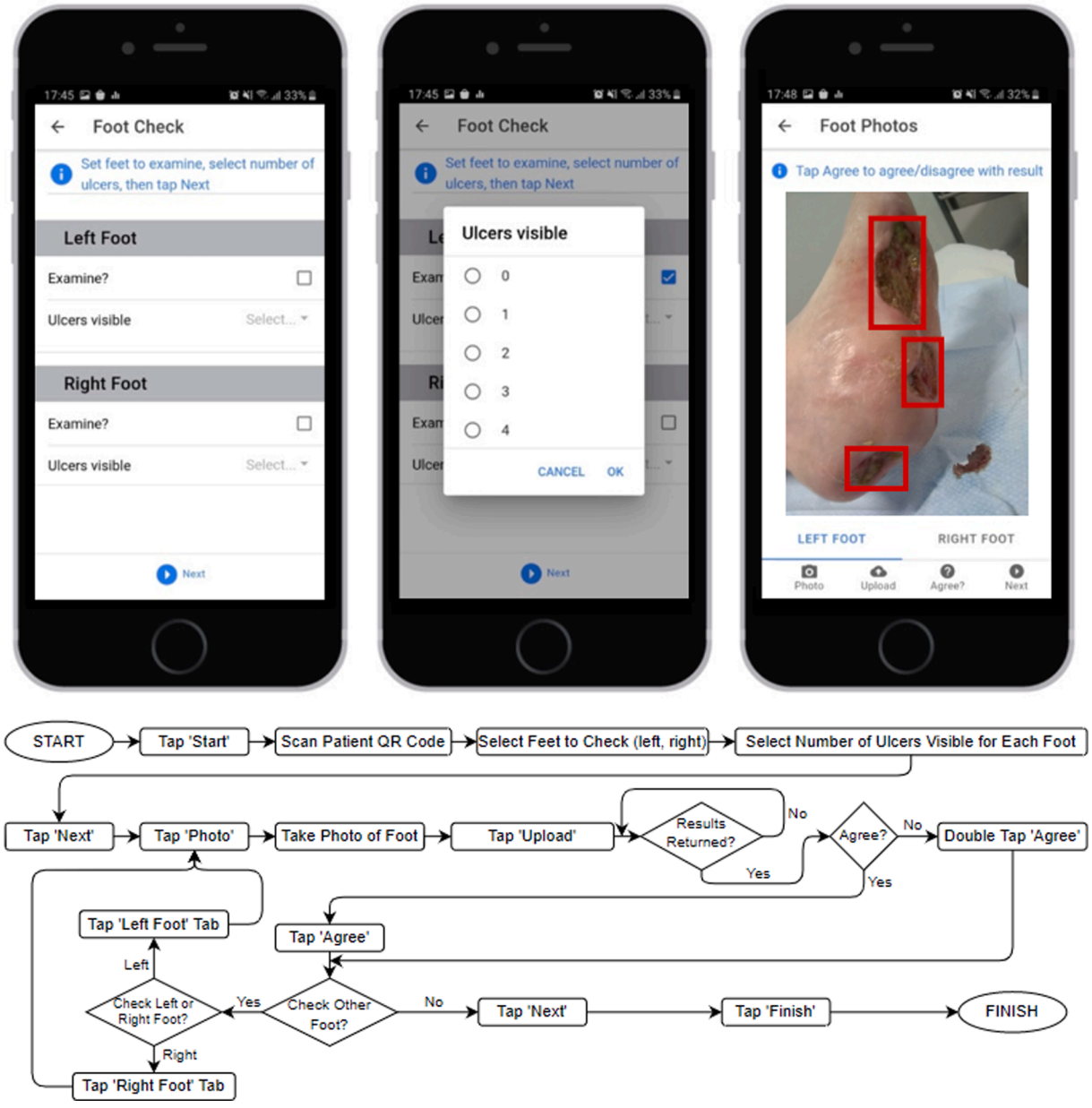


Fig. 1. Illustration of the main data-capture screens in the smartphone app and the clinical procedure for its use in patient appointments during the proof-of-concept clinical evaluation.

