

Comprehensive analysis of lung cancer pathology images to discover tumor shape features that predict survival outcome

Shidan Wang^{1, #}, Alyssa Chen^{1, 2, #}, Lin Yang^{1, 3}, Ling Cai^{1, 4}, Yang Xie^{1, 5, 6}, Junya Fujimoto⁷, Adi Gazdar^{6, 7, 8}, and Guanghua Xiao^{1, 5, 6, *}

¹Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas, 75390, USA

²Department of Computer Sciences, Massachusetts Institute of Technology

³Department of Pathology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China

⁴Children's Medical Center Research Institute at UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, 75390 USA

⁵Department of Bioinformatics, UT Southwestern Medical Center, Dallas, Texas, USA

⁶Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, Texas, USA

⁷Department of Translational Molecular Pathology, Division of Pathology/Lab Medicine, University of Texas MD Anderson Cancer Center, Houston, TX

⁸Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas, 75390, USA

⁹Hamon Center for Therapeutic Oncology Research, UT Southwestern Medical Center, Texas, 75390, USA

These authors contributed equally to this manuscript.

Correspondence to:

Guanghua Xiao, Quantitative Biomedical Research Center, University of Texas Southwestern Medical Center, Dallas, Texas, 75390, USA. Email: Guanghua.Xiao@utsouthwestern.edu

Tel: +1-214-648-5178

ABSTRACT

Pathology slide images capture tumor histomorphological details in high resolution. However, manual detection and characterization of tumor regions in pathology slides is labor intensive and subjective. Using a deep convolutional neural network (CNN), we developed an automated tumor region recognition system for lung cancer pathology slides. From the identified regions, we extracted 22 well-defined tumor shape features and found that 15 of them were significantly associated with patient survival outcome in lung adenocarcinoma patients from the National Lung Screening Trial. A tumor shape-based prognostic model was developed and validated in an independent patient cohort (n=389). The predicted high-risk group had significantly worse survival than the low-risk group (p value = 0.0029). Predicted risk group serves as an independent prognostic factor (high-risk vs. low-risk, hazard ratio = 2.25, 95% CI 1.34-3.77, p value = 0.0022) after adjusting for age, gender, smoking status, and stage. This study provides new insights into the relationship between tumor shape and patient prognosis.

Abstract word count: 148

Lung cancer is the leading cause of death from cancer, with about half of all cases comprised of lung adenocarcinoma (ADC), which is remarkably heterogeneous in morphological features^{1, 2} and highly variable in prognosis. Through sophisticated visual inspection of tumor pathology slides, ADC can be further classified into different subtypes with drastically different prognoses. Some contributing morphological features have been recognized, such as tumor size or vascular invasion in lung ADC. However, there is a lack of systematic studies on the relationship between tumor shape in pathology slides and patient prognosis.

Tumor tissue slide scanning is becoming part of routine clinical practice for the acquisition of high resolution tumor histological details. In recent years, several computer algorithms for hematoxylin and eosin (H&E) stained pathology image analysis have been developed to aid pathologists in objective clinical diagnosis and prognosis³⁻⁷. Examples include an algorithm to extract stromal features⁸ and an algorithm to assess cellular heterogeneity⁶ as a prognostic factor in breast cancer. More recently, studies have shown that morphological features are associated with patient prognosis in lung cancer as well^{4, 5, 7}. Deep learning methods, such as convolution neural networks (CNNs), have been widely used in image segmentation, object classification and recognition⁹⁻¹¹ and are now being adapted in biomedical image analysis to facilitate cancer diagnosis. To some extent, the performances of deep learning algorithms are similar to, or sometimes even better than, those of humans^{12, 13}. For analysis of H&E-stained pathology images, deep learning methods have been developed to distinguish tumor regions¹⁴, detect metastasis¹⁵, predict mutation status¹⁶, and classify tumors¹⁷ in breast cancer as well as in other cancers. However, due to the complexity of lung cancer tissue structures (such as microscopic alveoli and micro-vessel), deep learning methods for automatic lung cancer region detection from H&E-stained pathology images are not currently available.

Automatic tumor region detection allows for tumor size calculation and tumor shape estimation. Tumor size is a well-established lung cancer prognostic factor¹⁸; the effect of tumor shape has also been investigated in regard to its relationship with drug delivery^{19, 20} and prognosis prediction²¹⁻²⁶. In

X-Ray and computer tomography (CT) image studies, the rough tumor boundary has been reported as a marker for malignant tumor in breast cancer²⁷, and found to be associated with local tumor progression and worse prognosis in lung cancer patients^{22, 26}. Compared with CT images, which are most commonly used to evaluate tumor shape, pathology images have much higher spatial resolution²⁸. Thus, automatic tumor region detection in pathology images allows us to characterize tumor shape accurately and extract tumor shape-based features.

In this study, we developed a deep CNN model to automatically recognize tumor regions for lung ADC from H&E pathology images. More importantly, based on a systematic study of the detected tumor regions of lung ADC patients from the National Lung Screening Trial (NLST) cohort (n=150), we found that many features that characterize the shape of the tumors are significantly associated with tumor prognosis. Finally, we developed a risk-prediction model for lung cancer prognosis using the tumor shape-related features from the NLST lung ADC patient cohort. The prognostic model was then validated in lung ADC patients from The Cancer Genome Atlas (TCGA) dataset. The design of this study is summarized in a flow chart (**Figure 1**).

RESULTS

CNN model distinguishes tumor patches from non-malignant and white (empty region) patches

5344 tumor, non-malignant, and white image patches were extracted from 27 lung ADC H&E pathology slide images (**Supplemental Figure 1**). The imaging patches were split into training, validation and testing datasets (see Methods Section). The CNN model was trained on the training set. The training process stopped at the 28th epoch after validation accuracy failed to improve after 10 epochs. The learning curves for the CNN model in the training and validation sets are shown in **Supplemental Figure 2**. The overall prediction accuracy of the CNN model in the testing set was 89.8%; the accuracy was 88.1% for tumor patches and 93.5% for non-malignant patches (**Supplemental Table 1**).

Tumor region recognition for pathology images

In the NLST dataset, the pathology images have sizes ranging from 5280×4459 pixels to 36966×22344 pixels (median 24244×19261 pixels). To identify tumor regions at slide level, each whole slide image was partitioned into 300×300 image patches. To speed up slide-level prediction, tissue regions were first identified (see Methods Section) and only the image patches within the tissue regions were predicted by the CNN model (**Supplemental Figure 3**). The predicted probabilities of the image patches were summarized into heatmaps of tumor probability (**Figure 2**). An example of a tumor probability heatmap is shown in **Figure 2B**. The tumor region heatmap, predicted as the category with highest probability, is shown in **Figure 2C**. Each pixel in the heatmaps corresponds to a 300×300 pixel image patch in the original 40X pathology image.

