

Final Report

Development and validation of prediction models for lung cancer using AI

Name: Do Tien Bach Thanh

ID: 208330

Class: Advanced project for AI convergence

Company: Gwangju Artificial Intelligence Inc. Company

Abstract

In the Advanced project for AI convergence class, we worked on Development and validation of prediction models for lung cancer using AI. Doctors use traditional machine learning or math to predict the survival day of their patients. This task is to help doctors understand their patients well and earlier. In this work, we used MedicalNet for segmentation and we combined the results with clinical information, then used ANN method to predict survival days. We got 437.429, 424467.283, 651.512, 0.5159 by MAE, MSE, RMSE, C-index, respectively.

Introduction

Lung cancer is the most common cancer-related death worldwide. Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) are the two main histological forms of lung cancer, accounting for roughly 85 percent and 15 percent of lung cancer cases, respectively [1][2]. Prognostic information derived from genetic or medical imaging data is useful, but not all patients have access to this information. In lung cancer patients, however, clinical data such as age, gender,... are typically acquired during diagnostic examination. The purpose of this study was to examine the utilization of clinical data in the analysis of NSCLC patient's survival.

Dataset

NSCLC dataset has 2 kinds of dataset. One is 2D dcom image, the other is clinical information of patients. For segmentation, NSCLC has 2 dataset, the smaller one has 154 patients with segmentation mask, while the bigger one has 2687 patients without segmentation mask, but it has clinical information. The dataset combines PET and CT image for segmentation. Clinical parameters (age, gender, smoking status, smoking amount,...) were recorded for all patients in a bigger dataset.

Preprocessing

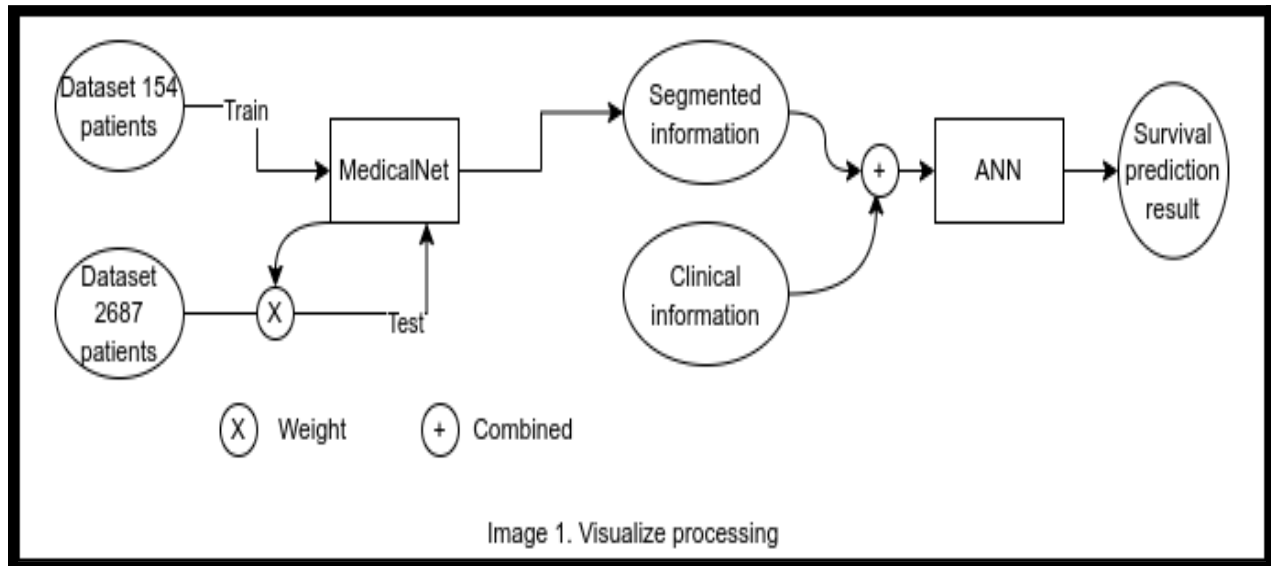
NSCLC dataset combines a lot of 2D dcom images of patients. For better work, we transferred 2D images to 3D image nii.gz. Then we resize to 256*256*depth of image.

