Computer-assisted Diagnosis (CAD) in Colposcopy: Evaluation of a Pilot Study

GRIT MEHLHORN[1*], ANDREAS KAGE[2*], CHRISTIAN MÜNZENMAYER[2], MICHAELA BENZ[2],
MARTIN C. KOCH[1], MATTHIAS W. BECKMANN[1] and THOMAS WITTENBERG[2]

[1]Department of Gynecology, Erlangen University Hospital, Erlangen, Germany;
[2]Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany

Abstract. *Background: A prototype system for computer-assisted colposcopic diagnosis (CAD) currently achieves a high level of accuracy of 80% (sensitivity 85%, specificity 75%) for the automatic assessment of colposcopic images. This pilot study investigated whether this type of CAD system is, in principle, capable of influencing the quality of the examiner's assessment. Materials and Methods: In this observer study, 24 digitized colposcopic images from patients attending a dysplasia clinic were assessed by 90 participants. All participants had attended a colposcopy training workshop so that they acquired the same basic information and skills. Results: Wide variation was seen among the non-experts, in contrast to the experts. An overall improvement in diagnostic accuracy was noted when the CAD system was used (non-experts: sensitivity 78%, specificity 70%; experts: sensitivity 74%, specificity 70%). Conclusion: The CAD system may serve as an aid in the further diagnosis of cervical intraepithelial neoplasia, and has the potential to improve the diagnostic process.*

Worldwide cervical cancer is the third most common type of cancer in women, and the seventh overall, with an estimated 530,000 new cases in 2008; Northern America (US, Canada) and Europe contribute 75,000 of these cases. Medical follow-up is recommended for approximately 7 million women with abnormal Pap smears (1). The international gold-standard for the diagnostic assessment is histological analysis (2). The markedly improved visual conditions provided by colposcopy, along with the use of various contrast-enhancing stains (such as acetic acid) may lead to more precise diagnostic clarification, on the basis of a targeted-biopsy (3, 4). Nevertheless, colposcopy had a sensitivity of 86% and a specificity of 30% in distinguishing a healthy cervix from that with cervical intraepithelial neoplasia or carcinoma; in distinguishing healthy cervix with low-grade lesions (CIN 1) from that with high-grade lesions (CIN 2/3) or carcinoma, colposcopy had a sensitivity of 61% and a specificity of 94% (5). Particular attention needs to be given to the implications of this type of differential colposcopy for subsequent procedures, as positive findings lead to a punch biopsy being taken – so that the colposcopy examination determines to some extent whether a histological diagnosis will be carried out and, thus, also influences the type of treatment that will be offered (6, 7). Against this background, the question arises of whether differential diagnosis of cervical neoplasias can be improved with the aid of semi-automatic evaluation of colposcopic images by a computer-assisted diagnosis (CAD) system. Pre-clinical studies have shown that this type of CAD system can be quite helpful to the physician as a diagnostic aide in assessing and establishing the diagnosis (12-17). The present pilot study investigated, to on knowledge, for the first time whether the use of this type of CAD system can improve or provide support for the diagnosis of cervical neoplasia.

Materials and Methods

CAD system. A prototype evidence-based system for CAD was developed to allow for interactive differentiation between abnormal and unremarkable intraepithelial neoplasias in the uterine cervix based on colposcopic images (13, 14). In a previously unknown colposcopic image, which can be captured from the colposcopy camera using a frame-grabber or loaded from a storage device and displayed on the monitor, for example, a questionable lesion is interactively marked by the user with a computer mouse. In the image areas marked in this way (regions of interest, ROIs), specially developed procedures for color texture analysis are used to calculate numerical features characterizing the corresponding surface structures on the basis of intensities, structures, textures, and color (15, 16). The unknown lesion is classified by comparing the similarity of the numerical

*These Authors contributed equally to this research.

Correspondence to: Grit Mehlhorn, MD, Department of Gynecology, Erlangen University Hospital, Universitätsstrasse 21-23, 91054 Erlangen, Germany. Tel: +49 91318533553, Fax: +49 91318536890, e-mail: grit.mehlhorn@uk-erlangen.de

Table I. *Overview of the mean classification results (means and variance) for non-experts and experts, with/without the aid of the computer-assisted diagnosis (CAD).*

Participants	n	CAD	Mean overall classification (%)	Mean sensitivity (%)	Mean specificity (%)
Non-experts	60	Without	71.9±7.8	75,2±12,6	68.6±12.7
		With	76.0±8.6	79.7±12.7	72.9±13.7
Experts	30	Without	69.4±6.9	70.8±16.6	68.1±13.6
		With	73.3±6.3	74.2±13.1	72.5±13.5

features calculated from it with the features of lesions in a histologically-validated reference data collection of colposcopies (17). On the basis of this similarity classification, the screen shows the user whether the unknown lesion is abnormal or unremarkable, along with a value for the degree of certainty with which the CAD system has reached this decision. In addition to this suggestion, the CAD system also displays the most similar images from the reference data collection, so that the proposed classification is transparent to the user (Figure 1). This principle of looking for similarities and presenting the most similar cases, is known in the literature as content-based image retrieval (CBIR) (18).

