EBioMedicine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images

Pegah Khosravi[a,b,1], Ehsan Kazemi[c,1], Marcin Imielinski[d,e,f,g], Olivier Elemento[a,b,d,g,*], Iman Hajirasouliha[a,b,d,g,*]

[a] Institute for Computational Biomedicine, Weill Cornell Medical College, NY, USA
[b] Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA
[c] Yale Institute for Network Science, Yale University, New Haven, CT, USA
[d] Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medical College, NY, USA
[e] Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, NY, USA
[f] The New York Genome Center, NY, USA
[g] The Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA

ARTICLE INFO

Article history:
Received 27 October 2017
Received in revised form 21 December 2017
Accepted 21 December 2017
Available online xxxx

Keywords:
Biomarkers
Classification
Convolutional Neural Network
Deep learning
Digital pathology imaging
Tumor heterogeneity

ABSTRACT

Pathological evaluation of tumor tissue is pivotal for diagnosis in cancer patients and automated image analysis approaches have great potential to increase precision of diagnosis and help reduce human error.
In this study, we utilize several computational methods based on convolutional neural networks (CNN) and build a stand-alone pipeline to effectively classify different histopathology images across different types of cancer.
In particular, we demonstrate the utility of our pipeline to discriminate between two subtypes of lung cancer, four biomarkers of bladder cancer, and five biomarkers of breast cancer. In addition, we apply our pipeline to discriminate among four immunohistochemistry (IHC) staining scores of bladder and breast cancers.
Our classification pipeline includes a basic CNN architecture, Google's Inceptions with three training strategies, and an ensemble of two state-of-the-art algorithms, Inception and ResNet. Training strategies include training the last layer of Google's Inceptions, training the network from scratch, and fine-tunning the parameters for our data using two pre-trained version of Google's Inception architectures, Inception-V1 and Inception-V3.
We demonstrate the power of deep learning approaches for identifying cancer subtypes, and the robustness of Google's Inceptions even in presence of extensive tumor heterogeneity. On average, our pipeline achieved accuracies of 100%, 92%, 95%, and 69% for discrimination of various cancer tissues., subtypes, biomarkers, and scores, respectively. Our pipeline and related documentation is freely available at https://github.com/ih-_lab/CNN_Smoothie.

1. Introduction

Evaluation of microscopic histopathology slides by experienced pathologists is currently the standard procedure for establishing a diagnosis and identifying the subtypes of different cancers. Visual-only assessment of well-established histopathology patterns is typically slow, and is shown to be inaccurate and irreproducible in certain diagnosis cases of tumor subtypes and stages (Mosquera-Lopez et al., 2015).

Several recent studies attempted to employ machine learning approaches for determining subtypes of malignancies (Esteva et al., 2017; Yu et al., 2016). These computational approaches can be complementary with other clinical evaluation methods to improve pathologists' knowledge of the disease and improve treatments (Felipe De Sousa et al., 2013; Beck et al., 2011). For example, previous studies have shown more accurate diagnosis results are derived by integrating information extracted from computational pathology with patients' clinical data for various cancer types such as prostate cancer (Bhargava et al., 2011; Doyle et al., 2012), lung cancer (Hamilton et al., 2015), breast cancer (Wang et al., 2013; Dong et al., 2014), colorectal cancer (Korbar et al., 2017), and ovarian cancer (Janowczyk et al., 2012). In particular, computerized image

* Corresponding authors.
[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2017.12.026

processing technology has been shown to improve performance, correctness, and robustness in histopathology assessments (Lemaître et al., 2015).

While new advanced approaches have improved image recognition (e.g., normal versus cancerous), the image interpretation of heterogeneous populations still suffers from lack of robust computerization approaches (Razzak et al., 2017; Carneiro et al., 2017; Gurcan, 2016; Jiang et al., 2016). Current available automatic methods focus on classification of just one type of cancer versus the corresponding normal condition. Although these studies achieved reasonable accuracy in detecting normal or cancerous conditions in specific kind of cancers, leveraging methods such as training Convolutional Neural Networks (CNNs)(LeCun et al., 1998), they have certain limitations which we address in this work:

1. Developing *ensemble* deep learning methods to employ state-of-the-art algorithms for improving training approaches in diagnosis and detection of various cancer subtypes (e.g., adenocarcinoma versus cell squamous lung cancer).
2. Improving the speed of deep learning, and investigating the trade-offs between performance (i.e., the size of the training set) and efficiency (i.e., the training speed).
3. Making decisions on selecting proper neural networks for different types of datasets.

One of the main challenges of computational pathology is that tumor tissue images often vary in color and scale batch effects across different research laboratories and medical facilities due to differences in tissue preparation methods and imaging implements (Kothari et al., 2013). Furthermore, erroneous evaluation of histopathology images and decision-making using tissue slides containing millions of cells can be time-consuming and subjective (Yu et al., 2016; Kothari et al., 2013). In this regard, utilization of the deep learning approaches with sufficient number of images to untangle color information can improve the computational approaches within a reasonable amount of time.

