RESEARCH ARTICLE

Editorial Process: Submission:11/01/2021 Acceptance:04/15/2022

Comparison of Fine-Tuned Deep Convolutional Neural Networks for the Automated Classification of Lung Cancer Cytology Images with Integration of Additional Classifiers

Tetsuya Tsukamoto[1]*, Atsushi Teramoto[2], Ayumi Yamada[2], Yuka Kiriyama[1,3], Eiko Sakurai[1], Ayano Michiba[1], Kazuyoshi Imaizumi[4], Hiroshi Fujita[5]

Abstract

Objective: It is essential to accurately diagnose and classify histological subtypes into adenocarcinoma (ADC), squamous cell carcinoma (SCC), and small cell lung carcinoma (SCLC) for the appropriate treatment of lung cancer patients. However, improving the accuracy and stability of diagnosis is challenging, especially for non-small cell carcinomas. The purpose of this study was to compare multiple deep convolutional neural network (DCNN) technique with subsequent additional classifiers in terms of accuracy and characteristics in each histology. **Methods:** Lung cancer cytological images were classified into ADC, SCC, and SCLC with four fine-tuned DCNN models consisting of AlexNet, GoogLeNet (Inception V3), VGG16 and ResNet50 pretrained by natural images in ImageNet database. For more precise classification, the figures of 3 histological probabilities were further applied to subsequent machine learning classifiers using Naïve Bayes (NB), Support vector machine (SVM), Random forest (RF), and Neural network (NN). **Results:** The classification accuracies of the AlexNet, GoogLeNet, VGG16 and ResNet50 were 74.0%, 66.8%, 76.8% and 74.0%, respectively. Well differentiated typical morphologies were tended to be correctly judged by all four architectures. However, poorly differentiated non-small cell carcinomas lacking typical structures were inclined to be misrecognized in some DCNNs. Regarding the histological types, ADC were best judged by AlexNet and SCC by VGG16. Subsequent machine learning classifiers of NB, SVV, RF, and NN improved overall accuracies to 75.1%, 77.5%, 78.2%, and 78.9%, respectively. **Conclusion:** Fine-tuning DCNNs in combination with additional classifiers improved classification of cytological diagnosis of lung cancer, although classification bias could be indicated among DCNN architectures.

Keywords: Liquid-based cytology- artificial intelligence- machine learning

Asian Pac J Cancer Prev, **23 (4)**, 1315-1324

Introduction

Lung cancer remains a primary cause of death in both males and females worldwide(American Cancer Society, 2015). In order to improve the survival rate of lung cancer patients, it is essential to identify the characteristics of lung cancer. Lung cancer has been classified into two categories, namely non-small cell lung cancer and small cell lung cancer. Recent improvements in chemotherapy and radiation therapy (Baas et al, 2006) have resulted in the former being further classified into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Travis et al, 2015). In clinical practice, it is often difficult to precisely differentiate adenocarcinoma and squamous cell carcinoma in terms of their morphological characteristics, thus requiring immunohistochemical evaluation. Cytodiagnosis is advantageous for the cytological evaluation of small cell carcinoma compared to histological specimen, often showing crushed small cell cancer cells. To obtain accurate diagnosis result, a combination of cytological evaluation and histopathological diagnosis is essential. However, the cell morphology of these three types of cancer cells varies. Here we focused on the automated classification technique of cancer types using cytological images. Among the four major types of carcinoma, the large cell carcinoma is the easiest to detect because of its cell morphology. We therefore concentrate on the classification of adenocarcinoma, squamous cell carcinoma, and small cell carcinoma that are sometimes confused with each

*[1]Department of Diagnostic Pathology, Fujita Health University School of Medicine, Toyoake, Japan. [2]Faculty of Radiological Technology, Fujita Health University School of Health Sciences, Toyoake, Japan. [3]Department of Diagnostic Pathology, Narita Memorial Hospital, Toyohashi, Japan. [4]Department of Respiratory Medicine, Fujita Health University School of Medicine, Toyoake, Japan. [5]Faculty of Engineering, Gifu University, Gifu, Japan. *For Correspondence: ttsukamt@fujita-hu.ac.jp. Tetsuya Tsukamoto and Atsushi Teramoto have equal contribution in this study.*

other in the cytological specimen.

In our previous studies (Teramoto et al, 2017; Teramoto et al, 2019), an automated classification scheme of lung cancer types in the cytology images using DCNN was developed. The original DCNN model was employed and trained using 15,000 cytological images, and the classification accuracy was found to be 71%. However, for clinical use, further improvement in the classification ability is necessary. As an improvement method, increase in the number of data used for training can be mentioned. The lung cytological specimen has a three-dimensional structure in which cells overlap each other, acquisition of images by a uniform scanning method such as a whole slide scanner is difficult, and observation by a conventional microscope is mainstream. Therefore, collection of many cytological images is more difficult than that of histopathological images.

Transfer learning is a technique known for improving the performance of deep learning (Ravishankar et al, 2016). This is a method of diverting the majority of the deep learning architecture learned using a large number of natural images and performing another classification task. To further improve the accuracies, additional supervised classifiers (Amancio et al, 2014) have been utilized to support the DCNNs. In this study, we evaluated four pretrained DCNN architectures plus subsequent four supervised classifiers for comparison of classification accuracy of lung cancer tissue types.

Materials and Methods

Image Dataset

Cancer cells were collected from 55 cases by exfoliative or interventional cytology under bronchoscopy or CT-guided fine-needle aspiration cytology. The 55 cases comprised 36 cases of adenocarcinoma, 14 cases of squamous cell carcinoma, and five cases of small cell carcinoma. Final diagnosis was made in all cases via a combination of histopathological and immunohistochemical diagnosis. Specifically, biopsy tissues, simultaneously collected with cytology specimen, were fixed using 10% neutral buffered formalin, dehydrated, and embedded in paraffin. The 3-μm tissue sections were subjected to immunohistochemical analysis for some cases. Cancer lesions were immunohistochemically identified as adenocarcinoma if TTF-1 (8G7G3/1, DAKO, Glostrup, Denmark) and/or napsin A (IP64, Novocastra, Leica Biosystems, New Castle upon Tyne, UK) were positive and as squamous cell carcinoma if p40 (BC28, Roche Diagnostics, Basel, Switzerland) and/or cytokeratin 5/6 (D5/16 B4, DAKO) were present. Positivity of neuroendocrine markers including chromogranin A (rabbit polyclonal, DAKO), synaptophysin (MRQ-40, Roche), and CD56 (MRQ-42, Roche) was suggestive of small cell carcinoma. Immunohistochemical procedures were performed using the iView DAB Universal Kits run on Ventana Benchmark Ultra (Roche) as described elsewhere (Tsukamoto et al., 2017).