Reference data collection and accuracy of the CAD system. The CAD system was trained and methodologically validated on the basis of a reference from a collection of 198 colposcopic images (with 376 representative image areas marked) of the Department of Gynecology of the Erlangen University Hospital. On the basis of the histological findings of dysplasia from results of the punch biopsies, the images for the selected cases were classified into groups, with one image for each patient (19). At the current stage of the system's development, the accuracy for identifying the correct correlation between abnormal and normal findings in comparison with conventional colposcopic diagnosis has a sensitivity of 85% and a specificity of 75% (19).

Observer study. An observer study with 90 participants was conducted during the annual Erlangen Colposcopy Courses from 2009-2011 (Figure 1). The study participants consisted of 30 experts, defined as having more than 3 years' experience in colposcopy; and 60 non-experts, defined as having fewer than 3 years' experience. All of the participants had acquired the same basic information and skills during the two-day colposcopy training course. For the observer study, 24 new digital colposcopic images from 24 patients attending the dysplasia clinic at the University of Erlangen's Department of Gynecology were selected in relation to the premalignant or benign nature of the cervix depicted. As the automatic analysis of the colposcopic images is currently based exclusively on the visual characteristics of the images, only images that had a magnification step identical to that used in the reference data collection were included in the study (19). The test images include six images from patients with normal imaging findings, six images with low-grade intraepithelial lesions corresponding to grade 1 cervical intraepithelial neoplasia (CIN I) and 12 test images from patients with higher-grade intraepithelial lesions (CIN II–III). The image was captured one to two minutes after the application of acetic acid. The cytological, colposcopic, and dysplastic histological findings were all consistent. The 24 test images were randomly arranged, although the sequence was identical for all participants. The task for the study

participants was to assign the total of 24 test images with which they were presented to one of the two groups, A: normal, CIN I; or B: CIN II-III. In each case, the user was initially presented with a colposcopic image (top left in Figure 1) with the relevant lesion marked on it in color. In addition, users were also asked, in each case, to give details of their own assessment of the diagnostic certainty. Three options (uncertain, neutral, certain) were provided for this. When the decision-making process was completed by the user pressing on the Confirm button, the CAD system showed a diagnostic suggestion, the degree of certainty associated with the classification, and the five most similar reference images from the database (Figure 2, right), with the reference lesions marked on them. On the basis of the additional information shown, users were then able to revise their diagnoses or to have their original opinion supported by the system's suggestion. The associated degree of diagnostic certainty perceived by the user ("uncertain", "neutral", "certain") was then again requested. The diagnostic process for the image was then started by pressing the Confirm button. In addition to using the prototypic CAD system for the observer study, all participants completed a questionnaire.

Results

Results of the observer study are shown in Table I. In comparison with the initially studies mentioned, the mean overall certainty in the group of experts, without use of the CAD system was relatively high, at 70%. However, there was more variation among the non-experts in comparison with the experts. The mean overall certainty of the non-experts, without the CAD system was 72% – also relatively high, but hardly different in comparison with the experts. When using the CAD system, both the expert and non-expert groups showed an overall improved mean classification performance (non-experts from 72% to 76%; experts from 69% to 73%). Improved mean rates for sensitivity and specificity were also observed in both groups (non-experts: sensitivity from 75% to 79%, specificity from 69% to 73%; experts: sensitivity from 71% to 74%, with specificity from 68% to 72%). Analysis of the questionnaires showed that more than 70% of the study participants were able to envisage using the CAD system as an aid to diagnosis. More than 81% of the participants felt greater certainty during the diagnostic work-up when using the CAD system (Figure 3a). More than 70% of the participants thought they were able to complete the diagnostic procedure more quickly when using

Figure 1. *The observer study, with participants of the colposcopy course at the METEAN Medical Technology Test and Demonstration Center by the Fraunhofer Institute (IIS).*

Figure 2. *The user interface in the observer study.*

this type of system (Figure 3b). Most of the participants (78%) stated that they could imagine using a CAD system as an aid to diagnosis in the future (Figure 3c). More than 89% of the participants said they could imagine using this type of system particularly for training purposes (Figure 3d).

