Automatic Histopathology Image Analysis with CNNs

Le Hou[a], Kunal Singh, Dimitris Samaras[a], Tahsin M. Kurc[b,d], Yi Gao[b,a,c], Roberta J. Seidman[e], Joel H. Saltz[b,a,e,f]
[a]Dept of Computer Science, Stony Brook University
[b]Dept of Biomedical Informatics, Stony Brook University
[c]Dept of Applied Mathematics and Statistics, Stony Brook University
[d]Oak Ridge National Laboratory
[e]Dept of Pathology, Stony Brook Hospital
[f]Cancer Center, Stony Brook Hospital

***Abstract*—We define Pathomics as the process of high throughput generation, interrogation, and mining of quantitative features from high-resolution histopathology tissue images. Analysis and mining of large volumes of imaging features has great potential to enhance our understanding of tumors. The basic Pathomics workflow consists of several steps: segmentation of tissue images to delineate the boundaries of nuclei, cells, and other structures; computation of size, shape, intensity, and texture features for each segmented object; classification of images and patients based on imaging features; and correlation of classification results with genomic signatures and clinical outcome. Executing a Pathomics workflow on a dataset of thousands of very high resolution (gigapixels) and heterogeneous histopathology images is a computationally challenging problem. In this paper, we use Convolutional Neural Networks (CNN) for automatic recognition of nuclear morphological attributes in histopathology images of glioma, the most common malignant brain tumor. We constructed a comprehensive multi-label dataset of glioma nuclei and applied two CNN based methods on this dataset. Both methods perform well recognizing some but not all morphological attributes and are complementary with each other.**

Keywords—Pathomics; Nucleus Classification; Convolutional Neural Network

I. INTRODUCTION

Radiomics has emerged as a highly promising approach for providing a comprehensive quantification of tumor properties at macro-scales through high-throughput generation and interrogation of medical imaging features [15-17]. We define *Pathomics* as the process of high throughput generation, interrogation, and mining of quantitative features from high-resolution tissue images – the histopathology equivalent of Radiomics.

Integrative, quantitative analyses of relationships among histopathology, spatially mapped molecular data, and clinical data have great potential to significantly enhance our understanding of disease mechanisms. Such analyses are motivated by studies that investigate tumor initiation, progression, heterogeneity, therapeutic target validation, cancer proliferation and metastasis and by research to characterize outcome and response to treatment using integrated morphology and molecular data.

Basic Pathomics workflows consist of a series of image segmentation, feature computation, classification, and correlation steps. The image segmentation step delineates the boundaries of nuclei, cells, and meso-scale structures such as crypts and ducts. Advanced digital microscopes can capture very high-resolution images (ranging from 20Kx20K to 100Kx100K pixel resolutions) from whole slide tissue specimens. Segmentation of a single image can generate hundreds of thousands to millions of nuclei. The feature computation step computes a set of quantitative (size, shape, intensity, texture) attributes for each segmented structure and images. The classification step makes use of the features and image data to categorize objects, images and subjects from which the images are obtained. The correlation step compares and looks for relationships between classifications based on imaging features and genomic signatures and clinical outcome data. The workflow is generally an iterative one (as shown in Figure 1), because many image analysis methods are sensitive to input parameters and input data. The goal of the iterative process is to provide feedback and feedforward information to generate robust results efficiently.

In this paper we describe an application of Convolutional Neural Networks (CNNs) in the feature computation and classification step for automatic recognition of nuclear morphological features (also referred to here as attributes) in whole slide tissue images from Glioma cancer patients. Glioma is a malignant brain tumor that rises from glial cells [2] and is the leading cause of cancer-related deaths in people under age 20 [1]. In glioma histopathology images, morphological attributes of nuclei provide rich information for diagnosing and classifying glioma patients into respective subtypes and grades [3]. CNN [10] is the backbone of state-of-the-art methods in many automatic image recognition applications. Given a training set of images with ground truth labels, a CNN can be trained to recognize the labels automatically given an input image. For example, in a ductal

This work was supported in part by 1U24CA180924-01A1 from the NCI, R01LM011119-01 and R01LM009239 from the NLM, and the Subsample DIGITEO project.