The cytological specimens were prepared with a liquid-based cytology (LBC) system using BD SurePath liquid-based Pap Test (Beckton Dickinson, Durham, NC, USA) and were stained using the Papanicolaou method (Teramoto et al, 2017). Some LBC specimens were immunostained like histological sections. Using a digital still camera (DP70, Olympus, Tokyo, Japan) attached to a microscope (BX51, Olympus) with ×40 objective lens, 82 images of adenocarcinoma, 125 images of squamous cell carcinoma, and 91 images of small cell carcinoma were collected in JPEG format. The initial matrix size of each JPEG image was 2,040 × 1,536 pixels. Subsequently, 768 × 768 pixels square images were generated by cropping and were further resized to 256 × 256 pixels. Duplicate 768 × 768 pixels square images were cut from the original image to avoid overlap therefrom. Finally, they were resized to 256 × 256 pixels.

Data augmentation

DCNN training requires a large amount of data, as a small dataset may cause overfitting. To prevent overfitting, we augmented the training dataset using image processing. The microscopic images are direction invariant, and the sharpness of the target cell in each image varies according to the position of the focal plane of the microscope. Moreover, staining varies depending on the specimen, and color balance fluctuates by the automated white balance function. Therefore, image data were augmented by rotating, flipping, and filtering the original images as shown in Figure 1 (Teramoto et al., 2017).

First, images were flipped to obtain a twofold increase in the number of images. A Gaussian filter, with standard deviation of Gaussian kernel being three pixels, and a convolutional edge enhancement filter, with center weight 5.4 and the eight surrounding weights of −0.55, were applied to the images. Then, images were rotated at 90°, 180°, and 270°. Furthermore, color augmentation was applied by adding random RGB offsets to the images, generating images with different color balance. The number of images augmented by color adjustment was determined such that the total number of augmented images was the same for the three output classes.

Transfer learning using major DCNN models

DCNN shows excellent performance for image classification. In the recent ImageNet Large Scale Visual Recognition Challenge (ILSVRC), CNN architectures compete to correctly classify objects. In the competition so far, several excellent DCNN architectures have been proposed. In this study, we compared the following four architectures originally trained to classify color photographs using ImageNet database and retrained them using cytological images to classify the lung cancer types (Ravishankar et al., 2016).

AlexNet

AlexNet was designed by the SuperVision group, consisting of Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever (Krizhevsky et al., 2012). It significantly outperformed all the prior competitors and won the ILSVRC 2012. This network has a deeper structure compared with the conventional feed forward neural networks. It consists of five convolution layers, three pooling layers, and three fully connected layers as shown

in Figure 2A.

GoogLeNet (Inception V3)

The winner of the ILSVRC 2014 was GoogLeNet developed by Google (Szegedy et al, 2015). The key feature of GoogLeNet is to introduce the small network called inception module. Several inception modules are connected together to go deeper as shown in Figure 2B.

VGG16

VGG16 was designed by Simonyan et al. in the Visual Geometry Group (Simonyan and Zisserman, 2015), and it showed excellent performance in the ILSVRC 2014. It consists of 13 convolution and pooling layers and 3 fully connected layers as shown in Figure 2C. Because the architecture is simple and publicly available with the weight configuration, VGG16 has been used in many other applications (Bychkov et al., 2018; Qu et al., 2018).

ResNet50

ResNet won the ILSVRC 2015 developed by Kaiming et al. in the Microsoft research (He et al., 2016). It employed a residual structure that has shortcut connection to ease the training of deeper networks. Kaiming proposed several architectures with numerous network layers. In this study, we employed ResNet50 with 50 layers as shown in Figure 2D.

Transfer learning using pretrained networks

Based on the pretrained model of the above-mentioned four DCNNs trained using ImageNet, which is a database composed of a large number of natural images, retrain was performed using collected and augmented cytological images. For classification, we replaced the fully connected layers of the original DCNN models with three layers having 1024, 256, and 3 units (corresponding to adenocarcinoma, squamous cell carcinoma, and small cell carcinoma). The output was given by a Softmax layer. The parameters of the replaced layers were randomly initialized, and our fine-tuning was performed only for these layers.

Training and evaluation

By dividing the images randomly into three datasets, the classification performances of the four DCNN models were evaluated using three-fold cross-validation. Here, images of the same case were included in the same dataset. The number of cases and images of the three datasets are listed in Table 1.

The DCNN training and evaluation were implemented using the Keras 2.12 and TensorFlow 1.4 on a computer with XEON CPU and NVIDIA Quadro P5000 for GPU, taking 10 h per dataset for training using the augmented images. Training was conducted using a mini-batch size of 32, optimization based on stochastic gradient descent, learning rate of 0.0001, and momentum of 0.9.

Inter-architectural agreements

Cohen's kappa coefficients were calculated between every two architectures as described (Cohen, 1960).

Integration of additional supervised classifiers

The output value of each histological type drawn in 4 DCNNs were utilized in additional classifiers, including Naïve Bayes, Support vector machine, Random forest, and Neural network, to further classify the lung cancer types to improve the classification accuracy using Orange software version 3.18 (Demsar et al, 2013). The parameters of these classifiers are listed in Table 3.

Results

Comparison of classification accuracy of DCNNs

For the classification of the three cancer types, four DCNN models were trained and evaluated using augmented data. Table 2 shows the confusion matrices of each DCNN model. From these results, the classification accuracy was found to be exceeded 70% in AlexNet, VGG16, and ResNet50, whereas as low as 66.8% in the case of Inception V3. The classification accuracy of VGG16 was the highest at 76.8%, approximately 5% improvement compared with our previous method (Teramoto et al, 2017). Regarding each histological type, AlexNet was the best for adenocarcinoma and small cell carcinoma. On the other hand, VGG16 was the best for squamous cell carcinoma.