Discussion

This study regarding the use of a prototype CAD system in the diagnosis of colposcopy images showed that – both for inexperienced, non-expert participants and also for more experienced expert users – the system made it easier to distinguish between unremarkable and slightly abnormal cervical findings of little clinical relevance, on the one hand, and higher-grade cervical findings on the other. The levels of diagnostic certainty observed (Table I), were much higher here in comparison with the findings reported by Hammes *et al*. for diagnosis by two experts in extended colposcopy, who were only able to achieve a maximum sensitivity of 61% with a specificity of 94% (5). One possible explanation for this observation might be the intensive two-day colposcopy course given immediately before the study, in which new methods
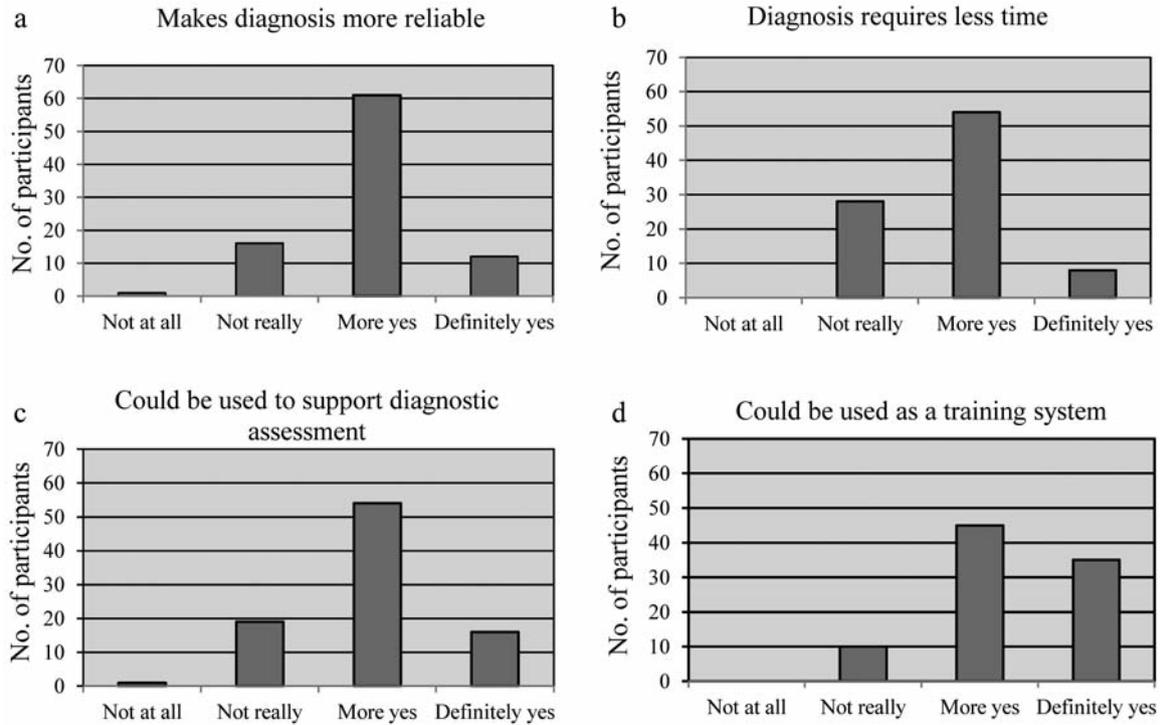
Figure 3. *Responses showing the degree of agreement with the statements "I can imagine that working with the system in clinical practice would lead to diagnoses being made more reliably" (a); "I can imagine that working with the system in clinical practice would lead to diagnoses being made more efficiently, with less time needed" (b); "I can imagine using this type of system (as shown here) as an aid in diagnosis" (c); and "I can imagine using this type of system (as shown here) for courses/regular training events" (d).*

were presented and participants were able to improve their existing knowledge and skills. Examining colposcopic images using the prototype CAD system (GynCAD) for computer-assisted diagnosis gave both non-experts and experts increased diagnostic certainty, with a mean increase of three to five percentage points. The majority of participants were non-experts. The trend towards improved sensitivity when the CAD system was used emphasizes the system's ability to play a supportive role. Although it is not yet possible to measure statistical significance with the results of this pilot study, the enhancement of conventional, purely visual colposcopy demonstrated here is comparable with results reported in the literature and can be regarded as showing a positive trend.

