Predicting Non-Small-Cell Lung Cancer Survival Outcome from Pretreatment CT-scans

Anisha Parsan

August 1, 2020

Abstract

Lung cancer is one of the leading causes of cancer-related deaths worldwide, and non-small-cell lung cancer (NSCLC) accounts for 80-85% of these cases. Determining an accurate prognosis is key in developing a personalized approach to treatment of NSCLC, which could ultimately aid in improving survival rate. Here, we developed a 2D, end-to-end model for NSCLC survival outcome prediction. We trained and evaluated the model on pretreatment CT-scans of 422 patients from The Cancer Imaging Archive. Our methodology results in an AUC of 0.59 in predicting survival outcome. Our framework presents a new application of a particular approach, and demonstrates avenues for further improvement.

1 Introduction

Lung cancer is one of the leading causes of cancer-related deaths worldwide [2]. Non-small cell lung cancer (NSCLC) results in approximately 80-85% of lung cancer diagnosis, with a 5-year survival rate of 24%. Improving disease outcomes of NSCLC relies largely on implementing personalized treatment plans early in diagnosis [1]. Prognosis of NSCLC is key in determining this personalized approach to treatment, as a more-informed decision on treatment can be made after assessing the risk of the patient [4].

Evaluation of prognosis is typically performed by physicians and oncologists, and consists of a largely qualitative approach that may be prone to human error. However, machine learning offers an accurate alternative to quantitatively evaluate prognosis of a patient. The widespread use of ML techniques in disease research has led to an increase in the use of deep learning for cancer prognosis, resulting in a 15%-20% increase of prediction accuracy from human accuracy to deep learning in the past few years [4]. Current techniques for cancer survival prediction involving machine learning largely consist of analysis of genomic, clinical, and histological information surrounding a patient [10]. However, tumor environments include many factors located not only within cancer cells, but also stromal and immune cells. This can often lead to the inclusion of unnecessary information that limits the predictive accuracy of ML algorithms [7]. Such approaches also typically require extensive tests and tissue sampling, and therefore are fairly invasive.

Radiomics for cancer prognosis offers a non-invasive and straightforward approach to cancer survival prediction. Connections between image-related biomarkers and NSCLC progression have largely been established, as CT scans are widely used for TNM staging [8]. Medical imaging can provide a picture of tumor phenotype and the surrounding environment, offering a promising avenue for prediction of survival through ML. In recent years, radiomics has been extensively investigated for the purpose of outcome prediction, along with other aspects of cancer research (e.g. diagnosis, classifying subtypes, nodule detection) [10]. A number of radiomic features have prognostic power in datasets of lung cancer patients, many of which were not previously noted [9, 10].

Approaches towards cancer prognosis through radiomics and machine learning largely consist of two main approaches. The first is a feature extraction and selection by hand or with a convolutional neural network (CNN), followed by outcome prediction through the use of ML algorithms (e.g. Random Forest, Support Vector Machine) [1, 9]. This particular method is relatively well studied, and has already been established for NSCLC cancer prognosis. Many of these methodologies require human input for feature extraction and selection, resulting in relatively low automation. However, the relatively new approach of constructing end-to-end CNNs for cancer prognosis offers a much more independent and highly automated method of survival prediction. A recent study utilized this approach to predict survival of head and neck cancer patients, solely based on pretreatment CT scans [5]. Results were compared to a benchmark study that employed the approach of feature extraction and separate outcome prediction [6]. The end-to-end approach demonstrated a higher AUC for overall survival prediction relative to the benchmark, and the same dataset was used in both studies. Various recent studies have employed a 2-dimensional end-to-end approach towards survival outcome prediction of NSCLC [11, 12],

with a reported AUC of 0.75-0.90. Utilizing a 3-dimensional approach allowing for incorporation of the whole tumor volume could offer options for further improvement [5, 13].

The primary focus of this work was to develop a 3D, end-to-end model that is capable of determining survival outcome of diagnosed NSCLC patients solely using computed tomography (CT) scans. We develop a U-net model to segment and output nodule segmentations of each scan. Segmented nodules are fed into an end-to-end 3D convolutional neural network (CNN) for binary classification. Similar methodologies have been utilized in segmentation of lung nodules and subsequent classification of malignancy [14], as well as for survival prediction in brain tumors (MRI scans) [17]. To our knowledge, this particular approach has not been implemented in the context of NSCLC.

Example citation [1].

2 Purpose

1. The primary purpose of this study is to develop a convolutional neural network to effectively predict non-small cell lung cancer survival based on pre-treatment CT-scans of a patient.

2. The model will be trained and tested on a dataset consisting of annotated CT-scans of NSCLC patients, and will then be evaluated through various statistical techniques.

