Effect of specimen processing technique on cell detection and classification by artificial intelligence

Brief title: Effect of processing technique in AI

Author names

Sayumi Maruyama[a], Nanako Sakabe[a], PhD, Chihiro Ito[a], Yuka Shimoyama[a], Shouichi Sato[b], PhD, Katsuhide Ikeda[a], PhD

Affiliation and postal address

[a] Pathophysiology Sciences, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Aichi, JAPAN

[b] Clinical Engineering, Faculty of medical sciences, Juntendo University, Urayasu, Chiba, JAPAN

Corresponding author:

Katsuhide Ikeda, CMIAC, PhD

Pathophysiology Sciences, Department of Integrated Health Sciences

Nagoya University Graduate School of Medicine

1-1-20 Daiko-Minami, Higashi-ku, Nagoya 461-8673, JAPAN

Tel: +81-52-719-3152

E-mail: k-ikeda@met.nagoya-u.ac.jp

KEY WORDS

cytopathology, artificial intelligence, deep learning, cell detection, cell classification,

CONFLICT OF INTEREST:

The authors declare no competing financial interests.

FUNDING STATEMENT:

This work was supported by JSPS KAKENHI (grant number JP21K18077).

ACKNOWLEDGEMENTS:

We would like to thank Editage (www.editage.jp) for their assistance with English language editing.

ABSTRACT

Objectives: Cytomorphology is known to differ depending on the processing technique, and these differences pose a problem for automated diagnosis using deep learning. We examined the as-yet unclarified relationship between cell detection or classification using AI and the AutoSmear and LBC processing techniques.

Methods: YOLOv5x was trained on the AutoSmear and LBC preparations of four cell lines, lung cancer (LC), cervical cancer (CC), malignant pleural mesothelioma (MM), and esophageal cancer (EC) cell lines. Detection and classification rates were used to evaluate the accuracy of cell detection.

Results: When preparations of the same processing technique were used for training and detection in the one cell (1C) model, the AutoSmear model had a higher detection rate than the LBC model. When different processing techniques were used for training and detection, detection rates of LC and CC were significantly lower in the four cell (4C) model than in the 1C model, and those of MM and EC were approximately 10% lower in the 4C model.

Conclusions: In AI-based cell detection and classification, attention should be paid to cells whose morphologies change significantly depending on the processing technique, further suggesting the creation of a training model.

KEY POINTS

- Cytomorphology is known to differ depending on the processing technique, and these differences pose a problem for automated diagnosis using deep learning.
- Accuracy of cell detection using deep learning is affected by the specimen processing technique, and its accuracy is reduced when different processing techniques are used for training and detection.
- In AI-based cell detection and classification, cells whose morphologies change significantly depending on the processing technique should be observed, suggesting the creation of a training model.

INTRODUCTION

With the success of deep learning and artificial intelligence (AI) in personal devices, social media, self-driving cars, and Go games,[1] its utilization in the medical field is anticipated. In the pathological field, AI-based research has accelerated due to digitization and the widespread use of whole slide imaging.[2,3] AI is a technology that allows computers to mimic human behavior or processes according to certain rules.[4] In 1992, as a practical application of AI-based automated diagnosis in cytology, PAPNET™ (Neuromedical Systems Inc. (NSI®), Suffern, NY, USA) was approved for commercial use as the first automated screening system in the world. In 2004 and 2008, the ThinPrep Imaging System™ (HOLOGIC®, Marlborough, MA, USA) and Focal Point™ GS Imaging System (BD, Franklin Lakes, NJ, USA) were launched and continue to dominate the market today,[5,6] however there is no utilization of this system except for in Papanicolaou tests. Previous automated diagnostic systems have detected atypical cells based on information such as cell size, staining, and human-defined algorithms.[7] The complexity of such information hinders the use of automated diagnostic systems for other cytological fields, suggesting a need for deep learning applications.

AI includes both machine learning and deep learning. Machine learning techniques allow a computer to “learn” from data without being the need for explicit programming.[4] Deep learning extrapolates the idea of machine learning by allowing algorithms to train themselves by exposing neural networks to large quantities of data.[4,8] Convolutional neural networks (CNN) are mainly utilized

in pathological deep learning studies[7] and have a strength in complex image interpretation. Algorithms using CNN locate multiple objects in an image. R-CNN (Region-based CNNs), fast R-CNN, single-shot multibox detector (SSD), and YOLO (You Only Look Once) are known as object detection algorithms. While most of these algorithms often include two steps of extracting candidate areas where objects exist from images and classifying what kind of objects they are,[9] YOLO can achieve region estimation and classification simultaneously; therefore, its processing speed is very high. Furthermore, YOLO rarely falsely detects the background as an object and has a high classification ability.[10] Cytopathological AI research also uses YOLO.[11,12]

Along with advances in automated diagnostic systems, liquid-based cytology (LBC) was developed in the 1990s. Conventional smears have adversely affected the evaluation of cytomorphology due to cell-to-cell overlaps and accumulation of non-cellular materials. LBC has been shown to solve this problem as well as improve the time and accuracy of manual screening.[6] The number of clinical laboratories using LBC is increasing, and LBC has emerged as a preferred alternative for cytologic specimens along with conventional smear and cytocentrifugation-based methods. Cytologic morphology is known to differ depending on the processing techniques and the LBC preservative solutions.[13] Recently, this difference in cytologic form is regarded as a problem of automated diagnosis using deep learning.[4,13,14] Previous AI studies have been small-scale and limited in the processing technique and the specimen type, where differences among laboratories were not

