

Article

A Deep Learning Model for Cervical Cancer Screening on Liquid-Based Cytology Specimens in Whole Slide Images

Fahdi Kanavati ^{1,†}, Naoki Hirose ², Takahiro Ishii ², Ayaka Fukuda ², Shin Ichihara ³  and Masayuki Tsuneki ^{1,*,†} ¹ Medmain Research, Medmain Inc., Fukuoka 810-0042, Fukuoka, Japan; fkanavati@medmain.com² Department of Clinical Laboratory, Sapporo Kosei General Hospital, 8-5 Kita-3-jo Higashi, Chuo-ku, Sapporo 060-0033, Hokkaido, Japan; okh@hotmail.co.jp (N.H.); 141takahiro8260@gmail.com (T.I.); ayaka3843@gmail.com (A.F.)³ Department of Surgical Pathology, Sapporo Kosei General Hospital, 8-5 Kita-3-jo Higashi, Chuo-ku, Sapporo 060-0033, Hokkaido, Japan; trampolineboy@mbn.nifty.com

* Correspondence: tsuneki@medmain.com; Tel.: +81-92-707-1977

† These authors contribute equally to this work.

Simple Summary: In this pilot study, we aimed to investigate the use of deep learning for the classification of whole-slide images of liquid-based cytology specimens into neoplastic and non-neoplastic. To do so, we used a large training and test sets. Overall, the model achieved good classification performance in classifying whole-slide images, demonstrating the promising potential use of such models for aiding the screening processes for cervical cancer.

Abstract: Liquid-based cytology (LBC) for cervical cancer screening is now more common than the conventional smears, which when digitised from glass slides into whole-slide images (WSIs), opens up the possibility of artificial intelligence (AI)-based automated image analysis. Since conventional screening processes by cytoscreeners and cytopathologists using microscopes is limited in terms of human resources, it is important to develop new computational techniques that can automatically and rapidly diagnose a large amount of specimens without delay, which would be of great benefit for clinical laboratories and hospitals. The goal of this study was to investigate the use of a deep learning model for the classification of WSIs of LBC specimens into neoplastic and non-neoplastic. To do so, we used a dataset of 1605 cervical WSIs. We evaluated the model on three test sets with a combined total of 1468 WSIs, achieving ROC AUCs for WSI diagnosis in the range of 0.89–0.96, demonstrating the promising potential use of such models for aiding screening processes.

Keywords: liquid-based cytology; deep learning; cervical screening; whole slide image



Citation: 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* **2022**, *14*, 1159. <https://doi.org/10.3390/cancers14051159>

Academic Editor: Samuel C. Mok

Received: 24 January 2022

Accepted: 22 February 2022

Published: 24 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the Global Cancer Statistics 2020 [1], cervical cancer is the fourth leading cause of cancer death in women, with an estimated 342,000 deaths worldwide in 2020. However, incidence and mortality rates have declined over the past few decades due to either increasing average socioeconomic levels or a diminishing risk of persistent infection with high risk human papillomavirus (HPV) [1]. In developed countries, cervical cytology screening systems have been organised to reduce mortality from cervical cancer [2–9].

The introduction of cervical cancer screening led to a fall in associated mortality rates; however, there is some evidence that the conventional smear method for screening is not consistent in reliably detecting cervical intraepithelial neoplasia (CIN) [10–12]. This is because conventional cervical smears, when spread on glass slides, tend to have the cells of interest mixed with blood, debris, and exudate. A number of new technologies and procedures are becoming available in various screening programs (e.g., liquid-based cytology (LBC), automated screening devices, computer-assisted microscopy, digital colposcopy with automated image analysis, HPV testing). The LBC technique preserves the cells of interest in a liquid medium and removes most of the debris, blood, and exudate

either by filtering or density gradient centrifugation. The other advantages in LBC are the availability of residual material for HPV and other molecular tests and the connection with automated screening devices. ThinPrep (Hologic, Inc., Marlborough, MA, USA) and SurePath (Becton Dickinson, Inc., Franklin Lakes, NJ, USA) for LBC specimen preparation have been approved by the US Food and Drug Administration (FDA), and it has also been adopted by the cervical screening programme in the UK. Moreover, the ThinPrep collection vial has been approved by the FDA for direct testing for HPV, which is particularly useful for managing women whose Pap smear tests show atypical squamous cells (ASCs) [4,13].

In 1998, the FDA approved the FocalPoint Slide Profiler (Becton Dickinson, Inc.) as a primary automated screener for cervical smears, followed by approval in 2002 for use with SurePath slides. In 2003, the FDA approved the ThinPrep Imaging System (Hologic, Inc.) as a primary screener for ThinPrep Pap slides. The FocalPoint uses algorithms to measure cellular features (e.g., nuclear size, integrated optical density, nuclear to cytoplasmic ratio, and nuclear contour) for the diagnosis of squamous and glandular lesions [14]. In the US, the American Society of Cytopathology (ASC) established guidelines for automated Pap test screening using the ThinPrep Imaging System and the FocalPoint GS Imaging System [15]. However, there are some issues with the current automated screening support systems. A multi-institutional feasibility study in Japan validated the usefulness of FocalPoint for cervical cytology automated screening quality control and showed that it was useful for NILM (Negative for Intraepithelial Lesion or Malignancy) cases, but on the other hand, 2174 (18.1%) of 12,000 specimens were judged to be unmeasurable and were not evaluated [16]. In the US, unmeasured rates were reported to be as low as 2.5% [17], 5.9% [18], and 4.8% [19], while in Brazil, the unmeasured rate was very high at 30.8% [20]. In order to use FocalPoint, it was reported that the unmeasured ratio can be suppressed to a low value by adjusting a specimen preparation method(s) including staining [16]. However, in routine clinical practice, there are many screening facilities that do not (or cannot) stain specimens accordingly to adjust for FocalPoint, as reported in Japan and Brazil [16,20].

The sensitivity of conventional cytology cervical cancer screening for detecting pre-invasive squamous and glandular lesions (pre-invasive intraepithelial lesions) is clearly far from perfect. It has been reported that most studies of the conventional Pap test were severely biased, and it was only moderately accurate and did not achieve concurrently high sensitivity and specificity (i.e., sensitivity ranged from 30% to 87% and specificity ranged from 86% to 100%) [21]. Moreover, the sensitivity of conventional cervical cytology is less than ideal for invasive cancers, with a wide range (45% to 76%), and false-negative or false-unsatisfactory rate in conventional smears was 50% [22]. These studies indicate that many women with cervical cancer have a history of one or more negative cervical cytology reports. As a background of these results, the interobserver reproducibility of cervical cytology is less than perfect. The reproducibility of 4948 monolayer cytologic interpretations was moderate ($\kappa = 0.46$; 95% confidence interval (CI), 0.44–0.48) among four categories of diagnosis (i.e., negative, ASC-US, LSIL, and over HSIL) by multiple well-trained observers [23]. In the same study, the greatest disagreement in monolayer cervical cytology involved ASC-US interpretations. Of the 1473 original interpretations of ASC-US, the second reviewer concurred in only 43.0% [23].

Whole-slide images (WSIs) are digitisations of the conventional glass slides obtained via specialised scanning devices (WSI scanners), and they are considered to be comparable to microscopy for primary diagnosis [24]. A routine scanning of LBC slides in a single layer of WSIs would be suitable for further high throughput analysis (e.g., automated image based cytological screening and medical image analysis) [25]. The advent of WSIs led to the application of medical image analysis techniques, machine learning, and deep learning techniques for aiding pathologists in inspecting WSIs. Deep-learning-based applications ranged from tasks, such as cancer diagnosis from WSIs, cell classification, and segmentation of nuclei, to patient stratification and outcome prediction [26–44]. For cytology, in particular, only recently have there been investigations for applying deep learning on large datasets of cervical WSIs Holmström et al. [45], Lin et al. [46], Cheng et al. [47].

