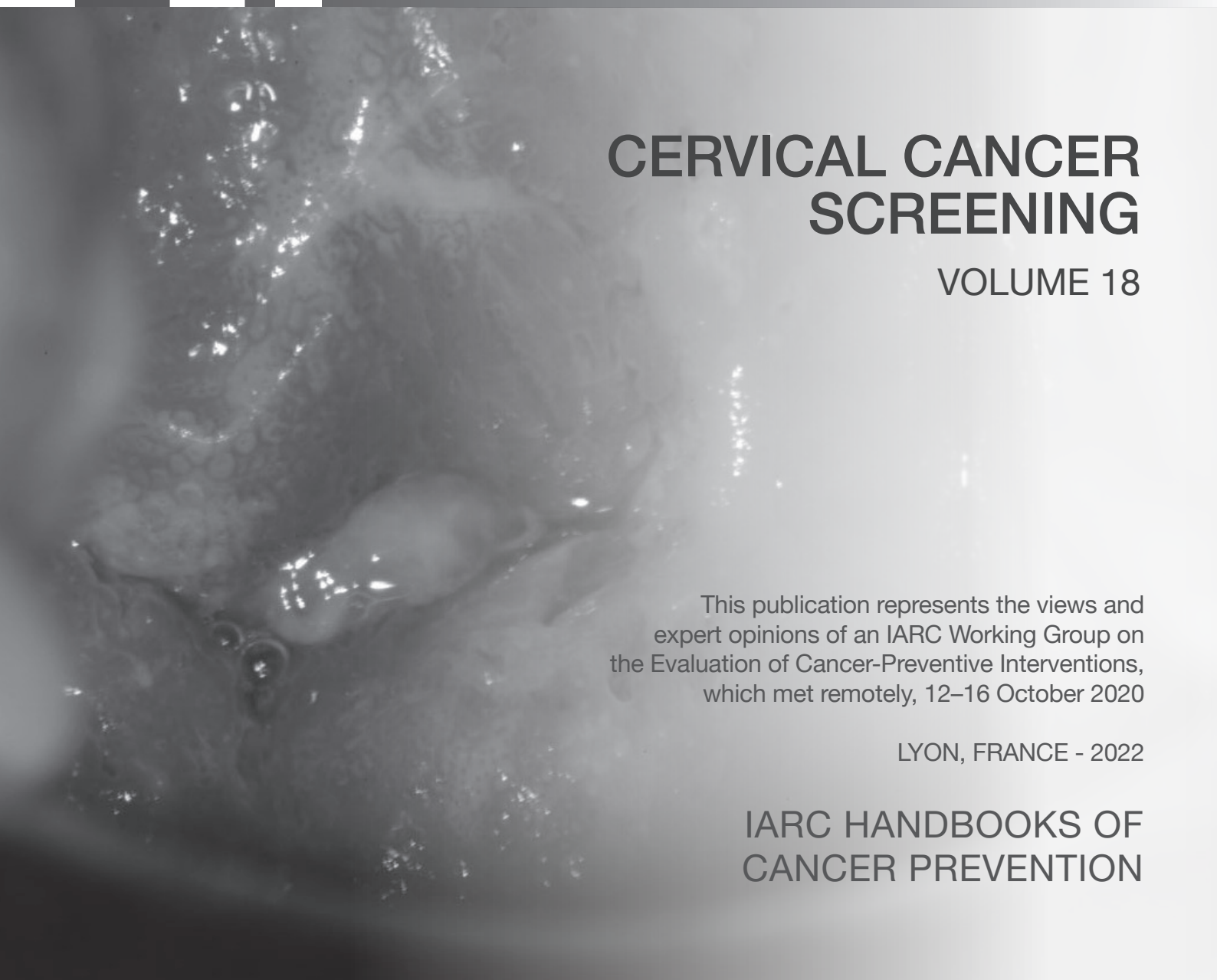
IARC HANDBOOKS



CERVICAL CANCER SCREENING

VOLUME 18

This publication represents the views and expert opinions of an IARC Working Group on the Evaluation of Cancer-Preventive Interventions, which met remotely, 12–16 October 2020

LYON, FRANCE - 2022

IARC HANDBOOKS OF CANCER PREVENTION

International Agency for Research on Cancer

World Health Organization

4.6 Emerging technologies

Recent advances in understanding of HPV-associated carcinogenesis have led to the development and evaluation of many new technologies and approaches for cervical cancer screening, triage, management, and diagnosis. Three types of approaches for the detection of cervical precancer are distinguished: those based on visual, cytological, and molecular technologies.