62 considered; therefore, the generalization of deep learning models was insufficient.[4,15,16–18] For
63 cytopathologists, the difference among the processing techniques does not affect the diagnosis,[19]
64 however for AI cytology, differences among the LBC preservative solutions affects the accuracy of
65 cell detection.[14] Laboratory cytologic techniques include smears, cytocentrifugation-based methods,
66 aspiration cytology, and staining methods such as Papanicolaou and Giemsa stain. These methods may
67 also affect AI cell detection. In this study, we used cell lines to clarify differences in cytologic form
68 and to eliminate biases associated with clinical samples; furthermore, although it lacks clinical
69 applications, we examined the relationship between cell detection or cell classification by AI and
70 processing techniques, which has not been clarified thus far.
71
72 ## MATERIAL AND METHOD
73 ***Image dataset preparation***
74 Cytological preparations were obtained from cultured A549 human lung cancer cell line (LC; RIKEN
75 Cell Bank, Tsukuba, Japan), HeLa human cervical cancer cell line (CC; RIKEN Cell Bank), ACC-
76 MESO-1 human malignant pleural mesothelioma cell line (MM; RIKEN Cell Bank)[20], and KYSE30
77 human esophageal cancer cell line (EC; JCRB Cell Bank, Kanagawa, Japan)[21]. Cell samples were
78 centrifuged at 600 × *g* for 5 min and divided into equal parts to prepare the cytocentrifugation-based
79 preparation (AutoSmear; Sakura Finetek Japan Co., Tokyo, Japan), which is the same principle as

80 cytospin, or LBC preparation using the SurePath manual method. The AutoSmear preparation was
81 centrifuged at 264 × *g* for 2 min, fixed overnight in 95% ethanol, and stained with Papanicolaou. The
82 SurePath manual method was performed as follows: 5 mL CytoRich™ Red (Becton, Dickinson and
83 Company, Franklin Lakes, NJ, USA) was added to the cell sediment and allowed to stand for 1 h. After
84 centrifugation at 600 × *g* for 10 min, 6 mL distilled water was added to the cell sediment and mixed.
85 After centrifugation at 600 × *g* for 5 min, the supernatant was removed, and 1.8 mL of distilled water
86 was added to the sediment and mixed. This solution (300 μL) was dispensed into settling chambers
87 adapted for BD SurePath™ PreCoat slides and allowed to stand for 10 min. The settling chambers
88 were then inverted to discard the supernatant, and the interior of the chambers was washed with 100%
89 ethanol. The settling chambers were inverted again, then removed, and the glass preparations were
90 fixed overnight in 95% ethanol. Prepared specimens were stained with Papanicolaou stain.
91 Cytological images were obtained with Basler USB3 Vision (Basler AG, Ahrensburg,
92 Germany) at 400 × magnification and collected in a 2,592 × 1,944 pixels JPEG format. Images for
93 training, validation, and test sets were obtained consecutively from the same slide without overlapping
94 fields, and no image selection was performed. Images of 1,991 cells were prepared for each cell type,
95 1,440 of which were annotated using the open-source graphical annotation tool labelImg (ver. 1.8.6).
96 Out-of-focus, degenerated, and mitotic cells were excluded.
97 The created training models were: one cell (1C) and four cell (4C) model, in which only one

type of cell line and four types of cell lines were trained, respectively. Details of the models created for each cell type are listed in Table 1. For the 1C model, a training model was created with 900 cells as the training set and 540 cells as the validation set. The 4C model consists of 3,600 cells for the training sets and 2,160 cells for the validation sets, which combine the four types of cell lines used for 1C model creation.

Network Architecture

The object detection algorithm YOLOv5 was used for deep learning. The workstation environment consisted of Windows 10 software (Microsoft, Redmond, WA, USA), Intel Core i9-11900K central processing unit (Intel, Santa Clara, CA, USA), graphics processing unit NVIDIA RTX 3080 (10GB; NVIDIA, Santa Clara, CA, USA), and 64GB of memory. The training conditions were as follows.

YOLOv5 architecture: YOLOv5x

Image size: 640 pixels per inch

Confidence scores: 0.25 (standard value of YOLOv5)

Batch size: 1, 2, 4, 8

In each preparation, four different batch sizes were trained 10 times, and the model with the highest F_1-score from the 40 models was used for this study. F_1-score is the following equations:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Data augmentation, such as vertical and horizontal mirroring, displacement, rotation, and filtering, was not performed in this study.

In the detection, the recognition of a region as the precise location and the appropriate cell type was considered correct, and a detection other than correct was considered incorrect. The detection and classification rates were evaluated by using the following equations:

$$\text{Detection rate} = \frac{\text{Correct}}{\text{Total number of cells}} \times 100$$

$$\text{Classification rate} = \frac{\text{Correct}}{\text{Correct} + \text{Incorrect}} \times 100$$

Statistically significant differences in the detection and classification rates were calculated using Ryan's method, which examines significant differences in the proportion of the population among three or more groups, where $p < 0.01$ was considered significant. Statistical analyses were performed using the StatFlex software (version 6.0; Artech Co., Ltd., Osaka, Japan).