In this pilot study, we trained a deep learning model, based on convolutional and recurrent neural networks, using a dataset of 1605 cervical WSIs. We evaluated the model on three test sets with a combined total of 1468 WSIs, achieving ROC AUCs for WSI diagnosis in the range of 0.89–0.96.

2. Materials and Methods

2.1. Clinical Cases and Cytopathological Records

This is a retrospective study. A total of 3121 LBC ThinPrep Pap test (Hologic, Inc.) conventionally prepared cytopathological slide glass specimens of human cervical cytology were collected from a private clinical laboratory in Japan after cytopathological review of those specimens by cytoscreeners and pathologists. The cases were selected mostly at random so as to reflect a real clinical scenario as much as possible; we have also collected cases so as to compile a test set with an equal balance of neoplastic and NILM. The cytoscreeners and pathologists excluded cases that had poor scanned quality ($n = 32$). Each WSI diagnosis was observed by at least two cytoscreeners and pathologists, with the final checking and verification performed by a senior cytoscreener or pathologist. All WSIs were scanned at a magnification of $\times 20$ using the same Aperio AT2 digital whole-slide scanner (Leica Microsystems, Osaka, Japan) and were saved in SVS file format with JPEG2000 compression.

2.2. Dataset

Table 1 breaks down the distribution of the dataset into training, validation, and test sets. The split was carried out randomly taking into account the proportion of each label in the dataset. A clinical laboratory that provided LBC cases was anonymised. The test sets were composed of WSIs of full agreement, clinical balance, and equal balance LBC specimens. The full agreement test set consisted of NILM and neoplastic LBC cases whose obtained diagnoses were fully agreed by two independent cytoscreeners in different institutes. The clinical balance test set consisted of 95% NILM and 5% neoplastic LBC cases based on a real clinical setting [48,49]. The equal balance test set consisted of 50% NILM and 50% neoplastic LBC cases. NILM and neoplastic LBC cases for clinical and equal balance test sets were collected based on the diagnoses provided by the clinical laboratory. The cases in the clinical and equal balance test sets were only based on the diagnostic reports. From these two test sets, we have also created their reviewed counterparts (clinical balance reviewed and equal balance reviewed), where two independent cytoscreeners viewed all the cases and the ones they had a disagreement on were removed (see Table 1).

Table 1. Distribution of WSIs into training, test, and validation sets.

	Total	Neoplastic	NILM
training	1503	302	1201
validation	150	50	100
test: full agreement	300	20	280
test: equal balance	750	375	375
test: equal balance-rev.	643	279	364
test: clinical balance	750	38	712
test: clinical balance-rev.	525	35	490

2.3. Annotation

Senior cytoscreeners and pathologists who perform routine cytopathological screening and diagnoses in general hospitals and clinical laboratories in Japan manually annotated 352 neoplastic WSIs from the training sets. Coarse annotations were obtained by free-hand drawing. (Figure 1 using an in-house online tool developed by customising the open-source OpenSeadragon tool at <https://openseadragon.github.io/> (accessed on 10 January 2020),

which is a web-based viewer for zoomable images.) On average, the cytoscreeners and pathologists annotated 150 cells (or cellular clusters) per WSI.

Neoplastic WSIs consisted of ASC (atypical squamous cell), LSIL (low-grade squamous intraepithelial lesion), HSIL (high-grade squamous intraepithelial lesion), CIS (carcinoma in situ), ADC (adenocarcinoma), and SCC (squamous cell carcinoma), except for the NILM. For example, on the HSIL (Figure 1A–D) and SCC (Figure 1E–H) WSIs, cytoscreeners and pathologists performed annotations around the neoplastic cells (Figure 1B–D,F–H) based on the representative neoplastic epithelial cell morphology (e.g., increased nuclear/cytoplasmic ratio, abnormalities of nuclear shape, hyperchromatism, irregular chromatin distribution, and prominent nucleolus). On the other hand, the cytoscreeners and pathologists did not annotate areas where it was difficult to cytologically determine that the cells were neoplastic. The NILM subset of the training and validation sets (1301 WSIs) was not annotated and the entire cell spreading areas within the WSIs were used.

The average annotation time per WSI was about an hour. Annotations performed by the cytoscreeners and pathologists were modified (if necessary), confirmed, and verified by a senior cytoscreener.

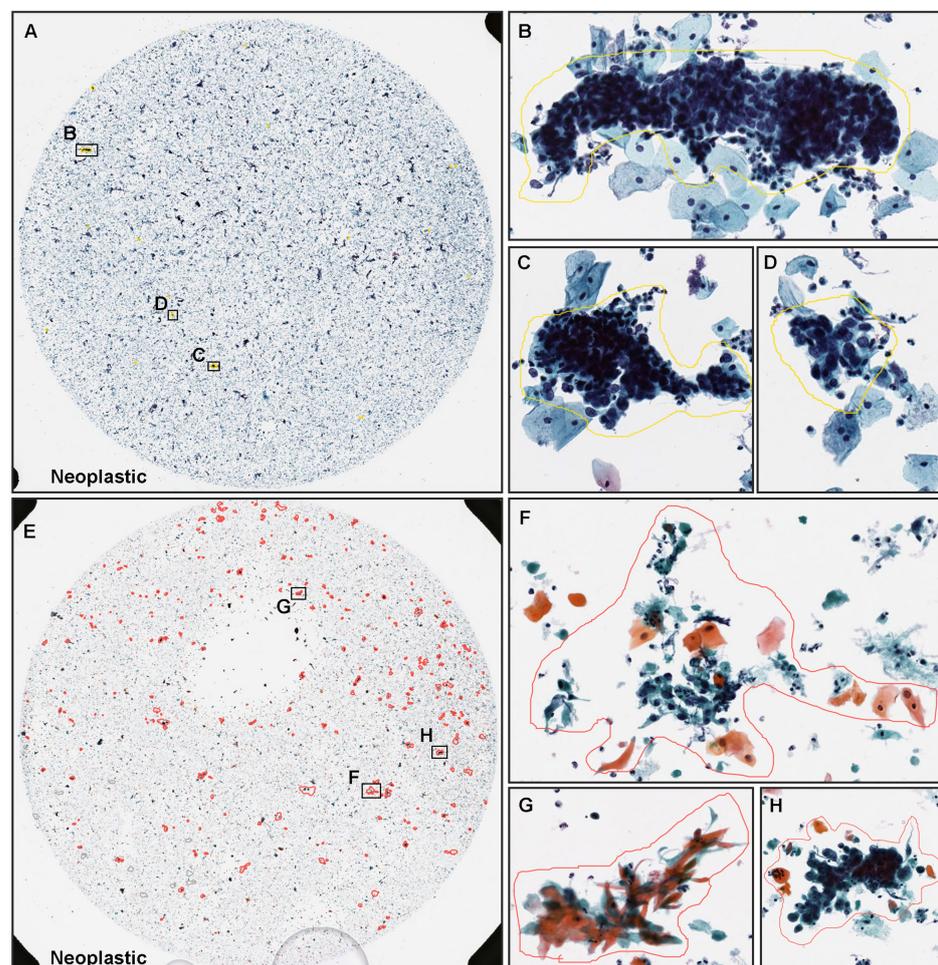


Figure 1. Representative manually drawing annotation images for neoplastic labels on liquid-based cytology (LBC) slides. The LBC case (A) was diagnosed as HSIL (high-grade squamous intraepithelial lesion) based on the representative neoplastic squamous epithelial cells with increase in nuclear/cytoplasmic ratio and nuclear atypia (B–D). The LBC case (E) was diagnosed as SCC (squamous cell carcinoma) based on the representative neoplastic squamous epithelial cells with HSIL features (F–H). Representative neoplastic cells were roughly annotated using in-house on-line drawing tools.

