COMPREHENSIVE ANALYSIS OF LUNG CANCER PROGNOSTIC FACTORS

APPROVED BY SUPERVISORY COMMITTEE

Yang Xie, Ph.D. (Mentor)

Guanghua Xiao, Ph.D. (Mentor)

David Gerber, M.D. (Chair)

Xiaowei Zhan, Ph.D.

Yujin Hoshida, M.D., Ph.D.

DEDICATION

I would like to thank my mentors Drs. Yang Xie and Guanghua Xiao. Their guidance and help provided me the great opportunity to learn and apply many wonderful data science theories to my PhD thesis. I also would like to thank my thesis committee members, Drs. David Gerber, Xiaowei Zhan, Yujin Hoshida, and Adi Gazdar, whose advices on both research and presentation gave me insight into my thesis topic. Next, I would like to thank all my collaborators from different fields for their knowledge and willingness to accomplish the projects together. Further, I would like to thank all my friends for their encouragement and help on my daily life. Finally yet importantly, I would like to thank my family and my boyfriend for their company.

COMPREHENSIVE ANALYSIS OF LUNG CANCER PROGNOSTIC FACTORS

by

SHIDAN WANG

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2019

COMPREHENSIVE ANALYSIS OF LUNG CANCER PROGNOSTIC FACTORS

Shidan Wang, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2019

Supervising Professor: Yang Xie, Ph.D. & Guanghua Xiao, Ph.D.

Lung cancer is the leading cause of death from cancer. It is remarkably heterogeneous in histopathological features and highly variable in prognosis. Analysis of prognostic factor is anticipated to guide clinicians for treatment selection, enhance patient care, and help understanding biological mechanism of tumor progression. To extend current knowledge about lung cancer prognosis, this dissertation analyzed lung cancer prognostic factors in three levels. First, in tumor level, deep learning aided pathology image analysis was used to extract tumor geometry and microenvironment features, upon which an image-based survival prediction model was built and independently validated for lung adenocarcinoma. Second, in patient level, a nomogram was built with demographic and clinical variables for patients with small cell lung cancer. The nomogram was implemented online for public usage. Third, in population level, how facility type and volume affect survival outcome and surgery selection for early stage non-small cell lung cancer was analyzed.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

First-author publications

1. **S. Wang**, H. Liu, X. Zhan, et al. Deep learning in lung cancer pathological image analysis. *Manuscript in preparation*.
2. **S. Wang**, R. Rong, L. Cai, et al. Computational staining of pathology images to study tumor microenvironment in lung cancer. Under review for ***Cancer Research***.
3. **S. Wang ***, T. Wang *, L. Yang *, et al. ConvPath: a software tool for lung adenocarcinoma digital pathological image analysis aided by convolutional neural network. Under review for ***EbioMedicine***.
4. **S. Wang**, S. Lai, M. Itzstein, et al. Type and Patient Volume of Healthcare Facilities Influence Survival and Surgery Selection of Early Stage Non-Small Cell Lung Cancer Patients. ***Cancer***, 2019.
5. **S. Wang**, R. Rong, X. Zhan, Y. Xie, G. Xiao. Pathology image analysis using segmentation deep learning algorithms. ***American Journal of Pathology***, *2019.*
6. **S. Wang ***, A. Chen *, L. Yang, et al. Comprehensive analysis of lung cancer pathology images to discover tumor shape features that predict survival outcome, ***Nature Scientific Reports***, 2018.
7. **S. Wang**, L. Yang, G. Xiao, et al. Development and validation of a nomogram prognostic model for small-cell lung cancer patients. ***Journal of Thoracic Oncology***, 2018.

8. **S. Wang**, R. Khera, S. Das, et al. Usefulness of a simple algorithm identifies hypertensive patients who benefit from intensive blood pressure lowering. ***American Journal of Cardiology***, 2018.

Co-author publications

9. F. Yi, L. Yang, **S. Wang**, et al. Microvessel prediction in H&E stained pathology images using fully convolutional neural networks. ***BMC bioinformatics,*** 2018.
10. L. Yang, **S. Wang**, D. Gerber, et al. Main bronchus location is a predictor for metastasis and prognosis in lung adenocarcinoma: a large cohort analysis. ***Lung Cancer***, 2018.
11. H. Tang, **S. Wang**, G. Xiao, et al. Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies. ***Annals of Oncology***, 2017.
12. L. Yang, **S. Wang**, Y. Zhou, et al. Evaluation of the 7th and 8th editions of the AJCC/UICC TNM staging systems for lung cancer in a large North American cohort. ***Oncotarget***, 2017.
13. Y. Xie, W. Lu, **S. Wang**, et al. Validation of the 12-Gene Predictive Signature for Adjuvant Chemotherapy Response in Lung Cancer. ***Clinical Cancer Research***, 2017.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADC – adenocarcinoma

AJCC – American Joint Committee on Cancer

ANN – artificial neural network

ARP – Academic/Research Program

AUC – Area under the Curve

BH – Benjamini-Hochberg

BIC – Bayesian Information Criterion

CCCP – Comprehensive Community Cancer Program

CCP – Community Cancer Program

CI – confidence interval

CNN – Convolutional Neural Network

GSEA – gene set enrichment analysis

H&E – Hematoxylin and Eosin

HR – hazard ratio

INCP – Integrated Network Cancer Program

IoU – Intersection over Union

K-M – Kaplan-Meier

LUAD – lung ADC

Mask-RCNN – Mask Regional Convolutional Neural Network

mRNA – messenger Ribonucleic acid

NCCN – National Comprehensive Cancer Network

NCDB – National Cancer Database

NSCLC – Non-Small Cell Lung Cancer

NLST – National Lung Screening Trial

OR – odds ratio

PA ratio – perimeter^2 to area ratio

PD1 – Programmed cell death protein 1

RGB – red, green, and blue

ROC – Receiver Operating Characteristics

ROI – Region of Interest

SCC – squamous carcinoma

SCLC – Small Cell Lung Cancer

STAS – tumor spread through air spaces

TF – teaching facility

TIL – tumor-infiltrating lymphocytes

TCGA – the Cancer Genome Atlas

TCR – T-cell receptor

TME – tumor microenvironment

TNM – Tumor, Node, and Metastases

UICC – Union for International Cancer Control

WSI – whole slide imaging

CHAPTER ONE – INTRODUCTION

Lung cancer is the leading cause of death from cancer[1]. It is remarkably heterogeneous in histopathological features and highly variable in prognosis. Lung cancer is mainly composed with two histologic types: Non-Small Cell Lung Cancer (NSCLC, ~ 85%) and Small Cell Lung Cancer (SCLC, ~ 14%)[1-3], which are different in treatment and prognosis. NSCLC is a group of several subtypes, including adenocarcinoma (ADC, ~ 40% of all lung cancers), squamous carcinoma (SCC, ~ 30%), large cell carcinoma (~ 10%), and other much less common subtypes. Those NSCLC subtypes originate from different types of lung cells but share similar treatment approach.

One of the most critical question is how to treat lung cancer patients properly and effectively. According to the National Comprehensive Cancer Network (NCCN) clinical practice guideline[4], treatment of lung cancer should be planned by doctors and clinicians using multiple source of information, including medical history, imaging, and pathology report. The treatment decision should be made upon comprehensive consideration of response, toxicity, and the life expectancy[5]. Thus, building prognostic model for lung cancer patients is anticipated to be an important tool for individualized treatment selection. In addition to guiding treatment selection, building survival-predicting model can provide guideline to patients, such as facility selection. Furthermore, identifying new prognostic factors can help researchers to understand biological mechanisms of lung cancer progression.

1.1 FACTORS AFFECTING PROGNOSTIC OUTCOMES IN DIFFERENT ASPECTS

Lung cancer prognosis is affected by multi-level factors (**FIGURE 1**), i.e., tumor level, patient level, and population level. In this dissertation, three studies were conducted with different methodologies to analyze those prognostic factors separately. For tumor level factors, we specifically focused on histopathological features.

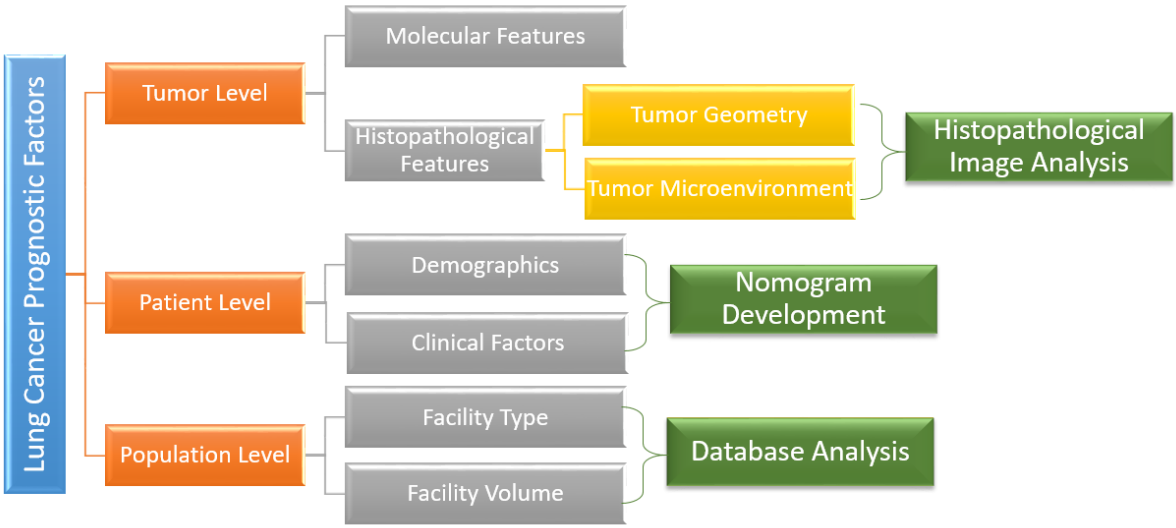


FIGURE 1 Overview of lung cancer prognostic factors and corresponding analysis methodologies.

1.1.1 Histopathological Features

Hematoxylin and Eosin (H&E)-stained tumor tissue slide scanning into electronic images is a routine procedure in lung cancer diagnosis. The pathological images produced by this procedure capture histological details in high resolution. Digital pathology images of tumor

tissues not only contain essential information for tumor grade and subtype classifications[6], but also information on the growth patterns and TME, such as the spatial distributions and organization of different types of cells. Out study aimed to automatically extract pathological image features and evaluate their prognostic value. We developed two deep learning algorithms to recognize tumor region and TME, separately. Well-defined feature extraction methods were then implemented to quantify histopathological features. Finally, based on image features, a prognostic model was built to evaluate risk score for each individual patient.

Tumor Diagnosis

Adequate TMN staging is critical for clinicians to choose accurate therapeutic methods[7]. Pathology image analysis acts as an important tool to determine the patients' TMN stage. For example, pathological N stage, which aims to describe the presence of lymph node invasion, relies on lymph node dissection. However, detecting tumor metastasis to lymph node is laborious and requires highly skilled pathologists, especially when amount of dissected lymph node is high and metastasis region is small. In addition to TMN staging, tumor spread through air spaces (STAS)[8] detection also requires detailed inspection of a whole slide image. STAS has been clinically illustrated as significantly negative prognostic factor for recurrence and survival[8]. Thus, fast and accurate quantification of STAS through automatic lung cancer detection is urgently desired. Our automatic tumor recognition model is anticipated to facilitate the laborious process required for tumor diagnosis.

TME Characteristics

In addition to lung cancer diagnosis, pathology images also provide tremendous information on TME. Specifically, lymphocyte, stromal cells, macrophages, and blood vessels mainly compose the lung tumor microenvironment (TME) in addition to tumor cells. It is especially noteworthy that in lung cancer, tumor-infiltrating lymphocytes (TIL) has been reported as positive prognostic factors[9], stromal cells has been reported to have complex prognostic effects[10,11], and angiogenesis have been reported as negatively associated with survival outcome[12]. Thus, in our TME segmentation model, 6 cell types were considered: tumor, stroma, lymphocyte, macrophage, karyorrhexis, and red blood cells.

1.1.2 Patient Level Factors

Both demographics and clinical information have been reported as prognostic factors. Demographics include age, gender, race, family history, smoking status, and social-economic status[2]. Clinical information include Tumor, Node, and Metastases (TNM) stage, tumor grade, and comorbidity[13]. It is noteworthy that these factors are usually correlated with each other, which should be taken into consideration in survival analysis. Our study aimed to use patient level factors collected in a national-wide database to produce an individualized prognostic model. Thus, we trained a prognostic nomogram, upon which we can easily get risk score and corresponding survival curve for an individual patient by simply adding risk scores contributed by clinical factors.

1.1.3 Population Level Factors

Population level factors refer to the factors that are defined or calculated in a group of patients instead of individuals, such as state-level diabetes rate and zip-code level household

income. Out study particularly focused on the factors related to patient care: facility type and facility volume. Although how teaching facility (TF) status[14-17] and facility volume[15,18-22] affect care quality has been widely investigated, the effect of non-TF type on survival outcome, whether the correlation between high-volume and better survival outcome persists in all types of facilities, and what the underlying reasons are remain unknown. Cox proportional hazard model was used to evaluate influence of facility type and volume on survival outcome; logistic regression model was used to evaluate influence of facility type and volume on probability of surgery selection.

1.2 ADVANCES IN DEEP LEARNING ALGORITHM FOR IMAGE RECOGNITION

To quantify histopathological features from H&E stained images, the challenge is to automatically classify and segment different cell types in pathology slides. Since deep learning methods has shown great power in handling image recognition tasks, we developed two deep learning algorithms for image recognition in pathology images. We will introduce the concepts in machine learning and deep learning, and summarize current applications and our contribution in this section.

1.2.1 The Concepts in Machine Learning

When solving a problem, we are actually predicting of the answer (the output) to the problem (the input). Sometimes we can manually define a fixed computer (the machine) algorithm to solve a problem, but in many cases, the algorithms are implicitly contained in the data; i.e., they are not easily to be theoretically defined without utilizing the data. The goal of “machine learning” is to learn the algorithm and to approximate the desired output[23]. To perform

machine learning, we need data that consists of multiple input instances and corresponding true outputs (the ground truth). Data used to derive the algorithm is called “training set”; data used to decide which algorithm to use from multiple trained ones is called “validation set”; data used to evaluate the prediction performance is called “testing set”.

1.2.2 Deep Learning Algorithms

Deep learning is a branch of machine learning field. Since 2012, deep learning has made significant improvements in all image recognition benchmarks[24-26]. The applications of deep learning algorithms in digital pathology have had remarkable success in traditional pathology tasks. For example, deep learning algorithms achieved performance comparable to pathologists in interpreting whole-slide images for the detection of tumor regions[27-29] and lymph node metastases[30].

To understand how deep learning excels in these areas, we build conceptual connections of deep learning in the machine learning literature. In essence, deep learning is a special kind of ANN, which is one category of machine learning algorithm. Deep learning and other ANNs are inspired by biological neural networks and mathematically construct a network model with multiple connected layers. The first network layer (called the “input” layer) receives inputs (e.g. slide images). It has a set of parameters and can use them to compute outputs. Similarly, each successive network layer receives inputs from its previous layers, uses its parameters, and computes outputs**.** At the end, the last network layer (called the “output” layer) calculates the outputs of the whole model. The layers between the “input” and “output” layers are not visible as they do not directly receive model input or generate model

outputs, and thus are called the “hidden” layers. In this process, prediction outputs from a good neural network can well approximate the observed outputs. Although ANNs claim excellent performances based on theoretical work[31], historically, it has been notoriously hard to calculate the network parameters when the total number of network layers exceeded three, which limited the performance of the model. Fortunately, this is no longer a severe bottleneck, owing to the advancements in computational hardware, the scale of data accumulation, and the improvements in algorithms. Nowadays, popular ANNs can have hundreds of layers. The machine learning community refers to these algorithms as “deep learning” to distinguish them from the conventional “shallow” artificial neural network (ANN) algorithm.

CNN

Convolutional Neural Network (CNN) is a form of deep learning model specifically designed to deal with high-dimensional data, such as 2-D and 3-D images. The basic structure of CNN is composed by a group of convolution layers and pooling layers, followed by several fully connected layers. The goal is to predict the class of input object, such as images. This network structure enables the extraction of representational features for prediction. The design of CNN is inspired by the functional mechanism of the visual cortex[24]: instead of using all outputs from the previous layer, a convolution kernel only focuses on a certain area, the so-called “receptive field”, to compute a feature at the corresponding spatial position. By spatially sliding the “receptive field” along the input dimensions (e.g., along the width and height directions for 2-D images), a “feature map” is computed as the outputs from the

convolution layer. As the number of parameters is determined by the area of the receptive field, convolution layers have much fewer parameters than the image size. This design thus effectively reduces the number of parameters within a neural network and greatly improves its computational efficiency.

Mask-RCNN

Through utilizing the building block of convolution layers and pooling layers, CNN has many derivatives, one of which is Mask Regional Convolutional Neural Network (Mask-RCNN)[32]. Different from the simple goal of image classification as CNN, Mask-RCNN aims to simultaneously identify the bounding boxes of all objects within the input image, classify each object, and segment the object within the bounding boxes. Thus, when multiple objects in the same category are close to each other, Mask-RCNN is still able to distinguish each object with its own boundary.

1.2.3 Applications in Tumor Pathologic Image Analysis

The applications of deep learning algorithms in digital pathology have had remarkable success in traditional pathology tasks. Deep learning algorithms achieved performance comparable to pathologists in interpreting whole-slide images for the detection of tumor regions[27-29] and lymph node metastases[30] by predicting each region of input pathology image as tumor or non-malignant. In addition to malignant region detection, deep learning models to distinguish different lung cancer subtypes were also developed. Coudray et al. trained a CNN to classify lung cancer image patches into normal, ADC, or SCC[33]. Coudray et al also trained a CNN to predict mutation status of 6 frequently mutated genes in lung ADC patients

based on 512*512 pixels pathology image patches[33]. To estimate the classification accuracy of mutated vs. non-mutated, the Area under the Curve (AUC) was between 0.733 and 0.856 for each of the 6 genes in validation dataset.

As TME is widely accepted as important factor affecting tumor progression and immunotherapy response, several deep learning models were also trained to characterize lung TME. Saltz et al. developed a CNN model to distinguish lymphocyte image patches against necrosis or other tissue patches across pathology slides of multiple cancer types including lung ADC and SCC[34]. Through quantifying spatial organization of detected lymphocyte image patches in whole slide imaging (WSI), they reported the relationship among TIL distribution patterns, prognosis, and lymphocyte fractions. Wang et al. developed another CNN model to distinguish tumor cell, stroma cell, and lymphocytes in cell level in lung ADC pathology images[35]. In Wang's study, basic image processing methods were used to extract small image patches centered with cell nuclei; the image patches were then categorized in different cell types using CNN. A prognostic model using image features describing the proportion and distribution of detected cells was then trained and validated in two independent datasets. Another important application is automatic microvessel segmentation to quantify pathological angiogenesis in lung ADC using Fully Convolutional Neural Network (FCN)[36], which is a derivative of CNN and aims at image segmentation[37]. The microvessel segmentation model by Yi et al is trained in lung ADC H&E stained images and showed generalizability to breast cancer and kidney cancer pathology images. While manual segmentation is laborious and error-prone, such automatic microvessel segmentation enables fast and quantitative characterization of area and spatial distribution of microvessels.

1.2.4 Contribution and Innovation of Our Methodology

Although traditional image processing methods have been applied to segment lymphocyte nuclei and to analyze spatial organization of cells in TME, such as TIL[34] and stroma cells[38], accurate and efficient lung tumor detection and TME segmentation remain big computational challenges because of the following reasons. 1) The composition of lung cancer microenvironment is complex; in addition to the aforementioned cell types, other structures including bronchus, cartilage, and pleura are also often detected in lung cancer pathology image. Such complexity makes manual segmentation laborious and traditional feature definition hard. 2) For H&E stained slides, the color could vary a lot according to different staining conditions and the length of period from slide making to scanning.

In this dissertation, through cooperation with experienced pathologists, we developed two individual deep learning algorithms for automatic tumor detection and TME computational staining. They were the first models to recognize tumor and different types of cell nuclei in lung WSI, separately. Based on these pioneer models, image features that were hard to extract using traditional image analysis tools were analyzed and proved prognostic. By applying the model development pipeline to other cancer types, our methodologies are to be generalized easily and readily. Thus, our research serves as first attempts to apply cutting-edge deep learning models to pathology image analysis in lung cancer and provides smooth pipeline to generalize our models to other cancer types.

