Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/42/20240655

Classification of single-cell routine Pap smear images based on deep learning algorithms

Guanhao Zhang

Information technology, UNSW sydney, New South Wales, 2134, Australia

king8glory@gmail.com

Abstract. In this study, we classified single-cell routine Pap smear images by applying deep learning algorithms such as AlexNet, VggNet, GoogleNet and MobileNet and compared their classification effects. The results show that the loss of all four models on both the training and test sets shows a trend of gradually decreasing and stabilising. Specifically, the loss of AlexNet gradually decreases from 0.637 to 0.212, VggNet from 0.777 to 0.278, GoogleNet from 1.77 to 0.31, and MobileNet from 0.809 to 0.267. At the same time, MobileNet exhibits the highest maximum and average accuracies which reached 93.9% and 88.3%, respectively, followed by GoogleNet model with 92.9% and 88.0%, AlexNet with 92% and 88.0%, and VggNet with 90.1% and 86.7%. The results show that MobileNet exhibits superior classification results in this task, which provides strong support for its potential application in the classification of single-cell routine Pap smear images. These findings are of great significance for further exploring the application of deep learning in the field of medical imaging and provide a useful reference for future related research.

Keywords: Image classification, MobileNet, GoogleNet.

1. Introduction

Single-cell routine Pap smear image classification is an important area of medical research dedicated to the diagnosis and treatment of diseases by analysing and classifying cell morphology, structure and staining characteristics. The background of this research stems from the need for early diagnosis and treatment of diseases such as cancer and infectious diseases.

In traditional medical imaging diagnosis, doctors need to rely on visual observation or microscopic observation of cell morphology features, followed by manual classification and diagnosis [1]. This method has problems such as high subjectivity and low efficiency. And with the development of digital medical imaging technology, deep learning algorithms began to be introduced for single-cell routine Pap smear image classification [2].

Deep learning algorithm plays an important role in single-cell routine Pap smear image classification. Firstly, deep learning algorithms can automatically extract feature information in images, including morphological, textural, staining and other features, avoiding the subjectivity and instability of traditional manual feature extraction [3]. Secondly, deep learning algorithms can be trained by a large amount of data and have a strong generalisation ability, which enables more accurate cell classification and diagnosis. In addition, deep learning algorithms can combine multimodal data (e.g., optical

DOI: 10.54254/2753-8818/42/20240655

microscope images, magnetic resonance imaging, etc.) for comprehensive analysis, which improves the accuracy and comprehensiveness of diagnosis.

In practical applications, deep learning algorithms can also be combined with the clinical experience of doctors to assist them in rapid and accurate diagnosis. At the same time, in the big data environment, deep learning algorithms can also help doctors discover some potential new clinical markers or molecular mechanisms, and provide support for individualised treatment.

In conclusion, there is an urgent need for deep learning algorithms in the field of single-cell routine Pap smear image classification. In this paper, we classify single-cell routine Pap smear images based on deep learning algorithms such as AlexNet [4], VggNet [5], GoogleNet [6], and MobileNet [7], and compare the classification effects of the four classification models to explore the effects of their classification on single-cell routine Pap smear images. routine Pap smear image classification and compare the classification effect of the four classification models to explore their potential application on single-cell routine Pap smear image classification.

2. Introduction to the dataset

In this paper, we choose the open source dataset of SIPaKMeD database, the URL of the dataset is (https://www.kaggle.com/datasets/mohaliy2016/papsinglecell/data), in this paper, we select the Dyskeratotic, Koilocytotic and Metaplastic images, Dyskeratotic contains 713 images, Koilocytotic contains 725 images, and Metaplastic contains 693 images. Some of the images are shown in Figure 1.

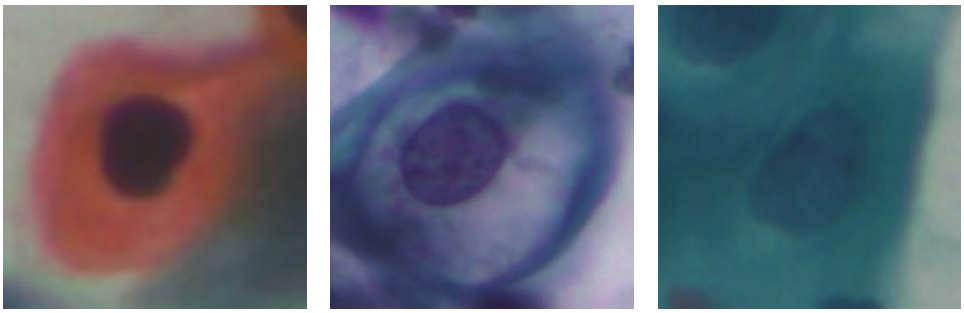


Figure 1. The change of the traffic flow. (a) Dyskeratotic. (b) Koilocytotic. (c) Metaplastic.