RESULT

Comparison of cytomorphological analysis

The cytological findings for the four cell lines are represented in figure 1. In all cell types, the AutoSmear preparation showed numerous anisocytosis, flat cells, and clear intranuclear structures. Conversely, the LBC preparation revealed three-dimensional cells with rounded edges, and the nuclear-cytoplasmic boundaries and intranuclear structures were unclear.

134

135 ***1C models***

136 When preparations using the same processing technique were used for both training and detection in
137 the 1C model, the detection rates differed for the different cell lines, and that of LC was lower than
138 that of the other cell lines ($p < 0.01$) (Figure 2). The models created using the AutoSmear preparation
139 (auto-model) had higher detection rates than those created using the LBC preparation (LBC-model),
140 with a significant difference in the CC, MM, and EC ($p < 0.01$).

141

142 ***4C models***

143 The results of the LCME-Auto model, used for training the four types of cell lines with AutoSmear
144 preparations, and the LCME-LBC model, used for training the four types of cell lines with LBC
145 preparations, are shown in Tables 2 and 3, respectively. The classification rates of the LCME-Auto
146 and LCME-LBC models were 99.6% and 92.8%, respectively. The LCME-Auto model's classification
147 rates were over 99% for all types of cell lines, and the LCME-LBC model's rates were 80.9–100%
148 depending on the type of cell line. The detection rates of the LCME-Auto model were higher than
149 those of the LCME-LBC model for all the cell lines ($p < 0.01$).

150 Comparative analysis of the detection rates of the 1C and 4C models, when different
151 processing techniques were used for training and detecting preparation, revealed that the detection

rates of LC and CC were significantly lower in the 4C model than that in the 1C model, and those of MM and EC were approximately 10% lower in the 4C model (Figure 3).

DISCUSSION

The SurePath method is a specimen-processing technique based on the density gradient, wherein cytological forms are characterized in three-dimensions and there are fewer anisocytosis forms. The cytocentrifugation-based method includes flat cells, anisocytosis, and various cytological forms differentiated by centrifugal force. Using four types of cell lines, cytologic forms in the LBC preparation were small, three-dimensional, and round with a deeper depth of focus and indistinct nucleocytoplasmic boundaries. Conversely, AutoSmear preparations presented a clear boundary between nucleus-cytoplasm, and nucleochromatin and nucleoli were observed in detail. Moreover, there was abundant anisocytosis in the AutoSmear preparation. AutoSmear preparations have been reported to possess a larger cytoplasmic and nuclear area than that in the LBC preparation, and anisocytosis and anisokaryosis are frequently seen.[13] The cytologic form of the four types of cell lines were the same as previously reported.

Recently, automated diagnosis using deep learning has been actively studied.[11,22,23] Nambu et al. developed a system that detects atypical cells in cervical cytology samples and classifies them using the Bethesda system. This system is a two-step algorithm based on two different deep learning

algorithms.[11] Teramoto et al. studied the automated classification of lung cancer from cytological images using deep convolutional neural network (DCNN), and reported that the classification accuracy was 71%, which is equivalent to pathologists.[23] However, most studies have used specimens collected from a limited number of laboratories, and there are concerns about generalizing the developed AI algorithms.[4,6] In this study, we focused on specimen processing techniques that differ between laboratories, and examined the cytologic form in the AutoSmear and LBC preparations, and accuracies of cell detection and classification by deep learning model. The effects of the specimen processing technique on cell detection and classification were also clarified.

In the 1C model, the detection rates, when the same processing techniques were used for training and detection, were higher in the auto-models than the LBC-models in all types of cell lines. In machine learning for digital pathology, if various colors and cytological forms of the target cells are trained, the robustness of the model is increased because of striking cell characteristics.[11,24] As a result, it is considered that the detection rates of the auto-models were high because of the variety of cytological forms on the preparation and clear cytological findings. In this study, differences in detection rates were observed not only among the preparation processing techniques but also among the types of cell lines. Although the detection rate of LC was significantly lower than that of other cell lines, there was no remarkable difference in the LC cytological form compared to that of the other cell lines. However, LC preparations present various cytologic forms. In this study, 1,440 cells were used

188 to create a deep learning model, and the number of training cells may be insufficient to train these
189 forms.

190 In cell detection using a deep learning model, the relationship between the cytologic forms,
191 which is affected by the LBC preservative solution, has been clarified.[14] When the LBC preservative
192 solutions for the training and detection differ, the detection rate is lower, and the accuracy of cell
193 detection using the deep learning model is affected by differences in the cytologic forms. Our results
194 also showed low detection rates when different processing techniques were used for training and
195 detection, and the detection rates of the auto-model and LBC-model decreased 1.5–20.8% and 13.8–
196 14.7%, respectively. In deep learning, there is a significant loss of accuracy if the network is trained
197 on a dataset containing images processed differently than the test set results.[16] Furthermore, if an
198 algorithm with low bias, high generality, and high accuracy is to be created, it should be trained on a
199 dataset from different resources.[14,16,25,26] Therefore, combining the two preparation techniques may
200 improve the accuracy of AI cell detection.

