A Survey on Deep Learning-based Cervical Cytology Screening: from Cell Identification to Whole Slide Image Analysis

Peng Jiang
Institute of Artificial Intelligence, School of Computer Science, Wuhan University
Xuekong Li
Wei County Maternal and Child Health Hospital
Hui Shen
Wei County Maternal and Child Health Hospital
Yuqi Chen
Institute of Artificial Intelligence, School of Computer Science, Wuhan University
Lang Wang
Institute of Artificial Intelligence, School of Computer Science, Wuhan University
Hua Chen
Institute of Artificial Intelligence, School of Computer Science, Wuhan University
Jing Feng
Institute of Artificial Intelligence, School of Computer Science, Wuhan University
Juan Liu (liujuan@whu.edu.cn)
Institute of Artificial Intelligence, School of Computer Science, Wuhan University

Research Article

Keywords: Cervical cytology, Deep learning, cancer screening, Artificial intelligence, Cytopathology

Posted Date: March 15th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2680912/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Artificial Intelligence Review on October 5th, 2023. See the published version at https://doi.org/10.1007/s10462-023-10588-z.

A Survey on Deep Learning-based Cervical Cytology Screening: from Cell Identification to Whole Slide Image Analysis

Peng Jiang[1], Xuekong Li[2], Hui Shen[2], Yuqi Chen[1], Lang Wang[1], Hua Chen[1], Jing Feng[1] and Juan Liu[1*]

[1]Institute of Artificial Intelligence, School of Computer Science, Wuhan University, Wuhan, 430072, China.
[2]Wei County Maternal and Child Health Hospital, Handan, 056000, China.

*Corresponding author(s). E-mail(s): liujuan@whu.edu.cn;

Abstract

Cervical cancer is one of the most common cancers in daily life. Early detection and diagnosis can effectively help facilitate subsequent clinical treatment and management. With the growing advancement of artificial intelligence (AI) and deep learning (DL) techniques, an increasing number of computer-aided diagnosis (CAD) methods based on deep learning have been applied in cervical cytology screening. In this paper, we survey more than 70 publications since 2016 to provide a systematic and comprehensive review of DL-based cervical cytology screening. First, we provide a concise summary of the medical and biological knowledge pertaining to cervical cytology, since we hold a firm belief that a comprehensive biomedical understanding can significantly contribute to the development of CAD systems. Then, we collect a wide range of public cervical cytology datasets. Besides, image analysis approaches and applications including cervical cell identification, abnormal cell or area detection, cell region segmentation and cervical whole slide image diagnosis are summarized. Finally, we discuss the present obstacles and promising directions for future researches in automated cervical cytology screening.

Keywords: Cervical cytology, Deep learning, cancer screening, Artificial intelligence, Cytopathology

1 Introduction

Cervical cancer is a common malignancy that poses a serious threat to women's health. It is the fourth most common cancer in terms of both incidence and mortality. In 2020, approximately 600,000 new cases of cervical cancer were diagnosed and more than 340,000 people died from this disease globally [1]. The incidence and mortality of cervical cancer may vary among countries and regions, which is related to the level of health services, the implementation of screening and prevention measures, lifestyle, and environmental factors in that region [2]. Fortunately, cervical cancer has a long precancerous stage, and annual screening programs can help detect and treat it in a timely manner. If cervical cancer is detected early, it can be completely eradicated.

Early-stage cervical cancer may not have obvious symptoms, but as the disease progresses, symptoms such as abnormal vaginal bleeding, vaginal discharge, and pelvic pain may appear. Early diagnosis is crucial for the treatment and prognosis of cervical cancer [3]. At present, the most widely used and effective screening scheme around the world is cervical cytology screening. However, the traditional cervical cytology screening program requires the manual identification of abnormal cells under a microscope, which is time-consuming, tedious, and error-prone [4]. In this context, an increasing number of automatic screening systems have been proposed to reduce the burden on cytopathologists and improve diagnosis efficiency [5–7]. With the advancement of artificial intelligence (AI) and digital image processing, machine learning (ML) technology has been widely applied in cervical cytology screening to analyze cytological images due to its high-performance results [8–10]. Nevertheless, traditional machine learning approaches have complex image preprocessing and feature selection steps that limit the further progress of human-machine collaboration.

In the past few years, deep learning (DL), a branch of machine learning, has exploded in the field of computer vision [11–15]. The end-to-end automatic feature extraction and learning process of DL eliminates the need for manual feature design and selection. DL has made a breakthrough in various fields of image processing, medical image analysis is no exception [16–18]. DL solutions have been successfully applied in many medical imaging tasks, such as thoracic Imaging, neuroimaging, cardiovascular imaging, abdominal imaging, and microscopy imaging[19]. The development of DL has also greatly accelerated automatic image analysis in cervical cytology screening. To understand the popularity and development trend of deep learning in cervical cytology, multiple literature databases (Google Scholar, dblp, PubMed, and Web of Science) are searched using the keywords related to cervical cytology screening (cervical cytology, cervical cancer diagnosis, deep learning, Pap smear, etc.). Fig. 1 illustrates the number of related publications from 2016 to 2022. Since 2016, there has been a notable surge in the use of DL for cervical cytology screening. Moreover, the object detection task has experienced significant growth since 2019, while the task of whole slide image (WSI) analysis has emerged in 2021 and shown impressive expansion recently.

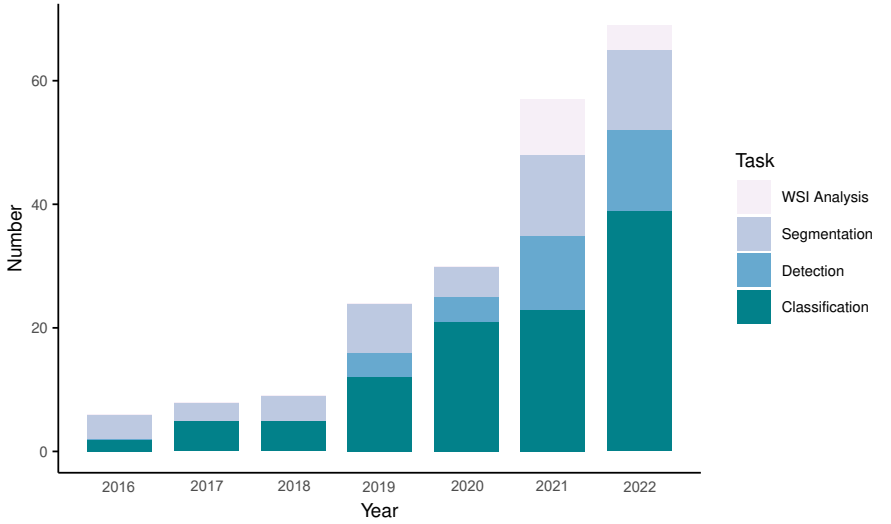


Fig. 1 Number of publications in DL-based classification, detection, segmentation and WSI analysis for automated cervical cytology.

There exist several surveys in the field of automated cervical cytology [10, 20–24]. Although these reviews provide valuable insights into automated cervical cytology, they are not exhaustive and some areas remain unexplored, calling for a further comprehensive investigation. First of all, the above reviews focus on classification and segmentation tasks at the cell level, and none of them investigate the application of object detection algorithms in automated cervical cytology screening. Secondly, the majority of these reviews primarily concentrate on conventional machine learning approaches, with comparatively limited coverage of DL-based methods. Moreover, few reviews provide biomedical context pertaining to cervical cytology, which is relevant for understanding the applicability of DL-based methods in this field. Last but not the least, there is currently no review specialized for automatic WSI analysis of cervical cytology as the related works have only recently started to emerge in 2021. Automatic WSI analysis of cervical cytology holds great promise for improving the efficiency and accuracy of cervical cancer screening. Staying abreast of the latest developments and advancements in this field will be important for researchers and practitioners.

To address the above issues, a comprehensive overview of relevant works for automated cervical cytology is presented in this survey including over 70 publications since 2016. For researchers just entering this field, this survey provides background knowledge on cervical cytology such as a brief introduction to cervical cancer, popular cervical cytology screening procedures, and definite cell categories in the Bethesda system (TBS). It is worth noting that the comparison of different reporting terms is elaborated as well. This can often cause confusion and impact the construction of a correct and reasonable DL

model. Besides, this survey has also compiled the most extensive collection of publicly available cervical cytology image datasets. Moreover, this survey summarizes the latest DL-based classification, detection, segmentation, and WSI analysis methods in automated cervical cytology screening. Towards the end of this paper, several challenges and opportunities (stain normalization, image super-resolution, incorporating medical domain knowledge, annotation-efficient learning, internet of medical things, etc.) are presented that may provide promising research directions in cervical cytology screening.

The paper is organized as follows: Section 1 introduces the background and objective of this survey. In Section 2, an overview of cervical cytology with detailed biomedical knowledge is provided. Section 3 lists the public datasets in cervical cytology screening and summarizes the detailed progress in the DL-based automated cervical cytology from cell identification to WSI analysis. In Section 4, existing challenges and potential opportunities in automated cervical cytology screening are discussed. Finally, Section 5 concludes this review paper.

2 Overview of cervical cytology

Before the review of deep learning-based method for cervical cytology screening, an preliminary overview of cervical cytology is presented in this section. We believe that medical and biological domain knowledge has a critical impact on the construction of computational models and the design of Computer-aided diagnosis (CAD) systems. We first introduce the basic knowledge of cervical cancer in Section 2.1. Then, a detailed procedure of cervical cytology screening is described in Section 2.2. After that, we introduce the history of reporting terminology for cervical cytology and explain the corresponding relations and difference between four reporting systems in Section 2.3. Next, we elaborate the cell categories in TBS in Section 2.4.

2.1 Introduction of cervical cancer

Cervical cancer is a kind of malignant tumor arising from the cervix and threatens the life and health of women. There are two main types of cervical cancer: (1) squamous cell carcinoma (SCC); and (2) adenocarcinoma. About 90% of cervical cancer cases are SCC, most of which begin in the transformation zone and develop from cells in the outer part of the cervix [25]. Cervical cancer is by far the most common HPV-related disease, and almost all cervical cancers (more than 95%) are caused by persistent infection with some types of HPV. There are at least 13 known types of HPV that can persist and progress to cancer, called high-risk HPV, the most common being HPV 16 and 18 strains. Cervical cancer has a long period of the precancerous stage, and its development is continuous, as shown in Figure 2. The main characteristics of precancerous cells focus on changes in the nucleus. For example, nuclear enlargement results in an increased nuclear to cytoplasmic ratio. It's common to see Binucleation and multinucleation. Besides, nucleoli are generally absent or inconspicuous if present. And the contour of the nuclear membrane is quite

irregular. Early forms of cervical cancer may have no symptoms or signs, but there is compelling evidence that cervical cancer is one of the most preventable and treatable cancers if detected early and managed effectively through regular screening programs. There are currently three World Health Organization (WHO) recommended screening tests for cervical cancer: (1) HPV testing for high-risk HPV types; (2) cervical cytology screening; and (3) visual inspection with acetic acid (VIA). Cervical cytology screening has been the basic method worldwide since the cytological features are significant indications of cervical cancer.

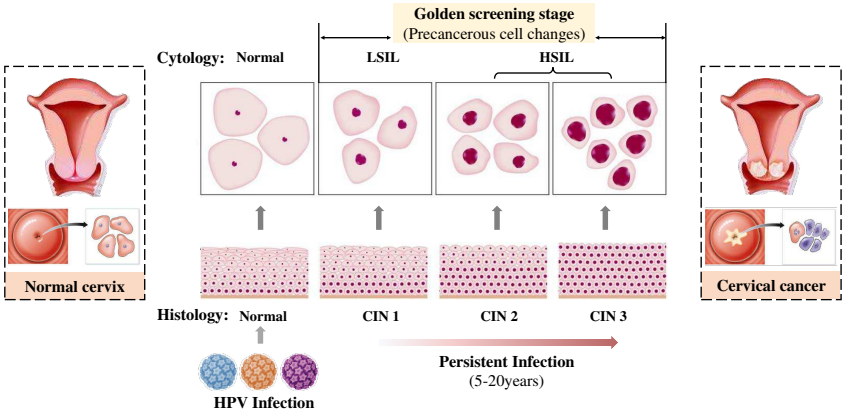


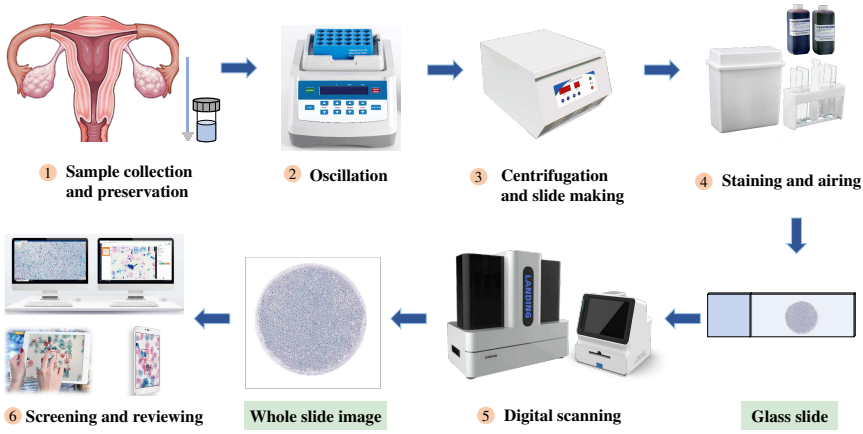
Fig. 2 The natural evolution of HPV infected cervical cancer.

2.2 Procedure of cervical cytology screening

Cervical cytology screening is the most effective and widely used screening program for discovering cancerous or precancerous lesions. The primary goal of screening is to identify abnormal cervical cells with severe cell changes so that they can be monitored or treated in time to prevent the development of invasive cancer [26]. A large number of medical organizations suggest conducting routine cervical cytology screening every few years. Currently, conventional Papanicolaou smear (CPS) test and liquid-based cytology (LBC) are performed for cervical cytology screening worldwide [27]. CPS is a procedure in which cervical cells are scraped and observed under a microscope. Fig. 3(a) illustrates the whole process of CPS. Under the guidance of a vaginal speculum, a soft brush will insert into the vagina to collect cells from the cervix. Then a pap smear can be acquired by evenly spreading the cells from the brush onto the glass slide. After staining, cytologists can observe the sample under a microscope and make a diagnosis.

Due to the influence of blood, mucus, inflammation, and other factors, CPS often acquires blurred samples, resulting in poor imaging results and detection

(a) Conventional Papanicolaou smear test



(b) Liquid-based cytology using digital slides

Fig. 3 Two prevalent procedures of cervical cytology screening

errors. In recent years, with the development of digital pathology scanners and the improvement of sample preparation level, LBC can significantly improve the imaging quality of cervical cell samples, and thus has gradually become the mainstream implement for cervical cytology screening. As shown in Fig. 3(b), the collected cells will be placed in a preservation solution for further process. After oscillation and centrifugation, a liquid-based glass slide can be obtained by natural sedimentation. Then, the liquid-based sample preparation is completed via staining and air drying. To facilitate retrospective examination, a digital slide is usually generated via pathological scanners. Digital pathology brings a positive and profound impact on traditional pathological diagnosis, which digitizes glass slides into whole slide images (WSIs) to greatly reduce the workload of pathologists and improve the diagnosis efficiency compared to microscope-based visual observation [28, 29]. Liquid-based preparation together with digital slides is a satisfactory alternative to conventional smear and has a great application prospect for nowadays' large-scale cervical cancer screening programs.

2.3 History of reporting terminology

The establishment of a standard cervical cytology report system plays a vital role in the universality of diagnosis methods and the acceptance of diagnosis results. In practice, the standard report system can cross the gap between different regions and different countries, strengthen the exchange of relevant scientific research results, and greatly improve the efficiency of cervical cancer diagnosis [30]. The earliest report system for cervical cytological diagnosis was the Papanicolaou classification system, which developed a numeric classification terminology to grade cervical cells for 5 levels [31]. Class I to Class V respectively indicated the absence of abnormal or atypical cells; atypical cells, but no evidence of malignancy; cytology suggestive of but not conclusive for malignancy; cytology strongly suggestive of malignancy; and cytology conclusive for malignancy. However, many pointed out that the Papanicolaou classification system was strongly subjective and there was no strict objective standard for the difference between Class II, III, and IV. In addition, the Papanicolaou classification system did not have a clear definition of precancerous lesions and was not able to correspond to histopathological diagnosis terms.

With the development and refinement of both cytological and histological diagnoses of cervical cancer, an understanding of the natural history of cervical intraepithelial neoplasms (CIN) has developed progressively. The term dysplasia was introduced to refer to precancerous abnormalities of squamous cells and the 3-tiers dysplasia system (mild/ moderate/severe dysplasia, or carcinoma in situ) was proposed [32]. Recognizing the difficulty in differentiating severe dysplasia and carcinoma in situ (CIS), in 1966 [33], the CIN classification system was developed to describe CIN as a continuum of neoplastic change with progressively increasing risk of invasion, which was subdivided into grades I, II, and III. The advantage of both the 3-tiers dysplasia system and CIN classification was the ability to use it for cytological as well as histological samples.

In the 1970s and 1980s, as HPV testing became more available, vast epidemiological and biochemical evidence manifested the link between HPV and cervical dysplasia, which supported the role of high-risk HPV as a necessary factor in the development of cervical cancer [34, 35]. As a result, the first edition of the Bethesda system (TBS) for reporting cervical cytology was promulgated in 1988 [36]. TBS aims to provide a uniform interpretation of cervical cytology, thereby facilitating communication between the clinician and the laboratory. With the change in practice to increased utilization of new technologies and findings in the last few decades, such as further insights into HPV biology and the development of liquid-based preparations, TBS has been updated three times to meet the evolving cervical cytology. The newest TBS 2014 guideline [37] offers comprehensive terminology for the reporting of cervical cytology. Nowadays, dysplasia and CIN systems remain standard reporting terms for cervical histopathology. However, for reporting cervical cytology, TBS is the

Table 1 Different cervical pathology reporting systems

Histopathology		Cytopathology	
3-tier dysplasia	CIN	The Pap classification	TBS
Normal	Normal	I (Normal)	NIML
Atypical cells	Squamous atypia	II (Atypical)	ASCUS ASCH AGC
mild dysplasia	CIN1	III (Suspicious)	LSIL
moderate dysplasia severe dysplasia carcinoma in situ	CIN2 CIN3	IV (Suggestive)	HSIL
Invasive cancer	Invasive cancer	V (Indicative)	SCC Adenocarcinoma

most commonly used and appropriate criterion. Referring to [38, 39], the specific classification criteria and corresponding relations of these four systems are shown in Table 1.

2.4 Cell categories in TBS

TBS lays the foundation for our further comprehending of HPV biology and provides the necessary framework for the development of systematic evidence-based guidelines for cervical cancer screening and management. Since TBS is the widely recognized standard for cervical cytology reporting, in this section, cell categories in the latest version of TBS (TBS 2014) [37] are introduced for better understanding of cervical cytology.

The specimen is reported as negative for intraepithelial lesion or malignancy (NIML) when there is no cellular evidence of neoplasia or epithelial abnormalities. Normal cellular elements include normal squamous cells and glandular cells. Squamous cells located in different positions of cervical epithelium have different characteristics, from shallow to deep can be divided into the superficial cell, intermediate cell, parabasal cell and basal cell. In a stained sample, the cytoplasm of superficial cells are pink or orange while the cytoplasm of all of the less mature cells are light green or cyan. Superficial cells and intermediate cells are large polygonal with very low nuclear-to-cytoplasmic ratio (N/C ratio) while parabasal cells and basal cells are generally round or oval with relatively high N/C ratio. Basal cells are small, and undifferentiated cells which are rarely seen in a Pap smear unless there is severe atrophy. Glandular cells consist of endocervical cells and endometrial cells. Viewed from above, sheets of endocervical cells have a honeycomb appearance, whereas when viewed from side line up like “picket-fence” palisades. Endocervical glandular cells exhibit polarity with nuclei at one end of the cytoplasm and mucus present at the other. Endometrial cells which are spontaneously shed are derived from epithelial or stromal and often in a 3-dimentional cluster referred to as an “exodus”

ball, which generally present at the end of menstrual flow. Figure 4 exhibits various normal cervical cells.

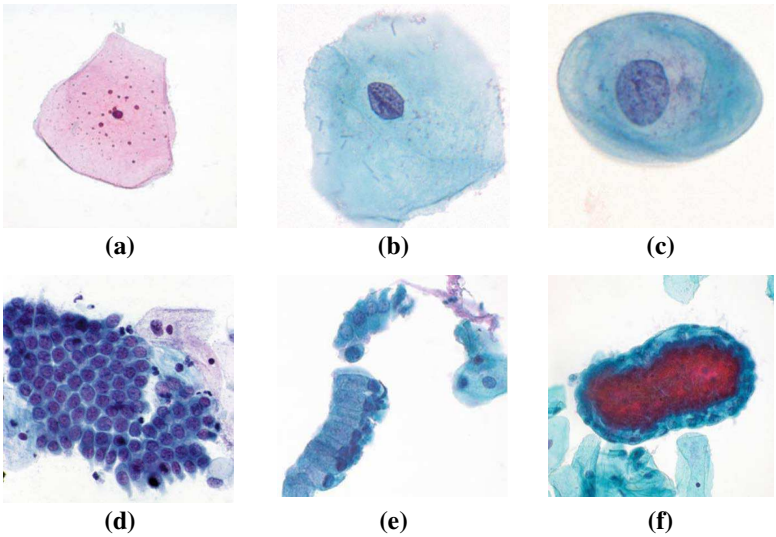


Fig. 4 Different normal cervical cells: (a) superficial cell, (b) intermediate cell, (c) parabasal cell, (d) basal cell, (e) endocervical cell, (f) endometrial cell.