2.4. Deep Learning Models

Our deep learning models consisted of a convolutional neural network (CNN) and a recurrent neural network (RNN) that were trained simultaneously end to end. For the CNN, we have used the EfficientNetB0 architecture [50] with a modified input size of 1024×1024 px to allow a larger view; this is based on cytologists' input that they usually need to view the neighbouring cells around a given cell in order to diagnose more accurately. We then performed 7×7 max pooling with a stride of 5×5 . The output of the CNN was reshaped and provided as input to an RNN with a gated recurrent unit Cho et al. [51] model of size 128, followed by a fully connected layer. We used the partial fine-tuning approach [52] for the tuning the CNN component, where only the affine weights of the batch normalisation layers are updated while the rest of the weights in the CNN remain frozen. We used the pre-trained weights from ImageNet as starting weights. Figure 2 shows a simplified overview of the model. The RNN component was initialised with random weights.

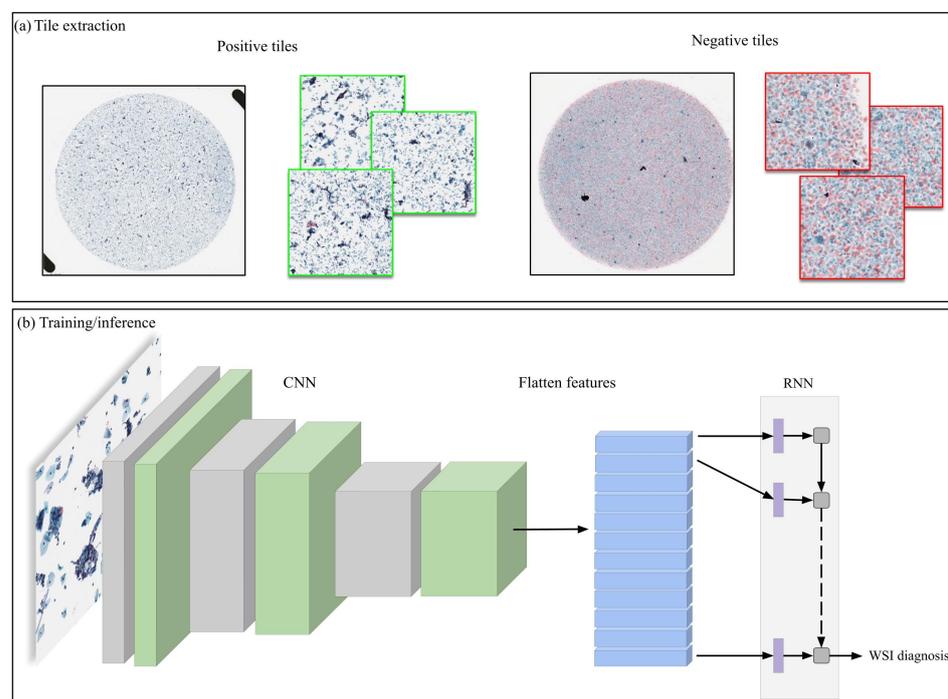


Figure 2. Method overview. (a) Large 1024×1024 are extracted from the WSIs; for the neoplastic WSIs, tiles are extracted only from annotated regions, while from NILM WSIs, tiles are extracted randomly from any region. (b) The tiles are then used to create random balanced batches used to train the model, which is composed of a CNN and an RNN and are trained simultaneously. During inference, the model is applied on all of the tiles of the WSIs in a sliding window fashion, and the WSI label is predicted based on the maximum probability from all of the tiles.

WSIs tend to contain a large white background that is not relevant for the model. We therefore start the preprocessing by eliminating the white background using Otsu's method [53] applied to the greyscale version of the WSIs.

For training and inference, we then proceeded by extracting 1024×1024 px tiles from the tissue regions. We performed the extraction in real-time using the OpenSlide library [54]. To perform inference on a WSI, we used a sliding window approach with a fixed-size stride of 512×512 px (half the tile size). This results in a grid-like output of predictions on all areas that contained cells, which then allowed us to visualise the prediction as a heatmap of probabilities that we can directly superimpose on top of the WSI. Each tile had a probability of being neoplastic; to obtain a single probability that is representative of the WSI, we computed the maximum probability from all the tiles.

During training, we maintained an equal balance of positively and negatively labelled tiles in the training batch. To do so, for the positive tiles, we extracted them randomly from the annotated regions of neoplastic WSIs, such that within the 1024×1024 px, at least one annotated cell was visible anywhere inside the tile. For the negative tiles, we extracted them randomly anywhere from the tissue regions of NILM WSIs. We then interleaved the positive and negative tiles to construct an equally balanced batch that was then fed as input to the CNN. In addition, to reduce the number of false positives, given the large size of the WSIs, we performed a hard mining of tiles, whereby at the end of each epoch, we performed full sliding window inference on all the NILM WSIs in order to adjust the random sampling probability such that false positively predicted tiles of NILM were more likely to be sampled.

During training, we performed real-time augmentation of the extracted tiles using variations of brightness, saturation, and contrast. We trained the model using the Adam optimisation algorithm [55], with the binary cross entropy loss, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate of 0.001. We applied a learning rate decay of 0.95 every 2 epochs. We used early stopping by tracking the performance of the model on a validation set, and training was stopped automatically when there was no further improvement on the validation loss for 10 epochs. The model with the lowest validation loss was chosen as the final model.

2.5. Interobserver Concordance Study

For the interobserver concordance study, a total of 10 WSIs (8 NILM cases and 2 neoplastic cases) of cervical LBC already reported by a clinical laboratory were retrieved from the records. Using the in-house on-line web virtual slide application, a total of 16 cytoscreeners (8 have over 10 years experiences and 8 have less than 10 years experiences) have reviewed the 10 WSIs and reported in subclasses (NILM, ASC-US, ASC-H, LSIL, HSIL, SCC, ADC).

2.6. Software and Statistical Analysis

The deep learning models were implemented and trained using the open-source TensorFlow library [56].

To assess the cytopathological diagnostic concordance of cytoscreeners, we performed the Fleiss' kappa statistic, which is a measure of inter-rater agreement of a categorical variable [57] between two or more raters. We calculated the kappa values using Microsoft Excel 2016 MSO (16.0.13029.20232) 64 bit. The scale for interpretation is as follows: ≤ 0.0 , poor agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, almost perfect agreement. AUCs were calculated in python using the scikit-learn package [58] and plotted using matplotlib [59]. The 95% CIs of the AUCs were estimated using the bootstrap method [60] with 1000 iterations.

The true positive rate (*TPR*) was computed as

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

and the false positive rate (*FPR*) was computed as

$$FPR = \frac{FP}{FP + TN} \quad (2)$$

where *TP*, *FP*, and *TN* represent true positive, false positive, and true negative, respectively. The ROC curve was computed by varying the probability threshold from 0.0 to 1.0 and computing both the *TPR* and *FPR* at the given threshold.

2.7. Code Availability

We adapted the training code from https://github.com/tensorflow/models/tree/master/official/vision/image_classification (accessed on 14 February 2020).