CHAPTER TWO – DIGITAL PATHOLOGY IMAGE ANALYSIS

2.1 BACKGROUND AND RATIONALE

With the advance of technology, H&E stained tumor tissue slide scanning into electronic images is a routine procedure for cancer diagnosis. This procedure produces pathological images that capture histological details in high resolution. With the development of computational algorithms, automatic tumor region detection allows for tumor size calculation and tumor shape estimation, and automatic nuclei recognition allows for TME quantification. For analysis of H&E stained pathology images, deep learning methods have been developed to distinguish tumor regions[27], detect metastasis[28], predict mutation status[39], and recognize TIL regions[34] 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), neither automatic lung cancer region detection nor nuclei segmentation/classification deep learning algorithms from H&E stained pathology images were available. Thus, in this section, we aimed to develop the first deep CNN algorithm to automatically recognize tumor regions and the first Mask-RCNN algorithm to automatically recognize different cell types in TME of lung ADC (LUAD) H&E pathology images. The image features from two different aspects, the tumor shape level and the cellular level, were then quantified and used to build an image-based prognostic model. The flow-chart of this study in summarized in **FIGURE 2**.

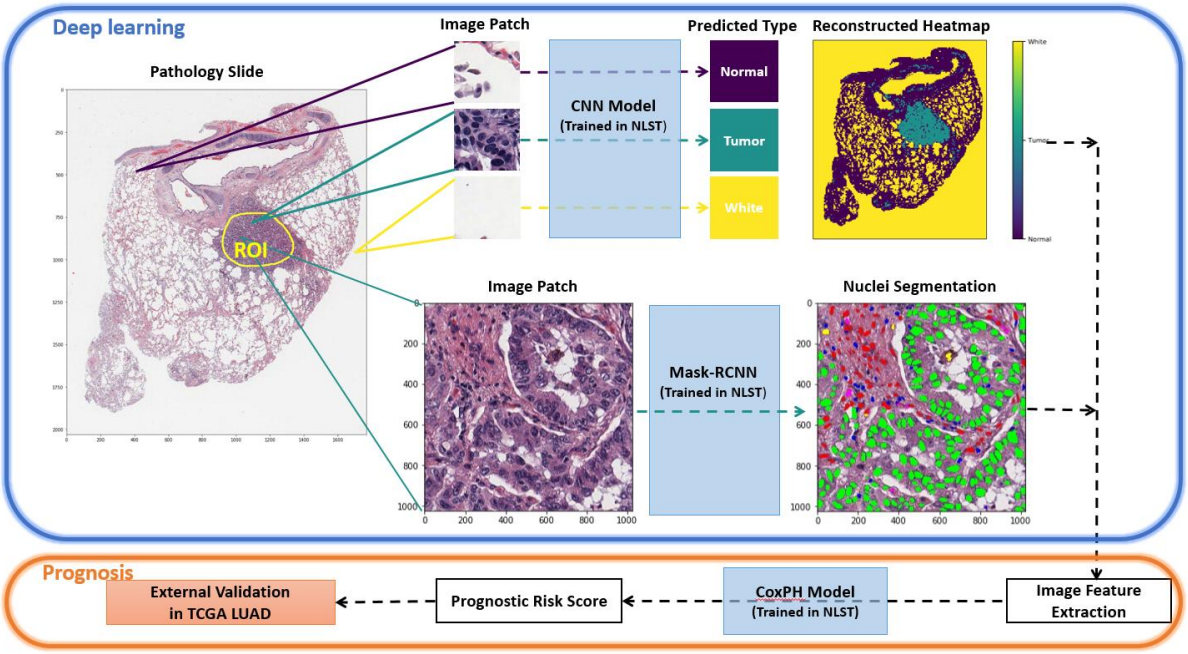


FIGURE 2 Flow chart of pathological image analysis pipeline for tumor detection and nuclei segmentation.

2.2 METHODS

2.2.1 Data Collection

NLST LUAD

208 40X H&E stained pathology images for 135 LUAD patients were acquired from the National Lung Screening Trial (NLST) dataset (https://biometry.nci.nih.gov/cdas/nlst/). The NLST is a randomized trial to screen lung cancer in high-risk patients with CT or single-view chest radiography. 54,000 participants were enrolled between 2002 and 2004; the median follow-up time was 6.5 years. In the NLST dataset, the H&E stained images were sampled from lung tumor tissues that were resected during diagnosis and treatment

of lung cancer. Corresponding clinical information for the LUAD patients from the NLST dataset were acquired. Clinical variables include age, gender, smoking history, stage, and information for overall survival.

TCGA LUAD

431 40X images for 372 LUAD patients were acquired from the Cancer Genome Atlas (TCGA) dataset (https://wiki.cancerimagingarchive.net/display/Public/TCGA-LUAD). The TCGA dataset comprehensively collected molecular and histopathological data spanning 33 cancer types. Corresponding clinical information for the 372 LUAD patients from the TCGA dataset were acquired. Clinical variables include age, gender, smoking history, stage, and information for overall survival. The patient characteristics were summarized in **TABLE 1**. Messenger Ribonucleic acid (mRNA) expression data for the TCGA dataset were available online at http://firebrowse.org. Gene expression data of the 372 patients from the TCGA LUAD dataset were downloaded and preprocessed. All gene sets from the Reactome database were used in the following analysis of image-genomic association[40]

TABLE 1 Patient characteristics for the NLST LUAD and TCGA LUAD datasets.

		NLST LUAD	**TCGA LUAD**
Number of patients		135	372
Number of pathology slides		208	431
Age at diagnosis (years, median (min - max))		64 (55 - 74)	66 (33 - 88)
Follow-up (years, median (min - max))		4.1 (0.1 – 7.1)	0.5 (0.0 - 5.9)
Vital status (%)	Alive	94 (69.6)	297 (79.8)
	Deceased	41 (30.4)	75 (20.2)
Gender (%)	M	77 (57.0)	208 (55.9)
	F	58 (43.0)	164 (44.1)
Cancer stage (%)	I	90 (66.7)	208 (55.9)
	II	12 (8.9)	95 (25.5)

	III	22 (16.3)	47 (12.6)
	IV	10 (7.4)	21 (5.6)
	NA	0 (0.0)	1 (0.3)
Smoking status (%)	Smoker	74 (54.8)	254 (68.3)
	Non-smoker	61 (45.2)	118 (31.7)

For both of the NLST and the TCGA dataset, a specialized lung cancer pathologist, Dr. Lin Yang, M.D., labeled the Region of Interest (ROI) for each of the pathology images. Another lung cancer pathologist, Dr. Adi Gazdar, M.D., Professor, confirmed the labelling.

2.2.2 Deep Learning for Tumor Detection

Image Patch Generation

For tumor detection, 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. 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. 2139 non-malignant, 2475 tumor and 730 white patches were generated in total. Images were scaled to the range [0, 1] by dividing by 255 before being fed into the model.

CNN Training Process

The Inception (V3) architecture[41] 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 backend. 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 normalize 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 validation accuracy failed to improve for 10 epochs.

Prediction Heatmap Generation

To avoid prediction on a large empty image area and to speed up the prediction process, the Otsu thresholding method followed by morphological operations such as dilation and erosion was first applied to pathology images to generate the tissue region mask[42,43]. A 300×300 pixel window was then slide 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 image (**FIGURE 3**). For each image patch, the class with the highest probability was determined as the predicted class.

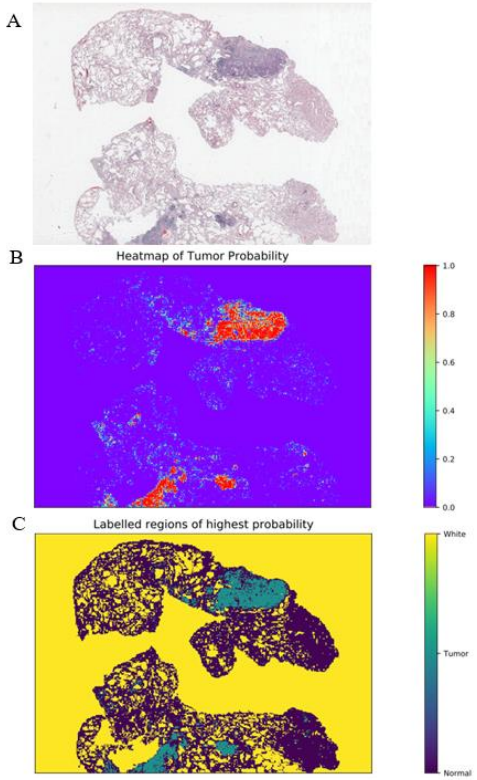


FIGURE 3 Whole-slide tumor region detection. (A) Original slide. **(B)** Predicted tumor probability. Each point in the heatmap corresponds to 300 × 300 pixels image patch in original 40x

slide. **(C)** Predicted region labels. Yellow: non-tissue background; green: tumor region; blue: normal region.

2.2.3 Deep Learning for Nuclei Segmentation

Training, Validation, and Testing Sets Preparation

In order to construct the training set for the Mask-RCNN algorithm, 127 image patches (500 × 500 pixels) from 39 pathological ROIs were extracted from the NLST dataset (**FIGURE 2**). In these patches, different types of cell nuclei were labeled for each image patch under supervision of expert pathologists. Every pixel within the mask was labeled as 1 of 7 categories: tumor nuclei, stroma nuclei, lymphocyte nuclei, macrophage nuclei, red blood cells, karyorrhexis, and others (background). These labels (also called as the mask) were then used as the ground truth for training Mask-RCNN model and evaluating the model performances. The labeled images were randomly divided into training, validation, and testing sets. To ensure independence among these datasets, image patches from the same ROI were assigned together. More than 12,000 cell nuclei were included in the training set, while 1227 and 1086 nuclei were included in the validation and testing set, respectively.

Training Process

A neural network model was developed using the Mask-RCNN architecture. The pre-trained model was fine-tuned on our training dataset from the NLST study. Images were standardized (centered and scaled to have zero mean and unit variance) for each red, green, and blue (RGB) channel. To increase generalizability and avoid bias from different H&E staining conditions, we performed extensive augmentations on the image patches. Specially,

random projective transformations were applied to images and their corresponding masks; each image channel was randomly shifted using linear transformation. For the training process, the batch size was set to 2, the learning rate was set to 0.01 and decreased to 0.001 after 500 epochs, the momentum was set to 0.9, and the maximum number of epochs to train was set to 1000. In the validation set, the model trained at the 707th epoch reached the lowest loss. As a result, this model was selected and used in the following analysis to avoid overfitting. Python (version 3.5.2) and python libraries (Keras, version 2.1.5; openslide-python, version 1.1.1; tensorflow-gpu, version 1.8.0) were used[44].

Segmentation Performance Evaluation

Since the Mask-RCNN model simultaneously segments and classifies cell nuclei, three criteria were used to evaluate the segmentation performance in the validation and testing datasets respectively. First, detection coverage was calculated as the ratio between the detected nuclei and the total ground truth nuclei. Each ground truth nuclei was matched to a segmented nucleus, which generated the maximum Intersection over Union (IoU). If the IoU for a ground truth nuclei were > 0.5, this nuclei would be labeled “matched”; otherwise it was labeled “unmatched”. Second, nuclei classification accuracy was determined for the matched nuclei by comparing the predicted nuclei type with the ground truth. Third, segmentation accuracy was evaluated by the IoUs, which were calculated for each detected nuclei and averaged in different nuclei categories.

2.2.4 Quantification of Pathology Image Features

Tumor Level Features

In a pathology image, sometimes there are multiple tissue samples. To distinguish different tissue samples in the same image, disconnected tissue regions were first identified by morphological operations on heatmaps of predicted classes[43]. 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 image were removed from analysis. Within each tissue region, the tumor region with the largest area was regarded as the "main tumor region". Three classes of 22 tumor region features were estimated for each tissue sample: tumor size description, tumor shape description, and negative control. When multiple tissue samples were available for one patient, either due to multiple tissues within one image or multiple images for one patient, the 22 image features were averaged to generate patient-level image features.

Features Describing Tumor Size

Area was used to estimate the number of image patches predicted as tumor. Convex area was used to estimate the number of patches within the convex hull of predicted tumor region. Filled area was used to estimate the number of patches within the predicted tumor region with all the holes filled in. Perimeter was used to estimate the length of tumor region borders. Major/minor axis length was used to estimate the axis length of ellipse that has the same normalized second central moments with the predicted tumor region.

Features Describing Tumor Shape

Number of regions was used to estimate the number of disconnected tumor regions. Number of holes was used to describe the Euler characteristics of tumor region. Here, 8-connectivity was used to determine disconnected tumor regions and disconnected holes[45]. Perimeter^2 to area ratio was used to describe the roundness of tumor border. Eccentricity was the ratio of the focal distance over the major axis length of the ellipse; eccentricity=0 indicates a circle. Extent was the ratio of the area to the number of patches in the bounding box of tumor region. Solidity was the ratio of the area to the convex area.

Features Designed as Negative Control

Angle between the X-axis and the major axis for the main tumor region only depended on how pathologists put the tissue onto the pathology slide, and thus served as a negative control.

Cell Level Features

In order to make the nuclei segmentation model computationally more efficient while getting a good representation of each ROI, instead of applying the Mask-RCNN model to the whole slide, 100 image patches (1024 × 1024 pixels) were randomly sampled and analyzed for each pathologist-labeled ROI. These 100 image patches provide a good coverage of the ROI. Nuclei were then segmented and classified through the Mask-RCNN model developed from section 2.2.3.

Nuclei Density of Different Cell Types

The density of each type of nuclei were calculated as the nuclei amount per 1024 × 1024 pixels image patch (yielding 6 image features).

Nuclei Distribution of Different Cell Types

In order to characterize the spatial organization of cells using a graph, we calculated the centroids of nuclei and used them as vertices to construct a Delaunay triangle graph for each image patch[46]. The Delaunay triangle graph connects nuclei into a graph, and the number of connection and the average length (i.e. spatial distance) between two types of nuclei summarize the spatial organization of different types of cell. Since 6 nuclei categories were included in this study, the edges of the graph were classified into 21 categories [i.e. 6 × (6-1)/2 + 6 = 21] according to their vertices pairs. For each image patch, the number of connections (i.e. edges) for different categories were counted (which leads to 21 features), the lengths of the connections were averaged for each edge category (yielding another 21 image features).

Thus, in total, 48 cell level image features were extracted. The image features were averaged across the 100 patches for each ROI in pathology image. When 2 or more pathology slides were available for 1 patient, the features from the slides are averaged for each patient.

2.2.5 Prognostic Model Based on Pathological Image Features

Model Development

Overall survival, defined as the date of diagnosis till death or last contact, was used as the response for survival analysis. An elastic-net-regularized Cox proportional hazard model was

developed in the NLST LUAD dataset and independently validated on the TCGA LUAD dataset. Given a set of input image features for each patient, the output of the Cox regression model was a risk score, with a higher risk score indicating worse prognosis.

Model Validation

Based on the risk scores, the patients in TCGA LUAD cohort were dichotomized into high- and low-risk groups using the median risk score as a cutoff. A log-rank test was used to compare survival differences between predicted high- and low-risk groups. The survival curves were estimated using the Kaplan-Meier (K-M) method. A multivariable Cox proportional hazard model was used to test the prognostic value and determine the hazard ratio of risk groups defined by image features after adjusting for other clinical characteristics, including age, gender, smoking status, and stage. R software, version 3.4.2, and R packages (survival, version 2.41-3; glmnet, version 2.0-13; spatstat, version 1.55-1) were used[47,48]. The results were considered significant if two-tailed p value <0.05.

2.2.6 Association Analysis between Image Features and Genetic Pathways

Gene expression data of 372 patients from the TCGA LUAD dataset were downloaded and preprocessed. Spearman rank correlation was used to evaluate the correlation between mRNA expression levels and image features. For each image feature, Spearman rank correlations for all mRNA expression levels were used for gene set enrichment analysis (GSEA). All gene sets from the Reactome database were used[40]. For multiple testing correction, Benjamini-Hochberg (BH)-adjusted p values were used to detect significantly

enriched gene sets. Gene sets with BH-adjusted two-tailed p values < 0.05 were regarded as significantly enriched. R packages Hmisc (version 4.1-1), fgsea (version 1.4.1), and gplots (version 3.0.1) were used[49].

2.3 AUTOMATIC TUMOR REGION RECOGNITION FOR PATHOLOGY IMAGES

2.3.1 CNN Model Distinguishes Tumor Patches from Non-malignant and Empty Region Patches

5344 tumor, non-malignant, and white image patches were extracted from 27 LUAD H&E stained pathology images. The image patches were split into training, validation, and testing datasets. 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 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.

2.3.2 Tumor Region Recognition in Giga-pixel Pathology Images

In the NLST LUAD dataset, the pathology images have sizes ranging from 5280×4459 pixels to 36966×22344 pixels (median 24244×19261 pixels). To identify tumor regions, each image was partitioned into 300×300 image patches. To speed up prediction, tissue regions were first identified and only the image patches within the tissue regions were predicted by the CNN model (**FIGURE 4**). The predicted probabilities of the image patches were summarized into heatmaps of tumor probability (**FIGURE 3**). An example of a tumor probability heatmap is shown in **FIGURE 3**. The tumor region heatmap, predicted as the category with highest probability, is shown in **FIGURE 3**. Each pixel in the heatmaps corresponds to a 300×300 pixel image patch in the original 40X pathology image.

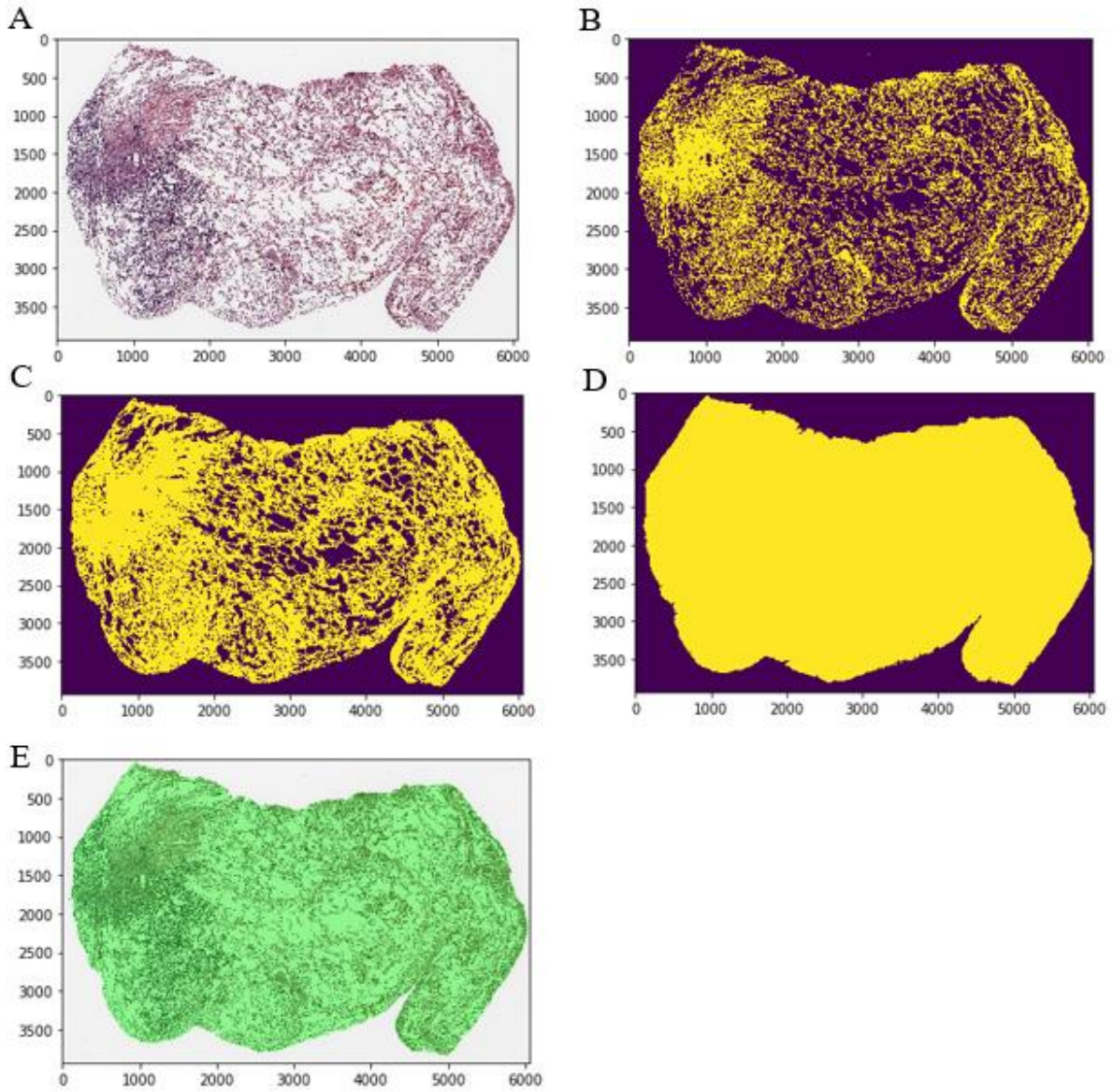


FIGURE 4 Operations to speed up slide-level prediction process. (A) The original slide. **(B)** The image mask after Otsu thresholding. **(C)** The image mask after dilation and removing small objects of the mask in **(B)**. **(D)** The final mask after dilation, erosion, and filling up holes of mask in **(C)**. **(E)** Overlap final image mask and original pathology slide.

