Osteosarcoma detection using Deep Learning on Histopathology Images

Hosein Barzekar, D M Anisuzzaman, Zeyun Yu

Department of Computer Science

Introduction

Primary bone tumor accounts for 5-10% of all new pediatric cancer diagnosis in the United States. Osteosarcoma is the most common form of malignant primary bone tumor. Osteosarcoma cancer usually occurs in the metaphysis of long bones on lower limbs (femur and proximal tibia), consisting of 40%-50% of the total cases [1]. The symptom usually begins with mild localized bone pain, redness, and warmth at the site of the tumor [2].

With the advent of whole slide imaging (WSI), microscopic image-based analysis in contemporary medicine has been the foundation of cancer diagnosis [3]. Pathologists must spend an extremely long time examining a large number of slides. Detecting the nuances of histological images can be difficult [4].

The rapid growth of AI in health and pathology has affected several areas, and image analysis is not an exception. By using the high-resolution histological image, AI techniques can differentiate the nuances of images between benign tissues and malignant tissues, while maintaining very high accuracy [5]. Since 2010, remarkable progress has been achieved in medical images, primarily due to the availability of large-scale datasets and deep convolutional neural networks (CNNs) in the computer science area [6]. We believe the adoption of computer-aided technology using CNN can significantly reduce the pathologist's workload and achieve a better prognosis of patients.

Methods

Dataset:

The dataset used in the study was obtained from the work of Arunachalam et al. Tumor samples from the Children's Medical Center, Dallas, were collected from the pathology reports of the osteosarcoma resection for 50 patients treated between 1995 and 2015. 40 WSIs of the digitized images representing tumor heterogeneity and response properties have been selected in the study. In each WSI, 30 1024×1024-pixel image tiles were randomly selected at the 10X magnification factor.

Methods

Model Architecture:

We have used Keras applications for importing VGG 19 model. Pre-trained weights have been used for model training. We have discarded the fully connected layer along with output layer of the VGG 19 model. We have added two fully connected layers after the last "maxpool" layer.

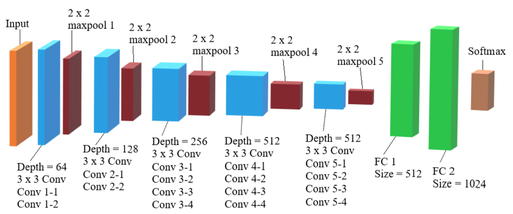
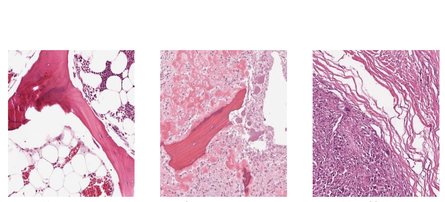


Figure 1: VGG 19 Network Architecture



(a) Non tumor (b) Necrotic tumor (c) Viable tumor

Figure 2: Data Samples

Results

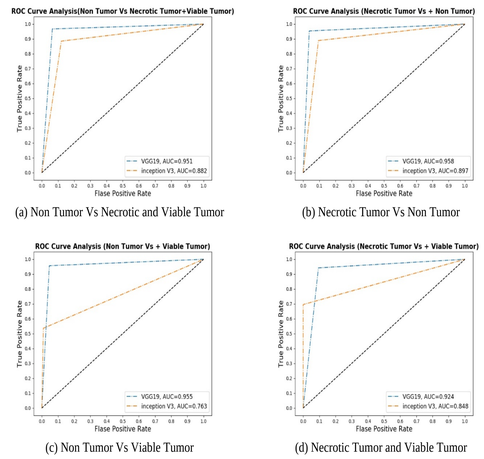
Setup

We performed four binary classifications and a multiclass (three classes) classification. In each classification, we applied two models: VGG 19 and Inception V3, where original images of 1024 × 1024 pixels were used for model training, validation, and evaluation.

Two-class classifications are evaluated on the following datasets: 1) Non-Tumor (NT) versus Necrotic Tumor (NCT) and Viable Tumor (VT), 2) Necrotic Tumor versus Non-Tumor, 3) Viable Tumor versus Non-Tumor, and 4) Necrotic Tumor versus Viable Tumor. We also performed the multiclass classification among the three classes: NT, NCT and VT.