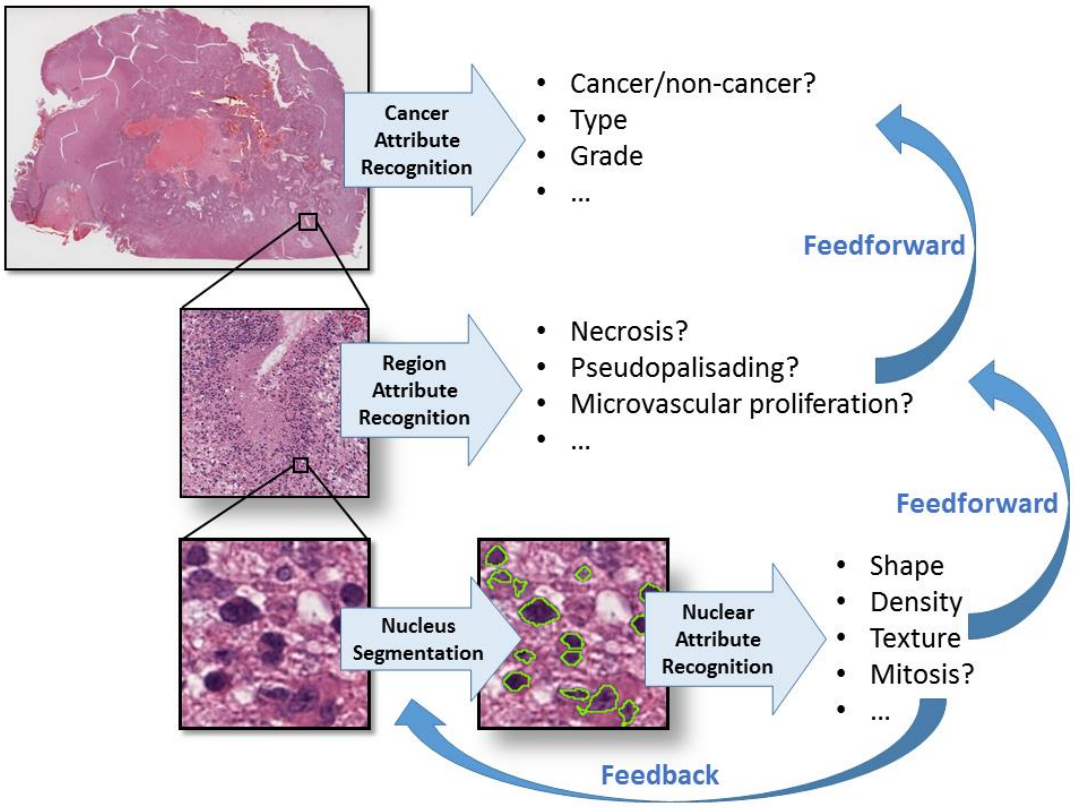


Figure 1. A core Pathomics workflow for segmentation, feature computation and classification steps.

carcinoma detection problem, a CNN model has been shown to achieve an F-measure of 71.80% which significantly outperforms previous methods [12]. Recently, CNN models have achieved the best results in multiple MICCAI challenges; for example, a multi-column CNN detection method won the MICCAI mitosis detection challenge in breast cancer [13]. CNNs with multiple-instance learning [14] achieved state-of-the-art performance recognizing cancer subtypes.

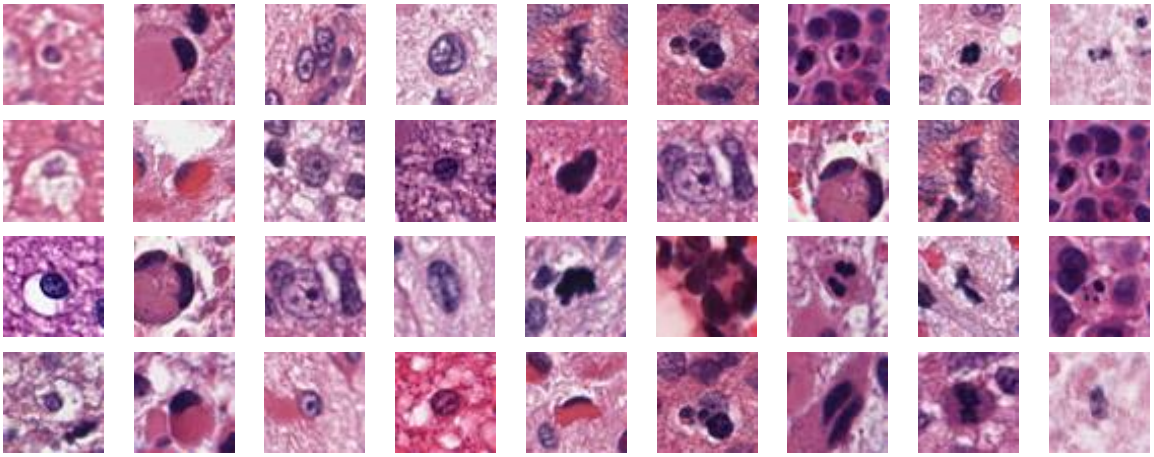
Our approach is designed to automatically recognize the nuclear morphological features of a given glioma image using CNN. These features are Perinuclear halos, Gemistocyte, Nucleoli, Grooved, Hyperchromasia, Overlapping nuclei, Multinucleation, Mitosis and Apoptosis.