2.4 AUTOMATIC TME CHARACTERIZATION FOR PATHOLOGY IMAGES

2.4.1 Mask-RCNN Simultaneously and Accurately Classifies and Segments Cell Nuclei

The developed Mask-RCNN model detects the bounding box and category for the individual nucleus, and segments the nucleus at the same time. **FIGURE 5** demonstrates some of the segmentation results. In total, there were 6 nuclei categories: tumor, stroma, lymphocyte, macrophage, red blood cell, and karyorrhexis, and all other remaining structures or spaces were considered as background. Different nuclei were colored according to the predicted categories (**FIGURE 5**). For detected objects, the overall classification accuracy was 85% and 85% in the validation set and the testing set, respectively, while the accuracy for tumor nuclei was 88% in validation and 90% in testing, respectively. The detection coverage was 75% and 77% in the validation set and the testing set, respectively. The mean IoU was 76% overall and 76% for tumor nuclei in the testing set.

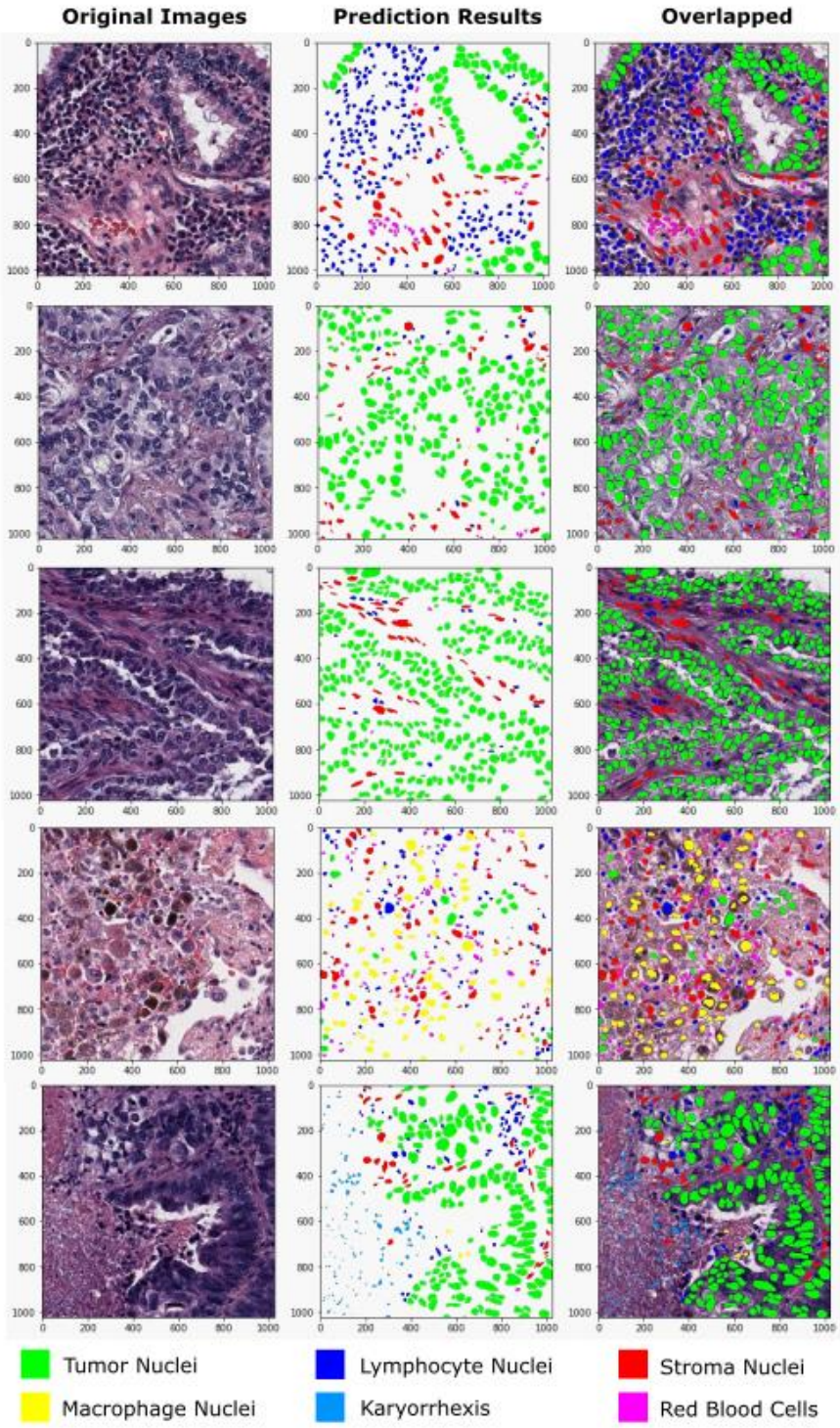


FIGURE 5 Deep-learning based nuclei segmentation results in pathological images. The image patches are 1024*1024 pixels under 40X magnification and from different patients in the TCGA LUAD dataset. **2.4.2 Generation of Whole Slide Nuclei Segmentation Map**

The developed Mask-RCNN model can be applied to the entire digital pathology image to generate a whole-slide map of nuclei segmentation map, where tumor region and lymphocyte infiltration areas were clearly illustrated (**FIGURE 6**).

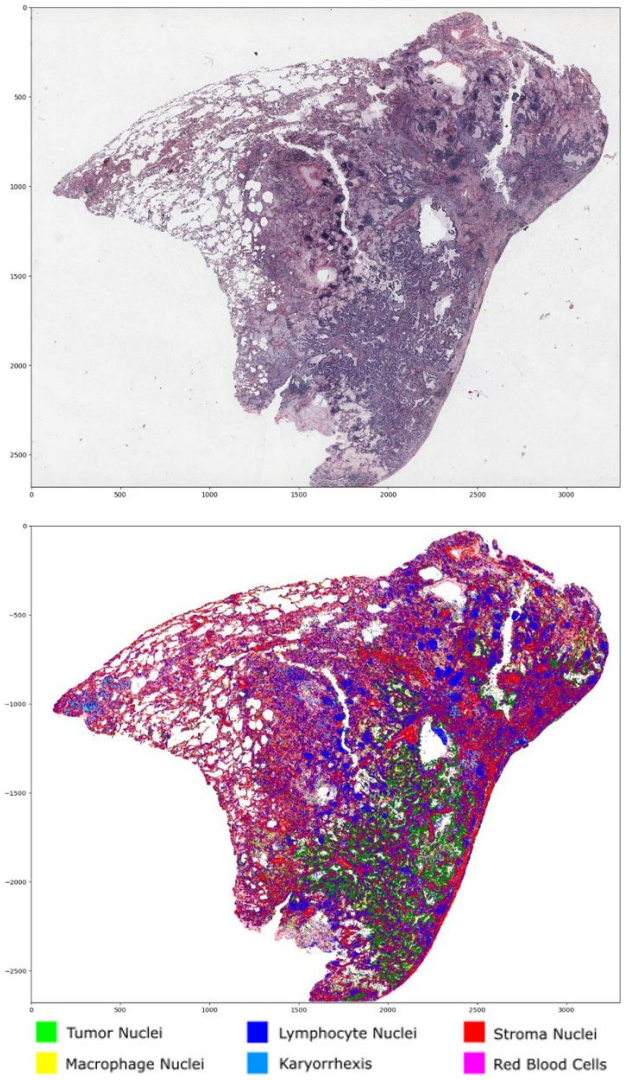


FIGURE 6 Whole-slide nuclei segmentation. Original pathology slide (upper) and detected nuclei (bottom) are shown.

2.5 PROGNOSTIC VALUE OF PATHOLOGY IMAGE FEATURES

2.5.1 Features of Predicted Tumor Regions Correlate with Survival Outcome

Based on the predicted tumor region heatmap, tissue samples were identified and 22 shape and boundary-based features were extracted for each tissue sample. 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 were summarized in **TABLE 2** in the NLST LUAD dataset. It shows that many features were associated with survival outcome. Most tumor area-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 perimeter^2 to area ratio (an estimation of circularity and boundary roughness), were also associated with poor survival outcome (for all tumor regions: per 100 number of holes, HR = 1.087, p value = 0.033; per 1000 perimeter^2 to area ratio, HR = 1.15, p value = 0.016; similar results for main tumor region). Examples comparing tumor regions with high and low values of eccentricity and perimeter2 to area ratio of main tumor region were illustrated in **FIGURE 7**. As expected, the angle between the X-axis and the major axis of the main region was not correlated with survival, which served as a negative control of the feature extraction process.

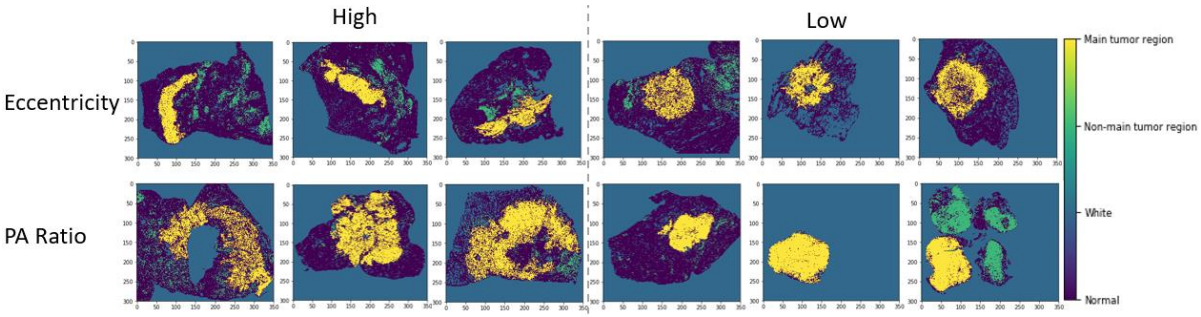


FIGURE 7 Comparison of tumor shapes with different values of eccentricity and PA ratio. Original heatmaps are cropped to the same size with same image scale. Yellow, main tumor region; green, non-main tumor region (as defined in section 2.2.4); dark blue, normal tissue; blue, blank part of pathology slide. PA ratio, perimeter^2 to area ratio.

TABLE 2 Univariable survival analysis of tumor region features. * 1 pixel in heatmap = 1 patch in 40X pathological slide. Patch size, 300 * 300 pixels.

Tumor region features	**HR (95% CI)**	**p value**
Number of regions (per 1000)	1.29 (0.64-2.58)	0.48
Area sum of all regions (per 1000 pixel*)	1.03 (1.01-1.05)	0.003
Perimeter sum of all regions (per 1000 pixel)	1.09 (1.03-1.15)	0.003
Sum of convex area for all regions (per 1000 pixel)	1.02 (1.01-1.03)	0.005
Sum of filled area for all regions (per 1000 pixel)	1.03 (1.01-1.05)	0.003
Sum of hole numbers of all regions (per 100)	1.09 (1.03-1.16)	0.003
Sum of major axis length of all regions (per 1000 pixel)	1.40 (1.00-1.96)	0.051
Sum of minor axis length of all regions (per 1000 pixel)	2.65 (1.10-6.40)	0.030
Perimeter^2/area of all regions (per 1000)	1.18 (1.03-1.35)	0.019
Area of main region (per 1000 pixel)	1.03 (1.01-1.05)	0.009
Convex area of main region (per 1000 pixel)	1.02 (1.00-1.03)	0.010
Eccentricity of main region	6.37 (0.57-71.56)	0.13
Hole number of main region (per 100)	1.09 (1.02-1.15)	0.006
Extent of main region	4.90 (0.19-126.30)	0.34
Filled area for main region (per 1000 pixel)	1.03 (1.01-1.04)	0.007
Major axis length for main region (per 100 pixel)	1.57 (1.11-2.21)	0.010
Minor axis length for main region (per 100 pixel)	1.73 (1.05-2.83)	0.031
Angle between the X-axis and the major axis of main region	0.98 (0.64-1.50)	0.92
Perimeter of main region (per 1000 pixel)	1.09 (1.02-1.15)	0.007
Solidity of main region	7.24 (0.45-117.40)	0.16
Average tumor probability of the main region. (per 0.10)	1.11 (0.53-2.24)	0.78
Perimeter^2/area for main region (per 1000)	1.21 (1.03-1.42)	0.021

2.5.2 Nuclei Recognition-based Features of TME Correlate with Survival Outcome

In univariable analysis in the NLST LUAD dataset, several cell level features were significantly correlated with survival outcome (**TABLE 3**). For example, higher karyorrhexis density, karyorrhexis-karyorrhexis connections and karyorrhexis-red blood cell connections were associated with worse survival outcome, which was expected as both features indicate a high rate of necrosis. Furthermore, stromal nuclei density and stromal-stromal connection were associated with better survival outcome, which agreed with our current knowledge that more stromal tissues corresponds to better prognosis.

TABLE 3 Significantly prognostic cell level features in univariable survival analysis. Features are dichotomized by the median value. P values are calculated using Ward test.

Cell level features (high vs. low)	HR (95% CI)	p value
Stroma-stroma edges number	0.32 (0.16 – 0.63)	0.001
Karyorrhexis-karyorrhexis edges number	2.54 (1.32 – 4.91)	0.006
Stroma nuclei number	0.41 (0.21 – 0.80)	0.009
Karyorrhexis number	2.27 (1.19 – 4.33)	0.013
Red blood cell-karyorrhexis edges number	1.98 (1.05 – 3.74)	0.035

2.5.3 Development and Validation of Prognostic Model

The tumor level feature and cell level features were pooled together to build a prognostic model. The image feature-based model was developed in NLST LUAD dataset and then independently validated in the TCGA LUAD dataset. Coefficients of the prognostic model were summarized in **TABLE 4**. For each patient in the TCGA LUAD dataset, the model predicts a risk score, based on which, the patient was assigned to either high- or low-risk group. The survival curves of the predicted high and low-risk groups were shown in **FIGURE 8**, where the high-risk group showed significantly worse survival than the low-risk group (log-rank test, p value = 0.00067). The risk group predicted by image features also served as an independent prognostic factor (high- vs. low-risk, hazard ratio [HR] = 2.23, 95%

confidence interval [CI] = 1.34-3.70, p value=0.002), after adjusting for clinical variables including age, gender, smoking status, and stage (
TABLE 5).

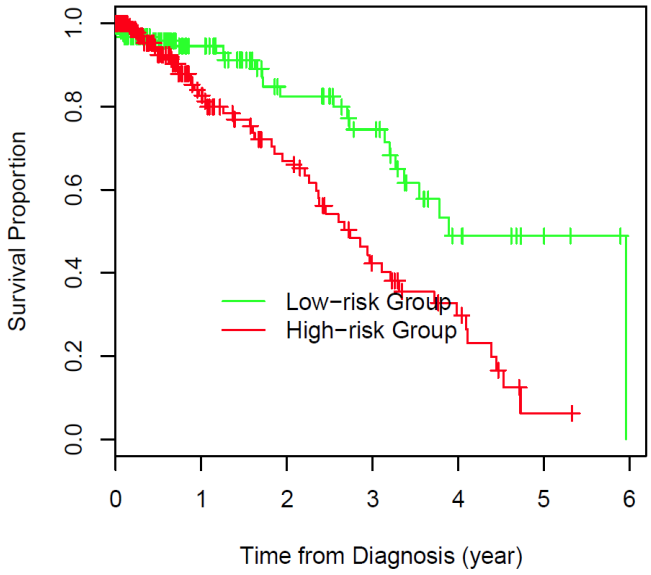


FIGURE 8 K-M plot of predicted high- and low-risk group in the TCGA LUAD dataset. Log-rank test, p value = 0.00067.

TABLE 4 Variable coefficients in the image-feature based risk prediction model.

Features	**Coefficients**
Perimeter sum of all regions (per 1000 pixel)	$1.38*10^{-2}$
Perimeter^2/area of all regions (per 1000)	$5.51*10^{-2}$
Stroma cell density	$-7.59*10^{-5}$
Tumor-lymphocyte connection density	$4.25*10^{-6}$
Tumor-necrosis connection density	$-8.89*10^{-4}$
Stroma-lymphocyte connection density	$-2.29*10^{-7}$
Stroma-macrophage connection density	$-1.16*10^{-4}$
Lymphocyte-lymphocyte connection density	$5.10*10^{-8}$
Average length of tumor-stroma connection	$3.05 *10^{-2}$
Average length of tumor-red blood cell connection	$-2.34*10^{-3}$
Average length of tumor-macrophage connection	$-1.45*10^{-5}$
Average length of stroma-red blood cell connection	$-7.95*10^{-3}$
Average length of stroma-macrophage connection	$-2.66*10^{-2}$
Average length of lymphocyte-macrophage connection	$-1.08*10^{-2}$

TABLE 5 Multivariable analysis of image feature-based risk in the validation dataset.

TCGA dataset (n=371)	HR (95% CI)	p value
High- vs. low-risk	2.23 (1.34 – 3.70)	0.002
Age (year)	1.02 (0.99 – 1.05)	0.097
Male vs. female	0.79 (0.48 – 1.30)	0.35
Smoker vs. non-smoker	0.99 (0.60 – 1.64)	0.98
Stage		
Stage I	ref	-
Stage II	2.26 (1.28 – 4.00)	0.005
Stage III	4.06 (2.16 – 7.63)	<0.001
Stage IV	3.52 (1.38 – 8.98)	0.008

2.6 ASSOCIATION BETWEEN IMAGE FEATURES AND TRANSCRIPTIONAL ACTIVITY OF BIOLOGICAL PATHWAYS

GSEA was performed to study the association between image-derived TME features and the mRNA expression generated from bulk tumor tissues in the TCGA LUAD dataset. Our analysis identified biological pathways whose mRNA expression profiles were significantly correlated with image-derived features (such as densities of tumor cell nuclei, stromal cell nuclei, lymphocytes and karyorrhexis) (**FIGURE 9A-D**). For example, we observed that transcription activation of both T-cell receptor (TCR) and Programmed cell death protein 1 (PD1) pathways were positively correlated with the lymphocytes density in the tumor tissue of each patient (**FIGURE 9A**). This observation is consistent with previous reports that genes involved in TCR and PD1 pathways were expressed in immune cells[50,51], supporting for the reliability of the TME features derived by this deep-learning based approach. In addition, the expression of the extracellular matrix organization gene set, for which fibroblasts act as an important source[52], is positively correlated with the density of stromal

cells density in tumor tissue of each patient (**FIGURE 9B**). This correlation also supports the image-derived cell composition features. It is noteworthy that after we randomly shuffled the patient IDs and repeated the same analysis, such correlation was no longer observed, which serves as a negative control.

Furthermore, in our results, cell cycle gene expression was positively enriched with both the tumor cell density and the karyorrhexis density for each patient (**FIGURE 9C, D**). To look into this enrichment, expression levels of genes within the cell cycle gene set with p value <0.001 in Spearman rank correlation analysis for the TCGA LUAD dataset were visualized in **FIGURE 9E**, where patients were sorted and grouped according to their tumor nuclei density. Positive correlations between gene expression and tumor nuclei density can be observed for most of the genes. Only one cell cycle related gene, *POLD4*, showed an inverse trend (**FIGURE 9E**). Furthermore, this pattern of *POLD4* compared with other genes in the cell cycle gene set was reported previously in lung cancer[53]: while most cell cycle genes were upregulated in lung cancer, *c* is usually downregulated. The observation of the upregulated cell cycle gene set with the density of tumor cells indicates that the image-derived tumor cell density is representative of tumor status.