201 In the 4C model, when the specimen prepared using the same processing technique was used
202 for training and detection, the classification rates were over 90%, and slight differences in the cytologic
203 forms could be recognized. In the detection rates of 4C models, 67.7% for the LCME-LBC model was
204 significantly lower than 91.5% for the LCME-Auto model. This may be attributed to the cytological
205 forms of the LBC preparations that have no variety, similar to the 1C model. LBC technology enables

206 the standardization of cytological preparation, but the cytologic form become smaller and
207 rounder.[13,27,28] The cytoplasm of LBC specimens are 63–75% smaller than that of AutoSmear
208 specimens in the mathematical analysis.[13] As LBC is now used in many laboratories, the possibility of
209 cytomorphological changes reducing the accuracy of AI cell detection and classification may impede
210 the development of AI cytology. Wu et al. attempted to classify ovarian cancer into four types: serous,
211 mucinous, endometrioid, and clear cell carcinoma using the DCNN algorithm, and reported that many
212 misclassified cells displayed a common feature lacking cytological characteristics.[24] In another study,
213 cervical cancer was classified into three types: keratinizing squamous, non-keratinizing squamous,
214 and basaloid squamous using DCNN, and concluded that most correctly classified images had a certain
215 number of cells with notable pathological features, such as cell morphology, tissue color, and cell
216 distribution, while misclassified images had poor features.[29] This study also suggested that AI models
217 can easily distinguish cell types using characteristic cytologic form.

218 The detection rates were markedly lower for LC and CC when the different processing
219 techniques were used for training and detection. This is probably because the cytomorphology of LC
220 and CC differed significantly between the AutoSmear and LBC preparations. The AutoSmear
221 preparation of LC was characterized by large, thin, pale cytoplasm and irregular cell edges, whereas
222 the LBC preparation had small and more three-dimensional nuclei and cytoplasm. These differences
223 in morphological characteristics appear to be the cause of the significantly reduced accuracy of cell

discrimination. In addition to the preparation technique, stain/color are also known to affect supervised algorithm performance to a considerable extent.[16] Thus, it is necessary to understand that algorithm performance is affected by many factors, and the standardization of preparation technique and staining are major challenges in AI cytology.

Ozturk et al. developed a deep learning model called COVID Net model and compared a three-class classification of COVID-19 pneumonia, pneumonia, and no findings with a binary classification of COVID-19 pneumonia and no findings, where they reported that the accuracy of binary classification was superior to that of three-class classification.[30] In our study, the 1C model is a binary classification of cell or non-cell, while the 4C model is four-class classification. Therefore, the detection rate of the 4C model was significantly reduced, because it was necessary to extract the characteristics of each cell type. In contrast, the detection rates of MM and EC showed no significant differences between the 1C and 4C models. This suggests that the cytomorphological changes in MM and EC are slight due to the type of processing technique used, indicating that the degree of change in morphology varies according to the type of cell line. However, numerous cell types (e.g., normal cells, malignant cells, and non-cellular elements) must be classified for practical applicability in clinical cytology. A two-step algorithm has also been developed,[11] as accuracy decreases as more cell types are classified.[30] Task-specific algorithms, such as those that only detect malignant cells, those that detect non-cellular components, and those that differentiate specific cells, may be needed to implement

AI cytology.

Although AI cell detection depends on the cell type, the accuracy of cell detection using deep learning is affected by the specimen processing technique, and its accuracy is reduced when different processing techniques are used for training and detection. Additionally, as the number of cell types used to train the model increases, the detection rate decreases significantly. In the cell detection and classification using a deep learning model, attention should be paid to cells whose cytological form changes depending on the processing technique, and the processing technique for creating the training model should be considered.

FIGURE LEGENDS

Fig. 1. Cytological features of four types of cell line.

Upper panels show the AutoSmear preparation and lower panels show the LBC preparation (Papanicolaou stain, ×1,000). LC, lung cancer cell line; CC, cervical cancer cell line; MM, malignant pleural mesothelioma cell line; EC, esophageal cancer cell line; LBC, liquid-based cytology.

Fig. 2. Detection rates of the one cell model.

When the same processing technique preparations were used for training and detection, the detection rates of the AutoSmear models were higher than those of the LBC model for all types of cell lines, and there was a significant difference between the CC, MM, and EC ($p < 0.01$). LC had a lower detection rate than the other cell lines ($p < 0.01$). LC, lung cancer cell line; CC, cervical cancer cell line; MM, malignant pleural mesothelioma cell line; EC, esophageal cancer cell line; LBC, liquid-based cytology.

Fig. 3. Comparison of detection rates of the one-cell and four-cell models.

When different processing techniques were used for training and detection, the detection rates of LC and CC were significantly lower in the 4C model, whereas those of MM and EC tended to be approximately 10% lower in the 4C model. LC, lung cancer cell line; CC, cervical cancer cell line; MM, malignant pleural mesothelioma cell line; EC, esophageal cancer cell line; LBC, liquid-based

268 cytology; 1C, one cell; 4C, four cell.

FUNDING STATEMENT:

This work was supported by JSPS KAKENHI [JP21K18077].

ACKNOWLEDGEMENTS:

We would like to thank Editage (www.editage.jp) for their assistance with English language editing.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethics approval was not applicable to the present study as clinical samples were not used. The cell lines used in this study were obtained from the RIKEN Cell Bank and JCRB Cell Bank. Their use did not require approval from the ethical committee because the study did not involve genetic analysis or modification.