Table 1: Precision, Recall, and F1-Score

Non-Tumor versus Necrotic Tumor and Viable Tumor						
	Non-Tumor			Necrotic and Viable Tumor		
Networks	Precision	Recall	F1	Precision	Recall	F1
VGG 19	**0.96**	**0.94**	**0.95**	**0.94**	**0.97**	**0.96**
Inception V3	0.87	0.88	0.88	0.89	0.89	0.89
Necrotic Tumor versus Non-Tumor						
	Necrotic Tumor			Non-Tumor		
Networks	Precision	Recall	F1	Precision	Recall	F1
VGG 19	**0.91**	**0.96**	**0.94**	**0.98**	**0.95**	**0.97**
Inception V3	0.80	0.91	0.85	0.95	0.89	0.92
Viable Tumor versus Non-Tumor						
	Non-Tumor			Viable Tumor		
Networks	Precision	Recall	F1	Precision	Recall	F1
VGG 19	**0.97**	0.95	**0.96**	0.93	**0.96**	**0.94**
Inception V3	0.77	**0.99**	0.87	**0.97**	0.54	0.69
Necrotic Tumor versus Viable Tumor						
	Necrotic Tumor			Viable Tumor		
Networks	Precision	Recall	F1	Precision	Recall	F1
VGG 19	**0.92**	0.91	**0.91**	0.93	**0.94**	**0.94**
Inception V3	0.72	**1.00**	0.83	**1.00**	0.70	0.82
Multiclass						
	Necrotic Tumor			Non-Tumor			Viable Tumor		
Networks	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
VGG 19	**0.92**	**0.91**	**0.91**	**0.95**	**0.95**	**0.95**	**0.93**	**0.94**	**0.94**
Inception V3	0.59	0.83	0.69	0.88	0.86	0.87	0.88	0.62	0.73



(a) Non Tumor Vs Necrotic and Viable Tumor (b) Necrotic Tumor Vs Non Tumor

(c) Non Tumor Vs Viable Tumor (d) Necrotic Tumor and Viable Tumor

Figure 3: ROC and AUC for all classifications.

Discussion

From Table 1, we can see that for some cases precision and recall are high for Inception V3. But the interesting fact is that all the f1 scores are higher for VGG 19 model. Hence, in balance VGG 19 beats inception V3 by a huge margin.

From Figure 3, we can see that VGG 19 has the highest AUC value for all binary (two-class) classifications. The AUC values are impressive (0.95, 0.96, 0.96, and 0.92 for non-tumor versus necrotic tumor and viable tumor, necrotic tumor versus non-tumor, viable tumor versus non-tumor, and necrotic tumor versus viable tumor classifications respectively), which assures us with great reliability. So, from all the above analytical discussion, it is safe to say that VGG 19 works well for all classifications.

To the best of our knowledge, this is the first pipeline that have been used in VGG19 and Inception architecture in Deep learning to recognize the osteosarcoma malignancy. The adjusted model can identify the minimal differences of images to predict the early signs of cancer. If the pipeline was deployed in various medical facilities, our model can help histologists as an adjunct tool reducing their extensive work.

Conclusion

In medical image processing, It is important to automate the classification of histological images by computer-aided systems. Automatic diagnosis of histology alleviates the workload and enables pathologists to focus on critical cases. A technique for the automated classification of histologic images has been applied for this study. To classify histological images in the osteosarcoma dataset, the latest VGG 19, and Inception V3 are used.

Literature cited

[1] A. J. Chou, D. S. Geller, and R. Gorlick, "Therapy for osteosarcoma," Pediatric Drugs, vol. 10, no. 5, pp. 315-327, 2008.

[2] P. P. Lin and S. Patel, "Osteosarcoma," in Bone Sarcoma: Springer, 2013, pp. 75-97.

[3] D. S. Geller and R. Gorlick, "Osteosarcoma: a review of diagnosis, management, and treatment strategies," Clin Adv Hematol Oncol, vol. 8, no. 10, pp. 705-718, 2010

[4] P. Picci, "Osteosarcoma (osteogenic sarcoma)," Orphanet journal of rare diseases, vol. 2, no. 1, p. 6, 2007.

[5] S. S. Aboutalib, A. A. Mohamed, W. A. Berg, M. L. Zuley, J. H. Sumkin, and S. Wu, "Deep learning to distinguish recalled but benign mammography images in breast cancer screening," Clinical Cancer Research, vol. 24, no. 23, pp. 5902-5909, 2018.

[6] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review, Neural computing," ed: MIT Press Journals, 2017.

Acknowledgments

Thanks the Children's Medical Center, Dallas, for providing the dataset and making it public.

For further information

A further study is to compare our model's performance with expert histologists. The comparison can make sure this tool is capable of detecting new malignant cases in clinical practices.