Our work has two contributions. **The first contribution** is the multi-label modeling of the feature recognition and classification problem. Existing classification methods are single-label learning algorithms [5,6]. In other words, they only recognize one nuclear attribute at a time. However, one nucleus can have multiple morphological attributes. Thus, we model this as a multi-label learning algorithm [4]. **The second contribution** is that the proposed approach recognizes nine subtle, important and common morphological attributes of nuclei in glioma histopathology images with good accuracy. Our experimental results show an averaged Area Under the ROC Curve (AUC) of 0.8712. We build a multi-label glioma nuclei dataset that covers nine nuclear morphological attributes in six glioma subtypes. The dataset contains 2078 glioma nucleus images each with nucleus in the image center. Figure 2 shows examples for each of the nine morphological attributes. Existing automatic nuclear attribute recognition methods only recognize a subset of important nuclear morphological attributes for a subset of glioma subtypes. For example, Thibault et al. [5] recognizes healthy and pathological nuclei. Kong et al. [6] classifies six morphological attributes in diffuse glioma only.

The rest of this paper is organized as follows. Section 2 introduces our CNN-based approach. Section 3 presents the experimental results. Section 4 concludes the paper.

II. CNN-BASED METHODS FOR NUCLEAR ATTRIBUTE CLASSIFICATION

We present two CNN-based methods for nucleus attribute classification: a semi-supervised CNN and a pre-trained CNN with Support Vector Machine (SVM). To model inter-attribute correlations in our multi-label learning problem, we adopt the two-round training method [11]. In particular, instead of training one classification model, we train two models. The predicted class distributions of the first classification model are used as features for the second classification model. In this way, the second model can learn the inter-attribute correlations.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 2. Examples of glioma nucleus images. Each column shows two images of one morphological attribute. The morphological attribute describes the nucleus in the center of image. The attributes are (a) Perinuclear halos (b) Gemistocyte (c) Nucleoli (d) Grooved (e) Hyperchromasia (f) Overlapping nuclei (g) Multinucleation (h) Mitosis (i) Apoptosis. Note that images can have multiple labels. For example, Mitosis are usually also Hyperchromasia.

A. Semi-supervised CNN

In contrast to fully-supervised methods, semi-supervised methods [18] utilize unlabeled instances (instances without ground truth labels) to boost classification performance. In the case of nuclear attribute classification, we have millions of nucleus images with unknown ground truth. To utilize these images, we use a Convolutional Auto-Encode (CAE) method [19] to initialize a classification CNN. A CAE is a specific type of neural network. In general, a CAE encodes input images as a set of activations of encoding neurons. It then decodes these activations into output (reconstructed) images. In order to achieve nontrivial solutions, in most cases, the number of encoding neurons is significantly smaller than the number of input image pixels. By minimizing the error (difference) between the output and input images, one trains a CAE to capture an intrinsic representation of input images. In other words, to reconstruct input images with encoding neurons, the CAE learns the patterns (appearance, texture, etc.) of input images. Figure 3 shows examples of input (original) and output (reconstructed) images.

In our work, we train an unsupervised CAE on millions of unlabeled nucleus images. Then the model parameters of our supervised nuclear classification CNN are initialized as the same as the model parameters in CAE. Figure 4 illustrates this process. The appearance and texture information of nuclei is learnt by the CAE model. This information is passed to the supervised classification CNN during its initialization step. The architecture of the CAE and CNN is similar to the VGG network [8] with fewer convolutional filters.