Abnormal squamous cells or glandular cells can be discovered during cervical cytology screening which can be categorized as following types according to TBS reporting terminology:

- **Atypical squamous cells - undetermined significance (ASC-US):** This type refers to changes that are suggestive of the low-grade squamous intraepithelial lesion (LSIL). The nuclei of ASC-US are about 2.5 to 3 times the area of a normal intermediate squamous cell nucleus (approximately 35 mm2) and the N/C ratio is slightly increased.
- **Atypical squamous cells - cannot exclude a high-grade squamous intraepithelial lesion (ASC-H):** ASC-H primarily affects the squamous metaplastic cells and the nuclei are usually approximately 1.5-2.5 times larger than normal metaplastic cells' nuclei. The cytological changes of ASC-H are suggestive of the high-grade squamous intraepithelial lesion (HSIL) but are insufficient for a definitive diagnosis of HSIL.
- **Low-grade squamous intraepithelial lesion (LSIL):** To render an LSIL diagnosis, explicit abnormal changes must be found in the squamous cells. Cytological changes of LSIL usually occur in mature intermediate or superficial squamous cells and the nuclear enlargement are more than three times the area of normal intermediate nuclei. Additional characteristics of LSILs include hyperchromatic nuclei, absent or inconspicuous nucleoli, binucleation or multinucleation, and increased koilocytosis.

- **How-grade squamous intraepithelial lesion (HSIL):** In general, the cells affected by HSIL are immature parabasal or basal cells. HSIL cells can appear in sheets, singly, or in syncytial clusters which may result in hyperchromatic crowded groups (HCG). The nuclear enlargement and small size of HSIL cells lead to a marked increase in the N/C ratio. The nucleoli are generally absent and the contour of the nuclear membrane is quite irregular.
- **Squamous cell carcinoma (SCC):** SCC is defined as "an invasive epithelial tumor composed of squamous cells of varying degrees of differentiation" according to 2014 WHO terminology [40], which is the most common malignant tumor of cervical cancer. Cytological features of SCC usually include pleomorphic hyperchromatic nuclei, irregularly dispersed chromatin with nuclear clearing, prominent irregular often multiple nucleoli, keratinization of cells, and keratinous debris.
- **Atypical glandular cells (AGC):** AGC is a generic terminology for atypical endocervical cells or atypical endometrial cells when there is difficulty in locating the origin of the cells. Atypical endocervical cells may be further qualified as "NOS" or "favor neoplasia", while atypical endometrial cells don't need it. The cytological features of AGC may include nuclear enlargement, crowding, variation in size, hyperchromasia, chromatin heterogeneity, and evidence of proliferation.
- **Endocervical adenocarcinoma in situ (AIS):** AIS is considered to be the glandular counterpart of HSIL and the precursor to invasive endocervical adenocarcinoma. The criteria of AIS comprise of the following aspects: The cells present as sheets, pseudostratified strips or clusters – with loss of well-defined honeycomb patterns; The nuclei tend to be enlarged, variably sized, oval or elongated, and the loosely superficial cells of the cell groups incline to be tapered and spread out, referred to as "feathering"; Nucleoli are usually small or inconspicuous and may not be present; The quantity of cytoplasm is diminished and N/C ratio is increased; Nuclear hyperchromasia with evenly dispersed, coarsely granular chromatin; The chromatin pattern is coarsely granular with even distribution and mitoses are common.
- **Adenocarcinoma:** The Cytological criteria for adenocarcinoma may overlap those outlined for AIS. There are abundant abnormal cells, typically with columnar configuration. Nuclei tend to be enlarged, pleomorphic with nuclear membrane irregularities, and may be hypochromatic with irregularly distributed chromatin or chromatin clearing. Multinucleation and Macronucleoli are common features. Adenocarcinoma may coexist with squamous lesions.

Figure 5 shows an illustration of various abnormal cervical cells. In a large-scale cervical cell screening program for the general population, the number of abnormal squamous cases is far more than abnormal glandular cases and ASC-US, LSIL, ASC-H, and HSIL are the four most common types. ASC-US and LSIL lesions usually occur in superficial cells or intermediate cells while ASC-H and HSIL lesions usually occur in parabasal cells and basal cells.

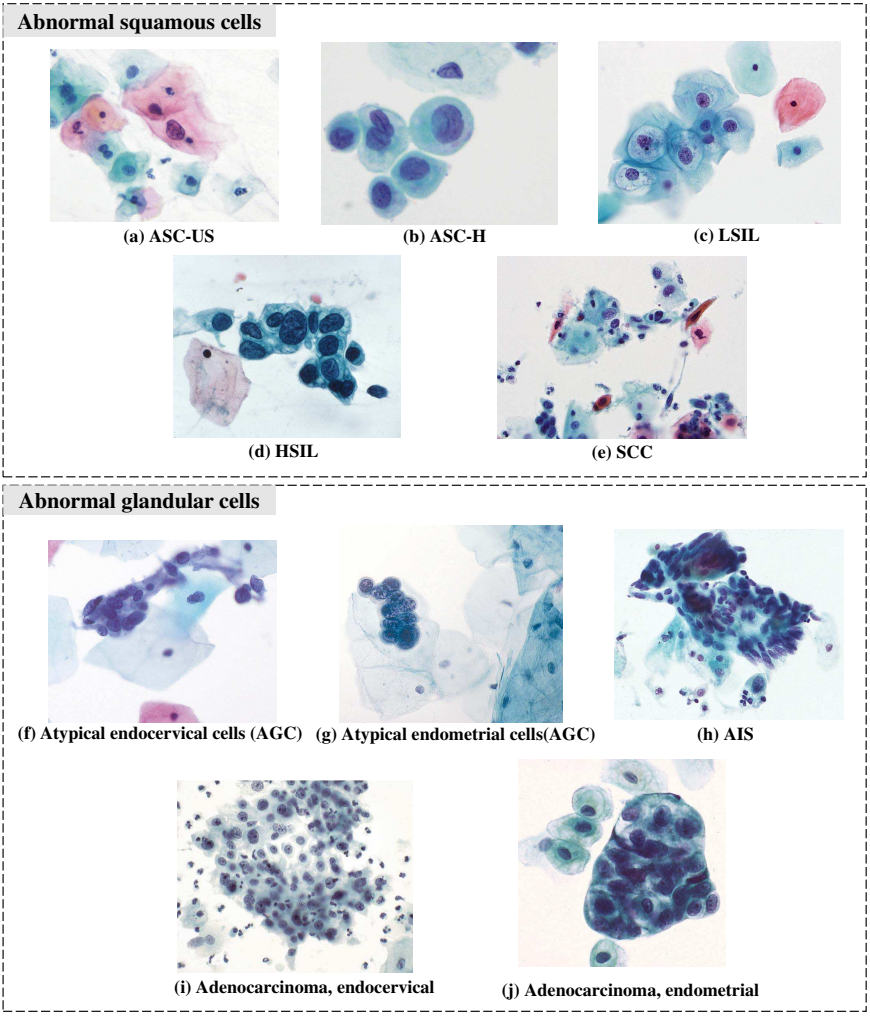


Fig. 5 Illustration of various abnormal cervical cells.

3 Deep learning in cervical cytology

The aim of automated cervical cytology screening is to automatically diagnose digital slides of subjects by computer modeling. This analysis procedure involves searching the region of interest (ROI), segmenting cells, and classifying precancerous or cancerous cells. With the development of such techniques as medical imaging, computer vision, and machine learning, automated analysis receives increasing attention. The realization and application of such methods are helpful to improve the efficiency and accuracy of cytologists performing WSI examinations during the cervical cytology screening process. In this section, we first survey publicly available cervical cytology datasets (Section

3.1) and then summarize the literature on various deep learning methods applied in cervical cytology, including several representative clinical tasks: cell-level identification (Section 3.2), detection (Section 3.3), segmentation (Section 3.4), and slide-level diagnosis(Section 3.5).

3.1 Public datasets of cervical cytology

At the beginning of developing automatic methods for cervical cytology screening, many human and material resources are devoted to the collection of cervical cytological images because automatic analysis methods rely on large amounts of labeled data and there is few public datasets available. We summarize publicly available datasets for cervical cytology screening, as listed in Table 2. These public cervical cytology datasets can be utilized to develop automatic analysis algorithms for multiple tasks, including image classification, object detection, semantic segmentation, etc.

Herlev [41]. Herlev is the most widely used dataset for the analysis of cervical cytology, which consists of 917 Papanicolaou (Pap) smear cervical images in 7 classes (3 normal classes and 4 abnormal classes) based on the classification rule of the 3-tiers dysplasia system. Each cell image is segmented manually into the background, cytoplasm, and nucleus for further feature extraction.

ISBI 2014 [42]. This dataset is released for the first Overlapping Cervical Cytology Image Segmentation Challenge under the auspices of the IEEE International Symposium on Biomedical Imaging (ISBI 2014). The main target of this challenge is to extract the boundaries of individual cytoplasm and nucleus from overlapping cervical cytology images. The dataset consists of 16 Extended Depth Field (EDF) cervical cytology images and 945 synthetic images. Each image consists of 20 to 60 Papanicolaou-stained cervical cells with different degrees of overlap.

ISBI 2015 [43]. This dataset is used for the second cervical cell segmentation challenge in ISBI 2015, consisting of a collection of 17 multi-layer cervical cell volumes, from which 8 will be used for training and 9 for testing. The main difference between ISBI 2015 and ISBI 2014 is that the input data will consist of a multi-layer cytology volume, which means that the input data is now a volume consisting of a set of multi-focal images acquired from the same specimen. This richer input dataset may provide more information on the task of detecting and segmenting cervical cells, thus enabling more accurate cytoplasmic and nuclear detection and segmentation of cervical cells.

SIPaKMeD [44]. This database consists of 4,049 images of isolated cervical cells. The cells are annotated by experienced cytopathologists into five different classes (superficial-intermediate, parabasal, koilocytotic, dyskeratotic, and meta-plastic cells), depending on their cytological appearance and morphology. Among these five classes, superficial-intermediate and parabasal are normal cells. Koilocytes and dyskeratotic are abnormal but not malignant cells while metaplastic belongs to benign. In each image of the SIPaKMeD database, the areas of the cytoplasm and the nucleus are manually defined.

Table 2 Summary of publicly available datasets for cervical cytology screening

Dataset	year	Preparation	Terminology	Task	Description	Link
Herlev [41]	2005	CPS	three-tiers dysplasia	Classification Segmentation	917 cells in 7 classes	http://mde-lab.aegean.gr/downloads
ISBI 2014 [42]	2014	LBC	-	Segmentation	16 EDF cell images and 945 sythetic images	https://cs.adelaide.edu.au/~carneiro/isbi14_challenge/index.html
ISBI 2015 [43]	2015	LBC	-	Segmentation	17 multi-layer cervical cell volumes	https://cs.adelaide.edu.au/~carneiro/isbi15_challenge/index.html
SIPaKMeD [44]	2018	CPS	cytological morphology	Classification	4,049 cervical cells in 5 catogories	https://www.cs.uoi.gr/-marina/sipakmed.html
CERVIX93 [45]	2018	LBC	TBS	Classification Detection	93 stacks of images with 2,705 annotated nuclei	https://github.com/parham-ap/cytology_dataset
BHS [46]	2019	CPS	-	Segmentation	194 images	https://sites.google.com/view/centercric
BTTFA [47]	2019	LBC	-	Segmentation	104 cervical LBC images	https://data.mendeley.com/datasets/jks43dkjj7/1
Mendeley LBC [48]	2020	LBC	TBS	Classification	963 LBC images in 4 classes	https://data.mendeley.com/datasets/zddtpgzv63/4
Cric [49]	2021	CPS	TBS	Classification	400 images with 11,534 cells	https://database.cric.com.br
Comparison Detector [50]	2021	LBC	TBS	Detection	7,410 images with 48,587 objects in 11 categories	https://github.com/kuku-sichuan/ComparisonDetector
RepoMedUNM [51]	2021	LBC + CPS	TBS	Classification	6,168 Pap smear cell images	http://repomed.nusamandiri.ac.id/
CCEDD [52]	2022	LBC	-	Edge Detection Segmentation	686 raw cervical images and 33,614 cut images	https://github.com/nachifur/LLPC
Cx22 [53]	2022	LBC	-	Segmentation	1,320 images of 14,946 cellular instances	https://github.com/LGQ330/Cx22

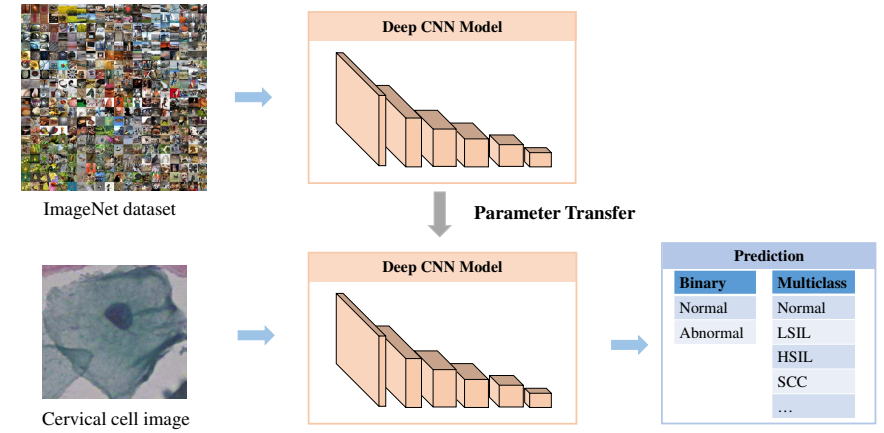
CERVIX93 [45]. This dataset consists of 93 stacks (frames) of images at 40× magnification. Each of the stacks has 10-20 images and all images are size 1280 × 960 pixels. Based on TBS, all frames are examined by cytologists and graded with three categories (Negative, LSIL, HSIL). A total of 2705 nuclei are manually annotated with bounding boxes according to all grade categories.
BHS [46]. This database collects 194 conventional pap smears from the Brazilian Health System (BHS). The collected glass slides are digitized with a magnification of 40× to construct the training dataset (26 images) and test dataset (168 images). The images are labeled into two classes (normal and abnormal) and abnormal images contain 5 different types of precancerous or cancerous cells (Carcinoma, HSIL, LSIL, ASCUS, and ASCH).
BTTFA [47]. This dataset contains 104 cervical LBC images with the size of 1024 × 768 scanned via the Olympus microscope B x 51 with a magnification of 200×. conventional pap smears from the Brazilian Health System (BHS). All collected images are manually segmented by a professional pathologist to get the pixel-level segmentation label.
Mendeley LBC [48]. This dataset consists of a total of 963 liquid-based cytology (LBC) images which can be used for a comparative assessment of one's experimental findings against publicly available conventional pap smear datasets. The dataset has been subdivided into four categories: NIML (613), LSIL (163), HSIL (113), and SCC (74).
CRIC [49]. The collection of the CRIC dataset has 400 images of conventional cervical pap smears and 11,534 classified cells. CRIC Cervix collection covers cervical cells with six types based on TBS nomenclature: NILM (6,779), ASC-US (606), LSIL (1,360), ASC-H(925), HSIL (1,703), and SCC (161).
Comparison Detector [50]. This database consists of 7,410 cervical images cropped from the WSIs. There is a total of 48,587 object instance bounding boxes labeled by experienced cytopathologists. According to TBS categories, the annotated objects belong to 11 categories: ASC-US, ASC-H, LSIL, HSIL, SCC, AGC, trichomonas (TRICH), candida (CAND), flora, herps and actinomyces (ACTIN).
RepoMedUNM [51]. This database is comprised of 6,168 Pap smear cell images including both non-ThinPrep Pap test images and ThinPrep Pap test images. For non-ThinPrep images, there are 3,083 images in total containing two categories, normal and LSIL. ThinPrep images are divided into three categories: normal cells (1,513), koilocyt cells (434), and HSIL (410).
CCEDD [52]. This dataset collects 686 cervical images with the a size of 2048 × 1536. The captured images contain overlapping cervical cell masses in various complex backgrounds and are labeled by 6 experienced cytologists to outline the contours of the cytoplasm and nucleus. The original images are divided into training set, validation set, and test set using a ratio of 6:1:3. All raw image are cut into 512 × 384 pixels and 33,614 cut images in total are obtained.
Cx22 [53]. This dataset is an extension of CCEDD dataset that more precise instances (cytoplasm and nucleus) are annotated. A total of 14,946 cell

instances in 1,320 images are collected and divided into two sub-sets, Cx22-Multi (containing multiple instances) and Cx22-Pair (only containing a pair of instances).

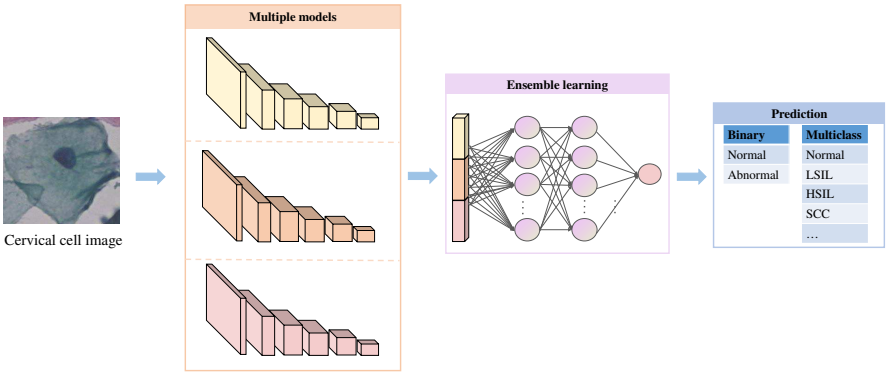
3.2 Cervical cell identification

Cell-level identification is one of the most successful tasks applied by deep learning in cervical cytology screening. Traditional machine learning methods need to accurately segment the cell outline and even the nucleus, and then manually design the features (nucleus area, cytoplasm area, nucleus perimeter, cytoplasm perimeter, N/C ratio, etc.). The extracted hand-crafted features are fused and utilized for final classification, to realize the identification of cervical cells. Most of the traditional machine learning-based methods rely on the accuracy of cell segmentation, which is the key to feature extraction. However, in actual clinical practice, the complex background and fuzzy overlapping cells bring serious difficulties to the accurate segmentation of cervical cells. Conversely, a DL-based identification scheme in the form of a convolutional neural network (CNN) avoids complex image preprocessing steps such as pixel-level cell segmentation, feature selection, and extraction. Owing to the learning of abundant training data, DL-based approaches have gradually become a promising research direction use can realize end-to-end and high-performance identification of cervical cells. The most straightforward approach is to feed the cell image directly into a deep CNN model to extract the feature maps, then use the output layer and a classifier to obtain the predicted category. Shanthi et al. [54] designed a CNN architecture composed of three convolutional layers, three max-pooling layers and one fully connected layer. They evaluated the proposed network on four different datasets using different settings (2 class, 3 class, 4 class, and 5 class), showing its ability for cervical cell identification. Chen et al. [55] proposed a novel network CompactVGG, which is adapted from VGGNet to realize the high-performance classification of cervical cells. On public datasets Herlev and SIPaKMeD, and their collected private dataset, CompactVGG achieved the best performance compared to some classical CNN models. Similarly, DCAVN is proposed to identify cervical cells as normal or abnormal by using deep convolutional and variational autoencoder network [56].

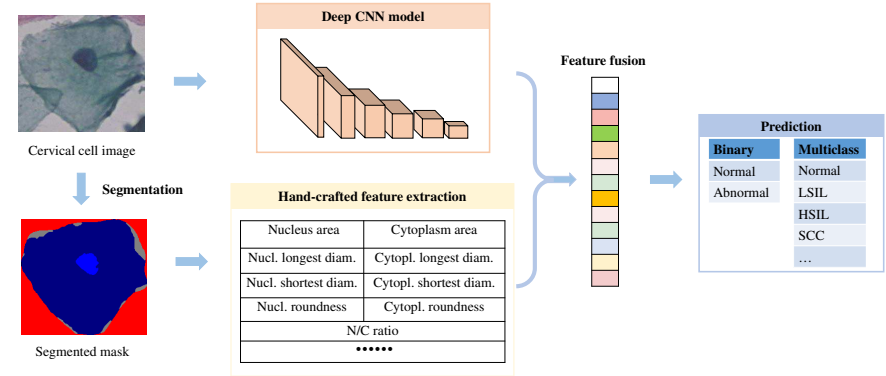
In addition to using the classical CNN architectures or self-designed models, there are three commonly used approaches for cervical cell identification: transfer learning, multi-model ensemble, and hybrid feature fusion, as shown in Fig. 6. A summary of the deep learning-based methods for cervical cell classification is exhibited in the table 3.



(a) Transfer learning based identification



(b) Multi-model ensemble based identification



(c) Multi-type feature fusion based identification

Fig. 6 Three prevalent deep learning based approaches for cervical cell identification.

Table 3: Summary of deep learning-based studies for cervical cell classification. Accuracy (Acc), Precision (Pre), Recall (Rec), Specificity (Spec), Sensitivity (Sens), F1-score (F1).