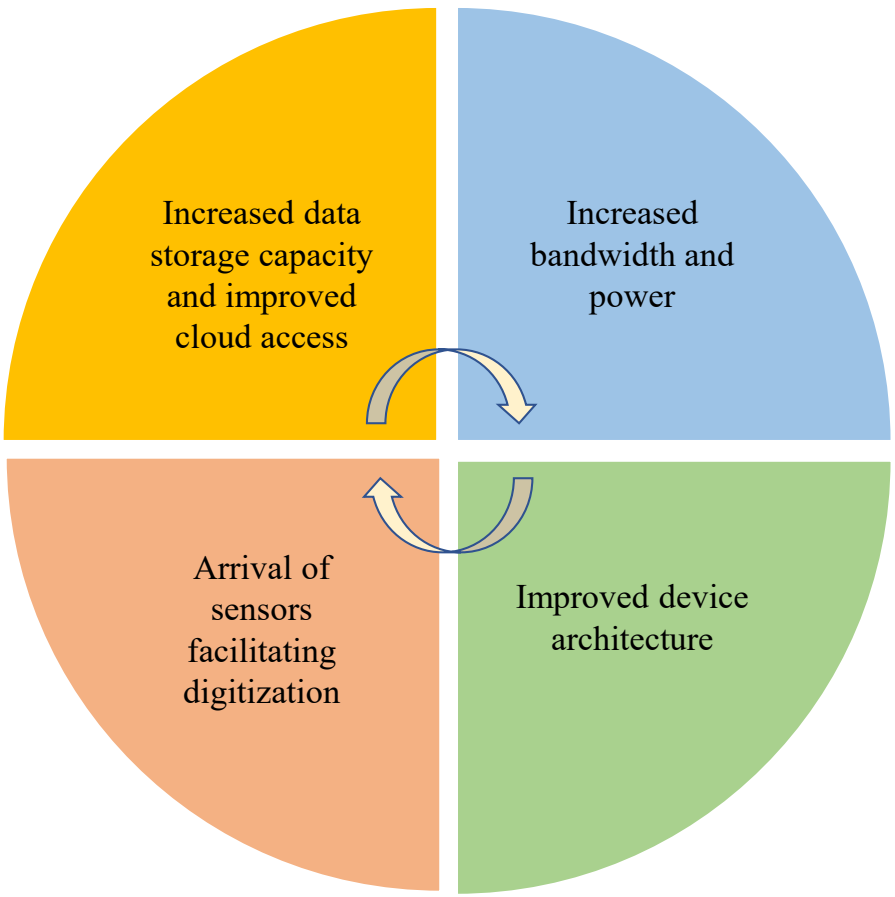
Several systematic approaches to assess the potential use of a biomarker in cervical cancer screening and management have been proposed (Arbyn et al., 2009; Wentzensen & Wacholder, 2013). Established guidelines for diagnostic research (the Standards for Reporting of Diagnostic Accuracy Studies [STARD] statement) have been adapted for technology development for cervical cancer screening (Arbyn et al., 2009). Five phases of technology evaluation are formally distinguished: (1) preclinical exploratory studies, (2) clinical validation studies, (3) retrospective biobank studies in the target population, (4) prospective screening studies, and (5) prospective intervention studies. Although this framework provides important guidance for technology development, not all of these steps are required for all technologies, and the sequence may vary depending on the clinical indication and the availability of suitable research studies. The evaluation of a technology must occur in the context of its potential use, because diagnostic accuracy requirements differ depending on whether the technology is used in screening, triage, or disease confirmation. Here, the term "emerging technology" is used when the discovery processes have been completed and the early steps of technology evaluation are under way (i.e. phases 1–3).

The process from discovery and development to clinical implementation is complex and involves many stakeholders, including researchers, industry, regulatory authorities, and professional societies that develop guidelines (Wentzensen & Silver, 2016). It can take a long time from initial discovery to clinical implementation. For example, HPV DNA testing was initially developed in the 1980s but did not enter clinical practice until 20 years later. The timeline from discovery to clinical practice is now shorter, because of the better understanding of the natural history of cervical cancer and the much accelerated technology development.

Because most discovered biomarkers do not make it into clinical practice, it is important to identify likely failures early in the evaluation process, enabling researchers to focus on the most promising leads (Wentzensen & Wacholder, 2013). The most important criterion for a biomarker is whether the test result will improve clinical management; if not, the test may be useless. Successful biomarker development usually relies on a commercial party to invest in assay development and regulatory approval. Therefore, barriers to bringing a promising biomarker into clinical practice may be the lack of intellectual property, or relatively limited clinical indication, which may result in too small a commercial market.

Of the molecular technologies summarized here, some were developed several years ago but have not been sufficiently validated for consideration of clinical use or have not been translated from the research setting to a commercially available test, for various reasons. Other novel technologies are rapidly progressing through the evaluation process, such as AI-based visual and cytological methods, as well as host and viral DNA methylation markers, which can be expected to appear in extensive clinical validation studies very soon.

Fig. 4.8 Pathway to the development of new technologies



Created by the Working Group.

4.6.1 Emerging technologies using artificial intelligence

AI is having an impact on many scientific disciplines, including medicine. As the power of computer software has increased, the size of the hardware has decreased, and as Internet bandwidth and electronic storage capacity have improved, it has become possible to deliver accurate image-recognition systems in very small, cloud-independent devices that incorporate comprehensive systems for management of clinical data and images (Fig. 4.8). Convolutional neural networks (CNNs) are commonly used for the analysis and classification of visual images; they are increasingly being used in medical diagnostics, such as in the classification of benign or malignant lung tumours (Hussein et al., 2017), in skin cancer (Esteva et al., 2017), in retinopathy (Ting et al., 2017), in the classification of colorectal polyps (Wei et al., 2020), in breast cancer (McKinney et al., 2020), and in the detection of cardiological abnormalities (Islam et al., 2017). Recently, these approaches have also been applied to automated and biomarker-enhanced cervical cytology (Schiffman et al., 2017; Wentzensen et al., 2021).

(a) AI-based automated visual evaluation

Even with adequate training and quality assurance measures in place, visual inspection of the cervix is a highly subjective procedure, including determining the adequacy of the examination, the type of the TZ, and the diagnostic impression. Furthermore, comprehensive training to the level of independent practice can take 6–18 months. A major but not exclusive part of this training is in image recognition, which to date has been learned largely within a live clinical setting. The concept of training a computer to recognize abnormality by "learning" the relevant features from a large image bank of known histopathology has obvious appeal. If that computational power can be harnessed in small, inexpensive, and user-friendly image-capture systems, the inadequacies of current visual examination methods could be addressed without the need for expensive training or adjunctive systems. As a laboratory-independent and reusable device, this technology could replace or complement current visual-based screening and triage approaches in LMICs. It may also negate the need for individual colposcopy expertise in screen-positive women who are not suitable for ablative treatment as part of a screen-and-treat protocol. AI can be used innovatively to train service providers and for quality control. Currently, no system has been properly evaluated in a live or real-world setting.