Inter-architectural concordance was evaluated with Cohen's kappa coefficients. All the combinations with Inception V3 revealed values less than 0.6 compared with other combinations that revealed values as high as 0.715 ± 0.035 between AlexNet and VGG16 (Table 4).

Correlation of morphology and prediction

Representative images of adenocarcinomas and the predicted results of four architectures are shown in Figure 3. Images with moderate sized cells/nuclei, single nucleoli, and tubular or papillary cell mass were identified as adenocarcinoma (Figure 3, A-D). On the other hand, images with small cells/nuclei (Figure 3E) were misjudged as small cell carcinoma, whereas images with larger cells/nuclei with rough nuclear matrix (Figure 3F) were misdiagnosed as squamous cell carcinoma. Images

Table 1. Number of Images in Each Dataset for Cross Validation

	SET 1		SET 2		SET 3		Total number of original images
	Original	Augmented	Original	Augmented	Original	Augmented	
Adenocarcinoma	23	5000	27	5000	25	5000	75
Squamous cell carcinoma	44	5000	40	5000	46	5000	130
Small cell carcinoma	26	5000	26	5000	32	5000	84
Total							289

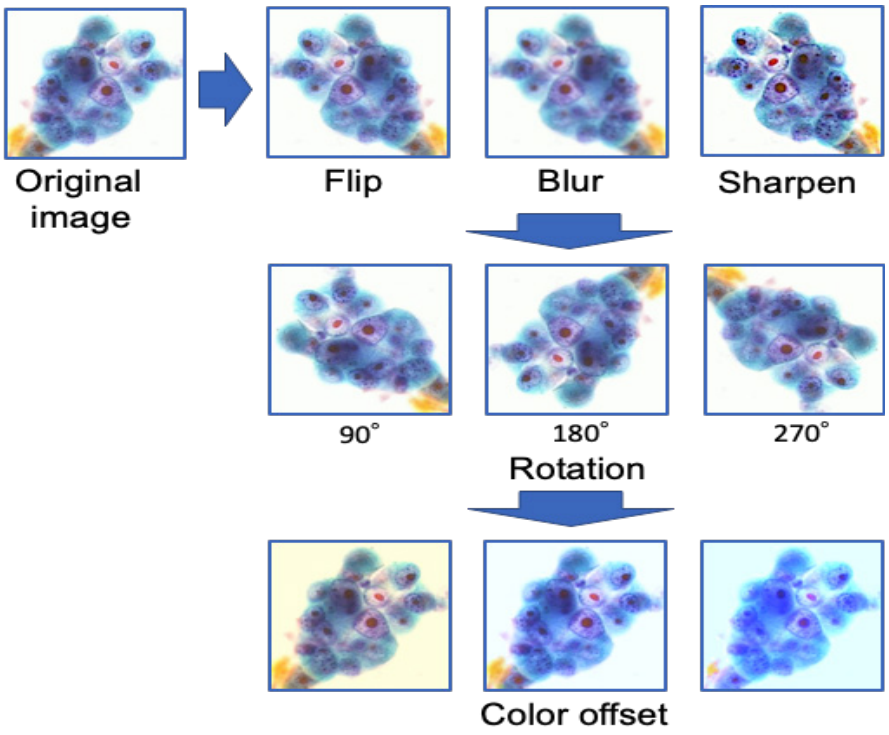

Figure 1. Data Augmentation Methods

with focused and unfocused cells might be predicted as adenocarcinoma by AlexNet.

Squamous cell carcinomas (Figure 4) are characterized by Orange G prone colored keratinized cytoplasm in a well differentiated type. Cancer cells with larger and rougher nuclear matrix, multiple nucleoli, flat cell mass, and spindle cell morphology were correctly diagnosed as squamous cell carcinoma (Figure 4, A-C). Those with smaller cell/nuclei were misjudged as small cell carcinoma (Figure 4E). Three-dimensional cell mass or prominent nucleoli (Figure 4, D-F) were misclassified as adenocarcinoma.

Small isolated cells with granular nuclear matrix were identified correctly as small cell carcinoma (Figure 5, A-C). Planar or three-dimensional cell masses were identified as squamous cell carcinoma or adenocarcinoma,

Table 2. Confusion Matrix of Classification Results by Fine-Tuned DCNN Architectures

AlexNet			Predicted		Total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	57 (76.0%)	10 (13.3%)	8 (10.7%)	75 (100%)	
	SqCC	37 (28.5%)	75 (57.7%)	18 (13.8%)	130 (100%)	
	SmCLC	1 (1.2%)	2 (2.4%)	81 (96.4%)	84 (100%)	73.7%
GoogLeNet (InceptionV3)			Predicted		Total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	46 (61.3%)	17 (22.7%)	12 (16.0%)	75 (100%)	
	SqCC	36 (20.0%)	80 (61.5%)	14 (10.8%)	130 (100%)	
	SmCLC	13 (15.5%)	4 (4.8%)	67 (79.8%)	84 (100%)	66.8%
VGG16			Predicted		Total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	56 (74.7%)	12 (16.0%)	7 (9.3%)	75 (100%)	
	SqCC	25 (19.2%)	89 (68.5%)	16 (12.3%)	130 (100%)	
	SmCLC	3 (3.6%)	4 (4.8%)	77 (91.7%)	84 (100%)	76.8%
ResNet50			Predicted		Total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	56 (74.7%)	14 (18.7%)	5 (6.7%)	75 (100%)	
	SqCC	30 (23.1%)	88 (67.7%)	12 (9.2%)	130 (100%)	
	SmCLC	5 (6.0%)	9 (10.7%)	70 (83.3%)	84 (100%)	74.0%

ADC, adenocarcinoma; SqCC, squamous cell carcinoma; SmCLC, small cell lung carcinoma

