

University of Groningen

Artificial intelligence and visual inspection in cervical cancer screening

Nakisige, Carolyn; de Fouw, Marlieke; Kabukye, Johnblack; Sultanov, Marat; Nazrui, Naheed; Rahman, Aminur; de Zeeuw, Janine; Koot, Jaap; Rao, Arathi P; Prasad, Keerthana

Published in:
International Journal of Gynecological Cancer

DOI:
[10.1136/ijgc-2023-004397](https://doi.org/10.1136/ijgc-2023-004397)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Nakisige, C., de Fouw, M., Kabukye, J., Sultanov, M., Nazrui, N., Rahman, A., de Zeeuw, J., Koot, J., Rao, A. P., Prasad, K., Shyamala, G., Siddharta, P., Stekelenburg, J., & Beltman, J. J. (2023). Artificial intelligence and visual inspection in cervical cancer screening. *International Journal of Gynecological Cancer*, 33, 1515-1521. <https://doi.org/10.1136/ijgc-2023-004397>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



OPEN ACCESS

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/ijgc-2023-004397>).

For numbered affiliations see end of article.

Correspondence to

Dr Carolyn Nakisige, Gynaecologic Oncology, Uganda Cancer Institute, Kampala, Uganda; carolyn.nakisige@uci.or.ug

Received 23 February 2023

Accepted 7 August 2023

Published Online First

4 September 2023



© IGCS and ESGO 2023. Re-use permitted under CC BY-NC. No commercial re-use. Published by BMJ.

To cite: Nakisige C, de Fouw M, Kabukye J, et al. *Int J Gynecol Cancer* 2023;**33**:1515–1521.

Artificial intelligence and visual inspection in cervical cancer screening

Carolyn Nakisige ¹,^{ID} Marlieke de Fouw,² Johnblack Kabukye,³ Marat Sultanov,⁴ Naheed Nazrui,⁵ Aminur Rahman,⁶ Janine de Zeeuw,⁴ Jaap Koot,⁴ Arathi P Rao,⁷ Keerthana Prasad,⁸ Guruvare Shyamala,⁹ Premalatha Siddharta,¹⁰ Jelle Stekelenburg,⁴ Jogchum Jan Beltman¹¹

ABSTRACT

Introduction Visual inspection with acetic acid is limited by subjectivity and a lack of skilled human resource. A decision support system based on artificial intelligence could address these limitations. We conducted a diagnostic study to assess the diagnostic performance using visual inspection with acetic acid under magnification of healthcare workers, experts, and an artificial intelligence algorithm.

Methods A total of 22 healthcare workers, 9 gynecologists/experts in visual inspection with acetic acid, and the algorithm assessed a set of 83 images from existing datasets with expert consensus as the reference. Their diagnostic performance was determined by analyzing sensitivity, specificity, and area under the curve, and intra- and inter-observer agreement was measured using Fleiss kappa values.

Results Sensitivity, specificity, and area under the curve were, respectively, 80.4%, 80.5%, and 0.80 (95% CI 0.70 to 0.90) for the healthcare workers, 81.6%, 93.5%, and 0.93 (95% CI 0.87 to 1.00) for the experts, and 80.0%, 83.3%, and 0.84 (95% CI 0.75 to 0.93) for the algorithm. Kappa values for the healthcare workers, experts, and algorithm were 0.45, 0.68, and 0.63, respectively.

Conclusion This study enabled simultaneous assessment and demonstrated that expert consensus can be an alternative to histopathology to establish a reference standard for further training of healthcare workers and the artificial intelligence algorithm to improve diagnostic accuracy.

INTRODUCTION

The WHO global strategy to accelerate the elimination of cervical cancer addressed targets that must be met by 2030. The strategy aims to upscale vaccination, screening, and treatment of cervical pre-cancer and cancer.¹ Only 41% of low-income countries administer the human papillomavirus (HPV) vaccine as part of their immunization program, yet these countries suffer from the highest burden of cervical cancer morbidity and mortality.² Therefore, screening and treatment of pre-cancer remain top priorities. In limited resource settings, visual inspection with acetic acid (VIA) is the most common screening method. It is relatively simple to use, cheap, and allows screening and treatment in one visit.^{3–5}

Screening programs based on VIA need a functioning healthcare infrastructure and well-trained

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Visual inspection with acetic acid is limited by subjectivity and a lack of skilled human resource. Artificial intelligence has been applied to improve diagnostic accuracy.

WHAT THIS STUDY ADDS.