Reference	Method	Dataset	Classes	Result
		Classical or Self-designed Model		
Shanthi et al. (2019) [54]	Simple CNN (3 Conv + 1 pooling + 1 FC)	Herlev	2-class, 3-class, 4-class, 5-class	(2-class): Acc = 96.11%, (3-class): Acc = 94.80%, (4-class): Acc = 94.62%, (5-class): Acc = 95.31%
Chen et al. (2021) [55]	CompactVGG (Adapted from VGGNet)	Herlev, SIPaKMed and private dataset (60,238 positive cells, 25,001 negative cells, and 113,713 junk images)	2-class, 3-class	Herlev (2-class): Acc = 94.81%, Sens = 95.52%, Spec = 92.76%, F1 = 96.46%; SIPaKMed (3-class): Acc = 98.94%, Sens = 97.80%, Spec = 99.17%, F1 = 98.28%; Private (3-class): Acc = 88.30%, Sens = 92.83%, Spec = 91.03%, F1 = 87.04%
Khamparia et al. [56]	Deep convolutional and variational autoencoder network (DCAVN)	Herlev	2-class	Acc = 99.4%, Pre = 99.4%, Rec = 99.1%, F1 = 99%
		Transfer Learning		
Zhang et al. (2017) [57]	Simple ConvNet + Transfer learning (Fine-tune)	Herlev and private dataset HEMLBC (989 abnormal cells and 989 normal cells)	2-class	Herlev: Sens=98.2%, Spec=98.3%, Acc=98.3%, F1=98.8%, AUC=99.8%; HEMLBC: Sens = 98.3%, Spec = 99.0%, Acc = 98.6%
Hyeon et al. (2017) [58]	VGG-16 + Transfer learning (Feature extraction)	Private dataset (8,373 abnormal cells and 8,373 normal cells)	2-class	Pre = 78.17%, Rec = 78.17%, F1 = 78.17%
Ghoneim et al. (2019) [59]	CNN (shallow CNN, VGG-16 and CaffeNet) + ELM-based classifier + Transfer learning (Fine-tune)	Herlev	2-class, 7-class	(2-class): Acc = 99.5%, (7-class): Acc = 91.2%
Khamparia et al. (2020)[60]	CNN (InceptionV3, VGG-19, SqueezeNet, ResNet-50) + Transfer learning (Feature extraction)	Herlev	2-class	InceptionV3: Acc = 94.88%; VGG-19: Acc = 96.76%; SqueezeNet: Acc = 95.17%; ResNet-50: Acc = 97.89%

Table 3: Continued

Reference	Method	Dataset	Classes	Result
Wang et al. (2020)[61]	Adaptive pruning deep transfer learning model (PsiNet-TAP)	Private dataset (120 normal cell images, 206 uninvolved cell images and 63 abnormal cell images)	2-class	Normal vs abnormal: Acc = 98.41%, Sens = 97.83%, Spec = 98.75%; Normal vs uninvolved: Acc = 98.18%, Sens = 98.57%, Spec = 97.50%; Normal vs abnormal + uninvolved: Acc = 98.49%, Sens = 99.64%, Spec = 95.83%
Bhatt et al. (2021)[62]	Progressive resizing technique + CNN (EfficientNet, VGGNet, ResNet) + Transfer learning (Fine-tune)	Herlev and SIPaKMed	2-class, 5-class, 7-class	SIPaKMed (2-class): Acc = 99.01%, Pre = 99.15%, Rec = 98.89%; SIPaKMed (5-class): Acc = 99.70%, Pre = 99.70%, Rec = 99.72%; Herlev (7-class): Acc = 93.14%, Pre = 94.56%, Rec = 93.98%
		Multi-Model Ensemble		
Rahaman et al. (2021) [63]	CNN (VGG-16, VGG-19, XceptionNet, and ResNet-50) + Hybrid deep feature fusion	Herlev and SIPaKMed	2-class, 3-class, 5-class, 7-class	SIPaKMed: (2-class) Acc = 99.85%, (3-class) Acc = 99.38%, (5-class) Acc = 99.14%; Herlev: (2-class) Acc = 98.32%, (7-class) Acc = 90.32%
Manna et al. (2021) [64]	CNN (Inception v3, Xception-Net and DenseNet-169) + Fuzzy rank-based ensemble	Mendeley LBC and SIPaKMed	2-class, 4-class, 5-class	SIPaKMed (2-class): Acc = 98.55%, Pre = 98.57%, Rec = 98.52%, F1 = 98.54%; SIPaKMed (5-class): Acc = 95.43%, Pre = 95.34%, Rec = 95.38%, F1 = 95.36%; Mendeley LBC (4-class): Acc = 99.23%, Pre = 99.13%, Rec = 99.23%, F1 = 99.18%
Diniz et al. (2021) [65]	CNN (EfficientNet, MobileNet, XceptionNet and Inception v3) + Vote from three best-trained models	CRIC	2-class, 3-class, 6-class	(2-class): Acc = 96%, Pre = 96%, Rec = 96%; (3-class): Acc = 96%, Pre = 94%, Rec = 94%; (6-class): Acc = 95%, Pre = 85%, Rec = 85%
Liu et al. (2022) [66]	CNN (Xception) + Vision transformer (tiny DeiT) + MLP module	Herlev, CRIC and SIPaKMeD combined dataset	2-class, 11-class	Herlev (2-class): Acc = 92.35%, Pre = 88.90%, Rec = 93.50%, F1 = 90.70%; Combined dataset (11-class): Acc = 91.72%, Pre = 91.80%, Rec = 91.60%, F1 = 91.70%

Table 3: Continued

Reference	Method	Dataset	Classes	Result
Kundu et al. (2022) [67]	CNN (GoogLeNet and ResNet-18) + genetic algorithm for feature selection	Mendeley LBC and SIPaKMed	2-class, 4-class, 5-class	SIPaKMed (2-class): Acc = 99.65%, Pre = 99.60%, Rec = 99.58%, F1 = 99.59%; SIPaKMed (5-class): Acc = 98.94%, Pre = 98.79%, Rec = 98.80%, F1 = 98.79%; Mendeley LBC (4-class): Acc = 99.07%, Pre = 98.39%, Rec = 98.18%, F1 = 98.31%
		Hybrid Feature Fusion		
Jia et al. (2020) [68]	Strong features (extracted by GLCM and Gabor) + Abstract features (Extracted from LeNet-5) + Multi feature serial fusion	Herlev and private dataset (1000 positive cell images and 1000 negative cell images)	2-class, 7-class	Herlev (2-class): Acc = 99.3%, Sens = 98.9%, Spec = 99.4%; Herlev (7-class): Acc = 93.8%, Sens = 93.7%, Spec = 93.7%; Private dataset (2-class): Acc = 94.9%, Sens = 93.3%, Spec = 93.3%
Dong et al. (2020) [69]	Hand-crafted features (Color, texure and morphological featurs) + Inception v3	Herlev	2-class, 7-class	Herlev (2-class): Acc = 98.23%, Sens = 99.44%, Spec = 96.73%; Herlev (7-class): Acc = 94.68%
Zhang et al. (2021) [70]	MDHDN (spectrum images + VGG-19 + Hand-crafted features)	Herlev, SIPaKMed and private BJTU dataset (735 normal cells and 1,756 abnormal cells)	2-class, 5-class, 7-class	Herlev (2-class): Acc = 98.7%, Sens = 98.2%, Spec = 98.9%; Herlev (7-class): Acc = 94.8%, Sens = 93.7%, Spec = 91.1%
Yaman et al. (2022) [71]	Multi-resolution images + DarkNet + NCA + SVM	Mendeley LBC and SIPaKMed	4-class, 5-class	SIPaKMed (5-class): Acc = 98.26%, Pre = 98.27%, Rec = 98.28%, F1 = 98.73%; Mendeley LBC (4-class): Acc = 99.47%, Pre = 99.26%, Rec = 98.21%, F1 = 98.73%
Qin et al. (2022) [72]	Multi-task feature fusion model (manual features fitting brance + multi-task classification branch)	SIPaKMeD and private HUSTC dataset (70,197 single-cell images in 5 categories)	2-class, 5-class	HUSTC (2-class): Acc = 99.52%, Sens = 98.90%, Spec = 98.12%, F1 = 99.71%; HUSTC (5-class): Acc = 81.88%, Sens = 79.58%, Spec = 94.97%, F1 = 79.10%; SIPaKMeD (2-class): Acc = 98.96%, Sens = 98.60%, Spec = 99.21%, F1 = 98.72%; SIPaKMeD (5-class): Acc = 98.67%, Sens = 98.65%, Spec = 99.67%, F1 = 98.67%

3.2.1 Transfer learning based identification

The success of deep learning is closely related to large amounts of data, which means that insufficient training data can seriously affect the performance of deep learning models. However, one problem with applying deep learning to medical image analysis is the lack of effective annotation. Limited labels result in limited available data, which makes deep learning models difficult to train well and brings overfitting problems. Therefore, transfer learning is an effective alternative in this case [73]. In contrast to general deep learning algorithms that solve isolated tasks, transfer learning attempts to transfer learned knowledge in the source task and apply it to improve learning in the target task, such as transferring knowledge from a large public dataset (e.g. ImageNet) to a dome-specific task (e.g. Cervical cell identification), as shown in Fig. 6(a). The application of transfer learning in the field of cervical cell identification can save a significant amount of labeling effort, reduce overfitting problems and improve the generalization ability of deep learning models. To transfer deep learning models, fine-tuning and feature extraction are two common strategies [74]. Fine-tuning needs to train the pre-trained model which is obtained from the source dataset on the target dataset to fine-tune all parameters in the learnable layers of the networks. Feature extraction remains the same parameters in all layers except the top layer. The top layer connects with the classifier and is related to the specific classification task.

Zhang et al. [57] first introduced a transfer learning approach to cervical cytology screening for both conventional Pap smear and liquid-based cytology datasets. They proposed a simple ConvNet, DeepPap, to classify the cervical cells into healthy and abnormal, as illustrated in Fig. 7. The proposed ConvNet was firstly pre-trained on a natural image dataset, ImageNet and then fine-tuned on cervical cytological datasets. On both the CPS dataset, Herlev, and the LBC dataset, HEMLBC, the proposed ConvNet presented high-performance classification results.

Hyeon et al. [58] utilized VGGNet-16 which was pre-trained on the ImageNet dataset to extract features of cervical cells and then trained an SVM classifier to perform the prediction. They collected 71,344 Pap smear microscopic images classified into six categories according to TBS criteria. To mitigate the imbalanced distribution they downsampled and regrouped all images into two classes: normal (8,373) and abnormal (8,373). Using 80% of the images for training and the rest for testing, the SVM classifier achieved the best performance with a 0.7817 F1 score when compared to logistic regression, random forest, and AdaBoost.

Ghoneim et al. [59] introduced CNNs and extreme learning machine (ELM)-based classifier in cervical cell classification. They compared the shallow CNN model with two deep CNN models, VGG-16 and CaffeNet. Three deep learning models were fine-tuned on the Herlev dataset, and the proposed CNN-ELM-based system achieved 99.5% accuracy in the 2-class classification and 91.2% in the 7-class classification.

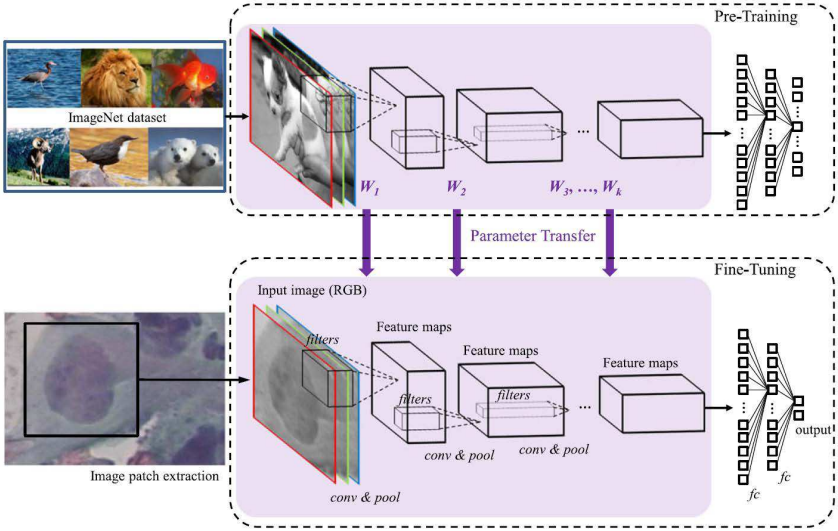


Fig. 7 The architecture of DeepPap [57].

Khamparia et al. [60] proposed a novel internet of health things (IoHT)-driven diagnostic system for cervical cancer. To classify abnormal cervical cells, they leveraged several classical CNN models (InceptionV3, VGG19, SqueezeNet, and ResNet50) as the feature extractor in conjunction with multiple machine learning classifiers (K nearest neighbor, naive Bayes, logistic regression, random forest, and support vector machines.) for final prediction. ResNet50 together with the random forest classifier achieved the highest classification accuracy of 97.89%. They also developed a web application for the prediction of uploaded test images and the proposed IoHT system can greatly improve the diagnosis efficiency of cytologists.

Wang et al. [61] presented an adaptive pruning deep transfer learning model (PsiNet-TAP) to classify Pap smear images. PsiNet-TAP consists of 10 convolution layers and is firstly pre-trained on the ImageNet dataset. After that, transfer learning is applied by using the pre-trained weights as the initialized weights to fine-tune the model on Pap smear images. Furthermore, to discard all unimportant convolution kernels, they designed an adaptive pruning method based on the product of l1-norm and output excitation mean. Using their collected 389 cervical Pap smear images, PsiNet-TAP achieved a remarkable performance of more than 98% accuracy.

Bhatt et al. [62] utilized progressive resizing together with a transfer learning technique to train several generic CNN models for the identification of cervical cells. They performed binary and multiclass experiments on Herlev and SIPaKMed datasets. The experimental results demonstrated the high performance of the proposed method and the activation results of GradCam highlights the pre-malignant or malignant lesions located by the proposed model.

3.2.2 Multi-model ensemble based identification

Ensemble learning is a machine learning technology that exploits multiple base learners to produce predictive results and fuse results with various voting mechanisms to achieve better performances of the learning systems [75]. The basic guiding principle of ensemble learning is 'many heads are better than one'. In recent years, with the rapid development of deep learning, ensemble deep learning has been widely applied in biomedical and bioinformatic fields [76, 77]. The multi-model ensemble is the most straightforward way to realize ensemble deep learning. The diversity of individual networks is the essential characteristic of multi-model ensemble learning and various integration strategies can assist the basic model for better performance. The ensemble across multiple models has been a promising direction to improve accuracy for cervical cell identification, as illustrated in Fig. 6(b).

Rahaman et al. [63] proposed a hybrid deep feature fusion (HDFF) approach, DeepCervix for the multiclass classification task of cervical cells. Four deep learning networks, VGG16, VGG19, XceptionNet, and ResNet50 were used to extract the features and the subsequent feature fusion network was utilized to concatenate the extracted features to perform the final prediction. The HDFF Network achieved an accuracy of 99.85% for 2-class, 99.38% for 3-class, and 99.14% for 5-class classification on the SIPaKMeD dataset. For the Herlev dataset, the proposed method achieved 98.32% and 90.32% for 2-class and 7-class classification respectively.

Manna et al. [64] developed an ensemble-based model for cervical cell classification using three general CNN models, Inception v3, Xception, and DenseNet-169. They presented a novel ensemble technique that the prediction scores of three CNN models were taken into account to make the final decision. The proposed ensemble method leveraged a fuzzy ranking-based approach, where two non-linear functions were applied to the probability scores of each base learner to determine the fuzzy ranks of the classes. The ranks assigned by the two non-linear functions are multiplied and the ranks of the three base learners were added and the lowest rank was assigned as the predicted class. Extensive experiments on two public datasets, SIPaKMeD, and Mendeley LBC demonstrated the high performance of the proposed method in terms of classification accuracy and sensitivity.

Diniz et al. [65] proposed a simple but effective ensemble method to classify cervical cells. After the selection of the three best-trained models from all models, the final prediction was generated by the vote of these three models' predictions. Using the public CRIC dataset, the proposed ensemble method outperformed EfficientNet, MobileNet, InceptionNetV3, and XceptionNet, showing its effectiveness in cervical cell classification.

Liu et al. [66] proposed a DL-based framework, CVM-Cervix for cervical cell classification. CVM-Cervix first combined a CNN module with a visual transformer module to extract local and global features from cervical cell images (See Fig. 8). The Xception model was used as the CNN module to generate 2048-dimensional local features and the tiny DeiT model was used

as a vision transformer module to generate 192-dimensional global features. Then a multilayer perceptron module fused the local and global features to perform the final identification. CVM-Cervix was evaluated on the combination of CRIC and SIPaKMeD datasets, which included 11 categories in total. The experimental results demonstrated the effectiveness of the proposed CVM-Cervix to classify cervical Pap smear images. To meet the practical needs of clinical work, they also introduced a lightweight post-processing to compress the model by using a quantization technique to reduce the storage space of each weight from 32 to 16 bits. The model parameter size was greatly reduced while the classification accuracy remained almost unchanged.

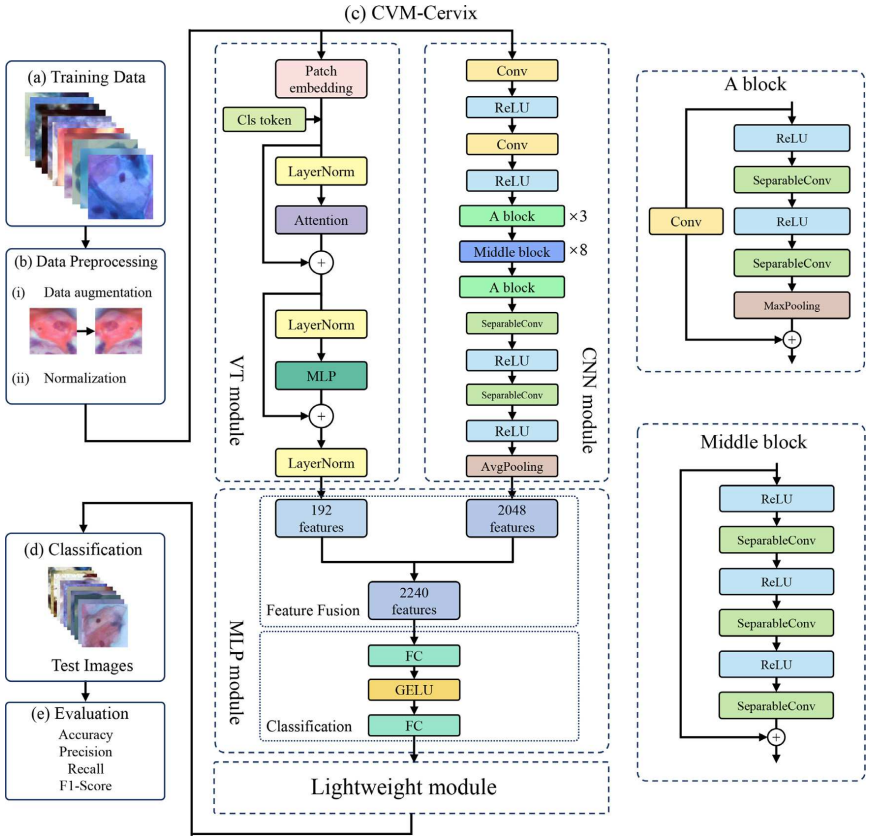


Fig. 8 Detailed network of CVM-Cervix [66].

Kundu et al. [67] employed an evolutionary metaheuristic algorithm, named Genetic Algorithm to select the features which were extracted from GoogLeNet and ResNet-18 models. After feature selection, an SVM served as the classifier to perform the final prediction. The proposed method achieved 99.07% accuracy and 98.31% F1-score on the Mendeley LBC dataset. For the SIPaKMeD

dataset, the proposed method achieved 99.65% and 98.94% for 2-class and 5-class classification.

3.2.3 Hybrid feature fusion based identification

Although the DL-based model has achieved good results in the task of cervical cell classification, there is still a lot of room for improvement. Hand-crafted features, especially some features related to cell morphology, contain rich domain knowledge in the medical field. Incorporating medical domain knowledge with the deep learning network can promote the effective attention of the network and further improve the network performance. Fig. 6(c) shows a general example of combining DL-based features with manual cytological characteristics.

Jia et al. [68] proposed a novel deep learning-based framework called strong feature CNN-SVM. Gray-Level Co-occurrence Matrix (GLCM) and Gabor were used to calculate the strong features. The strong features were fused with abstract features extracted by CNN and then they were sent into the SVM for final prediction. The experimental results on two independent datasets indicated the effectiveness of the strong feature CNN-SVM model in cervical cytology screening.

Dong et al. [69] proposed an innovative cell recognition algorithm that combines hand-crafted features with automatically extracted features via Inception v3 network. To address the low universality of artificial feature extraction while maintaining the cervical cell domain knowledge, they extracted both deep features and hand-crafted features and leveraged a fully connected layer to fuse these features. Furthermore, this paper also utilized an image enhancement algorithm to reduce noise generated during image acquisition and conversion and improve the overall performance. Based on the public Herlev dataset, the proposed method achieved an accuracy of 98.23% for 2-class classification and an accuracy of 94.68% for 7-class classification.