AUTHOR CONTRIBUTION STATEMENT

K.I. conceptualized and designed the study; K.I. and S.M. wrote the manuscript; S.M. collected specimens, annotated data, and analyzed data; K.I. and S.S. contributed to algorithm development and programming; N.S. and S.M. performed cell culture techniques; C.I. and Y.S. reviewed and edited the manuscript. All authors read and approved the final paper.

DATA AVAILABILITY STATEMENT:

All data supporting the findings of this study are available from the corresponding authors upon reasonable request.

REFERENCES

1. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. Adv Exp Med Biol. 2020;1213:3-21. DOI: 10.1007/978-3-030-33128-3_1.

2. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J Pathol Inform. 2016;7:29. DOI: 10.4103/2153-3539.186902.

3. Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. J Pathol Inform. 2018;9:38. DOI: 10.4103/jpi.jpi_53_18.

4. McAlpine ED, Pantanowitz L, Michelow PM. Challenges developing deep learning algorithms in cytology. Acta Cytol. 2021;65(4):301-309. DOI: 10.1159/000510991.

5. Landau MS, Pantanowitz L. Artificial intelligence in cytopathology: a review of the literature and overview of commercial landscape. J Am Soc Cytopathol. 2019;8(4):230-241. DOI: 10.1016/j.jasc.2019.03.003.

6. Lew M, Wilbur DC, Pantanowitz L. Computational cytology: lessons learned from Pap test computer-assisted screening. Acta Cytol. 2021;65(4):286-300. DOI: 10.1159/000508629.

7. Dey P. The emerging role of deep learning in cytology. Cytopathology. 2021;32(2):154-160. DOI: 10.1111/cyt.12942.

8. Chollet F. Deep Learning with Python. 2nd. Greenwich, United Kingdom: Manning

Publications Co; 2021.

9. Wataya T, Nakanishi K, Suzuki Y, Kido S, Tomiyama N. Introduction to deep learning: minimum essence required to launch a research. Jpn J Radiol. 2020;38(10):907-921. DOI: 10.1007/s11604-020-00998-2.

10. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. Preprint at https://arxiv.org/abs/1506.02640.

11. Nambu Y, Mariya T, Shinkai S, et al. A screening assistance system for cervical cytology of squamous cell atypia based on a two-step combined CNN algorithm with label smoothing. Cancer Med. 2022;11(2):520-529. DOI: 10.1002/cam4.4460.

12. Ye MH, Chen WY, Cai BJ, Jin CH, He XL. A convolutional neural network based model for assisting pathological diagnoses on thyroid liquid-based cytology. Zhonghua Bing Li Xue Za Zhi. 2021;50(4):358-362. DOI: 10.3760/cma.j.cn112151-20200802-00613.

13. Ikeda K, Oboshi W, Hashimoto Y, et al. Characterizing the effect of processing technique and solution type on cytomorphology using liquid-based cytology. Acta Cytol. 2022;66(1):55-60. DOI: 10.1159/000519335.

14. Ikeda K, Sakabe N, Maruyama S, et al. Relationship between liquid-based cytology preservative solutions and artificial intelligence: Liquid-based Cytology specimen cell detection using YOLOv5 deep convolutional neural network. Acta Cytol. 2022:1-9. DOI:

10.1159/000526098.

15. Abels E, Pantanowitz L, Aeffner F, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. J Pathol. 2019;249(3):286-294. DOI: 10.1002/path.5331.

16. Marée R. The need for careful data collection for pattern recognition in digital pathology. J Pathol Inform. 2017;8:19. DOI: 10.4103/jpi.jpi_94_16.

17. Laine RF, Arganda-Carreras I, Henriques R, Jacquemet G. Avoiding a replication crisis in deep-learning-based bioimage analysis. Nat Methods. 2021;18(10):1136-1144. DOI: 10.1038/s41592-021-01284-3.

18. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23(1):40-55. DOI: 10.1038/s41580-021-00407-0.

19. Luo Y, She DL, Xiong H, Yang L, Fu SJ. Diagnostic value of liquid-based cytology in urothelial carcinoma diagnosis: A systematic review and meta-analysis. PLOS ONE. 2015;10(8):e0134940. DOI: 10.1371/journal.pone.0134940.

20. Usami N, Fukui T, Kondo M, et al. Establishment and characterization of four malignant pleural mesothelioma cell lines from Japanese patients. Cancer Sci. 2006;97(5):387-394. DOI: 10.1111/j.1349-7006.2006.00184.x.

21. Shimada Y, Imamura M, Wagata T, Yamaguchi N, Tobe T. Characterization of 21 newly

established esophageal cancer cell lines. Cancer. 1992;69(2):277-284. DOI: 10.1002/1097-0142(19920115)69:2<277::aid-cncr2820690202>3.0.co;2-c.

22. Guan Q, Wang Y, Ping B, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J Cancer. 2019;10(20):4876-4882. DOI: 10.7150/jca.28769.

23. Teramoto A, Tsukamoto T, Kiriyama Y, Fujita H. Automated classification of lung cancer types from cytological images using deep convolutional neural networks. BioMed Res Int. 2017;2017:4067832. DOI: 10.1155/2017/4067832.