3. Results

3.1. High AUC Performance of WSI Evaluation of Neoplastic Cervical Liquid-Based Cytology (LBC) Images

The aim of this retrospective study was to train a deep learning model for the classification of neoplastic cervical WSIs. We trained a model that consists of a convolutional and a recurrent neural network using a dataset of 1503 WSIs for training and 150 for validation. We evaluated the model on three test sets with a combined total of 1468 WSIs. Figure 3 shows the resulting ROC curves, and Table 2 lists the resulting ROC AUC and log loss, as well as the accuracy, sensitivity, and specificity computed at a probability threshold of 0.5. Table 3 shows the confusion matrix. The model achieved a good performance overall, with ROC AUCs of 0.96 (0.92–0.99) on the full agreement, 0.89 (0.81–0.96) on the clinical balance reviewed, and 0.92 (0.89–0.94) on the equal balance reviewed test sets.

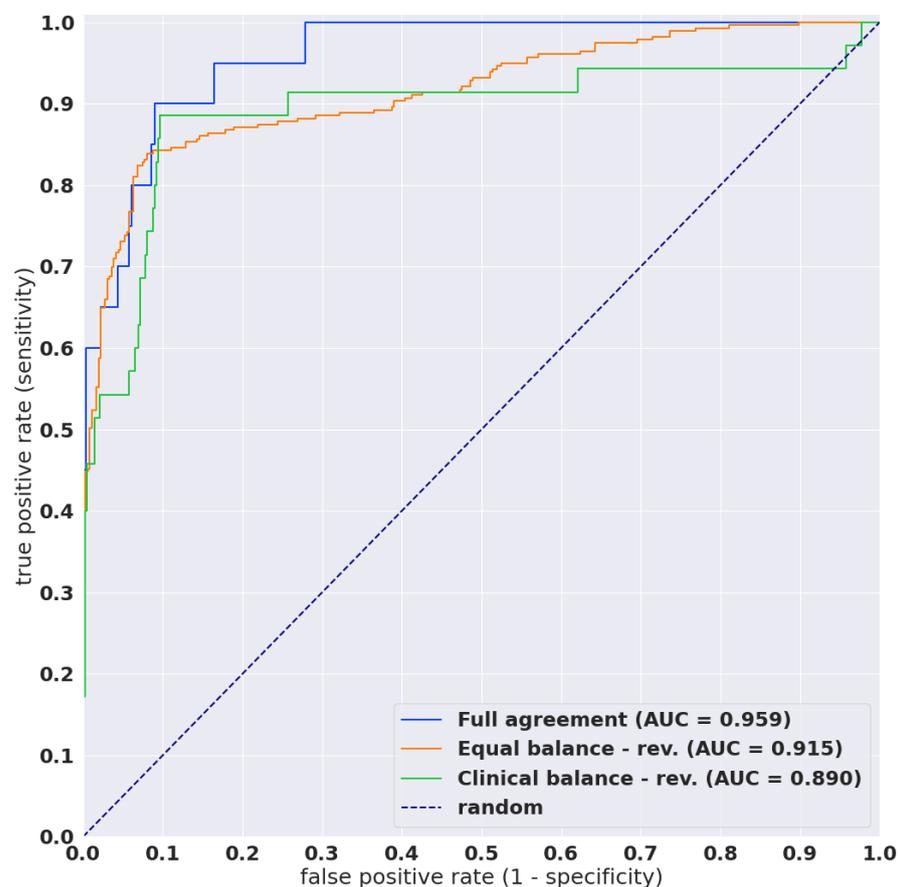


Figure 3. ROC curves for the three test sets.

Table 2. ROC AUC, log loss, accuracy, sensitivity, and specificity results on the test sets.

	Full Agreement	Clinical Balance	Clinical Balance-rev.	Equal Balance	Equal Balance-rev.
ROC AUC	0.960 [0.921–0.988]	0.774 [0.679–0.841]	0.890 [0.808–0.963]	0.827 [0.795–0.852]	0.915 [0.892–0.937]
log loss	2.244 [2.021–2.458]	2.272 [2.141–2.412]	1.347 [1.238–1.465]	1.126 [0.994–1.264]	0.913 [0.794–1.055]
accuracy	0.907 [0.873–0.937]	0.629 [0.591–0.660]	0.903 [0.876–0.924]	0.759 [0.725–0.785]	0.885 [0.859–0.908]
sensitivity	0.850 [0.667–1.000]	0.816 [0.686–0.923]	0.886 [0.774–0.978]	0.624 [0.573–0.668]	0.839 [0.794–0.880]
specificity	0.911 [0.877–0.942]	0.619 [0.579–0.652]	0.904 [0.877–0.926]	0.893 [0.862–0.924]	0.920 [0.890–0.945]

Table 3. Confusion matrix.

			Predicted Label	
			NILM	Neoplastic
Full agreement	True label	NILM	255	25
		Neoplastic	3	17
Clinical balance	True label	NILM	441	271
		Neoplastic	7	31
Clinical balance-rev.	True label	NILM	443	47
		Neoplastic	4	31
Equal balance	True label	NILM	335	40
		Neoplastic	141	234
Equal balance-rev.	True label	NILM	335	29
		Neoplastic	45	234

3.2. True Positive Prediction

Our deep learning model satisfactorily predicted neoplastic epithelial cells (Figure 4C–G) in cervical LBC (Figure 4A,B) specimen. The heatmap image shows true positive predictions (Figure 4B–D) of neoplastic epithelial cells. In contrast, in low probability tiles (Figure 4H,I), two independent cytoscreeners confirmed there were no neoplastic epithelial cells.

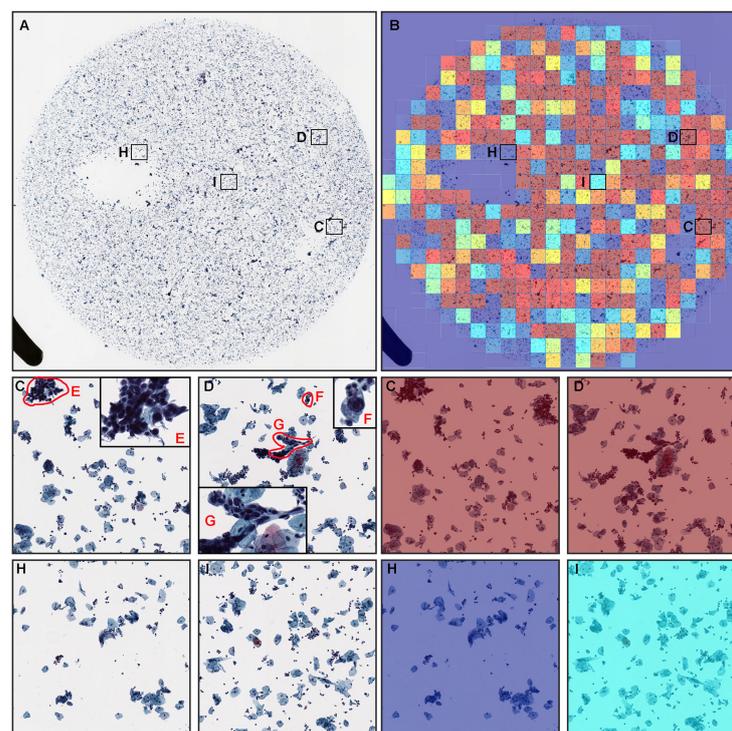


Figure 4. A representative example of neoplastic true positive prediction outputs on a liquid-based cytology (LBC) case from test sets. In the neoplastic whole-slide image (WSI) of LBC specimen (A), the heatmap image (B) shows a true positive prediction of neoplastic epithelial cells in high probability tiles (C,D), which correspond, respectively, to neoplastic epithelial cells (E–G) equivalent to HSIL (high-grade squamous intraepithelial lesion). On the other hand, in low probability tiles (H,I) of the same heatmap image (B), there are no evidence of neoplastic cells.