Method

Proposed method

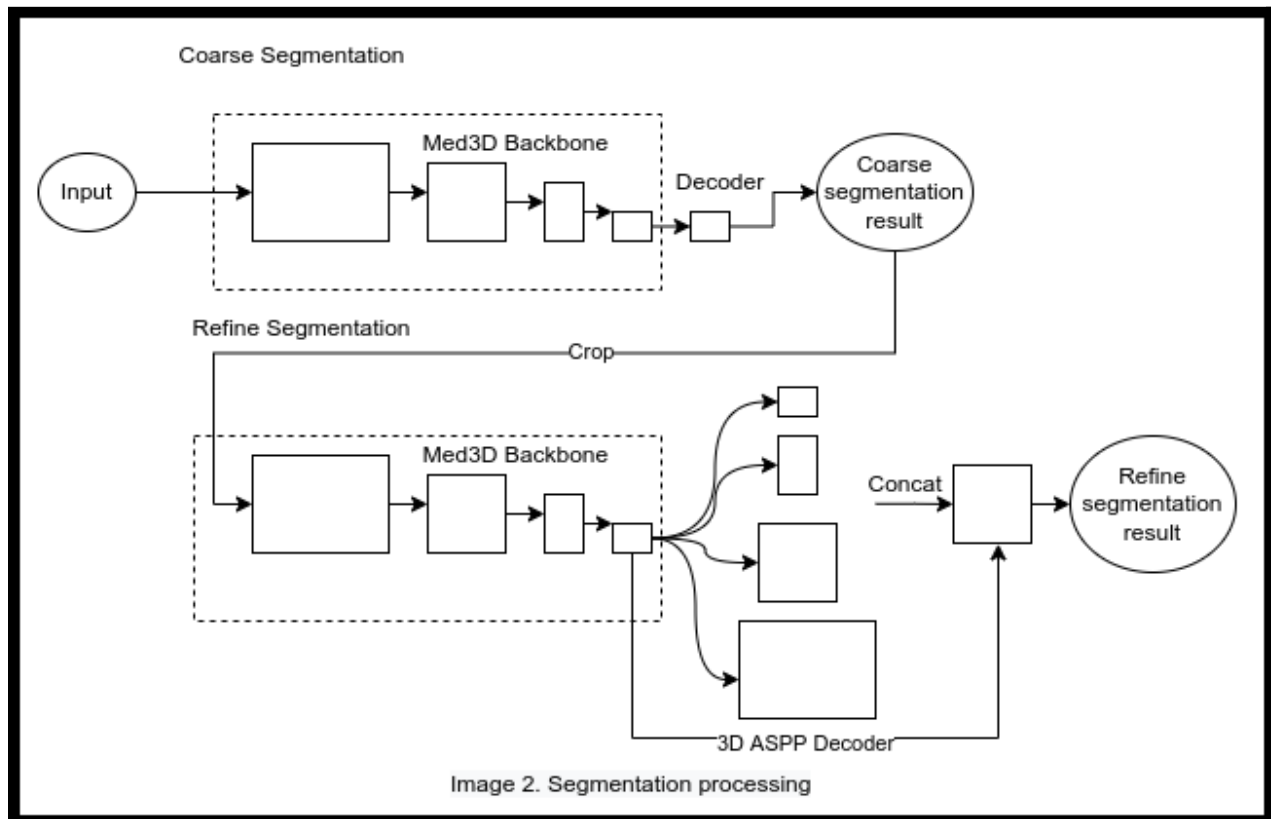
Image 1 shows the proposed method architecture. At first, we used a smaller dataset to train CT & PET image by MedicalNet [3], then we used the output weight to test on a bigger dataset.