⇒ In the absence of pathology, expert opinion can be used as a reference in training the artificial intelligence algorithm for cervical cancer screening in low- and middle-income countries. The algorithm has the potential to provide quality and objective decisional support in screening for cervical cancer in low resource settings.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Improving the diagnostic accuracy of the algorithm might enable task shifting of screening with the potential to increase coverage and adherence to follow-up. The algorithm will be required to differentiate cancers from pre-cancers and identify the squamo-columnar junction to guide treatment decisions for ablation or excision and referral.

personnel.^{6,7} The diagnostic accuracy of this method is dependent on the skills of the healthcare worker. Its sensitivity and specificity vary with a sensitivity of 62.5–80% and a specificity of 80–98.8% for the detection of histologically confirmed cervical intra-epithelial neoplasia (CIN) II or more advanced cervical lesions against the reference standard of histology or colposcopy followed by biopsy.^{8–10} Given its subjectivity, VIA is associated with a problem of over treatment in a single-visit approach and at the same time possibly failing to identify women who are at high risk for cervical cancer.^{11,12}

Recently, artificial intelligence (AI) systems have increasingly been applied in healthcare including cervical cancer screening.¹³ Studies in both high-income and low-income countries indicate that the application of AI in HPV testing,^{14–16} cytology,^{17–21} and colposcopy^{22–27} has achieved a good detection rate of pre-cancerous lesions with good accuracy. However, few studies investigated the potential of AI to improve the accuracy of VIA and function as a decision support system during screening in the primary care setting.

Table 1 Summary of characteristics of the participants

Team	Participants n (%)	Age (years) median (IQR)	VIA experience (years) median (IQR)
Healthcare workers			
Bangladesh	5 (23)	37 (30–42)	9 (8.50–11.0)
India	7 (32)	42 (35–51)	8 (3.0–15.0)
Uganda	10 (45)	27(25–33)	(0)*
Experts			
Bangladesh	2 (22)		
India	4 (44)		
Uganda	2 (22)		
Netherlands	1 (11)		
Total	9 (100)	52 (38–56)	23 (10.0–26.0)

There were 22 healthcare workers from Bangladesh, India, and Uganda. The median (IQR) were calculated per each country team of healthcare workers.

*Only one healthcare worker out of 10 in Uganda had had VIA experience of 3 years. The team of 9 experts from 4 countries had a median age of 52 years (IQR 38–56) and median years of experience of 23 years (IQR 10.0–26.0).

VIA, visual inspection with acetic acid.

We investigate an AI-Decision Support System for VIA in three countries as part of a European and Indian funded research project, namely Prevention and Screening Innovation Project Towards Elimination of Cervical cancer (PRESCRIP-TEC, www.prescriptec.org).²⁸ Besides the introduction of the AI Decision Support System, the project enhances country-specific cervical cancer prevention programs with focused awareness strategies, community mobilization, and HPV self-testing to increase uptake and quality of screening.

In Uganda, India, and Bangladesh we assessed the baseline quality of VIA of trained healthcare workers, experienced gynecologists, and the AI by measuring their diagnostic accuracy and intra- and inter-observer agreement on magnified VIA images.

METHODS

We conducted a cross-sectional diagnostic study to assess the diagnostic performance in VIA-M (VIA with magnification) by healthcare workers, experts, and the AI. The healthcare workers' team consisted of 10 members from Uganda, five from Bangladesh, and seven from India. They had been trained in VIA according to their respective national guidelines.

The expert team consisted of nine gynecologists with more than 10 years' expertise in screening and ablative treatment. Two experts originated from Bangladesh, four from India, one from the Netherlands, and two from Uganda. All were selected by their respective countries (Table 1).

The AI Decision Support System was developed by Manipal School of Information Sciences, Manipal Academy of Higher Education in India and trained on 100 images after application of acetic acid collected from routine VIA clinics. The images were divided into three sets: training, testing, and validation in the proportion 70:20:10, respectively. There was no overlap between the images in these three sets. The gold standard was a single best expert's report. The algorithm is built into an Android-based device to enable

healthcare workers to introduce the device during screening in field conditions.²⁹ The algorithm will generate an instant report after capturing the image, and can differentiate between a normal cervix (negative) and an abnormal cervix requiring further evaluation (positive).