3.3. True Negative Prediction

Our model satisfactorily predicted NILM cases (Figure 5A,B) in cervical LBC specimen. The heatmap image shows true negative predictions (Figure 5B,D,E) of neoplastic

epithelial cells. In both zero (Figure 5C) and very low probability tiles (Figure 5D,E), there are no neoplastic epithelial cells.

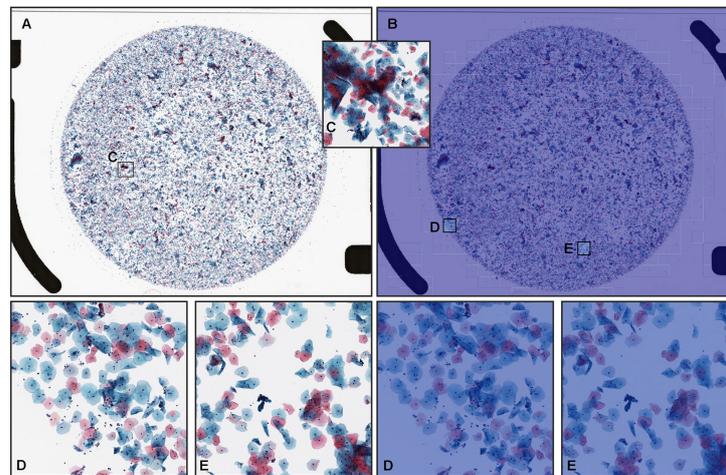


Figure 5. A representative example of neoplastic true negative prediction outputs on a liquid-based cytology (LBC) case from test sets. In the NILM (negative for intraepithelial lesion or malignancy) whole slide image (WSI) of LBC specimen (A), the heatmap image (B) shows true negative prediction of neoplastic epithelial cells which correspond, respectively, to non-neoplastic epithelial cells (C). Moreover, in very low probability tiles (D,E) of the same heatmap image (B), there are no evidence of neoplastic cells.

3.4. False Positive Prediction

A cytopathologically diagnosed NILM case (Figure 6A) was false positively predicted for neoplastic epithelial cells (Figure 6B). The heatmap image (Figure 6B) shows false positive predictions of neoplastic epithelial cells (Figure 6C,E) with high probabilities. Cytopathologically, there are parabasal cells with a high nuclear cytoplasmic (N/C) ratio (Figure 6C,D) and cell clusters of squamous epithelial cells with cervical gland cells with high N/C ratios (Figure 6E), which could be a major cause of false positive.

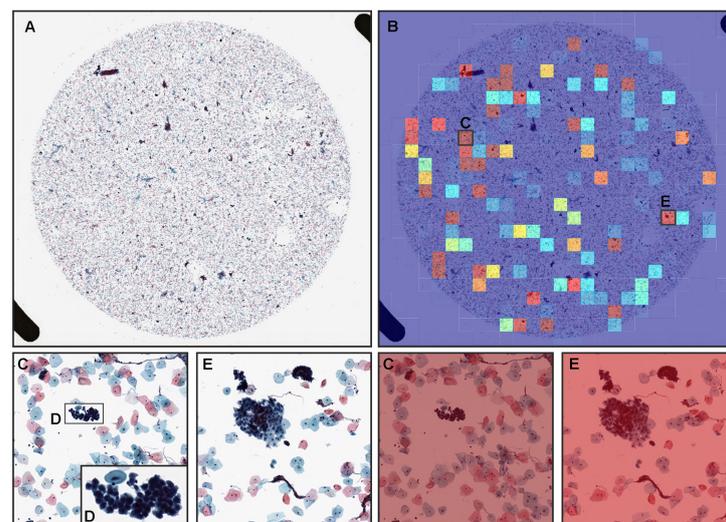


Figure 6. A representative example of neoplastic false positive prediction outputs on a liquid-based cytology (LBC) case from test sets. Cytopathologically, (A) is a NILM (negative for intraepithelial lesion or malignancy) whole-slide image (WSI) of LBC specimen. The heatmap image (B) exhibited false positive predictions of neoplastic tiles (C,E). In (C), there are parabasal cells with a slightly high nuclear cytoplasmic (N/C) ratio with dense chromatin appearance due to the cellular overlapping (D). In (E), there are cell clusters of squamous epithelial cells and cervical gland cells with slightly high N/C ratios and a dense chromatin appearance due to the cellular overlapping.

3.5. Interobserver Variability

To evaluate the practical interobserver variability among cytoscreeners, we have asked a total of 16 cytoscreeners (8 are over 10 years experiences and 8 are less than 10 years experiences) to review the same 10 LBC WSIs, which consist of 8 NILM and 2 neoplastic cases already diagnosed by a clinical laboratory. The results of each cytoscreener were summarised in Table 4. The Fleiss' kappa statistics were summarised in Table 5. There was poor to moderate concordance in assessing subclass, with Fleiss' kappas of NILM (range: 0.042–0.755), neoplastic (range: 0.098–0.500), and all cases (range: 0.364–0.716). On the other hand, there was poorly to almost perfect concordance in assessing binary class, with Fleiss' kappas of NILM (range: 0.073–0.815), neoplastic (1.000), and all cases (range: 0.568–0.861). Interestingly, there was a robust higher concordance in both subclass and binary class among cytoscreeners over 10-year experiences. However, overall, there was poor concordance in assessing NILM cases (range: 0.042–0.073).

Table 4. Cytopathological evaluations for 10 LBC WSIs by diagnostic report (Dx) and 16 cytoscreeners (CS) with their age and years of experience.

Age	Exp. (Years)		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10
		Dx	NILM	HSIL	LSIL							
30s	≥10	CS1	NILM	HSIL	ASC-H							
50s		CS2	NILM	NILM	NILM	ASC-H	NILM	NILM	HSIL	ASC-H	HSIL	HSIL
50s		CS3	NILM	ASC-US	HSIL	LSIL						
40s		CS4	NILM	NILM	NILM	ASC-US	NILM	NILM	NILM	ASC-US	HSIL	SCC
30s		CS5	NILM	HSIL	ASC-US							
30s		CS6	NILM	ASC-US	NILM	NILM	NILM	NILM	NILM	NILM	HSIL	HSIL
60s		CS7	NILM	HSIL	ASC-H							
40s		CS8	NILM	HSIL	ASC-US							
20s	<10	CS9	NILM	HSIL	LSIL							
20s		CS10	NILM	LSIL	LSIL							
30s		CS11	NILM	NILM	NILM	NILM	ASC-H	NILM	NILM	HSIL	LSIL	HSIL
20s		CS12	NILM	ASC-US	ASC-H	NILM	NILM	NILM	NILM	LSIL	SCC	HSIL
40s		CS13	NILM	NILM	HSIL	NILM	NILM	NILM	NILM	ASC-US	HSIL	ASC-H
30s		CS14	NILM	NILM	LSIL	NILM	NILM	NILM	NILM	NILM	HSIL	LSIL
20s		CS15	NILM	NILM	NILM	NILM	NILM	NILM	LSIL	NILM	HSIL	ASC-US
20s		CS16	NILM	NILM	NILM	ASC-US	LSIL	NILM	NILM	ASC-US	HSIL	SCC

Table 5. Interobserver variability: kappa.

