Diagnostic Classification of Cervical Cell Images from Pap Smear Slides

*[1]Muhammed Talo
*[1]Faculty of Engineering, Department of Software Engineering, Firat University, Elazig, Turkey

Abstract

Analysis of pap smear images under a microscope by experts is a laborious and time-consuming task. Computer aided diagnostic (CAD) systems can simplify this tedious process and allow experts to focus on more critical cases. Effective screening and early diagnosis can help to detect precancerous cells and allow early treatment. In this study, we have used a deep learning approach to classify cervical cell images obtained from pap smear slides. The proposed method automatically classifies cervical cell images into five categories without using any pre-processing on raw input images. We have obtained promising results as compared to the previous studies in the literature. The proposed deep learning model can give a second opinion to clinicians in their daily routines and help them to focus on more complex cases.

Key words: Deep learning, cervical cell classification, pap smear, convolutional neural networks.

1. Introduction

Cervical cancer is the fourth common cancer among women worldwide [1]. According to the World Health Organization (WHO), more than 570,000 women were diagnosed with cervical cancer in 2018 and every year approximately 300,000 women die from cervical cancer [2]. The countries with the highest mortality rates are low-income developing countries due to the shortage of skilled practitioners and limited resources.

The most common method to diagnose cervical cancer is effective screening. Early diagnosis can save lives [3]. Pap smear is a manual screening test for checking the cervix for abnormal cells. Practitioners take some samples from the cervix area using a brush or scrapper. The collected cells are observed under a microscope to diagnose and classify abnormalities in cervical cells [4]. During the examination of cervical cells, trained health workers examine the shape and color characteristics of the nucleus and cytoplasm regions. In developing countries, pap smear test has decreased mortality rates from 70% to 50% [5].

In the past decade, various machine learning approaches and CAD systems have been used for detection and classification of abnormal cell images such as K-nearest neighbors [6,7], support vector machines (SVM) [7, 8], artificial neural networks (ANN) [9] etc. Plissiti et al. [10] applied K-PCA (Gaussian kernel) method to classify cervical cells images to normal and abnormal classes using nucleus features by ignoring the features derived from the cytoplasm. In this study, the authors reported the best harmonic mean (H-mean) classification sensitivity of 90.58% using seven nucleus features.

*Corresponding author: Muhammed Talo, Address: Faculty of Engineering, Department of Software Engineering Firat University, 23300, Elazig TURKEY. E-mail address: mtalo@firat.edu.tr, Phone: +904242370000-5571

Chankong et al. [11] used a patch-based fuzzy C-means (FCM) clustering approach to segment cervical cell images into the cytoplasm, nucleus, and background. Then, the segmented images were fed into various classifiers, such as bayesian, SVM, K-nearest neighbor, linear discriminant analysis, and artificial neural networks to classify cervical cell images. They have reported the highest accuracy of 97.83% using the ERUDIT dataset to classify cell images into two classes. The authors also used Herlev and LCH datasets and obtained the highest accuracy of 99.27% for two class problems. Bora et al. [12] proposed an ensemble classifier using least squares support vector machines (LS-SVM), multilayer perceptron (MLP) and random forest (RF) to classify pap smear images. The proposed method has yielded a precision of 91.71% in cell level and 98.38% at smear level. They have tested the proposed ensemble model on the Herlev dataset and obtained classification accuracy of 96.51% and 91.73% for two and three class problems, respectively. In 2019, Mousser and Ouadfel [13] employed VGG, ResNet-50 and Inception-V3 pre-trained convolutional neural networks (CNN) architectures to extract features from pap smear images and used multi-layer perceptron (MLP) method to classify images into normal and abnormal categories. The authors obtained the best classification accuracy of 89% with ResNet-50 pre-trained CNN.

In 2018, Plissiti et al. [14] provided a publicly available dataset, SIPaKMeD, for the detection and classification of cervical cells. The authors used convolutional neural networks (CNN), multi-layer perceptron (MLP), and support vector machine (SVM) approaches to classify cervical cell images into five categories. They have achieved the highest classification accuracy of 95.35% using CNN features. In another study, Shi et al. [15] applied ResNet-161, DenseNet-121 pre-trained models, and graph convolutional networks (GCN) to classify cervical cell images into five and seven classes using SIPaKMeD and Motic datasets, respectively. The proposed GCN model achieved the best classification accuracy of %98.37 and 94.86% for SIPaKMeD and Motic datasets, respectively. In 2019, Kiran and Reddy [16] proposed ResNet-34 pre-trained CNN model using data augmentation technique to classify pap smear images. They have evaluated the performance of ResNet-34 model on SIPaKMeD and Herlev databases. The proposed method has yielded an accuracy of 96.38% for whole slide images (WSI) and 99.63% accuracy obtained using single-cell images in the SIPaKMeD dataset. The proposed pre-trained model has also tested on the Herlev database and achieved the highest classification accuracy of 98.76%. [17, 18] presents details of recent studies to diagnose cervical cancer.