1. How many DFU can you see in the photo?
2. How many predictions do you agree with in the photo?
3. How many predictions do you disagree with in the photo?
4. How many DFUs in the photo are undetected?

For each image, we showed the predicted DFU to both expert raters and recorded their clinical decisions and agreement in terms of predictions that they agreed with (True Positive; TP), predictions that they disagreed with (False Positive; FP), missing predictions (False Negative; FN) and photographs without DFU (True Negative; TN). We then derived a confusion matrix from these answers, which allowed for the validation of the AI predictions. We analysed the results in terms of sensitivity, specificity, positive precision and F1-score. We also conducted an inter- and intra-rater reliability analysis to observe reliability between rater's results when rating their agreement with predictions made by the AI algorithm. In medical image analysis, the application of post-processing to further refine prediction results has been shown to reduce the number of FPs [13]. To study the effect of this technique on our data in a second phase of analysis, we included results from this additional experiment to investigate if the AI algorithm accuracy can be further improved by this post-processing. Visual analysis of the rater's results showed that they would classify an intersecting prediction bounding box as a FP (see Fig. 2 (a)). Therefore, in this experiment, we applied a bounding box merging algorithm that merges intersecting prediction bounding boxes. We then adjusted the number of false positives derived from each rater in the sensitivity and specificity measures acquired previously. A total of 5 prediction bounding boxes where identified which intersected with other bounding boxes. We made an assumption that clinicians would identify at least 2 of these predictions and adjust the number of false positives accordingly. An example of prediction bounding box merging is shown in Fig. 2 (a) and (b). Fig. 2 (c) to (g) show other predictions reported by the system, and are discussed further later in the paper.

A summary of baseline characteristics of the data collected during the clinical evaluation is shown in Table 1.