Classification	Dx Report	16 Cytoscreeners	8 Cytoscreeners (≥10 Years of Exp.)
Subclass	NILM	0.042 (slight)	0.755 (substantial)
	Neoplastic	0.098 (slight)	0.500 (moderate)
	All cases	0.364 (fair)	0.716 (substantial)
Binary	NILM	0.073 (slight)	0.815 (almost perfect)
	Neoplastic	1.000 (complete)	1.000 (complete)
	All cases	0.568 (moderate)	0.861 (almost perfect)

4. Discussion

In this pilot study, we trained a deep learning model for the classification of neoplastic cells in WSIs of LBC specimens. The model achieved overall a good performance, with ROC AUCs of 0.96 (0.92–0.99) on the full agreement, 0.89 (0.81–0.96) on the clinical balance reviewed, and 0.92 (0.89–0.94) on the equal balance reviewed test sets.

Looking at the interobserver concordance among cytoscreeners in Table 4, it is obvious that there is considerable interobserver variability, with the poor concordance in NILM cases even for binary classification (NILM vs. neoplastic). In addition, there is the problem of human fatigue due to the continuous observation of a large number of cases. Therefore,

when considering future accuracy control, it may be necessary to conduct screening using deep learning model(s) with guaranteed accuracy, such as the results of this study, at least in the binary classification (NILM vs. neoplastic), and to conduct detailed assessments by cytoscreeners and cytopathologists in the subclassification (e.g., NILM, ASC-US, ASC-H, LSIL, HSIL, SCC, and ADC).

From our results in Figure 2, it was obvious that there was interobserver variability among cytoscreeners in different clinical laboratories and hospitals. Clinical balance and equal balance test sets were prepared based on diagnostic (screening) reports from a clinical laboratory. The only difference between clinical balance and clinical balance-reviewed (same as equal balance and equal balance-reviewed) was whether it was additionally reviewed by two more cytoscreeners in different clinical laboratories and hospitals or not. All scores (ROC-AUC, accuracy, sensitivity, and specificity) were increased in clinical balance-reviewed and equal balance-reviewed test sets as compared to clinical balance and equal balance test sets (Figure 2). Hence, our deep learning model would be helpful for standardising in the screening process.

In routine cervical cancer screening at clinical laboratories and hospitals, it is difficult to introduce a screening programme dependent on cervical smears due to poor human cytoscreener resources. LBC techniques opened new possibilities for a systemic cervical cancer screening. LBC slides are amenable to high throughput automated analysis. Especially for the detection of rare events on LBC slides, WSI and subsequent image analysis is of crucial importance for guaranteeing a standardised high-quality read out [25]. Practical automated cervical cytology screening devices have been under development since the 1950s. The technological development in semi-automated screening devices for cervical cancer screening is very rapid; however, currently, no machines are available to provide a fully automated screening by computer without human intervention. There are two FDA-approved semi-automated slide scanning devices on the market; these systems are the BD FocalPoint GS Imaging System and the HOLOGIC ThinPrep Imaging System. Both are designed to perform computer-assisted analysis of cellular images followed by location-guided screening of limited fields of view. FocalPoint-assisted smear reading has been proposed prior to conventional manual reading; the latter may be unnecessary for cases reported as No Further Review (NFR) and would be required for cases reported as Review (REV) [61]. FocalPoint-assisted practice showed statistically superior sensitivity and specificity when compared to conventional manual smear screening for the detection of HSIL and LSIL [14,62,63]. However, ASC-US sensitivity and specificity were not significantly different between FocalPoint-assisted practice and conventional screening [62]. Overall, in neoplastic slides (ASC-US, LSIL, and HSIL) by FocalPoint-assisted practice, sensitivity was in the range of 81.1–86.1% and specificity was in the range of 84.5–95.1% [62]. The other study showed that FocalPoint-assisted reading was comparable to conventional reading, and the very low observed negative predictive value of an NFR report (0.02%) suggested that these cases might safely return to periodic screening [61]. The ThinPrep Imaging System (TIS) is an automated system that uses location-guided screening to assist cytoscreeners in reviewing a ThinPrep Pap LBC slides [64]. TIS scans the LBC slides and identifies 22 fields of view (FOVs) on each slide based on optical density measurements and other features [64]. It has been reported that TIS was ideally suited to the rapid screening of negative cases; however, the sensitivity and specificity of the TIS (85.19% and 96.67%, respectively) were equivalent to those of manual screening (89.38% and 98.42%, respectively) [65]. In another study, for diagnostic categories of neoplastic slides (ASC-US, LSIL, and HSIL) by TIS practice, sensitivity was in the range of 79.2–82.0% and specificity was in the range of 97.8–99.6% [64].

As shown in Figure 2, our LBC cervical cancer screening deep learning model exhibited around 90% accuracy (in the range of 89–91%), 86% sensitivity (in the range of 84–89%), and 91% specificity (in the range of 90–92%) in full agreement, clinical balance-reviewed, and equal balance-reviewed test sets; those scores were as well or better than the existing assistance systems mentioned above.

5. Conclusions

In the present study, we have trained a deep learning model for the classification of neoplastic cervical LBC in WSIs. We have evaluated the model on three test sets achieving ROC-AUCs for WSI diagnosis in the range of 0.89–0.96. The main advantage of our deep learning model is that the model can be used to evaluate the cervical LBC at the WSI level. Therefore, our model is able to infer whether the cervical LBC WSI is NILM (non-neoplastic) (Figure 5) or neoplastic (Figure 4). This makes it possible to use a deep learning model such as ours as a tool to aid in the cervical screening process, which could potentially be used to rank the cases by order of priority. After which the cytoscreeners will need to perform full screening and subclassification (e.g., ASC-US, ASC-H, LSIL, HSIL, SCC, ADC) on neoplastic output cases after the primary screening by our deep learning model, which could reduce their working time as the model would have highlighted the potential suspected neoplastic regions, and they would not have to perform an exhaustive search through the entire WSI.

Author Contributions: F.K. and M.T. contributed equally to this study; F.K., S.I. and M.T. designed the studies; F.K., N.H., T.I., A.F., S.I. and M.T. performed experiments and analysed the data; N.H., T.I., A.F. and S.I. performed cytopathological diagnoses and reviewed cases; F.K. and M.T. performed computational studies; F.K., S.I. and M.T. wrote the manuscript; M.T. supervised the project. All authors reviewed and approved the final manuscript.

Funding: The authors received no financial supports for the research, authorship, and publication of this study.

Institutional Review Board Statement: The experimental protocol in this study was approved by the ethical board of the private clinical laboratory. All research activities complied with all relevant ethical regulations and were performed in accordance with relevant guidelines and regulations in the clinical laboratory. Due to the confidentiality agreement with the private clinical laboratory, the name of the clinical laboratory cannot be disclosed.

Informed Consent Statement: Informed consent to use cytopathological samples (liquid-based cytology glass slides) and cytopathological reports for research purposes had previously been obtained from all patients and the opportunity for refusal to participate in research had been guaranteed by an opt-out manner.

Data Availability Statement: The datasets used in this study are not publicly available due to specific institutional requirements governing privacy protection; however, they are available from the corresponding author and from the private clinical laboratory in Japan on reasonable request. Restrictions apply based on the data use agreement, which was made according to the Ethical Guidelines for Medical and Health Research Involving Human Subjects as set by the Japanese Ministry of Health, Labour, and Welfare.

Acknowledgments: We thank cytoscreeners and pathologists who have been engaged in reviewing cases, annotations, and cytopathological discussion for this study.

Conflicts of Interest: F.K. and M.T. are employees of Medmain Inc. The authors declare no conflict of interest.