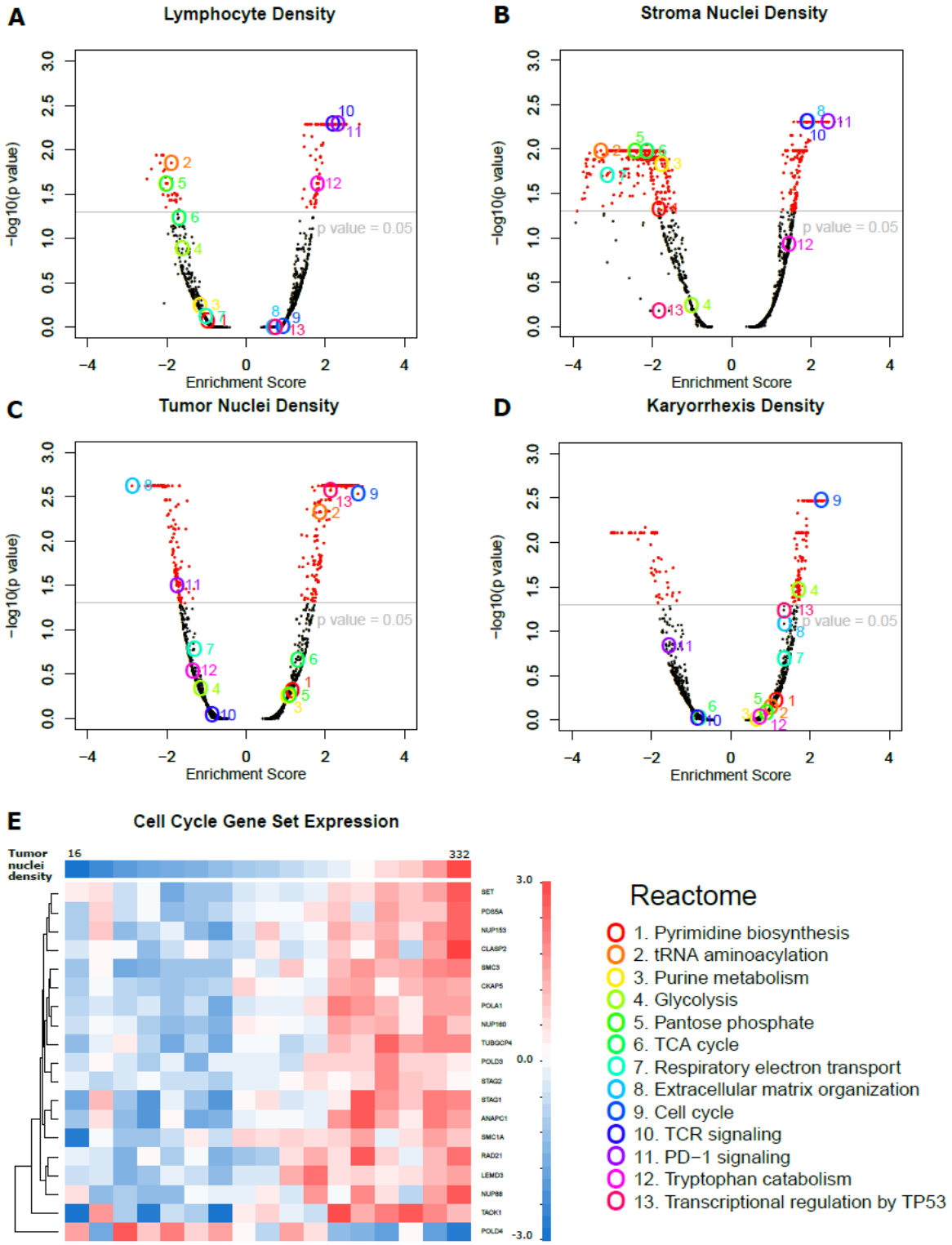

A
Lymphocyte Density
B
Stroma Nuclei Density
C
Tumor Nuclei Density
D
Karyorrhexis Density
−log10(p value)
Enrichment Score
p value = 0.05
E
Cell Cycle Gene Set Expression
Tumor nuclei density
16
332
SET
PDS5A
NUP153
CLASP2
SMC3
CKAP5
POLA1
NUP160
TUBGCP4
POLD3
STAG2
STAG1
ANAPC1
SMC1A
RAD21
LEMD3
NUP88
TACK1
POLD4
3.0
0.0
-3.0
Reactome
1. Pyrimidine biosynthesis
2. tRNA aminoacylation
3. Purine metabolism
4. Glycolysis
5. Pantose phosphate
6. TCA cycle
7. Respiratory electron transport
8. Extracellular matrix organization
9. Cell cycle
10. TCR signaling
11. PD−1 signaling
12. Tryptophan catabolism
13. Transcriptional regulation by TP53

FIGURE 9 Correlation between image features and mRNA expression in bulk tumor. (A-D) Volcano plots of gene set expression enrichment analysis results correlating mRNA expression level with tumor nuclei density **(A)**, stroma nuclei density **(B)**, lymphocyte density **(C)**, and karyorrhexis density **(D)** respectively. 13 interesting gene sets are highlighted. **(E)** To look into the significantly correlated gene sets, an example heatmap shows that most mRNA expression level in the cell cycle gene set are positively correlated with tumor nuclei density. Only genes with p value < 0.001 in spearman rank correlation with tumor cell number are shown. Patients are grouped according to tumor cell number per image patch showing on the top row. For best view, mRNA expression levels are centered across all patients for each gene.

2.7 PUBLICLY ACCESSIBLE PATHOLOGICAL IMAGE ANALYSIS TOOL

2.7.1 GitHub for Tumor Region Detection and Characterization Tool

All scripts of utilizing the CNN model for tumor region detection and feature extraction were shared in GitHub: https://github.com/sdw95927/pathology-images-analysis-using-CNN. The open-source scripts enables researchers to reproduce our results and develop new models in other cancer types.

2.7.2 Webserver for Pathology Image Segmentation Model

In order to facilitate the usage of this Mask-RCNN model, we developed an online tool (http://lce.biohpc.swmed.edu/maskrcnn/analysis.php) for this deep-learning based nuclei segmentation and classification model **(FIGURE 10)**. The only required input for this tool is a pathology image **(FIGURE 10A)**. Each uploaded input image will be assigned a job ID **(FIGURE 10B)**. The segmentation results will be automatically displayed and the spatial coordinates of each nucleus can be downloaded as an Excel table **(FIGURE 10C)**.

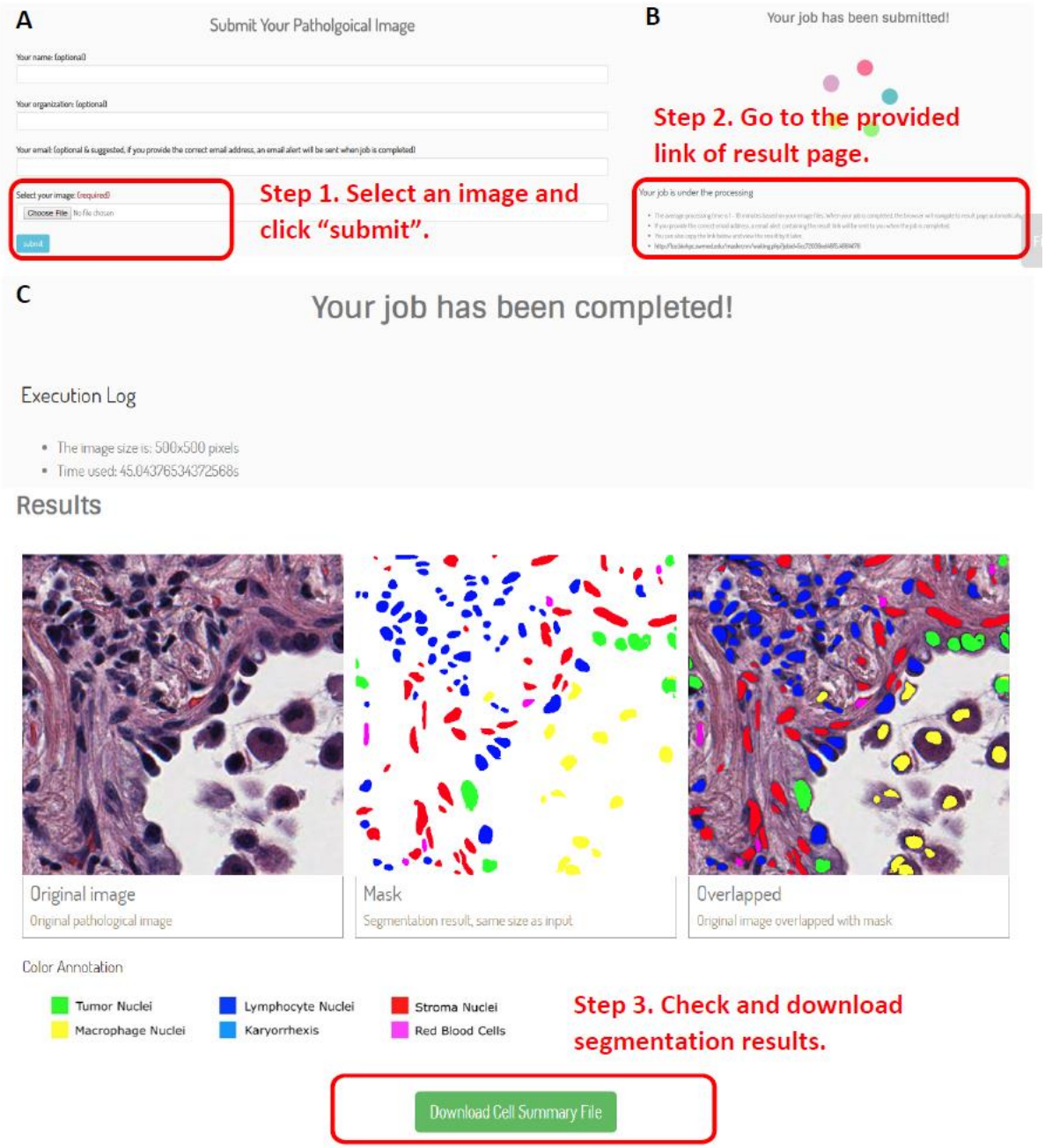


FIGURE 10 Illustration of performing pathological image segmentation online. Web-portal: http://lce.biohpc.swmed.edu/maskrcnn/analysis.php. Using 40X H&E stained pathology images smaller than 1024 * 1024 pixels as input will automatically generate segmentation result and cell information will be downloadable. Computation time for each image is usually shorter than 1 minute.

2.8 DISCUSSION

In this section, two deep-learning based analysis tools were developed using standard H&E stained pathology images to study the tumor shape features and TME, respectively. These tools successfully visualizes the presence of malignant regions and the spatial distribution of tumor cells, stromal cells, lymphocytes, and inflammatory cells in the TME from LUAD patients. The tumor region features and topological features of cell spatial organizations were quantified and found to be associated with patient survival outcome and transcription expression of biological pathways. Based on these tumor shape-based and TME-related image features, we developed a prognostic model for LUAD patients and the model was independently validated in another patient cohort. Our results show that the prognostic model predicts patient survival independent of other clinical variables in the validation cohort.

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]) can be easily detected in the heatmaps[8,54]. 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 the entire 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.

This is also the first study to automatically segment and classify nuclei in LUAD. Although several previous studies have tried to analyze the TME and discover prognostic image features, these studies involved time-consuming hand labeling by pathologists[38,55,56]. In contrast, we developed a fully automated and subjective nuclei segmentation and classification strategy that requires manual labeling only in constructing the training set. In addition, this deep-learning aided method enables segmentation of all nuclei within a whole slide image. Since the number of cells for a whole slide image could be tremendous (~2,000,000 on average), manually labelling all of them is impractical. Thus, this deep-learning method empowers quantification of the microenvironment across the whole slide image. Furthermore, although developed in LUAD, this method can be easily generalized to other cancer types by retraining the model using the tools on our website.

The associations between the extracted TME features and patient prognosis were evaluated in this study. Karyorrhexis, a representative of necrosis, has been reported as an aggressive tumor phenotype in lung cancer[57]. Consistently, the numbers of karyorrhexis and karyorrhexis-karyorrhexis edges were shown as negative prognostic factors in this study. On the other hand, the numbers of stromal cells and stromal cell-stromal cell edges were positive

prognostic factors, which is consistent with a recent report on LUAD patients[11]. Those consistencies support the validity of this segmentation neural network and the potentiality of using cell organization features as novel biomarkers for clinical outcomes.

The image-derived TME features show interesting correlations with the transcriptional activities of biological pathways. For example, gene expression levels of TCR and PD-1 pathways were positively correlated with the number of lymphocytes detected from tumor tissues. This indicates the image-derived TME features may be used to study or predict immunotherapy response, since several promising cancer immunotherapies rely on activation of tumor-infiltrated immune cells and blocking immune checkpoint pathways[51,58]. In addition, the gene expression extracellular matrix organization pathway is associated with the number of stromal cells in tumor tissues. Since traditional transcriptome sequencing is done in bulk tumor, accurate cell composition derived from pathological images could help to improve the evaluation of gene expression for each individual cell type. Moreover, the correlation between image features and transcriptional patterns of biological pathways hints at the potential usage of image features to study tumor bioprocesses, including cell cycle and metabolism status.

Gene expression patterns have been widely used to study the underlying biological mechanisms of different tumor types and subtypes[59,60]; moreover, genes with abnormal expression could become potential therapeutic targets of cancers[61,62]. However, traditional transcriptome profiling is usually done in bulk tumor[63], which contains multiple cell types, such as stromal cells and lymphocytes, in addition to tumor cells. This bulk tumor-based

sequencing could blur or diminish the mRNA expression changes arising from a single cell type or from different cell compositions in the TME. Currently, the relationship between the transcription activities of biological pathways and the tumor morphological microenvironment remains unclear. In this study, using the Mask-RCNN model, we derived TME-related features correlated with the expression patterns of biological pathways, indicating the biological relevance of these TME features. The Mask-RCNN model can be potentially used as tool to study TME in addition to gene expression analysis.

There were some limitations to these two models for tumor geometry quantification and TME dissection. First, information on the individual nucleus is not considered since this study focused on nuclei organization. Morphological and intensity features of nuclei have been reported as prognostic factors, which can be automatically extracted using this nuclei segmentation algorithm[64]. Second, some special structures, such as bronchus and cartilage, were not included in the TME dissection algorithm. This study handled this problem by avoiding such structures during ROI annotation. However, a more comprehensive training set would be desirable for whole slide analysis. Third, pathological images 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.

CHAPTER THREE – NOMOGRAM WITH CLINICAL INFORMATION

3.1 BACKGROUND AND RATIONALE

Lung cancer is the leading cause of death from cancer in the United States and worldwide. SCLC accounts for 13.6% of all lung cancer cases[1,65]. Compared to non-small-cell lung NSCLC, in which the 5-year survival rate is 18.0%, SCLC has only a 6.2% 5-year survival rate, and is characterized by a more rapid tumor growth rate and death from recurrent disease[66,67]. Over the last several decades, there have been only modest improvements in patient survival[68] and no molecularly targeted therapy has proven beneficial for SCLC patients[69]. Nomogram prognostic models that predict patient outcomes may facilitate better treatment stratification and outcome evaluation, as well as more refined patient enrollment criteria for clinical trials in SCLC. Furthermore, a recent study in breast cancer[5] showed that user-friendly online prognostic tools could greatly enhance patient care. However, currently there are no such online tools available for prognosis of SCLC.

To date there are three studies of nomograms in SCLC, published by Xie et al[67], Pan et al[70], and Xiao et al[71]. The nomograms developed from those studies provide useful tools for clinicians and researchers to stratify the risk of SCLC patients. However, two of the studies simply classified patients as limited or extensive stage without using the more accurate TNM staging proposed by the IASLC[72]. Furthermore, there is a lack of independent validation for these models, probably due to the limited sample size (n = 938[67], 275[70], and 647[71]

separately). Other non-nomogram prognostic models include the Manchester score and Spain score. However, both of these were developed on small sample sets (n = 407 for Manchester score and n = 341 for Spain score) and divided patients into only three risk groups[73,74].

The goal of this section was to identify prognostic factors for SCLC patients, and then develop and validate a new nomogram prognostic model in a large SCLC patient cohort. Compared to the previously published models, our model has the following advantages: 1) it was validated in an independent set; 2) it was developed and validated with a much larger sample size; 3) it was developed across multiple facilities and facility types, which greatly diminishes sample selection bias; 4) it utilizes accurate SCLC staging criteria: the American Joint Committee on Cancer (AJCC) 8th edition TNM staging system proposed by IASLC[75,76]; and 5) it provides an online webserver so that clinicians can use the nomogram model easily.

3.2 METHODS

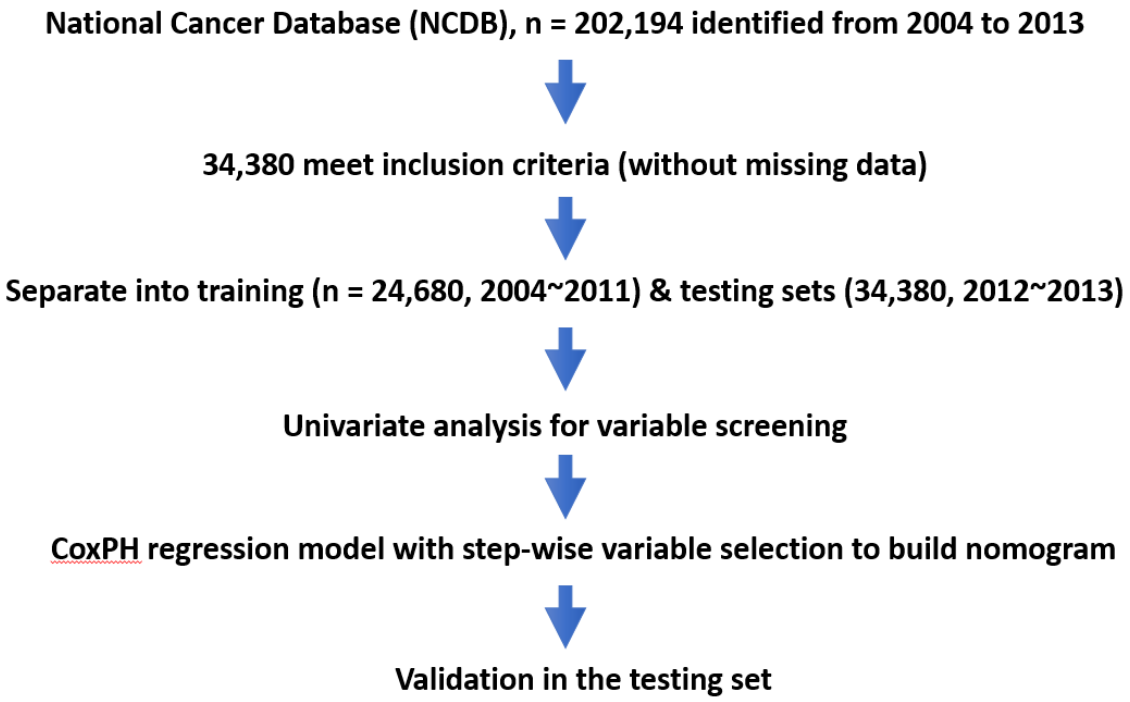


FIGURE 11 Flow chart of nomogram development. 3.2.1 Data Collection

NCDB SCLC

We identified 202,194 SCLC cases from National Cancer Database (NCDB); 34,380 of them met our inclusion criterion that they do not contain any missing data for the variables selected by univariable survival analysis (**FIGURE 11**). The cases are independent and recorded by annual reports from all the CoC-accredited programs from 2004 to 2013. 24,680 cases that were diagnosed from 2004 to 2011 were assigned to the training group and used to develop a nomogram prognostic model. The 9,700 cases diagnosed from 2012 to 2013 were assigned to the testing group and used to validate the model.

The variables collected were age, gender, race, Hispanic origin, Charlson/Deyo Score, sequence number, primary site, laterality, grade (tumor's resemblance to normal tissue), 8th edition TNM stage, and treatment type. Two extra variables were constructed based on the NCDB variables. 1) Treatment was defined as the stratification result of surgery, chemotherapy, and radiation therapy. 2) TNM stage was defined according to the coding guidelines of the Collaborative Staging Manual and Coding Instructions for the new 8th edition lung cancer staging system defined by the AJCC and the Union for International Cancer Control (UICC)[77-80], and followed Yang et al's method[81]. Stages IA1, IA2, and IA3 were combined together in our study as stage IA, since no significant prognostic differences were detected among the three sub-stages[76]. The patient characteristics were summarized in **TABLE 6**.

TABLE 6 Patient characteristics of the NCDB SCLC training and testing sets. P values were calculated by Chi-square test.