Zhang et al. [70] proposed a novel multi-domain hybrid deep learning framework (MDHDN) to classify cervical cells. It was the first time to apply cell spectrum for cervical cell classification. MDHDN was a three-path cooperative framework, in which two subpaths were used to extract deep features from the time and frequency domains respectively using the VGG-19 network, and the other subpath was used to extract and select hand-crafted features. The final classification results were obtained through the correlation analysis of the prediction of the three paths. On the Herlev dataset, MDHDN acquired an accuracy of 98.7% for 2-class classification and 94.8% for 7-class classification. The proposed framework also presented an excellent performance on the public SIPaKMeD dataset and their collected in-house dataset BJTU.

In [71], Yaman et al. designed an exemplar pyramid deep feature extraction model for the classification of cervical cells. They fed pap-smear images of different resolutions into the DarkNet19/DarkNet53 to get the pyramid features. Then, a Neighborhood Component Analysis (NCA) algorithm was deployed to select the most discriminative features and an SVM classifier was utilized to

execute the final classification. SIPaKMeD and Mendeley LBC datasets were used for method validation. Experimental results demonstrated that the proposed method outperformed some mainstream classification models such as ResNet, DenseNet, InceptionV3, Xception, etc.

Qin et al. [72] presented a multi-task feature fusion model which performed binary classification and 5-class classification for cervical cells (See Fig. 9). The whole model consisted of a manual features fitting branch and a multi-task classification branch. They utilized CE-Net [78] to segment cervical cells for further manual feature acquirement. Multiple discriminatively hand-crafted features including morphological features, integral optical Dens, and texture features were obtained and utilized in the manual features fitting branch to supply prior knowledge for more precise classification. They also utilized smoothing noisy label regularization and supervised contrastive learning strategy for model training. On the SIPaKMeD dataset, the proposed method achieved accuracy of 98.96% and 98.67% for 2-class and 5-class classification that surpassed other SOTA methods. On the self-built dataset, the proposed method also achieved the best performance.

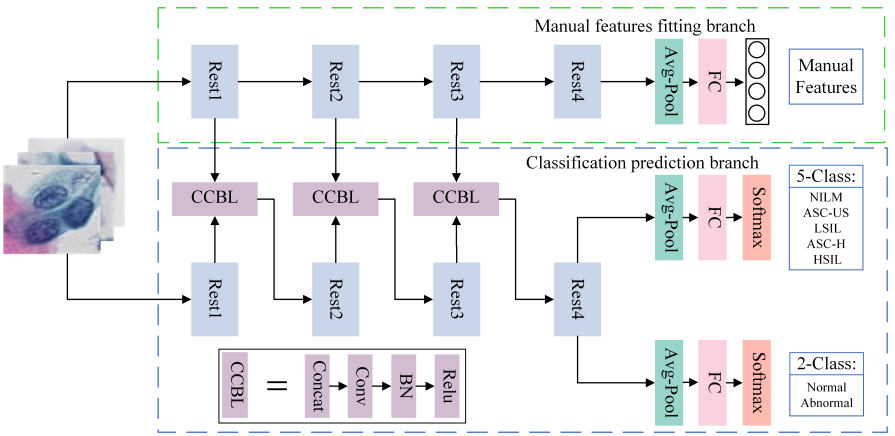


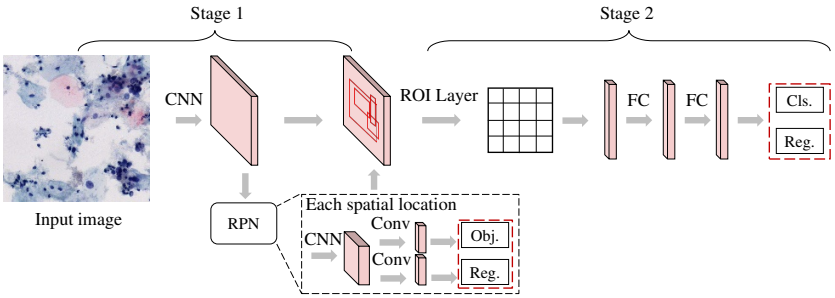
Fig. 9 Multi-task feature fusion model for cervical cell classification [72].

3.3 Abnormal cell detection

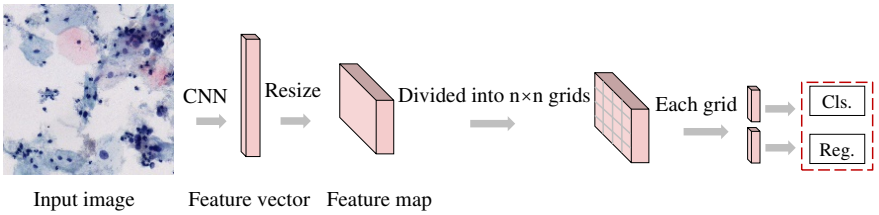
Identifying thousands of cells in a specimen using a classification network alone is time-consuming and inefficient. Thus, a fast search and localization of suspicious abnormal cervical cells are essential for cervical image analysis which further affects the slide-level diagnosis in cervical cytology screening. Object detection models from the computer vision field which simultaneously locate the objects and predict the categories have been well studied and applied in abnormal cervical cell detection.

After the first CNN-based object detection framework R-CNN [79] was put forward, a series of improved algorithms have been proposed which greatly promote the development of generic object detection [80, 81]. There are mainly two types of generic object detection methods: two-stage object detection which involves two stages of region proposal and object detection, and one-stage object detection which directly predicts object bounding boxes and class labels in a single pass [82]. The two-stage object detection is preferred in scenarios where high detection accuracy is required, and the object instances are small or densely packed. In the region proposal stage, the algorithm first generates a set of candidate regions of interest in the image. These regions are proposed as potential locations of objects, and the goal is to reduce the number of regions to be processed in the second stage. This is usually achieved by using algorithms like Selective Search [83], EdgeBoxes [84], or Region Proposal Networks (RPN) [15]. In the object detection stage, the algorithm processes the candidate regions generated in the previous stage and assigns object class labels and bounding boxes to each region. The object detection is usually performed using deep learning models, such as the popular Faster R-CNN [15](see Fig. 10(a)), R-FCN [85], FPN [86] or Cascade R-CNN [87], which use convolutional neural networks (CNNs) for feature extraction and classification. When considering the problem of detection speed, one-stage methods are better choices. One-stage object detection algorithms are typically faster and more efficient than two-stage approaches, as they don't require an initial region proposal step. However, they are generally less accurate, particularly for smaller objects or objects with high levels of occlusion. Some popular examples of one-stage object detection algorithms include YOLO [88] (see Fig. 10(b)), SSD (see Fig. 10(c)), RetinaNet [89] and RefineDet [90].

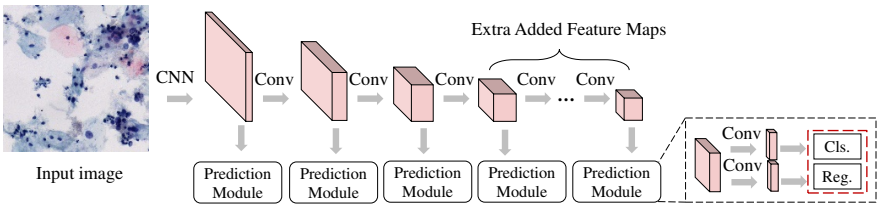
In clinical practice, it's hard to build a high-quality dataset for cervical cell detection since the annotation of cervical cells depends heavily on professional medical knowledge. Thus, some semi-supervised methods have also been explored to detect abnormal cervical cells. In this section, we not only review supervised learning based methods (Section 3.3.1 and Section 3.3.2) for cervical cell detection but survey the latest semi-supervised learning based methods as well (Section 3.3.3).



(a) Two-stage detection model: Faster R-CNN [15]



(b) One-stage detection model: YOLO [88]



(c) One-stage detection model: SSD [91]

Fig. 10 Three commonly used detection models.

Table 4: Summary of deep learning-based studies for abnormal cell detection. Accuracy (Acc), Precision (Pre), Recall (Rec), Specificity (Spec), Sensitivity (Sens), Average precision (AP), Mean average precision (mAP).

Reference	Method	Dataset	Classes	Result
		One-stage supervised learning		
Xiang et al. (2020) [92]	YOLOv3 + Inception V3+ Label smoothing	Private (12,909 cervical images with 58,995 annotations corresponded to 10 categories.)	10-class	mAP = 0.634
Nambu et al. (2021) [93]	YOLOv4 + ResNeSt	Private (919 cervical cell images)	6-class	Detection: AP = 0.542; Classification: Acc = 0.905, Pre = 0.718, Rec = 0.708
Liang et al. (2021) [94]	YOLOv3 + ILCB+ SSAM	Private (12,909 cervical images with 58,995 annotations corresponded to 10 categories.)	10-class	mAP = 0.6544
Jia et al. (2022) [95]	Improved SSD	Private	4-class	mAP = 0.8153
Jia et al. (2022) [96]	Improved YOLOv3	Private (2,000 cell images of 200 patients)	7-class	mAP = 0.7887
		Two-stage supervised learning		
Sompawong et al. (2019) [97]	Mask R-CNN	Private (2,734 normal cells, 494 atypical cells, 148 low-grade cells, and 84 high-grade cells.)	4-class	mAP = 0.578, Acc = 0.917, Sens=0.917, Spec = 0.917
Zhang et al. (2019) [98]	R-FCN (Net-22 as feature extractor)	Private (62 cervical cell images including 180 abnormal regions)	2-class	AP=0.932
Li et al. (2021) [99]	DGCA-RCNN	Tian-chi competition dataset	2-class	$AP_{.1} = 0.505$, $AP_{.3} = 0.486$, $AP_{.75} = 0.445$
Yan et al. (2021) [100]	HSDet	Private (1,000 WSIs)	2-class	mAP = 0.571
Liang et al. (2021) [50]	Comparison detector (Faster R-CNN + Prototype + Few-shot learning)	ComparisonDetector (7,410 cervical microscopical images with 48,587 object instance bounding boxes in 11 categories.)	11-class	Small dataset: mAP = 0.263, Rec = 0.357; Medium-sized dataset: mAP = 0.488, Rec = 0.640
Wang et al. (2022) [101]	3cDe-Net (DC-ResNet as backbone + FPN)	Tian-chi competition dataset and Herlev	2-class	mAP = 0.504
Xu et al. (2022) [102]	Faster R-CNN + FPN + Transfer Learning + Multi-scale Learning	ComparisonDetector	11-class	mAP = 0.616, Rec = 0.877
Liang et al. (2022) [103]	Faster R-CNN + FPN + RRAM + GRAM	Private (40,000 images with 194,880 annotated objects)	10-class	mAP = 0.342, $AP_{.5} = 0.586$, $AP_{.75} = 0.360$

Table 4: Continued

Reference	Method	Dataset	Classes	Result
Liu et al. (2022) [104]	RepPoints + FPN + Grad-Libra Loss	Private dataset	10-class	mAP = 0.531
Chen et al. (2022) [105]	TDCC-Net	Private (6,935 cervical cytologic images including 22,054 ground-truth boxes in 5 categories: ASCUS, LSIL, ASCH, HSIL and SCC)	5-class	mAP = 0.256, $AP_{.5}$ = 0.477, $AP_{.75}$ = 0.255
Semi-supervised learning				
Zhang et al. (2021) [106]	CLCR-STNet	Private (6 categories: Normal, ASCUS, LSIL, ASCH, HSIL and SCC)	6-class	30% labeled: mAP = 0.3207; 50% labeled: mAP = 0.3532; 70% labeled: mAP = 0.3556
Du et al. (2021) [107]	RetinaNet + MT + attention mechanism	Private (9,000 images and 2,050 testing images)	2-class	20% labeled: mAP = 0.579, Pre= 0.28, Rec = 0.78
Chai et al. (2022) [108]	Faster R-CNN + Proposal alignment + Prototype alignment	Private (240,860 images containing lesions of ASCUS, LSIL, HSIL and AGC)	4-class	25% labeled: mAP=0.170; 50% labeled: mAP = 0.195; 75% labeled: mAP = 0.254

3.3.1 One-stage supervised learning based detection

Xiang et al. [92] utilized CNN-based object detection to achieve the recognition of cervical cells. They exploited YOLOv3 as the baseline model and cascaded a further task-specifical classifier to improve the classification performance of hard examples. Furthermore, to relieve the problem of unreliable annotations, they smoothed the distribution of noisy labels. To evaluate the proposed method, they built a dataset composed of 12,909 cervical images with 58,995 ground truth boxes. All labels corresponded to 10 categories. The proposed method eventually achieved an mAP of 63.4% and improved the detection precision of hard samples.

Nambu et al. [93] proposed a two-step screening assistance system for detecting atypical cervical cells. The first step was a quick detection based on YOLOv4 and the second one was a further classification of the localized cells using a ResNeSt model. Experimental results showed that the developed system enabled highly sensitive with fast detection speed.

To relieve the problem that general CNN-based detectors might yield too many false positive predictions, Liang et al. [94] proposed a global context-aware framework based on YOLOv3 using an image-level classification branch (ILCB) and a weighted loss to filter false positive predictions, as shown in Fig. 11. Besides, they presented a soft scale anchor matching (SSAM) method to assign objects to anchors more softly. This paper carried out substantial experiments to evaluate the proposed method and the experimental results validated the effectiveness of the proposed method, which achieved an mAP of 65.44% and gained a 5.7% increase in mAP together with an 18.5% increase in specificity.

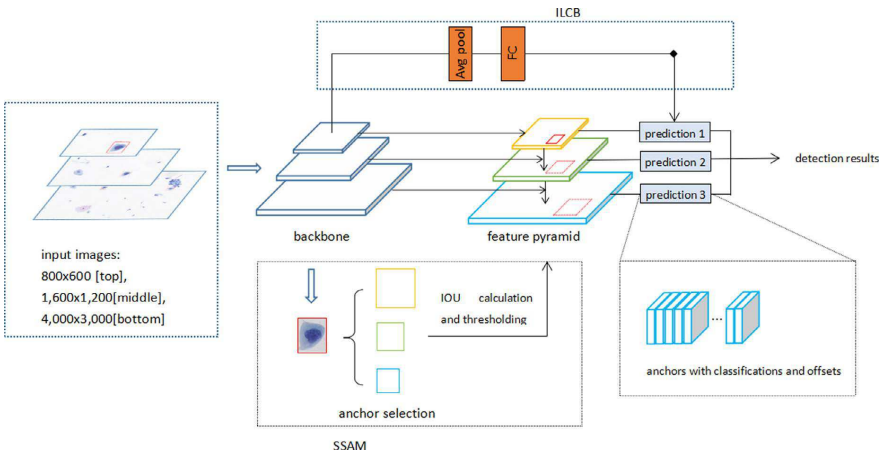


Fig. 11 Global context-aware framework for cervical cell detection [94].

Jia et al. studied one-stage detection method for cervical cancer cells carefully [95, 96]. They improved SSD model by fusing feature maps between different layers in the first work. For the second work, they improved YOLOv3 model by using dense blocks and S3Pool algorithm. To further enhance the performance for cervical cell detection, they did anchor cluster analysis based on k-means++ to select proper anchor size for cervical cells and adjusted the loss function for better training. Both of these two works achieved good detection accuracy for abnormal cervical cells.

3.3.2 Two-stage supervised learning based detection

Sompawong et al. [97] applied Mask Regional Convolutional Neural Network (Mask R-CNN) to detect cervical cancer. In detail, they leveraged ResNet-50 which was pre-trained from ImageNet as the backbone and used a feature pyramid network (FPN) as the detection neck to better select and fuse features. Based on their collected liquid-based dataset, the proposed method obtained an mAP of 57.8%, accuracy of 91.7%, sensitivity of 91.7%, and specificity of 91.7%.

Zhang et al. [98] utilized a region-based, fully convolutional network (R-FCN) for abnormal region detection in cervical cytology screening. Inspired by ResNet, they designed a new feature extractor called Net-22, which consisted of 22 convolutional layers including the structure of the residual block. Experimental results showed that the R-FCN gained an average precision of 93.2%.

Li et al. [99] proposed a novel detection model, deformable and global context aware Faster R-CNN (DGCA-RCNN), to detect abnormal cervical cells in cytology images. DGCA-RCNN improved the original FPN-based Faster R-CNN by introducing deformable convolutional layers and a global context aware (GCA) module. The proposed DGCA-RCNN was evaluated on the public Tian-chi competition dataset and achieved the best performance compared with other SOTA detectors.

In [100], Yan et al. proposed a novel cervical cell detector, HSDet to make better use of negative samples. They adopted HRNet [109] as a feature extractor to cooperate with the cascade R-CNN [110]. Besides, they proposed a pair sampling method to generate the sample pair images and a hybrid sampling strategy to balance hard samples with simple samples. Combining the above methods with HSDet, false detections were effectively decreased. On the in-house dataset consisting of 1000 WSIs, HSDet achieved an mAP of 57.1%, surpassing the Faster R-CNN and Cascade R-CNN models.

Liang et al. [50] proposed an end-to-end cervical cell/clumps detection method called Comparison detector. The Comparison detector utilized Faster R-CNN with FPN as the basic network and adapted the classifier to compare each proposal with the prototype representations of each category. They also investigated the generation method of prototype representations for the background category and considered different designs of the head model. For experiments, the Comparison detector obtained an mAP of 48.8% on its

collected dataset. It's worth noting that on the constructed small dataset, Comparison detector improved by about 20% accuracy than the baseline model.

Wang et al. [101] presented a cervical cancer cell detection algorithm called 3cDe-Net, to address the issue of cell overlap with blurred cytoplasmic boundaries in clinical practice. 3cDe-Net consisted of an improved backbone network named DC-ResNet by introducing dilated convolution and group convolution and a multiscale feature fusion based detection head. Based on the Faster R-CNN algorithm, this paper also generated adaptive anchors and defined a new balanced loss function. The proposed method was evaluated on two publicly available datasets, the Tian-chi competition dataset (Data-T) and the Herlev dataset. Extensive experiments demonstrated the effectiveness of a novel backbone network, DC-ResNet. Besides, the proposed detection algorithm 3cDe-Net achieved an mAP of 50.4%, which significantly improved the performance of the original Faster R-CNN for cervical cancer cell detection.

Xu et al. [102] studied a transfer learning-based method for the detection of cervical cells or clumps. Specifically, Faster R-CNN together with FPN was pre-trained on the COCO dataset and then fine-tuned on cervical cytological images for abnormal cell detection. The authors also utilized a multi-scale training strategy that randomly selected input scales to further improve the performance. The proposed method ultimately obtained an mAP of 0.616 and an average recall of 0.877.

To mimic cytopathologists' diagnostic behaviors that surrounding cells should be referred to identify whether a cervical cell is abnormal, Liang et al. [103] explored contextual relationships in cervical cytological images towards better abnormal cell detection. Based on Faster R-CNN equipped with FPN, they presented RoI-relationship Attention Module (RRAM) and Global RoI Attention Module (GRAM) to respectively capture the cross-cell contextual relationship and global context for context-rich features. Compared with various SOTA detection networks, the proposed method achieved overwhelming success with an mAP of 34.2%.

Liu et al. [104] proposed a Grad-Libra Loss to address the long-tailed data distribution problem in cervical cytology screening that normal or inflammatory cells were much more than cancerous or precancerous cells. Grad-Libra Loss considered the "hardness" of each sample and helped the detection model focus on hard samples in all categories. Various mainstream detectors were utilized to verify the performance of Grad-Libra Loss against the conventional cross-entropy loss. On the collected long-tailed CCA-LT dataset, Grad-Libra Loss presented excellent detection performance superior to other loss functions.

Chen et al. [105] proposed a novel task decomposing and cell comparing network, TDCC-Net for cervical lesion cell detection (Fig. 12). To cope with the large appearance variances between single-cell and multi-cell lesion regions, they decomposed the original detection task into two subtasks detecting single-cell and multi-cell regions, respectively. In addition, to better obtain lesion

features and conform with clinical practice, they designed a dynamic comparing module to perform normal-and-abnormal cells comparison adaptively and present an instance contrastive loss to perform abnormal-and-abnormal cells comparison. Extensive experiments on a large cervical cytology dataset demonstrated that TDCC-Net achieved state-of-the-art performance in cervical lesion detection.

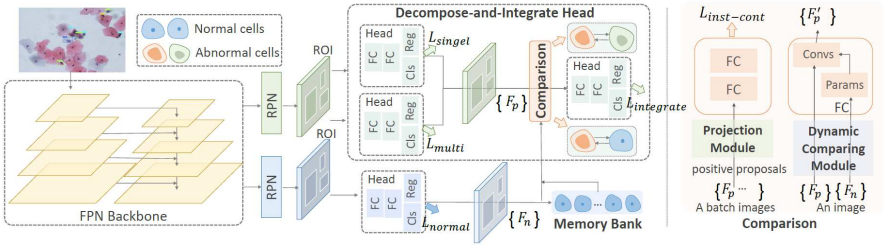


Fig. 12 The architecture of TDCC-Net [105].

3.3.3 Semi-supervised learning based detection

In general, cervical cell detection has been done using supervised learning, where a model is trained on a set of labeled images to learn the patterns that indicate the presence of abnormal cells. However, obtaining a large number of labeled images can be difficult and time-consuming, especially in areas where access to healthcare is limited. Semi-supervised learning based methods for cervical cell detection have been proposed in recent years to alleviate the above problem, which combines the use of labeled and unlabeled data to improve the accuracy of the model [111]. The model uses the labeled data to learn the patterns that indicate the presence of abnormal cells and then applies this knowledge to the unlabeled data to identify additional cases of abnormality. Semi-supervised learning for cervical cell detection has the potential to improve the accuracy of automated systems for detecting abnormal cells, especially in areas where labeled data is scarce.