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 Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Arbyn, M.; Anttila, A.; Jordan, J.; Ronco, G.; Schenck, U.; Segnan, N.; Wiener, H.; Herbert, A.; Von Karsa, L. European guidelines for quality assurance in cervical cancer screening.—Summary document. *Ann. Oncol.* **2010**, *21*, 448–458. [[CrossRef](#)] [[PubMed](#)]
3. Wright, T.C.; Schiffman, M.; Solomon, D.; Cox, J.T.; Garcia, F.; Goldie, S.; Hatch, K.; Noller, K.L.; Roach, N.; Runowicz, C.; et al. Interim guidance for the use of human papillomavirus DNA testing as an adjunct to cervical cytology for screening. *Obstet. Gynecol.* **2004**, *103*, 304–309. [[CrossRef](#)] [[PubMed](#)]
4. Wright, T.C., Jr.; Massad, L.S.; Dunton, C.J.; Spitzer, M.; Wilkinson, E.J.; Solomon, D. 2006 consensus guidelines for the management of women with abnormal cervical cancer screening tests. *Am. J. Obstet. Gynecol.* **2007**, *197*, 346–355. [[CrossRef](#)]

5. Saslow, D.; Runowicz, C.D.; Solomon, D.; Moscicki, A.B.; Smith, R.A.; Eyre, H.J.; Cohen, C. American Cancer Society guideline for the early detection of cervical neoplasia and cancer. *CA Cancer J. Clin.* **2002**, *52*, 342–362. [\[CrossRef\]](#)
6. Smith, R.A.; Andrews, K.S.; Brooks, D.; Fedewa, S.A.; Manassaram-Baptiste, D.; Saslow, D.; Wender, R.C. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J. Clin.* **2019**, *69*, 184–210. [\[CrossRef\]](#)
7. Sasieni, P.; Castanon, A.; Cuzick, J. Effectiveness of cervical screening with age: Population based case-control study of prospectively recorded data. *BMJ* **2009**, *339*, b2968. [\[CrossRef\]](#)
8. Hamashima, C.; Aoki, D.; Miyagi, E.; Saito, E.; Nakayama, T.; Sagawa, M.; Saito, H.; Sobue, T. The Japanese guideline for cervical cancer screening. *Jpn. J. Clin. Oncol.* **2010**, *40*, 485–502. [\[CrossRef\]](#)
9. ACOG, Committee on Practice Bulletins. ACOG Practice Bulletin Number 45, August 2003: Committee on Practice Bulletins-Gynecology. Cervical Cytology Screening. *Obstet. Gynecol.* **2003**, *102*, 417–427. [\[CrossRef\]](#)
10. Anttila, A.; Pukkala, E.; Söderman, B.; Kallio, M.; Nieminen, P.; Hakama, M. Effect of organised screening on cervical cancer incidence and mortality in Finland, 1963–1995: Recent increase in cervical cancer incidence. *Int. J. Cancer* **1999**, *83*, 59–65. [\[CrossRef\]](#)
11. McGoogan, E.; Reith, A. Would monolayers provide more representative samples and improved preparations for cervical screening? Overview and evaluation of systems available. *Acta Cytol.* **1996**, *40*, 107–119. [\[CrossRef\]](#)
12. Fahey, M.T.; Irwig, L.; Macaskill, P. Meta-analysis of Pap test accuracy. *Am. J. Epidemiol.* **1995**, *141*, 680–689. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Solomon, D.; Schiffman, M.; Tarone, R. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: Baseline results from a randomized trial. *J. Natl. Cancer Inst.* **2001**, *93*, 293–299. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Lee, J.; Kuan, L.; Oh, S.; Patten, F.W.; Wilbur, D.C. A feasibility study of the AutoPap system location-guided screening. *Acta Cytol.* **1998**, *42*, 221–226. [\[CrossRef\]](#) [\[PubMed\]](#)
15. 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. *Diagn. Cytopathol.* **2013**, *41*, 174–178. [\[CrossRef\]](#)
16. Sugiyama, Y.; Sasaki, H.; Komatsu, K.; Yabushita, R.; Oda, M.; Yanoh, K.; Ueda, M.; Itamochi, H.; Okugawa, K.; Fujita, H.; et al. A multi-institutional feasibility study on the use of automated screening systems for quality control rescreening of cervical cytology. *Acta Cytol.* **2016**, *60*, 451–457. [\[CrossRef\]](#)
17. Colgan, T.; Patten, S., Jr; Lee, J. A clinical trial of the AutoPap 300 QC system for quality control of cervicovaginal cytology in the clinical laboratory. *Acta Cytol.* **1995**, *39*, 1191–1198.
18. Patten, S.F., Jr; Lee, J.S.; Wilbur, D.C.; Bonfiglio, T.A.; Colgan, T.J.; Richart, R.M.; Cramer, H.; Moinuddin, S. The AutoPap 300 QC System multicenter clinical trials for use in quality control rescreening of cervical smears: I. A prospective intended use study. *Cancer Cytopathol. Interdiscip. Int. J. Am. Cancer Soc.* **1997**, *81*, 337–342. [\[CrossRef\]](#)
19. Marshall, C.J.; Rowe, L.; Bentz, J.S. Improved quality-control detection of false-negative pap smears using the Autopap 300 QC system. *Diagn. Cytopathol.* **1999**, *20*, 170–174. [\[CrossRef\]](#)
20. Saieg, M.A.; Motta, T.H.; Fodra, M.E.; Scapulatempo, C.; Longatto-Filho, A.; Stiepcich, M.M. Automated screening of conventional gynecological cytology smears: Feasible and reliable. *Acta Cytol.* **2014**, *58*, 378–382. [\[CrossRef\]](#)
21. Nanda, K.; McCrory, D.C.; Myers, E.R.; Bastian, L.A.; Hasselblad, V.; Hickey, J.D.; Matchar, D.B. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: A systematic review. *Ann. Intern. Med.* **2000**, *132*, 810–819. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Krane, J.F.; Granter, S.R.; Trask, C.E.; Hogan, C.L.; Lee, K.R. Papanicolaou smear sensitivity for the detection of adenocarcinoma of the cervix: A study of 49 cases. *Cancer Cytopathol.* **2001**, *93*, 8–15. [\[CrossRef\]](#)
23. Stoler, M.H.; Schiffman, M. Interobserver reproducibility of cervical cytologic and histologic interpretations: Realistic estimates from the ASCUS-LSIL Triage Study. *JAMA* **2001**, *285*, 1500–1505. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Mukhopadhyay, S.; Feldman, M.D.; Abels, E.; Ashfaq, R.; Beltaifa, S.; Cacciabeve, N.G.; Cathro, H.P.; Cheng, L.; Cooper, K.; Dickey, G.E.; et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *Am. J. Surg. Pathol.* **2018**, *42*, 39. [\[CrossRef\]](#)
25. Lahrmann, B.; Valous, N.A.; Eisenmann, U.; Wentzensen, N.; Grabe, N. Semantic focusing allows fully automated single-layer slide scanning of cervical cytology slides. *PLoS ONE* **2013**, *8*, e61441. [\[CrossRef\]](#)
26. Song, Y.; Zhang, L.; Chen, S.; Ni, D.; Li, B.; Zhou, Y.; Lei, B.; Wang, T. A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 2903–2906.
27. Yu, K.H.; Zhang, C.; Berry, G.J.; Altman, R.B.; Ré, C.; Rubin, D.L.; Snyder, M. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **2016**, *7*, 12474. [\[CrossRef\]](#)
28. Hou, L.; Samaras, D.; Kurc, T.M.; Gao, Y.; Davis, J.E.; Saltz, J.H. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2424–2433.
29. Litjens, G.; Sánchez, C.I.; Timofeeva, N.; Hermsen, M.; Nagtegaal, I.; Kovacs, I.; Hulsbergen-Van De Kaa, C.; Bult, P.; Van Ginneken, B.; Van Der Laak, J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **2016**, *6*, 26286. [\[CrossRef\]](#)

