International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Lung Cancer Segmentation and Detection using Machine Learning

Dr. Naveena C, Arfin Khan, Arjun Dwivedi, Ashish Anand, Diksha Priya

Computer Science and Engineering Department, SJB Institute of Technology

Abstract:

Early detection of lung cancer is important in improving a patient's life. Histopathological examination of tissue is a common procedure needed to obtain an early diagnosis. Tissue analysis is usually done by a pathologist review, however, this process is time-consuming and flawed. Early detection of cancerous regions will be greatly accelerated the whole process and help the pathologist. In this paper, we suggest a completely automatic way to get lung cancer throughout the slide images of lung tissue samples. Separation is done on image correction rate using a convolutional neural network (CNN). Two CNN architects (VGG and ResNet) are trained along with their own performance compared. The results obtained show that CNN the established method has the potential to assist pathologists in lung cancer diagnosis.

Keywords: Convolutional neural networks, lung cancer

INTRODUCTION

Lung cancer is a serious disease that is a major cause of cancer-related deaths. The prognosis of patients with this disease is poor with a 5-year survival rate of less than 20% . Many patients have a severe prognosis as a result of diagnosis performed in the advanced stage of the disease. 70% Patients diagnosed with the first stage have a very high life span of 5 years. It is shown that low-dose computed tomography (LDCT) tests reduce mortality by 20% for high-risk individuals. These findings confirm early diagnosis and diagnosis as the most important step that affects the effectiveness of treatment. After discovering the tumor-suspect CT images, histopathological examination of detected tissue Bronchoscopy is a common procedure that is needed for early detection diagnosis. Biopsy tissue examination was performed by a pathologist it is a time-consuming and erroneous process in which diagnoses accuracy is less than 80%. Adjust the division into larger ones histological types (squamous carcinoma, adenocarcinoma), Small cell carcinoma, and undifferentiated carcinoma) are important treatment options. Presentation of digital pathology, where digital pathology scanners provide full-slide (WSI) images with high resolution (up to

160 nm per pixel), opens up the possibility of using computer vision to find automatic cancer in WSIs. Over the years, convolutional neural networks (CNNs) have had improved accuracy in various computer vision functions including medical images, and therefore prevailing at present way. In this paper, we suggest the default method detection of cancer cells in WSIs of lung tissue. The first step is the release of the WSI region containing the tissue, which is the region of interest (ROI), to reduce the calculation burden. This is followed by CNN-based classification of image patches tumor and normal class.

LITERATURE SURVEY

In recent years, cumulative neural networks have become a mainstream approach in medical image analysis, outperforming homemade feature-based algorithms. Esteva et al. [6] proposed a CNN-based system for skin lesion classification achieving performance comparable to that of a dermatologist. For the diagnosis of lung cancer, the methods proposed to date mainly focus on radiology. In [7], imaging-based radiographic features closely related to viability were extracted from positron emission tomography (PET/CT). In [8], CNN was used to classify lung nodule images with an accuracy of 86.4%. In digital pathology tasks, CNNs have been used at the cellular level to detect mitosis [9] and detect cell nuclei [10]. CAMELYON16 is the first challenge using WSI to detect breast cancer metastases in lymph nodes. Thanks to the availability of a large set of annotated training courses in this challenge, it is possible to train deeper and more powerful CNN architectures such as GoogLeNet [11], VGGNet [12], and ResNet[2]. 13]. The method that gives the best results in this challenge is described in [14]. It performs patch-based classification to distinguish tumor patches from normal patches using a combination of 2 GoogLeNet architectures where one of them is trained and the remaining architecture has no negative exploits. The goal of the TUPAC challenge is to detect WSI-based mitosis in breast cancer tissue and predict tumor classification. In the most efficient methodIn recent years, cumulative neural networks have become a mainstream approach in medical image analysis, outperforming homemade feature-based algorithms. Esteva et al. [6] proposed a CNN-based system for skin lesion classification achieving performance comparable to that of a dermatologist. For the diagnosis of lung cancer, the methods proposed to date mainly focus on radiology.