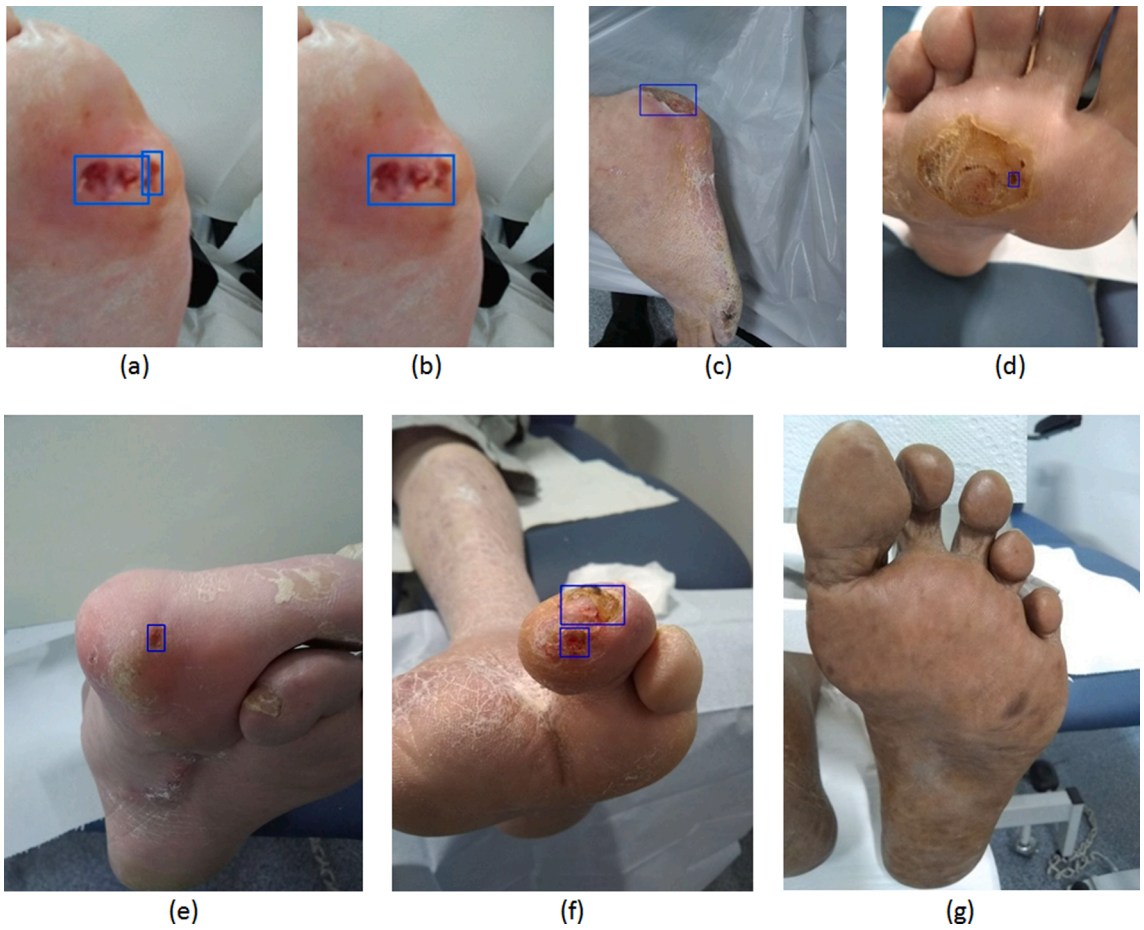


Fig. 2. Illustration showing a selection of notable AI prediction results from the proof-of-concept clinical evaluation.

2.3. Study protocol

Clinicians who participated in the clinical evaluation ($n = 6$) were provided with printed QR Codes (QRCs). Each QRC contained a randomly generated Globally Unique Identifier (GUID) - a series of random alpha-numeric characters. During the patient's initial appointment, the patient QRC was scanned using the app. The clinician then indicated which of the patient's feet they were examining, and how many DFU were visible on each foot (see Fig. 1). The clinician would take a photograph of each foot being examined, which would then be uploaded to the cloud platform and stored in a database, with the diagnosis returned to the app in < 6 s. Results were displayed as red rectangles drawn around each DFU identified by the system (see Fig. 1).

2.4. Qualitative analysis

For the qualitative analysis, we evaluate the following:

1. Sensitivity and specificity measures acquired from two raters who rated their agreement / disagreement with individual AI prediction results.
2. Adjusted specificity measures for the prediction results from the AI algorithm using measures acquired from two raters following an additional post-processing stage to merge intersecting prediction bounding boxes. This step is an experimental adjustment of the results from step 1 where we adjust the number of false positives by a value representing the minority of the total number of intersecting bounding boxes. In this case, 2 out of 5 bounding boxes were merged. We perform this step to emphasise how post-processing could be used for real-world application to enhance the results returned by the AI algorithm.
3. The inter-rater reliability of clinician agreement / disagreement with individual diagnostic results (prediction bounding boxes) returned by the AI algorithm, of which there may be more than one per case.
4. The intra-rater reliability of clinician agreement / disagreement with individual diagnostic results (prediction bounding boxes) returned by the AI algorithm, of which there may be more than one per case.

2.5. Statistical analysis

Statistical analysis to obtain inter-rater (between different clinicians) and intra-rater (between individual clinicians) reliability measures taken from clinicians rating AI diagnostic results was completed using IBM SPSS version 28.0.1.0 (SPSS Inc., Chicago, Illinois). Krippendorff's alpha was used to analyse inter- and intra-rater reliability, and was chosen due to its ability to provide more stable estimates in cases of missing data [27]. Our intra-rater results were missing a small number of ratings ($n = 4$ for rater 1; $n = 3$ for rater 2), therefore Krippendorff's alpha was deemed a more suitable measure so as to provide less biased results. To provide Krippendorff's alpha confidence intervals, a bootstrap value of 10,000 was used. The relevant mathematical expressions for assessing diagnostic accuracy and rater reliability are presented in the appendix.