Zhang et al. [106] proposed a novel semi-supervised cervical cell detection method, called Classification and Localization Consistency Regularized Student-Teacher Network (CLCR-STNet), as shown in Fig. 13. Since it was difficult to acquire large amounts of labeled data in the field of medical image analysis, this paper introduced a novel semi-supervised method that utilized both labeled and unlabeled data with online pseudo label mining. Faster R-CNN was employed as the backbone network and Jensen-Shannon (JS) divergence was used to compute the consistency loss between student and teacher models. The experimental results demonstrated that the proposed CLCR-STNet effectively exerted the potential of unlabeled data and outperformed the supervised methods counterpart.

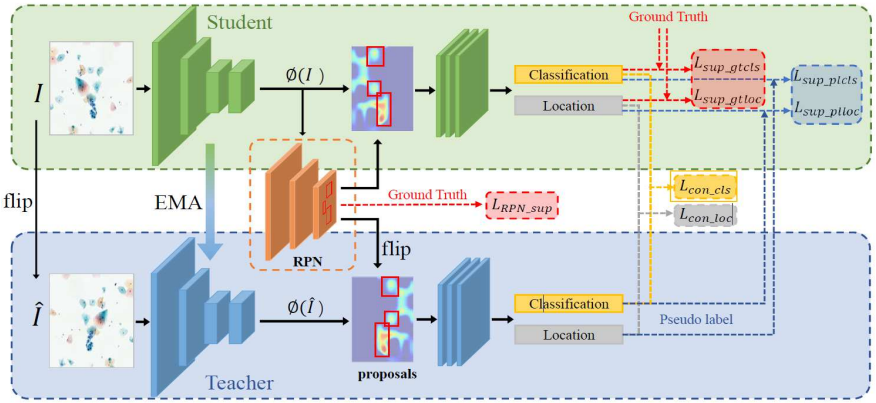


Fig. 13 Overview of semi-supervised model, CLCR-STNet, for cervical cell detection [106].

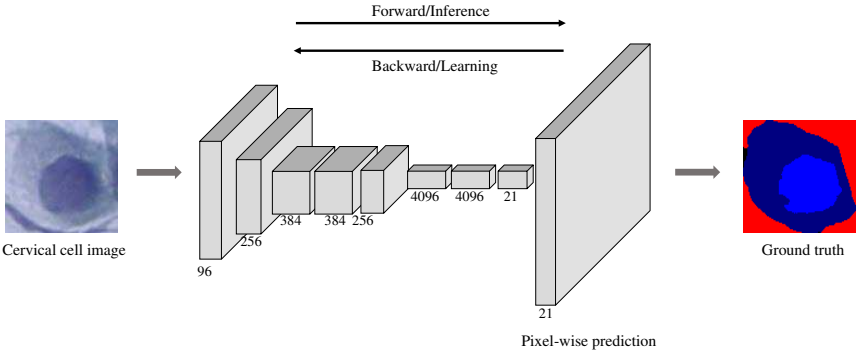
In [107], Du et al. devised a semi-supervised detection network to reduce the false positive rate in cervical cytology screening. To be specific, a RetinaNet was first employed to find the suspicious abnormalities and then a false positive suppression network based on Mean Teacher (MT) model was utilized to execute the further fine-grained classification and decrease the false positive samples. MT model utilized both labeled and unlabeled data for training via the enforced consistency between the teacher network and student network. Moreover, the authors used the generated mask as an attention map to further improve the MT model. Using 20% labeled data and 80% unlabeled data for training, the proposed method achieved 88.6% accuracy which was comparable with the fully supervised method. Besides, the proposed method successfully reduced the false positive rate after using false positive suppressing.

Chai et al. [108] delved into the semi-supervised method for cervical cancer cell detection. To learn more discriminative features, they proposed a deep semi-supervised metric learning network that performed a dual alignment of semantic features on both the proposal level and the prototype levels. Concretely, the pseudo labels were generated for the unlabeled data to align the proposal features with the class proxy derived from the labeled data. Besides, to reduce the influence of possibly noisy pseudo labels, they further aligned the labeled and unlabeled prototypes. They also utilized a memory bank to store the labeled prototypes. The proposed method achieved an average mAP of 27.0% and surpassed another two state-of-the-art semi-supervised object detection methods, the consistency-based semi-supervised detection (CSD) model and Mean Teacher model. Extensive experiments showed that the proposed method could improve the fully-supervised baseline through the use of metric learning.

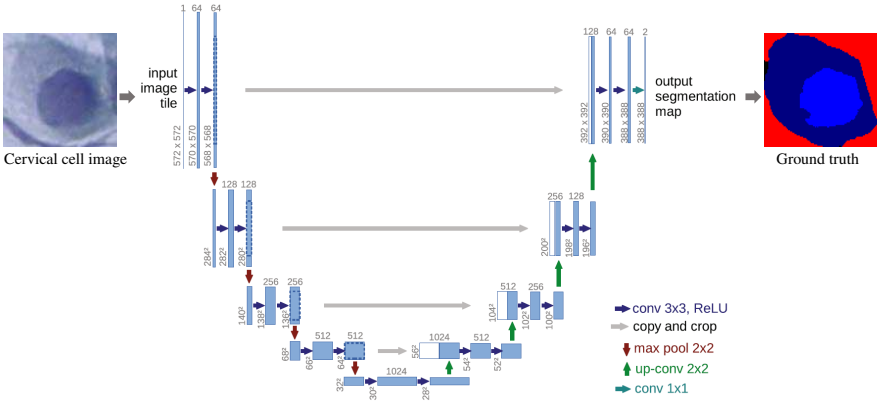
3.4 Cell region segmentation

Cervical cell segmentation is a process to identify and separate individual cervical cells from a digital image in cervical cytology screening. Even though the DL-based classification method has been widely applied to cervical cell identification, which does not need to accurately segment the contours of cervical cells, the segmentation of cervical cell regions is still the fundamental link to carrying out quantitative cell analysis (shape, size, texture, etc.). Besides, a precise segmentation of cell regions can provide detailed cytological features of clinical significance which can further support fine-grained cervical cell identification.

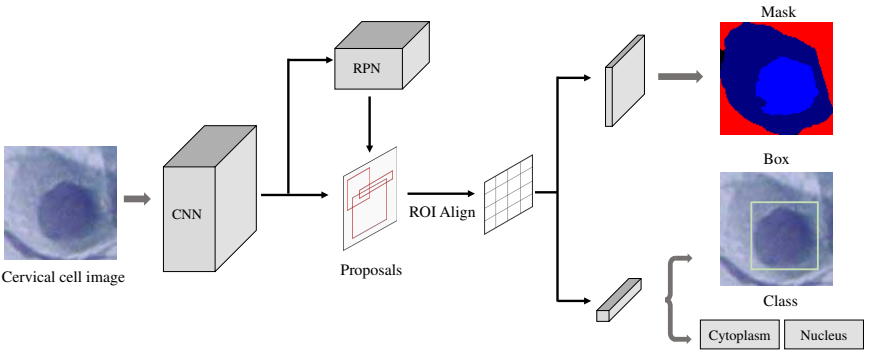
Cervical cell segmentation can be expressed as the problem of classifying pixels with semantic labels (semantic segmentation), especical the differentiation of the nucleus and cytoplasm, or the division of individual cells (instance segmentation). Traditional segmentation methods are generally based on thresholding, edge detection, region growing, k-means, clustering, or watershed methods [112]. With the development of deep learning and CNN, a new generation of deep learning-based segmentation models have been yielded with remarkable performance improvements, and gradually present their potential for medical image segmentation [113]. Fully Convolutional Network (FCN) is a milestone in DL-based segmentation models which firstly introduce CNN into the task of semantic segmentation [114]. Inspired by FCNs, U-Net [115] has been proposed for biomedical image segmentation and received a good reputation and promotion. In addition to the above models, SegNet [116], Mask R-CNN [117], DeepLab [118], and a series of improved methods have been developed to further enhance the performance of image segmentation. Fig. 14 presents three commonly used models for cervical cell segmentation. In this section, the reviewed works encompass segmentation of both cell components (See Section 3.4.1) and overlapping cells (See Section 3.4.2), with the most relevant DL-based approaches being summarized in Table 5.



(a) The FCN model [114]



(b) The U-Net model [115]



(c) The Mask R-CNN model [117]

Fig. 14 Three commonly used commonly used models for cervical cell segmentation.

Table 5: Summary of deep learning-based studies for cervical cell segmentation. Nucleus (Nuc), Cytoplasm (Cyt), Accuracy (Acc), Precision (Pre), Recall (Rec), Specificity (Spec), Sensitivity (Sens), False negative rate (FNR), True positive rate (TPR), Zijdenbos similarity index (ZSI), Dice similarity coefficient (DSC), Average Jaccard Index (AJI), Mean intersection over union (mIoU).

Reference	Method	Dataset	Segmented region	Result
Segmentation of nucleus and cytoplasm				
Zhang et al. (2017) [119]	FCN + graph-based approach	Herlev dataset	Nucleus	ZSI = 0.92
Gautam et al. (2018) [120]	VGGNet-like network + selective pre-processing	Herlev dataset	Nucleus	ZSI = 0.90, Pre = 0.89, Rec = 0.91
Liu et al. (2018) [121]	Mask R-CNN + LFCCRF	Herlev dataset	Nucleus	ZSI = 0.95, Pre = 0.96, Rec = 0.96
Zhang et al. (2019) [47]	BTTFA (Binary-tree-like network topology + Two-path fusion attention)	ISBI 2014 and a new released dataset	Nucleus	ISBI 2014 dataset: DS C= 0.931; Released dataset: DSC = 0.91
Zhao et al. (2019) [122]	D-MEM (Unet + Dense block + Deformable convolution + Multi-path ensemble model)	Herlev dataset	Nucleus	ZSI = 0.933, Pre = 0.946, Rec = 0.984
Zhao et al. (2019) [123]	PGU-net+ (Unet + Residual module + Progressive growing method)	Herlev dataset	Nucleus	ZSI = 0.925, Pre = 0.901, Rec = 0.968
Hussain et al. (2020) [124]	Shape context FCN	Combination of public Herlev, private LBC dataset (1670 raw images) and private conventional dataset (1320 raw images)	Nucleus	ZSI = 0.97
Zhao et al. (2022) [125]	LFANet (UNet + LFA module)	Herlev dataset	Nucleus and cytoplasm	Cyt DSC = 0.9454, AJI = 0.8979, Pre = 0.9453, and Rec = 0.9762; Nuc DSC = 0.9743, AJI = 0.9401, Pre = 0.9592, and Rec = 0.9748
Luo et al. (2022) [126]	DSSNet (Dual-Supervised Sampling Network)	Mendeley-LBC, ISBI2014, and private TJ_sparse (3,600 images), TJ_dense(1,400images)	Nucleus	TJ_sparse: mIoU = 0.7923; TJ_dense: mIoU = 0.6927; Mendeley-LBC: mIoU = 0.7458; ISBI 2014: mIoU = 0.8
Segmentation of overlapping cells				
Song et al. (2016) [127]	Multi-scale CNN and deformation model	ISBI 2015 and private dataset (21 cervical cytology images and each image has 30-80 H&E stained cervical cells distributed in 7 clumps)	Overlapping cell	ISBI 2015 dataset: Cyt DSC=0.91, Nuc DSC = 0.93; Private SZU dataset: Cyt DSC = 0.90, Nuc DSC = 0.9

Table 5: Continued

Reference	Method	Dataset	Segmented region	Result
Tareef et al. (2017) [128]	Super pixel-wise CNN + dynamic shape modeling	ISBI 2014	Overlapping cell	Cyt ZSI = 0.90, Nuc ZSI = 0.94
Xu et al. (2018) [129]	Light CNN model + SLIC + multi-cell labeling	ISBI 2014, ISBI 2015 and private dataset (14 scanned cervical cytology images)	Overlapping cell	ISBI 2014 dataset: DSC = 0.91, FNRo = 0.13, TPRp = ISBI 2015 dataset: DSC = 0.90, FNRo = 0.21, TPRp = Private dataset: DSC = 0.90, FNRo = 0.24, TPRp = 0.
Wan et al. (2019) [130]	TernausNet model + double-window based cell localization + DeepLab V2 model + CRFs + DRLSE	ISBI 2014, ISBI 2015 and private dataset (14 cervical cytology images and each image has 2−20 cells)	Overlapping cell	ISBI 2014 dataset: DSC = 0.93; ISBI 2015 dataset: DSC = 0.92; Private dataset: DSC = 0.92
Zhou et al. (2019) [131]	IRNet (Mask R-CNN + DRM + IRM)	Private dataset (413 images with 4,439 cytoplasm and 4,789 nuclei annotated)	Overlapping cell	Cyt AJI = 0.7185, F1 = 0.7497; Nuc AJI = 0.5496, F1 = 0.7554
Zhang et al. (2020) [132]	Attention U-Net + graph-based Random Walk	ISBI 2014	Overlapping cell	Cyt DSC = 0.917, TP = 0.937, FP = 0.003; Nuc DSC = 0.93, Pre = 0.94, Rec = 0.95
Zhou et al. (2020) [133]	MMT-PSM (Mean Teacher framework + Perturbation-Sensitive Samples Distillation + Mask-Guided Feature distillation)	Private dataset (413 labeled and 4,371 unlabeled images)	Overlapping cell	100% labeled: AJI = 0.6643, mAP = 0.4052; 80% labeled: AJI = 0.6692, mAP = 0.4013; 40% labeled: AJI = 0.6449, mAP = 0.3726.
Mahyari et al. (2022) [134]	Residual CNN model + multi-layer random walker image segmentation + Hungarian algorithm	ISBI 2014 and extended ISBI 2014 dataset (100,000 overlapped cervical cell images are newly created)	Overlapping cell	ISBI 2014 dataset: DSC = 0.89, FNR = 0.10,TPR = Extended ISBI 2014 dataset: DSC = 0.97, FNR = 0.

3.4.1 Segmentation of nucleus and cytoplasm

According to TBS [37], the morphological features, especially variations in the nucleus, are decisive factors supporting the precancerous lesions. There are a number of important specific cytological features that need a precise segmentation of nucleus and cytoplasm, such as nucleus area, cytoplasm area, nucleus/cytoplasm ratio, nucleus roundness, cytoplasm roundness, distribution of nucleus, etc. Thus, the accuracy and reliability of the segmentation algorithm can greatly affect the accuracy of subsequent cell feature extraction. Besides, the segmentation of the nucleus and cytoplasm plays a crucial role in the quantitative analysis of abnormal cells and the accurate diagnosis of cervical cancer. Numerous DL-based studies for the segmentation of cervical cell components have been investigated below.

Zhang et al. [119] combined fully convolutional networks (FCN) and a graph-based approach for the automatic segmentation of cervical nuclei. The overall framework included two steps. FCN was first employed to coarsely split the background, cytoplasm, and nuclei in cervical cell images. Later, the graph-based approach was applied and incorporated with the FCN-learned nucleus probability map to yield fine-grained cell nucleus segmentation results. The proposed method finally obtained a ZSI of 0.92 on the Herlev dataset, superior to several traditional machine learning-based segmentation methods.

Gautam et al. [120] put forward a novel approach for nuclei segmentation in Pap smear images based on deep CNN and selective pre-processing. They emphasize the importance of selective pre-processing since there were significant differences in the image characteristics (e.g. object sizes, chromatin pattern variability) for normal and abnormal cells. Using a VGGNet-like network, the proposed approach excellently accomplished nucleus segmentation on the Herlev dataset with a ZSI of 0.90.

Liu et al. [121] provided pixel-level prior information to train a Mask R-CNN for cervical nucleus segmentation. ResNet together with the feature pyramid network (FPN) was utilized as the backbone of the Mask R-CNN to extract multi-scale features of the nuclei. To refine the segmentation result from Mask R-CNN's output, the authors leveraged a local fully connected conditional random field (LFCCRF). The experimental results on the Herlev dataset showed that the proposed method outperformed other prevailing methods with a precision of 0.96 and an average ZSI of 0.95.

Zhang et al. [47] proposed a binary tree-like network with two-path fusion attention feature (BTTFA) for segmenting cervical cell nuclei. Due to the lack of real-world data for the cervical nucleus segmentation task, at the beginning of the work, they constructed a real-world clinical dataset including 104 LBC-based images with pixel-wise labels manually annotated by professional pathologists. BTTFA model selected ResNeXt as the backbone and utilized a binary tree-like network together with a two-path fusion attention to incorporate multi-level features which compensated for the information loss caused by the pooling layers. The proposed BTTFA was evaluated on the collected real-world dataset and ISBI 2014 dataset. BTTFA obtained a DSC

score of 0.91 on the released dataset, and a DSC score of 0.931 on the ISBI 2014 dataset which outperformed three classical segmentation networks, U-Net, FCN, and DeepLabv3+. The experimental results demonstrated BTTFA provided a feasible method for cervical cell nucleus segmentation.

Zhao et al. [122] suggested a unique method to segment cervical nuclei using Deformable Multipath Ensemble Model (D-MEM). To build the D-MEM, U-Net was adopted as the basic network and dense blocks were exploited to transfer feature information more effectively. To capture the irregular shape of abnormal cervical nuclei and make the network sensitive to subtle changes in objects, deformable convolutions were employed. Moreover, this paper created the multi-path ensemble model by training several networks simultaneously and integrating all paths' predictions for final results.

In [123], a progressive growing U-net (PGU-net+) model was presented to segment nuclei of cervical cells. Residual modules ware inserted into different stages of the U-net to enhance the extraction ability of multi-scale features. Furthermore, the authors adopted the progressive growing method as the network training strategy that could significantly reduce computational consumption and effectively improve the segmentation performance. PGU-net+ gained a ZSI of 0.925 on the Herlev dataset and outperformed the original U-net.

Hussain et al. [124] proposed a shape context fully convolutional neural network (FCN) to extract cervical nuclei which accomplished instance segmentation and classification on Pap smear images simultaneously. Based on standard Unet architecture, they added residual blocks, densely connected blocks, and a bottleneck layer to build the final segmentation network. Besides, a stacked auto-encoder based shape representation model (SRM) was introduced to enhance the strength and robustness of the proposed FCN. To evaluate the performance of the proposed method, extensive experiments were carried out on the combination of three datasets (two clinical datasets and one public dataset, Herlev). The proposed method realized an average Zijdenbos similarity index (ZSI) of 0.97 and surpassed another two deep learning-based models Unet and Mask R-CNN.

To meet the needs of clinical application in practice, Zhao et al. [125] proposed a lightweight feature attention network (LFANet) for abnormal cervical cell segmentation. Two plug-and-play modules, the lightweight feature extraction (LFE) module, and the feature layer attention (FLA) module were introduced to improve the feature extraction ability and reduce the computational consumption. The proposed LFANet achieved the best segmentation results on the Herlev dataset with a low computational complexity showing that LFANet was effective for splitting the nucleus and cytoplasm regions of cervical cells. Besides, the authors also carried out comparative experiments on three other medical image segmentation datasets to further verify the robustness of LFANet.

Luo et al. [126] proposed a dual-supervised sampling network (DSSNet) to accelerate the speed of cervical nucleus segmentation, as illustrated in Fig. 15.

Via the supervised-down sampling module using the compressed images rather than raw images, the amount of the convolution computation was dramatically reduced. Besides, a boundary detection network was exploited to supervise the up-sampling process of the decoding layer to ensure segmentation accuracy. The proposed DSSNet achieved the same level of accuracy as UNet while speeding up for 5 times.

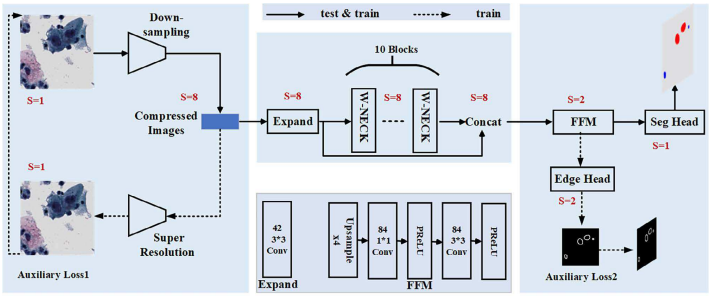


Fig. 15 Detailed structure of DSSNet [126].

3.4.2 Segmentation of overlapping cells

Segmentation of overlapping cervical cells refers to the process of separating individual cells that are overlapped in a cervical cytological image. Early systems focus on segmenting the nucleus and cytoplasm of isolated cells, which is not entirely practical. In clinical practice, the overlap of cervical cells is a very common phenomenon. The large degree of overlap and poor cytoplasmic boundary contrast increase the complexity of the cell segmentation task, which may lead to incorrect diagnoses. In recent years, with the successful organization of the Overlapping Cervical Cytology Image Segmentation Challenge in ISBI 2014 and 2015, an increasing number of works pay attention to this topic. To address this issue, researchers tend to adopt multi-stage approaches in which a coarse segmentation of cell elements is performed first and then the extraction and refinement of overlapping regions followed.

To segment individual cells from overlapping clumps in Pap smear images, Song et al. [127] proposed a novel framework based on a multi-scale CNN and deformation model. The overall segmentation framework consisted of the following three parts: cell component segmentation part to classify the region of nuclei, cytoplasm or background; multiple cells labeling part for splitting of the detected overlapping cytoplasm; and cell boundary refinement and inference part to achieve accurate segmentation results. They evaluated the proposed method with two different datasets, ISBI 2015 Challenge Dataset and Shenzhen University (SZU) Dataset. The experimental results demonstrated that the proposed method outperformed state-of-the-art methods and achieved the highest dice coefficient (DSC) value on both two datasets.