Image features from predicted tumor regions correlate with survival outcome

Based on the predicted tumor region heatmap, tissue samples were identified (**Supplemental Figure 4**) and 22 image features were extracted for each tissue sample (see Methods Section). For each patient the image features from multiple tissue samples of the same patient were averaged. The associations between tumor region features and prognostic outcome are summarized in **Table 2** in the NLST dataset. It shows that many features were associated with survival outcome. Most tumor area/size-related features, including area, perimeter, convex area, filled area, major axis length, and minor axis length, both for all tumor regions and for the main tumor region, were associated with poor survival outcome. Interestingly, the number of holes and the perimeter² to area ratio, an estimation of circularity, were also associated with poor survival outcome (for all tumor regions: per 100 number of holes, hazard ratio [HR] = 1.087, p value = 0.033; per 1000 perimeter² to area ratio, HR = 1.15, p value = 0.016; similar results for main tumor region; **Table 2**). Examples comparing tumor shapes with high and low values of eccentricity and perimeter² to area ratio of main tumor region are illustrated in **Figure 3**. As expected, the angle between the X-axis and the major axis of

the main region was not correlated with survival, which serves as a negative control of the feature extraction process.

Development and validation of prognostic model

Utilizing the tumor shape features extracted from the pathology images in the NLST dataset, we developed a prognostic model to predict patient survival outcome. The model was then independently validated in the TCGA cohort. Each patient was assigned into a predicted high- or low-risk group based on the extracted tumor shape features of the patient (see Methods Section). The survival curves for the predicted high- and low-risk groups are shown in **Figure 4**. The patients in the predicted high-risk group had significantly worse survival outcome than those in the predicted low-risk group (log rank test, p value = 0.0029). The multivariate analysis shows that the predicted risk groups independently predicted survival outcome (high- vs. low-risk, HR = 2.25, 95% CI 1.34-3.77, p value = 0.0022, **Table 3**) after adjusting for age, gender, smoking status and stage. This indicates the risk group defined by tumor shape features is an independent prognostic factor, in addition to other clinical variables.

DISCUSSION

In this study, we developed image processing, tumor region recognition, image feature extraction, and risk-score prediction algorithms for pathology images of lung ADC. The algorithms successfully visualized the slide-level tumor region, and serve as a prognostic method independent of other clinical variables. The patient prognostic model was trained in the NLST cohort and independently validated in the TCGA cohort, which indicates the generalizability of the models to other lung ADC patient cohorts.

For tumor region detection, the whole slide image was divided into 300×300 pixel image patches, which were then classified into tumor, non-malignant, or white categories using a CNN model. The

CNN model was trained on 3,848 images patches and tested on 1,068 patches, with 89.8% accuracy in the testing sets. Within the 109 incorrectly predicted patches, 27 contained an insufficient number of cells, which caused confusion between the tissue and background. For the other 82 cases where non-malignant patches were misclassified as tumor or vice versa, the cause seems to be interference from red blood cells, stroma, macrophages, and necrosis (**Supplemental Table 1**). The prediction errors related to out-of-focus tissues (such as macrophages and stroma cells) could be reduced by improved image scanning quality and training set labelling. A similar problem has also been reported in breast cancer recognition¹⁴.

The patch-level tumor prediction results were then arranged to generate tumor region heatmaps (**Figure 2**). In total, 22 well-defined image features were extracted for each tissue region, and averaged to generate patient-level features (**Supplemental Figure 3**). 15 of the 22 features were significantly correlated with survival outcome in NLST dataset (**Table 2**). The features related to tumor area and perimeter were associated with worse prognosis, as tumor size is a well-established prognostic factor, and size-based features were also reported as prognostic from lung cancer CT images^{29, 30}. Interestingly, both for all tumor regions and for the main tumor region, the perimeter² to area ratio was negatively correlated with survival outcome. The perimeter² to tumor area ratio is a quantification of the smoothness of the tumor surface; a large perimeter² to tumor area ratio indicates a large tumor surface and thus a rough tumor boundary. The negative correlation between perimeter² to area ratio and survival outcome is consistent with studies conducted on lung cancer CT images, which reported that a more irregular shape predicted worse survival^{22, 26}. To date, several genes have been reported to be associated with tumor shape^{31, 32}; understanding the relationship among gene expression, tumor shape, and survival outcome can provide insight into tumor development and guide therapeutic decision.

This is the first study to quantify tumor shape-related features using a CNN-based model in lung cancer. In addition, both the main tumor body and the tumor spread through air spaces (STAS,

sometimes referred as aerogenous spread with floating cancer cell clusters [ASFC]), which can be easily detected in the heatmaps^{33, 34}. Since the median size of 40X pathology images is 24244×19261 pixels and the STASs usually only occupy 1 image patch (300×300 pixels) in the NLST dataset, it is labor intensive for human pathologists to circle accurate tumor boundaries and indicate all the tumor STASs. Thus, automatically generating the tumor region heatmap will facilitate pathologists in finding tumor regions and quantifying STASs. More importantly, our study has developed a computation-based method to quantify tumor shape, circularity, irregularity and surface smoothness, which can be an essential tool to study the underlying biological mechanisms. Although tumor size is a well-known prognostic factor, quantifications of the tumor area and perimeter-related features from pathology images are challenging and time-consuming for human pathologists. Thus, it is a natural step to extract image features directly from the predicted tumor heatmaps, thereby avoiding a subjective assessment by a human pathologist.

There are several limitations to our slide-level tumor region detection and image feature extraction pipeline. First, as mentioned before, our CNN model is sensitive to out-of-focus tissue such as red blood cells, macrophages, and stroma cells. Better pathology image scanning quality and more comprehensive labeling of the training set will help solve the problem. Second, the image features can be affected by slide preparation artifacts, such as artificially damaged tumor tissues and failure to select the slides that faithfully represent the tumor. Thus, to ensure the representability of the predicted risk score, a representative tumor slide is required. Third, the image slides are 2-dimensional, which loses the 3-dimensional spatial information. Combining the tumor prediction and feature extraction algorithms with other imaging techniques, such as CT or X-Ray, may produce more comprehensive descriptions of the tumor region and improve the performance of the current risk prediction model.

CONCLUSION

Our pipeline for tumor region recognition and risk-score prediction based on tumor shape features serves as an objective prognostic method independent of other clinical variables, including age, gender, smoking status and stage. The tumor region heatmaps generated by our model can help pathologists locate tumor regions in gigapixel pathology slides swiftly and accurately. The model development pipeline can also be used in other cancer types, such as breast and kidney cancer.

METHODS

Datasets

The pathology images together with the corresponding clinical data were obtained from two independent datasets: 267 40X images for 150 lung ADC patients were acquired from the NLST dataset; 457 40X images for 389 lung ADC patients were acquired from the TCGA dataset. Clinical characteristics of patients in this study are presented in **Table 1**. The prognostic model was trained on the NLST dataset and independently validated on the TCGA dataset.

Image patch generation

A CNN model was trained to classify non-malignant tissues, tumor tissues, and white regions based on image patches of H&E stained pathology images. The patch size was determined as 300×300 pixels under 40X magnification, to ensure at least 20 cells within one patch (**Supplemental Figure 1**). Tumor and non-malignant patches were randomly extracted from tumor regions and non-malignant regions labeled by a pathologist, respectively. The patches were classified as white if the mean intensity of all pixel values was larger than a threshold determined from sample images. Examples of each patch class are shown in **Supplemental Figure 1**. 2139 non-malignant, 2475 tumor and 730 white patches were generated in total.