3. Results

3.1. Experts ratings on the AI algorithm predictions

High sensitivity ($m = 0.92$, $sd = 0.0099$), specificity ($m = 0.89$, $sd =$

Table 1
Baseline characteristics for the data acquired during the proof-of-concept clinical evaluation based on clinical assessment.

Category	Number
No. of patients	81
No. of cases with DFU	162
No. of cases without DFU	41
Pathology	
Control	145
Infection	37
Ischemia	4
Both (infection & ischemia)	2
Undetermined	1
Visible comorbidities	
Amputation	25
Charcot neuropathic osteoarthropathy	3
Calosity	55
Cellulitis	4
Oedema	6
Onychomycosis	5
Cyanosis	15
DFU anatomical location	
Midfoot	11
Lateral midfoot	1
Forefoot	18
Forefoot midfoot	2
Lateral forefoot	21
Lateral midfoot forefoot	1
Medial forefoot	30
Anterior forefoot	9
Anterior lateral forefoot	13
Anterior medial forefoot	38
Hindfoot	12
Hindfoot midfoot	4
Lateral hindfoot	4
Posterior	2
Posterior midfoot	1
Posterior hindfoot	19

0.0080), positive precision and F1-score measures were identified for the diagnostic results returned by the AI algorithm compared to expert clinical decision (Table 2).

3.2. Experts ratings on the AI algorithm predictions with Post-Processing

High sensitivity ($m = 0.92$, $sd = 0.0099$), specificity ($m = 0.93$, $sd = 0.0107$), positive precision, and F1-score measures were identified for the diagnostic results returned by the AI algorithm compared to expert clinical decision following a post-processing stage to merge intersecting prediction results (Table 2).

3.3. Inter-rater reliability

High inter-rater reliability measures were identified for question 1 ($K\alpha = 0.84$), question 2 ($K\alpha = 0.90$), question 3 ($K\alpha = 0.94$), and question 4 ($K\alpha = 0.84$), as shown in Table 3. These measures show the reliability between the two clinician ratings of the AI algorithm predictions for each question. The author of the $K\alpha$ advises that variables with reliabilities above $K\alpha = 0.8000$ can be considered reliable, and that variables with a reliability value between $K\alpha = 0.6670$ and $K\alpha = 0.8000$ should only be used for drawing tentative conclusions [17]. Hence, given the values obtained are all > 0.8000 we are confident in the reliability of clinical decision-making.

Table 3
Inter- and intra-rater reliability measures for the proof-of-concept clinical evaluation. Note - number of cases is 203; Q = question number; Kα = Krippendorff alpha; LB = lower bound; UB = upper bound; CI = Confidence interval; Inter = inter-rater; Intra = intra-rater.

Rater	Q	Kα	Lower Bound 95% CI	Upper Bound 95% CI	Analysis
–	1	0.8329	0.7626	0.9033	Inter
–	2	0.8986	0.8341	0.9539	
–	3	0.9320	0.8640	0.9864	
–	4	0.8338	0.7452	0.9114	
1	1	0.8837	0.8210	0.9374	Intra
1	2	0.9347	0.8881	0.9813	
1	3	0.9256	0.8635	0.9752	
1	4	0.8798	0.8077	0.9519	
2	1	0.8396	0.7641	0.9056	
2	2	0.9132	0.8554	0.9614	
2	3	0.8734	0.7890	0.9437	
2	4	0.8086	0.7130	0.8924	

3.4. Intra-rater reliability

High intra-rater reliability measures were identified for both raters ($m = 0.89$, $sd = 0.0430$), as shown in Table 3. These measures show the reliability of within-clinician ratings of the AI algorithm predictions. This analysis was performed using repeated-measures from the same two raters in the previous inter-rater reliability analysis.