In this study, an end-to-end deep learning-based approach is presented for the classification of cervical cell images. We have used DenseNet [19] architecture using transfer learning method to classify cervical cell images into five different categories. The proposed method allows automatic classification of raw cervical cell images without the use of any pre-processing and feature extraction techniques.

The rest of the paper is organized as follows: In the materials and methods section, we explain transfer learning method used in the training of the proposed model and provide architectural details of DenseNet model. In addition, detailed information is provided about SIPaKMeD cervical cell image database used in the training and testing phase of the model. In the results and discussion sections, the training stages are presented and the performance of the model is compared with the existing studies in the literature. The last section is devoted to the conclusion.

2. Materials and Method

2.1. Deep Learning Based Transfer Learning Approach

In the last decade, deep learning, which is the sub-branch of artificial intelligence, have been used many times in the analysis of medical data and successful results have been obtained for tasks such as disease detection [20, 21], classification [22, 23], and segmentation [24]. Deep learning methods, especially CNN, automatically extract features from raw data and allow automatic classification using the extracted features in an end to end manner.

The transfer learning method, which is a widely used technique for training of new deep learning models, transfers the gained information from the model which trained on a large dataset to the newly constructed model to solve a different problem. A large amount of data is often needed for the training of deep CNN models from scratch. Medical data annotation by specialists is a tedious task and expensive. In addition, since data used in the medical field is personal, it is prohibited to use it without consent of patients. Therefore, it is often not possible to find a large number of labeled data in the medical field. Transfer learning approach provides solution to the problems which has small dataset and requires less computation power.

We have employed DenseNet deep learning model to classify cervical cells from pap smear images. In regular convolutional neural networks, each layer is connected to the next layer. However, in DenseNet, each layer is linked to all consecutive layers. Therefore, each layer collects information from other layers (see details in [19]).

2.3. Dataset

In this study, we have used the open-source SIPaKMeD [14] pap smear image database which contains 4049 cell images. The cells in the SIPaKMeD dataset are divided into five different classes according to morphology and cellular appearance. Fig.1 shows sample images from the SIPaKMeD dataset. Normal cells composed of two classes: parabasal and superficial-intermediate. Metaplastic cells are the benign cells. Dyskeratotic and koilocytotic classes are abnormal but non-malignant cells. The distribution of the cells in the SIPaKMeD dataset is given in Table 1.

Table 1. SIPaKMeD dataset cell distribution.

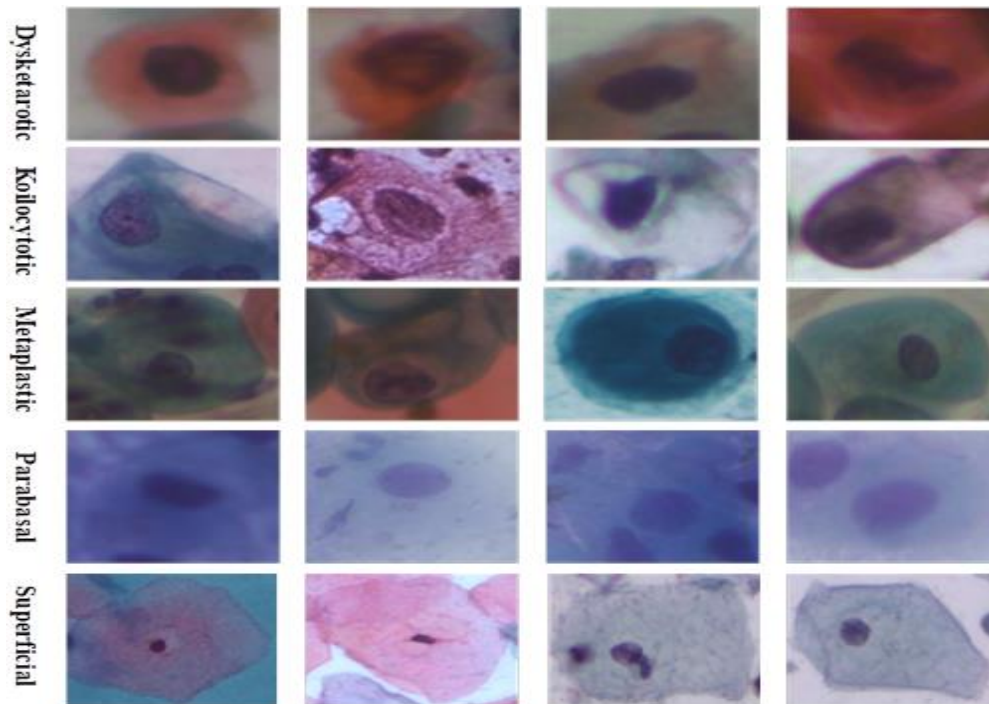| Types	Classes	Number of cells
Normal	Parabasal	787
Normal	Superficial-intermediate	813
Benign	Metaplastic	793
Abnormal	Dyskeratotic	813
Abnormal	Koilocytotic	825