In [7], imaging-based radiographic features closely related to viability were extracted from positron emission tomography (PET/CT). In [8], CNN was used to classify lung nodule images with an accuracy of 86.4%. In digital pathology tasks, CNNs have been used at the cellular level to detect mitosis [9] and detect cell nuclei [10]. CAMELYON16 is the first challenge using WSI to detect breast cancer metastases

in lymph nodes. Thanks to the availability of a large set of annotated training courses in this challenge, it is possible to train deeper and more powerful CNN architectures such as GoogLeNet [11], VGGNet [12], and ResNet[2]. 13]. The method that gives the best results in this challenge is described in [14]. It performs patch-based classification to distinguish tumor patches from normal patches using a combination of 2 GoogLeNet architectures where one of them is trained and the remaining architecture has no negative exploits. The goal of the TUPAC challenge is to detect WSI-based mitosis in breast cancer tissue and predict tumor classification. In the most efficient method [15], ROI regions were first extracted from WSI based on cell density. This was followed by mitosis detection using the ResNet CNN architecture. Finally, each WSI is represented by a feature vector that includes the number of mitoses and cells in each patch as well as other statistical derived features. This feature vector is passed to the SVM classifier to predict tumor proliferation. [15], ROI regions were first extracted from WSI based on cell density. This was followed by mitosis detection using the ResNet CNN architecture. Finally, each WSI is represented by a feature vector that includes the number of mitoses and cells in each patch as well as other statistical derived features. This feature vector is passed to the SVM classifier to predict tumor proliferation.

Methodology

PROPOSED SYSTEM:

Histological slides, stained with hematoxylin and eosin (H&E), were scanned under an automatic microscope at 20x magnification. The cancerous regions of the slide were annotated by experienced pathologists. Since WSI has a resolution of about 10000x200000, the first step is to extract regions of interest (ROI) to reduce computational load. Similar to [16], the extracted tissue area is an area with a gray level less than ± 0.8. Training samples were generated from ROI by extracting 256x256 patches with a step of 196, achieving a sufficient degree of patch overlap. The patch is labeled as a tumor if 75% of its pixels are labeled as a tumor. Two CNN architectures were tested to classify patches into normal and tumor classes: VGG, winner of the 2014 ImageNet contest, and ResNet, which won the ImageNet challenge in 2015. , here we used VGG16 (with 16 layers) and ResNet50 (with 50 layers).). In the VGG architecture, all-composite layers use filters with a small receptive field of size 3x3. Part of the layers of accumulation is followed by maximum spatial clustering on a 2x2 and 2-step window, which effectively reduces the resolution by a factor of 2. 2 layers are fully connected, each layer has 4096 channels, and one layer has 2 sigmoid activated outputs corresponding to normal and tumor layer connected to the last group

layer. An overview of the VGG16 CNN is shown in Figure 2.

The ResNet architecture is built with resnet blocks to solve the problem of accuracy degradation when building deeper architectures with more layers. Instead of learning the mapping function, the residual block adjusts the residual mapping. The main assumption here is that it is easier to find a remaining map than the original map. If H(x) represents the original mapping (Figure 3), the layers overlap the mapping pattern F(x) = H(x)- x. At the top of the ResNet50 network, we added a middle group layer, followed by a fully-connected layer with 2 sigmoid enabled outputs. For both tested architectures, pre-trained weights on ImageNet are used for initialization in order to speed up convergence. Using the CNN output, a 256-fold subsampled tumor heatmap was generated by assigning a probability of being a tumor to each patch.