Databank and Selection of Images

Cervical images from existing databanks of the Uganda Cancer Institute, International Agency for Research on Cancer,³⁰ and Leiden University Medical Center were used to create a new dataset after obtaining consent from each institute. A total of 96 images were selected, in the proportions of 13, 73 and 10, respectively. Four images were purposively duplicated to assess intra-observer agreement making a total of 100 images that were presented for assessment. The expert team of nine gynecologists did not participate in the selection of these images to avoid recall bias.

Data Collection

An online tool to upload cervical images and questionnaires was developed by the Marconi AI laboratory in Makerere University, Uganda.³¹ The images were uploaded to this annotation platform which the experts and healthcare workers independently accessed on computer monitors. The paired cervical images for each case (before and after application of acetic acid) were presented in a single frame along with a brief questionnaire on the quality, VIA assessment, and if rated 'positive' eligibility for ablative therapy (Online supplemental table S1). These images were viewed on a computer monitor which enables VIA-M.

The AI ran on a computer instead of the Android-based device, assessed the images only after application of acetic acid, and provided a binary result (positive or negative).

Procedure

The healthcare workers and experts each created a personal account and provided personal data, including age, country, and

years of experience in screening. Participants were invited to complete the assessment of cervical images individually within a period of 1 week. Before starting they received a video and letter with guiding instructions. The project manager of each country team ensured all healthcare workers filled out the form individually without consulting sources like the internet or the screening manual. After completion, access to the forms was automatically locked.

The images were stored in a folder in a computer and the AI analyzed those images and reported them as being negative or positive.

The gold standard or reference was based on expert consensus due to lack of pathology. An expert consensus meeting was conducted after the images had been assessed individually by all the healthcare workers, experts, and the algorithm. Consensus was reached when at least five out of the nine experts agreed on the VIA assessment. Throughout the consensus meeting, the experts were not aware of the gold standard of the original databanks. In four images the expert consensus was different from the gold standard of the original databanks. All experts collectively agreed that the four images were VIA negative while the original databank stated VIA positive. Therefore, the VIA assessment of these four images was changed to VIA negative. It turned out that the histology of all four images was normal.

In four other images of the final dataset, results were inconclusive based on the individual expert assessment. The cut-off of at least five of nine was not attained to determine their final grades. These were re-evaluated by the consensus group and all experts agreed on grades which turned out to be similar to the original grades.

During the meeting, three images were excluded from analysis; one image was of insufficient quality for evaluation due to poor lighting, and two images were taken at follow-up after initial treatment with the Loop Electro-surgical Excision Procedure. These two images were excluded because women with previous cervical treatment will be excluded from the implementation study. Additionally, the algorithm could not assess 10 out of the 93 images due to file format issues. Therefore, a total of 83 unique images were included for final analysis (Figure 1). Of the 83 images, 24 had histopathological diagnoses: six normal, four CIN I, two CIN II, three CIN III, two with features of HPV infection, and seven squamous cell carcinomas. The expert consensus ended up with 48 VIA negative images, 26 VIA positive images, and nine suspected cancers. All cancer images were tagged positive to enable comparison with the binary outcome of the AI, resulting in 35 VIA positive images.

Analysis

Analysis was done using Statistical Package for Social Science version 26 where diagnostic performance and intra/inter-observer agreement were analyzed using sensitivity, specificity, area under the curve (AUC), false positives and false negatives as well as Fleiss kappa (κ) values.

Diagnostic Performance

Diagnostic performance was assessed by comparing the individual assessment, the majority vote of the assessment within the teams of healthcare workers, and the assessment of the algorithm to that of the expert consensus.

We evaluated the diagnostic accuracy of healthcare workers, individual experts, and the AI using sensitivity, specificity, receiver operating characteristics (ROC), and its summary statistic AUC. True positives, true negatives, false positives, and false negatives of the individual country teams were also reported. Feedback on their performance was given and the healthcare workers re-trained on the same images in their settings.

Intra-/Inter-Observer Agreement

We determined the intra-observer and inter-observer agreement within teams of healthcare workers, experts, and the algorithm using Fleiss κ values.³² Activities under the PRESCRIP-TEC were approved by the institutional review boards of Bangladesh, India, and Uganda and written informed consent obtained from both the healthcare workers and experts.

RESULTS

Characteristics of the Participants

A total of 31 individuals and the AI participated in this study. Table 1 illustrates their country of work, age, and VIA experience. The experience varied among the teams, from median 8 years (IQR 3.0–15.0) in India to 9 years (IQR 8.50–11.0) in Bangladesh, while Uganda had only one healthcare worker with 3 years of experience. The rest of the Uganda team had no experience prior to the training for the project.