CNN training process

The Inception (V3) architecture³⁵ with input size 300×300 and weights pre-trained on ImageNet was used to train our CNN model. The network was trained with stochastic gradient descent algorithms in Keras with TensorFlow backend³⁶. The batch size was set to 32, the learning rate was set to 0.0001 without decay, and the momentum was set to 0.9. From the extracted 5,344 image patches, 3,848 patches (72%) were allocated to the training set, 428 patches (8%) to the validation set, and the remaining 1,068 patches (20%) to the testing set, with equal proportions among the three classes. Keras Image Generators were used to non-malignantize and flip the images, both horizontally and vertically, to augment the training and validation datasets. The maximum number of epochs to train was set to 50. To avoid overfitting, the training process automatically stopped after the validation accuracy failed to improve for 10 epochs.

Prediction heatmap generation

To avoid prediction on a large empty image area and speed up the slide-level prediction process, the Otsu thresholding method followed by morphological operations such as dilation and erosion was first applied to whole pathology slides to generate the tissue region mask (**Supplemental Figure 2**)³⁷.³⁸ A 300×300 pixel window was then slid over the entire mask without overlapping between any two windows. The image patches were predicted with batch size 32, and one image patch was predicted only once without rotation or flipping. For each image patch, probabilities of being in each of the three classes were predicted, and a heatmap of the predicted probability was generated for each pathology slide (**Figure 2**). For each image patch, the class with the highest probability was determined as the predicted class.

Image feature extraction

In a pathology slide, sometimes there are multiple tissue samples. To distinguish different tissue samples in the same slide, disconnected tissue regions were first identified by morphological operations on heatmaps of predicted classes (**Supplemental Figure 3A-C**)³⁸. To remove the effects

of some very small tissue samples, the tissue regions with area smaller than half of the largest tissue region in the same slide were removed from analysis. Within each tissue region, the tumor region with the largest area was regarded as the “main tumor region” (**Supplemental Figure 3D**). The following features of tumor regions were estimated for each tissue sample: number of regions, area, convex area, filled area, perimeter, major axis length, minor axis length, number of holes, and perimeter² to area ratio for all tumor regions and the main tumor region separately; eccentricity, extent, solidity, and angle between the X-axis and the major axis for the main tumor region (22 features in total)³⁹. Here, 8-connectivity was used to determine disconnected tumor regions and disconnected holes³⁹. When multiple tissue regions were available for one patient, either due to multiple tissues within one slide or multiple slides for one patient, the 22 image features were averaged to generate patient-level image features.

Prognostic model development

A univariate Cox proportional hazard model was used to study the association between the 22 tumor shape features and patient survival outcome in the NLST dataset. The image features that were significantly associated with survival outcome were selected to build the prediction model for patient prognosis. To avoid overfitting, a Cox proportional hazard model with an elastic-net penalty⁴⁰ was used; the penalty coefficient λ was determined through 10-fold cross-validation in the NLST cohort.

Model validation in an independent cohort

The model developed from the NLST cohort was then independently validated in the TCGA cohort (n=389) for prognostic performance. First, the tumor region(s) in each pathology image slide from the TCGA dataset were detected using the CNN model, and the tumor shape features were extracted based on the detected tumor region(s) in each image slide and summarized to patient level. Finally a risk score was calculated based on the extracted tumor features. The patients were dichotomized into high- and low-risk groups based on their predicted risk score, using the median as the cutoff. A log-

rank test was used to compare survival difference between predicted high- and low-risk groups. The survival curves were estimated using the Kaplan-Meier method. A multivariate Cox proportional hazard model was used to test the prognostic value of the predicted risk score after adjusting for other clinical factors, including age, gender, tobacco history and stage. Overall survival, defined as the period from diagnosis until death or last contact, was used as response. Survival analysis was performed with R software, version 3.3.2⁴¹. R packages “survival” (version 2.40-1) and “glmnet” (version 2.0-5) are used^{40, 42}. The results were considered significant if the two-tailed p value ≤ 0.05 .

Data availability

Pathology images that support the findings of this study are available online in NLST (<https://biometry.nci.nih.gov/cdas/nlst/>) and The Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD, <https://wiki.cancerimagingarchive.net/display/Public/TCGA-LUAD>).

Code availability

The codes are available upon request. We will share the codes through Github following manuscript acceptance.

Author Contributions:

SW and GX designed the study. SW, AC, LC, YX, JF, AG, and GX analyzed the results and wrote the manuscript. SW and AC implemented the algorithm. LY labeled the data. All authors commented on the manuscript.

Competing financial interests:

The authors declare that they have no competing interests.

Materials & Correspondence:

Correspondence and material requests should be addressed to GX.

Source of Funding:

This work was partially supported by the National Institutes of Health [5R01CA152301, P50CA70907, 5P30CA142543, 1R01GM115473, and 1R01CA172211], and the Cancer Prevention and Research Institute of Texas [RP120732].

References

1. Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, Cho H. Seer cancer statistics review, 1975–2008. . Bethesda, MD: National Cancer Institute, 19. 2011
2. Matsuda T, Machii R. Morphological distribution of lung cancer from cancer incidence in five continents vol. X. *Jpn J Clin Oncol*. 2015;45:404
3. Tabesh A, Teverovskiy M, Pang H-Y, Kumar VP, Verbel D, Kotsianti A, Saidi O. Multifeature prostate cancer diagnosis and gleason grading of histological images. *Medical Imaging, IEEE Transactions on*. 2007;26:1366-1378
4. Wang H, Xing F, Su H, Stromberg A, Yang L. Novel image markers for non-small cell lung cancer classification and survival prediction. *BMC Bioinformatics*. 2014;15:310
5. Luo X, Zang X, Yang L, Huang J, Liang F, Rodriguez-Canales J, Wistuba, II, Gazdar A, Xie Y, Xiao G. Comprehensive computational pathological image analysis predicts lung cancer prognosis. *J Thorac Oncol*. 2017;12:501-509
6. Yuan Y, Failmezger H, Rueda OM, Ali HR, Graf S, Chin SF, Schwarz RF, Curtis C, Dunning MJ, Bardwell H, Johnson N, Doyle S, Turashvili G, Provenzano E, Aparicio S, Caldas C, Markowitz F. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med*. 2012;4:157ra143
7. Yu KH, Zhang C, Berry GJ, Altman RB, Re C, Rubin DL, Snyder M. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun*. 2016;7:12474
8. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, West RB, van de Rijn M, Koller D. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med*. 2011;3:108ra113
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444
10. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun Acn*. 2017;60:84-90
11. Schmidhuber J. Deep learning in neural networks: An overview. *Neural networks : the official journal of the International Neural Network Society*. 2015;61:85-117
12. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118
13. Li H, Giger ML, Huynh BQ, Antropova NO. Deep learning in breast cancer risk assessment: Evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms. *Journal of medical imaging*. 2017;4:041304
14. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, Venugopalan S, Timofeev A, Nelson PQ, Corrado GS. Detecting cancer metastases on gigapixel pathology images. *arXiv preprint arXiv:1703.02442*. 2017
15. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. <https://arxiv.org/abs/1606.05718>. 2016
16. Vandenberghe ME, Scott ML, Scorer PW, Soderberg M, Balcerzak D, Barker C. Relevance of deep learning to facilitate the diagnosis of her2 status in breast cancer. *Scientific reports*. 2017;7:45938
17. Cruz-Roa A, Gilmore H, Basavanahally A, Feldman M, Ganesan S, Shih NNC, Tomaszewski J, Gonzalez FA, Madabhushi A. Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific reports*. 2017;7:46450
18. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, Nicholson AG, Groome P, Mitchell A, Bolejack V, International Association for the Study of Lung Cancer S, Prognostic Factors Committee AB, Participating I, International Association for the Study of Lung Cancer S, Prognostic Factors Committee Advisory B, Participating I. The iaslc lung cancer staging project: Proposals for revision of the tnm stage groupings in the forthcoming (eighth) edition of the tnm classification for lung cancer. *J Thorac Oncol*. 2016;11:39-51