30. Kraus, O.Z.; Ba, J.L.; Frey, B.J. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics* **2016**, *32*, i52–i59. [[CrossRef](#)]
31. Korbar, B.; Olofson, A.M.; Mirafior, A.P.; Nicka, C.M.; Suriawinata, M.A.; Torresani, L.; Suriawinata, A.A.; Hassanpour, S. Deep learning for classification of colorectal polyps on whole-slide images. *J. Pathol. Inform.* **2017**, *8*, 30.
32. 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 Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia, 18–21 April 2017; pp. 406–409.
33. Luo, X.; Zang, X.; Yang, L.; Huang, J.; Liang, F.; Rodriguez-Canales, J.; Wistuba, I.I.; Gazdar, A.; Xie, Y.; Xiao, G. Comprehensive computational pathological image analysis predicts lung cancer prognosis. *J. Thorac. Oncol.* **2017**, *12*, 501–509. [[CrossRef](#)]
34. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [[CrossRef](#)] [[PubMed](#)]
35. Wei, J.W.; Tafe, L.J.; Linnik, Y.A.; Vaickus, L.J.; Tomita, N.; Hassanpour, S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci. Rep.* **2019**, *9*, 3358. [[CrossRef](#)] [[PubMed](#)]
36. Jith, O.N.; Harinarayanan, K.; Gautam, S.; Bhavsar, A.; Sao, A.K. DeepCerv: Deep neural network for segmentation free robust cervical cell classification. In *Computational Pathology and Ophthalmic Medical Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 86–94.
37. Lin, H.; Hu, Y.; Chen, S.; Yao, J.; Zhang, L. Fine-grained classification of cervical cells using morphological and appearance based convolutional neural networks. *IEEE Access* **2019**, *7*, 71541–71549. [[CrossRef](#)]
38. Gupta, M.; Das, C.; Roy, A.; Gupta, P.; Pillai, G.R.; Patole, K. Region of interest identification for cervical cancer images. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 1293–1296.
39. 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. *J. Comput. Sci. Technol.* **2021**, *36*, 347–360. [[CrossRef](#)]
40. Gertych, A.; Swiderska-Chadaj, Z.; Ma, Z.; Ing, N.; Markiewicz, T.; Cierniak, S.; Salemi, H.; Guzman, S.; Walts, A.E.; Knudsen, B.S. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci. Rep.* **2019**, *9*, 1483. [[CrossRef](#)]
41. Bejnordi, B.E.; Veta, M.; Van Diest, P.J.; Van Ginneken, B.; Karssemeijer, N.; Litjens, G.; Van Der Laak, J.A.; Hermsen, M.; Manson, Q.F.; Balkenhol, M.; et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **2017**, *318*, 2199–2210. [[CrossRef](#)]
42. Saltz, J.; Gupta, R.; Hou, L.; Kurc, T.; Singh, P.; Nguyen, V.; Samaras, D.; Shroyer, K.R.; Zhao, T.; Batiste, R.; et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **2018**, *23*, 181–193. [[CrossRef](#)] [[PubMed](#)]
43. Campanella, G.; Hanna, M.G.; Geneslaw, L.; Mirafior, A.; Silva, V.W.K.; Busam, K.J.; Brogi, E.; Reuter, V.E.; Klimstra, D.S.; Fuchs, T.J. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **2019**, *25*, 1301–1309. [[CrossRef](#)]
44. Iizuka, O.; Kanavati, F.; Kato, K.; Rambeau, M.; Arihiro, K.; Tsuneki, M. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci. Rep.* **2020**, *10*, 1504. [[CrossRef](#)]
45. Holmström, O.; Linder, N.; Kaingu, H.; Mbuuko, N.; Mbete, J.; Kinyua, F.; Törnquist, S.; Muinde, M.; Krogerus, L.; Lundin, M.; et al. Point-of-Care Digital Cytology With Artificial Intelligence for Cervical Cancer Screening in a Resource-Limited Setting. *JAMA Netw. Open* **2021**, *4*, e211740. [[CrossRef](#)]
46. 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. *Med. Image Anal.* **2021**, *69*, 101955. [[CrossRef](#)] [[PubMed](#)]
47. 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. *Nat. Commun.* **2021**, *12*, 5639. [[CrossRef](#)] [[PubMed](#)]
48. Li, W.; Liu, L.L.; Luo, Z.Z.; Han, C.Y.; Wu, Q.H.; Zhang, L.; Tian, L.S.; Yuan, J.; Zhang, T.; Chen, Z.W.; et al. Associations of sexually transmitted infections and bacterial vaginosis with abnormal cervical cytology: A cross-sectional survey with 9090 community women in China. *PLoS ONE* **2020**, *15*, e0230712. [[CrossRef](#)] [[PubMed](#)]
49. Duby, J.M.; DiFurio, M.J. Implementation of the ThinPrep Imaging System in a tertiary military medical center. *Cancer Cytopathol. J. Am. Cancer Soc.* **2009**, *117*, 264–270. [[CrossRef](#)] [[PubMed](#)]
50. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
51. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.
52. Kanavati, F.; Tsuneki, M. Partial transfusion: On the expressive influence of trainable batch norm parameters for transfer learning. *arXiv* **2021**, arXiv:2102.05543.
53. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man, Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
54. Goode, A.; Gilbert, B.; Harkes, J.; Jukic, D.; Satyanarayanan, M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J. Pathol. Inform.* **2013**, *4*, 27.

55. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
56. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: [tensorflow.org](https://www.tensorflow.org) (accessed on 3 February 2019).
57. Artstein, R.; Poesio, M. Inter-coder agreement for computational linguistics. *Comput. Linguist.* **2008**, *34*, 555–596. [[CrossRef](#)]
58. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
59. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
60. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994.
61. Troni, G.M.; Cariaggi, M.P.; Bulgaresi, P.; Houssami, N.; Ciatto, S. Reliability of sparing Papanicolaou test conventional reading in cases reported as no further review at AutoPap-assisted cytological screening: Survey of 30,658 cases with follow-up cytological screening. *Cancer Cytopathol. Interdiscip. Int. J. Am. Cancer Soc.* **2007**, *111*, 93–98. [[CrossRef](#)] [[PubMed](#)]
62. Wilbur, D.C.; Black-Schaffer, W.S.; Luff, R.D.; Abraham, K.P.; Kemper, C.; Molina, J.T.; Tench, W.D. The Becton Dickinson FocalPoint GS Imaging System: Clinical trials demonstrate significantly improved sensitivity for the detection of important cervical lesions. *Am. J. Clin. Pathol.* **2009**, *132*, 767–775. [[CrossRef](#)]
63. Wilbur, D.C.; Prey, M.U.; Miller, W.M.; Pawlick, G.F.; Colgan, T.J. The AutoPap system for primary screening in cervical cytology. Comparing the results of a prospective, intended-use study with routine manual practice. *Acta Cytol.* **1998**, *42*, 214–220. [[CrossRef](#)] [[PubMed](#)]
64. 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. *Am. J. Clin. Pathol.* **2005**, *123*, 281–287. [[CrossRef](#)] [[PubMed](#)]
65. Bolger, N.; Heffron, C.; Regan, I.; Sweeney, M.; Kinsella, S.; McKeown, M.; Creighton, G.; Russell, J.; O’Leary, J. Implementation and evaluation of a new automated interactive image analysis system. *Acta Cytol.* **2006**, *50*, 483–491. [[CrossRef](#)]