In the Authors' view, a further increase in the overall recognition rate, sensitivity, specificity, and capacity to differentiate low-grade from severe changes in the cervix would be possible both by standardizing the image capture procedure and by further developing the specific imaging features used to characterize the lesions (5). In particular, the potential increase in sensitivity provided by supplementary usage of a CAD system suggests decisive clinical relevance with regard to biopsy clarification and subsequent surgical treatment, as the numbers of unnecessary biopsies could

thereby be reduced. Using digital colposcopy, Schädel *et al.*, among others, investigated similar approaches to automated means of improving diagnostic processes and thereby increasing quality control (20).

No differences between the two groups (non-experts/experts) were seen when they were directly compared with regard to the maximum overall diagnostic certainty achieved when using the CAD system. The quality of diagnosis improved in both groups, but the overall results were similar (76% *vs.* 73.0%) in assessing the colposcopy images (Table I). However, the greater level of inter-observer variation among the non-experts in comparison with the experts, clearly showed that the group consisted of beginners. The difference in accuracy between the participants of both groups is only marginal. One possible cause is the number of colposcopic test images. Using a greater number of images, a significant difference between diagnoses with and without CAD would be expected.

To the best of our knowledge, this is the first observer study to have been carried out in the field of differential colposcopy using a prototype CAD system. It shows that the supplementary use of the CAD system makes it possible to achieve a clear improvement in overall diagnostic certainty for clinically-relevant pre-cancerous lesions in the cervix.

Standardization of the image capture procedure – producing colposcopy images with a constant image quality, standard lighting, and standardized sectional views – might help improve the CAD approach, and the development and use of new, dedicated features for characterizing tissue might also reveal further differences between low-grade and high-grade tissue changes in the cervix. This was confirmed by the study participants in their completed questionnaires. Most of the participants were able to use the support provided by the system successfully and considered that they might use this type of CAD system, with its greater reliability and shorter assessment time, as a diagnostic support in the future.

Computer-assisted colposcopic diagnosis of the type used for the first time by the participants at the Erlangen Colposcopy Course is a new supplementary diagnostic method in which the CAD system is used to provide a second opinion. The opportunity to check one's own conclusions on the basis of the reference images displayed also means that it is possible for the system described here to be used as a training and teaching system (Figure 3). This type of CAD system cannot and should not replace examiners, compete with them, nor substitute for their extensive experience. The colposcopic image seen after the acetic acid solution is applied, is still, and will continue to be, merely a snapshot. A complete colposcopic examination using acetic acid staining, however, is a *dynamic* process of a type that is not currently capable of being analyzed and supported by a CAD colposcopy system. However, CAD systems such as the one used in the present study, have tremendous potential in a supportive role in further diagnosis, as shown here in relation to CIN. CAD systems will not be able to compete with the diagnostic confirmation provided by histology, but they may be able to reduce the need for histological examinations. In principle, CAD can make it possible in cases of diagnostic uncertainty to use an objective analysis of features in order to systematically improve colposcopy-based diagnosis. The use of standardized imaging techniques to acquire images, reproducible cases, and the availability of immediate feedback from the user to the system make the application of this type of CAD system appear feasible both for direct diagnosis and also indirectly for training purposes, and further intensive research on such systems should be carried out.

Acknowledgements

References

1 Ferlay J, Shin HR, Bray F, Forman D, Mathers C and Parkin DM: GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10.Lyon, France: International Agency for Research on Cancer; Available from: http://globocan.iarc.fr.2010. Last accessed 2012.

2 Deutsche Gesellschaft für Gynäkologie und Geburtshilfe: S2k-Leitlinie zur Prävention, Diagnostik und Therapie der HPV-Infektion und präinvasiven Läsionen des weiblichen Genitale. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) 2008, no. 015/027. Available at: http://leitlinien.net/. Last accessed 2012.

3 Bornstein J, Bentley J, Bosze P, Girardi F, Haefner H, Menton M, Perrotta M, Prendiville W, Russell P, Sideri M, Strander B, Torne A and Walker P: 2011 colposcopic terminology of the international Federation for Cervical Pathology and Colposcopy. Obstet Gynecol *120(1)*: 166-72, 2012.

4 Jeronimo J and Schiffman M: Colposcopy at a crossroads. Am J Obstet Gynecol *195(2)*: 349-353, 2006.

5 Hammes LS, Naud P, Passos EP, Matos J, Brouwers K, Rivoire W and Syrjänen KJ: Value of the international Federation for Cervical Pathology and Colposcopy (IFCPC) Terminology in predicting cervical disease. J Low Genit Tract Dis *11(3)*: 158-165, 2007.