	Training set (%)	***Testing set (%)***	***p value***
No. of cases	24,680	9,700	
Year of diagnosis	2004-2011	2012-2013	
Age			0.09
< 65y	9,559 (38.7)	3,855 (39.7)	
≥ 65y	15,121 (61.3)	5,845 (60.3)	
Gender			0.9
Male	12,240 (49.6)	4,803 (49.5)	
Female	12,440 (50.4)	4,897 (50.5)	
Race			0.73
White	22,276 (90.3)	8,779 (90.5)	
Black	1,912 (7.7)	727 (7.5)	
Other	492 (2.0)	194 (2.0)	
Hispanic origin			0.91
Non-Hispanic	24,084 (97.6)	9,463 (97.6)	
Hispanic	596 (2.4)	237 (2.4)	
Charlson/Deyo score			<0.001
0	13,288 (53.8)	5,031 (51.9)	
1	7,629 (30.9)	3,061 (31.6)	
≥ 2	3,763 (15.2)	1,608 (16.6)	
*Sequence number**			0.82
0	24,084 (97.6)	9,463 (97.6)	

1	527 (2.1)	213 (2.2)	
≥ 2	69 (0.3)	24 (0.2)	
AJCC V8 TNM stage			<0.001
IA	1,207 (4.9)	160 (1.6)	
IB	463 (1.9)	74 (0.8)	
IIA	140 (0.6)	18 (0.2)	
IIB	853 (3.5)	97 (1.0)	
IIIA	1,548 (6.3)	156 (1.6)	
IIIB	902 (3.7)	89 (0.9)	
IIIC	208 (0.8)	27 (0.3)	
IVA	14,699 (59.6)	6,655 (68.6)	
IVB	4,660 (18.9)	2,424 (25.0)	
Treatment			<0.001
No surgery, no chemo, no radiation	5,025 (20.4)	2,213 (22.8)	
No surgery, no chemo, radiation done	1,230 (5.0)	520 (5.4)	
No surgery, chemo done, no radiation	7,668 (31.1)	3,473 (35.8)	
No surgery, chemo done, radiation done	7,901 (32)	3,050 (31.4)	
Surgery done, no chemo, no radiation	856 (3.5)	116 (1.2)	
Surgery done, no chemo, radiation done	64 (0.3)	8 (0.1)	
Surgery done, chemo done, no radiation	1,000 (4.1)	165 (1.7)	
Surgery done, chemo done, radiation done	936 (3.8)	155 (1.6)	
Primary site			<0.001
C340	2,298 (9.3)	911 (9.4)	
C341	11,019 (44.6)	4,152 (42.8)	
C342	968 (3.9)	368 (3.8)	
C343	4,959 (20.1)	1,923 (19.8)	
C348	485 (2.0)	200 (2.1)	
C349	4,951 (20.1)	2,146 (22.1)	
Laterality †			<0.001
Not a paired site	2,298 (9.3)	911 (9.4)	
Only one side involved	20,447 (82.8)	8,016 (82.6)	
Bilateral involvement	624 (2.5)	154 (1.6)	
Paired site but lateral origin unknown; midline tumor	1,311 (5.3)	619 (6.4)	
Grade			<0.001
Well differentiated	88 (0.4)	8 (0.1)	
Moderately differentiated	179 (0.7)	39 (0.4)	
Poorly differentiated	2,795 (11.3)	899 (9.3)	
Undifferentiated	5,037 (20.4)	1,457 (15)	
Cell type not determined, not stated or not applicable	16,581 (67.2)	7,297 (75.2)	

* Sequence number: 0 means the tumor diagnosis is the only one over the lifetime of the patient. If the patient has multiple tumor diagnoses, sequence number refers to the sequence of this diagnosis, with 1 refers to the 1st diagnosis.

† Laterality: the side of a paired organ (for lung cancer, the organ is lung) on which the primary tumor originated.

3.2.2 Prognostic Model Based on Clinical Characteristics

Nomogram Development

A nomogram was developed using the training cohort of 24,680 patients diagnosed from 2004 to 2011 in the NCDB SCLC dataset. Overall survival was defined as the length of time from diagnosis to death or last contact, and used as the primary outcome. Univariable Cox regression and Wald test were used to screen for variables that were significantly correlated with overall survival in the training group. Predictors with a p value less than 0.05 were fed to a multivariable Cox regression model. Backward stepwise selection based on Bayesian Information Criterion (BIC) was used to further eliminate redundant variables. The resulting multivariable Cox regression model was used to calculate risk score and build the final nomogram prognostic model. The assumptions were made here that the timing and sequence of the treatments were interchangeable, and none of these are salvage treatment due to recurrence/progression.

Model Validation

To validate our model, four criteria were used to evaluate prediction performance in the testing set. 1) The cases were grouped according to their predicted risk score, and K-M survival curves and Wald test were used to compare survival differences among the groups. 2) A concordance index (c-index) was calculated to estimate the similarity between the ranking of true survival time and of predicted risk score. The theoretical value of the c-index is between 0 and 1; a c-index larger than 0.5 indicates prediction performance better than

random guessing. When evaluating the performances of different models, c-indexes from different models were compared using z-test. 3) The AUC of time-dependent ROC[82,83] was calculated at each month from the 1st to the 30th month. Integrated AUC was calculated by averaging the AUC values. 4) Calibration curves were plotted to evaluate the consistency between predicted survival probability and actual survival proportion at 1 and 2 years, separately[84].

The other two models, the AJCC 8th edition TNM staging system and the traditional limited/extensive staging system, were also tested for prognostic performance in the testing group. C-index and integrated AUC were used to compare this nomogram with the two staging systems. Here, extensive stage was defined based on the presence of distant metastases (M1 stage)[85,86]. All other cases (M0 stage) were grouped as limited stage.

All computations in this section were conducted in the R environment, version 3.3.2[48]. R packages “survival” (version 2.40-1), “timeROC” (version 0.3), and “rms” (version 5.1-2) were used. Results with p value ≤ 0.05 were considered statistically significant.

3.3 RESULTS

3.3.1 Characteristics of the Training and Validation Cohorts

In total, 202,194 SCLC cases were identified in NCDB, among which, 34,380 cases that did not contain any missing variables were included in this study. Based on year of diagnosis, included cases were divided into two distinct groups: cases that were diagnosed from 2004 to 2011 (n = 24,680) were used as the training cohort, while cases that were diagnosed from

2012 to 2013 (n = 9,700) were used as the validation cohort. The follow-up time ranged from 0 to 10.76 years (median 0.64 year) for the training cohort and from 0 to 2.92 years (median 0.53 year) for the testing cohort. Characteristics of the two sets were shown in **TABLE 6**. In comparing the training and testing sets, the demographic variables were similar, while the clinical variables, including Charlson/Deyo score, 8th AJCC stage, and laterality, were significantly different.

3.3.2 Building Nomogram Prognostic Model in Training

In univariable analysis, age, gender, race, Hispanic origin, Charlson/Deyo score, TNM stage by AJCC 8th edition, treatment type, primary site, laterality, and grade were significantly associated with overall survival in the training group (**TABLE 7**). After stepwise selection to further remove potential redundancy, age, sex, race, ethnicity, Charlson/Deyo score, TNM stage by AJCC 8th edition, treatment type, and laterality were used in the final nomogram model (coefficients summarized in **TABLE 8**). The final risk score was calculated by adding up the score of each item using the nomogram depicted in **FIGURE 12A**. The TNM stage defined by the AJCC 8th edition showed the largest range of risk scores, followed by the treatment type and age. The predicted survival probability using the Cox regression model of risk scores was plotted in **FIGURE 12B**.

TABLE 7 Univariable survival analysis of clinical features in SCLC.

Variable	***HR (95% CI)***	***p value***
Age	1.02 (1.02-1.02)	< 0.001
Sex (Female vs. Male)	0.84 (0.83-0.85)	< 0.001
Race		
White	1 (reference)	-
Black	0.97 (0.95-0.99)	0.006
Other	0.94 (0.90-0.97)	0.001
Hispanic origin (Yes vs. No)	0.95 (0.92-0.99)	0.028

Charlson/Deyo score		
0	1 (reference)	-
1	1.22 (1.20-1.24)	< 0.001
≥ 2	1.59 (1.56-0.61)	< 0.001
Sequence number		
0	1 (reference)	-
1	1.00 (0.98-1.01)	0.82
≥ 2	1.01 (0.92-1.11)	0.83
AJCC V8 TNM stage		
IA	1 (reference)	-
IB	1.22 (1.07-1.39)	< 0.001
IIA	1.63 (1.34-1.98)	< 0.001
IIB	1.60 (1.45-1.78)	< 0.001
IIIA	2.12 (1.94-2.31)	< 0.001
IIIB	2.55 (2.32-2.81)	< 0.001
IIIC	3.26 (2.81-3.78)	< 0.001
IVA	5.25 (4.88-5.65)	< 0.001
IVB	7.04 (6.51-7.61)	< 0.001
Treatment		
No surgery, no chemo, no radiation	1 (reference)	-
No surgery, no chemo, radiation done	0.72 (0.70-0.74)	< 0.001
No surgery, chemo done, no radiation	0.46 (0.45-0.47)	< 0.001
No surgery, chemo done, radiation done	0.26 (0.25-0.26)	< 0.001
Surgery done, no chemo, no radiation	0.19 (0.18-0.20)	< 0.001
Surgery done, no chemo, radiation done	0.28 (0.24-0.33)	< 0.001
Surgery done, chemo done, no radiation	0.13 (0.13-0.14)	< 0.001
Surgery done, chemo done, radiation done	0.13 (0.12-0.14)	< 0.001
Primary site		
C340	1 (reference)	-
C341	0.89 (0.88-0.91)	< 0.001
C342	0.90 (0.87-0.92)	< 0.001
C343	0.96 (0.94-0.98)	< 0.001
C348	1.07 (1.03-1.11)	< 0.001
C349	1.13 (1.11-1.16)	< 0.001
Laterality		
Not a paired site	1 (reference)	-
Only one side involved	0.94 (0.93-0.96)	< 0.001
Bilateral involvement	1.47 (1.40-1.54)	< 0.001
Paired site but lateral origin unknown; midline tumor	1.18 (1.15-1.21)	< 0.001
Grade		
Well differentiated	1 (reference)	-
Moderately differentiated	0.99 (0.86-1.14)	0.86
Poorly differentiated	1.29 (1.15-1.46)	< 0.001
Undifferentiated	1.39 (1.23-1.56)	< 0.001
Cell type not determined, not stated or not applicable	1.44 (1.28-1.62)	< 0.001

TABLE 8 Variable HRs in the clinical-feature based risk prediction nomogram.

Variable	***HR (95% CI)***	***p value***
Age	1.01 (1.01-1.02)	< 0.001
Sex (Female vs. Male)	0.88 (0.85-0.90)	< 0.001
Race		
White	1 (reference)	-
Black	0.88 (0.84-0.92)	< 0.001
Other	0.89 (0.80-0.98)	0.02
Hispanic origin (Yes vs. No)	0.75 (0.68-0.82)	< 0.001
Charlson/Deyo score		
0	1 (reference)	-
1	1.18 (1.14-1.21)	< 0.001
>= 2	1.36 (1.31-1.41)	< 0.001
AJCC V8 TNM stage		
IA	1 (reference)	-
IB	1.17 (1.02-1.35)	0.02
IIA	1.49 (1.20-1.84)	< 0.001
IIB	1.70 (1.52-1.90)	< 0.001
IIIA	2.04 (1.83-2.26)	< 0.001
IIIB	2.38 (2.11-2.68)	< 0.001
IIIC	2.97 (2.50-3.54)	< 0.001
IVA	3.86 (3.48-4.27)	< 0.001
IVB	5.62 (5.06-6.24)	< 0.001
Treatment		
No surgery, no chemo, no radiation	1 (reference)	-
No surgery, no chemo, radiation done	0.67 (0.63-0.71)	< 0.001
No surgery, chemo done, no radiation	0.35 (0.33-0.36)	< 0.001
No surgery, chemo done, radiation done	0.25 (0.24-0.26)	< 0.001
Surgery done, no chemo, no radiation	0.31 (0.28-0.35)	< 0.001
Surgery done, no chemo, radiation done	0.35 (0.27-0.46)	< 0.001
Surgery done, chemo done, no radiation	0.21 (0.19-0.23)	< 0.001
Surgery done, chemo done, radiation done	0.18 (0.17-0.20)	< 0.001
Laterality		
Not a paired site	1 (reference)	-
Only one side involved	0.95 (0.91-0.99)	0.02
Bilateral involvement	0.72 (0.66-0.79)	< 0.001
Paired site but lateral origin unknown; midline tumor	1.05 (0.98-1.13)	0.19

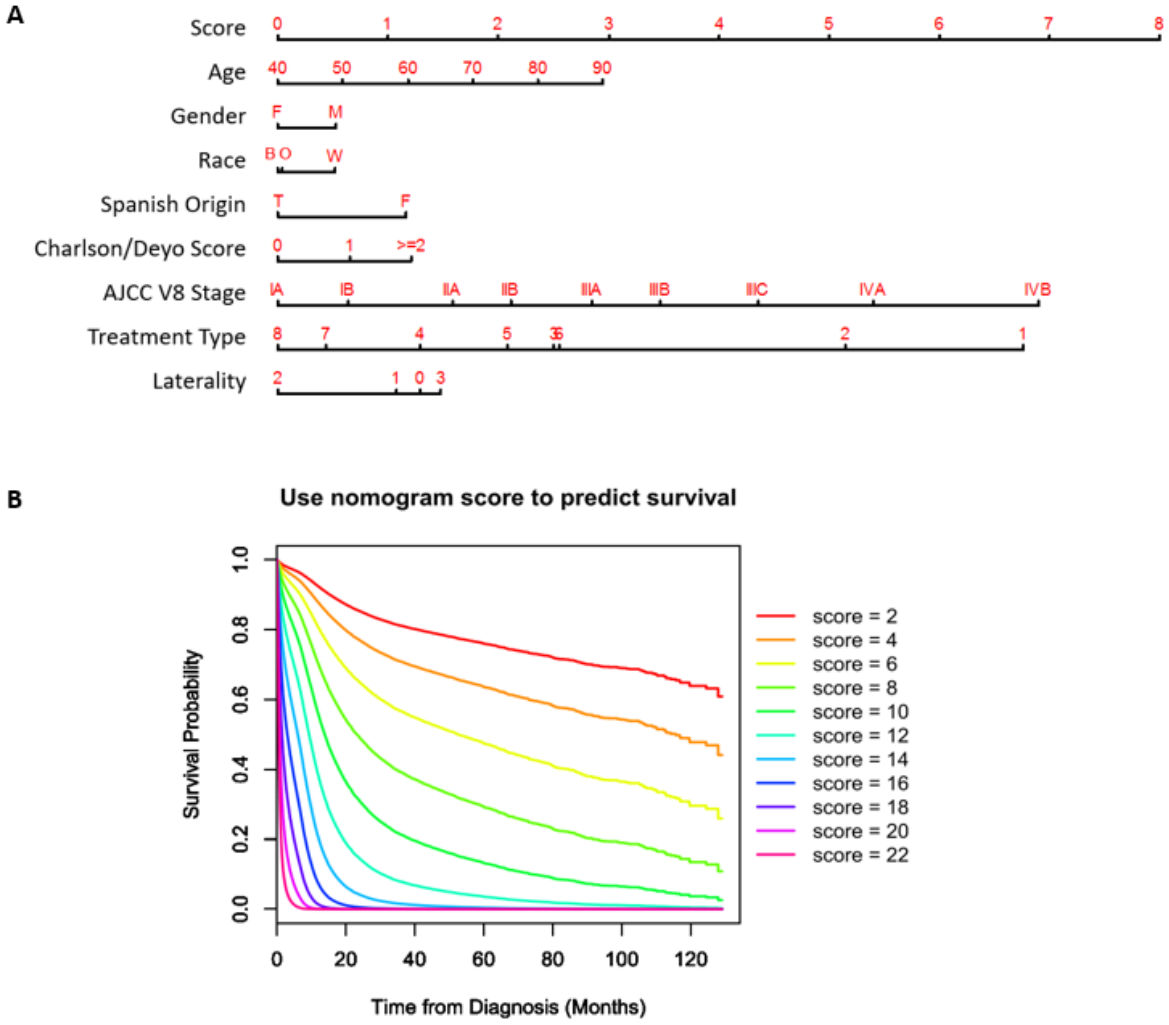


FIGURE 12 Nomogram to calculate risk score and survival probability for SCLC patients. (A) Race includes black (B), white (W) and other (O). Treatment types include: no surgery, no chemo, no radiation (1); no surgery, no chemo, radiation done (2); no surgery, chemo done, no radiation (3); no surgery, chemo done, radiation done (4); surgery done, no chemo, no radiation (5); surgery done, no chemo, radiation done (6); surgery done, chemo done, no radiation (7); and surgery done, chemo done, radiation done (8). Laterality of tumor origin includes: not a paired site (0), only one side (either left or right) is involved (1), bilateral involvement (2), paired site with unknown origin side or midline tumor (3). **(B)** Predicted survival probability curve corresponding to risk scores ranging from 2 to 22.

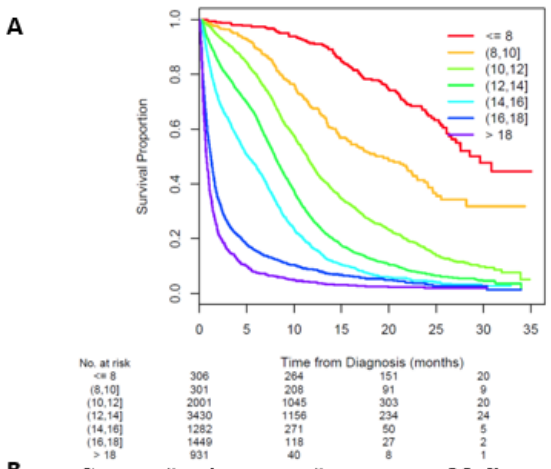
3.3.3 Validation and Sensitivity Analysis

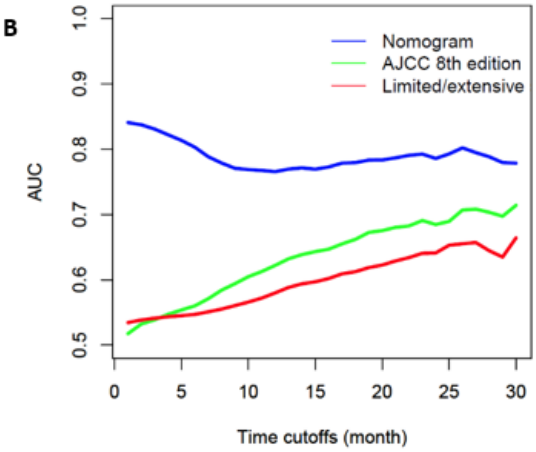
The proposed nomogram was validated in the independent testing set (n=9,700). The survival difference between any two adjacent groups, which were grouped by predicted risk score, was significant (p value < 0.05, **FIGURE 13A & B**). The median survival times of score groups ranged from 0.7 months (when risk score > 18) to 30.9 months (when risk score < 6). The c-index was 0.722 ± 0.004 and the integrated AUC was 0.79 from the 1st month to the 30th month (**FIGURE 13C**). A calibration curve at 1 year (**FIGURE 13D**) or 2 years (**FIGURE 13E**) also showed high consistency between predicted survival probability and actual survival proportion.

With regard to prognostic ability, the proposed nomogram performed better than the two commonly used SCLC staging systems, the AJCC TNM system and limited/extensive staging system (**FIGURE 13C**). The AUC of the nomogram was the highest throughout the 1st to the 30th month, followed by the 8th edition TNM staging system. The integrated AUC of the proposed nomogram was 0.789, while those of the 8th edition TNM staging system and the limited/extensive staging system were 0.634 and 0.598, respectively. The c-index of this nomogram (0.722 ± 0.004) was also significantly higher than the c-indexes of the 8th edition TNM staging system (0.550 ± 0.003) and the limited/extensive staging system (0.539 ± 0.002), confirming the strong prognostic power of this proposed nomogram.

To evaluate the robustness of our model to missing data, a sensitivity analysis was performed on the excluded cases diagnosed from the year 2012 to 2013 (n = 11,020). The missed variables were imputed using corresponding modes in the training cohort (**TABLE 6**): missed stages (n = 10,416) were imputed as “stage IVA”; missed treatment types (n = 508)

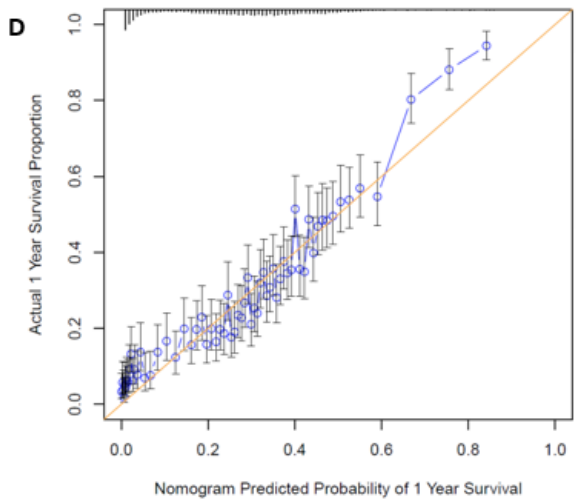
were imputed as “No Surgery, Chemo Done, Radiation Done”; missed Hispanic origins (n = 819) were imputed as “False”. Under the circumstance of having at least one variable imputed, the survival difference between any two adjacent predicted risk groups was still significant. The c-index was 0.691 ± 0.004, and the integrated AUC was 0.734. A calibration curve at 1 year or 2 years still showed high consistency between predicted survival probability and actual survival proportion, proving the robustness of this nomogram to missing data.