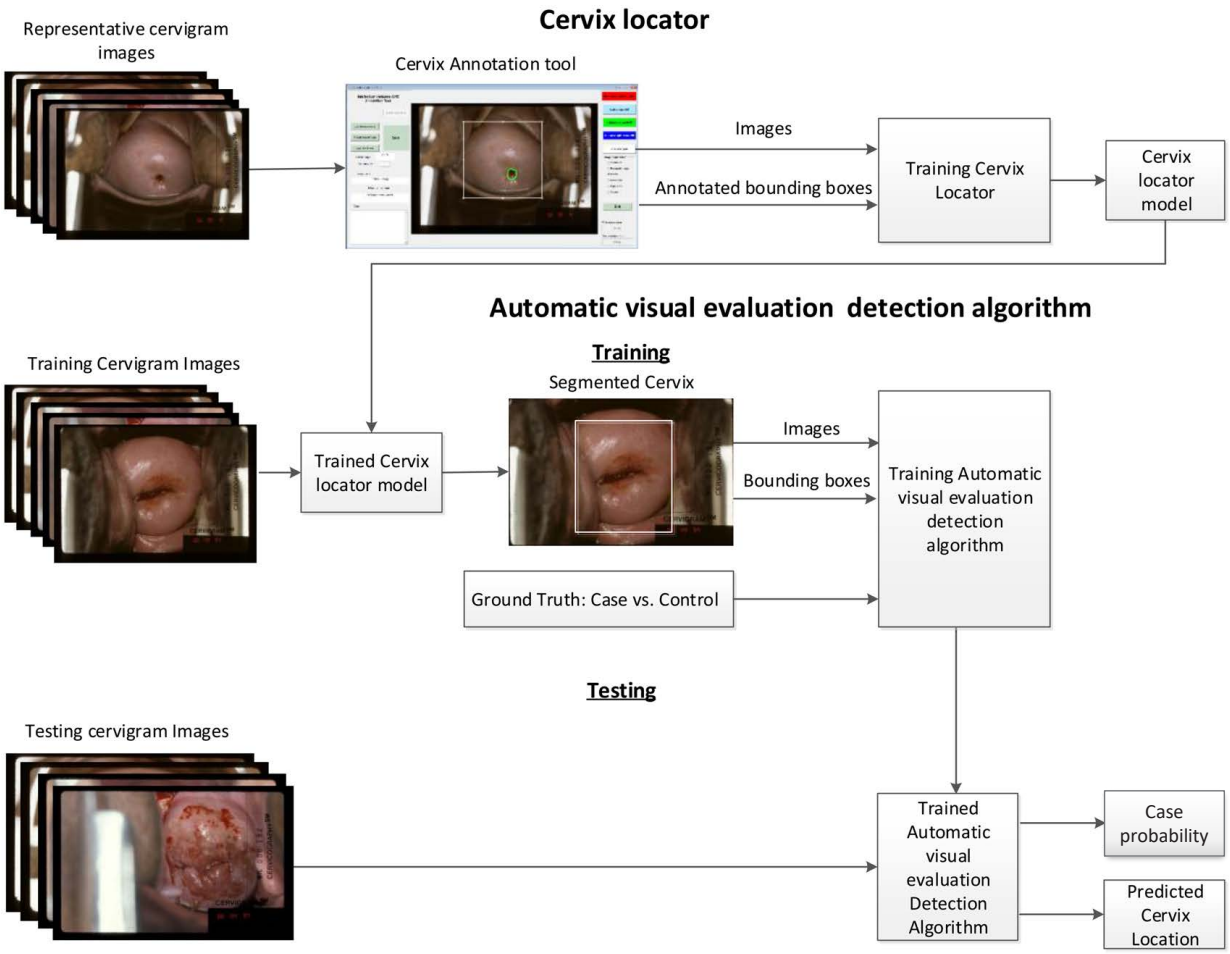
(i) Technical description

Training a model to discriminate between one image and another is now feasible, thanks to improved technology. Also, computing power has increased exponentially, and large, appropriately labelled image banks are available. Currently, for the detection of squamous cervical precancer, the clinically important discriminatory threshold is between normal or LSIL and HSIL. Therefore, algorithms in cervical precancer detection have focused on this dichotomous division. Training a CNN to discriminate between two distinct epithelial appearances within the squamous epithelium of the TZ involves exposing the model to a large series of adequate cervical images of known severity (i.e. supported by histopathology). Moreover, specific features on the cervical image may also be labelled by experts for a model to process. The CNN may then categorize cervical images into one of the two categories (≤ LSIL or HSIL) by outputting the probability that a given image belongs to either category.

During training, the CNN receives as inputs images from the training data set and adjusts its parameters to minimize the error between its predictions and the ground truth (i.e. colposcopically or histologically verified disease status) of the training set. Thus, the CNN is fitted to the training data set, learning the relevant features from the training data set, which enables it to increase the number of correct predictions. This process is illustrated in Fig. 4.9 (Hu et al., 2019). While the model is being trained on the training data set, the discriminative performance of the model is evaluated in a validation set. The purpose of the validation set is to evaluate the performance of the model on data that it has not been fitted to during the training process. Models with different selected hyperparameters can be trained in this way until a model that performs optimally on the validation set is determined. This yields a final trained model that can then be evaluated on a test set of images to assess its generalizability to predict cervical disease.

In general, the larger the training set, the higher the accuracy of the model. A viable model is often only as good as the quality of the images on which it is trained and the labels, or the robustness of the disease end-points, associated with these images. In many medical applications, there is often an imbalance between the number of images in each category; for example, in most cervical precancer image banks there are more images of ≤ LSIL than of HSIL. This imbalance can affect the training and validation process for the development of the model. The

Fig. 4.9 System architecture of the automated visual evaluation algorithm used by Hu et al. (2019)



From Hu et al. (2019).

scarcity of accurately labelled medical data, or robust disease end-points, with which to train CNNs for certain medical problems is a challenge to computational analysis. Although large image repositories may be available in some cases, relevant labelling of these images or information about the methods used to determine disease may be unclear or limited, leading to risk of disease misclassification. In addition, the quality of the available images depends on the sophistication of the image-capture system used. However, several specialized techniques (e.g. augmentation, transfer learning) can be used to address these issues and improve the performance of the model.