6 Massad LS and Collins YC: Strength of correlations between colposcopic impression and biopsy histology. Gynecol Oncol *89(3)*: 424-428, 2003.

7 Pretorius RG, Bao YP, Belinson JL, Burchette RJ, Smith JS and Qiao YL: Inappropriate gold standard bias in cervical cancer screening studies. Int J Cancer *121(10)*: 2218-2224, 2007.

8 Münzenmayer C, Mühldorfer S, Mayinger B, Volk H, Grobe M and Wittenberg T: Farbtexturbasierte optische Biopsie auf hochauflösenden endoskopischen Farbbildern des Ösophagus. *In*: Bildverarbeitung für Medizin. Wittenberg T, Hastreiter P, Hoppe U, Handels H, Horsch A and Meinzer HP (eds.). Berlin, Springer, pp. 191-195, 2003.

9 Rajan P, Canto M, Gorospe E, Almario A, Kage A, Winter C, Hager T, Wittenberg T and Münzenmayer CL: Automated diagnosis of Barrett's esophagus with endoscopic images. *In*: Proceedings of the World Congress on Medical Physics and Biomedical Engineering. Dössel O and Schlegel WC (eds). Berlin, Springer, vol 25/4, pp. 2189-2192, 2009.

10 Zopf S, Kage A, Münzenmayer C, Wittenberg T, Hahn EG and Raithel M: Narrow-band imaging for the computer-assisted diagnosis in patients with Barrett's esophagus. Gastrointest Endosc *69(5)*: AB376, 2009.

11 Zopf S, Kage A, Münzenmayer C, Wittenberg T, Neurath T and Raithel M: Narrow-band imaging für die Computer-unterstüzte Diagnose bei Patienten mit Barrett-Ösophagus. Endosk Heute *23(1)*: 77, 2010.

12 Kothe C, Münzenmayer C, Wittenberg T and Hess M: Erfahrungen mit optischen Stimmlippenbiopsien bei Leukoplakien. Laryngo-Rhino-Otologie *84(2)*: 92-95, 2005.

13 Kage A, Münzenmayer C and Wittenberg T: A knowledge-based system for the computer assisted diagnosis of endoscopic images. *In*: Bildverarbeitung für die Medizin 2008: Algorithmen - Systeme - Anwendungen. Proceedings des Workshops vom 6. bis 8. April in Berlin. Tolxdorff T, Braun J, Deserno TM, Handels H, Horsch A and Meinzer HP (eds.). Berlin, Springer, pp. 272-276, 2008.

14 Wittenberg T, Kage A, Benz M, Winter C, Koch M, Beckmann M, Münzenmayer C and Mehlhorn G: GynCAD – a computer-assisted diagnosis approach for colposcopy. *In*: Abstracts from the Demonstration Workshop on Computer-Aided Detection and Diagnosis, Computer Assisted Radiology and Surgery (CARS). 24th International Congress and Exhibition, Geneva, 23-26 June 2010. Horsch A (ed.). Geneva, CARS, pp. 7-8, 2010.

15 Manjunath B, Ohm JR, Vasudevan VV and Yamada A: Color and texture descriptors. IEEE Trans Circuits Syst Video Technol *11(6)*: 703-715, 2001.

16 Münzenmayer C, Kage A, Wittenberg T and Mühldorfer S: Computer-assisted diagnosis for precancerous lesions in the esophagus. Methods Inf Med *48(4)*: 324-330, 2009.

17 Michie D, Spiegelhalter DJ and Taylor CC: Machine Learning, Neural and Statistical Classification. New York, Ellis Horwood, 1994.

18 Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, Ney H, Kohnen M, Schubert H and Wein BB: Content-based image retrieval in medical applications. Methods Inf Med *43(4)*: 354-361, 2004.

19 Mehlhorn G, Münzenmayer C, Benz M, Kage A, Beckmann MW and Wittenberg T: Computer-assisted diagnosis (CAD) in colposcopy – results of a preliminary experiment? Acta Cytologica *56(5)*: 554-559, 2012.

20 Schädel D, Coumbos A, Willrodt RG, Roggan A, Jochum T, Müller G, Albrecht H und Kühn W: Digitale Kolposkopie bei Läsionen der Cervix uteri – eine Pilotstudie unter Berücksichtigung telematischer Fragestellungen. Geburtshilfe Frauenheilkd *64(12)*: 1205-1212, 2004.