B

Score	*#patients*	*#events*	*Median survival (mos)*	*HR*	*p-value*
≤ 8	306	98	29.63 (26.64-NA)	-	-
(8,10]	301	156	19.42 (15.67-22.54)	(8,10] vs ≤ 8: 2.3 (1.78-2.96)	< 0.001
(10,12]	2001	1512	11.4 (10.94-11.73)	(10,12] vs (8,10]: 2 (1.7-2.36)	< 0.001
(12,14]	3430	2929	7.98 (7.75-8.21)	(12,14] vs (10,12]: 1.63 (1.53-1.74)	< 0.001
(14,16]	1282	1144	5.16 (4.67-5.85)	(14,16] vs (12,14]: 1.44 (1.35-1.55)	< 0.001
(16,18]	1449	1273	1.22 (1.12-1.31)	(16,18] vs (14,16]: 1.93 (1.78-2.1)	< 0.001
> 18	931	867	0.72 (0.66-0.82)	> 18 vs (16,18]: 1.37 (1.25-1.49)	< 0.001



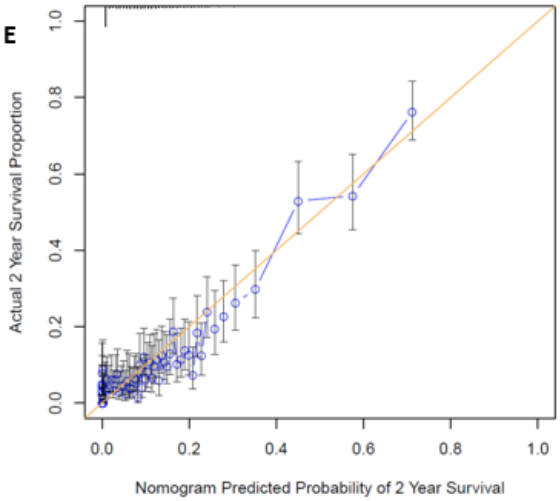


FIGURE 13 Validation of proposed nomogram prognostic model in the testing set. (A) Risk scores of testing set cases were calculated according to the model in Figure 1 and grouped into 8 subgroups. K-M plot was depicted for each group. **(B)** Summary of groups in **(A)**. HR was calculated using Coxph regression model between each two adjacent lines. P-value was calculated using Wald test. **(C)** AUC was calculated for three prognostic models for every month from the 1st to the 30th month. Blue: nomogram developed in this study; green: AJCC 8th TNM staging system; red: limited/extensive staging system. **(D, E)** Calibration curves compare predicted and actual survival proportions at 1 year **(D)** and 2 years **(E)**, separately. Distributions of predicted survival probabilities are plotted at the top. Error bars represent 95% CIs.

3.3.4 Webserver for Easy Access of Our Own and Previously Published Models

An online version of our nomogram can be accessed at http://lce.biohpc.swmed.edu/lungcancer/sclc_nomogram, to assist researchers and clinicians. Online implementation of the other nomograms from Pan et al, Xiao et al, and Xie et al are also available (**FIGURE 14B-D**). Predicted survival probability across time can be easily determined by inputting clinical features and reading output figures and tables generated by the webserver.

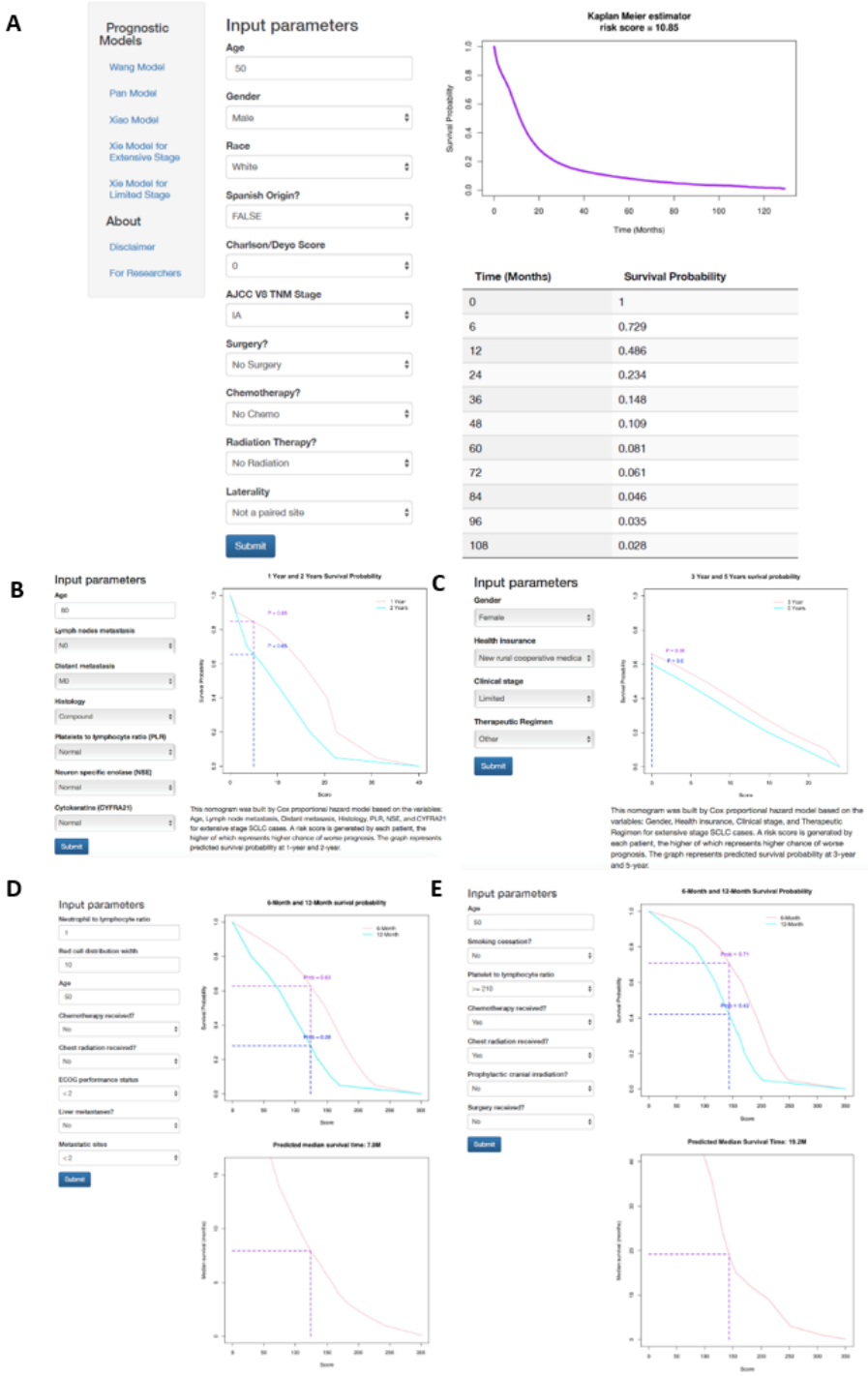


FIGURE 14 Illustration of performing prognostic nomogram online.(A) The newly developed nomogram in this study (Wang model). **(B-D)** Published nomograms by Pan et al **(B)**, Xiao et al **(C)**, and Xie et al (**D**: Extensive Stage; **E**: Limited Stage).

3.4 DISCUSSION

In this study, a prognostic nomogram was developed and validated using a large cohort of SCLC cases across the United States. This nomogram, based on routinely available demographic, staging and treatment information, predicts the survival probability for individual SCLC patients. The publicly accessible online implementation will assist clinicians in making treatment decisions.

Compared with other prognostic indexes, such as the Manchester Score[74] and the Spain prognostic index[73], our model calculates individualized survival probability rather than assigning cases into a few risk groups, thus better capturing heterogeneity across patients. Compared with the previously published nomogram by Xie et al., this model used a much larger training dataset and involved multiple treatment facilities, which allowed for smaller sampling bias. The internal c-index of this model was 0.744 ± 0.002, higher than in previously published models (0.73 for both nomograms in [67]). Independent validation of our model showed significantly different outcomes among different score groups (**FIGURE 13A&B**). A high concordance index (0.722 ± 0.004) and integrated AUC score (0.789, **FIGURE 13C**) in the testing set also indicated the strong predictive ability of our nomogram model. In addition, combining demographic, clinical and treatment information together produced a nomogram with better performance than using staging information alone (**FIGURE 13B**). Thus, this comprehensive and individualized risk score calculation method could be used as stratification criteria in randomized studies and clinical trials.

In this nomogram, age, gender, race, ethnicity, Charlson/Deyo score, AJCC 8th edition stage, treatment type and laterality were kept after univariable Cox regression screening and backward stepwise selection. Age, gender, and Charlson/Deyo score have previously been shown significantly relevant to survival of SCLC patients[67,87]. Noticeably, AJCC 8th edition stage contributed the most to the final risk score (**FIGURE 12A**), with clear distinctions between each two adjacent TNM stages (**TABLE 8**), and showed better prognostic performance than the limited/extensive staging system with higher c-index and AUC (**FIGURE 13B**). The significant contribution of TNM stage to this nomogram externally validates the performance of the 8th edition TNM lung cancer classification system, and highlights the importance of applying this more accurate staging system to SCLC rather than using the traditional limited/extended staging[72,75,88].

This proposed nomogram also illustrates the prognostic implications of using different treatment methods (**FIGURE 12A**, **TABLE 8**). As expected, cases treated with both surgery and chemo-radiation therapy have the lowest risk score and cases not treated with any method have the highest risk score. Furthermore, the nomogram (**FIGURE 12B**) is consistent with current research in that it predicts better survival for surgery with chemo-radiation (treatment type 7 and 8) than for surgery with chemotherapy alone (type 3 and 4). However, the risk scores of different treatment methods are not recommended for direct use as a guideline for treatment selection, since clinical treatment decisions should be made based on multiple factors such as TNM stage and patient comorbidities[66].

There were several limitations in the development of this nomogram. The first limitation was a lack of some routinely available clinical data, such as the neutrophil to lymphocyte ratio (NLR) and platelet to lymphocyte ratio (PLR). The absence of this information prevented direct comparison of performance between our model and another published nomogram[67]. Constructing a prognostic model using both the factors identified in our model and other lab tests such as NLR would thus be beneficial in creating an even more accurate prognostic prediction. The second limitation was the inability to capture interaction terms among the predictors. For example, patients with early stage disease (stage I & II) were more likely to receive surgery than patients with late stage disease (stage III and IV). The interactions between stage and treatment strategies are worth further investigation. To satisfy the requirement for convenience and interpretability of the nomogram, interaction terms were not considered in this model. However, a more complex model considering all potential interaction terms would be expected to have better prognostic performance. The third limitation was that the sequence of treatment was not considered. Since neither recurrence nor progression is recorded in the dataset, we have to consider the treatment as baseline variables instead of time-varying covariates. Finally, out of 200,000 SCLC patients from the NCDB, there are only 34,380 patients without missing values. This large percent of missing data might introduce some selection bias.

CHAPTER FOUR – FACILITY INFLUENCES SURVIVAL AND SURGERY SELECTION

4.1 BACKGROUND AND RATIONALE

Lung cancer screening trials, such as the International Early Lung Cancer Action Program (I-ELCAP)[89] and the NLST[90], have shown great benefits in early-stage disease detection and 10-year survival rate improvement. The implementation of low-dose computed tomography (CT) screening is expected to increase the incidence of diagnosed early-stage NSCLC[91]. Other recent developments in treatment advances for early-stage NSCLC include stereotactic body radiotherapy (SBRT)[92] and robotic/minimally invasive surgery[93].

Early-stage NSCLC is a potentially curable condition, so treatment selection especially receipt of potentially curative surgery is a well-known determinant of survival, which has been included as the standard care for eligible early stage NSCLC into the European Society for Medical Oncology (ESMO) guidelines[94]. How teaching facility (TF) status affects care quality has been widely investigated, showing a positive relationship between TF status and patient outcomes[14-17]. In reported studies of facility type and surgical lung cancer treatment[15,17,95,96], only some studies report overall survival outcome, while others report resection outcomes. The limited population size in previous studies meant that stratification of patients by early and late stage has not been performed. Moreover, the effect of non-TF type on survival outcome has not yet been reported.

The relationship between facility volume and care quality has also been investigated for over 40 years, with mixed results[18]. For cancer treatment, most studies found that higher volume was associated with lower operative morality rate and better survival outcomes for patients[15,19,20], while others found no difference[21,22]. One possible reason is that estimation of low-volume hospitals' performance would be unstable because of small sample size[97]. Also, it is still unknown whether this association persists in all types of facilities, and if so, what the underlying reasons are.

To investigate these knowledge gaps, this study aimed to evaluate the influence of facility type and volume on surgery selection and survival outcome in early stage NSCLC. Although facility volume-outcome and TF status-outcome have been widely investigated independently, to our knowledge, this was the first study to stratify the volume effect by facility type.

4.2 METHODS

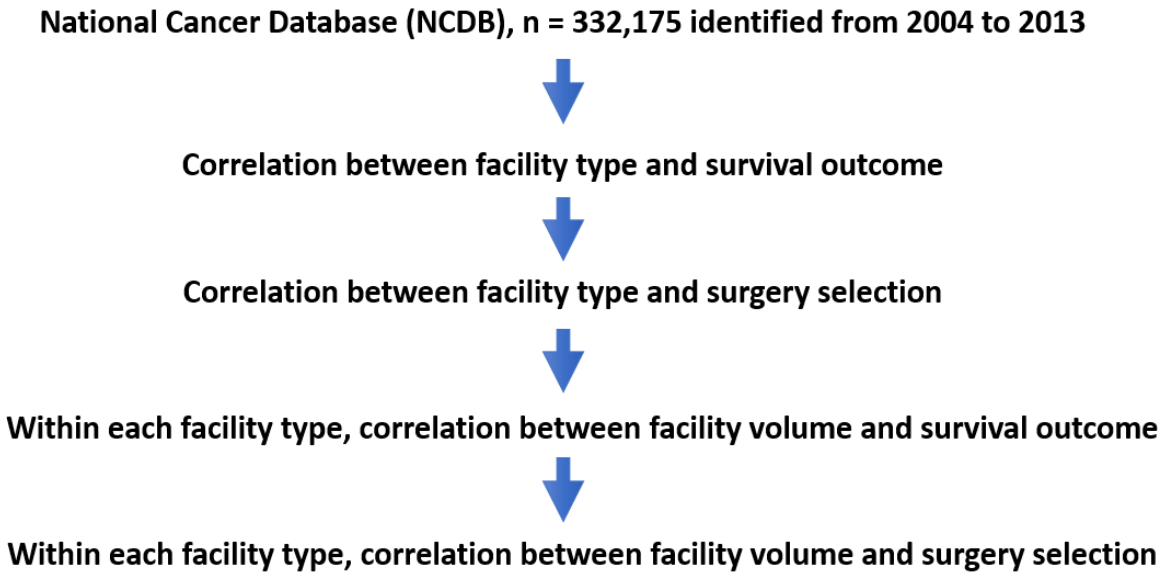


FIGURE 15 Flow chart of investigating how facility type and volume affect patient survival outcome.

4.2.1 Data Collection

332,175 early stage (stage I or II) NSCLC patients were identified from the NCDB. The cases are independent and recorded by annual reports from all the CoC Accreditation Programs from 2004 to 2013 (Error! Reference source not found.). 1,299 facilities were involved and had been categorized by the CoC Accreditation Program into 4 types of program as follows: Community Cancer Program (CCP), which accessions 100 to 500 newly diagnosed cancer cases per year; Comprehensive Community Cancer Program (CCCP), which accessions 500 or more cases per year; Academic/Research Program (ARP), which participates in cancer-related clinical research and mandates postgraduate education, including residency training; Integrated Network Cancer Program (INCP), defined by American College of Surgeons website (https://www.facs.org/), as "an organization that owns, operates, leases, or is part of a joint venture with multiple facilities providing

integrated cancer care and comprehensive services". In this study, facility volumes were determined by counting the number of NSCLC patients reported annually according to the NCDB. Median volume for all facilities was determined and used to categorize each facility as high- or low-volume to study the joint effects of facility type and volume. To separate the effect of patient volume from facility type, median facility volume for each facility type was determined and used to assign each facility as high or low volume within each facility type.

In addition to facility type information, other demographic and clinical variables collected included: age at diagnosis, gender, race, insurance status, comorbidity, treatment, and resident zip code-level characteristics (median household income collected from 2008 to 2012, proportion without high school diploma collected from 2008 to 2012, and urban/rural area collected in 2013). Comorbidity was represented by Charlson/Deyo score, where 0 means no comorbid conditions. Treatment information, including surgery, chemotherapy, and radiation therapy, was collected. Treatment information was stratified into four groups: 1) surgery performed; 2) no surgery but radiation therapy performed; 3) no surgery or radiation therapy but chemotherapy performed; 4) no treatment received.

4.2.2 Statistical Analysis

To study the surgery selection, a multivariable logistic regression model adjusted by potential confounders was used to calculate the odds ratio (OR) of surgery selection among different facility types or volumes. Wald tests were used to check if OR = 1. In survival analysis, overall survival was defined as time from diagnosis to death from any reason or last contact. K-M survival curves were used to visualize overall survival; Cox regression models and

Wald tests were used to compare survival difference among different facility types or volumes in both univariable and multivariable analysis adjusted for potential confounders. To further rule out the effect of potential confounders, propensity score matching was used to weight and balance patient groups with different clinical characteristics[98]. Propensity score estimates the conditional probability of selecting a certain treatment condition given all covariates that may affect this selection; thus, weighting patients under different conditions to balance propensity scores helps infer the relationship between treatment condition and outcome independent of other covariates. All variables listed in section 4.2.1 were considered in propensity score matching. All p values were two-sided; results were considered significant at p value ≤ 0.05. All analyses were performed with R software, version 3.4.2[99]. R packages "survival" (version 2.41-3) and "twang" (version 1.5) were used.

4.3 RESULTS

4.3.1 Patient Characteristics Differed across Facility Types

In total, 332,175 early stage NSCLC patients reported by 1,299 facilities were included in our analysis. Compared with other facility types, NSCLC patients reported by ARP were younger, more likely female, more likely non-white, had lower Charlson/Deyo comorbidity scores, and more likely to have private insurance (Error! Reference source not found.). Patients reported by INCP were more likely to come from higher income and higher education areas, and metropolitan counties.

TABLE 9 Characteristics of 1,299 facilities studied in NCDB NSCLC. Chi-square test is used to calculate p values.