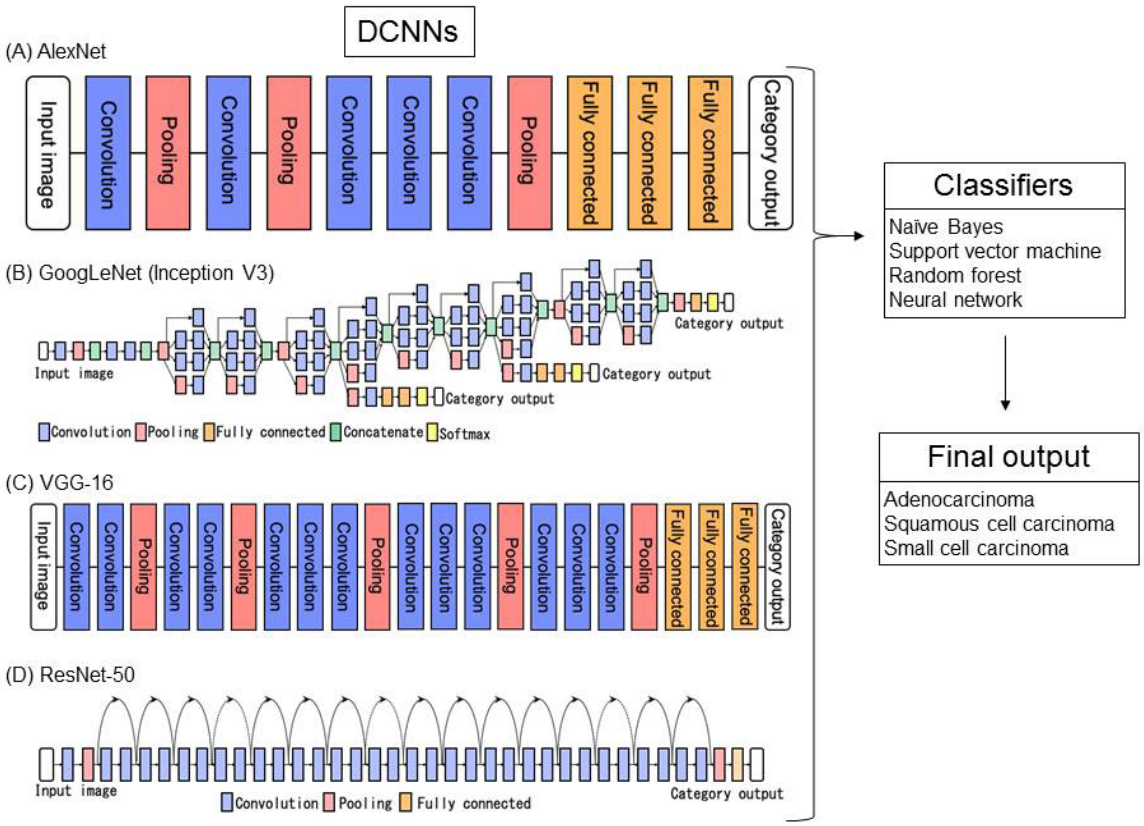

Figure 2. Architectures of Deep Convolutional Neural Networks and Subsequent Classifiers Used in this Study

respectively (Figure 5, D-F), even though being obtained from a same patient.

Practically, some adenocarcinomas and squamous cell carcinoma could not be distinguished correctly and would be compromised as non-small cell carcinomas. Figure 6 shows the resembling cases for adenocarcinoma and squamous cell carcinoma. Figure 6A has planar distribution and wide cytoplasm but has peripheral nuclei and prominent single nucleoli. Immunohistochemical analysis revealed this case as an adenocarcinoma. Figure 6B, squamous cell carcinoma being positive for p40, was morphologically misjudged as adenocarcinoma by all four architectures.

Some cases show variable cell morphologies even from one patient. Figure 7 shows Orange G colored cells, a characteristic of squamous cell carcinoma, whereas the second and the third cells are not typical of squamous cell carcinoma, being misjudged by all four DCNNs. Immunohistochemical analysis confirmed the cell type as squamous cell carcinoma.

Table 3. Parameters of the Used Classifiers

	Parameter	Value
Naïve Bayes	-	-
Support vector machine	Regression loss epsion	0.1
	Cost	1.00
	Kernel	Polynomial
Random forest	Number of trees	20
Neural network	Number of Hidden layers	2 (100-100)
	Activation function	tanh
	Optimizer	SGD
	Number of iteration	300

Effect of additional machine learning classifiers

After processing the image classification with pre-trained DCNNs, additional machine learning techniques were further employed to improve the accuracy (Table 5). Neural network revealed best overall accuracy of 78.9% among 4 classifiers.

Discussion

In this study, we have evaluated four fine-tuned DCNNs for the classification of lung cancer types. AlexNet, VGG16, and ResNet50 revealed that more than 73% of cases were classified correctly compared with our previous result (71.1%), whereas Inception V3 showed the worse figure of 66.8%. The combinations with the former three architectures showed better concordance

Table 4. Kappa Coefficients between Every 2 Fine-Tuned DCNNs.

	AlexNet	Inception V3	VGG16	ResNet50
AlexNet	–	0.565±0.040	0.715±0.035	0.654±0.037
Inception V3	–	–	0.543±0.041	0.548±0.041
VGG16	–	–	–	0.672±0.037
ResNet50	–	–	–	–