PyRadiomics [4] was used to extract radiomic from segmented results. We combined radiomic features from segmented information with clinical information and used ANN model to have the final survival result.



Segmentation method

We used MedicalNet [3] for a segmentation task.



Result

Table 1. Results for proposed method

	MAE	MSE	RMSE	C-index
Our method	414.245	366797.501	605.105	0.5159

When we worked, we found that the result on segmentation is not good and very difficult to train, due to no label provision. We had to use weight from 154 patients with labels. In the dataset, there is a lack of information to visualize the right position of the tumor, which consumes a lot of time to find the right one. We will continue training to improve our results.

There are two results from segmentation task.

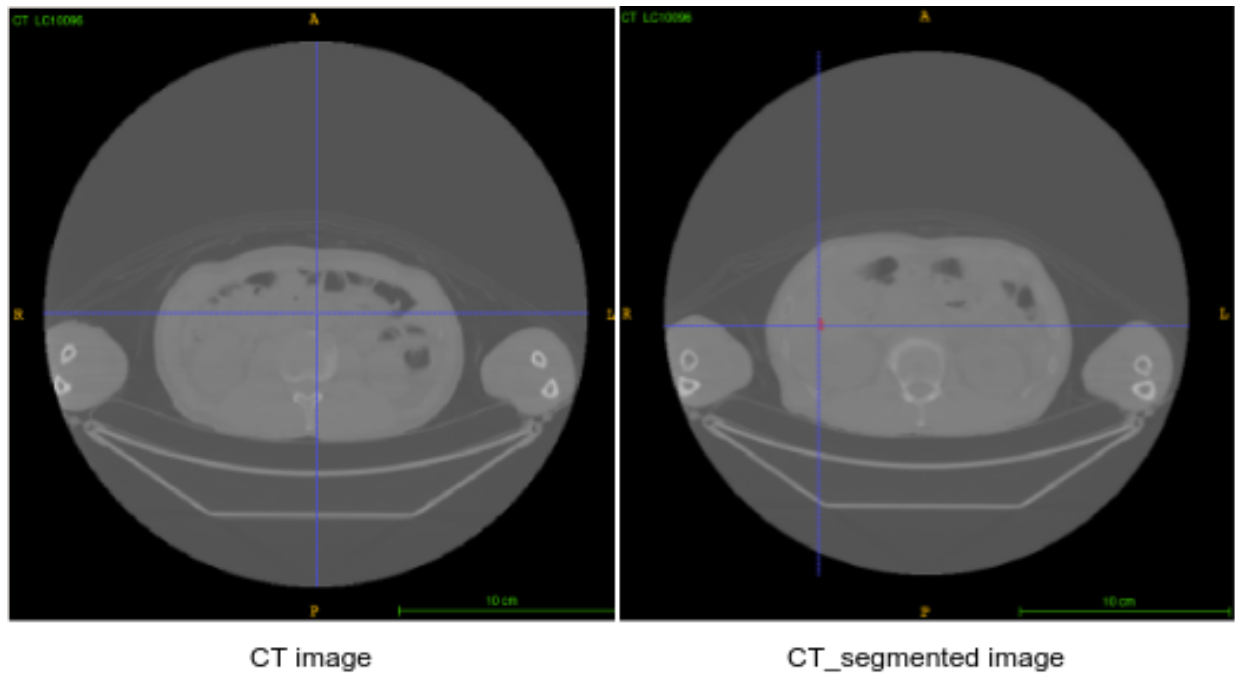


Image 3. CT image result

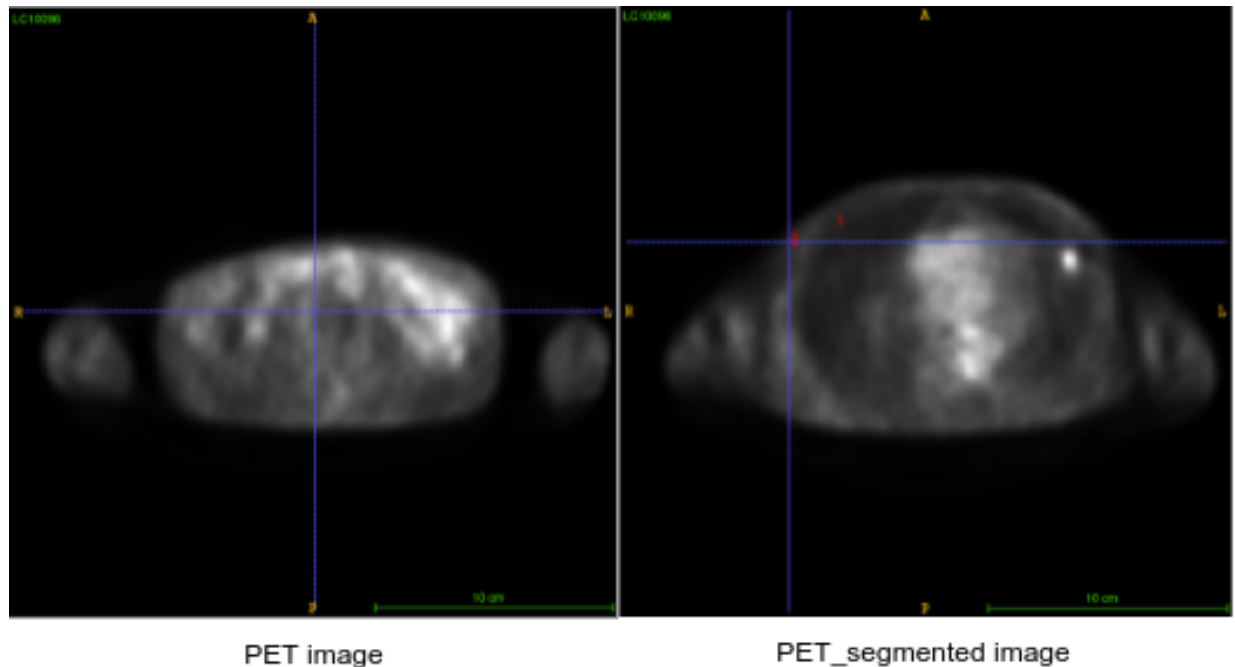


Image 4. PET image result

Conclusion

We worked on this research to answer the challenge of NSCLC survival prediction using clinical data. Due to privacy concerns and the management process, data in medicine is difficult to get, resulting in a scarcity of samples. To deal with this problem, we combined information from the segmentation process with clinical data and used it to generate a survival prediction result. We got 437.429, 424467.283, 651.512, 0.5159 by MAE, MSE, RMSE, C-index, respectively. In addition, we use a combination of numeric and categorical data, which is lacking in recent studies on survival analysis.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2018R1D1A3B05049058) and This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF)& funded by the Korean government (MSIT) (NRF-2019M3E5D1A02067961) and This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2020R1A4A1019191).

References

- [1] Oser, M.G.; Niederst, M.J.; Sequist, L.V.; Engelman, J.A. Transformation from non-small-cell lung cancer to small-cell lung cancer: Molecular drivers and cells of origin. *Lancet Oncol.* **2015**, *16*, e165–e172.
- [2] Lu, T.; Yang, X.; Huang, Y.; Zhao, M.; Li, M.; Ma, K.; Yin, J.; Zhan, C.; Wang, Q. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. *Cancer Manag. Res.* **2019**, *11*, 943.
- [3] Chen, Sihong; Ma, Kail Zheng, Yefeng. Med3D: Transfer Learning for 3D Medical Image Analysis. arXiv preprint arXiv: 1904.00625. **2019**.

[4] van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. Computational Radiomics System to Decode the Radiographic Phenotype. (2017) *Cancer Research*, 77(21), e104–e107.