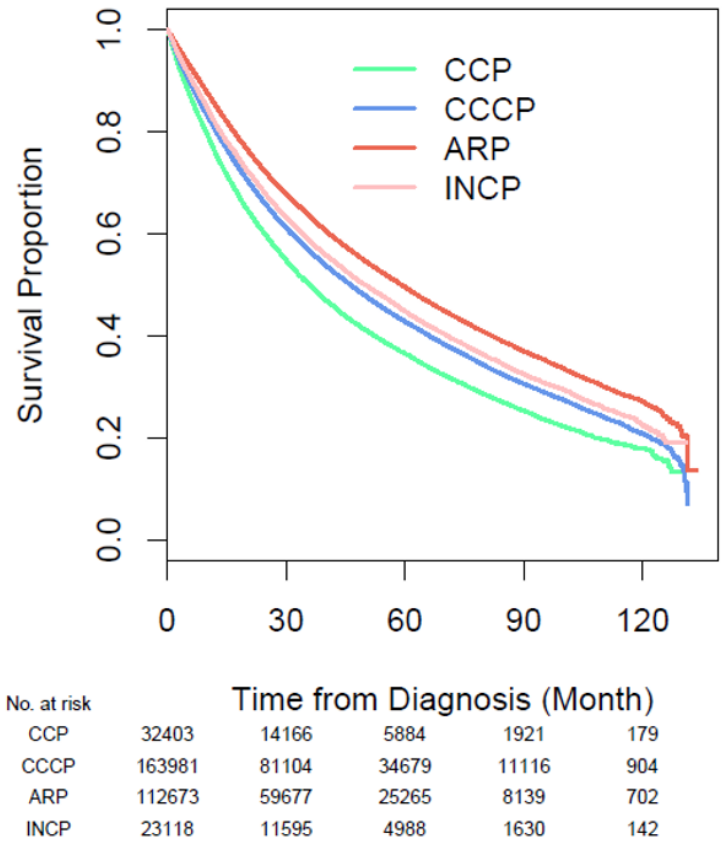
	All	CCP	CCCP	ARP	INCP	p value
No. of hospitals	1,299	433	598	232	36	

No. of patients	332,175	32,403	163,981	112,673	23,118	
Age (%)						<0.001
<65 yr	29.2	28.5	27.2	32.5	27.8	
>= 65 yr	70.8	71.5	72.8	67.5	72.2	
Sex (%)						<0.001
Male	49.6	51.7	50	48.5	48.7	
Female	50.4	48.3	50	51.5	51.3	
Race (%)						<0.001
White	88.4	89.2	91.2	84.2	87.1	
Black	8.6	8.1	6.5	11.7	9.5	
Other	3.0	2.7	2.3	4.1	3.3	
Stage of disease (%)						<0.001
I	76.5	72	76.1	78.3	77	
II	23.5	28	23.9	21.7	23	
Charlson/Deyo score (%)						<0.001
0	53.7	52.5	51.8	57.4	50.2	
1	32.6	32.7	33.7	30.4	34.8	
>= 2	13.7	14.8	14.4	12.2	15	
Insurance status (%)						<0.001
Not insured	1.6	1.7	1.4	1.9	1.7	
Private insurance	26.1	21.9	25.5	28.4	24.8	
Medicaid	3.9	4.9	3.3	4.6	3.8	
Medicare	65.3	68.6	67.5	60.6	68.4	
Other government	1.2	1	1	1.5	0.7	
Unknown	1.9	1.8	1.2	3.1	0.6	
Surgery rate (%)						< 0.001
Received	69.9	60.8	68.4	74.8	69.8	
Not received	30.1	39.2	31.6	25.2	30.2	
Resident zip code-level characteristics						
Median household income, USD (%)						<0.001
< $38,000	19.4	20.8	19.6	19.4	15.9	
$38,000 - $47,999	25.4	31.7	26.6	22.3	23.5	
$48,000 - $62,999	26.8	26.7	27.8	24.9	30	
≥ $63,000	28.3	20.9	26	33.4	30.6	
Proportion without high school diploma (%)						<0.001
≥ 21%	17.0	19.4	16.7	17.3	14.6	
13% - 20.9%	27.9	31.9	28.6	26.2	25.1	
7% - 12.9%	33.6	35.3	33.9	32.1	36.2	
< 7%	21.5	13.4	20.8	24.4	24.2	
Urban/Rural (%)						<0.001
Metropolitan counties	81.7	70.1	79.9	85.8	90.6	
Urban counties	16.1	26.7	17.3	12.8	8.7	
Rural counties	2.2	3.2	2.8	1.4	0.7	

4.3.2 Facility Type is Associated with Patients' Survival Outcomes

Among different categories of facilities, ARP had the best survival outcome with median survival time of 59.1 months, followed by INCP (49.9 months), CCCP (46.3 months), and

CCP (36.0 months) **(FIGURE 16)**. To further investigate whether the facility type was an independent factor in survival outcome, a multivariable Cox regression was performed to study the association between facility type and patient survival adjusted by treatment selection, age, gender, race, stage, Charlson/Deyo score, insurance status, income, education, and urban/rural (**TABLE 10**). Propensity score matching through these confounders also showed similar results as using multivariable Cox regression model as shown in **TABLE 10**. The significant differences in patient survival persisted among all different facility types, indicating that facility type was an independent factor affecting lung cancer prognosis.



Facility type	#patients	Median Survival (month)	Hazard Ratio (95% CI)	p-value
CCP	32,403	36.0	Reference	
CCCP	163,981	46.3	0.83 (0.82-0.85)	< 0.001
ARP	112,673	59.1	0.68 (0.66-0.69)	< 0.001
INCP	23,118	49.9	0.78 (0.76-0.8)	< 0.001

FIGURE 16 K-M plot of patients from different facility types. ARP showed the lowest HR and the longest median survival time.

TABLE 10 HR of NSCLC patients in different facility types in multivariable Cox regression.

	HR (95% CI)	p-value		HR (95% CI)	p-value
Individual-level characteristics			Stage		
Facility type			Stage I	reference	
CCP	reference		Stage II	1.60 (1.58-1.62)	< 0.001
CCCP	0.94 (0.92-0.95)	< 0.001	Treatment approach		
ARP	0.86 (0.84-0.87)	< 0.001	Surgery received	reference	
ICNP	0.91 (0.88-0.93)	< 0.001	No surgery; RT received	2.48 (2.45-2.51)	< 0.001

Age at diagnosis		
< 65 yr	reference	
≥ 65 yr	1.40 (1.37-1.42)	< 0.001
Gender		
Male	reference	
Female	0.74 (0.73-0.75)	< 0.001
Race		
Black	reference	
Other	0.86 (0.83-0.89)	< 0.001
White	1.05 (1.03-1.07)	< 0.001
Insurance status		
Not insured	reference	
Private insurance	0.81 (0.78-0.85)	< 0.001
Medicaid	1.08 (1.03-1.13)	< 0.001
Medicare	0.99 (0.95-1.04)	0.73
Other government	0.94 (0.88-1.00)	0.05
Unknown	0.92 (0.87-0.97)	< 0.001
Charlson/Deyo score		
0	reference	
1	1.17 (1.15-1.18)	< 0.001
≥ 2	1.44 (1.42-1.46)	< 0.001
No surgery or RT; CT received	3.67 (3.57-3.77)	< 0.001
No treatment received	4.43 (4.36-4.50)	< 0.001
Resident zip code-level characteristics		
Median household income, USD		
< $38,000	reference	
$38,000 - $47,999	0.97 (0.87-0.97)	< 0.001
$48,000 - $62,999	0.94 (0.92-0.95)	< 0.001
≥ $63,000	0.87 (0.85-0.89)	< 0.001
Proportion without high school diploma		
≥ 21%	reference	
13% - 20.9%	1.00 (0.99-1.02)	0.76
7% - 12.9%	1.00 (0.99-1.02)	0.79
< 7%	0.97 (0.95-0.99)	< 0.001
Urban/Rural		
Metropolitan counties	reference	
Urban counties	1.01 (1.00-1.02)	0.15
Rural counties	1.04 (1.00-1.07)	0.03

4.3.3 Facility type is Associated with Surgery Selection

To determine if different facility types had a different tendency to perform surgery, multivariable logistic regression was used to calculate the OR of surgery selection, adjusted by all other available demographic and clinical variables, including stage (**TABLE 11**). The likelihood of performing surgery was significantly different between different facility types. Specifically, the ARP facilities were the most likely to treat patients with surgery (OR = 1.81, ARP vs. CCP), followed by INCP (OR = 1.44, INCP vs. CCP), CCCP (OR = 1.36, CCCP vs. CCP) and CCP (used as reference). Interestingly, this was the same order of facility types, from best to worst, when analyzed for patient survival. Propensity score matching across the four facility types showed similar results and support this finding.

TABLE 11 Adjusted OR of performing surgery for each facility type.

Facility types	OR (95% CI)*	p value
CCP	1 (Reference)	
CCCP	1.36 (1.33-1.40)	< 0.001
ARP	1.81 (1.76-1.87)	< 0.001
INCP	1.44 (1.39-1.50)	< 0.001

4.3.4 Joint Effects of Facility Type and Patient Volume on Patient Survival Outcome

The distribution of facility volume is shown in **FIGURE 17**. When further stratified by the median annual volume of all facilities (17.9 cases/year), patient survival outcomes and facility types demonstrated a similar trend, but within a facility type, high-volume facilities had better survival than low-volume facilities (**FIGURE 18**). Thus, we further investigated the effects of facility type as a primary factor and facility volume as a secondary factor on the surgery selection and patient outcome in the following analysis.

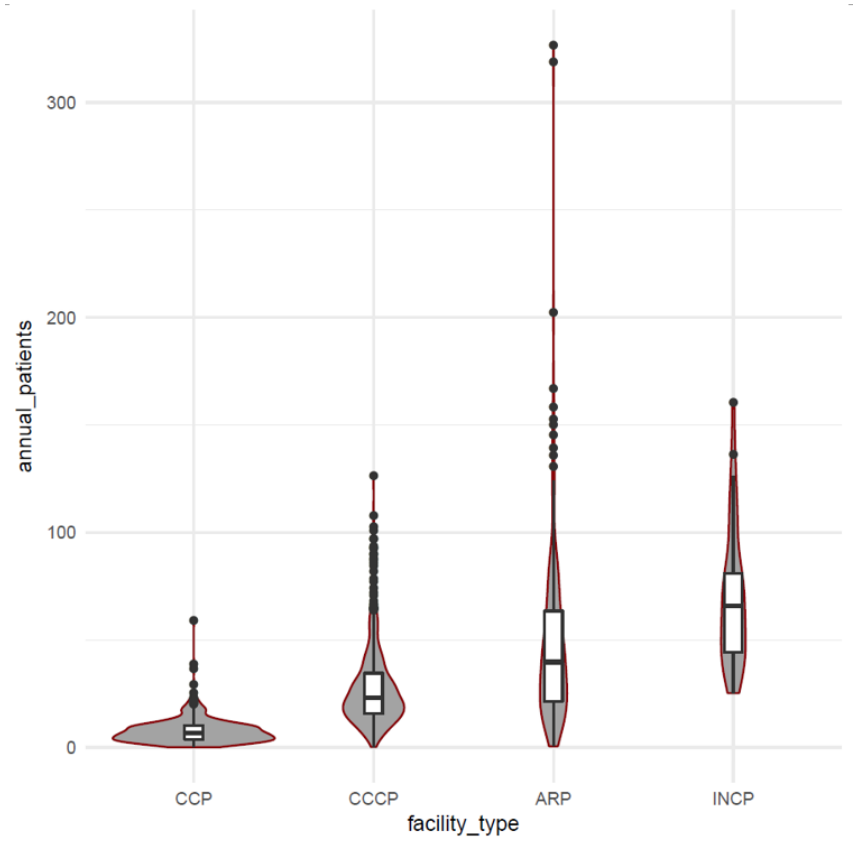
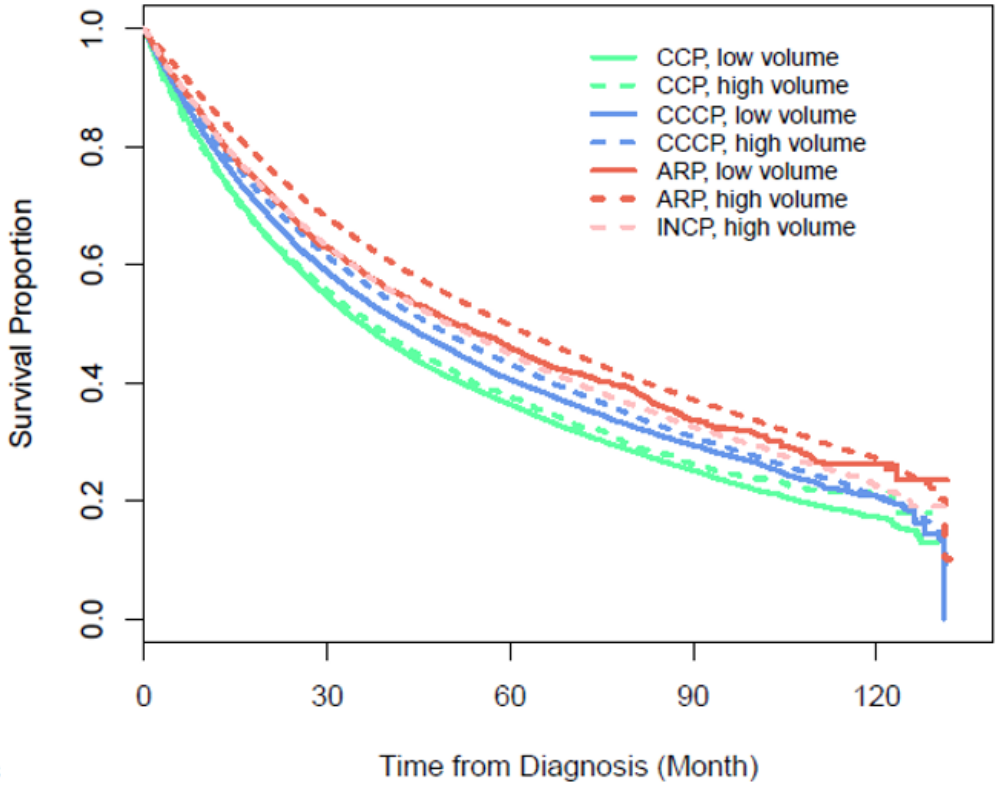


FIGURE 17 Violin plot of annual patients for each facility grouped by different facility types. Distribution of facility volumes was shown with density plot. Median, 1st quantile, and 3rd quantile of facility volumes were shown with box plot for each facility type.



No. at risk					
CCP, low volume	27171	11854	4879	1595	148
CCP, high volume	5232	2312	1005	326	31
CCCP, low volume	22912	10682	4460	1449	131
CCCP, high volume	141069	70422	30219	9667	773
ARP, low volume	5138	2391	943	273	31
ARP, high volume	107535	57286	24322	7866	671
INCP, high volume	23118	11595	4988	1630	142

Facility type	No. of facilities	No. of patients	Median Survival (month)	Hazard Ratio (95% CI)	p-value
CCP-low	399	27,171	35.6	1 (reference)	
CCP-high	34	5,232	37.5	0.97 (0.93-1.01)	0.11
CCCP-low	190	22,912	42.5	0.88 (0.86-0.9)	< 0.001
CCCP-high	408	141,069	46.9	0.82 (0.81-0.83)	< 0.001
ARP-low	45	5,138	51.1	0.76 (0.73-0.79)	< 0.001
ARP-high	187	107,535	59.4	0.67 (0.66-0.68)	< 0.001
INCP-high	36	23,118	49.9	0.78 (0.76-0.8)	< 0.001

FIGURE 18 K-M plot of patients from different facility types and volumes. High volume was defined as more than 17.9 stage I/II cases annually, which was the median volume across all facilities. No data of INCP low-volume facilities was shown because volumes of all INCP facilities were more than 17.9 stage I/II cases annually.

4.3.5 The Effects of Patient Volume within the Same Facility Type

As a secondary factor, the effect of facility volume on overall survival was then investigated individually for each facility type. Within each facility type, the facilities were dichotomized into high- and low-volume groups using the median facility volume as a cutoff. Consistent with previous reports, for CCP, CCCP, and ARP, the high-volume group showed better survival outcome than low volume group in univariable analysis, while no significant survival difference was detected for INCP. However, multivariable analysis showed that only in ARP and CCCP was higher volume found to be an independent factor associated with better survival outcome. In INCP, no difference was found between high- and low-volume facilities. Surprisingly, in CCP, high-volume facilities had modestly worse survival outcomes than low-volume facilities (HR = 1.05, 95% CI 1.02-1.10). Propensity score matching also showed similar HR results, which led to the same conclusion. To investigate the differences in surgery selection between high- and low-volume facility groups, multivariable logistic regression was used to calculate the OR of surgery selection. Interestingly, for each facility type, high-volume facilities were more likely to select surgery than low-volume facilities. Propensity score matching results closely followed the trends.

4.4 DISCUSSION

This study examined a large cohort of 332,175 early stage (stage I and II) NSCLC patients from the NCDB to study the relationship between facility type, surgery selection, and patient outcomes. The large sample size and multi-facility data collection greatly improved the statistical power and generalizability of this study. Although facility volume-outcome and TF

status-outcome had been widely investigated independently, as far as we are aware, this is the first study to stratify the volume effect by facility types, including TF status.

ARP showed the best overall survival, consistent with previous reports[14,15,95]. To our knowledge, this is the first study to identify a gradient in adjusted, long term survival with the best outcomes among patients treated in ARP sites followed by those in INCP, CCCP and CCP (with the worst outcomes). This ranking persisted after multivariable adjustment by age, gender, race, treatment, and other socioeconomic status information, indicating facility type is an independent predictor of survival outcomes for NSCLC patients.

To understand the factors accounting for the survival difference among different facilities, it is noteworthy that facilities with improved overall survival were more likely to perform surgery. Since surgery is the preferred treatment modality for resectable patients[94,100], the correlation between surgery rate and survival outcomes rank strongly suggests that survival outcomes among facility types largely result from selected treatment modality. This result should alert hospital facilities to ensure the most appropriate treatment modality is chosen when evaluating patients, and surgery should be selected when clinically appropriate. The correlation found between surgery rate and survival outcomes also invokes an urgent need for researchers to determine the specific reasons why surgery is performed more frequently at ARP and if there are any other factors that explain outcome differences between facility types, which requires future analysis. A possible explanation for these findings is that high volume and ARP facilities may have greater availability of cardiothoracic surgeons experienced in lung cancer surgery, and are therefore more likely to offer surgical treatment

modalities to early-stage NSCLC patients. This result could also arise from improper staging in community facilities, which have higher modality rate, leading to selection of inappropriate treatment modality. In any case, the survival analysis combined with surgery selection tendency verified the existence of care quality differences between facility types.

In our multivariable analysis, the effect of volume was modest and smaller than the effect of facility type. Interestingly, for CCP facilities, high-volume facilities performed even worse than low-volume ones after adjusted for other confounders, which is inconsistent with previous studies[15,95]. Adjustment by clinical and demographic confounders and using a nationwide database largely reduced potential bias due to patients' own characteristics, which might explain this inconsistency with other reports. However, it was still possible that other quality measures, such as surgical mortality ratio, were improved in high-volume CCP facilities, which requires further study.

Our results support the idea that for cancer patients or health plans, selecting hospitals with ARP facility type performing a high volume number of lung surgeries would be reasonable when no other quality measures are available[101]. However, such a recommendation should be approached carefully in context of a few limitations. First, cost was not considered in our research as cost information was not available from NCDB database. Since it has been reported that the average cost was 60% higher in teaching hospitals but lower in high-volume facilities[102,103], it is an important consideration in addition to facility type and volume when considering healthcare outcomes. Second, the survival outcome was not in favor of high-volume facilities in multivariable analysis in CCP, which means if a CCP facility was being

considered, high-volume facilities would not necessarily be a better choice. Third, facility volume rather than procedure volume was considered in our research. It might be reasonable to choose a hospital according to surgeon volume or type rather than facility volume, as some previous studies have found improved outcomes when cardiothoracic surgeons with high individual case volumes perform lung cancer surgery[104,105]. This information was not available from NCDB and was therefore not included in our study.

CHAPTER FIVE – CONCLUSION AND FUTURE WORK

In summary, we adapted three different methodologies to analyze lung cancer prognostic factors in tumor, patient, and population levels respectively. 1) To analyze tumor level histopathological features, deep learning based pathology image analysis served as a powerful tool to replace error-prone and laborious work of pathologists. Two deep learning models were developed to automatically recognize tumor region and segment cell components in TME respectively. Based on computation staining results from deep learning, tumor-level features were proved significant prognostic factors independent of clinical variables. 2) To analyze patient-level factors, a nomogram was built to incorporate both demographics and clinical variables for individualized prognosis. The online implementation of the nomogram would help both clinicians and patients in making treatment decisions. 3) As population-level factors, facility selection, including different types and volumes, was proved to affect both overall survival and surgery selection. The different tendency to select surgery can largely explain the survival difference existed among different facility types.

In the future, we will: 1) validate performance of our models in independent cohorts. Especially, to validate nuclei recognition accuracy of the mask-RCNN model, H&E stained images with corresponding immunohistochemistry (IHC) stained slides will be collected; the nuclei recognition results for the paired slides will be compared. The inconsistency, if there is any, will be added to the training set to refine our model with IHC stained slides serving as

ground truth. 2) Fully utilize TME information through developing novel deep learning methods. Specifically, a new model, "Graph Convolution Network (GCN)"[106], will be adapted to fit into the graph structure of TME, where each nucleus serves as each vertices, and geometric relationship between a pair of nuclei serves as each edge. Since GCN only utilizes nucleus type and spatial position information, using GCN to classify different histologic subtypes and patient outcomes will be more explainable than algorithms with direct image input and thus providing insights of how nuclei components and distributions affect patient outcome. 3) In addition to prognostic prediction, we will utilize the features in three different levels to predict treatment response. Other cohorts with both pathology images, clinical data, and therapeutic response data will be collected and analyzed. We will especially focus on quantification of innate immune activity through pathological image analysis. Since immunotherapy such as anti-PD1 and anti-PDL1 antibodies has shown durable response in certain NSCLC patients[107], it is important to estimate whether the TME is in immunosuppressive state and select the patients who benefits the most from certain immunotherapy. This individualized predictive model will be validated and used to guide clinical trials by assigning patient treatment groups according to predicted treatment response. Overall, we are aiming to develop novel and powerful algorithms to facilitate pathological diagnosis, prognosis, and prediction of therapeutic response.