19. Soltani M, Chen P. Effect of tumor shape and size on drug delivery to solid tumors. *J Biol Eng*. 2012;6:4
20. Sefidgar M, Soltani M, Raahemifar K, Bazmara H, Nayinian SM, Bazargan M. Effect of tumor shape, size, and tissue transport properties on drug delivery to solid tumors. *J Biol Eng*. 2014;8:12
21. Chatzistamou I, Rodriguez J, Jouffroy T, Girod A, Point D, Sklavounou A, Kittas C, Sastre-Garau X, Klijanienko J. Prognostic significance of tumor shape and stromal chronic inflammatory infiltration in squamous cell carcinomas of the oral tongue. *J Oral Pathol Med*. 2010;39:667-671
22. Vogl TJ, Worst TS, Naguib NN, Ackermann H, Gruber-Rouh T, Nour-Eldin NE. Factors influencing local tumor control in patients with neoplastic pulmonary nodules treated with microwave ablation: A risk-factor analysis. *AJR Am J Roentgenol*. 2013;200:665-672
23. Hashiba T, Hashimoto N, Maruno M, Izumoto S, Suzuki T, Kagawa N, Yoshimine T. Scoring radiologic characteristics to predict proliferative potential in meningiomas. *Brain Tumor Pathol*. 2006;23:49-54
24. Miller TR, Pinkus E, Dehdashti F, Grigsby PW. Improved prognostic value of 18f-fdg pet using a simple visual analysis of tumor characteristics in patients with cervical cancer. *J Nucl Med*. 2003;44:192-197
25. Yokoyama I, Sheahan DG, Carr B, Kakizoe S, Selby R, Tzakis AG, Todo S, Iwatsuki S, Starzl TE. Clinicopathologic factors affecting patient survival and tumor recurrence after orthotopic liver transplantation for hepatocellular carcinoma. *Transplant Proc*. 1991;23:2194-2196
26. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Velazquez ER, Lambin P, Gu Y, Balagurunathan Y, Eikman E, Gatenby RA, Eschrich S, Gillies RJ. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One*. 2015;10:e0118261
27. Rangayyan RM, El-Faramawy NM, Desautels JE, Alim OA. Measures of acutance and shape for classification of breast tumors. *IEEE Trans Med Imaging*. 1997;16:799-810
28. Farahani N, Parwani AV, Pantanowitz L. Whole slide imaging in pathology: Advantages, limitations, and emerging perspectives. *Pathol Lab Med Int*. 2015;7:23-33
29. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, Postmus PE, Rusch V, Sobin L, International Association for the Study of Lung Cancer International Staging C, Participating I. The iaslc lung cancer staging project: Proposals for the revision of the tnm stage groupings in the forthcoming (seventh) edition of the tnm classification of malignant tumours. *J Thorac Oncol*. 2007;2:706-714
30. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, Goldgof DB, Hall LO, Korn R, Zhao B, Schwartz LH, Basu S, Eschrich S, Gatenby RA, Gillies RJ. Test-retest reproducibility analysis of lung ct image features. *J Digit Imaging*. 2014;27:805-823
31. Tripodis N, Demant P. Genetic analysis of three-dimensional shape of mouse lung tumors reveals eight lung tumor shape-determining (ltsd) loci that are associated with tumor heterogeneity and symmetry. *Cancer Res*. 2003;63:125-131
32. Kida H, Takano Y, Yamamoto K, Mori M, Yanaga K, Tanaka J, Kudo SE, Mimori K. A single nucleotide polymorphism in fibronectin 1 determines tumor shape in colorectal cancer. *Oncol Rep*. 2014;32:548-552
33. Kadota K, Nitadori J, Sima CS, Ujiie H, Rizk NP, Jones DR, Adusumilli PS, Travis WD. Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage i lung adenocarcinomas. *J Thorac Oncol*. 2015;10:806-814
34. Shiono S, Ishii G, Nagai K, Yoshida J, Nishimura M, Murata Y, Tsuta K, Nishiwaki Y, Kodama T, Ochiai A. Histopathologic prognostic factors in resected colorectal lung metastases. *Ann Thorac Surg*. 2005;79:278-282; discussion 283
35. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *Proc Cvpr Ieee*. 2016:2818-2826
36. Chollet Fao. Keras. 2015

37. Otsu N. Threshold selection method from gray-level histograms. *Ieee T Syst Man Cyb.* 1979;9:62-66
38. Vincent L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans Image Process.* 1993;2:176-201
39. Gonzalez RC, Woods R. *Digital image processing*. Pearson Education; 2002.
40. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model via coordinate descent. *J Stat Softw.* 2011;39:1-13
41. Team RC. R: A language and environment for statistical computing. 2016
42. T T. A package for survival analysis in s. 2015

Table 1. Patient characteristics of training and validation datasets.

	<i>NLST (training)*</i>	<i>TCGA (validation)</i>	<i>p-value</i>
<i>No. of patients</i>	150	389	
<i>Age</i>	64.03 ± 5.12	64.98 ± 10.33	0.16
<i>Gender</i>			0.055
<i>Male</i>	82 (54.7)	175 (45.0)	
<i>Female</i>	68 (45.3)	214 (55.0)	
<i>Smoking status</i>			0.0020
<i>Yes</i>	81 (54.0)	267 (68.6)	
<i>No</i>	69 (46.0)	122 (31.4)	
<i>Stage</i>			0.0048
<i>I</i>	101 (67.3)	222 (57.1)	
<i>II</i>	16 (10.7)	96 (24.7)	
<i>III</i>	23 (15.3)	49 (12.6)	
<i>IV</i>	10 (6.7)	22 (5.7)	

NLST, the National Lung Screening Trial; TCGA, the Cancer Genome Atlas.

* Values are either mean ± standard deviation, or number (percentage)

Table 2. Univariate analysis of tumor region features in NLST training dataset.