–, not applicable

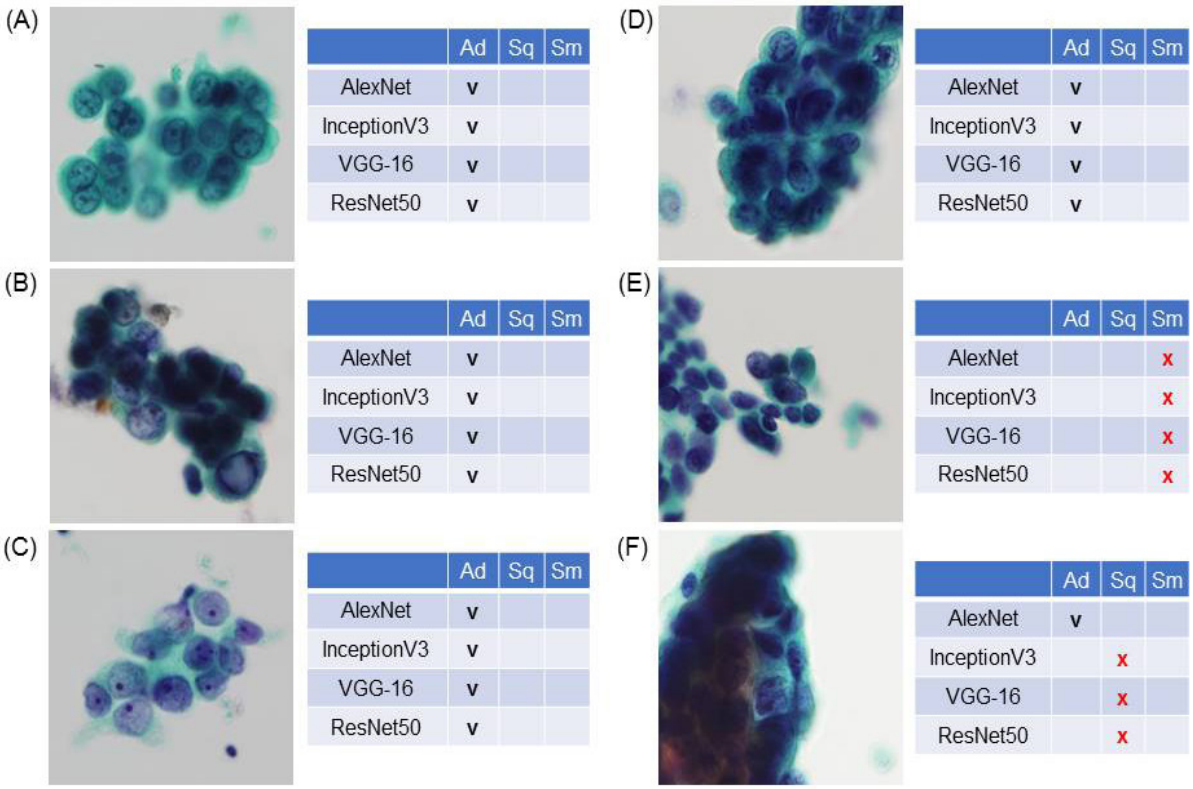

Figure 3. Representative Images of Adenocarcinoma. v and x represent correct- and mis-classifications, respectively. Papanicolaou staining. Ad, adenocarcinoma; Sq, squamous cell carcinoma; Sm, small cell carcinoma

revealing good kappa values of more than 0.6 compared to the combinations with Inception V3 providing values less than 0.6 (i.e., 0.5s).

Patel et al., (2017) recently reported the accuracy of subtyping of non-small cell lung carcinomas. Fine-needle aspiration cytology allowed tumor typing in 83 (77.6%)

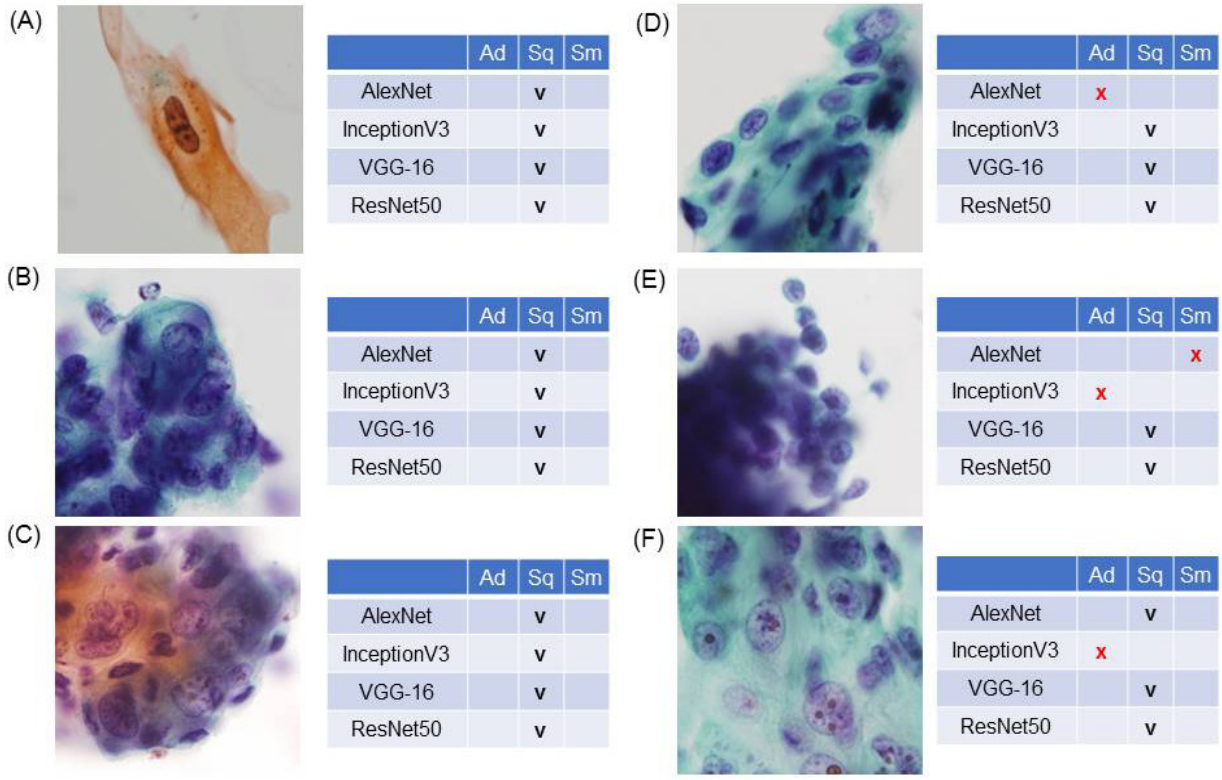

Figure 4. Representative Images of Squamous Cell Carcinoma. v and x represent correct- and mis-classifications, respectively. Papanicolaou staining.