3. Method

3.1. AlexNet

AlexNet employs a deep convolutional neural network structure with 8 transform layers (5 convolutional and 3 fully connected layers) and uses a ReLU activation function to enhance the nonlinear representation of the network. This deep network structure better captures the high-level features in the image and improves the accuracy of image recognition. In addition, AlexNet introduces a Local Response Normalisation (LRN) layer to enhance the generalisation ability, while Dropout technique is used to reduce the risk of overfitting.The modular structure of AlexNet is shown in Fig. 2.

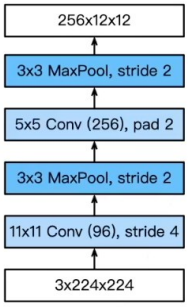


Figure 2. The modular structure of AlexNet.

DOI: 10.54254/2753-8818/42/20240655

During the training process, AlexNet employs regularisation methods such as data augmentation and stochastic inactivation to improve the generalisation ability of the model. Data augmentation expands the training dataset by rotating, flipping, scaling and other operations on the training samples, thus reducing the sensitivity of the model to the input data; while random deactivation reduces the dependency between neurons by randomly setting the neuron outputs to 0, which helps prevent overfitting. In terms of optimisation algorithms, AlexNet uses a momentum-based stochastic gradient descent (SGD) algorithm and introduces a learning rate decay strategy. These optimisation methods effectively accelerate the model training process and improve the speed of model convergence.

In terms of hardware devices, AlexNet significantly accelerates the training speed with the help of GPU parallel computing capability. This lays the foundation for the subsequent application of deep learning models on large-scale image data [8].

3.2. VGG

The principle of VGG model is mainly based on the design and feature extraction of deep convolutional neural network, which gradually extracts the features of the input image through multi-layer convolution and pooling operations, and classifies them through the fully connected layer.

Firstly, the VGG model adopts a uniform convolutional kernel size (3x3) and step size (1), and the ReLU activation function is used after each convolutional layer to enhance the nonlinear fitting ability of the network. This design allows the VGG model to have a deeper network structure, which helps to extract more complex and abstract image features. In addition, in order to reduce the number of parameters and decrease the risk of overfitting, the VGG model employs multiple consecutive convolutional layers followed by a maximum pooling layer to gradually reduce the feature map size. Secondly, the VGG model structurally employs stacking multiple convolutional and pooling layers as well as fully connected layers to construct the entire network. Among them, VGG16 and VGG19 are two commonly used versions, consisting of 16 and 19 convolutional layers, respectively. These convolutional layers can effectively capture local features in the input image and gradually combine them into more abstract and high-level feature representations. Finally, these high-level features are mapped to different classes for classification through fully connected layers.

In addition, the VGG model uses a smaller convolutional kernel and a deeper network structure to enhance the recognition of subtle features as well as complex textures in images.

3.3. GoogLeNet

GoogLeNet uses multiple Inception modules to build the entire network. Each Inception module consists of four parallel paths that perform 1x1, 3x3, 5x5 convolution and max pooling operations and connect their outputs. Doing so allows the network to learn richer and more complex feature representations at different scales, which improves the image recognition performance.The module structure diagram of GoogLeNet is shown in Fig. 3.

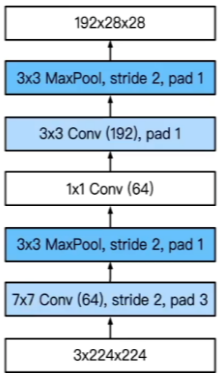


Figure 3. The module structure diagram of GoogLeNet.

DOI: 10.54254/2753-8818/42/20240655

In order to reduce the number of parameters and computational effort, GoogLeNet uses 1x1 convolution to reduce the number of channels in each Inception module.1x1 convolution reduces the number of channels in the input data while keeping the spatial dimensions unchanged, which reduces the amount of data that needs to be processed in the subsequent 3x3 and 5x5 convolution operations [9].

In terms of network structure, GoogLeNet also uses a global average pooling layer instead of a fully connected layer. Global average pooling allows the feature maps output from the last convolutional layer to be subjected to an average pooling operation, yielding a fixed-size feature vector as input to the final classifier. This reduces the number of parameters and effectively prevents overfitting.