(ii) Performance of method

This technology may be appropriate for both screening and triage of screen-positive women. Early work using deep learning in cervical imagery has been encouraging (Xu et al., 2017). A deep-learning-based object detection method (Ren et al., 2017) was used to develop a visual evaluation algorithm for the detection of cervical precancer. Digitized cervigrams were collected as part of a population-based longitudinal cohort

study in 9406 women in Costa Rica; 241 of the women had histopathological confirmation of precancer (CIN2/3), and 38 had cancer over 7 years of follow-up in 1993–2001 (Hu et al., 2019). Despite limitations in image quality and images without full visualization of the squamocolumnar junction, the algorithm showed high accuracy for the identification of cervical precancers (Fig. 4.10). Automated visual evaluation of cervigrams collected at enrolment identified the cumulative number of cases of precancer or cancer with greater accuracy (AUC, 0.91; 95% CI, 0.89–0.93) than interpretation of the same images by a colposcopist (cervicography; AUC, 0.69; 95% CI, 0.63–0.74; $P < 0.0001$) or conventional cytology (AUC, 0.71; 95% CI = 0.65–0.77; $P < 0.0001$).

AI or deep-learning algorithms may be developed in different ways. Because the discriminative model "reads" images, the image-capture technique is relevant. Using this approach, Xue et al. (2020) developed an algorithm to interpret images captured by the smartphone-based MobileODT system. Automated visual evaluation can classify images of the cervix taken using smartphone camera image-capture systems. Alternatives to this approach include the development of a dedicated high-quality image-capture device that can capture multiple images to mimic a thorough colposcopic evaluation. Such systems can incorporate all the necessary computational power within a single device that is independent of the cloud; this makes them useful in low-resource settings. Both approaches have yet to be evaluated in the field.

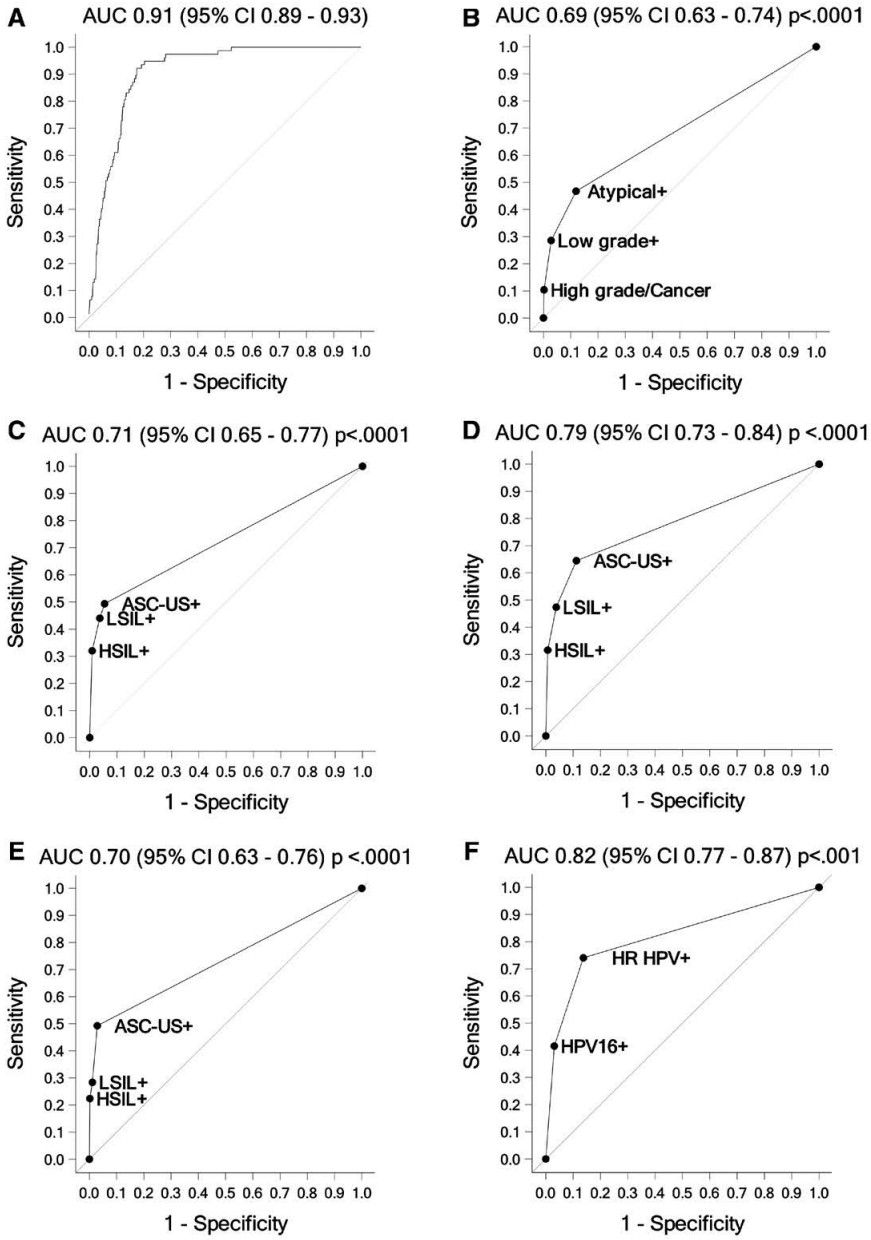
(b) Automated cytology technologies

Computer-assisted cytology systems have previously been developed for the reading of conventional or liquid-based cytology slides and are currently used in some settings. For the technical description and performance of these technologies, see Section 4.3.1(c). Recently, new AI-based approaches have been developed for automated evaluation of Pap cytology and dual-stain cytology.

A fully automated approach to evaluate Pap cytology was developed and validated in two studies in the USA. The training and validation data set included 1178 cervical cytology slides from HPV-positive women in Oklahoma who were referred for colposcopy for cytological abnormalities or for treatment of previously diagnosed precancer or cancer. The automated cytology algorithm achieved a performance for detection of CIN2+ (sensitivity, 0.91; specificity, 0.30) similar to that of conventional cytology with a threshold of ASC-US+ (sensitivity, 0.94; specificity, 0.30) (Schiffman et al., 2017). A subsequent study in 1839 HPV-positive women in the KPNC cohort, of whom 310 had precancer (181 with CIN2 and 129 with CIN3/AIS), similarly reported comparability of automated cytology and LBC with a threshold of ASC-US+ and LSIL+ (Yu et al., 2018).