Tareef et al. [128] proposed a variational segmentation framework for cervical cells using super pixel-wise CNN and dynamic shape modeling. The cellular components were first classified into background, nuclei, and cytoplasm based on a CNN model. Then, individual cytoplasm was separated from overlapping cellular mass using Voronoi segmentation and learned shape prior-based evolution. On both versions of the ISBI 2014 datasets (preliminary version and final challenge version), the proposed framework achieved the highest segmentation performance.

Xu et al. [129] presented a novel method for automated segmentation of overlapping cervical cells using a light CNN model and fast multi-cell labeling. They first leveraged a light CNN model which is composed of a convolutional layer a pooling layer and a fully connected layer, to discriminate nuclei part as accurate initialization. Then, for the segmentation of overlapping cytoplasm, they utilized the simple linear iterative clustering (SLIC) method to generate a superpixel map and devised a fast multi-cell labeling method to roughly split clumped cytoplasm. Finally, the cell boundary was refined by an improved distance regularized level set method. The proposed method was validated on three datasets including ISBI 2014 dataset, ISBI 2015 dataset and an in-house dataset. The experimental results showed the effectiveness of the proposed method for the segmentation of overlapping cervical cells.

Wan et al. [130] presented a unique DCNN-based framework to automatically segment overlapping cervical cells. The workflow of the proposed method included cell detection, cytoplasm segmentation, and boundary refinement, as shown in Fig. 16. TernausNet model and the double-window based cell localization method were first utilized to extract the individual cells for cell detection. Then, a modified DeepLab V2 model was constructed to segment the cytoplasm. To refine the cell outer contours, fully connected conditional random fields (CRFs) and distance regularized level set evolution (DRLSE) served as post-processing methods. Three datasets including one in-house dataset and public datasets, ISBI 2014 and ISBI 2015, were served to evaluate the proposed method. The developed DCNN method achieved DSCs of 0.93, 0.92, and 0.92 on ISBI 2014, ISBI 2015, and the in-house dataset, respectively. The high-performance segmentation results showed the effectiveness and potential of the proposed method to be applied for automatic cervical cancer diagnosis.

Zhou et al. [131] proposed Instance Relation Network (IRNet) to segment overlapping cervical cells which explored instance relation interaction, as illustrated in Fig. 17. Based on Mask R-CNN, IRNet introduced Instance Relation Module (IRM) and Duplicate Removal Module (DRM) to improve the network's ability for cell-instance segmentation. IRM could make good use of contextual information and enhance semantic consistency. DRM benefited candidates selection which calibrated the misalignment between classification score and localization accuracy. A large cervical Pap smear (CPS) dataset was built to validate the performance of IRNet and the experimental results demonstrated the effectiveness of IRNet for overlapping cervical cell segmentation.

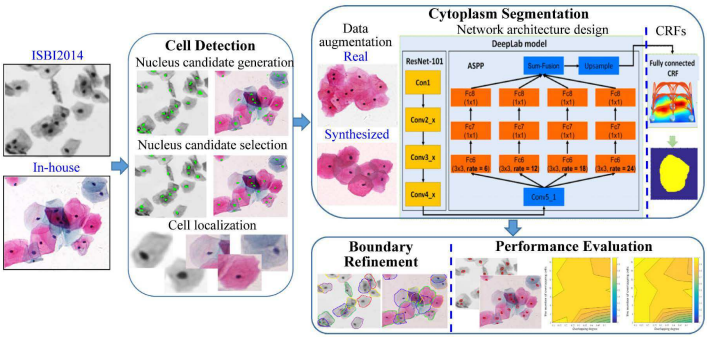


Fig. 16 The workflow of accurate overlapping cell segmentation in [130].

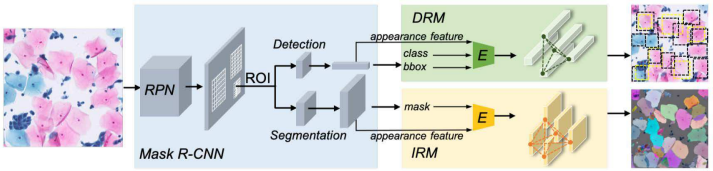


Fig. 17 Overview of IRNet [131].

Zhang et al. [132] proposed a polar coordinate sampling-based approach for overlapping cervical cell segmentation using Attention U-Net and graph-based Random Walk (RW). Attention U-Net was utilized to separate nuclei from the cellular clumps and graph-based RW was exploited to extract the cytoplasm. On ISBI 2014 dataset, the proposed approach gained DSC scores of 0.93 and 0.917 for the nucleus and cytoplasm, respectively. The experimental results demonstrated that the proposed approach was effective and reliable for segmenting overlapping cervical cells.

To address the problem of limited data for cervical cell segmentation since the instance segmentation task required voluminous pixel-level annotations, Zhou et al. [133] proposed a novel semi-supervised method, Mask-guided Mean Teacher framework with Perturbation-sensitive Sample Mining (MMT-PSM), which utilized both labeled and unlabeled data for cervical cell segmentation. MMT-PSM consisted of a teacher network and a student network using the same backbone. The teacher's self-ensemble predictions from augmented samples were used to generate reliable pseudo-labels to supervise the student network. Moreover, mask-guided feature distillation was leveraged to reduce the interference of the background noise. Experiments demonstrated the proposed MMT-PSM outperformed other semi-supervised methods and significantly improved the segmentation accuracy.

Mahyari et al. [134] designed a three-phase scheme for the segmentation of overlapping cells. In the first phase, a self-created residual CNN model was used to generate probabilistic image maps for cell components. In the second

phase, high-probability nuclei nodes were used as seeds for a multi-layer random walker image segmentation for nuclei-seeded region growing. In addition, a cytoplasm approximation could be acquired by thresholding the cytoplasm probabilistic output maps. In the last phase, the Hungarian algorithm was applied to refine the individual pixel locations for the final cell segmentation. On the extended ISBI 2014 dataset, the proposed three-phase method achieved the highest segmentation performance with a DSC of 0.97 over nine different segmentation techniques.

3.5 Whole slide image analysis

Automated WSI analysis has been widely studied in digital histopathological images for cancer diagnosis since the histopathological examination is the most reliable diagnostic basis and the gold standard for clinical diagnosis of cancer [135, 136]. In general, automated WSI analysis is realized by multiple instance learning (MIL) [137], in which each tissue specimen is represented as a bag of instances and each instance is a small image patch extracted from the WSI. MIL belongs to weakly-supervised learning and there is only the slide-level label for all patches in the same WSI. The core of MIL algorithms is to associate the slide-level label (e.g., normal specimen or cancerous specimen) with patch-level features. MIL-based WSI analysis has the potential to improve diagnostic accuracy and has been well studied in histopathology [138, 139].

However, it is still arduous work to perform WSI analysis in cytopathology since the lesion area is continuous in histopathological WSI, and even the presence of a single isolated diseased cell may lead to an abnormal sample in cytopathological WSI. Thus, it is important to leverage an object detection algorithm to search abnormal cells and collect cell-level features. Both cell-level features and patch-level features are crucial for the final slide-level prediction, as shown in Fig. 18. It was not until 2021 that automated cervical cytology screening entered the thorough WIS analysis stage with the presence of the first DL-based WSI analysis methods in cervical cytology screening [140]. In the past two years, several DL-based WSI analysis method for cervical cytology successively emerged.

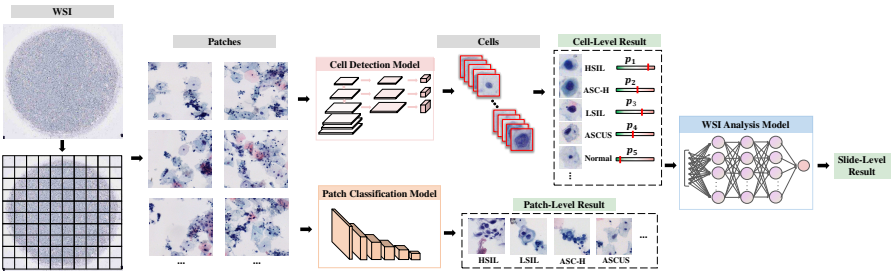


Fig. 18 The process of cervical WSI analysis.

Table 6: Summary of deep learning-based studies for cervical WSI analysis. Accuracy (Acc), Precision (Pre), Recall (Rec), Specificity (Spec), Sensitivity (Sens), Area Under Curve (AUC), F1-score (F1).

Reference	Method	Dataset	Result
Chen et al. (2021) [141]	Unit-level CNN (VGG16, ResNet50) + attention module	Private dataset (264 positive slides and 108 negative slides in total)	AUC = 0.851
Lin et al. (2021) [140]	DP-Net + SGL + RRS	Private dataset (19,303 WSIs in 6 classes from 4 centers)	Sens = 0.907, Spec = 0.80, AUC = 0.925.
Zhou et al. (2021) [142]	Hierarchical pathology screening (RetinaNet + Patch Encoder Module + SVM)	Private dataset (237 WSIs for traing and 361 WSIs for testing)	Acc = 0.905, F1 = 0.867, Sens = 0.891
Zhu et al. (2021) [143]	AIATBS (YOLOv3 + Xception + Patch-based models + U-Net + XGBoost + logical decision tree)	Private dataset (81,727 WSIs retrospective samples for training and 34,403 prospective clinical samples for clinical validation)	Sens = 0.9474
Cao et al. (2021) [144]	AttFPN + ResNet50	Private dataset (325 cases)	AUC =0 .934, Sens = 0.913, Spec = 0.906, and Acc = 0.909.
Cheng et al. (2021) [145]	LR and HR model (ResNet50) + RNN	Private dataset (3,545 WSIs with 79,911 annotations for training and validation, 1,170 WSIs for testing)	Spec = 0.935, Sens = 0.951
Wei et al. (2021) [146]	YOLCO + Transformer	Private dataset (2,019 WSIs)	AUC = 0.872
Pirovano et al. (2021) [147]	Tile classification and localizaiton + Slide level aggregation	Private dataset (40 slides)	Acc = 0.775, Spec = 0.83
Kanavati et al. (2022) [148]	EfficientNet + RNN	Private dataset (1,503 WSIs for training, 150 for validation, and 1,468 WSIs of different settings for testing)	Full agreement set: AUC = 0.960, Acc = 0.907, Sens = 0.850, Spec = 0.911; Equal Balance-rev set: AUC = 0.915, Acc = 0.885, Sens = 0.839, Spec = 0.920; Clinical Balance-rev set: AUC = 0.890, Acc = 0.903, Sens = 0.886, Spec = 0.904
Geng et al. (2022) [149]	FCOS + ResNet34	Private dataset (2,625 WSIs)	2-category task: Sens = 0.9784, Spec=0.8550, Pre = 82.55; 5-category task: Acc = 79.74
Zhang et al. (2022) [150]	RetinaNet + SE-ResNeXt-50 + graph attention network + supervised contrastive learning	Private dataset (3,485 negative WSIs and 3,462 positive WSIs)	Acc = 0.8579, AUC = 0.9252, Rec = 0.8263, Pre = 0.8815, F1 = 0.8528

In [141], an automatic WSI diagnosis was proposed using unit stochastic selection and attention fusion. Chen et al. first constructed a unit-level CNN based on VGG16b and ResNet50 to extract features of each unit (patch or cell). Next, they leveraged a UOI selection method to select the representative features of the WSI and employed an attention module to fuse all units' features for WSI diagnosis. The authors evaluated the proposed framework on three different types of pathological images. For the diagnosis of cervical cytological WSIs, the proposed method achieved good performance with a mean AUC of 0.851.

Lin et al. [140] presented the first work for the specific analysis of cervical whole slide images. Firstly, an efficient deep learning-based dual-path network (DP-Net) was designed for lesion detection. Inspired by medical domain knowledge that different precancerous cervical cells belonged to different groups (epidermal group and basal group), a synergistic grouping loss (SGL) was proposed for fine-grained cell classification. Then, a slide-level classifier called rule-based risk stratification (RRS) was introduced to perform the final WSI diagnosis, which simulated the clinical diagnostic criteria of cytopathologists. To evaluate the proposed method, a large number of samples were collected from multiple medical centers to construct the cervical WSI dataset (19,303 WSIs). The proposed method achieved a high sensitivity of 0.907 and a specificity of 0.80, showing strong robustness for practical cervical cytology screening.

Zhou et al. [142] proposed a hierarchical framework for case-level automatic diagnosis of cervical smears, which consisted of three stages. In the first stage, a large number of cytological images were extracted from the scanned WSI, and cell-level detection was performed for each image using RetinaNet. In the second stage, top-k regions with the highest confidence were selected and fed into the subsequent Patch Encoder Module (PEM) for image-level classification. In the last stage, the confidence scores of all images in each case were collected and used as the feature vectors to train an SVM classifier for final case-level diagnosis. Experiments showed that the proposed framework presented better accuracy than applying object detection and classification network directly.

Zhu et al. [143] developed an AI-aided diagnostic system for automated cervical cytology screening, called AIATBS, which could help cytologists interpret in strict accordance with TBS standards. This system integrated five AI models including YOLOv3 for object detection, Xception for further fine-grained classification, DenseNet-50 for patch-based classification, U-Net for nucleus segmentation, and XGBoost model together with the logical decision tree for final slide-level diagnostic decisions. This paper also presented a digital pathology image quality control (DPIQC) system to ensure the quality of digitized images. AIATBS system was validated at 11 medical centers, and the outstanding performance demonstrated its adoption applicability and robustness for routine assistive diagnostic screening which could reduce the workload of cytologists, and improve the accuracy of cervical cancer screening.

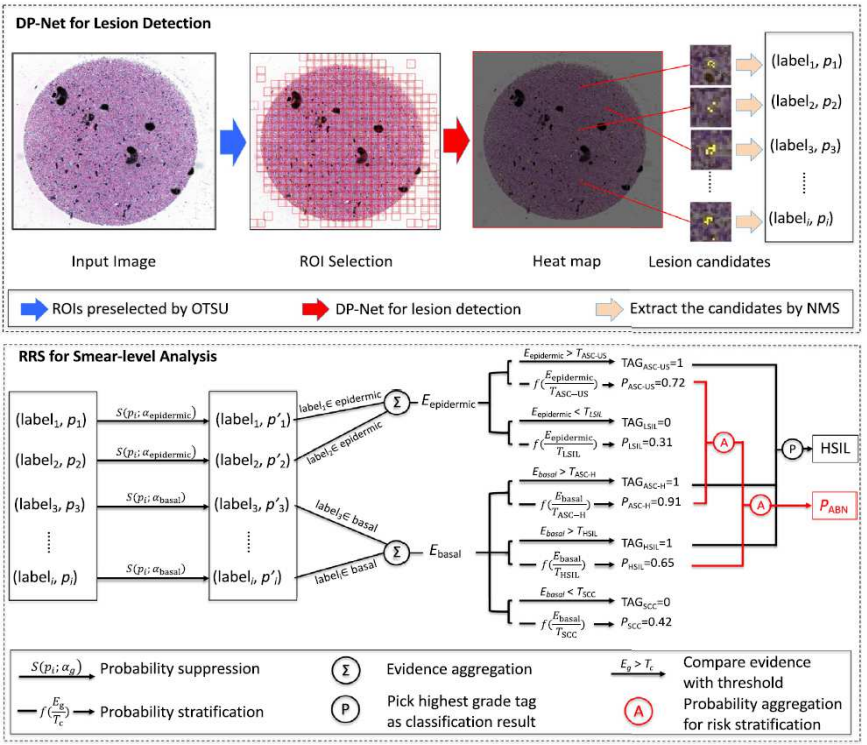


Fig. 19 First WSI analysis framework in cervical cytology screening, DP-Net with synergistic grouping loss and rule-based risk stratification [140].

Cao et al. [144] devised a three-phase framework for automatic cervical cytology screening. Firstly, they proposed a novel attention feature pyramid network (AttFPN) to automatically detect abnormal cervical cells. AttFPN leveraged both channel and spatial attention for multi-scale feature fusion to improve the accuracy of abnormal cervical cells at different scales. Then, the image-level classification results were obtained by using the ResNet50 according to the corresponding probability prediction of detected abnormal cervical cells. At last, The classification results of all image patches in the same WSI were summarized to determine the ultimate case-level result. Extensive experiments demonstrated that AttFPN was effective for abnormal cell detection and the whole system had the potential for routine cervical cancer screening programs.

In [145], Cheng et al. proposed a robust WSI analysis method for cervical cancer screening by imitating the diagnosis process of cytopathologists, in which suspicious cells were found at low magnification and then scrutinized for confirmation at high magnification, as illustrated in Fig. 20. They utilized a low-resolution model cascaded with a high-resolution model to recommend the 10 most suspicious lesion cells in each WSI. Then, an RNN-based

WSI classification model was constructed by integrating the extracted feature representations of the top 10 lesion cells. The proposed system achieved 93.5% specificity and 95.1% sensitivity on multi-center WSI datasets with 1170 samples.

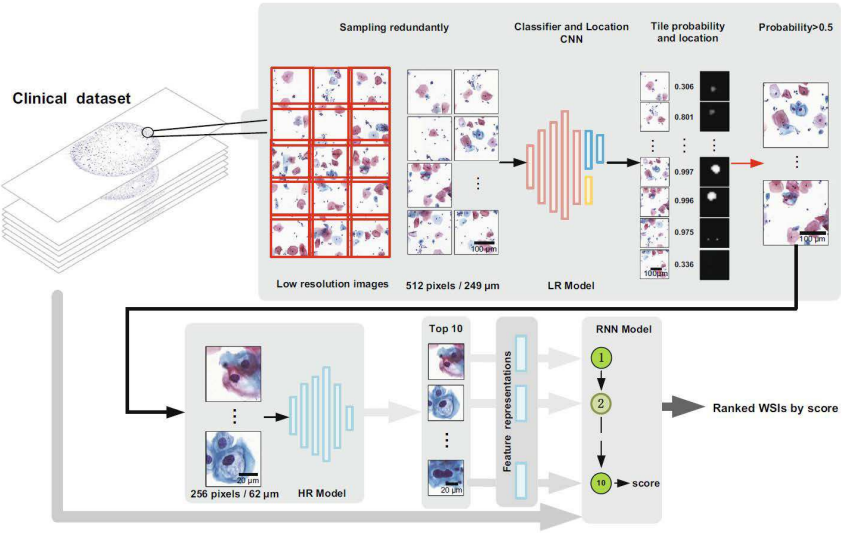


Fig. 20 Robust WSI analysis method using a combination of low resolution model and high resolution model [145].

Wei et al. [146] proposed a progressive framework for cervical WSI analysis based on affluent semantic and location features. They devised a novel lightweight detection model, YOLCO which was adapted from the YOLOv3 detection algorithm to acquire the cell-level and patch-level predictions at the same time. To make the network lighter for practical application, they utilized depthwise separable convolution and inline connection network (InCNet) to replace general CNN. Then, these local predictions were integrated as an input WSI-level feature vector to a transformer architecture for diagnosis results. The experimental results on 2,019 samples demonstrated that the framework achieved a high AUC score of 0.872 with higher detection speed showing its effectiveness and efficiency.

Pirovano et al. [147] devised an explainable region classifier in cervical cytological WSIs. A created dataset and a novel loss were proposed to train an efficient region classifier to perform weakly supervised localization for malignancy regions in WSIs. Besides, they extended their approach to a more general detection task for cell abnormality and a real clinical slide dataset. The results demonstrated its effectiveness and potential to be applied in the current workflow of cytopathologists.

Kanavati et al. [148] developed a DL-based method for WSI analysis of LBC specimens. They utilized a CNN model together with an RNN model to realize the slide-level classification. EfficientNetB0 [151] model was employed to extract features of all tiles in one WSI. The output of the CNN model was adjusted as the input of the RNN model, which then gave a final WSI diagnosis. On 1468 collected test WSIs, the proposed method achieved AUCs in the range of 0.89–0.96, which fully demonstrated its effectiveness for cervical WSI diagnosis.

Geng et al. [149] developed a two-stage learning framework for analyzing gigapixel cervical WSIs including a patch-level feature learning module and a WSI-level feature learning module. Patch-level model leveraged a one-stage object detector FCOS [152] and WSI-level feature learning module utilized a modified ResNet34. The proposed approach achieved state-of-the-art classification performance on both 2-class and 5-class tasks.

Zhang et al. [150] developed a deep learning-based framework for cervical cancer screening which explored the relationships between the suspicious cells and took advantage of other cells for comparison. This system comprised of a ranking and feature extractor based on RetinaNet [89] and SE-ResNeXt-50 [153] model, and a graph attention network (GAT) to model the intrinsic relationships between different patches. They also proposed a supervised contrastive learning strategy to enhance the feature learning capacity for better classification. Extensive experiments validated the effectiveness of the proposed GAT and contrastive learning strategy, which outperformed other prevalent WSI classification approaches.

4 Challenges and opportunities

Despite significant progress in automated cervical cytology screening in recent years, there are still considerable challenges and opening issues that need to be resolved. Furthermore, the development of DL technology and computational cytology is accelerating the advancement of this field. This section further discusses the prospects and potential research directions in automated cervical cytology screening.