3.5. AI system performance

The mean response time for the cloud-based system to return results to the user on the smartphone identifying any DFU present, was 5.866 s per-case ($sd = 0.747$ s). This indicates a fast and consistent response time for per-case requests in a real-world situation.

4. Discussion

In this paper, we show for the first time, proof-of-concept efficacy for accurate automated DFU detection using a smartphone app employing an AI algorithm operating as part of a cloud-based architecture. The AI algorithm was able to automatically detect DFU with a sensitivity and specificity of 0.92 and 0.89, respectively. These findings come from hospital environments with AI automated decisions/outcomes validated against expert clinical judgement on DFU identification. Furthermore, this process was quick, with only 5.9 s between the foot image being sent to the cloud by the user and identification of any DFU present returned to their smartphone.

The hospital environment was used to provide a high number of DFUs and associated pathology to test the AI algorithm. However, this real-world proof-of-concept clinical evaluation demonstrates the efficacy of the AI system for remote patient screening and monitoring. The future implications being that a patient's spouse or carer could take

Table 2
Measures derived from the expert raters on AI algorithm predictions with and without post-processing. TP – true positive, FP – false positive, TN – true negative, FN – false negative, Sen – sensitivity, Spe – specificity, PP – positive precision, F1 – F1-score.

Rater	TP	FP	TN	FN	Sen	Spe	PP	F1	Post-proc
1	189	5	41	19	0.9087	0.8913	0.9742	0.9403	No
2	155	6	44	13	0.9226	0.8800	0.9627	0.9422	No
1	189	3	41	19	0.9087	0.9318	0.9844	0.9450	Yes
2	155	4	44	13	0.9226	0.9167	0.9748	0.9480	Yes

photographs enabling remote screening and providing patient reassurance. Potentially, such a system could function fully remotely, even without connection to the cloud, whereby the model is installed onto the smartphone itself. In the event a DFU was detected, the system could in theory trigger an alert with appropriate treatment signposting enabling early treatment, reduction of DFU severity and reduced risk of associated infection/amputation. This evaluation purposely used low-end smartphone devices with varying photographic quality to challenge the AI algorithm and demonstrate its efficacy in real-world scenarios.

The high sensitivity of 0.92 indicates that the system correctly identified DFU occurrences when present. Specificity was 0.89, slightly lower than the sensitivity value of 0.92, indicating that the system was better at correctly identifying DFU when present in contrast to correctly rejecting when a patient did not have a DFU, resulting in a slightly higher number of FPs. For the experts' ratings on AI algorithm predictions (see Table 3), the reported SD values indicate that there is little variance between the mean values in each group.

Using post-processing of the AI algorithm predictions in a second phase of analysis (Table 3), we observed that the modest reduction of FPs (by 2) resulted in an increase in specificity from the primary analysis. Specificity increased from 0.89 to 0.93, thereby reducing the number of FPs. Positive precision and F1-score measures also showed a corresponding increase in this secondary analysis.

For the inter- and intra-rater reliability results, we observed that $K\alpha > 0.8$ for all raters in Table 3. The inter- and intra-rater results can therefore be viewed as indicating a high level of reliability between and amongst raters.

We observe from the prediction results that the AI algorithm was capable of accurately detecting DFU wounds on different regions of the foot at various stages of development. Such cases include examples where the AI algorithm was able to detect partially visible DFUs present on the curvature of the foot (shown on the hindfoot in Fig. 2 (c)) and DFUs surrounded by tissue affected by large areas of debridement (shown on the anterior support in Fig. 2 (d)). Note that the bruising on the forefoot was not miss-detected in Fig. 2 (c). Fig. 2 (d) and (e) show two challenging cases (with (e) exhibiting Charcot neuropathic osteoarthropathy) which indicate that the AI algorithm is capable of accurately detecting very small early stage DFUs. Fig. 2 (f) shows another challenging case, exhibiting partial amputation, where the AI algorithm has accurately detected a DFU, but also detects a partially visible toenail that is located next to the wound presenting ambiguous nail and wound boundaries. This may indicate that the AI algorithm requires a larger and more diverse set of training data. Fig. 2 (g) shows the left plantar aspect from a non-white patient in a visually complex medical setting. This result indicates that the AI algorithm is capable of returning true negative results where there may be newly epithelialised skin from a recently healed DFU, shown here on the lateral support region. Of the AI detection results shown in Fig. 2 (c to g), clinicians indicated agreement with (c), (d), (e), and (g), and disagreement with (f). These promising examples indicate that the AI system will be able to detect early stage DFUs, which would reduce complications if used as part of an early screening system.