Cytology with p16/Ki-67 dual staining (see Section 4.3.1(e)), which is used as a triage marker for HPV-positive women (see Section 4.4.7), can also be read by an automated system. A CNN deep-learning-based automated algorithm has been developed to evaluate p16/Ki-67 dual-stained slides (CYTOREADER software). The system uses a whole-slide scan followed by a machine-learning algorithm to detect and quantify p16/Ki-67 dual-stain-positive cells. A deep-learning classifier for automated dual-stained slides was compared with manual dual staining and conventional cytology for the detection of precancer in 602 women in Oklahoma who were referred for colposcopy, of whom 53 (8.8%) had CIN3+ (Wentzensen et al., 2021). The automated dual-staining algorithm had marginally lower positivity than manual dual staining (58% vs 63%; $P = 0.06$), with comparable sensitivity for the detection of CIN3+ (automated dual staining: 87%; 95% CI, 76–94%; manual dual staining: 87%; 95% CI, 76–94%; $P = 1.0$) and marginally higher specificity (automated

Fig. 4.10 ROC curve of automated visual evaluation of cervical images, and comparison of performance in identification of CIN2+



ROC-like curves are shown for the categorical variables for simple visual and statistical comparison with automated visual evaluation (two-sided χ^2 tests). The thresholds are listed on each curve, showing the sensitivity and 1 – specificity applicable to that threshold. Automated visual evaluation was as accurate as or more accurate than all of the screening tests used in the cohort study: (A) automated visual evaluation, (B) cervicography, (C) conventional cytology, (D) liquid-based cytology, (E) first-generation neural network-based cytology, and (F) MY09/MY11 PCR-based hrHPV testing.

ASC-US+. atypical squamous cells of undetermined significance or worse; AUC, area under the curve; CI, confidence interval; CIN2+, cervical intraepithelial neoplasia grade 2 or worse; HPV, human papillomavirus; hrHPV, high-risk human papillomavirus; HSIL+, high-grade squamous intraepithelial lesion or worse; LSIL+, low-grade squamous intraepithelial lesion or worse; PCR, polymerase chain reaction; ROC, receiver operating characteristic.

From Hu et al. (2019).

dual staining: 46%; 95% CI, 41–51%; manual dual staining: 41%; 95% CI, 36–46%; $P = 0.07$). Similarly, in 3095 HPV-positive women undergoing routine cervical cancer screening in the KPNC cohort, of whom 218 (7.0%) had CIN3+, the test positivity of the automated dual-staining algorithm was significantly lower than that of manual dual staining or conventional cytology with a threshold of ASC-US+ (42%, 50%, and 60%, respectively), with comparable sensitivity (88%, 90%, and 86%, respectively) and higher specificity (62%, 53%, and 42%, respectively). The automated dual-staining algorithm led to a substantial reduction in the colposcopy referral rate compared with conventional cytology, paired with better disease detection, and provided additional risk stratification compared with manual dual staining in HPV-positive women.

4.6.2 Emerging molecular technologies

HPV-based testing may soon replace cytology as the primary screening method for cervical cancer in many parts of the world. However, the lower specificity of HPV DNA-based tests means that some screen-positive women are referred for colposcopy unnecessarily. Novel methods are required to identify which HPV-positive women need to be referred for colposcopy (Cuschieri et al., 2018). Although infection with carcinogenic HPV is necessary for the development of cervical cancer, other molecular changes occur with carcinogenic HPV infection, which result from DNA nucleotide mutations, structural genomic variations, or epigenetic alterations, such as DNA methylation (Steenbergen et al., 2014). Aberrant DNA methylation may help to distinguish non-progressive HPV infections from those that will progress to cervical cancer. It may thus be used as a strategy to triage HPV-positive women.

(a) DNA methylation

(i) Technical description

DNA methylation occurs after the addition of a methyl group to position 5 of the cytosine (C) ring immediately preceding a guanine (G) in the DNA sequence. It occurs mainly at CpG dinucleotide sites (C and G separated by one phosphate), known as CpG islands, which are present in about 60% of human genes (Laird, 2010). Controlled DNA methylation is essential for normal biological processes, such as the regulation of cellular processes including embryonic development, chromosomal instability, and protection from invading foreign viral DNA. However, aberrant DNA methylation can lead to alterations in the functions of gene products that regulate tumour suppression, DNA repair, apoptosis, metastasis, and invasion (Steenbergen et al., 2014; Lorincz, 2016). DNA methylation of some human genes and of the genome of hrHPV genotypes has been shown to be associated with increasing persistence of hrHPV genotypes (Mirabello et al., 2012), precancer (Wentzensen et al., 2009; Bierkens et al., 2013), and invasive cervical cancer (Bowden et al., 2019; Cook et al., 2019; Kelly et al., 2019). DNA methylation of more than 100 human genes and up to 12 carcinogenic HPV genotypes has been evaluated as a possible biomarker for the detection of cervical precancer and cancer using clinician-collected or self-collected cervical samples (Wentzensen et al., 2009; Lorincz, 2016).

(ii) Host DNA methylation

The most widely studied human gene DNA methylation targets have been evaluated as triage tests in HPV-positive women in cross-sectional, case–control, or convenience studies. Most studies evaluated the DNA methylation of the human genes *CADM1*, *MAL*, and *miR-124-2* in different combinations, and of *PAX-1*, *SOX-1*, *POU4F3*, and *FAM19A4*, alone or in combination with *miR-124-2*, for the detection of CIN2+

or CIN3+. Several studies evaluated the DNA methylation of the human gene *EPB41L3*, alone or in combination with DNA methylation of HPV16 (late coding regions L1 and L2), HPV18 (L2), HPV31 (L1), and HPV33 (L2), which is defined as the S5 classifier. The sensitivity and specificity of DNA methylation assays for the detection of prevalent CIN2+ have been shown to vary widely depending on the human gene target, the CpG targets of the gene studied, variations in the thresholds used to define methylation positivity, and the study design (Lorincz, 2016; Kelly et al., 2019).

RCTs comparing detection of CIN2+ in women undergoing testing with DNA methylation compared with cytology, and prospective studies evaluating baseline DNA methylation status to predict the risk of cervical cancer over time have been informative in clarifying the value of DNA methylation as a triage test.