During the training process, to further improve performance, Google also used auxiliary classifiers to guide the gradient back propagation. These auxiliary classifiers were added to certain layers in the middle and combined with an overall loss function to facilitate gradient propagation and accelerate convergence.

3.4. MobileNet

MobileNet employs a depth-separable convolution structure that breaks down the standard convolution operation into two steps: deep convolution and point-by-point convolution. Deeply separable convolution first performs independent spatial convolution for each input channel, and then linearly combines the output channels through point-by-point convolution. This structure effectively reduces computational complexity and maintains relatively good feature extraction capability.The schematic structure of MobileNet is shown in Figure 4.

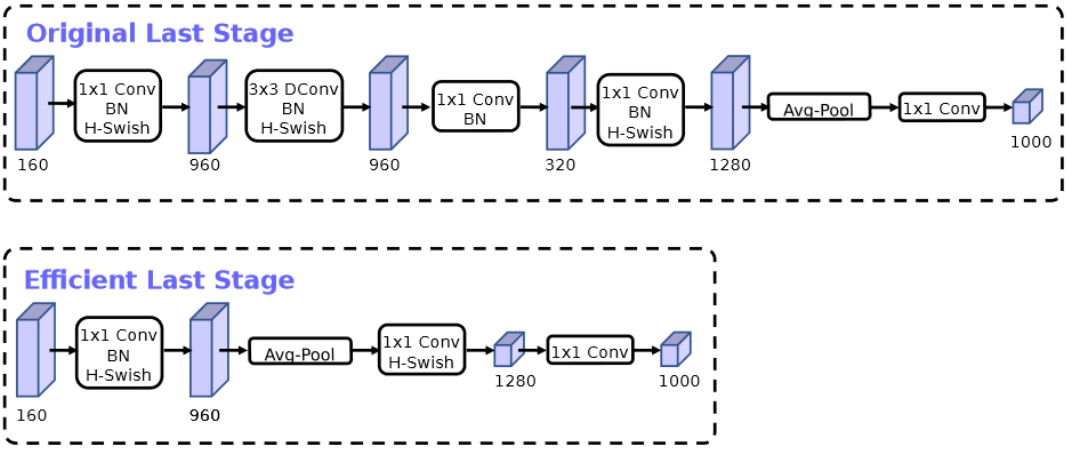


Figure 4. The schematic structure of MobileNet.

MobileNet introduces the concepts of width multiscale and resolution multiscale. Width multiscaling allows the user to control the number of channels in each layer of the network, thus striking a balance between model size and speed, while resolution multiscaling allows the input image to be scaled so that the model can adapt to image inputs of different resolutions.

In addition, global average pooling is used in the last few layers instead of fully connected layers, further reducing the number of parameters. This design allows MobileNet to significantly reduce the model size and computational load while maintaining relatively high accuracy, making it ideal for deployment on mobile devices, embedded systems, and edge-end devices [10].

4. Result

Firstly, the data is configured, and a total of three types of images, Dyskeratotic, Koilocytotic and Metaplastic, are divided into the training set and the test set according to the ratio of 7:3, with 70% of the data being used for training, and 30% of the data being used for testing. The graphics card used for the experiment is 3090, and the memory is 32G. During the experiment, the changes of loss and accuracy

DOI: 10.54254/2753-8818/42/20240655

in the validation set are recorded, and the change curves of loss and accuracy are outputted; the change curves of loss for the four models are shown in Fig. 5, and the change curves of accuracy for the four models are shown in Fig. 6.