In addition, cancer is known to be a heterogeneous disease. i.e., a high degree of genetic and phenotypic diversity exists “within tumors” (intra-tumor) and “among tumors” (inter-tumor) (Polyak, 2011). Tumor heterogeneity leads to an important effect of disease progression and resistant responses to targeted therapies (Hardiman et al., 2016). We also aim to evaluate deep learning approaches for discrimination of digital pathology images from intra- and inter-tumor heterogeneous samples.

Deep learning approaches are emerging as leading machine-learning tools in medical imaging where they have been proven to produce precious results on various tasks such as segmentation, classification, and prediction (Greenspan et al., 2016). In this paper, we present an innovative deep learning based pipeline, CNN_Smoothie, to discriminate various cancer tissues, subtypes, and their relative staining markers and scores. We utilize the pathological images which stained by immunohistochemical markers of tumor differentiation to train CNNs for analyzing and identifying specific clinical patterns in different staining markers and scores of breast and bladder cancers. In addition, we applied deep learning methods on immunohistochemistry (IHC) and hematoxylin & esoin (H&E) stained images of squamous cell carcinoma and lung adenocarcinoma to investigate the performance of various classifiers.

This is a comprehensive stud of applying a wide range of CNN architectures (all integrated in a single pipeline) on histopathology images from multiple different datasets. We evaluate performance of different architectures to detect and diagnosis of tumor images. Our results clearly demonstrate the power of deep learning approaches for distinguishing different cancer tissues, subtypes, IHC markers and their expression scores. Source codes and documentation of our pipeline containing training, evaluation and prediction methods are publicly available at https://github.com/ih-_lab/CNN_Smoothie.

2. Materials and Methods

2.1. Histopathology Images Resource

Our datasets come from a combination of open-access histopathology images, The Stanford Tissue Microarray Database (TMAD) and The Cancer Genome Atlas (TCGA). A total of 12,139 whole-slide stained histopathology images were obtained from TMAD (Marinelli et al., 2007). TMA database enables researchers have access to bright field and fluorescence images of tissue microarrays. This archive provide thousand human tissues which are probed by antibodies simultaneously for detection of protein abundance (immunohistochemistry; IHC), or by labeled nucleic acids (in situ hybridization; ISH) to detect transcript abundance. The extracted data included samples from three cancer tissues: (1) lung, (2) breast, comprising five biomarker types (EGFR, CK17, CK5/6, ER, and HER2), and (3) bladder with four biomarker types (CK14, GATA3, S0084, and S100P). Characteristics of all three cohorts and the comprised classes are summarized in Table 1. From the extracted TMA datasets, one dataset is stained by H&E method (BladderBreastLung) and one dataset is stained by both H&E and IHC methods (TMAD-InterHeterogeneity). The remaining datasets (BladderBiomarkers, BreastBiomarkers, BladderScores, and BreastScores) are stained by IHC markers including different polyclonal antiserums such as CK14, GATA3, S0084, S100P, EGFR, CK17, CK5/6, ER, and HER2 for their related proteins which play critical roles in tumor progression.

The markers are widely used in clinical immunohistochemistry as biomarkers for detection of various neoplasm types (Higgins et al., 2007; Vandenberghe et al., 2017). Several studies have acquired the expressions of biomarkers in biopsy samples of various cancer types to improve the distinction of specific pathological subtyping and understanding of molecular pathways of different cancers. For example, we can refer to the attempts made to discriminate morphologic subtyping of non-small call lung carcinoma (NSCLC), lung adenocarcinoma (LUAD) versus lung squamous cancer (LUSC) (Scagliotti et al., 2008; Khayyata et al., 2009; Conde et al., 2010; Fatima et al., 2011). Antiserums staining tissue are sub-classified according to the staining grade. Each tissue sample in this cohort was scored by a trained pathologist using a discrete scoring system (0, 1, 2, 3). A score zero represents no significant protein expression (negative) because there is no staining color, whereas a score three indicates high expression. Positive results were scored based on both the extent and the intensity of staining. For score three, intense staining was required in more than 50% of the cells. Other scores including one and two staining comprise in fewer than 50% of the total cells (Higgins et al., 2007).

We also obtained the TCGA (Network et al., 2012, 2014) images by extracting them from the Cancer Digital Slide Archive (CDSA) (Gutman et al., 2013) that is accessible to the public and, at the time of writing this, hosts 31,999 whole-slide images from 32 cancer tissues. For the purpose of this study, we analyze 1520 H&E stained whole-slide histopathology images as well as 1629 H&E stained high resolution image patches (40× magnification) of two TCGA lung cancer subtypes (i.e., LUAD versus LUSC).

2.2. Classification and Diagnostic Framework

This study presents a framework (see Fig. 1) to discriminate different cancer types, subtypes, immunohistochemistry markers, and marker staining scores of histopathology images (Table 1). For the first step of our study, the stained whole-slide images with 1504 × 1440 and 2092 × 975 pixels were obtained from TMA and TCGA databases, respectively. Note that we did not use any pre-processing

Please cite this article as: P. Khosravi et al., Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images, EBioMedicine (2017), https://doi.org/10.1016/j.ebiom.2017.12.026