In a non-inferiority RCT (Protection by Offering HPV Testing on Self-Sampled Cervicovaginal Specimens Trial 3 [PROHTECT-3]) in the Netherlands, HPV-positive women registered in the national cervical cancer screening programme who submitted a self-collected sample were randomly allocated to either triage with cytology (509 women) or triage with DNA methylation analysis of the *MAL* and *miR-124-2* genes (515 women) (Verhoef et al., 2014). Detection of CIN2+ with triage by methylation was non-inferior to that by cytology (17% vs 15%; RR, 1.19; 95% CI, 0.90–1.57), and the sensitivity for detection of CIN2+ was equivalent (adjusted sensitivity, 71%; 95% CI, 66–75% for both DNA methylation and cytology), although the sensitivity for detection of CIN3+ was slightly lower with DNA methylation (68%; 95% CI, 63–72%) than with cytology (75%; 95% CI, 70–79%). Also, because of a lower specificity to distinguish < CIN2, referral for colposcopy was more common in the methylation group than in the cytology group (55% vs 29%; $P < 0.0001$) (Verhoef et al., 2014). In a 14-year longitudinal study in 1040 HPV-positive women enrolled in the POBASCAM screening trial in the Netherlands, all of whom underwent testing with DNA methylation and cytology, a negative *FAM19A4/miR-124-2* methylation test indicated lower risk of cervical cancer incidence over a 14-year follow-up period compared with a negative cytology result (< ASC-US) at enrolment (risk ratio, 0.71; 95% CI, 0.16–1.40) (De Strooper et al., 2018).

Previous studies have shown high agreement between clinician-collected and self-collected samples and between lavage-based and brush-based self-collected samples for several human gene DNA methylation targets (Boers et al., 2014; De Strooper et al., 2016); this offers the possibility of conducting screening and triage on the same self-collected specimen.

(iii) Viral DNA methylation

DNA methylation of the early (E2) and late (L1 and L2) coding regions of the HPV viral genome has been reported to increase with increasing CIN grade for 12 carcinogenic HPV types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59 (Clarke et al., 2012; Wentzensen et al., 2012; Lorincz et al., 2013; Mirabello et al., 2013; Bowden et al., 2019). The diagnostic accuracy of DNA methylation of HPV genotypes, alone or in various combinations, has been evaluated for detection of CIN2+. In a meta-analysis of seven studies evaluating DNA methylation of the E2, L1, and/or L2 coding regions of HPV16 in HPV16-positive women, the pooled sensitivity for detection of CIN2+ was 74% (95% CI, 57–85%) and the pooled specificity was 73% (95% CI, 66–79%), although there was significant heterogeneity in the observed estimates, because of differences in the CpG sites targeted (Kelly et al., 2019). A second, independent meta-analysis on the diagnostic accuracy of the HPV16 *L1* and/or *L2* genes in 10 studies reported similar findings, with a pooled sensitivity of 77% (95% CI, 63–87%) and a pooled specificity of 64% (95% CI, 55–71%) (Bowden et al., 2019).

The addition of HPV type-specific methylation (HPV types 16, 18, 31, and 33) to a human gene target (*EPB41L3*) as part of the S5 classifier enables testing in all women, irrespective of HPV type positivity. In three studies conducted in HPV-positive women in Canada, Colombia, and the United Kingdom, the sensitivity of the S5 classifier varied from 74% to 82% for detection of CIN2+ and from 84% to 93% for detection of CIN3+, suggesting that the combination of viral and host gene targets may increase detection of CIN2+/CIN3+ (Lorincz et al., 2016; Cook et al., 2019; Ramírez et al., 2021). However, the specificity for < CIN2 varied from 35% to 65%. Compared with either cytology with a threshold of ASC-US+ or HPV16/18 partial genotyping, the S5 classifier had a consistently higher sensitivity for the detection of CIN2+ or CIN3+ but a lower specificity (Lorincz et al., 2016; Cook et al., 2019; Ramírez et al., 2021).

A multiplex DNA methylation test targeting the L1/L2 regions of a wider range of HPV types (HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) was evaluated in a case–control study in 299 women with precancer (CIN3/AIS) and 360 women who had normal cytology but who were positive for any one of the targeted HPV types (i.e. 30 controls for each of the 12 carcinogenic HPV types evaluated) (Clarke et al., 2018). Methylation was positively associated with CIN3/AIS for all 12 types. The diagnostic accuracy of the 12-type DNA methylation assay was simulated by applying type-specific sensitivity and specificity estimates for the DNA methylation test to a population of 30 000 women using data from a cohort of women undergoing routine cervical screening in the USA. The simulated sensitivity and specificity of the 12-type DNA methylation assay were 80% and 66%, respectively; both were higher than for cytology with a threshold of ASC-US+ (77% and 54%, respectively).

(b) Detection of HPV E6 oncoprotein

Elevated expression of the HPV oncoproteins E6 and E7 is associated with the development of HPV-associated cervical cancer. E6 oncoprotein from HPV16/18/45 can be detected by the Onco*E6* test (Wentzensen et al., 2016). Zhao et al. (2013) reported the test performance when E6 oncoprotein was used as a primary screening method. Another study in China assessed the test performance of E6 oncoprotein for the detection of CIN3+ as triage for HPV-positive women (Qiao et al., 2014). The sensitivity of E6 oncoprotein from HPV16/18/45 was about 50% and the specificity was more than 90% in both clinician-collected and self-collected samples. Compared with HPV16/18/45 DNA testing, the sensitivity was lower but the specificity was higher.

A recent study reported the cumulative incidence of CIN3+ in 1742 women at 10-year follow-up (Dong et al., 2020). The cumulative incidence of CIN3+ was higher in women harbouring methylation at six sites (CpG 5602, 6650, 7034, 7461, 31, and 37) with and without E6 oncoprotein than in women with abnormal cytology. For triage of HPV16-positive women with detection of CIN3+, the sensitivity of E6 oncoprotein was lower than that of cytology (57.1% vs 92.9%), but the specificity was higher (86.5% vs 43.2%). A higher AUC was obtained with the methylation test at the six sites (0.82; 95% CI, 0.69–0.91) than with E6 oncoprotein detection (0.72; 95% CI, 0.58–0.82) and with cytology (0.68; 95% CI, 0.54–0.80).