Diagnostic Performance

Table 2 shows that the sensitivity of healthcare workers assessment was 80.4% and specificity 80.5%, and the sensitivity of expert assessment was 81.6% and specificity 93.5%. The algorithm demonstrated a sensitivity of 80.0% and a specificity of 83.3%. Online supplemental figure S1 demonstrates the AUC, which was 0.80 (95% CI 0.70 to 0.90) for the healthcare workers, 0.93 (95% CI 0.87 to 1.00) for the experts, and 0.84 (95% CI 0.75 to 0.93) for the AI.

Table 3 illustrates the relationship between the AUC and diagnostic accuracy.³³ Individual performance is presented in Online supplemental table S2.

Intra-/Inter-Observer Agreement

There was 100% intra-observer agreement of all healthcare workers and experts concerning the four duplicate images. Healthcare worker agreement within teams was moderate with κ values of 0.43, 0.44, and 0.48 in Bangladesh, India, and Uganda, respectively. Overall κ values for all healthcare workers, experts, and the algorithm were 0.45, 0.68, and 0.63, respectively. When excluding outliers in the team of healthcare workers, defined as having κ scores <0.5, agreement among healthcare workers was substantial with $\kappa=0.63$ (Table 4).

DISCUSSION

Summary of Main Results

We found that the diagnostic performance of healthcare workers was adequate and enabled sufficient quality of screening. The algorithm performed better than the healthcare workers but lower than the experts. Agreement between teams was comparable.