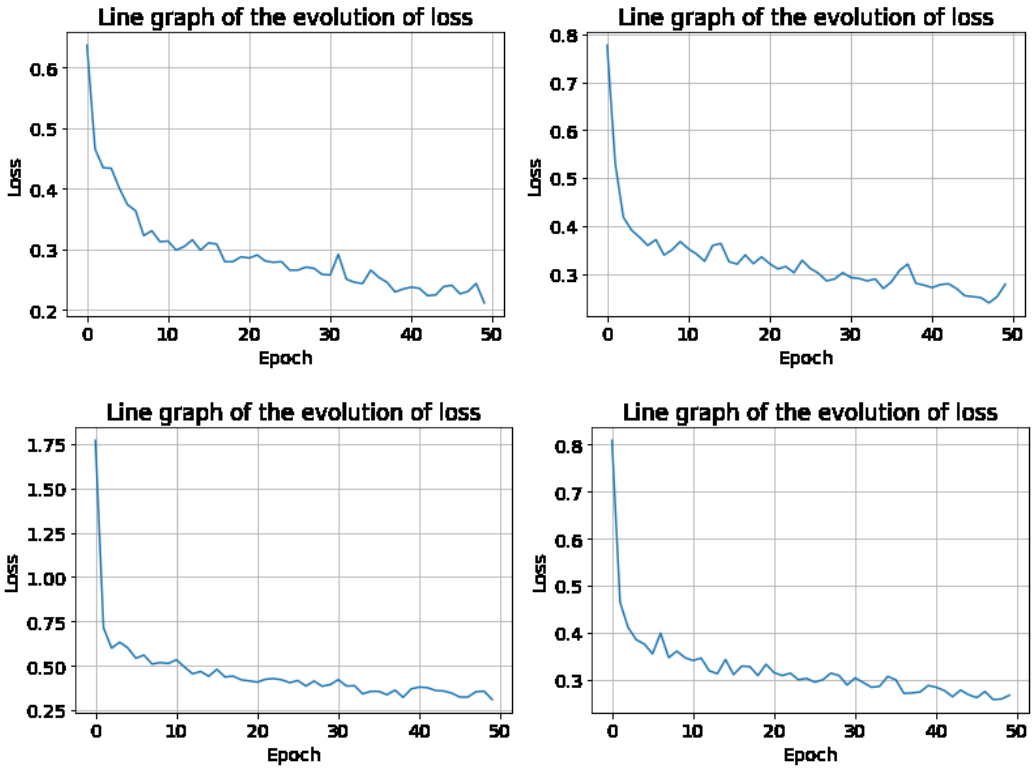
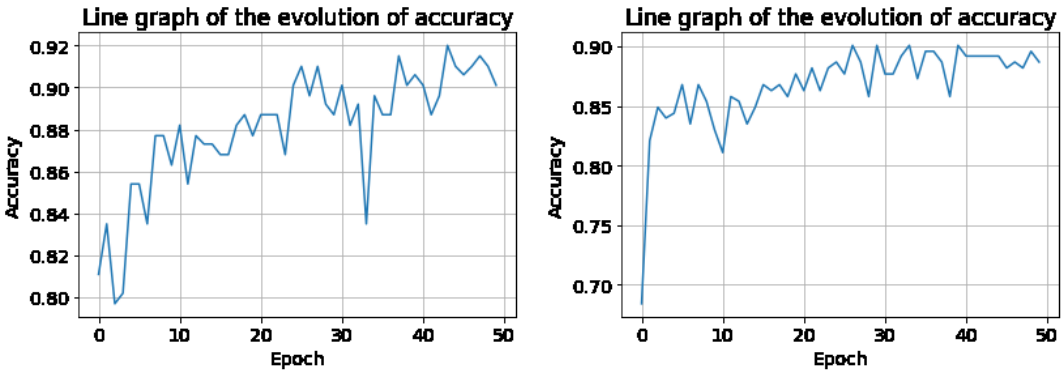


Figure 5. The change curves of loss for the four models.

As can be seen from the change of loss of the four models, the loss value of AlexNet gradually decreases from 0.637 and converges to 0.212, the loss value of Vgg gradually decreases from 0.777 and converges to 0.278, the loss value of GoogleNet gradually decreases from 1.77 and converges to 0.31, and the loss value of MobileNet gradually decreases from 0.80 to 0.31, and the loss value of Vgg gradually decreases from 0.80 to 0.31. for GoogleNet decreases from 1.77 and converges to 0.31, and for MobileNet the loss value decreases from 0.809 and converges to 0.267.



DOI: 10.54254/2753-8818/42/20240655

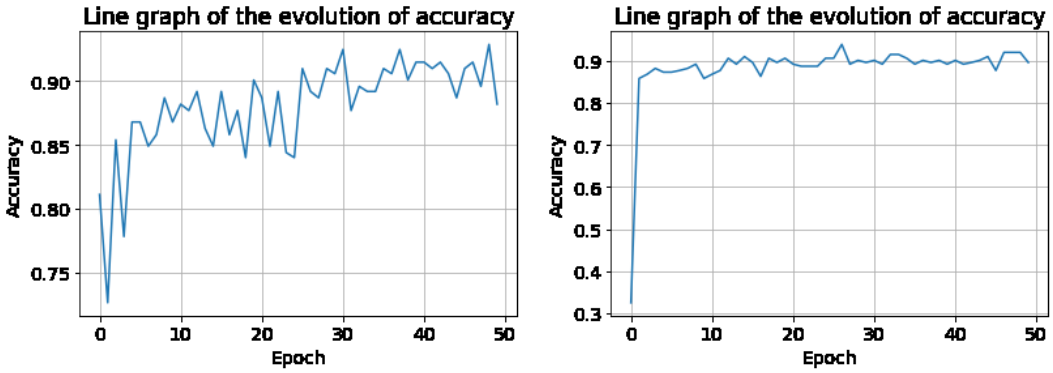


Figure 6. The change curves of accuracy for the four models.