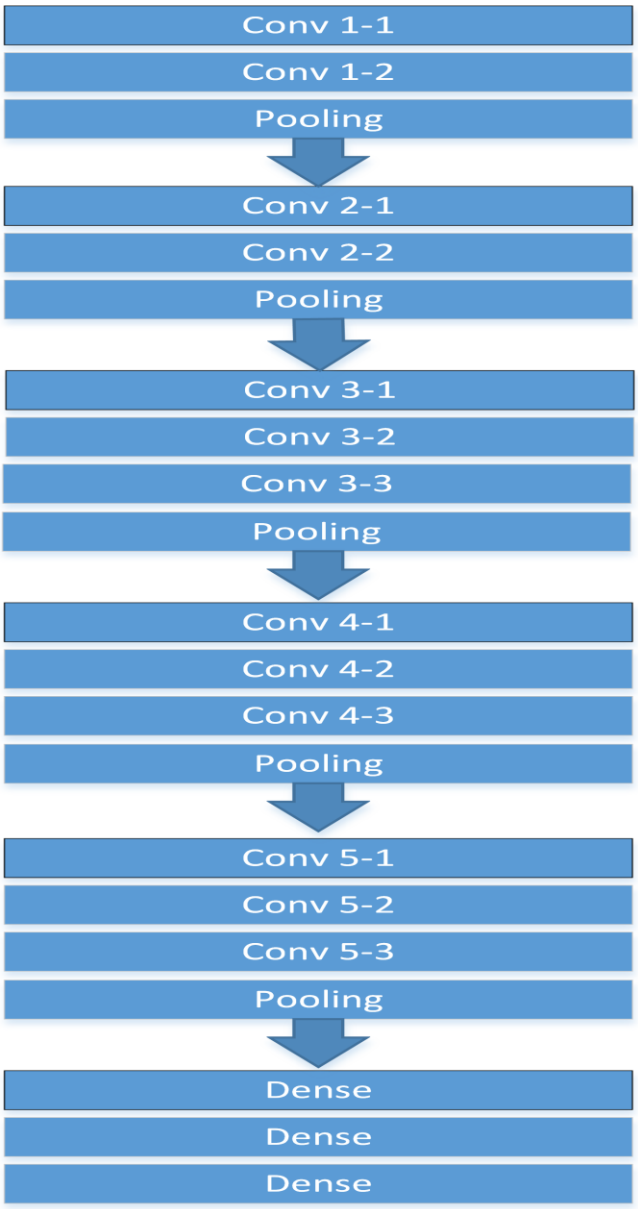

Fig. 2. VGG16 network structure

EXPERIMENTAL RESULTS

The proposed method was trained and evaluated using ACDC @ LUNGHP (automatic detection and classification of cancer in free histopathology databases) [17]. Specifically, we selected the first 25 images in this data set to produce a training set containing 124,434 standard sites and 97,588. 10% of this set is used for verification. The test set contains 28784 patches (14392 per layer) produced in 8 images. SGD was used as a draft and the bulk size was set to 16. The reading level was set at 0.0001 and binary cross-entropy was selected as a loss function. Two networks were built during the 17th century. The performance of the patch level receiver (ROC) was used as a test matrix (Figure 3). It is evident that VGG16 is slightly higher than ResNet50 according to AUC (below the ROC curve). This result is contrary to expected behavior because ResNet achieves the highest accuracy in the ImageNet database (75.2% vs 70.5% with Top1 accuracy and 93% vs 91.2% with Top1 accuracy). namely Top 5). One possible explanation is that the digital pathology domain differs from the ImageNet domain. In other words, the distinction between cancer and plant areas is a different matter than recognizing visual categories such as "dogs", "boards" or "cars" used for verification.

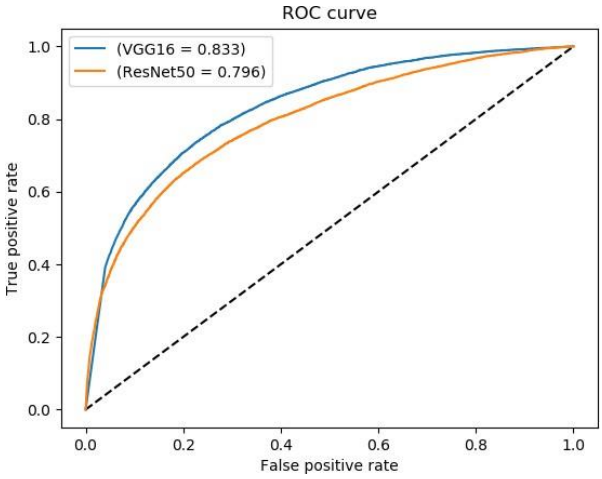

Fig. 4. ROC for tested CNN architectures

TABLE I
TRUE POSITIVE RATES WITH RESPECT TO DIFFERENT FALSE POSITIVE RATES

Method	FP@0.05	FP@0.1	FP@0.5
VGG16	0.427	0.5617	0.9086
ResNet50	0.3780	0.5031	0.8583

Therefore, the better results obtained with ImageNet do not necessarily imply an advantage in lung histopathology. The AUC values obtained for VGG16 correspond to the results presented in [5], while in the same paper the ResNet50 architecture for AUC was 0.8687.

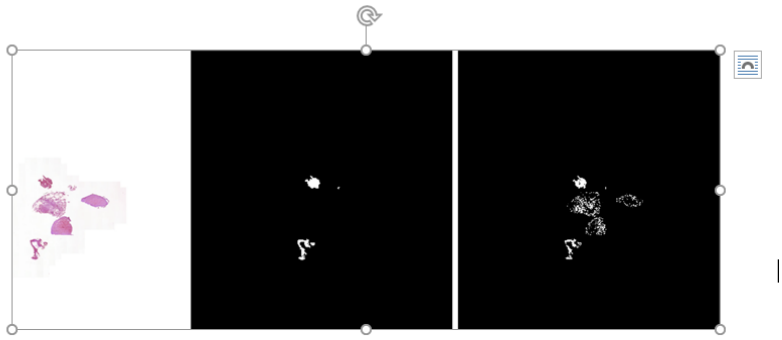

Fig. 5. Visualization of tumor region detection. Original slide image (left), gorund truth tumor regions (middle) and detected regions (right).