Accurate detection is the first key stage in automated monitoring of DFU wounds. Without it, longitudinal monitoring and other forms of automated analysis of DFU healing status may be limited. Detection, as described in this paper, allows for the wound region to be isolated from background details and can act as a vital preprocessing stage in automated DFU delineation, which involves the detection of the detailed outline of the wound [22].

4.1. Recommendations and future work

Following the completion of this proof-of-concept clinical evaluation, a more fully-featured version of this AI system is being developed which will integrate into existing NHS healthcare systems in the United Kingdom. DFU has worse outcomes for people living in rural localities when compared to those living in urban areas and remote screening and monitoring of DFU may be able to reduce this health inequality.

By 2035 the global prevalence of diabetes is estimated to rise to almost 600 million, and around 80% of these people will live in developing countries [5]. This proof-of-concept clinical evaluation was conducted primarily on white participants. Therefore, to make the model more relevant globally, we plan to collect substantial datasets from a range of ethnicities to retrain and further enhance the model.

DFU and resulting amputation attracts less public concern, research effort, political consideration, and clinical attention than other conditions with similar impacts on quality of life and survival. Substantial geographic variations exist in service provision, with delays in assessment directly associated with increased DFU severity and longer healing times [16]. Tools to increase early intervention, such as the present AI system, could be important contributors to addressing the growing global challenges presented by DFU.

The intention of this research is to augment medical expertise. This can be especially useful in scenarios where clinicians lack specialty training to accurately diagnose conditions [18] To help combat the trend of overburdened healthcare systems and to maintain high standards of quality in patient care, new technologies could provide a net benefit in terms of both time and costs.

4.2. Summary

This paper presents the results of a proof-of-concept clinical evaluation for a framework that is capable of automatic detection of DFUs using smartphone, cloud and AI technologies. The automated system was shown to provide high sensitivity and specificity, together with excellent inter- and intra-rater reliability. To the best of our knowledge, this is the first system capable of fully automated DFU detection using smartphone, cloud and AI technologies that has been evaluated in clinical settings [21],Cassidy et al. [7]. The intention is to build on this work to bring this technology into the hands of patients and their carers. This solution will be used as a tool to promote regular remote screening of diabetic feet.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Oracle Research for providing the cloud technologies that enabled our proof-of-concept clinical evaluation to take place. Gratitude is also extended to Salford Royal NHS Foundation Trust, Manchester University NHS Foundation Trust and Lancashire Teaching Hospitals NHS Foundation Trust for their extensive support.

Appendix

The relevant mathematical expressions for assessing diagnostic accuracy are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

$$PositivePrecision = \frac{TP}{TP + FP} \tag{3}$$

$$F1 - score = \frac{Sensitivity \times PositivePrecision}{Sensitivity + PositivePrecision} = \frac{2TP}{2TP + FP + FN} \tag{4}$$

where TP is the total number of true positives, TN is the total number of true negatives, FP is the total number of false positives and FN is the total number of false negatives.

The relevant mathematical expression for the general form of Krippendorff's alpha is as follows:

$$Ka = \frac{P_a - P_e}{1 - P_e} \tag{5}$$

where P_a represents the observed weighted percent agreement, and P_e represents the chance weighted percent agreement.