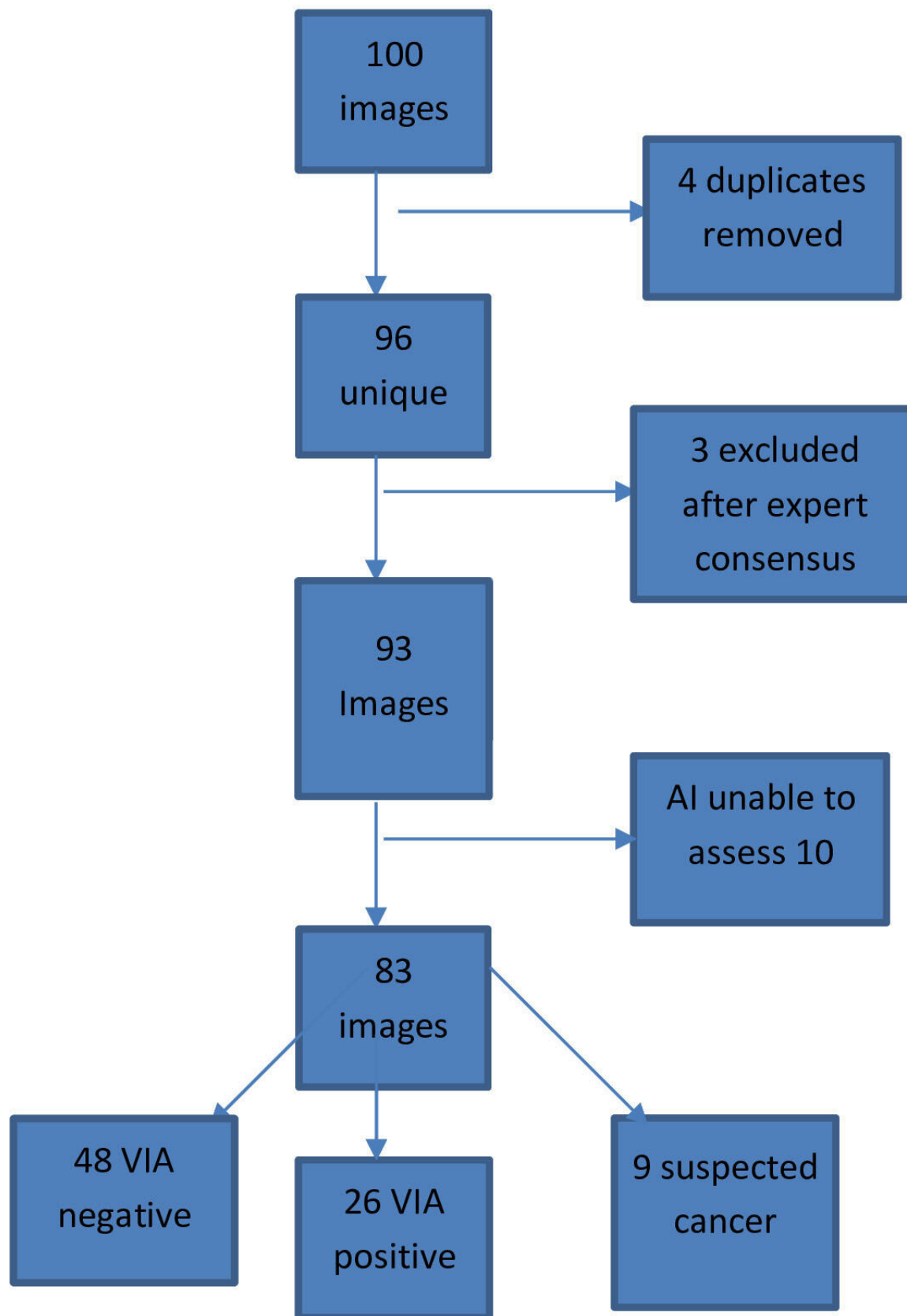


Figure 1 Flow chart of the VIA dataset assessed, comprising 96 unique images and four duplicates. Three were excluded at an expert consensus meeting, leaving 93. Ten were excluded after AI failed to assess them, leaving 83 images for final assessment. These comprised 48 VIA negative images, 28 VIA positive images, and nine suspected cancers. The nine suspected cancers were added to the VIA positives to make 35 VIA positive images and 48 VIA negatives to enable comparison with the binary outcome of the AI. AI, artificial intelligence; VIA, visual inspection with acetic acid.

Table 2 Diagnostic performance of respective teams

	Expert consensus	Original dataset*	Expert team	AI	India	Uganda	Bangladesh	All HCWs
Total number (n)	83	–	9	1	7	10	5	22
VIA positive images (n)†	–	39	351	39	273	390	195	858
VIA positive assessment (n)	35	35	257	28	210	248	161	619
VIA negative images (n)†	–	44	396	44	308	440	220	968
VIA negative assessment (n)	48	48	404	40	261	417	172	850
Sensitivity (%)	–	92.1	81.6	80.0	85.7	70.9	92.0	80.4
Specificity (%)	–	100	93.5	83.3	77.7	86.9	71.7	80.5
False positive (%)	–	7.9	6.5	16.7	22.3	13.1	28.3	19.5
False negative (%)	–	0	18.4	20.0	14.3	29.1	8.0	19.6
AUC	–	–	0.932	0.840	0.813	0.773	0.831	0.800

*Refers to images from the Uganda Cancer Institute and the International Agency for Research in Cancer datasets.

†The number of VIA positive and VIA negative images refers to the total number of VIA positive images that was presented to the respective group; for example, the expert team consisted of 9 experts, therefore they were presented with $9 \times 35 = 315$ VIA positive images. AI, artificial intelligence; AUC, area under the curve; HCWs, healthcare workers; VIA, visual assessment with acetic acid.

Results in the Context of Published Literature

The diagnostic performance of the healthcare workers compared well to the sensitivity and specificity described in two reviews which mentioned sensitivities of 80% and 73.2% and specificities of 92% and 86.7%.^{9,10} In clinical practice the diagnostic performance might be less comparable to research settings. It was remarkable that the Ugandan team had the highest specificity, leading to fewer false positive results and lower risk of over treatment, while the sensitivity was low compared with the team in India and Bangladesh. The high sensitivity of the Bangladesh and Indian teams translates to few false negative results at the cost of possible over diagnosis and over treatment. This could be attributed to the higher prevalence of HPV infections and higher VIA positivity rate in Uganda compared with India and Bangladesh (unpublished project data). Furthermore, the overall performance measured by the AUC reflected the number of years' experience with VIA which was highest in Bangladesh and lowest in Uganda.

The overall performance of the experts compared favorably with previous studies.^{8,9} The wide range of diagnostic performance was

similar to what Vidya et al reported among a group of gynecologists and residents where sensitivity was 57.1–92.9% and specificity was 54.3–94.5%. The wide variation could be due to varying experiences of the experts and the subjectivity of VIA assessment.²⁹ The sensitivity and specificity of the AI was lower than its previous performance after initial training.²⁹ The lesser performance could be due to being tested on images from various sources with expert consensus as reference compared with the assessment of a single expert who was the reference during the training of the AI. To mitigate the risk of over fitting during the development phase of the algorithm, images were divided into three sets of training, validation and testing without overlap however, the total number was small.