3. The performance of the developed model will be compared to other established models for cancer survival prediction that have utilized the same dataset.

3 Methods

Patient Cohorts:

This study utilizes a publicly available dataset on The Cancer Imaging Archive repository (TCIA). The Lung1 dataset has a total of 422 full CT-scans of NSCLC patients treated at the MAASTRO Clinic in the Netherlands. Additionally, the dataset contained clinical information regarding each patient (e.g. survival time, histology, stage at diagnosis). Figure 1 demonstrates a general overview of the relationships between percent survival and clinical information of patients in the dataset.

Percent survival tends to decrease with age, as well as later stages. Such trends in the dataset could affect model performance on specific groups of patients. The LUNA-16 dataset consisted of CT-scans of 888 cancerous and non-cancerous patients. In this study, CT-scans of only cancerous patients were used, along with nodule annotations for these scans.

Data Processing:

Processing of CT-scans was necessary to reduce unwanted noise and normalize images. Both datasets underwent thresholding for segmentation of

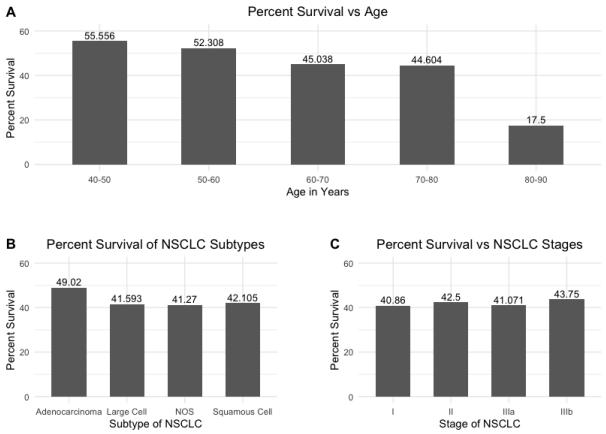

Figure 1: Demographics of patient cohort in comparison with survival. Demonstrates the percent of each group that survived within the given dataset. Of particular interest could be the imbalances between the survival classes in different groups. Could introduce potential variation in model accuracy.

lung structures. Additionally within the datasets consisted of lung thresholding along with erosion and dilation for noise reduction. Binary masks containing corresponding nodule segmentations for segmented lung images were produced for the LUNA16 dataset.

Current Approach:

Due to constraints on time, the current model utilized a different approach than listed above. For future research, the full approach detailed above would be implemented. The current approach utilizes a 2D CNN with an input of the central slice of a patient CT-scan and an output of overall survival. The model was trained on the segmented central slices of nodules. Visualization of the approach is provided in Figure 2. The overall architecture of the CNN is the same as the 2D CNN utilized in the following methodology.

Long-term Approach: would continue implementing after program, not currently completed.

In this study a modified version of the U-Net as described in Chon et. al. [14] was utilized. Initially, a 2D binary classification model was trained on simply the center images of CT-scans. However, this did not provide an accurate representation of the CT-scan or the patients. U-nets are widely used for lung and nodule segmentation [15], and would allow for more specific approach to analysis of CT-scans. Our modified U-Net takes an input of 512 512 2D CT slices, and was labeled with 512 512 binary masks containing nodule segmentations from the LUNA 16 dataset. The U-net was trained to output nodule segmentation binary masks for the Lung1 NSCLC-Radiomics dataset. Metrics used for evaluations were dice coefficient and overall accuracy.

Outputted 2D binary masks containing nodule segmentations for each

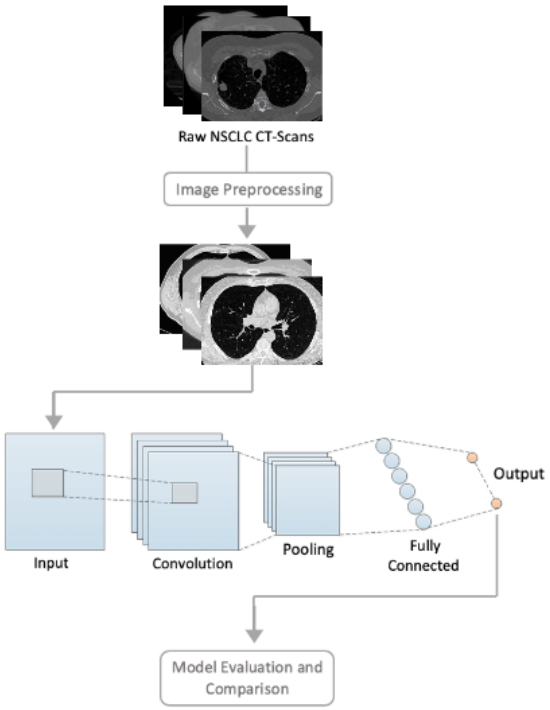

Figure 2: Overview of methodology in the current approach. Model consists of a 2D Convolutional Neural Network. Contains three convolutional blocks containing a convolutional layer, pooling layer, and relu layer.