	<i>HR (95% CI)</i>	<i>p-value</i>
<i>Number of regions (per 1000)</i>	1.29 (0.64-2.58)	0.48
<i>Area sum of all regions (per 1000 pixels*)</i>	1.030 (1.010-1.050)	0.0033
<i>Perimeter sum of all regions (per 1000 pixels)</i>	1.088 (1.028-1.151)	0.0034
<i>Sum of convex area for all regions (per 1000 pixels)</i>	1.020 (1.006-1.033)	0.0047
<i>Sum of filled area for all regions (per 1000 pixels)</i>	1.027 (1.009-1.045)	0.0029
<i>Sum of hole numbers of all regions (per 100)</i>	1.087 (1.031-1.16)	0.0033
<i>Sum of major axis length of all regions (per 1000 pixels)</i>	1.40 (1.00-1.96)	0.051
<i>Sum of minor axis length of all regions (per 1000 pixels)</i>	2.65 (1.10-6.40)	0.030
<i>Perimeter²/area of all regions (per 1000)</i>	1.18 (1.03-1.35)	0.019
<i>Area of main region (per 1000 pixels)</i>	1.027 (1.007-1.048)	0.0093
<i>Convex area of main region (per 1000 pixels)</i>	1.018 (1.004-1.032)	0.010
<i>Eccentricity of main region</i>	6.37 (0.57-71.56)	0.13
<i>Hole number of main region (per 100)</i>	1.087 (1.020-1.15)	0.0060
<i>Extent of main region</i>	4.90 (0.19-126.30)	0.34
<i>Filled area for main region (per 1000 pixels)</i>	1.025 (1.007-1.043)	0.0072
<i>Major axis length for main region (per 100 pixels)</i>	1.57 (1.11-2.21)	0.0099
<i>Minor axis length for main region (per 100 pixels)</i>	1.73 (1.05-2.83)	0.031
<i>Angle between the X-axis and the major axis of main region</i>	0.98 (0.64-1.50)	0.92
<i>Perimeter of main region (per 1000 pixels)</i>	1.087 (1.023-1.15)	0.0068
<i>Solidity of main region</i>	7.24 (0.45-117.40)	0.16
<i>Average tumor probability of the main region (per 0.10)</i>	1.11 (0.53-2.24)	0.78
<i>Perimeter²/area for main region (per 1000)</i>	1.21 (1.03-1.42)	0.021

NLST, the National Lung Screening Trial.

* 1 pixel in heatmap = 1 patch in 40X pathological slide. Patch size: 300 pixels * 300 pixels.

Table 3. Multivariate analysis of predicted risk and clinical variables in TCGA.

<i>Variable</i>	<i>HR (95% CI)</i>	<i>p-value</i>
<i>High risk vs. low risk</i>	2.25 (1.34-3.77)	0.0022
<i>Age</i>	1.02 (1.00-1.04)	0.12
<i>Male vs. female</i>	0.71 (0.43-1.16)	0.17
<i>Smoker vs. non-smoker</i>	0.95 (0.59-1.54)	0.85
<i>Stage II vs. stage I</i>	2.58 (1.47-4.51)	<0.001
<i>Stage III vs. stage I</i>	5.23 (2.85-9.59)	<0.001
<i>Stage IV vs. stage I</i>	2.69 (1.19-6.09)	0.017

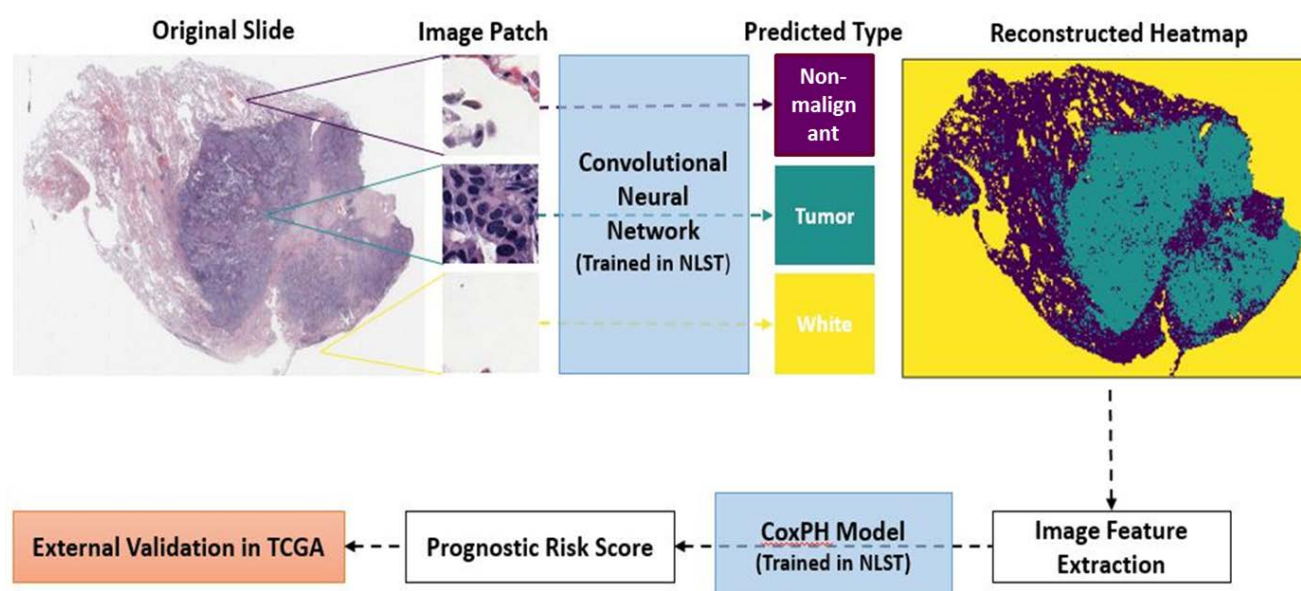


Figure 1. Flow chart of analysis process. CNN, convolutional neural network; NLST, the National Lung Screening Trial; TCGA, The Cancer Genome Atlas.