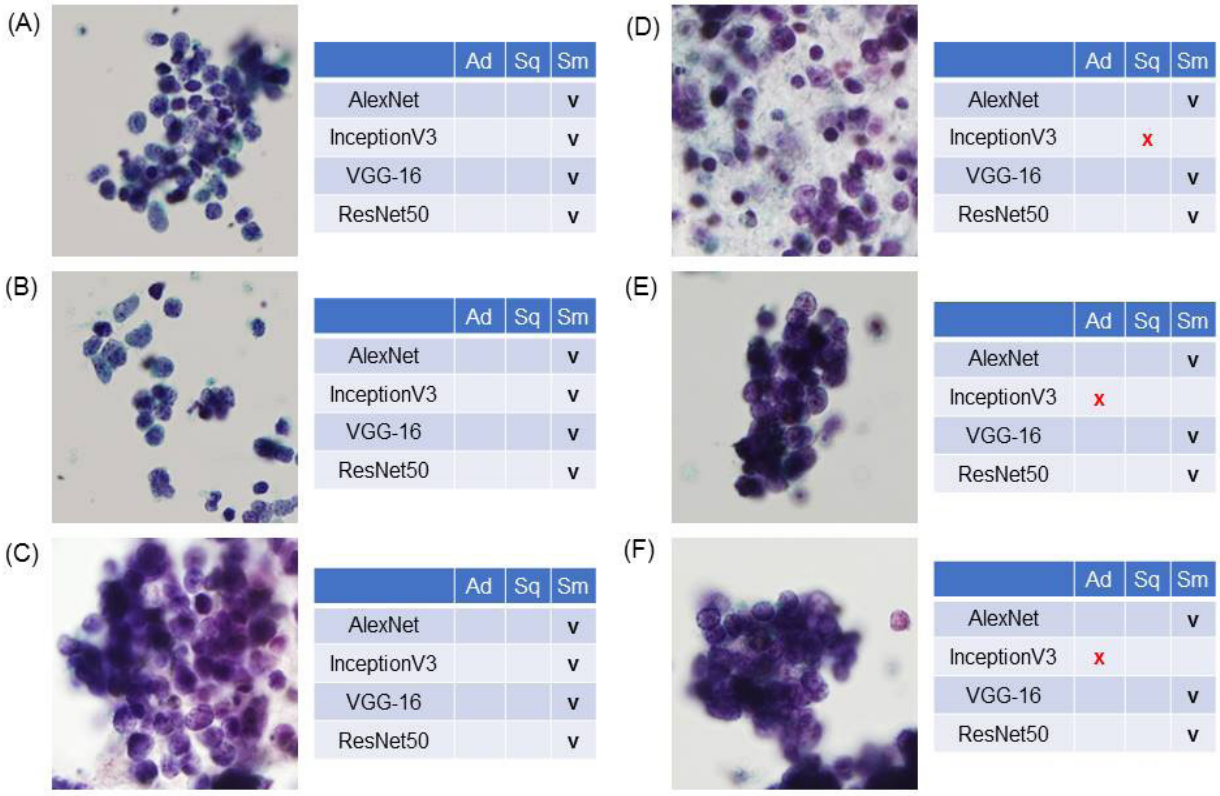

Figure. 5. Representative Images of Small Cell Carcinoma. v and x represent correct- and mis-classifications, respectively. Papanicolaou staining.

cases (36 adenocarcinomas and 47 squamous cell carcinomas) out of 107 cases. Twenty-four non-small cell carcinomas could not be further classified and remained as "non-small cell lung carcinoma, not otherwise specified." Twelve out of 14 cases were diagnosed immunohistochemistry using biopsy samples. Wallace and Rassl (Wallace and Rassl, 2011) also reported comparable result to distinguish adenocarcinoma and squamous cell carcinoma among non-small cell carcinomas. The overall accuracy of classification measured using

Table 5. Confusion Matrix of Classification Results by Additional Classifiers

Naïve Bayes			Predicted		total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	59 (78.7%)	11 (14.7%)	5 (6.7%)	75 (100%)	
	SqCC	30 (23.1%)	92 (70.8%)	8 (6.2%)	130 (100%)	
	SmCLC	5 (6.0%)	13 (15.5%)	66 (75.6%)	84 (100%)	75.1%
Support vector machine			Predicted		total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	44 (58.7%)	26 (36.7%)	5 (6.7%)	75 (100%)	
	SqCC	15(11.5%)	110 (84.6%)	5 (3.8%)	130 (100%)	
	SmCLC	0 (0%)	14 (16.7%)	70 (83.3%)	84 (100%)	77.5%
Random forest			Predicted		total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	51 (68.0%)	19 (25.3%)	5 (6.7%)	75 (100%)	
	SqCC	17 (13.1%)	107 (82.3%)	6 (4.6%)	130 (100%)	
	SmCLC	2 (2.4%)	14 (16.7%)	68 (81.0%)	84 (100%)	78.2%
Neural network			Predicted		total	Overall accuracy
		ADC	SqCC	SmCLC		
Actual	ADC	54 (72.0%)	17 (22.7%)	4 (5.3%)	75 (100%)	
	SqCC	21 (16.2%)	98 (75.4%)	11 (8.5%)	130 (100%)	
	SmCLC	4 (4.8%)	4 (4.8%)	76 (90.5%)	84 (100%)	78.9%

ADC, adenocarcinoma; SqCC, squamous cell carcinoma; SmCLC, small cell lung carcinoma

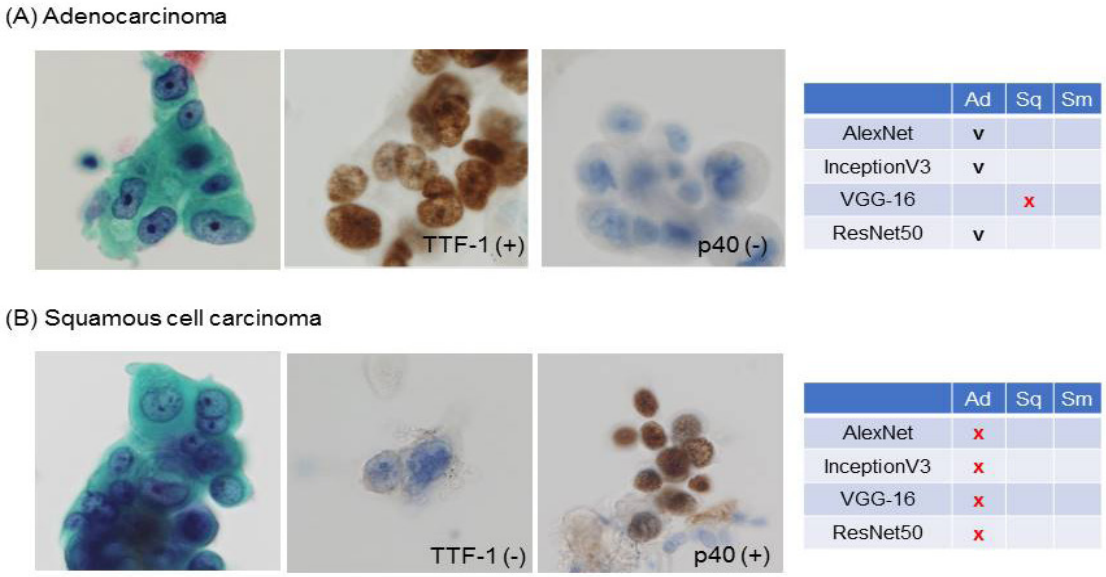

Figure 6. Two Resembling Non-Small Cell Lung Carcinomas, v and x Represent Correct- and mis-classifications, respectively. Papanicolaou and immunostaining for TTF-1 and p40.