Interestingly, the AI identified all images with suspected or invasive cancer as positive, although it was not trained on such images. It is trained only on images after application of acetic acid and this procedure is not performed when cancer is suspected. Further training of the AI including images of suspected cervical cancer will probably improve its diagnostic accuracy and applicability in the field.

The intra-observer agreement of the four images for all the participants was 100%, which is very rare but could be attributed

Table 3 Relationship between AUC and diagnostic accuracy of a test.³³

Area (AUC)	Diagnostic accuracy
0.9–1.0	Excellent
0.8–0.9	Very good
0.7–0.8	Good
0.6–0.7	Sufficient
0.5–0.6	Bad
< 0.5	Not useful

Statistic of AUC for a test ranges between 0–1. The closer to 0, the poorer the diagnostic accuracy of a test is, while the closer to 1 the more accurate the test is. This is calculated using receiver operator characteristics. AUC, area under the curve.

Table 4 Interpretation of agreement by kappa (κ) values³⁴

κ value	Strength of agreement
< 0.0	No agreement
0.01–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

The measure of agreement between and among entities was calculated using Fleiss κ values and ranges from no agreement to almost perfect agreement.

Original research

to the very small number of duplicate images. The inter-observer agreement of all healthcare workers was moderate, with an overall weighted κ value of 0.454. The moderate level of agreement could be due to the varying levels of experience between and within country teams, the different exposure to refresher training and supervision, and the very low specificity and sensitivity of a few individuals. After excluding the five outliers, overall weighted κ values improved from 0.454 to 0.625, thus from moderate to substantial agreement. Among the outliers with lower sensitivity and specificity were two healthcare workers without previous VIA experience. Therefore, we will conduct ongoing supervision and regular performance evaluations to continuously assess the quality of screening.

Strength and Weaknesses

This study is unique in evaluating the diagnostic performance of VIA-M among both experts and healthcare workers from different low resource settings simultaneously. The use of images from publicly available databases could have introduced recall bias. We mitigated this risk by adding pictures from different datasets, presenting the pictures in random order, and checking for intra-observer agreement with duplicate images, although the number was small. This presented varying gold standards, making comparison difficult. We developed a common expert consensus gold standard applicable to all the datasets.

We based our gold standard on expert consensus, given that pathology is not readily available in the study settings, while the desired reference in cervical cancer screening studies is histopathology. We used a team of nine experts to ensure the quality of the expert consensus; however, the bigger the number the more difficult it is to reach consensus. Screening by healthcare professionals will provide three possible outcomes, namely positive, negative, or suspected cancer. The AI provides only two possible outcomes, positive and negative. Given that detection of suspected cancers has great clinical importance, in future studies we hope to introduce a trinary outcome for the algorithm.

Implications for Practice and Future Research

During project implementation, the outcome of screening with VIA will be based on the assessment of healthcare workers with supervision provided by experts. At the same time, images will be captured on the Android-based device with the algorithm to further train the AI and validate its performance in field conditions.

Over time, the diagnostic performance of the AI and healthcare workers will be re-evaluated to assess the effect of training, regular feedback, and supervision. The data generated will confirm feasibility of the use of the AI as a decision support system in the field for cervical cancer screening. The AI could lead to task shifting of screening to less-trained healthcare providers, after rigorous validation in field conditions. Task shifting will increase the opportunities to provide quality screening to women, especially in resource constrained settings, and will potentially translate into increased uptake of screening and adherence to follow-up.

We recommend for future research to assess the feasibility of detection of suspected cancer and the eligibility for direct treatment with thermal ablation by the AI, and evaluate the perception of screened women and healthcare providers about the implementation of AI in screening programs. In addition, we recommend to

evaluate the use of expert consensus as an alternative to histology as the gold standard for training the AI, especially in settings where pathology is not readily accessible.

CONCLUSION

The diagnostic accuracy in VIA-M of the healthcare workers and the AI was good, though lower than the experts. Agreement within teams was moderate to substantial. This signifies need for further training of the healthcare workers and the algorithm to improve diagnostic accuracy and agreement, with strict validation of the AI to maximize its performance in the field and allow task shifting of screening to less trained healthcare workers. This study showed that in the absence of pathology, expert consensus can be used as a reference to train healthcare workers and the AI.