B. Pretrained CNN features with SVM

It has been shown that the activations of hidden neurons can be viewed as features extracted from input images [20]. One can use CNNs as feature extractors and apply other supervised models such as Support Vector Machine (SVM) [9] for image classification. The advantage of this method is that no CNN training is required on the application dataset. In this work, we use the VGG 16-layer network [8] as a feature extractor. In particular, activations from neurons previous to the output layer are used as features. We apply this feature extractor on all nucleus images. Each image is represented by a CNN feature vector of length 4096. We then use this 4K features to train and test an SVM with Radial Basis Function (RBF) kernel. For SVM hyperparameters parameters that must be predefined before model training (in contrast to parameters that are learnt by model training), we adjust those parameters to maximize the experimental cross-validation error. In particular, we split the training set into five non-overlapping subsets. We train SVMs sets of hyperparameter assignments on four subsets and validate the classification performance on the remaining one. This procedure is repeated for five times before the classification results are averaged. The best-performing set of hyperparameter assignment is then used to train the SVM on the entire training set. For each nuclear attribute, we train a binary SVM classifier that outputs either does or does not have the target attribute, ignoring other nuclear attributes. Therefore, we have ten binary SVMs for ten nuclear attributes.

Morphological Attributes	#. Present	#. Absent
Perinuclear halos	78	2000
Gemistocyte	51	2027
Nucleoli	77	2001
Grooved	14	2064
Hyperchromasia	505	1573
Overlapping nuclei	105	1973
Multinucleation	43	2035
Mitosis	53	2025
Apoptosis	20	2058
No nucleus	545	1533

Table 1. The distribution of nuclear attributes in our nuclear attribute classification dataset. In this multi-label dataset, one nucleus image can have multiple morphological attributes.

III. Experiments

We apply the semi-supervised CNN and the pretrained CNN with SVM [9] to recognize nuclear morphological attributes. Our experiments achieved promising results.

A. *Nuclear attribute classification dataset*

We aim to build a dataset of thousands of nucleus images with nuclear attribute ground truth. We first applied an automatic nucleus segmentation method to extract millions of nuclei. Segmented nuclei are stored as small images, which we refer to as *nucleus images*. Then a pathologist and a graduate student viewed these images together and assigned nuclear attributes to two thousand nucleus images. In the rest of this section, we describe this process in more detail.

To automatically segment nuclei, the color of a tissue image is normalized to a Hematoxylin and Eosin stained template image in the L*a*b color space. Then, the

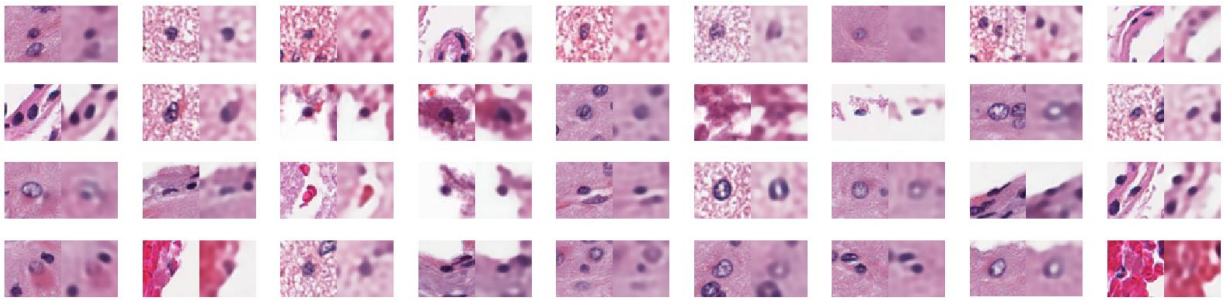


Figure 3. Randomly selected examples of original and CAE reconstructed images. Each 50 by 50 RGB images (shown on the left) is reconstructed (shown on the right) from decoding activations of 200 encoding neurons. We can see that though the reconstructed images are blur, most of the structural, appearance, and texture information are reconstructed. A classification CNN initialized by the CAE encodes this information.