BIBLIOGRAPHY

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66(1):7-30.

2. Alberg AJ, Brock MV, Samet JM. Epidemiology of lung cancer: Looking to the future. *J Clin Oncol* 2005;23(14):3175-3185.

3. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* 2018;553(7689):446-454.

4. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, D'Amico TA, DeCamp MM, Dilling TJ, Dobelbower M. Non–small cell lung cancer, version 5.2017, nccn clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network* 2017;15(4):504-535 %@ 1540-1405.

5. Shachar SS, Muss HB. Internet tools to enhance breast cancer care. *NPJ Breast Cancer* 2016;2:16011.

6. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, Geisinger K, Hirsch FR, Ishikawa Y, Kerr KM, Noguchi M, Pelosi G, Powell CA, Tsao MS, Wistuba I, Panel WHO. The 2015 world health organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* 2015;10(9):1243-1260.

7. 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(1):39-51.

8. 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(5):806-814.

9. Gooden MJ, de Bock GH, Leffers N, Daemen T, Nijman HW. The prognostic influence of tumour-infiltrating lymphocytes in cancer: A systematic review with meta-analysis. *Br J Cancer* 2011;105(1):93-103.

10. Bremnes RM, Donnem T, Al-Saad S, Al-Shibli K, Andersen S, Sirera R, Camps C, Marinez I, Busund LT. The role of tumor stroma in cancer progression and prognosis: Emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. *J Thorac Oncol* 2011;6(1):209-217.

11. Ichikawa T, Aokage K, Sugano M, Miyoshi T, Kojima M, Fujii S, Kuwata T, Ochiai A, Suzuki K, Tsuboi M. The ratio of cancer cells to stroma within the invasive area is a histologic prognostic parameter of lung adenocarcinoma. *Lung Cancer* 2018.

12. Han H, Silverman JF, Santucci TS, Macherey RS, dAmato TA, Tung MY, Weyant RJ, Landreneau RJ. Vascular endothelial growth factor expression in stage i non-small cell lung cancer correlates with neoangiogenesis and a poor prognosis. *Ann Surg Oncol* 2001;8(1):72-79.

13. Putila J, Remick SC, Guo NL. Combining clinical, pathological, and demographic factors refines prognosis of lung cancer: A population-based study. *PLoS One* 2011;6(2):e17493.

14. Allison JJ, Kiefe CI, Weissman NW, Person SD, Rousculp M, Canto JG, Bae S, Williams OD, Farmer R, Centor RM. Relationship of hospital teaching status with quality of care and mortality for medicare patients with acute mi. *JAMA* 2000;284(10):1256-1262.

15. Cheung MC, Hamilton K, Sherman R, Byrne MM, Nguyen DM, Franceschi D, Koniaris LG. Impact of teaching facility status and high-volume centers on outcomes for lung cancer resection: An examination of 13,469 surgical patients. *Ann Surg Oncol* 2009;16(1):3-13.

16. McDonald RJ, Cloft HJ, Kallmes DF. Impact of admission month and hospital teaching status on outcomes in subarachnoid hemorrhage: Evidence against the july effect. *J Neurosurg* 2012;116(1):157-163.

17. Meguid RA, Brooke BS, Chang DC, Sherwood JT, Brock MV, Yang SC. Are surgical outcomes for lung cancer resections improved at teaching hospitals? *Ann Thorac Surg* 2008;85(3):1015-1024; discussion 1024-1015.

18. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002;137(6):511-520.

19. Glasgow RE, Mulvihill SJ. Hospital volume influences outcome in patients undergoing pancreatic resection for cancer. *West J Med* 1996;165(5):294-300.

20. Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. *JAMA* 1998;280(20):1747-1751.

21. Wade TP, Halaby IA, Stapleton DR, Virgo KS, Johnson FE. Population-based analysis of treatment of pancreatic cancer and whipple resection: Department of defense hospitals, 1989-1994. *Surgery* 1996;120(4):680-685; discussion 686-687.

22. Kozower BD, Stukenborg GJ. The relationship between hospital lung cancer resection volume and patient mortality risk. *Ann Surg* 2011;254(6):1032-1037.

23. Alpaydin E. *Introduction to machine learning.* MIT press; 2009.

24. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436-444.

25. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun Acm* 2017;60(6):84-90.

26. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep learning.* Vol 1: MIT press Cambridge; 2016.

27. 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:170302442* 2017.

28. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:160605718* 2016.

29. Wang SD, Chen A, Yang L, Cai L, Xie Y, Fujimoto J, Gazdar A, Xiao GH. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. *Sci Rep-Uk* 2018;8.

30. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak J, the CC, Hermsen M, Manson QF, Balkenhol M, Geessink O, Stathonikos N, van Dijk MC, Bult P, Beca F, Beck AH, Wang D, Khosla A, Gargeya R, Irshad H, Zhong A, Dou Q, Li Q, Chen H, Lin HJ, Heng PA, Hass C, Bruni E, Wong Q, Halici U, Oner MU, Cetin-Atalay R, Berseth M, Khvatkov V, Vylegzhanin A, Kraus O, Shaban M, Rajpoot N, Awan R, Sirinukunwattana K, Qaiser T, Tsang YW, Tellez D, Annuscheit J, Hufnagl P, Valkonen M, Kartasalo K, Latonen L, Ruusuvuori P, Liimatainen K, Albarqouni S, Mungal B, George A, Demirci S, Navab N, Watanabe S, Seno S, Takenaka Y, Matsuda H, Ahmady Phoulady H, Kovalev V, Kalinovsky A, Liauchuk V, Bueno G, Fernandez-Carrobles MM, Serrano I, Deniz O, Racoceanu D, Venancio R.

Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318(22):2199-2210.

31. Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems* 1989;2(4):303-314.

32. He KM, Gkioxari G, Dollar P, Girshick R. Mask r-cnn. *Ieee I Conf Comp Vis* 2017:2980-2988.

33. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, Moreira AL, Razavian N, Tsirigos A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24(10):1559-1567.

34. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, Samaras D, Shroyer KR, Zhao TH, Batiste R, Van Arnam J, Shmulevich I, Rao AUK, Lazar AJ, Sharma A, Thorsson V, Network CGAR. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Reports* 2018;23(1):181-+.

35. Wang S, Wang T, Yang L, Yi F, Luo X, Yang Y, Gazdar A, Fujimoto J, Wistuba II, Yao B. Convpath: A software tool for lung adenocarcinoma digital pathological image analysis aided by convolutional neural network. *arXiv preprint arXiv:180910240* 2018.

36. Yi FL, Yang L, Wang SD, Guo L, Huang CL, Xie Y, Xiao GH. Microvessel prediction in h&e stained pathology images using fully convolutional neural networks. *Bmc Bioinformatics* 2018;19.

37. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proc Cvpr Ieee* 2015:3431-3440.

38. 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(108):108ra113.

39. 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. *Sci Rep* 2017;7:45938.

40. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H,

D'Eustachio P. The reactome pathway knowledgebase. *Nucleic Acids Res* 2016;44(D1):D481-487.

41. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *2016 Ieee Conference on Computer Vision and Pattern Recognition (Cpvr)* 2016:2818-2826.

42. Otsu N. Threshold selection method from gray-level histograms. *Ieee T Syst Man Cyb* 1979;9(1):62-66.

43. Vincent L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans Image Process* 1993;2(2):176-201.

44. Chollet F. Keras. 2015; Available at: https://keras.io.

45. Gonzalez RC, Woods R. *Digital image processing.* Pearson Education; 2002.

46. Cheng J, Mo X, Wang X, Parwani A, Feng Q, Huang K. Identification of topological features in renal tumor microenvironment associated with patient survival. *Bioinformatics* 2018;34(6):1024-1030.

47. *A package for survival analysis in s* [computer program]. 2015.

48. *R: A language and environment for statistical computing.* [computer program]. R Foundation for Statistical Computing, Vienna, Austria.2016.

49. Sergushichev A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv* 2016:060012.

50. Cantrell DA. T cell antigen receptor signal transduction pathways. *Cancer Surv* 1996;27:165-175.

51. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 2012;12(4):252-264.

52. Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer* 2006;6(5):392-401.

53. Huang QM, Tomida S, Masuda Y, Arima C, Cao K, Kasahara TA, Osada H, Yatabe Y, Akashi T, Kamiya K, Takahashi T, Suzuki M. Regulation of DNA polymerase pold4 influences genomic instability in lung cancer. *Cancer Research* 2010;70(21):8407-8416.

54. 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(1):278-282; discussion 283.

55. Wang C, Pecot T, Zynger DL, Machiraju R, Shapiro CL, Huang K. Identifying survival associated morphological features of triple negative breast cancer using multiple datasets. *J Am Med Inform Assoc* 2013;20(4):680-687.

56. 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, Markowetz F. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med* 2012;4(157):157ra143.

57. Swinson DEB, Jones JL, Richardson D, Cox G, Edwards JG, O'Byrne KJ. Tumour necrosis is an independent prognostic marker in non-small cell lung cancer: Correlation with biological variables. *Lung Cancer* 2002;37(3):235-240.

58. Topalian SL, Drake CG, Pardoll DM. Targeting the pd-1/b7-h1(pd-l1) pathway to activate anti-tumor immunity. *Curr Opin Immunol* 2012;24(2):207-212.

59. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. Multiclass cancer diagnosis using tumor gene expression signatures. *P Natl Acad Sci USA* 2001;98(26):15149-15154.

60. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale AL. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *P Natl Acad Sci USA* 2001;98(19):10869-10874.

61. Vermeulen K, Van Bockstaele DR, Berneman ZN. The cell cycle: A review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif* 2003;36(3):131-149.

62. Semenza GL. Targeting hif-1 for cancer therapy. *Nat Rev Cancer* 2003;3(10):721-732.

63. Li HP, Courtois ET, Sengupta D, Tan YL, Chen KH, Goh JJL, Kong SL, Chua C, Hon LK, Tan WS, Wong M, Choi PJ, Wee LJK, Hillmer AM, Tan IB, Robson P, Prabhakar S. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 2017;49(5):708-+.

64. Diamond DA, Berry SJ, Umbricht C, Jewett HJ, Coffey DS. Computerized image-analysis of nuclear shape as a prognostic factor for prostatic-cancer. *Prostate* 1982;3(4):321-332.

65. 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.

66. Kalemkerian GP, Akerley W, Bogner P, Borghaei H, Chow LQ, Downey RJ, Gandhi L, Ganti AK, Govindan R, Grecula JC, Hayman J, Heist RS, Horn L, Jahan T, Koczywas M, Loo BW, Jr., Merritt RE, Moran CA, Niell HB, O'Malley J, Patel JD, Ready N, Rudin CM, Williams CC, Jr., Gregory K, Hughes M, National Comprehensive Cancer N. Small cell lung cancer. *J Natl Compr Canc Netw* 2013;11(1):78-98.

67. Xie D, Marks R, Zhang M, Jiang G, Jatoi A, Garces YI, Mansfield A, Molina J, Yang P. Nomograms predict overall survival for patients with small-cell lung cancer incorporating pretreatment peripheral blood markers. *J Thorac Oncol* 2015;10(8):1213-1220.

68. Govindan R, Page N, Morgensztern D, Read W, Tierney R, Vlahiotis A, Spitznagel EL, Piccirillo J. Changing epidemiology of small-cell lung cancer in the united states over the last 30 years: Analysis of the surveillance, epidemiologic, and end results database. *J Clin Oncol* 2006;24(28):4539-4544.

69. Jett JR, Schild SE, Kesler KA, Kalemkerian GP. Treatment of small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest* 2013;143(5 Suppl):e400S-419S.

70. Pan H, Shi X, Xiao D, He J, Zhang Y, Liang W, Zhao Z, Guo Z, Zou X, Zhang J, He J. Nomogram prediction for the survival of the patients with small cell lung cancer. *J Thorac Dis* 2017;9(3):507-518.

71. Xiao HF, Zhang BH, Liao XZ, Yan SP, Zhu SL, Zhou F, Zhou YK. Development and validation of two prognostic nomograms for predicting survival in patients with non-small cell and small cell lung cancer. *Oncotarget* 2017;8(38):64303-64316.

72. The diagnosis and treatment of lung cancer (update). *National Collaborating Centre for Cancer (UK* 2011.

73. Maestu I, Pastor M, Gomez-Codina J, Aparicio J, Oltra A, Herranz C, Montalar J, Munarriz B, Reynes G. Pretreatment prognostic factors for survival in small-cell lung cancer: A new prognostic index and validation of three known prognostic indices on 341 patients. *Ann Oncol* 1997;8(6):547-553.

74. Cerny T, Blair V, Anderson H, Bramwell V, Thatcher N. Pretreatment prognostic factors and scoring system in 407 small-cell lung cancer patients. *Int J Cancer* 1987;39(2):146-149.

75. Rami-Porta R, Bolejack V, Giroux DJ, Chansky K, Crowley J, Asamura H, Goldstraw P, International Association for the Study of Lung Cancer S, Prognostic Factors Committee ABM, Participating I. The iaslc lung cancer staging project: The new database to inform the eighth edition of the tnm classification of lung cancer. *J Thorac Oncol* 2014;9(11):1618-1624.

76. Abdel-Rahman O. Validation of the ajcc 8th lung cancer staging system among patients with small cell lung cancer. *Clin Transl Oncol* 2017.

77. Eberhardt WE, Mitchell A, Crowley J, Kondo H, Kim YT, Turrisi A, 3rd, Goldstraw P, Rami-Porta R, International Association for Study of Lung Cancer S, Prognostic Factors Committee ABM, Participating I. The iaslc lung cancer staging project: Proposals for the revision of the m descriptors in the forthcoming eighth edition of the tnm classification of lung cancer. *J Thorac Oncol* 2015;10(11):1515-1522.

78. 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(8):706-714.

79. Rami-Porta R, Bolejack V, Crowley J, Ball D, Kim J, Lyons G, Rice T, Suzuki K, Thomas CF, Jr., Travis WD, Wu YL, Staging I, Prognostic Factors Committee AB, Participating I. The iaslc lung cancer staging project: Proposals for the revisions of the t descriptors in the forthcoming eighth edition of the tnm classification for lung cancer. *J Thorac Oncol* 2015;10(7):990-1003.

80. Postmus PE, Brambilla E, Chansky K, Crowley J, Goldstraw P, Patz EF, Jr., Yokomise H, International Association for the Study of Lung Cancer International Staging C, Cancer R, Biostatistics, Observers to the C, Participating I. The iaslc lung cancer staging project: Proposals for revision of the m descriptors in the forthcoming (seventh) edition of the tnm classification of lung cancer. *J Thorac Oncol* 2007;2(8):686-693.

81. Yang L, Wang S, Zhou Y, Lai S, Xiao G, Gazdar A, Xie Y. Evaluation of the 7th and 8th editions of the ajcc/uicc tnm staging systems for lung cancer in a large north american cohort. *Oncotarget* 2017.

82. Li J, Ma S. Time-dependent roc analysis under diverse censoring patterns. *Stat Med* 2011;30(11):1266-1277.

83. Heagerty PJ, Lumley T, Pepe MS. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics* 2000;56(2):337-344.

84. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008;26(8):1364-1370.

85. Slotman BJ, van Tinteren H, Praag JO, Knegjens JL, El Sharouni SY, Hatton M, Keijser A, Faivre-Finn C, Senan S. Use of thoracic radiotherapy for extensive stage small-cell lung cancer: A phase 3 randomised controlled trial. *Lancet* 2015;385(9962):36-42.

86. Ganti AK, Zhen W, Kessinger A. Limited-stage small-cell lung cancer: Therapeutic options. *Oncology (Williston Park)* 2007;21(3):303-312; discussion 312, 315-308, 323.

87. Combs SE, Hancock JG, Boffa DJ, Decker RH, Detterbeck FC, Kim AW. Bolstering the case for lobectomy in stages i, ii, and iiia small-cell lung cancer using the national cancer data base. *J Thorac Oncol* 2015;10(2):316-323.

88. Vallieres E, Shepherd FA, Crowley J, Van Houtte P, Postmus PE, Carney D, Chansky K, Shaikh Z, Goldstraw P, International Association for the Study of Lung Cancer International Staging C, Participating I. The iaslc lung cancer staging project: Proposals regarding the relevance of tnm in the pathologic staging of small cell lung cancer in the forthcoming (seventh) edition of the tnm classification for lung cancer. *J Thorac Oncol* 2009;4(9):1049-1059.

89. Henschke CI, Yankelevitz DF, Smith JP, Libby D, Pasmantier M, McCauley D, McGuinness G, Naidich DP, Farooqi A, Vasquez M, Miettinen OS. Ct screening for lung cancer assessing a regimen's diagnostic performance. *Clin Imaging* 2004;28(5):317-321.

90. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD, Team NLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New Engl J Med* 2011;365(5):395-409.

91. Tanoue LT, Tanner NT, Gould MK, Silvestri GA. Lung cancer screening. *Am J Respir Crit Care Med* 2015;191(1):19-33.

92. Davis JN, Medbery C, Sharma S, Pablo J, Kimsey F, Perry D, Muacevic A, Mahadevan A. Stereotactic body radiotherapy for centrally located early-stage

non-small cell lung cancer or lung metastases from the rssearch((r)) patient registry. *Radiat Oncol* 2015;10:113.

93. Yang HX, Woo KM, Sima CS, Bains MS, Adusumilli PS, Huang J, Finley DJ, Rizk NP, Rusch VW, Jones DR, Park BJ. Long-term survival based on the surgical approach to lobectomy for clinical stage i nonsmall cell lung cancer: Comparison of robotic, video-assisted thoracic surgery, and thoracotomy lobectomy. *Ann Surg* 2017;265(2):431-437.

94. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, Escriu C, Peters S, Committee EG. Early and locally advanced non-small-cell lung cancer (nsclc): Esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017;28(suppl_4):iv1-iv21.

95. Bach PB, Cramer LD, Schrag D, Downey RJ, Gelfand SE, Begg CB. The influence of hospital volume on survival after resection for lung cancer. *N Engl J Med* 2001;345(3):181-188.

96. Rosen JE, Hancock JG, Kim AW, Detterbeck FC, Boffa DJ. Predictors of mortality after surgical management of lung cancer in the national cancer database. *Ann Thorac Surg* 2014;98(6):1953-1960.

97. Ash AS, et al. Statistical issues in assessing hospital performance. 2012; Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf. Accessed Sep, 2016.

98. Burgette L, Griffin BA, McCaffrey D. Propensity scores for multiple treatments: A tutorial for the mnps function in the twang package. *R package Rand Corporation* 2017.

99. *R: A language adn environment for statistial computing, vienna, austria. Url https://r-project.Org* [computer program]. 2016.

100. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 2008;83(5):584-594.

101. Phillips KA, Luft HS. The policy implications of using hospital and physician volumes as "indicators" of quality of care in a changing health care environment. *Int J Qual Health Care* 1997;9(5):341-348.

102. Shortell SM, Hughes EFX. The effects of regulation, competition, and ownership on mortality-rates among hospital inpatients. *New Engl J Med* 1988;318(17):1100-1107.

103. Gordon TA, Burleyson GP, Tielsch JM, Cameron JL. The effects of regionalization on cost and outcome for one general high-risk surgical-procedure. *Annals of Surgery* 1995;221(1):43-49.

104. Smith CB, Wolf A, Mhango G, Wisnivesky JP. Impact of surgeon volume on outcomes of older stage i lung cancer patients treated via video-assisted thoracoscopic surgery. *Semin Thorac Cardiovasc Surg* 2017;29(2):223-230.

105. Goodney PP, Lucas FL, Stukel TA, Birkmeyer JD. Surgeon specialty and operative mortality with lung resection. *Ann Surg* 2005;241(1):179-184.

106. Zhou J, Cui G, Zhang Z, Yang C, Liu Z, Sun M. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:181208434* 2018.

107. Anagnostou VK, Brahmer JR. Cancer immunotherapy: A future paradigm shift in the treatment of non-small cell lung cancer. *Clinical Cancer Research* 2015;21(5):976-984.