Stain Normalization. Due to the variations in staining procedures, staining durations, imaging environments, and scanning instruments, there always exists diverse image styles of the collected cytological images. Such image style inconsistency makes it difficult to build robust and generalized DL-based models of cervical cytology since training data and testing data may have different image styles causing the low performance of trained models in actual deployment. Stain normalization is an ideal way to eliminate the differences in image style. Traditional stain normalization methods such as color transfer, stain spectral matching, color deconvolution, etc. need one or several template images to estimate stain parameters, but a few template images cannot represent the color distribution of the entire reference dataset. Therefore, DL-based stain normalization methods using generative adversarial networks (GANs) are

a better substitute because the whole dataset of the target style is leveraged as the template to execute color normalization by image-to-image translation. For example, Chen et al. [154] proposed a two-stage domain adversarial style normalization framework for cervical cytopathological images and Kang et al. [155] presented StainNet by using StainGAN [156] and distillation learning to complete the stain normalization of cervical cell images.

Image Super-Resolution. Image super-resolution is another promising research direction for cervical cytology images. Out-of-focus and low-resolution images will interfere with the precise diagnosis in cervical cytology screening. However, in real-world screening programs, blur field of view (FoV) caused by scanning too fast without proper focusing is a frequent occurrence in scanned images. In addition, the acquisition of high-resolution digital slides needs advanced scanners which increases the financial burden in remote and underdeveloped regions. To address this problem, single-image super-resolution (SISR) brings an effective solution by converting low-resolution slides into high-resolution slides. Two DL-based SISR methods, PathSRGAN [157] and STSRNet [158] have been proposed for cervical cytopathological images. Both stain normalization and image super-resolution are urgently needed image preprocessing tools to assist DL-based diagnoses to improve inter-laboratory comparability and facilitate the development of CAD systems in cervical cytology screening.

Effective Feature Extractor. Feature extractors are used to learn discriminative features of cytology images in computational cytology [159]. The feature representation capability of the feature extractor will greatly affect the downstream tasks (cervical cell identification, abnormal cell detection, and cell region segmentation). During the initial period of rapid development of deep learning, researchers aimed to enhance the feature extraction ability of deep neural networks by either increasing the depth or width of the network [13, 14, 160, 161]. In the past few years, attention mechanism has been introduced into the field of computer vision and various visual attention module has been proposed [162–164]. visual attention modules which make DL-based models focus on lesion-related parts while inhibiting irrelevant information have been widely employed in automated cervical cytology screening [132, 141, 144]. Most recently, with the successful practice of transformer [165] in multiple computer vision tasks [166–168], vision transformer quickly spread in various research fields. CVM-Cervix [66] demonstrates the superior performance of the vision transformer to serve as an effective feature extractor for cervical cell classification. More vision transformer-based approaches in automated cervical cytology screening are expected in the future.

Incorporating Medical Domain Knowledge. Since experienced cytopathologists can often give fairly accurate diagnoses, it's not surprising that their knowledge may guide DL-based models to do their assigned tasks better. The specialized knowledge of cytopathologists for cervical cytology refers to the cytological characteristics they learned, the way they browse the slides, the features they pay special attention to, and the training process

they experienced [169]. A simple way to realize the incorporation of medical domain knowledge is to combine the hand-crafted features with DL models since the manual features contain cytological characteristics related to diagnosis which are definitely pointed out in guidelines and criteria of cervical cytology, as mentioned in Section 3.2.3. Besides, Lin et al. presented a synergistic grouping loss and a rule-based risk stratification system using the cell grouping rules in TBS criterion[140]. Cheng et al. devised a DL-based model mimicking the cytopathologists' habits of viewing specimens [145]. Cao et al. attention-guided network, AttFPN, to pay special attention to lesion-related areas [144]. Moreover, Chen et al. [105] built TDCC-Net by leveraging the diagnosis experience of cytopathologists that normal cells in the same image should be used as a reference for better identification of abnormal ones. All the above studies make good use of medical domain knowledge to guide the construction of the DL model so as to achieve excellent results. There is a wealth of untapped medical knowledge that could be leveraged to develop high-performance and interpretable DL models.

Annotation-Efficient Learning. Unlike natural images, the annotation of medical images requires specialized medical knowledge. The extensive annotation work can be a heavy burden for cytologists, making it difficult to obtain a large-scale dataset of high quality in cervical cytology screening. To address limited and noisy labels, annotation-efficient learning has emerged which is generally accomplished by transfer learning, domain adaptation, weakly supervised learning(multiple instance learning), semi-supervised learning, and self-supervised learning [170]. Wang et al. proposed a novel annotation-efficient learning method for medical image segmentation based on noisy pseudo labels and adversarial learning [171]. Hu et al. utilized semi-supervised contrastive Learning to segment MRI and CT images [172]. For cervical cytology, several semi-supervised learning based methods have also been proposed to detect abnormal cell detection or segment overlapping cells [106–108, 133]. These approaches have successfully improved the labeling efficiency and exhibit high accuracies which are comparable with full-supervised methods. More methods deserve to be explored and studied in cervical cytology screening by using annotation-efficient learning.

Multi-modal Data Fusion. With the recent advancements in multi-modal deep learning technologies, significant progress has been made in the field of cancer diagnosis and prognosis analysis [173, 174]. Multi-modal data fusion aided decision is also a good choice to realize slide-level diagnosis in cervical cytology screening. In addition to cervical cytopathological images, clinical data such as electronic medical records (EMRs) is also a critical reference in the final slide-level diagnosis. EMR contains a great deal of helpful personal information (Age, duration of menstrual period, medical history, cytology screening record, etc.) that can be utilized to guide more accurate diagnosis. At present, there is no related work of multi-modal data-based diagnosis in cervical cytology but this is a potential task in the future. Combining natural language

processing (NLP) and computer vision (CV) technology to extract image features and clinical text features, and building a multi-modal classification model to realize the interactive fusion of multi-source data are meaningful to realize the precise diagnoses and personalized recommendations for cervical cytology.

Internet of Medical Things (IoMT). IoMT is an emerging challenge of the conventional internet of things (IoT) which enables the connection of medical devices, software applications, and health systems that collect and exchange healthcare data [175]. IoMT can provide significant benefits by increasing access to care, improving the quality of medical service, and reducing healthcare costs, particularly for patients in remote or underdeveloped regions. IoMT also enables the integration of deep learning algorithms into healthcare, allowing for real-time disease diagnosis and personalized treatment plans based on individual patient data. Liu et al. proposed a Dental IoMT system based on intelligent hardware, deep learning, and mobile terminals, aiming to explore the feasibility of in-home dental health [176]. Guo et al. proposed a hybrid intelligence-driven IoMT system to diagnose pathological myopia for remote patients by combining conventional machine learning with deep learning [177]. In automated cervical cytology screening, there are also some works that designing smart scanners and IoMT systems to promote digital pathology in rural areas and remote hospitals [178–181]. The design of a universal and efficient cytopathological IoMT system is the ultimate pursuit for automated cervical cytology screening and there is still a long way to go.

Federated learning. In order to successfully apply the DL-based model to the actual clinical screening programs, strong generalization is a guarantee. Currently, most DL methods can achieve considerable performance on their internal datasets whereas the results are less than satisfactory when applied to clinical environments [19]. With the improvement of medical services and the promotion of IoMT, there are increasing concerns about the security and privacy of healthcare data. The lack of data privacy has restricted data sharing among medical institutions and further affected the construction and verification of the DL-based model with superior generalization [18]. Recently, federated learning (FL) has been proposed to address the above issue which allows multiple parties to collaborate on the training of a shared model without sharing their data [182]. IoMT and FL can work together to improve the accuracy of DL-based models and guarantee generalization while maintaining patient data privacy and security. For example, a COVID-19 IoMT System is proposed by using FL and blockchain [183] and a novel skin disease detection system is presented with the integration of federated machine learning [184]. Overall, IoMT together with FL has the potential to revolutionize cervical cytology screening for cancer prevention and timely treatment.

5 Conclusion

In this survey, an overview of cervical cytology and its current screening procedures is first introduced. Then, we offer a comprehensive collection of public

image datasets for cervical cytology. Next, the most relevant DL-based image analysis methods in automated cervical cytology screening have been analyzed. From these summarized approaches, different learning paradigms (transfer learning, ensemble learning, semi-supervised and weakly supervised learning) have been applied to multiple tasks (cell identification, abnormal cell or suspicious area detection, cell component or overlapping cell segmentation, and WSI diagnosis) in cervical cytology screening. Since the primary objective of this review is to aid the advancement of automated tools that can effectively facilitate cervical screening procedures. The primary objective of this survey is to aid the advancement of CAD tools that can effectively facilitate automated cervical cytology screening programs. Additionally, this work provides insights into potential directions for future research including data preprocessing, feature representation, model design, clinical application, and privacy security.

Acknowledgments. This work was supported by the Major Projects of Technological Innovation in Hubei Province (2019AEA170), the Frontier Projects of Wuhan for Application Foundation (2019010701011381), the Translational Medicine and Interdisciplinary Research Joint Fund of Zhongnan Hospital of Wuhan University (ZNJC201919).

References

[1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **71**(3), 209–249 (2021)

[2] Cramer, D.W.: The role of cervical cytology in the declining morbidity and mortality of cervical cancer. Cancer **34**(6), 2018–2027 (1974)

[3] Cohen, P.A., Jhingran, A., Oaknin, A., Denny, L.: Cervical cancer. The Lancet **393**(10167), 169–182 (2019)

[4] Elsheikh, T.M., Austin, R.M., Chhieng, D.F., Miller, F.S., Moriarty, A.T., Renshaw, A.A.: American society of cytopathology workload recommendations for automated pap test screening: Developed by the productivity and quality assurance in the era of automated screening task force. Diagnostic cytopathology **41**(2), 174–178 (2013)

[5] Koss, L.G., Lin, E., Schreiber, K., Elgert, P., Mango, L.: Evaluation of the papnet™ cytologic screening system for quality control of cervical smears. American journal of clinical pathology **101**(2), 220–229 (1994)

[6] Biscotti, C.V., Dawson, A.E., Dziura, B., Galup, L., Darragh, T., Rahemtulla, A., Wills-Frank, L.: Assisted primary screening using

the automated thinprep imaging system. American journal of clinical pathology **123**(2), 281–287 (2005)

[7] Kardos, T.F.: The focalpoint system: Focalpoint slide profiler and focalpoint gs. Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society **102**(6), 334–339 (2004)

[8] Marinakis, Y., Dounias, G., Jantzen, J.: Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. Computers in Biology and Medicine **39**(1), 69–78 (2009)

[9] Chen, Y.-F., Huang, P.-C., Lin, K.-C., Lin, H.-H., Wang, L.-E., Cheng, C.-C., Chen, T.-P., Chan, Y.-K., Chiang, J.Y.: Semi-automatic segmentation and classification of pap smear cells. IEEE Journal of Biomedical and Health Informatics **18**(1), 94–108 (2013)

[10] William, W., Ware, A., Basaza-Ejiri, A.H., Obungoloch, J.: A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. Computer methods and programs in biomedicine **164**, 15–22 (2018)

[11] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)

[12] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017)

[13] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[14] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[15] Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)

[16] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)

[17] Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T.: Deep learning in medical ultrasound analysis: a review. Engineering

5(2), 261–275 (2019)

[18] Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J.: Ai in health and medicine. Nature medicine **28**(1), 31–38 (2022)

[19] Zhou, S.K., Greenspan, H., Davatzikos, C., Duncan, J.S., Van Ginneken, B., Madabhushi, A., Prince, J.L., Rueckert, D., Summers, R.M.: A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE **109**(5), 820–838 (2021)

[20] Rahaman, M.M., Li, C., Wu, X., Yao, Y., Hu, Z., Jiang, T., Li, X., Qi, S.: A survey for cervical cytopathology image analysis using deep learning. IEEE Access **8**, 61687–61710 (2020)

[21] Conceição, T., Braga, C., Rosado, L., Vasconcelos, M.J.M.: A review of computational methods for cervical cells segmentation and abnormality classification. International journal of molecular sciences **20**(20), 5114 (2019)

[22] Chitra, B., Kumar, S.: Recent advancement in cervical cancer diagnosis for automated screening: a detailed review. Journal of Ambient Intelligence and Humanized Computing, 1–19 (2022)

[23] Hou, X., Shen, G., Zhou, L., Li, Y., Wang, T., Ma, X.: Artificial intelligence in cervical cancer screening and diagnosis. Frontiers in oncology **12** (2022)

[24] Shanthi, P., Hareesha, K., Kudva, R.: Automated detection and classification of cervical cancer using pap smear microscopic images: a comprehensive review and future perspectives. Engineered Science **19**, 20–41 (2022)

[25] Waggoner, S.E.: Cervical cancer. The lancet **361**(9376), 2217–2225 (2003)

[26] Sankaranarayanan, R., Budukh, A.M., Rajkumar, R.: Effective screening programmes for cervical cancer in low-and middle-income developing countries. Bulletin of the World Health Organization **79**(10), 954–962 (2001)

[27] Siebers, A.G., Klinkhamer, P.J., Grefte, J.M., Massuger, L.F., Vedder, J.E., Beijers-Broos, A., Bulten, J., Arbyn, M.: Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: a randomized controlled trial. Jama **302**(16), 1757–1764 (2009)

[28] Al-Janabi, S., Huisman, A., Van Diest, P.J.: Digital pathology: current status and future perspectives. Histopathology **61**(1), 1–9 (2012)

[29] Niazi, M.K.K., Parwani, A.V., Gurcan, M.N.: Digital pathology and artificial intelligence. The lancet oncology **20**(5), 253–261 (2019)

[30] St Clair, C., Wright, J.: Cervical intraepithelial neoplasia: history and detection. Glob Libr Women's Med (ISSN: 1756-2228). doi **10**, 3843 (2009)

[31] Traut, H.F., Papanicolaou, G.N.: Cancer of the uterus: the vaginal smear in its diagnosis. California and western medicine **59**(2), 121 (1943)

[32] Reagan, J.W., Seidemann, I.L., Saracusa, Y.: The cellular morphology of carcinoma in situ and dysplasia or atypical hyperplasia of the uterine cervix. Cancer **6**(2), 224–235 (1953)

[33] RICHART, R.M.: Natural history of cervical intraepithelial neoplasia. Clinical Obstetrics and Gynecology **10**(4), 748–784 (1967)

[34] Hausen, H.z.: Human papillomaviruses and their possible role in squamous cell carcinomas. Current topics in microbiology and immunology, 1–30 (1977)

[35] Crum, C.P., Mitao, M., Levine, R.U., Silverstein, S.: Cervical papillomaviruses segregate within morphologically distinct precancerous lesions. Journal of Virology **54**(3), 675–681 (1985)

[36] Workshop, N.C.I.: The 1988 bethesda system for reporting cervical/vaginal cytological diagnoses. JAMA **262**(7), 931–934 (1989)

[37] Nayar, R., Wilbur, D.C.: The Bethesda System for Reporting Cervical Cytology: Definitions, Criteria, and Explanatory Notes. Springer, ??? (2015)

[38] Herbert, A., Bergeron, C., Wiener, H., Schenck, U., Klinkhamer, P., Bulten, J., Arbyn, M.: European guidelines for quality assurance in cervical cancer screening: recommendations for cervical cytology terminology. Cytopathology **18**(4), 213–219 (2007)

[39] Kedra, B., Chomczyk, M., Zlotkowski, M., Stokowska, W., Borsuk, A., Bicz, M., Pietruska, M., Tokajuk, G., Charkiewicz, R., Czajka, P., *et al.*: Cytological picture of the oral mucosa in patients with gastric and colon cancer. Folia Histochemica et Cytobiologica **50**(3), 375–380 (2012)

[40] Young, R.: Who classification of tumours of female reproductive organs.

Kurman RJ Carcangiu ML Herrington CS Young RH Monodermal teratomas and somatic-type tumours arising from a dermoid cyst, 63–66 (2014)

[41] Jantzen, J., Norup, J., Dounias, G., Bjerregaard, B.: Pap-smear benchmark data for pattern classification. Nature inspired Smart Information Systems (NiSIS 2005), 1–9 (2005)

[42] Lu, Z., Carneiro, G., Bradley, A.P.: An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. IEEE Transactions on Image Processing **24**(4), 1261–1272 (2015)

[43] Lu, Z., Carneiro, G., Bradley, A.P., Ushizima, D., Nosrati, M.S., Bianchi, A.G., Carneiro, C.M., Hamarneh, G.: Evaluation of three algorithms for the segmentation of overlapping cervical cells. IEEE journal of biomedical and health informatics **21**(2), 441–450 (2016)

[44] Plissiti, M.E., Dimitrakopoulos, P., Sfikas, G., Nikou, C., Krikoni, O., Charchanti, A.: Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 3144–3148 (2018). IEEE

[45] Phoulady, H.A., Mouton, P.R.: A new cervical cytology dataset for nucleus detection and image classification (cervix93) and methods for cervical nucleus detection. arXiv preprint arXiv:1811.09651 (2018)

[46] Araújo, F.H., Silva, R.R., Ushizima, D.M., Rezende, M.T., Carneiro, C.M., Bianchi, A.G.C., Medeiros, F.N.: Deep learning for cell image segmentation and ranking. Computerized Medical Imaging and Graphics **72**, 13–21 (2019)

[47] Zhang, J., Liu, Z., Du, B., He, J., Li, G., Chen, D.: Binary tree-like network with two-path fusion attention feature for cervical cell nucleus segmentation. Computers in biology and medicine **108**, 223–233 (2019)

[48] Hussain, E., Mahanta, L.B., Borah, H., Das, C.R.: Liquid based-cytology pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions. Data in brief **30**, 105589 (2020)

[49] Rezende, M.T., Silva, R., Bernardo, F.d.O., Tobias, A.H., Oliveira, P.H., Machado, T.M., Costa, C.S., Medeiros, F.N., Ushizima, D.M., Carneiro, C.M., *et al.*: Cric searchable image database as a public platform for conventional pap smear cytology data. Scientific data **8**(1), 151 (2021)

[50] Liang, Y., Tang, Z., Yan, M., Chen, J., Liu, Q., Xiang, Y.: Comparison detector for cervical cell/clumps detection in the limited data scenario.

Neurocomputing **437**, 195–205 (2021)

[51] Riana, D., Hadianti, S., Rahayu, S., Hasan, M., Karimah, I.N., Pratama, R.: Repomedunm: A new dataset for feature extraction and training of deep learning network for classification of pap smear images. In: Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part V, pp. 317–325 (2021). Springer

[52] Liu, J., Fan, H., Wang, Q., Li, W., Tang, Y., Wang, D., Zhou, M., Chen, L.: Local label point correction for edge detection of overlapping cervical cells. Frontiers in Neuroinformatics **16** (2022)

[53] Liu, G., Ding, Q., Luo, H., Sha, M., Li, X., Ju, M.: Cx22: A new publicly available dataset for deep learning-based segmentation of cervical cytology images. Computers in Biology and Medicine **150**, 106194 (2022)

[54] Shanthi, P., Faruqi, F., Hareesha, K., Kudva, R.: Deep convolution neural network for malignancy detection and classification in microscopic uterine cervix cell images. Asian Pacific journal of cancer prevention: APJCP **20**(11), 3447 (2019)

[55] Chen, H., Liu, J., Wen, Q.-M., Zuo, Z.-Q., Liu, J.-S., Feng, J., Pang, B.-C., Xiao, D.: Cytobrain: cervical cancer screening system based on deep learning technology. Journal of Computer Science and Technology **36**(2), 347–360 (2021)

[56] Khamparia, A., Gupta, D., Rodrigues, J.J., de Albuquerque, V.H.C.: Dcavn: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network. Multimedia Tools and Applications **80**(20), 30399–30415 (2021)

[57] Zhang, L., Lu, L., Nogues, I., Summers, R.M., Liu, S., Yao, J.: Deeppap: deep convolutional networks for cervical cell classification. IEEE journal of biomedical and health informatics **21**(6), 1633–1643 (2017)

[58] Hyeon, J., Choi, H.-J., Lee, B.D., Lee, K.N.: Diagnosing cervical cell images using pre-trained convolutional neural network as feature extractor. In: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 390–393 (2017). IEEE

[59] Ghoneim, A., Muhammad, G., Hossain, M.S.: Cervical cancer classification using convolutional neural networks and extreme learning machines. Future Generation Computer Systems **102**, 643–649 (2020)

[60] Khamparia, A., Gupta, D., de Albuquerque, V.H.C., Sangaiah, A.K.,

Jhaveri, R.H.: Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. The Journal of Supercomputing **76**(11), 8590–8608 (2020)

[61] Wang, P., Wang, J., Li, Y., Li, L., Zhang, H.: Adaptive pruning of transfer learned deep convolutional neural network for classification of cervical pap smear images. IEEE Access **8**, 50674–50683 (2020)

[62] Bhatt, A.R., Ganatra, A., Kotecha, K.: Cervical cancer detection in pap smear whole slide images using convnet with transfer learning and progressive resizing. PeerJ Computer Science **7**, 348 (2021)

[63] Rahaman, M.M., Li, C., Yao, Y., Kulwa, F., Wu, X., Li, X., Wang, Q.: Deepcervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. Computers in Biology and Medicine **136**, 104649 (2021)

[64] Manna, A., Kundu, R., Kaplun, D., Sinitca, A., Sarkar, R.: A fuzzy rank-based ensemble of cnn models for classification of cervical cytology. Scientific Reports **11**(1), 1–18 (2021)

[65] N. Diniz, D., T. Rezende, M., GC Bianchi, A., M. Carneiro, C., JS Luz, E., JP Moreira, G., M. Ushizima, D., NS de Medeiros, F., JF Souza, M.: A deep learning ensemble method to assist cytopathologists in pap test image classification. Journal of Imaging **7**(7), 111 (2021)