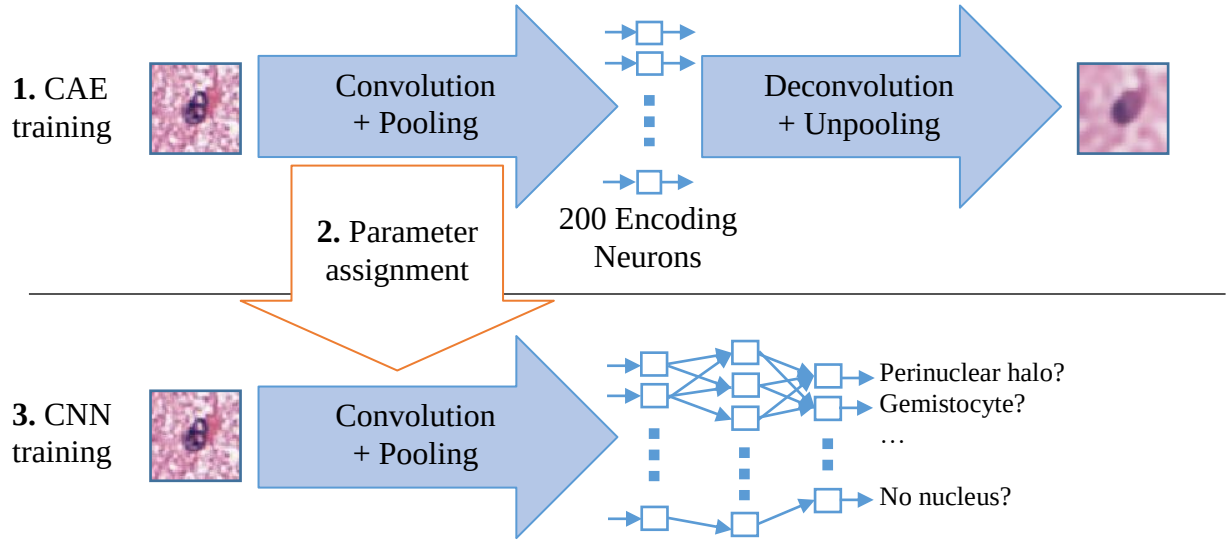


Figure 4. Initializing the nuclear classification CNN with Convolutional Auto-encoder (CAE). Top: training a convolutional auto-encoder. Bottom: initializing the nuclear classification CNN. The model parameters in the "convolution + pooling" part of the classification CNN are assigned by model parameters in the CAE.

hematoxylin channel is extracted through a color decomposition process. After that, the optimal threshold in the hematoxylin channel is computed and a localized region based level set method is used to determine the contour of each nucleus. In cases where several nuclei are clumped together, a hierarchical mean shift algorithm is used to separate the clump into individual nucleus. We then extract nucleus images of 50 by 50 pixels around the centers of automatically segmented nuclei. The extracted images are in RGB space (Figure. 2).

A graduate student and a pathologist view samples of extracted nucleus images together. They assign all suitable attributes to each image. In case there are multiple nuclei presented in a nucleus image, they assign attributes according to the nucleus that is closest to the image center. Note that because the nucleus segmentation method is not perfect, some images do not contain any nucleus. To address this problem, we introduce the "no nucleus" attribute. If an images is labeled as "no nucleus", no other attributes can be assigned. The distribution of attributes is summarized in Table 1.

B. Classifier implementation

For the semi-supervised CNN, we use Theano [22] for the CAE and CNN implementation. To avoid overfitting, we apply data augmentation. Input images are randomly rotated, flipped. The color of input images are randomly adjusted. We use stochastic gradient descent with momentum for optimization and backpropagation for computing the gradient in the parameters space. The learning rates of the CAE and CNN are 0.001 and 0.0005 respectively. The momentum is 0.975. It takes around 12 hours to train the CAE and 0.5 hours to train the CNN. Note that we only train one CAE. All CNNs are initialized by the same CAE.

For the pretrained CNN with SVM, we use MatConvNet [21] to extract VGG 16-layer network features and LIBSVM [9] for the SVM implementation.

C. Experimental results

We apply five random-split validation and average the results. We use the Area Under the ROC Curve (AUC) as the evaluation metric. The AUC ranges from 0.5 (random prediction) to 1.0 (perfect prediction). Table 2 shows the AUC results of both methods. We achieved an encouraging averaged AUC of 0.8712. Notice that each method performs well on some but not all morphological attributes. Therefore, we achieved a better AUC of 0.9109 by combining these two methods.