Author affiliations

- ¹Gynaecologic Oncology, Uganda Cancer Institute, Kampala, Uganda
- ²Gynecology, Leiden University Medical Center department of Gynecology, Leiden, Zuid-Holland, Netherlands
- ³Uganda Cancer Institute, Kampala, Uganda
- ⁴University Medical Center Groningen, University of Groningen, Groningen, Netherlands, Groningen, Netherlands
- ⁵FRIENDSHIP, Dhaka, Bangladesh
- ⁶ICDDR Public Health Sciences Division, Dhaka, Dhaka District, Bangladesh
- ⁷Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, India, Manipal, India
- ⁸Manipal Academy of Higher Education School of Life Sciences, Manipal, Karnataka, India
- ⁹Manipal Academy of Higher Education - Mangalore Campus, Mangalore, Karnataka, India
- ¹⁰Gynecological Oncology, St John's National Academy of Health Sciences, Bangalore, Karnataka, India
- ¹¹Gynaecology, LUMC, Leiden, Netherlands

Collaborators Collaborator group name: PRESCRIP-TEC. Individual author names: Carolyn Nakisige, Marlieke de Fouw, Johnblack Kabukye, Naheed Nazrul, Aminur Rahman, Marat Sultanov, Janine de Zeeuw, Jaap Koot, Arathi Ra, Keerthana Prasad, Shyamala Guruvare, Premalatha Siddharta, Ranajit Mandal, Jelle Stekelenburg, Jogchum Beltman.

CN* Guarantor Conceptualization: CN, MF, JbK, JZ, JK, JB. Data curation: CN, MS. Formal analysis: CN, MS. Methodology: CN, MF, JbK, JK, JB. Supervision: MF, JbK, JZ, JK, JS, JB. Review of manuscript: MF, JbK, JZ, JK, KP, MS, SG, JS, JB.

Funding Prevention and Screening Innovation Project – Towards Elimination of Cervical Cancer (PRESCRIP-TEC) is a research consortium project delivered through a collaboration of 15 consortium members. This project has received funding from the European Union's Horizon 2020 research and innovation program grant agreement No 964270 and from the Ministry of Science and Technology, Department of Biomedical Technology in India, grant No 13213, under the Global Alliance for Chronic Diseases. International Agency for Research on Cancer (IARC), Leiden University Medical Centre (LUMC), and Uganda Cancer Institute (UCI) for availing the images used. Manipal Academy for Higher Education (MAHE) in India for availing the algorithm. Marconi laboratory in Makerere University, Uganda for providing the online tool. All the healthcare workers and experts for their time and voluntary effort for this study.

Competing interests KP and SG were involved in development of the Artificial Intelligence algorithm.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by 1. Uganda Cancer Institution Research Ethics Committee (UCIREC) reference number UCI-2021-29; 2. Uganda National Council for Science and Technology (UNCST) reference number HS2222ES. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, an indication of whether changes were made, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Carolyn Nakisige <http://orcid.org/0000-0002-1497-8056>