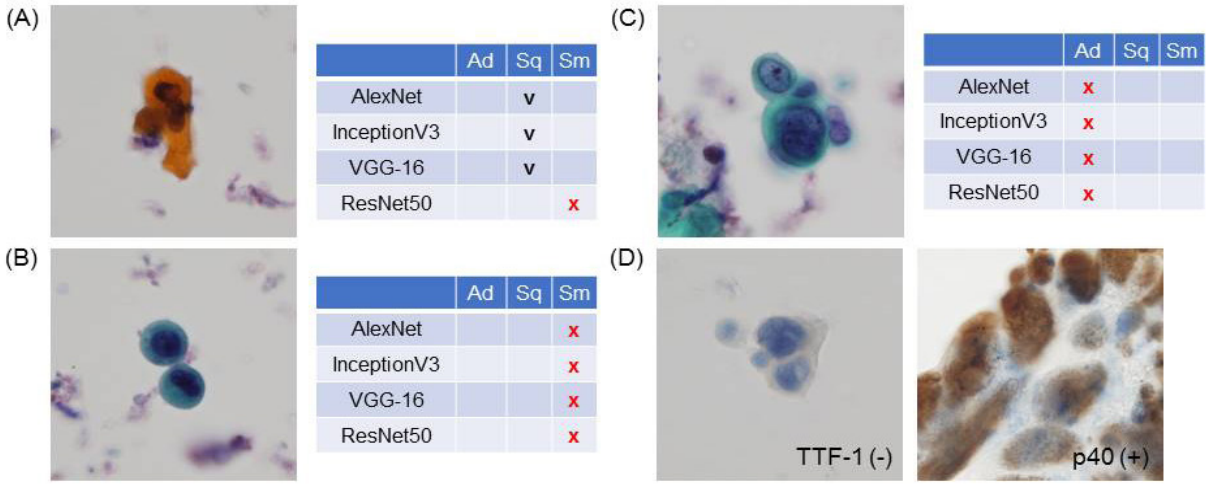

Figure 7. Non-Resembling Images of Squamous Cell Carcinomas from a Same Patient. v and x represent correct- and mis-classifications, respectively. Papanicolaou and immunostaining for TTF-1 and p40.

immunohistochemical methods of simultaneously obtained biopsy samples was 37 out of 48 cases (75.0%). Nizzoli et al. (Nizzoli et al., 2011) also reported that 85% cases (158 cases) were categorized as adenocarcinoma or squamous cell carcinoma, whereas 15% cases were categorized as non-small cell carcinoma with significant concordance of histology and cytology (kappa = 0.755). In the current analysis, DCNNs were forced to identify one of the three categories, namely adenocarcinoma, squamous cell carcinoma, and small cell carcinomas, and were not allowed to stay in non-small cell carcinomas. Although, it is not feasible to compare our results with the data from previous literatures, the classification performance of the current fine-tuned DCNNs may be considered to be comparable with human cytopathologists.

Zusman-Harach et al., (1991) stated that the presence of glands, cell balls, branching or papillary structures, cylindrical cells, and nuclear grooving were major diagnostic indicators for adenocarcinoma. On the other hand, the major cytological indicators for squamous cell carcinomas were the presence of keratin and eosinophilic spindle cells with glassy or laminated cytoplasm. Sigel et al., (2011) reported that non-small cell carcinomas were distributed into 69% definitive, 19% favored, and 12% unclassified categories as evaluated with diagnostic certainty by the individual pathologist. Definitive diagnoses for adenocarcinoma were rendered when tumor cells featured obvious glandular features or cytoplasmic mucin. For the identification of squamous cell carcinoma, keratinization was considered as a reliable morphologic feature. Images that were correctly identified by all four DCNNs tended to show such typical morphology tubular or papillary cell mass, moderate size of nuclei, single nucleoli with fine nuclear matrix. However, smaller or larger nuclei show difficulties for correct diagnosis. Intermediate cell types between adenocarcinoma and

squamous cell carcinoma, usually diagnosed as non-small cell carcinoma, revealed difficulties in identification as expected. However, a combination of several images would help more accurate diagnosis even though there are difficulties in each single figure. It is notable that DCNNs could recognize both single cell morphology and shape of cancer cell mass solely from images without prior knowledge and experience of biology and pathology. Thus, automated classification should be performed with multiple sampling of cellular images from one patient.

Transfer learning has recently attracted the attention of many researchers. Deniz et al., (2018) utilized AlexNet and VGG16 to diagnose breast cancer images. The evaluation results showed that the transfer learning produced better results than the deep feature extraction and support vector machine classification. For the comparison of various DCNN architectures, Too et al., (2019) compared VGG16, Inception V4, ResNet with 50, 101 and 152 layers and DenseNets with 121 layers to evaluate diseased images of plant leaves; DenseNet showed consistent improvement of accuracy without overfitting. The CNN methodology has been getting established; however, there have not been enough research publications for the comparison of different architectures. Although weight optimization of the artificial neural network would be performed during the learning, architecture itself should be determined empirically before the analysis (Shirakawa et al, 2018).