24. Wu M, Yan C, Liu H, Liu Q. Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. Biosci Rep. 2018;38(3):BSR20180289. DOI: 10.1042/BSR20180289.

25. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun (Lond). 2020;40(4):154-166. DOI: 10.1002/cac2.12012.

26. Lucas AM, Ryder PV, Li B, Cimini BA, Eliceiri KW, Carpenter AE. Open-source deep-learning software for BioImage segmentation. Mol Biol Cell. 2021;32(9):823-829. DOI: 10.1091/mbc.E20-10-0660.

27. Hoda RS. Non-gynecologic cytology on liquid-based preparations: a morphologic review of

facts and artifacts. Diagn Cytopathol. 2007;35(10):621-634. DOI: 10.1002/dc.20698.

28. Elsheikh TM, Kirkpatrick JL, Wu HH. Comparison of ThinPrep and cytospin preparations in the evaluation of exfoliative cytology specimens. Cancer. 2006;108(3):144-149. DOI: 10.1002/cncr.21841.

29. Wu M, Yan C, Liu H, Liu Q, Yin Y. Automatic classification of cervical cancer from cytological images by using convolutional neural network. Biosci Rep. 2018;38(6):BSR20181769. DOI: 10.1042/BSR20181769.

30. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020;121:103792. DOI: 10.1016/j.compbiomed.2020.103792.

Table 1. Deep Learning datasets and model metrics.

model	cell line	preparation	epoch	mAP	F_1-score
one cell model					
LC-Auto	A549	AutoSmear	418	0.843	0.827
LC-LBC	A549	LBC	364	0.800	0.794
CC-Auto	HeLa	AutoSmear	340	0.833	0.841
CC-LBC	HeLa	LBC	332	0.602	0.844
MM-Auto	ACC-MESO-1	AutoSmear	511	0.782	0.802
MM-LBC	ACC-MESO-1	LBC	223	0.871	0.644
EC-Auto	KYSE30	AutoSmear	435	0.894	0.911
EC-LBC	KYSE30	LBC	220	0.832	0.827
four cell model					
LCME-Auto	A549, HeLa, ACC-MESO-1, KYSE30	AutoSmear	460	0.848	0.854
LCME-LBC	A549, HeLa, ACC-MESO-1, KYSE30	LBC	249	0.735	0.746

LC, lung cancer cell line; CC, cervical cancer cell line; MM, malignant pleural mesothelioma cell line; EC, esophageal cancer cell line; Auto, AutoSmear; LBC, liquid-based cytology; mAP, mean average precision

Table 2. Detection and classification rate of four cell AutoSmear model.

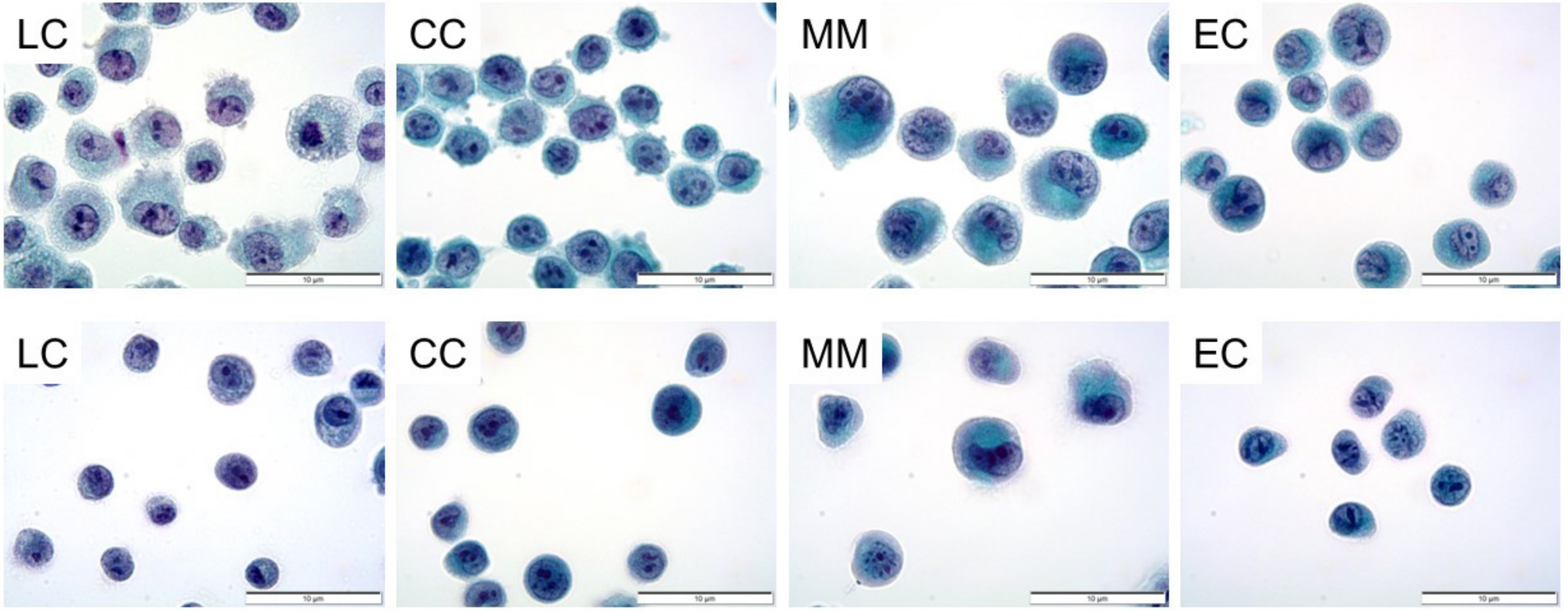
Model		LCME-Auto					
Preparation							
Training		AutoSmear				Detection rate (%)	Classification rate (%)
Detection		AutoSmear					
Cell line		LC	CC	MM	EC		
TRUE	LC	440	0	3	0	79.9 (440/551)	99.3 (440/443)
	CC	0	506	0	1	91.8 (506/551)	99.8 (506/507)
	MM	4	0	525	0	95.3 (525/551)	99.2 (525/529)
	EC	0	0	0	546	99.1 (546/551)	100.0 (546/546)
total						91.5 (2017/2204)	99.6 (2017/2025)