REFERENCES

- World Health Organization. Global strategy to accelerate the elimination of cervical cancer as a public health problem. 17 Nov 2020/global strategy;
- Tsu VD, LaMontagne DS, Atuhebwe P, *et al*. National implementation of HPV vaccination programs in low-resource countries: lessons, challenges and future prospects. *Prev Med* 2021;144:106335.
- Sankaranarayanan R, Shyamalakumary B, Wesley R, *et al*. Visual inspection with acetic acid in the early detection of cervical cancer and precursors. *Int J Cancer* 1999;80:161–3.
- Denny L, Kuhn L, Pollack A, *et al*. Evaluation of alternative methods of cervical cancer screening for resource-poor settings. *Cancer* 2000;89:826–33.
- Belinson JL, Pretorius RG, Zhang WH, *et al*. Cervical cancer screening by simple visual inspection after acetic acid. *Obstet Gynecol* 2001;98:441–4.
- Pollack AE, Tsu VD. Preventing cervical cancer in low-resource settings: building a case for the possible. *Int J Gynaecol Obstet* 2005;89 Suppl 2:S1–3.
- World Health Organization. *Cervical cancer screening in developing countries: report of a WHO consultation*. Geneva: World Health Organization, 2002.
- Sauvaget C, Fayette JM, Muwonge R, *et al*. Accuracy of VIA for Cervical cancer screening. *Review Int J Gynecol Obstet* 2011;Apr 113:14–24.
- Catarino R, Schäfer S, Vassilakos P, *et al*. Accuracy of combinations of visual inspection using acetic acid or Lugol's iodine to detect cervical pre-cancer: a meta-analysis. *BJOG* 2018;125:545–53.
- Qiao L, Li B, Long M, *et al*. Accuracy of visual inspection with Lugol's iodine for cervical cancer screening: a meta-analysis. *J Obstet Gynaecol Res* 2015;41:1313–25.
- Denny L, Kuhn L, De Souza M, *et al*. Screen and treat approaches for cervical cancer prevention in low-resource settings: a randomized controlled trial. *JAMA* 2005;294:2173–81.
- Sankaranarayanan R, Nene BM, Shastri SS, *et al*. HPV screening for cervical cancer in rural India. *N Engl J Med* 2009;360:1385–94.
- Hou X, Shen G, Zhou L, *et al*. n.d. Artificial intelligence in cervical cancer screening and diagnosis. *Front Oncol*;12.
- Wong OGW, Ng IFY, Tsun OKL, *et al*. Machine learning interpretation of extended human papillomavirus genotyping by Onclarity in an Asian cervical cancer screening population. *J Clin Microbiol* 2019;57:e00997–19.
- Pathania D, Landeros C, Rohrer L, *et al*. Point-of-care cervical cancer screening using deep learning-based microholography. *Theranostics* 2019;9:8438–47.
- Tian R, Cui Z, He D, *et al*. Risk stratification of cervical lesions using capture sequencing and machine learning method based on HPV and human integrated genomic profiles. *Carcinogenesis* 2019;40:1220–8.
- Holmström O, Linder N, Kaingu H, *et al*. Point of care digital cytology with artificial intelligence for cervical cancer screening in a resource limited setting. *JAMA Netw Open* 2021;4:e211740.
- Bao H, Sun X, Zhang Y, *et al*. The artificial intelligence-assisted cytology diagnostic system in large-scale cervical cancer screening: a population-based cohort study of 0.7 million women. *Cancer Med* 2020;9:6896–906.
- Bao H, Bi H, Zhang X, *et al*. Artificial intelligence-assisted cytology for detection of cervical intraepithelial neoplasia or invasive cancer: a multicenter, clinical-based, observational study. *Gynecol Oncol* 2020;159:171–8.
- Wang C-W, Liou Y-A, Lin Y-J, *et al*. Artificial intelligence-assisted fast screening cervical high grade squamous intraepithelial lesion and squamous cell carcinoma diagnosis and treatment planning. *Sci Rep* 2021;11.
- Zhu X, Li X, Ong K, *et al*. Hybrid AI-assistive diagnostic model permits rapid TBS classification of cervical liquid-based thin-layer cell smears. *Nat Commun* 2021;12.
- Xue P, Ng MTA, Qiao Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. *BMC Med* 2020;18:169.
- Hu L, Bell D, Antani S, *et al*. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J Natl Cancer Inst* 2019;111:923–32.
- Cho B-J, Choi YJ, Lee M-J, *et al*. Classification of cervical neoplasms on colposcopic photography using deep learning. *Sci Rep* 2020;10:13652.
- Asiedu MN, Simhal A, Chaudhary U, *et al*. Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, pocket colposcope. *IEEE Trans Biomed Eng* 2019;66:2306–18.
- Miyagi Y, Takehara K, Nagayasu Y, *et al*. Application of deep learning to the classification of uterine cervical squamous epithelial lesions from colposcopy images combined with HPV types. *Oncol Lett* 2020;19:1602–10.
- Xue P, Tang C, Li Q, *et al*. Development and validation of an artificial intelligence system for grading colposcopic impressions and guiding biopsies. *BMC Med* 2020;18:406.
- Sultanov M, Zeeuw J de, Koot J, *et al*. Investigating feasibility of 2021 WHO protocol for cervical cancer screening in underscreened populations: prevention and screening innovation project toward elimination of cervical cancer (PRESCRIP-TEC). *BMC Public Health* 2022;22:1356.
- Kudva V, Prasad K, Guruvare S. Andriod device-based cervical cancer screening for resource-poor settings. *J Digit Imaging* 2018;31:646–54.
- International Agency of Research on Cancer. Cervical cancer image bank to aid artificial intelligence cancer detection– IARC screening group;
- The Marconi Society machine learning laboratory, . *College of Engineering, Design Art and Technology*. Makerere University, Kampala, Uganda,
- Kudva V, Guruvare S, Prasad K, *et al*. Inter observer variability among gynecologists in manual cervix image analysis for detection of cervical epithelial abnormalities. *Clinical Epidemiology and Global Health* 2019;7:500–3.
- Simundic A-M. Measures of diagnostic accuracy: basic definitions. *eJIFCC* 2008;19:203–11.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.