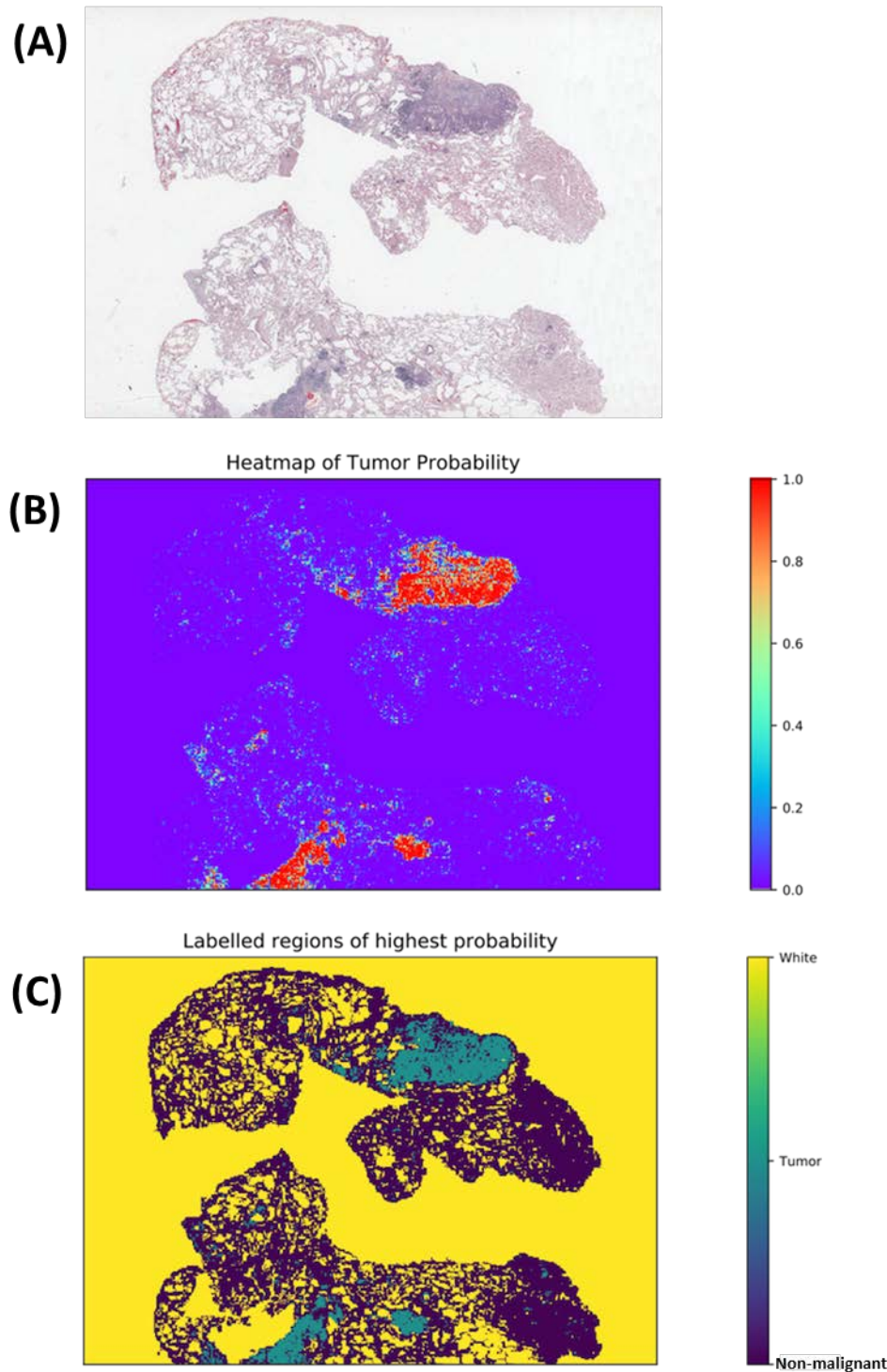


Figure 2. Example results of slide-level tumor region detection. (A) Original slide. (B) Predicted tumor probability. Each point in the heatmap corresponds to a 300×300 pixel image patch in original 40x slide. (C) Predicted region labels. Yellow: white (empty) region; green: tumor region; blue: non-malignant region.

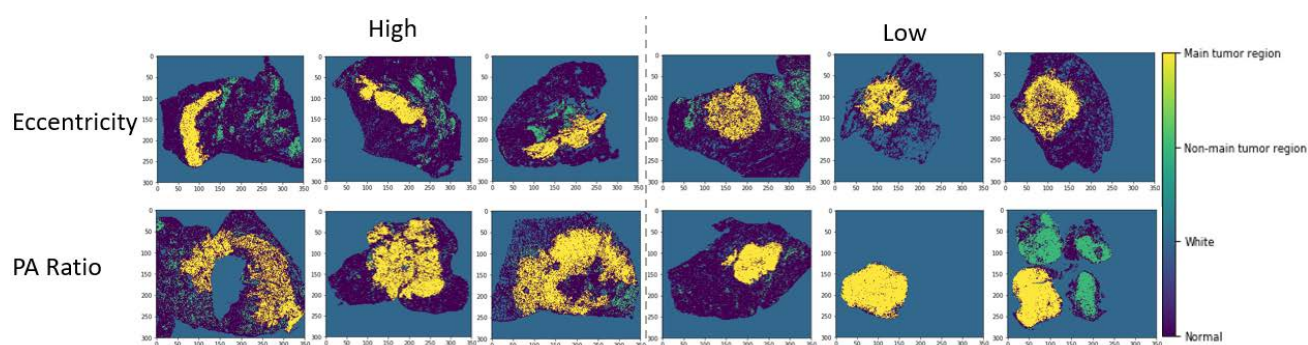


Figure 3. Comparison of tumor shapes with high or low values of eccentricity and PA ratio of main tumor region. Original heatmaps are cropped to the same size with the same image scale. Yellow, main tumor region; green, non-main tumor region; dark blue, non-malignant tissue; blue, blank part of pathology slide. P-A ratio, perimeter^2 to area ratio.