	AUC		
Morphological Attributes	Semi-supervised CNN	VGG16 + SVM	Best of two (per attribute)
Perinuclear halos	0.8789	**0.9257**	0.9257
Gemistocyte	0.8026	**0.9548**	0.9548
Nucleoli	0.8366	**0.9076**	0.9076
Grooved	**0.8956**	0.7296	0.8956
Hyperchromasia	**0.9450**	0.8854	0.9450
Overlapping nuclei	**0.8969**	0.8305	0.8969
Multinucleation	0.7329	**0.7507**	0.7507
Mitosis	**0.8731**	0.8559	0.8731
Apoptosis	0.8676	0.9767	0.9767
No nucleus	**0.9828**	0.9639	**0.9828**
Averaged AUC	0.8712	**0.8616**	**0.9109**

Table 2. The results of morphological attribute recognition. We achieved an encouraging averaged AUC of 0.8712. Notice that both methods perform well on some but not all morphological attributes and are complementary with each other.

IV. CONCLUSIONS

In this paper, we discussed the general workflow of Pathomics. We employed a Convolutional Neural Network (CNN) in the classification step of a Pathomics workflow for automatic nuclear attribute recognition in glioma histopathology images and achieved promising results. In particular, we constructed a comprehensive multi-label glioma (the most common brain cancer) nuclear morphological attribute recognition dataset and applied two CNN based methods on this dataset. Both CNN based methods perform well recognizing some but not all morphological attributes and are complementary with each other.

REFERENCES

[1] Brain tumor statistics. http://www.abta.org/about-us/news/brain-tumor-statistics/

[2] American brain tumor association. http://www.abta.org/brain-tumor-information/types-of-tumors/glioma.html

[3] Louis, David N., Hiroko Ohgaki, Otmar D. Wiestler, Webster K. Cavenee, Peter C. Burger, Anne Jouvet, Bernd W. Scheithauer, and Paul Kleihues. The 2007 WHO classification of tumours of the central nervous system. Acta neuropathologica. 2007.

[4] Zhang, Min-Ling, and Zhi-Hua Zhou. A review on multi-label learning algorithms. Knowledge and Data Engineering. 2014.

[5] Thibault, Guillaume, Caroline Devic, Jean-François Horn, Bernard Fertil, Jean Sequeira, and Jean-Luc Mari. Classification of cell nuclei using shape and texture indexes. 2008.

[6] Kong, Jun, Lee Cooper, Fusheng Wang, Candace Chisolm, Carlos Moreno, Tahsin Kurc, Patrick Widener, Daniel Brat, and Joel Saltz. A comprehensive framework for classification of nuclei in digital microscopy imaging: An application to diffuse gliomas. In Biomedical Imaging: From Nano to Macro. 2011.

[7] Masci, Jonathan, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In Artificial Neural Networks and Machine Learning. 2011.

[8] Simonyan, Karen, and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv. 2014.

[9] Chang, Chih-Chung, and Chih-Jen Lin. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology. 2011.

[10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. NIPS. 2012.

[11] Godbole, Shantanu, and Sunita Sarawagi. Discriminative methods for multi-labeled classification. In Advances in Knowledge Discovery and Data Mining. 2004.

[12] Cruz-Roa, Angel, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi. Automatic detection of invasive ductal

carcinoma in whole slide images with convolutional neural networks. SPIE Medical Imaging. International Society for Optics and Photonics. 2014.

[13] Cireşan, Dan C., Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2013.

[14] Hou, Le, Dimitris Samaras, Tahsin M. Kurc, Yi Gao, James E. Davis, and Joel H. Saltz. Efficient Multiple Instance Convolutional Neural Networks for Gigapixel Resolution Image Classification. Computer Vision and Pattern Recognition (CVPR). 2016.

[15] R. Gillies, Radiomics: informing cancer heterogeneity, in J. Nucl Med, 2013, p. 31.

[16] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, et al., Radiomics: the process and the challenges, Magn Reson Imaging, vol. 30, pp. 1234-48, Nov 2012.

[17] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat Commun, vol. 5, p. 4006, 2014.

[18] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews]. IEEE Transactions on Neural Networks 20, no. 3 (2009): 542-542.

[19] Masci, Jonathan, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. InInternational Conference on Artificial Neural Networks, pp. 52-59. Springer Berlin Heidelberg, 2011.

[20] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.

[21] Vedaldi, Andrea, and Karel Lenc. Matconvnet: Convolutional neural networks for matlab. In Proceedings of the 23rd ACM international conference on Multimedia, pp. 689-692. ACM, 2015.

[22] Team, The Theano Development, Rami Al-Rfou, Guillaume Alain, Amjad Almahairi, Christof Angermueller, Dzmitry Bahdanau, Nicolas Ballas et al. Theano: A Python framework for fast computation of mathematical expressions. arXiv preprint arXiv:1605.02688 (2016).