References

Arbyn M, Ronco G, Cuzick J, Wentzensen N, Castle PE (2009). How to evaluate emerging technologies in cervical cancer screening? *Int J Cancer*. 125(11):2489–96. doi:10.1002/ijc.24774 PMID:19626591

Bierkens M, Hesselink AT, Meijer CJLM, Heideman DAM, Wisman GBA, van der Zee AGJ, et al. (2013). *CADM1* and *MAL* promoter methylation levels in

hrHPV-positive cervical scrapes increase proportional to degree and duration of underlying cervical disease. *Int J Cancer*. 133(6):1293–9. doi:10.1002/ijc.28138 PMID:23456988

Boers A, Bosgraaf RP, van Leeuwen RW, Schuuring E, Heideman DAM, Massuger LFAG, et al. (2014). DNA methylation analysis in self-sampled brush material as a triage test in hrHPV-positive women. *Br J Cancer*. 111(6):1095–101. doi:10.1038/bjc.2014.392 PMID:25032730

Bowden SJ, Kalliala I, Veroniki AA, Arbyn M, Mitra A, Lathouras K, et al. (2019). The use of human papillomavirus DNA methylation in cervical intraepithelial neoplasia: a systematic review and meta-analysis. *EBioMedicine*. 50:246–59. doi:10.1016/j.ebiom.2019.10.053 PMID:31732479

Clarke MA, Gradissimo A, Schiffman M, Lam J, Sollecito CC, Fetterman B, et al. (2018). Human papillomavirus DNA methylation as a biomarker for cervical precancer: consistency across 12 genotypes and potential impact on management of HPV-positive women. *Clin Cancer Res*. 24(9):2194–202. doi:10.1158/1078-0432.CCR-17-3251 PMID:29420222

Clarke MA, Wentzensen N, Mirabello L, Ghosh A, Wacholder S, Harari A, et al. (2012). Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiol Biomarkers Prev*. 21(12):2125–37. doi:10.1158/1055-9965.EPI-12-0905 PMID:23035178

Cook DA, Krajden M, Brentnall AR, Gondara L, Chan T, Law JH, et al. (2019). Evaluation of a validated methylation triage signature for human papillomavirus positive women in the HPV FOCAL cervical cancer screening trial. *Int J Cancer*. 144(10):2587–95. doi:10.1002/ijc.31976 PMID:30412281

Cuschieri K, Ronco G, Lorincz A, Smith L, Ogilvie G, Mirabello L, et al. (2018). Eurogin roadmap 2017: triage strategies for the management of HPV-positive women in cervical screening programs. *Int J Cancer*. 143(4):735–45. doi:10.1002/ijc.31261 PMID:29341110

De Strooper LMA, Berkhof J, Steenbergen RDM, Lissenberg-Witte BI, Snijders PJF, Meijer CJLM, et al. (2018). Cervical cancer risk in HPV-positive women after a negative *FAM19A4/mir124-2* methylation test: a post hoc analysis in the POBASCAM trial with 14 year follow-up. *Int J Cancer*. 143(6):1541–8. doi:10.1002/ijc.31539 PMID:29663363

De Strooper LMA, Verhoef VMJ, Berkhof J, Hesselink AT, de Bruin HME, van Kemenade FJ, et al. (2016). Validation of the *FAM19A4/mir124-2* DNA methylation test for both lavage- and brush-based self-samples to detect cervical (pre)cancer in HPV-positive women. *Gynecol Oncol*. 141(2):341–7. doi:10.1016/j.ygyno.2016.02.012 PMID:26921784

Dong L, Zhang L, Hu SY, Feng RM, Zhao XL, Zhang Q, et al. (2020). Risk stratification of HPV 16 DNA methylation combined with E6 oncoprotein in cervical cancer screening: a 10-year prospective cohort study. *Clin Epigenetics*. 12(1):62. doi:10.1186/s13148-020-00853-1 PMID:32381054

Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 542(7639):115–8. doi:10.1038/nature21056 PMID:28117445

Hu L, Bell D, Antani S, Xue Z, Yu K, Horning MP, et al. (2019). An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J Natl Cancer Inst*. 111(9):923–32. doi:10.1093/jnci/djy225 PMID:30629194

Hussein S, Cao K, Song Q, Bagci U (2017). Risk stratification of lung nodules using 3D CNN-based multi-task learning. In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap PT, et al., editors. Information processing in medical imaging. IPMI 2017. (Lecture Notes in Computer Science, Volume 10265). Cham, Switzerland: Springer; pp. 249–60. doi:10.1007/978-3-319-59050-9_20

Islam MT, Aowal MA, Tahseen Minhaz A, Ashraf K (2017). Abnormality detection and localization in chest X-rays using deep convolutional neural networks. Ithaca (NY), USA: Cornell University. Available from: https://arxiv.org/pdf/1705.09850.pdf.

Kelly H, Benavente Y, Pavon MA, De Sanjose S, Mayaud P, Lorincz AT (2019). Performance of DNA methylation assays for detection of high-grade cervical intraepithelial neoplasia (CIN2+): a systematic review and meta-analysis. *Br J Cancer*. 121(11):954–65. doi:10.1038/s41416-019-0593-4 PMID:31616037

Laird PW (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet*. 11(3):191–203. doi:10.1038/nrg2732 PMID:20125086

Lorincz AT (2016). Virtues and weaknesses of DNA methylation as a test for cervical cancer prevention. *Acta Cytol*. 60(6):501–12. doi:10.1159/000450595 PMID:27806357

Lorincz AT, Brentnall AR, Scibior-Bentkowska D, Reuter C, Banwait R, Cadman L, et al. (2016). Validation of a DNA methylation HPV triage classifier in a screening sample. *Int J Cancer*. 138(11):2745–51. doi:10.1002/ijc.30008 PMID:26790008

Lorincz AT, Brentnall AR, Vasiljević N, Scibior-Bentkowska D, Castanon A, Fiander A, et al. (2013). HPV16 L1 and L2 DNA methylation predicts high-grade cervical intraepithelial neoplasia in women with mildly abnormal cervical cytology. *Int J Cancer*. 133(3):637–44. doi:10.1002/ijc.28050 PMID:23335178

McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. (2020). International evaluation of an AI system for breast cancer screening.