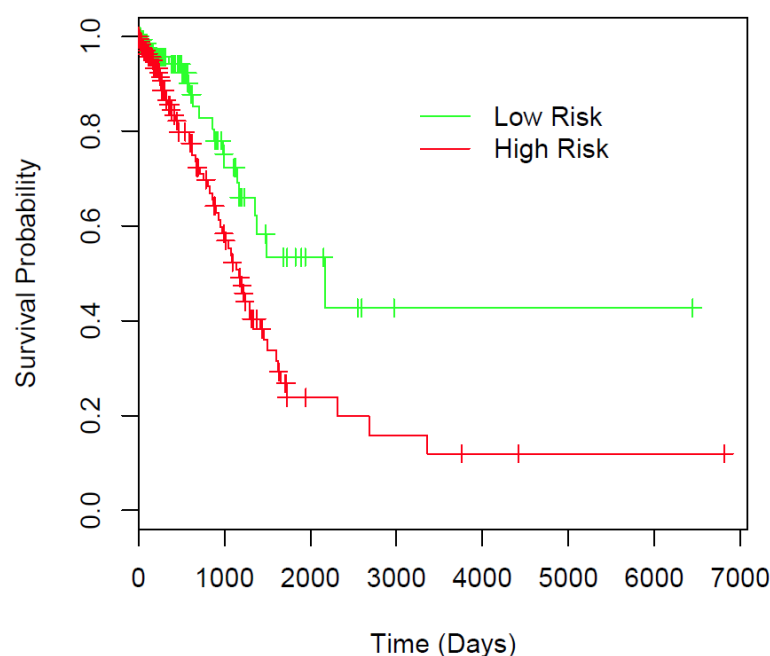
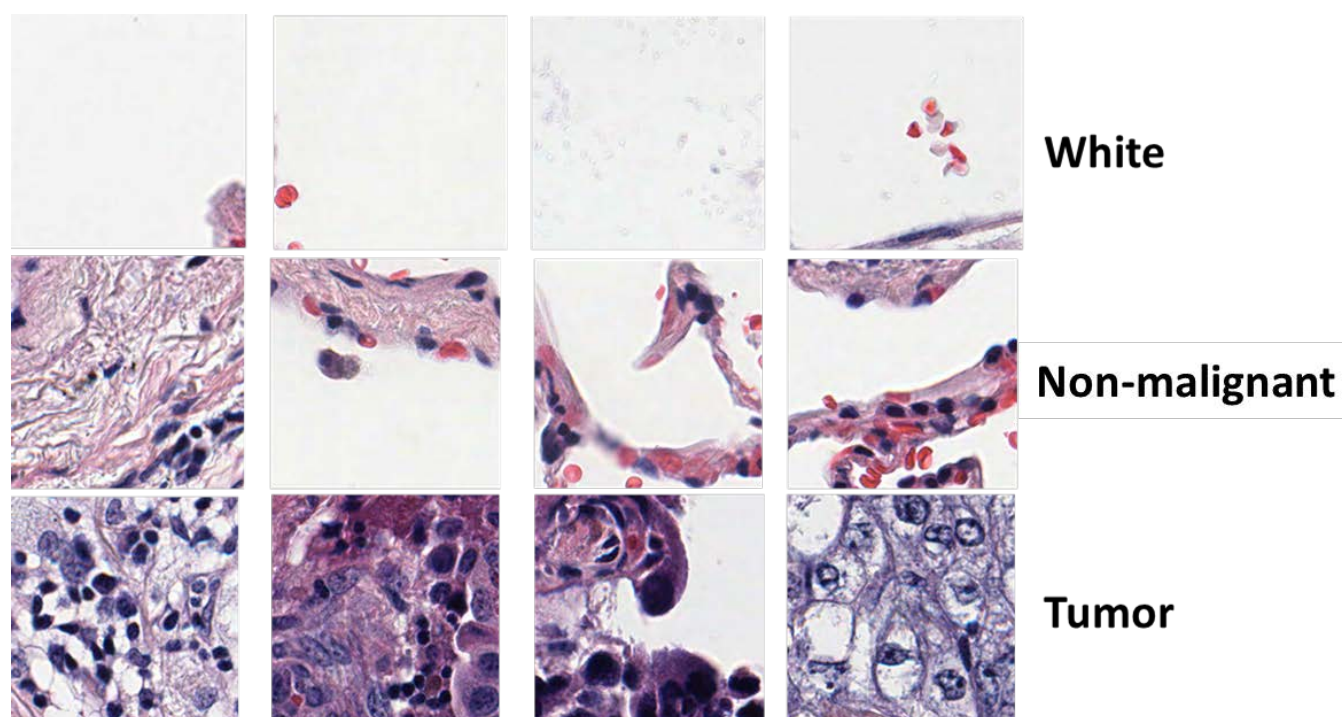
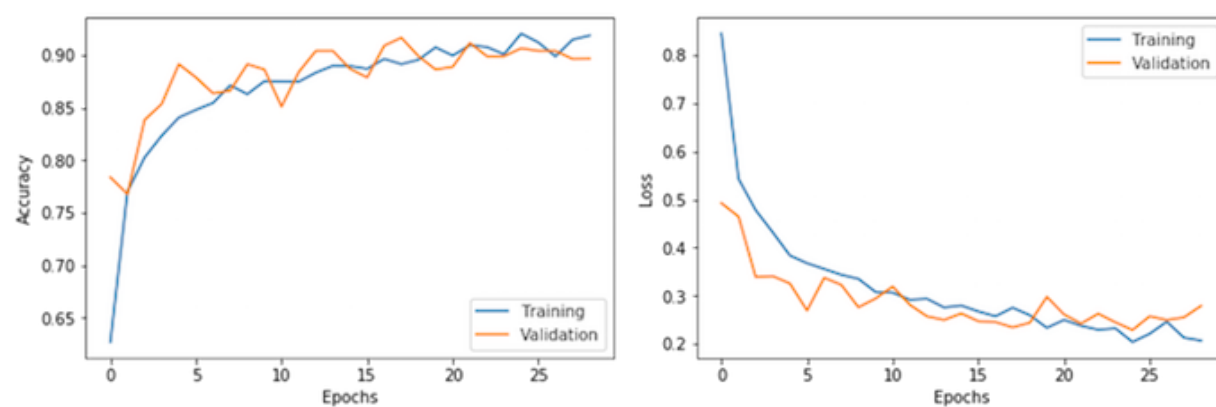


Figure 4. Prognostic performance in TCGA validation dataset illustrated by Kaplan-Meier plot. Patients are dichotomized according to median predicted risk score. Difference between the two risk groups: log-rank test, p value = 0.0029.