TABLE II
PATCH CLASSIFICATION ACCURACY

Method	Patch level accuracy
VGG16	0.7541
ResNet50	0.7205

Table I shows the percentages of true positives and false positives of 0.05, 0.1 , and 0.5. It can be seen that VGG16 has a higher true positive rate than ResNet50. The values obtained for the VGG16 network are the same as those presented in [5], but the true positive values are weaker for ResNet. Similar conclusions can be drawn about the accuracy of the trace classification (Table II). Another factor that can affect the performance of ResNet is that the relatively small training set cannot fully utilize the training capabilities of the deeper architecture. ResNet50 (50 levels) is a deeper network than VGG16 (16 levels), so a larger training dataset is required to achieve maximum classification accuracy. The results show the potential of CNN-based classification in detecting lung cancer, but the method presented is less effective

than that of detecting other types of cancer in the whole slide image. For example , the beast metastatic cancer detection method [14] based on the VGG16 network yielded a patch classification accuracy of 97.9%. This conclusion is also reached in [5] where it is shown that the diagnosis of lung cancer is more difficult. This can be explained by the large variety of templates on different slides. Figure 5 shows a visualization of the proposed cancer detection approach where a 256-fold subsampled heat map is generated by assigning a probability of being a tumor to each patch. Brighter pixels correspond to a higher probability of tumor areas.

CONCLUSIONS

In this paper, we have proposed a fully automated deep learning-based method to detect lung cancer on whole-slide histopathology images. The CNN architectures VGG16 and ResNet50 were compared and the former showed higher AUC and patch classification accuracy. The presented results suggest that the complex neural network has the potential to diagnose lung cancer from the whole slide image, but further efforts are needed to increase the classification accuracy. In future work, the next steps will be to increase the size of the training set, add a function to increase the image, and normalize the spot. In addition, we will try to train from scratch instead of using pre-trained weights on ImageNet.

REFERENCES

[1]Rebecca Siegel, Deepa Naishadham, and Ahmedin Jemal. Cancer statistics, 2013. CA: a cancer journal for clinicians, 63(1):11–30, 2013.

[2]National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine, 365(5):395–409, 2011.

[3]J St Thomas, D Lamb, T Ashcroft, B Corrin, CW Edwards, AR Gibbs, WE Kenyon, RJ Stephens, and WF Whimster. How reliable is the diagnosis of lung cancer using small biopsy specimens? report of a ukcccr lung cancer working party. Thorax, 48(11):1135–1139, 1993.

[4]Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. Annual review of biomedical engineering, 19:221–248, 2017.

[5]Zhang Li, Zheyu Hu, Jiaolong Xu, Tao Tan, Hui Chen, Zhi Duan, Ping Liu, Jun Tang, Guoping Cai, Quchang Ouyang, et al. Computer-aided diagnosis of lung carcinoma using deep learning-a pilot study. arXiv preprint arXiv:1803.05471, 2018.

[6]Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639):115, 2017.

[7]Stephen Baek, Yusen He, Bryan G Allen, John M Buatti, Brian J Smith, Kristin A Plichta, Steven N Seyedin, Maggie Gannon, Katherine R Cabel, Yusung Kim, et al. What does ai see? deep segmentation networks discover biomarkers for lung cancer survival. arXiv preprint arXiv:1903.11593, 2019.

[8]Wei Li, Peng Cao, Dazhe Zhao, and Junbo Wang. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. Computational and mathematical methods in medicine, 2016, 2016.

[9]Christopher D Malon and Eric Cosatto. Classification of mitotic figures with convolutional neural networks and seeded blob features. Journal of pathology informatics, 4, 2013.

[10] Yuanpu Xie, Fuyong Xing, Xiangfei Kong, Hai Su, and Lin Yang. Beyond classification: structured regression for robust cell detection using convolutional neural network. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 358–365. Springer, 2015.

[11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1– 9, 2015.

[12] Karen Simonyan and Andrew Zisserman. Very deep convolu- tional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[14] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H Beck. Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718, 2016.

[15] Kyunghyun Paeng, Sangheum Hwang, Sunggyun Park, and Minsoo Kim. A unified framework for tumor proliferation score prediction in breast histopathology. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pages 231–239. Springer, 2017.

[16] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of pathology informatics, 7, 2016.

[17] Automatic Cancer Detection and Classification in Whole-slide Lung Histopathology (ACDC@LUNGHP). https://acdc-lunghp. grand-challenge.org/Challenge/, 2019. [Online; accessed 1-April-2019].