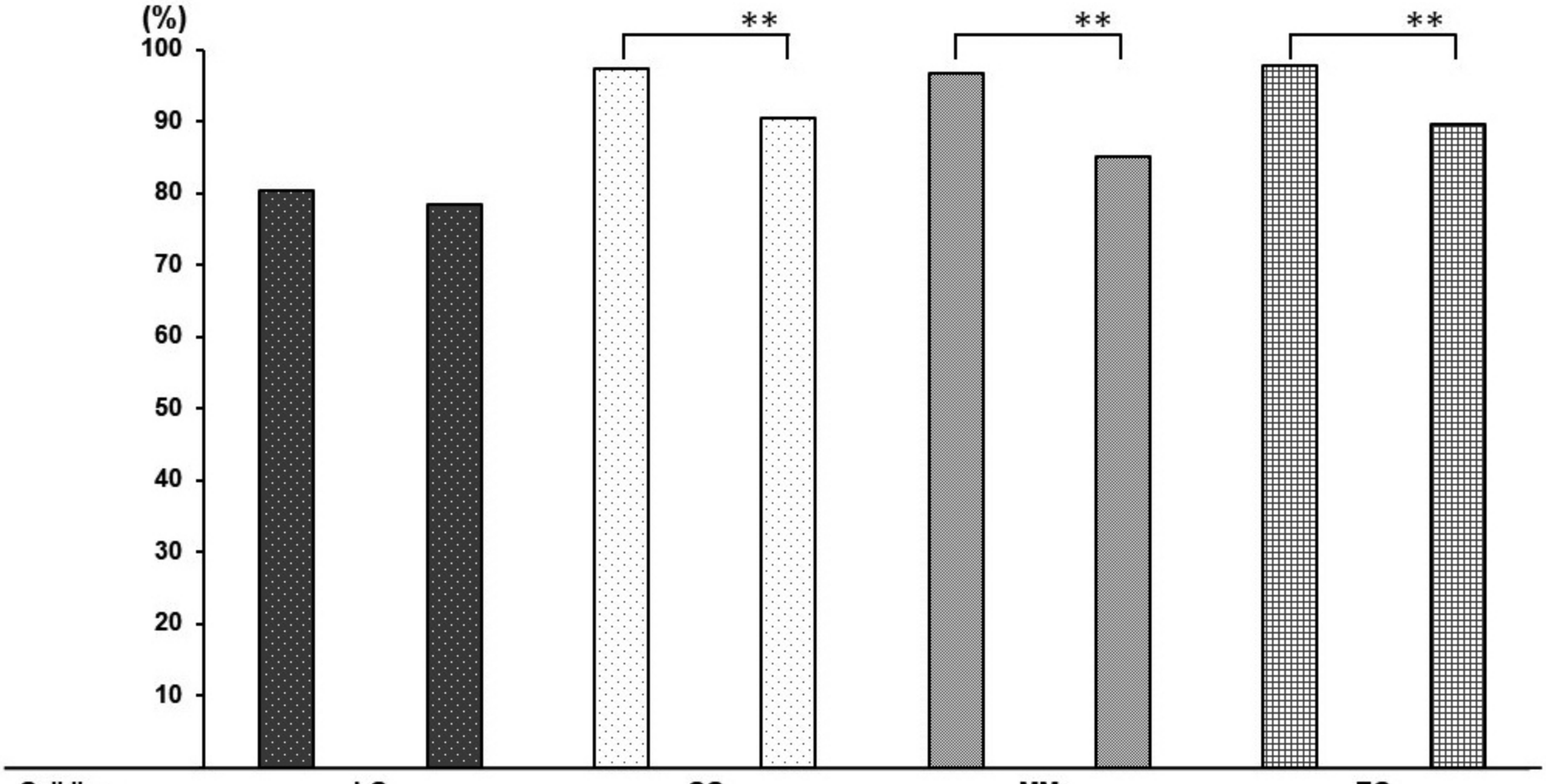
LC, lung cancer cell line; CC, cervical cancer cell line; MM, malignant pleural mesothelioma cell line; EC, esophageal cancer cell line; Auto, AutoSmear

Table 3. Detection and classification rate of four cell LBC model.