References

[1] Abbott CA, Chatwin KE, Foden P, Hasan AN, Sange C, Rajbhandari SM, et al. Innovative intelligent insole system reduces diabetic foot ulcer recurrence at plantar sites: a prospective, randomised, proof-of-concept study. Lancet Digital Health 2019;1:e308–18. https://doi.org/10.1016/S2589-7500(19)30128-1.

[2] Adam M, Ng EY, Tan JH, Heng ML, Tong JW, Acharya UR. Computer aided diagnosis of diabetic foot using infrared thermography: a review. Comput Biol Med 2017;91:326–36. https://doi.org/10.1016/j.compbiomed.2017.10.030.

[3] Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11:e0156164.

[4] Armstrong DG, Boulton AJ, Bus SA. Diabetic foot ulcers and their recurrence. N Engl J Med 2017;376:2367–75. https://doi.org/10.1056/NEJMra1615439.

[5] Bakker K, Apelqvist J, Lipsky B, van Netten J, Schaper N. The 2015 iwgdf guidance documents on prevention and management of foot problems in diabetes: development of an evidence-based global consensus. Diabetes Metab Res Rev 2016; 32:2–6. https://doi.org/10.1002/dmrr.2694.

[6] Brown R, Ploderer B, Seng L, Lazzarini P, Van Netten J. MyFootCare: a mobile self-tracking tool to promote self-care amongst people with diabetic foot ulcers. JMIR Diab 2017:462–6. https://doi.org/10.1145/3152771.3156158.

[7] Cassidy B, Reeves ND, Pappachan JM, Ahmad N, Haycocks S, Gillespie D, et al. A cloud-based deep learning framework for remote detection of diabetic foot ulcers. IEEE Pervasive Comput 2022;1–9. https://doi.org/10.1109/MPRV.2021.3135686.

[8] Cavanagh P, Attinger C, Abbas Z, Bal A, Rojas N, Xu ZR. Cost of treating diabetic foot ulcers in five different countries. Diabetes Metab Res Rev 2012;28:107–11. https://doi.org/10.1002/dmrr.2245.

[9] Chan KS, Chan YM, Tan AHM, Liang S, Cho YT, Hong Q, et al. Clinical validation of an artificial intelligence-enabled wound imaging mobile application in diabetic foot ulcers. Int Wound J 2021. https://doi.org/10.1111/iwj.13603.

[10] Chan KS, Lo ZJ. Wound assessment, imaging and monitoring systems in diabetic foot ulcers: A systematic review. Int Wound J 2020;17:1909–23. https://doi.org/10.1111/iwj.13481.

[11] Costa R, Cardoso N, Procópio R, Navarro T, Dardik A, Cisneros L. Diabetic foot ulcer carries high amputation and mortality rates, particularly in the presence of advanced age, peripheral artery disease and anemia. Diab Metabol Syndrom: Clin Res Rev 2017;11. https://doi.org/10.1016/j.dsx.2017.04.008.

[12] van Doremalen RF, van Netten JJ, van Baal J, Vollenbroek-Hutten MM, van der Heijden F. Validation of low-cost smartphone-based thermal camera for diabetic foot assessment. Diabetes Res Clin Pract 2019;149:132–9. https://doi.org/10.1016/j.diabres.2019.01.032.

[13] Goyal M, Reeves N, Rajbhandari S, Yap MH. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. IEEE J Biomed Health Inform 2018. https://doi.org/10.1109/JBHI.2018.2868656.

[14] Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. Diabet Med 2012;29:855–62. https://doi.org/10.1111/j.1464-5491.2012.03698.x.

[15] Jodheea-Jutton A, Hindocha S, Bhaw-Luximon A. Health economics of diabetic foot ulcer and recent trends to accelerate treatment. Foot 2022;52:101909. https://doi.org/10.1016/j.foot.2022.101909.