slice within CT-scans were then cropped and stacked for input. These images were then utilized to train a binary classification 3D CNN, to predict the overall survival of a patient. The overall architecture of this CNN was adapted from GoogleNet [16], which utilized a similar approach to predicting overall survival for head and neck cancer patients. Figure 3 depicts the architecture of this model, which consists of three convolutional blocks followed by multiple fully connected layers. These blocks consist of a convolutional layer, a pooling layer, and a ReLu layer. Data augmentation with random rotation (0-20), vertical/horizontal flipping, and shifting was performed to increase variability and volume of training data.

4 Results

The current progress made shows results of the end-to-end 2D CNN, with the input being the central slice of a given CT-scan and the output being the survival outcome. Figure 4b shows a plot containing the ROC of the model. The AUC was 0.59, and Figure 4 shows the ROC curve of the model.

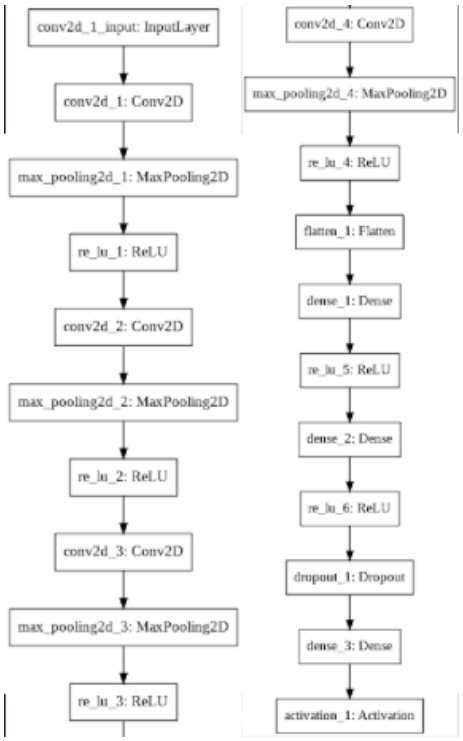

Figure 3: Architecture of 2D CNN for binary classification (with current approach, not long term approach)

5 Discussion and Conclusion

Determining an accurate prediction of survival outcome is key in determining a personalized approach to treatment for NSCLC. The need for high accuracy led to the widespread use of ML in prognosis, and demonstrated the prognostic power of radiomic biomarkers. Though many studies have noted the benefits of 3D CNN's and end-to-end models separately [4, 12], the implementation of both has not been thoroughly investigated in the context of NSCLC. The primary goal of this study was to develop an end-to-end, 3D CNN for survival outcome prediction. However, due to time constraints, the current developed model consists of a 2D, end-to-end approach. This approach had shown promising results in head and neck cancer [4], but had not been directly utilized for NSCLC.

The state-of-the-art model for this particular task achieves an AUC of 0.75-0.85. Though our AUC was relatively low in comparison, future optimization and fine tuning to the 2D model could improve this approach. These results could show potential for accurate prediction of survival outcome, considering various limitations such as training data size as well as time constraints. However, applying a 2D approach may not be the most optimal way to consider all factors that lead to an accurate prognosis of NSCLC.

The 2D approach utilized only considered the central tumor slice of each patient. This does not accurately represent the tumor or the patient, pos-

sibly adding to the cause for such a low AUC. Considering central slices from various axes could present an avenue for improvement for the 2D approach, though going forward, a 3D approach will be implemented. Another avenue for improvement would include consideration of clinical information of a patient. Collection of such information is much less invasive than genomic and histological data, and could possibly increase the accuracy of models.

In terms of the future model (2D U-Net, 3D-CNN approach), implementation of a method of false positive reduction after nodule segmentation would be beneficial. The outputted masks from the U-Net generally are able to segment the nodule present in the ground truth mask, but also include various other candidate regions.

Though this model did not present a high AUC in comparison to state-of-the-art methods, there is promise for improvement. Additionally, this study has demonstrated that implementing a 2-dimensional approach for NSCLC prognosis from solely CT-scans may also not be the most effective method, as many extraneous factors are excluded in this process. This suggests future work to implement more patient information, along with a holistic tumor approach to segmentation and survival prediction from radiomic information.

References

[1] Guangqi Cui, John Dickerson, Naveen Durvasula, William Gasarch, Erik Metz, Jacob Prinz, Naveen Raman, Daniel Smolyak, and Sung Hyun Yoo. A muffin-theorem generator. In *9th International Conference on Fun with Algorithms (FUN 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.