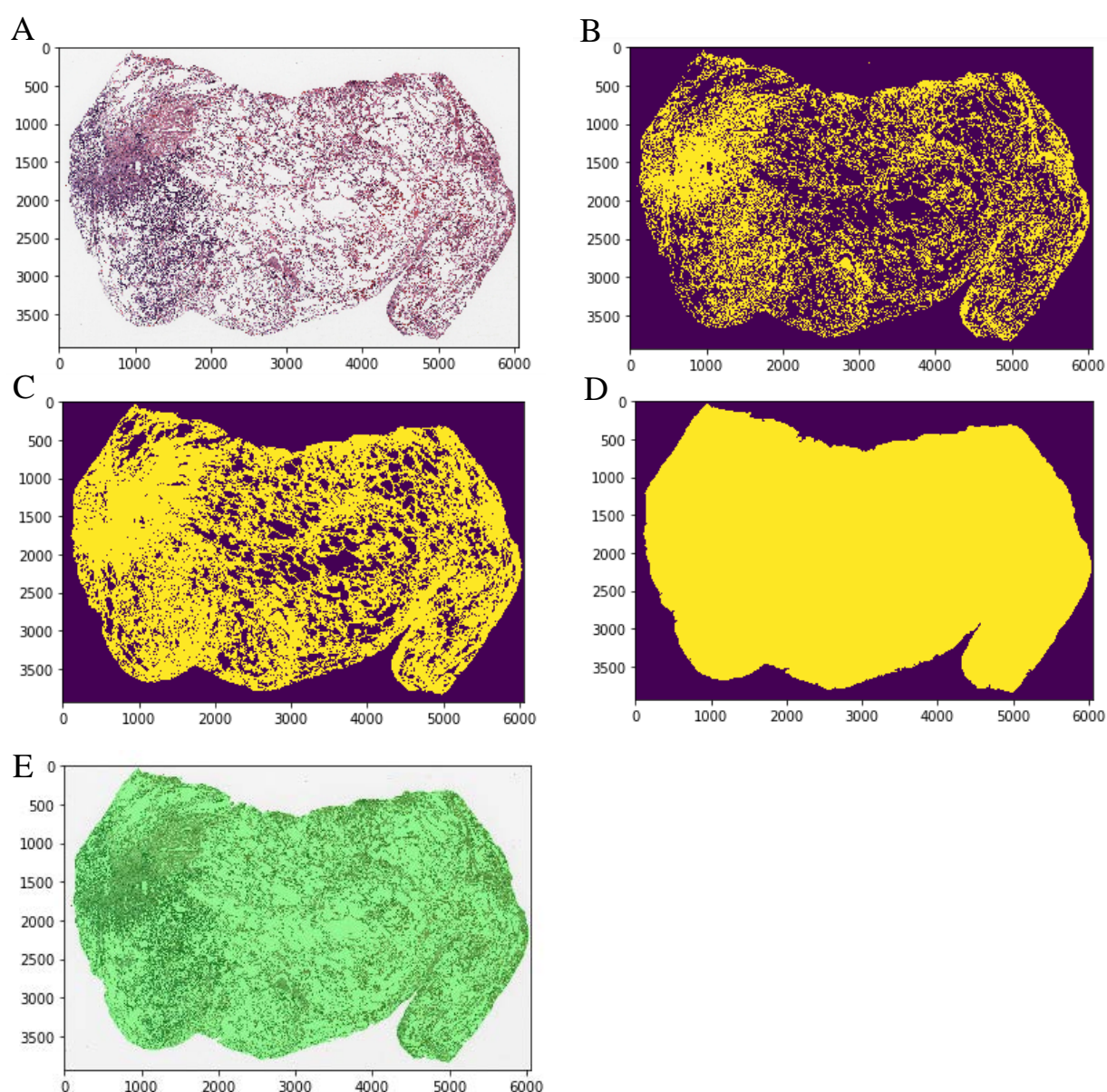
Supplementary material



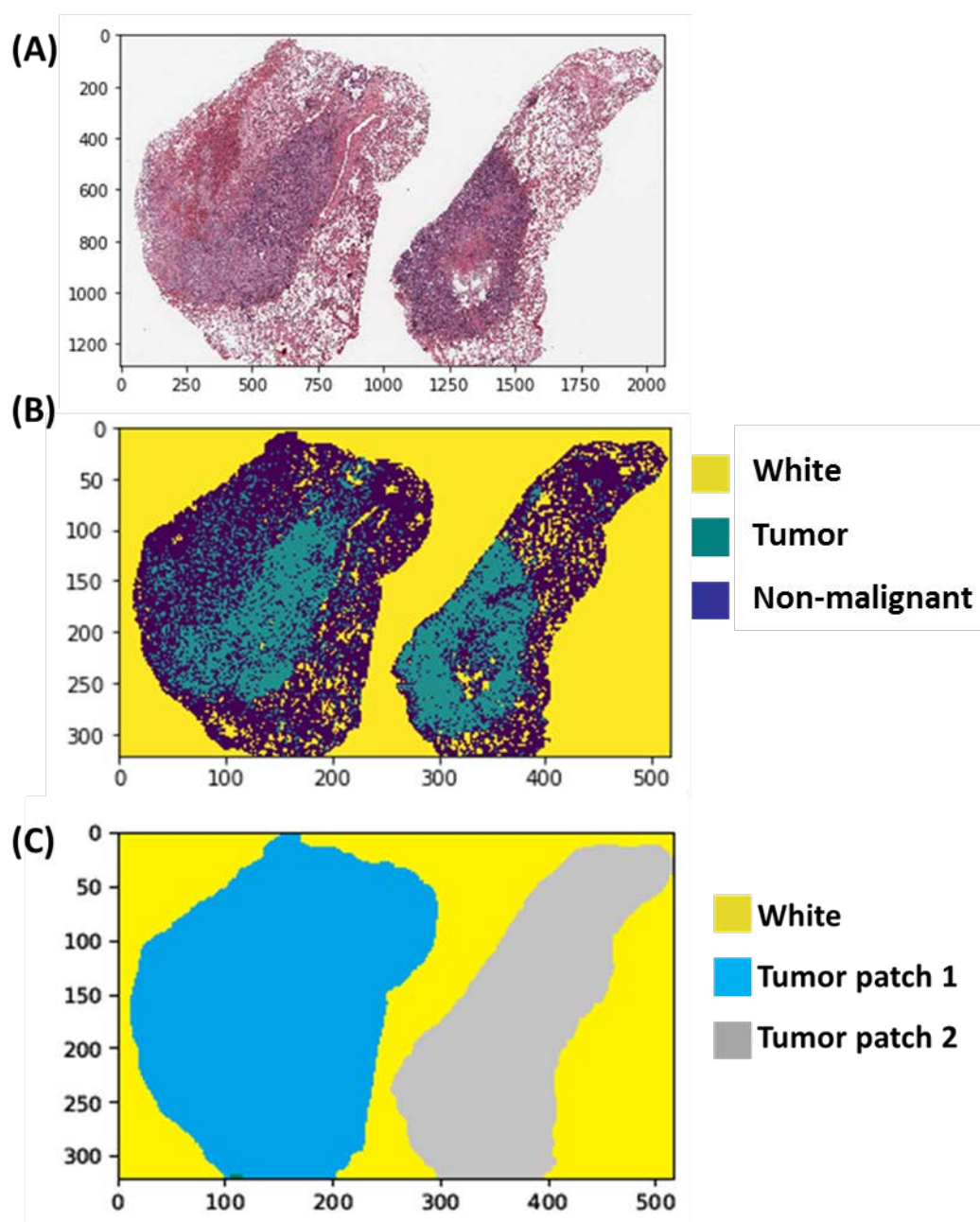
Supplemental Figure 1. Example of image patches from “white” (empty regions, upper panel), “non-malignant” (middle panel) and “tumor” (bottom panel) categories. Patch size: 300×300 pixels.



Supplemental Figure 2. Convolutional Neural Network learning curves in both training and validation datasets. Left, accuracy versus epochs; right, loss versus epochs.



Supplemental Figure 3. Otsu thresholding and image morphological operations to speed up slide level prediction process. (A) The original slide. (B) The image mask after Otsu thresholding. (C) The image mask after dilation and removing small objects of the mask in (B). (D) The final mask after dilation, erosion, and filling up holes of mask in (C). (E) Overlap final image mask and original pathology slide.



Supplemental Figure 4. Tissue region identification in case of multiple tissue samples within one slide. **(A)** Original slide. **(B)** Predicted patch-level tumor, non-malignant and white heatmap. **(C)** Disconnected tissue samples identified by image processing. Yellow, background; blue, first tissue patch; gray, second tissue patch.

Supplemental Table 1. Confusion matrix for image patch classification.

<i>Ground Truth \ Predicted Value</i>	<i>Non-malignant</i>	<i>Tumor</i>	<i>White</i>
<i>Non-malignant</i>	400 (93.5%)	24 (5.6%)	4 (0.9%)
<i>Tumor</i>	58 (11.7%)	436 (88.1%)	1 (0.2%)
<i>White</i>	21 (14.4%)	1 (0.7%)	124 (84.9%)

Supplemental Table 2. Comparison of patient characteristics between high-risk and low-risk groups in TCGA validation dataset.

	<i>Low-risk</i>	<i>High-risk</i>	<i>p-value</i>
<i>No. of patients</i>	195	194	
<i>Age</i>	65.50 ± 10.11	64.45 ± 10.54	0.31
<i>Gender</i>			0.20
<i>Male</i>	81 (41.5)	94 (48.5)	
<i>Female</i>	114 (58.5)	100 (51.5)	
<i>Smoking status</i>			0.89
<i>Yes</i>	135 (69.2)	132 (68.0)	
<i>No</i>	60 (30.8)	62 (32.0)	
<i>Stage</i>			0.005
<i>I</i>	127 (65.1)	95 (49.0)	
<i>II</i>	44 (22.6)	52 (26.8)	
<i>III</i>	17 (8.7)	32 (16.5)	
<i>IV</i>	7 (3.6)	15 (7.7)	