[66] Liu, W., Li, C., Xu, N., Jiang, T., Rahaman, M.M., Sun, H., Wu, X., Hu, W., Chen, H., Sun, C., et al.: Cvm-cervix: A hybrid cervical pap-smear image classification framework using cnn, visual transformer and multilayer perceptron. Pattern Recognition, 108829 (2022)

[67] Kundu, R., Chattopadhyay, S.: Deep features selection through genetic algorithm for cervical pre-cancerous cell classification. Multimedia Tools and Applications, 1–22 (2022)

[68] Jia, A.D., Li, B.Z., Zhang, C.C.: Detection of cervical cancer cells based on strong feature cnn-svm network. Neurocomputing **411**, 112–127 (2020)

[69] Dong, N., Zhao, L., Wu, C.-H., Chang, J.-F.: Inception v3 based cervical cell classification combined with artificially extracted features. Applied Soft Computing **93**, 106311 (2020)

[70] Zhang, C., Jia, D., Li, Z., Wu, N.: Auxiliary classification of cervical cells based on multi-domain hybrid deep learning framework. Biomedical Signal Processing and Control **77**, 103739 (2022)

[71] Yaman, O., Tuncer, T.: Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images. Biomedical Signal Processing and Control **73**, 103428 (2022)

[72] Qin, J., He, Y., Ge, J., Liang, Y.: A multi-task feature fusion model for cervical cell classification. IEEE Journal of Biomedical and Health Informatics **26**(9), 4668–4678 (2022)

[73] Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on knowledge and data engineering **22**(10), 1345–1359 (2009)

[74] Yu, X., Wang, J., Hong, Q.-Q., Teku, R., Wang, S.-H., Zhang, Y.-D.: Transfer learning for medical images analyses: A survey. Neurocomputing **489**, 230–254 (2022)

[75] Yang, Y., Lv, H., Chen, N.: A survey on ensemble learning under the era of deep learning. Artificial Intelligence Review, 1–45 (2022)

[76] Cao, Y., Geddes, T.A., Yang, J.Y.H., Yang, P.: Ensemble deep learning in bioinformatics. Nature Machine Intelligence **2**(9), 500–508 (2020)

[77] Ganaie, M.A., Hu, M., Malik, A., Tanveer, M., Suganthan, P.: Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence **115**, 105151 (2022)

[78] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: Ce-net: Context encoder network for 2d medical image segmentation. IEEE transactions on medical imaging **38**(10), 2281–2292 (2019)

[79] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)

[80] Zou, Z., Shi, Z., Guo, Y., Ye, J.: Object detection in 20 years: A survey. arXiv preprint arXiv:1905.05055 (2019)

[81] Wu, X., Sahoo, D., Hoi, S.C.: Recent advances in deep learning for object detection. Neurocomputing **396**, 39–64 (2020)

[82] Zhao, Z.-Q., Zheng, P., Xu, S.-t., Wu, X.: Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems **30**(11), 3212–3232 (2019)

[83] Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. International journal of computer

vision **104**, 154–171 (2013)

[84] Zitnick, C.L., Dollár, P.: Edge boxes: Locating object proposals from edges. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 391–405 (2014). Springer

[85] Dai, J., Li, Y., He, K., Sun, J.: R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processing systems **29** (2016)

[86] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)

[87] Cai, Z., Vasconcelos, N.: Cascade r-cnn: high quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence **43**(5), 1483–1498 (2019)

[88] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)

[89] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)

[90] Zhang, S., Wen, L., Bian, X., Lei, Z., Li, S.Z.: Single-shot refinement neural network for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4203–4212 (2018)

[91] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: Ssd: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp. 21–37 (2016). Springer

[92] Xiang, Y., Sun, W., Pan, C., Yan, M., Yin, Z., Liang, Y.: A novel automation-assisted cervical cancer reading method based on convolutional neural network. Biocybernetics and Biomedical Engineering **40**(2), 611–623 (2020)

[93] Nambu, Y., Mariya, T., Shinkai, S., Umemoto, M., Asanuma, H., Sato, I., Hirohashi, Y., Torigoe, T., Fujino, Y., Saito, T.: A screening assistance system for cervical cytology of squamous cell atypia based on a two-step combined cnn algorithm with label smoothing. Cancer Medicine **11**(2),

520–529 (2022)

[94] Liang, Y., Pan, C., Sun, W., Liu, Q., Du, Y.: Global context-aware cervical cell detection with soft scale anchor matching. Computer Methods and Programs in Biomedicine **204**, 106061 (2021)

[95] Jia, D., Zhou, J., Zhang, C.: Detection of cervical cells based on improved ssd network. Multimedia Tools and Applications **81**(10), 13371–13387 (2022)

[96] Jia, D., He, Z., Zhang, C., Yin, W., Wu, N., Li, Z.: Detection of cervical cancer cells in complex situation based on improved yolov3 network. Multimedia Tools and Applications **81**(6), 8939–8961 (2022)

[97] Sompawong, N., Mopan, J., Pooprasert, P., Himakhun, W., Suwannarurk, K., Ngamvirojcharoen, J., Vachiramon, T., Tantibundhit, C.: Automated pap smear cervical cancer screening using deep learning. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7044–7048 (2019). IEEE

[98] Zhang, J., He, J., Chen, T., Liu, Z., Chen, D.: Abnormal region detection in cervical smear images based on fully convolutional network. IET Image Processing **13**(4), 583–590 (2019)

[99] Li, X., Xu, Z., Shen, X., Zhou, Y., Xiao, B., Li, T.-Q.: Detection of cervical cancer cells in whole slide images using deformable and global context aware faster rcnn-fpn. Current Oncology **28**(5), 3585–3601 (2021)

[100] Yan, X., Zhang, Z.: Hsdet: A representative sampling based object detector in cervical cancer cell images. In: Bio-Inspired Computing: Theories and Applications: 15th International Conference, BIC-TA 2020, Qingdao, China, October 23-25, 2020, Revised Selected Papers 15, pp. 406–418 (2021). Springer

[101] Wang, W., Tian, Y., Xu, Y., Zhang, X.-X., Li, Y.-S., Zhao, S.-F., Bai, Y.-H.: 3cde-net: a cervical cancer cell detection network based on an improved backbone network and multiscale feature fusion. BMC Medical Imaging **22**(1), 1–13 (2022)

[102] Xu, C., Li, M., Li, G., Zhang, Y., Sun, C., Bai, N.: Cervical cell/clumps detection in cytology images using transfer learning. Diagnostics **12**(10), 2477 (2022)

[103] Liang, Y., Feng, S., Liu, Q., Kuang, H., Liu, J., Liao, L., Du, Y., Wang, J.: Exploring contextual relationships for cervical abnormal cell detection. arXiv preprint arXiv:2207.04693 (2022)

[104] Liu, M., Li, X., Gao, X., Chen, J., Shen, L., Wu, H.: Sample hardness based gradient loss for long-tailed cervical cell detection. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II, pp. 109–119 (2022). Springer

[105] Chen, T., Zheng, W., Ying, H., Tan, X., Li, K., Li, X., Chen, D.Z., Wu, J.: A task decomposing and cell comparing method for cervical lesion cell detection. IEEE Transactions on Medical Imaging (2022)

[106] Zhang, M., Li, X., Shen, L.: Classification and localization consistency regularized student-teacher network for semi-supervised cervical cell detection. In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), pp. 289–294 (2021). IEEE

[107] Du, X., Huo, J., Qiao, Y., Wang, Q., Zhang, L.: False positive suppression in cervical cell screening via attention-guided semi-supervised learning. In: Predictive Intelligence in Medicine: 4th International Workshop, PRIME 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 4, pp. 93–103 (2021). Springer

[108] Chai, Z., Luo, L., Lin, H., Chen, H., Han, A., Heng, P.-A.: Deep semi-supervised metric learning with dual alignment for cervical cancer cell detection. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–5 (2022). IEEE

[109] Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5693–5703 (2019)

[110] Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6154–6162 (2018)

[111] Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine learning **109**(2), 373–440 (2020)

[112] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D.: Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(7), 3523–3542 (2021)

[113] Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging **32**, 582–596 (2019)

[114] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)

[115] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer

[116] Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence **39**(12), 2481–2495 (2017)

[117] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)

[118] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)

[119] Zhang, L., Sonka, M., Lu, L., Summers, R.M., Yao, J.: Combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 406–409 (2017). IEEE

[120] Gautam, S., Bhavsar, A., Sao, A.K., Harinarayan, K.: Cnn based segmentation of nuclei in pap-smear images with selective pre-processing. In: Medical Imaging 2018: Digital Pathology, vol. 10581, pp. 246–254 (2018). SPIE

[121] Liu, Y., Zhang, P., Song, Q., Li, A., Zhang, P., Gui, Z.: Automatic segmentation of cervical nuclei based on deep learning and a conditional random field. IEEE Access **6**, 53709–53721 (2018)

[122] Zhao, J., Li, Q., Li, X., Li, H., Zhang, L.: Automated segmentation of cervical nuclei in pap smear images using deformable multi-path ensemble model. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1514–1518 (2019). IEEE

[123] Zhao, J., Dai, L., Zhang, M., Yu, F., Li, M., Li, H., Wang, W., Zhang, L.: Pgu-net+: progressive growing of u-net+ for automated cervical nuclei

segmentation. In: Multiscale Multimodal Medical Imaging: First International Workshop, MMMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 1, pp. 51–58 (2020). Springer

[124] Hussain, E., Mahanta, L.B., Das, C.R., Choudhury, M., Chowdhury, M.: A shape context fully convolutional neural network for segmentation and classification of cervical nuclei in pap smear images. Artificial Intelligence in Medicine **107**, 101897 (2020)

[125] Zhao, Y., Fu, C., Xu, S., Cao, L., Ma, H.-f.: Lfanet: Lightweight feature attention network for abnormal cell segmentation in cervical cytology images. Computers in Biology and Medicine **145**, 105500 (2022)

[126] Luo, D., Kang, H., Long, J., Zhang, J., Chen, L., Quan, T., Liu, X.: Dual supervised sampling networks for real-time segmentation of cervical cell nucleus. Computational and Structural Biotechnology Journal **20**, 4360–4368 (2022)

[127] Song, Y., Tan, E.-L., Jiang, X., Cheng, J.-Z., Ni, D., Chen, S., Lei, B., Wang, T.: Accurate cervical cell segmentation from overlapping clumps in pap smear images. IEEE transactions on medical imaging **36**(1), 288–300 (2016)

[128] Tareef, A., Song, Y., Huang, H., Wang, Y., Feng, D., Chen, M., Cai, W.: Optimizing the cervix cytological examination based on deep learning and dynamic shape modeling. Neurocomputing **248**, 28–40 (2017)

[129] Xu, S., Sang, C., Jin, Y., Wan, T.: Robust segmentation of overlapping cells in cervical cytology using light convolution neural network. In: Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part VII 25, pp. 387–397 (2018). Springer

[130] Wan, T., Xu, S., Sang, C., Jin, Y., Qin, Z.: Accurate segmentation of overlapping cells in cervical cytology with deep convolutional neural networks. Neurocomputing **365**, 157–170 (2019)

[131] Zhou, Y., Chen, H., Xu, J., Dou, Q., Heng, P.-A.: Irnet: Instance relation network for overlapping cervical cell segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22, pp. 640–648 (2019). Springer

[132] Zhang, H., Zhu, H., Ling, X.: Polar coordinate sampling-based segmentation of overlapping cervical cells using attention u-net and random walk. Neurocomputing **383**, 212–223 (2020)

[133] Zhou, Y., Chen, H., Lin, H., Heng, P.-A.: Deep semi-supervised knowledge distillation for overlapping cervical cell instance segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23, pp. 521–531 (2020). Springer

[134] Mahyari, T.L., Dansereau, R.M.: Multi-layer random walker image segmentation for overlapped cervical cells using probabilistic deep learning methods. IET Image Processing **16**(11), 2959–2972 (2022)

[135] Tellez, D., Litjens, G., van der Laak, J., Ciompi, F.: Neural image compression for gigapixel histopathology image analysis. IEEE transactions on pattern analysis and machine intelligence **43**(2), 567–578 (2019)

[136] Dimitriou, N., Arandjelović, O., Caie, P.D.: Deep learning for whole slide image analysis: an overview. Frontiers in medicine **6**, 264 (2019)

[137] Carbonneau, M.-A., Cheplygina, V., Granger, E., Gagnon, G.: Multiple instance learning: A survey of problem characteristics and applications. Pattern Recognition **77**, 329–353 (2018)

[138] Xiang, T., Song, Y., Zhang, C., Liu, D., Chen, M., Zhang, F., Huang, H., O'Donnell, L., Cai, W.: Dsnet: A dual-stream framework for weakly-supervised gigapixel pathology image analysis. IEEE Transactions on Medical Imaging **41**(8), 2180–2190 (2022)

[139] Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18802–18812 (2022)

[140] Lin, H., Chen, H., Wang, X., Wang, Q., Wang, L., Heng, P.-A.: Dual-path network with synergistic grouping loss and evidence driven risk stratification for whole slide cervical image analysis. Medical Image Analysis **69**, 101955 (2021)

[141] Chen, P., Liang, Y., Shi, X., Yang, L., Gader, P.: Automatic whole slide pathology image diagnosis framework via unit stochastic selection and attention fusion. Neurocomputing **453**, 312–325 (2021)

[142] Zhou, M., Zhang, L., Du, X., Ouyang, X., Zhang, X., Shen, Q., Luo, D., Fan, X., Wang, Q.: Hierarchical pathology screening for cervical abnormality. Computerized Medical Imaging and Graphics **89**, 101892 (2021)

[143] Zhu, X., Li, X., Ong, K., Zhang, W., Li, W., Li, L., Young, D., Su, Y.,

Shang, B., Peng, L., *et al.*: Hybrid ai-assistive diagnostic model permits rapid tbs classification of cervical liquid-based thin-layer cell smears. Nature communications **12**(1), 3541 (2021)

[144] Cao, L., Yang, J., Rong, Z., Li, L., Xia, B., You, C., Lou, G., Jiang, L., Du, C., Meng, H., *et al.*: A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening. Medical image analysis **73**, 102197 (2021)

[145] Cheng, S., Liu, S., Yu, J., Rao, G., Xiao, Y., Han, W., Zhu, W., Lv, X., Li, N., Cai, J., *et al.*: Robust whole slide image analysis for cervical cancer screening using deep learning. Nature communications **12**(1), 5639 (2021)

[146] Wei, Z., Cheng, S., Liu, X., Zeng, S.: An efficient cervical whole slide image analysis framework based on multi-scale semantic and spatial deep features. arXiv preprint arXiv:2106.15113 (2021)

[147] Pirovano, A., Almeida, L.G., Ladjal, S., Bloch, I., Berlemont, S.: Computer-aided diagnosis tool for cervical cancer screening with weakly supervised localization and detection of abnormalities using adaptable and explainable classifier. Medical image analysis **73**, 102167 (2021)

[148] Kanavati, F., Hirose, N., Ishii, T., Fukuda, A., Ichihara, S., Tsuneki, M.: A deep learning model for cervical cancer screening on liquid-based cytology specimens in whole slide images. Cancers **14**(5), 1159 (2022)

[149] Geng, R., Liu, Q., Feng, S., Liang, Y.: Learning deep pathological features for wsi-level cervical cancer grading. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1391–1395 (2022). IEEE

[150] Zhang, X., Cao, M., Wang, S., Sun, J., Fan, X., Wang, Q., Zhang, L.: Whole slide cervical cancer screening using graph attention network and supervised contrastive learning. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II, pp. 202–211 (2022). Springer

[151] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019). PMLR

[152] Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9627–9636 (2019)

[153] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)

[154] Chen, X., Yu, J., Cheng, S., Geng, X., Liu, S., Han, W., Hu, J., Chen, L., Liu, X., Zeng, S.: An unsupervised style normalization method for cytopathology images. Computational and Structural Biotechnology Journal **19**, 3852–3863 (2021)

[155] Kang, H., Luo, D., Feng, W., Zeng, S., Quan, T., Hu, J., Liu, X.: Stainnet: a fast and robust stain normalization network. Frontiers in Medicine **8**, 746307 (2021)

[156] Shaban, M.T., Baur, C., Navab, N., Albarqouni, S.: Staingan: Stain style transfer for digital histological images. In: 2019 Ieee 16th International Symposium on Biomedical Imaging (Isbi 2019), pp. 953–956 (2019). IEEE

[157] Ma, J., Yu, J., Liu, S., Chen, L., Li, X., Feng, J., Chen, Z., Zeng, S., Liu, X., Cheng, S.: Pathsrgan: Multi-supervised super-resolution for cytopathological images using generative adversarial network. IEEE transactions on medical imaging **39**(9), 2920–2930 (2020)

[158] Ma, J., Liu, S., Cheng, S., Chen, R., Liu, X., Chen, L., Zeng, S.: Stsrnet: Self-texture transfer super-resolution and refocusing network. IEEE Transactions on Medical Imaging **41**(2), 383–393 (2021)

[159] Jiang, H., Zhou, Y., Lin, Y., Chan, R.C., Liu, J., Chen, H.: Deep learning for computational cytology: A survey. Medical Image Analysis, 102691 (2022)

[160] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)

[161] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500 (2017)

[162] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)

[163] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on

Computer Vision (ECCV), pp. 3–19 (2018)

[164] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)

[165] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[166] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[167] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

[168] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H.: Training data-efficient image transformers & distillation through attention. In: International Conference on Machine Learning, pp. 10347–10357 (2021). PMLR

[169] Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., Yu, S.: A survey on incorporating domain knowledge into deep learning for medical image analysis. Medical Image Analysis **69**, 101985 (2021)

[170] Cheplygina, V., de Bruijne, M., Pluim, J.P.: Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical image analysis **54**, 280–296 (2019)

[171] Wang, L., Guo, D., Wang, G., Zhang, S.: Annotation-efficient learning for medical image segmentation based on noisy pseudo labels and adversarial learning. IEEE Transactions on Medical Imaging **40**(10), 2795–2807 (2020)

[172] Hu, X., Zeng, D., Xu, X., Shi, Y.: Semi-supervised contrastive learning for label-efficient medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24, pp. 481–490 (2021). Springer

[173] Cui, C., Yang, H., Wang, Y., Zhao, S., Asad, Z., Coburn, L.A., Wilson, K.T., Landman, B.A., Huo, Y.: Deep multi-modal fusion of image

and non-image data in disease diagnosis and prognosis: a review. arXiv preprint arXiv:2203.15588 (2022)

[174] Arya, N., Saha, S.: Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. IEEE/ACM transactions on computational biology and bioinformatics **19**(2), 1032–1041 (2020)

[175] Kakhi, K., Alizadehsani, R., Kabir, H.D., Khosravi, A., Nahavandi, S., Acharya, U.R.: The internet of medical things and artificial intelligence: trends, challenges, and opportunities. Biocybernetics and Biomedical Engineering (2022)

[176] Liu, L., Xu, J., Huan, Y., Zou, Z., Yeh, S.-C., Zheng, L.-R.: A smart dental health-iot platform based on intelligent hardware, deep learning, and mobile terminal. IEEE journal of biomedical and health informatics **24**(3), 898–906 (2019)

[177] Guo, Z., Shen, Y., Wan, S., Shang, W.-L., Yu, K.: Hybrid intelligence-driven medical image recognition for remote patient diagnosis in internet of medical things. IEEE journal of biomedical and health informatics **26**(12), 5817–5828 (2021)

[178] Huang, Y.-N., Peng, X.-C., Ma, S., Yu, H., Jin, Y.-B., Zheng, J., Fu, G.-H., *et al.*: Development of whole slide imaging on smartphones and evaluation with thinprep cytology test samples: Follow-up study. JMIR mHealth and uHealth **6**(4), 9518 (2018)

[179] Tang, H.-P., Cai, D., Kong, Y.-Q., Ye, H., Ma, Z.-X., Lv, H.-S., Tuo, L.-R., Pan, Q.-J., Liu, Z.-H., Han, X.: Cervical cytology screening facilitated by an artificial intelligence microscope: a preliminary study. Cancer cytopathology **129**(9), 693–700 (2021)

[180] Jiang, P., Liu, J., Xiao, D., Pang, B., Hao, Z., Cao, D.: A novel iomt system for pathological diagnosis based on intelligent mobile scanner and whole slide image stitching method. In: Intelligent Computing Methodologies: 18th International Conference, ICIC 2022, Xi'an, China, August 7–11, 2022, Proceedings, Part III, pp. 463–472 (2022). Springer

[181] Jiang, P., Liu, J., Luo, Q., Pang, B., Xiao, D., Cao, D.: Development of automatic portable pathology scanner and its evaluation for clinical practice. Journal of Digital Imaging, 1–13 (2023)

[182] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., *et al.*: The future of digital health with federated learning. NPJ digital medicine **3**(1), 119 (2020)

[183] Samuel, O., Omojo, A.B., Onuja, A.M., Sunday, Y., Tiwari, P., Gupta, D., Hafeez, G., Yahaya, A.S., Fatoba, O.J., Shamshirband, S.: Iomt: a covid-19 healthcare system driven by federated learning and blockchain. IEEE Journal of Biomedical and Health Informatics **27**(2), 823–834 (2022)

[184] Hossen, M.N., Panneerselvam, V., Koundal, D., Ahmed, K., Bui, F.M., Ibrahim, S.M.: Federated machine learning for detection of skin diseases and enhancement of internet of medical things (iomt) security. IEEE journal of biomedical and health informatics (2022)