Nature. 577(7788):89–94. doi:10.1038/s41586-019-1799-6 PMID:31894144

Mirabello L, Schiffman M, Ghosh A, Rodriguez AC, Vasiljevic N, Wentzensen N, et al. (2013). Elevated methylation of HPV16 DNA is associated with the development of high grade cervical intraepithelial neoplasia. *Int J Cancer*. 132(6):1412–22. doi:10.1002/ijc.27750 PMID:22847263

Mirabello L, Sun C, Ghosh A, Rodriguez AC, Schiffman M, Wentzensen N, et al. (2012). Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. *J Natl Cancer Inst*. 104(7):556–65. doi:10.1093/jnci/djs135 PMID:22448030

Qiao YL, Jeronimo J, Zhao FH, Schweizer J, Chen W, Valdez M, et al. (2014). Lower cost strategies for triage of human papillomavirus DNA-positive women. *Int J Cancer*. 134(12):2891–901. doi:10.1002/ijc.28616 PMID:24248915

Ramírez AT, Sánchez GI, Nedjai B, Agudelo MC, Brentnall AR, Cuschieri K, et al.; ASC-US-COL Trial Group (2021). Effective methylation triage of HPV positive women with abnormal cytology in a middle-income country. *Int J Cancer*. 148(6):1383–93. doi:10.1002/ijc.33314 PMID:33006394

Ren S, He K, Girshick R, Sun J (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 39(6):1137–49. doi:10.1109/TPAMI.2016.2577031 PMID:27295650

Schiffman M, Yu K, Zuna R, Terence Dunn S, Zhang H, Walker J, et al. (2017). Proof-of-principle study of a novel cervical screening and triage strategy: computer-analyzed cytology to decide which HPV-positive women are likely to have ≥CIN2. *Int J Cancer*. 140(3):718–25. doi:10.1002/ijc.30456 PMID:27696414

Steenbergen RDM, Snijders PJF, Heideman DAM, Meijer CJLM (2014). Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. *Nat Rev Cancer*. 14(6):395–405. doi:10.1038/nrc3728 PMID:24854082

Ting DSW, Cheung CYL, Lim G, Tan GSW, Quang ND, Gan A, et al. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 318(22):2211–23. doi:10.1001/jama.2017.18152 PMID:29234807

Verhoef VMJ, Bosgraaf RP, van Kemenade FJ, Rozendaal L, Heideman DAM, Hesselink AT, et al. (2014). Triage by methylation-marker testing versus cytology in women who test HPV-positive on self-collected cervicovaginal specimens (PROHTECT-3): a randomised controlled non-inferiority trial. *Lancet Oncol*. 15(3):315–22. doi:10.1016/S1470-2045(14)70019-1 PMID:24529697

Wei JW, Suriawinata AA, Vaickus LJ, Ren B, Liu X, Lisovsky M, et al. (2020). Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides. *JAMA Netw Open*. 3(4):e203398. doi:10.1001/jamanetworkopen.2020.3398 PMID:32324237

Wentzensen N, Lahrmann B, Clarke MA, Kinney W, Tokugawa D, Poitras N, et al. (2021). Accuracy and efficiency of deep-learning-based automation of dual stain cytology in cervical cancer screening. *J Natl Cancer Inst*. 113(1):72–9. doi:10.1093/jnci/djaa066 PMID:32584382

Wentzensen N, Schiffman M, Palmer T, Arbyn M (2016). Triage of HPV positive women in cervical cancer screening. *J Clin Virol*. 76(Suppl 1):S49–55. doi:10.1016/j.jcv.2015.11.015 PMID:26643050

Wentzensen N, Sherman ME, Schiffman M, Wang SS (2009). Utility of methylation markers in cervical cancer early detection: appraisal of the state-of-the-science. *Gynecol Oncol*. 112(2):293–9. doi:10.1016/j.ygyno.2008.10.012 PMID:19054549

Wentzensen N, Silver MI (2016). Biomarkers for cervical cancer prevention programs: the long and winding road from discovery to clinical use. *J Low Genit Tract Dis*. 20(3):191–4. doi:10.1097/LGT.0000000000000231 PMID:27243141

Wentzensen N, Sun C, Ghosh A, Kinney W, Mirabello L, Wacholder S, et al. (2012). Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *J Natl Cancer Inst*. 104(22):1738–49. doi:10.1093/jnci/djs425 PMID:23093560

Wentzensen N, Wacholder S (2013). From differences in means between cases and controls to risk stratification: a business plan for biomarker development. *Cancer Discov*. 3(2):148–57. doi:10.1158/2159-8290.CD-12-0196 PMID:23299199

Xu T, Zhang H, Xin C, Kim E, Long LR, Xue Z, et al. (2017). Multi-feature based benchmark for cervical dysplasia classification evaluation. *Pattern Recognit*. 63:468–75. doi:10.1016/j.patcog.2016.09.027 PMID:28603299

Xue Z, Novetsky AP, Einstein MH, Marcus JZ, Befano B, Guo P, et al. (2020). A demonstration of automated visual evaluation of cervical images taken with a smartphone camera. *Int J Cancer*. 147(9):2416–23. doi:10.1002/ijc.33029 PMID:32356305

Yu K, Hyun N, Fetterman B, Lorey T, Raine-Bennett TR, Zhang H, et al. (2018). Automated cervical screening and triage, based on HPV testing and computer-interpreted cytology. *J Natl Cancer Inst*. 110(11):1222–8. doi:10.1093/jnci/djy044 PMID:29659930

Zhao FH, Jeronimo J, Qiao YL, Schweizer J, Chen W, Valdez M, et al. (2013). An evaluation of novel, lower-cost molecular screening tests for human papillomavirus in rural China. *Cancer Prev Res (Phila)*. 6(9):938–48. doi:10.1158/1940-6207.CAPR-13-0091 PMID:23878179