Figure 1. Sample images from SIPaKMeD dataset.

3. Results

The training and testing of DenseNet model have been carried out using PyTorch libraries. We have performed all experiments on a Linux server which has Ubuntu operating system and eight NVIDIA GeForce GTX 1080 TI graphics cards, each with 11 GB of memory. We have applied the same 5-fold cross-validation experimental protocol as [14-16] for the performance evaluation of the pre-trained CNN model. In the training of the model, RMSprop optimization algorithm was used to update the weights of the model. The original DenseNet architecture has been trained in the ImageNet dataset to classify images belongs to 1000 different categories. Therefore, it has 1000 nodes in the output layer (fully-connected layer). Since the cell images used in this study have five different categories, a fully connected layer is added to the last layer of the model to give five different outputs. The training of the DenseNet-161 model has been conducted only in the new fully-connected layer. The rest of the layers took advantage of the knowledge that the model gained in the ImageNet database.

Using the transfer learning method, training of DenseNet-161 model was carried out in two phases as follows: In the first phase, we have trained the model using 256×256 image size for 12 epochs. In the second stage, the image size was increased to 512×512 and 40 epochs training was performed. The learning rates in the first and second stages were chosen as 1e-2 and 1e-3, respectively. The main reason for choosing a lower learning rate in the second phase is the retraining of the model on the SIPaKMeD dataset. Fig. 2 shows the training and validation loss graphs of DenseNet-161 model for Fold-3 and the accuracy graph for five folds.
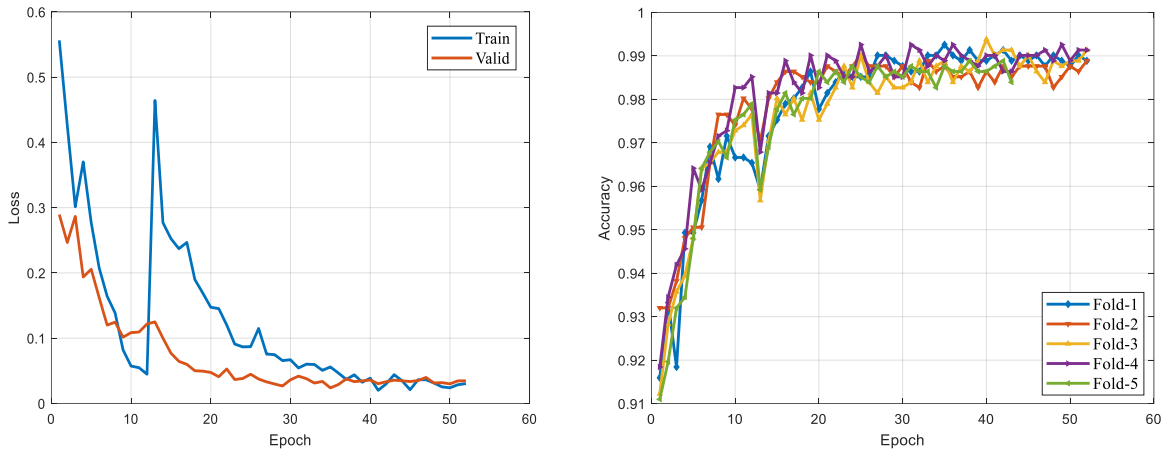
Figure 2. Accuracy and loss graphs for DenseNet-161 CNN model.
Left: training and validation loss for Fold-3. Right: 5-fold accuracy

The training process of DenseNet model has been continued for a total of 52 epochs. The proposed model has yielded an average classification accuracy of 98.96% ± 0.12. For each fold, the performance of the model has been evaluated using the corresponding validation set. 5-fold classification accuracies of DenseNet-161 are shown in Table 2.