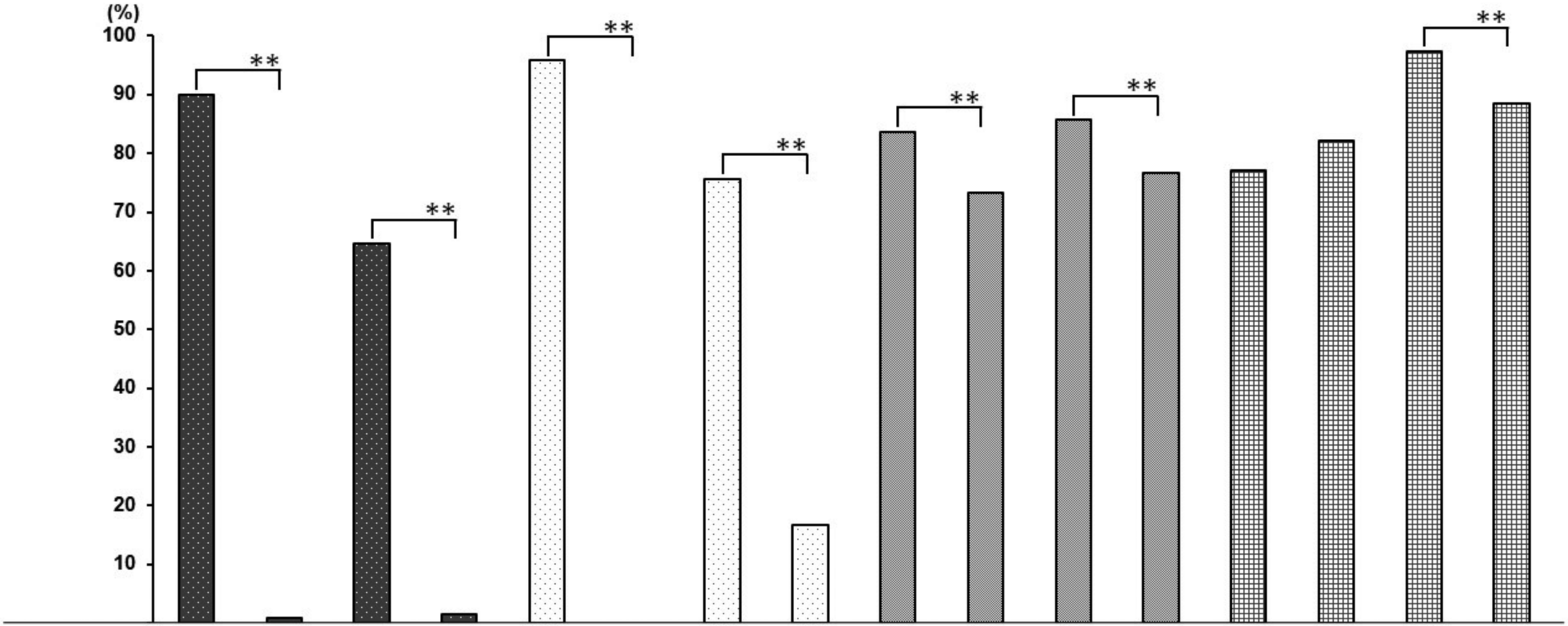
Model		LCME-LBC					
Preparation							
Training		LBC				Detection rate (%)	Classification rate (%)
Detection		LBC					
Cell line		LC	CC	MM	EC		
TRUE	LC	294	0	22	0	53.4 (294/551)	93.0 (294/316)
	CC	0	338	80	0	61.3 (338/551)	80.9 (338/418)
	MM	0	0	374	0	67.9 (374/551)	100.0 (374/374)
	EC	0	12	2	486	88.2 (486/551)	97.2 (486/500)
total						67.7 (1492/2204)	92.8 (1492/1608)

LC, lung cancer cell line; CC, cervical cancer cell line; MM, malignant pleural mesothelioma cell line; EC, esophageal cancer cell line; LBC, liquid-based cytology


LC
CC
MM
EC
LC
CC
MM
EC



Cell line	LC		CC		MM		EC	
Preparation								
Training	AutoSmear	LBC	AutoSmear	LBC	AutoSmear	LBC	AutoSmear	LBC
Detection	AutoSmear	LBC	AutoSmear	LBC	AutoSmear	LBC	AutoSmear	LBC
Detection rate	80.4	78.4	97.3	90.4	96.7	85.1	97.8	89.7
	(443/551)	(432/551)	(536/551)	(498/551)	(533/551)	(469/551)	(539/551)	(494/551)



Cell line	LC				CC				MM				EC			
Preparation																
Training	AutoSmear		LBC		AutoSmear		LBC		AutoSmear		LBC		AutoSmear		LBC	
Detection	LBC		AutoSmear		LBC		AutoSmear		LBC		AutoSmear		LBC		AutoSmear	
	1C	4C	1C	4C	1C	4C	1C	4C	1C	4C	1C	4C	1C	4C	1C	4C
Detection rate	90.0	0.9	64.6	1.5	95.8	0.0	75.7	16.7	83.7	73.3	85.7	76.6	77.0	82.2	97.3	88.4
	(496/551)	(5/551)	(356/551)	(8/551)	(528/551)	(0/551)	(417/551)	(92/551)	(461/551)	(404/551)	(472/551)	(422/551)	(424/551)	(453/551)	(536/551)	(487/551)