The accuray from the validation set of the four models are counted to calculate the maximum and average values of the model accuracy, and the results are shown in Table 1, from which it can be seen that MobileNet has the highest maximum and average accuracy, which are 93.9% and 88.3%, respectively; followed by GoogleNet model, with the maximum and average accuracy, which are 92.9% and 88.0%, respectively; AlexNet model, with the maximum and average accuracy, which are 92.9% and 88.0%, respectively; and GoogleNet model, with the maximum and average accuracy, which are 92.9% and 88.0%, respectively. 88.0%; AlexNet is next with 92% and 88.0% for maximum and average accuracy, respectively; and the worst prediction is the Vgg model with 90.1% and 86.7% for maximum and average accuracy, respectively.

Table 1. Model evaluation parameters.

Model	Max	Average
AlexNet	0.92	0.88058
Vgg	0.901	0.86774
GoogleNet	0.929	0.88038
MobileNet	0.939	0.8829

5. Conclusion

In this study, we have classified single-cell routine Pap smear images based on deep learning algorithms such as AlexNet, VggNet, GoogleNet and MobileNet, and compared the effectiveness of the four classification models. By dividing the data into training and test sets in the ratio of 7:3, we recorded and analysed the variations of loss and accuracy in the validation set. When observing the trend of the loss changes of the four models, it is found that the loss value of AlexNet gradually decreases from 0.637 and gradually converges to 0.212, the loss value of Vgg gradually decreases from 0.777 and gradually converges to 0.278, the loss value of GoogleNet gradually decreases from 1.77 and gradually converges to 0.31, and the loss value of MobileNet's loss value gradually decreases from 0.809 and gradually converges to 0.267.These results indicate that MobileNet exhibits faster convergence to lower loss values in the task of single-cell routine Pap smear image classification.

Further analysing in terms of maximum and average accuracies, we find that MobileNet exhibits the best performance with 93.9% and 88.3% maximum and average accuracies, respectively; followed by the GoogleNet model with 92.9% and 88.0% maximum and average accuracies, respectively; and AlexNet is ranked third, with 92.9% and 88.0% maximum and average accuracy of 92% and 88.0%, respectively; and the Vgg model presents the worst prediction results, with a maximum and average accuracy of only 90.1% and 86.7%.

In conclusion, the results of this study show that MobileNet exhibits high classification accuracy in the task of single-cell routine Pap smear image classification and converges faster during the training process. Therefore, it has the potential for wide application in this field.

DOI: 10.54254/2753-8818/42/20240655

References

[1] Kumar, Rakesh, et al. "Medical images classification using deep learning: a survey." Multimedia Tools and Applications 83.7 (2024): 19683-19728.
[2] Wang, Pin, et al. "Deep sample clustering domain adaptation for breast histopathology image classification." Biomedical Signal Processing and Control 87 (2024): 105500.
[3] Hou, Wenpin, and Zhicheng Ji. "GPT-4V exhibits human-like performance in biomedical image classification." bioRxiv (2024).
[4] Deo, Bhaswati Singha, et al. "An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification." International Journal of Data Science and Analytics (2024): 1-18.
[5] Alsaidi, Mostapha, et al. "Tackling the class imbalanced dermoscopic image classification using data augmentation and GAN." Multimedia Tools and Applications 83.16 (2024): 49121-49147.
[6] Hörst, Fabian, et al. "Cellvit: Vision transformers for precise cell segmentation and classification." Medical Image Analysis 94 (2024): 103143.
[7] Han, Qi, et al. "DM-CNN: Dynamic Multi-scale Convolutional Neural Network with uncertainty quantification for medical image classification." Computers in Biology and Medicine 168 (2024): 107758.
[8] Veeramani, Nirmala, et al. "DDCNN-F: double decker convolutional neural network'F'feature fusion as a medical image classification framework." Scientific Reports 14.1 (2024): 676.
[9] Zhang, Mengxuan, et al. "Tree-shaped multiobjective evolutionary CNN for hyperspectral image classification." Applied Soft Computing 152 (2024): 111176.
[10] Qu, Linhao, et al. "Rethinking multiple instance learning for whole slide image classification: A good instance classifier is all you need." IEEE Transactions on Circuits and Systems for Video Technology (2024).