[16] Kerr M, Barron E, Chadwick P, Evans T, Kong W, Rayman G, et al. The cost of diabetic foot ulcers and amputations to the national health service in England. Diabet Med 2019;36:995–1002. https://doi.org/10.1111/dme.13973.

[17] Krippendorff K. Content analysis: an introduction to its methodology. 2nd ed. Sage Publications; 2004.

[18] Liu Y, Jain A, Eng C, Way D, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. Nat Med 2020;26:1–9. https://doi.org/10.1038/s41591-020-0842-3.

[19] Mani R, Margolis DJ, Shukla V, Akita S, Lazarides M, Piaggesi A, et al. Optimizing technology use for chronic lower-extremity wound healing: A consensus document. Int J Low Extrem Wounds 2016;15:102–19. https://doi.org/10.1177/1534734616646261.

[20] Petrova NL, Donaldson NK, Tang W, MacDonald A, Allen J, Lomas C, et al. Infrared thermography and ulcer prevention in the high-risk diabetic foot: data from a single-blind multicentre controlled clinical trial. Diabet Med 2020;37:95–104. https://doi.org/10.1111/dme.14152.

[21] Reeves ND, Cassidy B, Abbott CA, Yap MH. Chapter 7 – novel technologies for detection and prevention of diabetic foot ulcers. In: Gefen A, editor. The Science, Etiology and Mechanobiology of Diabetes and its Complications. Academic Press; 2021. p. 107–22. https://doi.org/10.1016/B978-0-12-821070-3.00007-6.

[22] Scebba G, Zhang J, Catanzaro S, Mihai C, Distler O, Berli M, et al. Detect-and-segment: A deep learning approach to automate wound image segmentation. Inf Med Unlocked 2022;29:100884. https://doi.org/10.1016/j.imu.2022.100884.

[23] Statista, 2019. Smartphone sales in the united states from 2005 to 2019 (in billion u.s. dollars). URL: https://www.statista.com/statistics/191985/sales-of-smartphones-in-the-us-since-2005/ [last access: 03/12/19].

[24] Wang L, Pedersen PC, Agu E, Strong DM, Tulu B. Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification. IEEE Trans Biomed Eng 2017;64:2098–109. https://doi.org/10.1109/TBME.2016.2632522.

[25] Wound Desk Mobile Enhanced Wound Management last access 01/12/20 2021 URL https://wounddesk.com/.

[26] Yap MH, Chatwin KE, Ng CC, Abbott CA, Bowling FL, Rajbhandari S, et al. "FootSnap": A new mobile application for standardizing diabetic foot images. J Diabetes Sci Technol 2017;12. https://doi.org/10.1177/1932296817713761.

[27] Zapf A, Castell S, Morawietz L, Karch A. Measuring interrater reliability for nominal data - which coefficients and confidence intervals are appropriate? BMC Med Res Method 2016;16. https://doi.org/10.1186/s12874-016-0200-9.

[28] Qin Y, Wu J, Xiao W, Wang K, Huang A, Liu B, et al. Machine Learning Models for Data-Driven Prediction of Diabetes by Lifestyle Type. Int. J. Environ. Res. Public Health 2022;19:15027. https://doi.org/10.3390/ijerph192215027.

[29] Kushwaha S, Srivastava R, Jain R, Sagar V, et al. Harnessing machine learning models for non-invasive pre-diabetes screening in children and adolescents. Comput. Methods Programs Biomed. 2022;226:107180. https://doi.org/10.1016/j.cmpb.2022.107180.

[30] Bendifallah S, Puchar A, Suisse S, Delbos L, et al. Machine learning algorithms as new screening approach for patients with endometriosis. Sci. Rep. 2022;12:639. https://doi.org/10.1038/s41598-021-04637-2.

[31] Mensah MA, Ott CE, Horn D, Pantel JT. A machine learning-based screening tool for genetic syndromes in children. Lancet Digit Health 2022;4(5):e295. https://doi.org/10.1016/S2589-7500(22)00050-4.