Table 2. 5-fold classification accuracies (%)

Fold number	Accuracy
Fold-1	98.89
Fold-2	98.89
Fold-3	99.13
Fold-4	99.13
Fold-5	99.13
Average (mean ± standard deviation)	**98.96 ± 0.12**

For a detailed analysis, the confusion matrix obtained for Fold-3 using the validation set is shown in Fig. 3.

Accuracy: 99.13%

	Dysketarotic	Koilocytotic	Metaplastic	Parabasal	Superficial
Dysketarotic	98.8% / 159	0.0% / 0	0.0% / 0	0.0% / 0	0.0% / 0
Koilocytotic	1.2% / 2	99.4% / 169	1.4% / 2	0.0% / 0	0.5% / 1
Metaplastic	0.0% / 0	0.6% / 1	97.9% / 138	0.0% / 0	0.0% / 0
Parabasal	0.0% / 0	0.0% / 0	0.0% / 0	100.0% / 154	0.0% / 0
Superficial	0.0% / 0	0.0% / 0	0.7% / 1	0.0% / 0	99.5% / 182

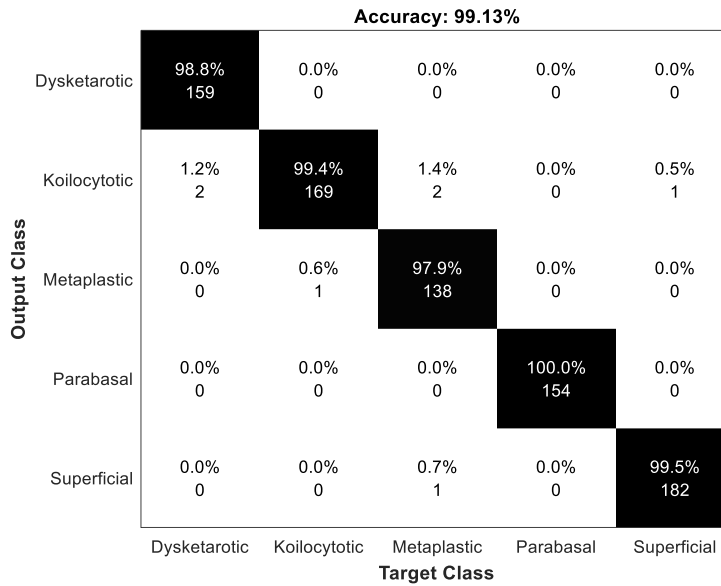Output Class (vertical axis) — Target Class (horizontal axis)

Figure 3. The confusion matrix obtained in Fold-3.

DenseNet-161 CNN model correctly classified 802 out of 809 cell images. Hence, the proposed CNN model has reached an accuracy of 99.13% for Fold-3.

4. Discussion

In Table 3, we compared the results of the other studies and the proposed DenseNet-161 deep learning model that performed cervical cell classification using the SIPaKMeD dataset. In order to have a fair comparison, we have presented the studies, that used the same dataset without data augmentation and applied the 5-fold cross-validation technique for model evaluation.

Table 3. Comparison of studies that used the SIPaKMeD dataset without data augmentation.

Study	Year	Method	Accuracy (%) (mean ± std)
Plissiti et al. [14]	2018	Deep Convolutional + SVM Deep fully-connected + SVM CNN	93.35 ± 0.62 94.44 ± 1.21 95.35 ± 0.42
Shi et al. [15]	2019	CNN ResNet-101 DenseNet-161 GNC	95.35 ± 0.42 94.86 ± 0.74 96.79 ± 0.42 98.37 ± 0.57
This study	**2019**	**DenseNet-161**	**98.96 ± 0.12**

In 2018, Plissiti et al. [14] achieved an average classification accuracy of 95.35% ± 0.42 to classify cervical cell images using CNN features. In 2019, Shi et al. [15] employed ResNet-161, DenseNet-121 pre-trained models, and graph convolutional networks (GCN) and reported the highest classification accuracy of 98.37% ± 0.57 with GNC method. In the same year, an outstanding study presented by Kiran and Reddy [16], that used ResNet-34 pre-trained CNN model with data augmentation and obtained average classification accuracy of 99.63% ± 0.19. It can be seen that from Table-3 that the proposed DenseNet-161 model has the current best classification accuracy without using data augmentation.