Subsequent use of additional classifiers following DCNNs improved the overall accuracy. McAllister et al., (2018) revealed that ResNet-152 with Support vector machine with RBF kernel detected food items with 99.4% accuracy. Wu et al., (2020) showed better performance in brain tumor segmentation in MRI images using DCNN in combination with Support vector machine. To analyze time domain vibration signal, Xu et al., (2019) converted the signals to gray scale images, which was classified with CNN model and further analyzed with multiple Random forests. This ensemble learning proved better performance.

In this study, we have evaluated and compared four automated classification schemes for lung cancers in microscopic images using fine-tuned DCNNs. Typical morphology was identified correctly by all four schemes. But poorly differentiated carcinomas showed difficulties in concordance between architectures. Certain DCNNs tended to judge certain histological types. Additional machine learning classifiers proved better performance compared to DCNNs alone. The study results warrant further analysis to develop a method to comprehensively classify cells and arrays of cells. So, we hope our method would help assisting cytological examination in lung cancer diagnosis.

Author Contribution Statement

The authors confirm contribution to the paper as follows: study conception and design: TT, AT, HF; data collection: AY, YK, ES, AM; analysis and interpretation of results: KI. Author; draft manuscript preparation: TT, AT. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

The authors thank Dr. Malcolm A. Moore for valuable discussion.

Funding Statement

This research was partially supported by a Grant-in-Aid for Scientific Research on Innovative Areas (Multidisciplinary Computational Anatomy, No. 26108005), a Grant-in-Aid for Scientific Research (No. 17K09070), Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, and a Grant-in-Aid for Scientific Research from Fujita Health University, Japan.

Ethics Approval Statement

This study was approved by an institutional review board, Fujita Health University, and patient informed consents were obtained under the condition that all data were anonymized (No. HM16–155).

Statement conflict of Interest

The authors have no conflict of interests.

References

Amancio DR, Comin CH, Casanova D et al (2014). A systematic comparison of supervised classifiers. *PLoS One*, **9**, e94137.

American Cancer Society. (2015). Cancer Facts & Figures 2015. American Cancer Society, Atlanta, GA.

Baas P, Belderbos JS, Senan S et al (2006). Concurrent chemotherapy (carboplatin, paclitaxel, etoposide) and involved-field radiotherapy in limited stage small cell lung cancer: a Dutch multicenter phase II study. *Br J Cancer*, **94**, 625-30.

Bychkov D, Linder N, Turkki R et al (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep*, **8**, 3395.

Cohen J (1960). A coefficient of agreement for nominal scales. *Educ Psychol Meas*, **20**, 37-46.

Demsar J, Curk T, Erjavec A, et al (2013). Orange: Data Mining Toolbox in Python. *J Mach Learn Res*, **14**, 2349-53.

Deniz E, Sengur A, Kadiroglu Z, et al (2018) Transfer learning based histopathologic image classification for breast cancer detection. *Health Inf Sci Syst*, **6**, 18.

He K, Zhang X, Ren S, et al (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp 770-8.

Krizhevsky A, Sutskever I, Hinton GE (2012), Imagenet classification with deep convolutional neural networks. *Adv Neur In*, **25**, 1106–14.

McAllister P, Zheng H, Bond R, et al (2018), Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Comput Biol Med*, **95**, 217-33.

Nizzoli R, Tiseo M, Gelsomino F, et al (2011) Accuracy of fine needle aspiration cytology in the pathological typing of non-small cell lung cancer. *J Thorac Oncol*, **6**, 489-93.

Patel TS, Shah MG, Gandhi JS et al (2017) Accuracy of cytology in sub typing non small cell lung carcinomas. Diagn Cytopathol, 45, 598-603.

Qu J, Hiruta N, Terai K et al (2018). Gastric Pathology Image Classification Using Stepwise Fine-Tuning for Deep Neural Networks. J Healthc Eng, 2018, 8961781.

Ravishankar H, Sudhakar P, Venkataramani R, et al (2016). Understanding the Mechanisms of Deep Transfer Learning for Medical Images. In Deep Learning and Data Labeling for Medical Applications DLMIA 2016, LABELS 2016 Lecture Notes in Computer Science, Carneiro Gea (ed), Vol. 10008. Springer, Cham.

Shirakawa S, Iwata Y, Akimoto Y (2018). Dynamic Optimization of Neural Network Structures Using Probabilistic Modeling. In The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). New Orleans, Louisiana, USA.

Sigel CS, Moreira AL, Travis WD et al (2011). Subtyping of non-small cell lung carcinoma: a comparison of small biopsy and cytology specimens. *J Thorac Oncol*, **6**, 1849-56.

Simonyan K, Zisserman A (2015). Very deep convolutional networks for large-scale image recognition. ICLR 2015; arXiv Preprint.

Szegedy C, Lie W, Jia Y et al (2015). Going deeper with convolutions. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp 1-9.

Teramoto A, Tsukamoto T, Kiriyama Y, et al (2017). Automated classification of lung cancer types from cytological images using deep convolutional neural networks. *Biomed Res Int*, **2017**, 4067832.

Teramoto A, Yamada A, Kiriyama Y, et al (2019). Automated classification of benign and malignant cells from lung cytologicalimages using deep convolutional neural network. *Inform Med Unlocked*, **16**, 100205.

Too EC, Yujian L, Njuki S, et al (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Comput Electron Agric*, **161**, 272-9.

Travis WD, Brambilla E, Burke A, et al (2015). WHO Classification of tumours of the lung, pIeura, thymus, and heart, 4th edn. Lyon: IARC.

Tsukamoto T, Nakagawa M, Kiriyama Y, et al (2017). Prevention of Gastric Cancer: Eradication of Helicobacter Pylori and Beyond. *Int J Mol Sci*, **18**, 1699.

Wallace WA, Rassl DM (2011). Accuracy of cell typing in nonsmall cell lung cancer by EBUS/EUS-FNA cytological samples. *Eur Respir J*, **38**, 911-7.

Wu W, Li D, Du J et al (2020). An intelligent diagnosis method of brain MRI tumor segmentation using deep convolutional neural network and SVM Algorithm. *Comput Math Methods Med*, **2020**, 6789306.

Xu G, Liu M, Jiang Z, et al (2019). Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. *Sensors (Basel)*, **19**.

Zusman-Harach SB, Harach HR, Gibbs AR (1991) Cytological features of non-small cell carcinomas of the lung in fine needle aspirates. *J Clin Pathol*, **44**, 997-1002.