Conclusions

In this study, we have used DenseNet CNN model to classify cervical cell images into five classes using transfer learning technique. SIPaKMeD dataset has been used for the evaluation of the proposed model. The deep learning model has achieved an average classification accuracy of 98.96% ± 0.12 without data augmentation. The obtained result is promising to develop an automatic cervical cell classification system. The proposed method automatically classifies cervical cell images in an end to end manner.

References

[1] Khazaei, Z., Jarrahi, A. M., Momenabadi, V., Ghorat, F., Adineh, H. A., Sohrabivafa, M., & Goodarzi, E. (2019). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide stomach cancers and their relationship with the human development index (HDI). World Cancer Research Journal, 6, 9.

[2] Okunade, K. S. (2019). Human papillomavirus and cervical cancer. Journal of Obstetrics and Gynaecology, 1-7.

[3] Catarino, R., Petignat, P., Dongui, G., & Vassilakos, P. (2015). Cervical cancer screening in developing countries at a crossroad: Emerging technologies and policy choices. World journal of clinical oncology, 6(6), 281.

[4] Cox, S. (2012). Guidelines for Papanicolaou Test Screening and Follow-Up. Journal of Midwifery & Women's Health, 57(1), 86-89.

[5] Hartikainen, J. (2001). The Papanicolaou test: its utility and efficacy in cancer detection. Contemporary nurse, 11(1), 45-49.

[6] M. Sharma, S. Kumar Singh, P. Agrawal, and V. Madaan, "Classification of Clinical Dataset of Cervical Cancer using KNN," Indian J. Sci. Technol., vol. 9, no. 28, 2016

[7] R. Kumar, R. Srivastava, and S. Srivastava, "Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features," vol. 2015, 2015

[8] B. Ashok and P. Aruna, "Comparison of Feature selection methods for diagnosis of cervical cancer using SVM classifier," vol. 6, no. 1, pp. 94–99, 2016.

[9] Chankong, T., Theera-Umpon, N., & Auephanwiriyakul, S. (2009). Cervical cell

classification using Fourier transform. In 13th International Conference on Biomedical Engineering (pp. 476-480). Springer, Berlin, Heidelberg.

[10] Plissiti, M. E., & Nikou, C. (2012, June). Cervical cell classification based exclusively on nucleus features. In International Conference Image Analysis and Recognition (pp. 483-490). Springer, Berlin, Heidelberg.

[11] Chankong, T., Theera-Umpon, N., & Auephanwiriyakul, S. (2014). Automatic cervical cell segmentation and classification in Pap smears. Computer methods and programs in biomedicine, 113(2), 539-556.

[12] Bora, K., Chowdhury, M., Mahanta, L. B., Kundu, M. K., & Das, A. K. (2017). Automated classification of Pap smear images to detect cervical dysplasia. Computer methods and programs in biomedicine, 138, 31-47.

[13] Mousser, W., & Ouadfel, S. (2019, April). Deep Feature Extraction for Pap-Smear Image Classification: A Comparative Study. In Proceedings of the 2019 5th International Conference on Computer and Technology Applications (pp. 6-10). ACM.

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

[15] Shi, J., Wang, R., Zheng, Y., Jiang, Z., & Yu, L. (2019). Graph Convolutional Networks for Cervical Cell Classification.

[16] Kiran, G. V., & Meghana Reddy, G. (2019). Automatic Classification of Whole Slide Pap Smear Images Using CNN with PCA Based Feature Interpretation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

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

[18] Singh, Y., Srivastava, D., Chandranand, P. S., & Singh, D. (2018). Algorithms for screening of Cervical Cancer: A chronological review. arXiv preprint arXiv:1811.00849.

[19] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[20] Talo, M. (2019, April). Pneumonia Detection from Radiography Images using Convolutional Neural Networks. In 2019 27th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[21] Ay, B., Yildirim, O., Talo, M., Baloglu, U. B., Aydin, G., Puthankattil, S. D., & Acharya, U. R. (2019). Automated Depression Detection Using Deep Representation and Sequence Learning with EEG Signals. Journal of medical systems, 43(7), 205.

[22] Talo, M. Meme Kanseri Histopatalojik Görüntülerinin Konvolüsyonal Sinir Ağları ile Sınıflandırılması. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 31(2), 391-398.

[23] Talo, M. (2019). Convolutional Neural Networks for Multi-class Histopathology Image Classification. arXiv preprint arXiv:1903.10035.

[24] Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 Fourth International Conference on 3D Vision (3DV) (pp. 565-571). IEEE.