Predicting the Chance of Lung Disease using Patient Data and Identifying the Type of Lung Disease by CT scans

Paradorn Watcharasemakul

Massive Information & Knowledge Engineering Laboratory
Department of Computer Engineering
Faculty of Engineering, Kasetsart University, Thailand.
E-mail: paradorn.wa@ku.th

Abstract - Lung diseases have been the cause of death for numerous people throughout history. Due mostly to doctors' misdiagnoses and patients worrying about the exorbitant expense of scanning, the majority of patients receive their diagnosis too late. The aforementioned factors may avoid patients from receiving treatment on a timely basis, which might be fatal. As a result, We see the significance of developing a program that will aid in closing the gaps in the erroneous identification of lung disorders. To clarify, this program will assess patients' risk of developing lung cancer and decrease excessive CT scan expenditures for patients. Finally, let me express my earnest hope that this program will be somewhat helpful to both doctors and patients.

Keywords—Lung diseases classification, CT scan identification, lung cancer risk

1. Introduction

Cancer is among the top 5 deadliest diseases in Thailand and lung cancer is one of the deadliest cancers [1, 8]. The primary causes of lung cancer in Thai people include smoking, pollution, a hazardous environment, and heredity [6].

However, other lung disorders can also be quite dangerous and can even be deadly, so it is not simply lung cancer that is lethal. CoronaVirus disease is the lung condition that has killed more than 6.39 million people worldwide and been plaguing individuals for more than two years [7]. Other lung conditions including bacterial pneumonia, viral pneumonia, tuberculosis, and many more are also extremely hazardous.

Although these lung disorders are dangerous, it is difficult to diagnose them. Annual health exams often place a greater emphasis on detecting chronic non-communicable diseases (NCDs), such as diabetes and high blood pressure, than on a comprehensive cancer screening. Most cancer patients discover they have the disease after it is well advanced [2]. As a result, we examined user input to model for the risk analysis of lung cancer. Users who are at high risk of lung cancer should get a CT scan so that any lung problems they may have may be checked out on the film. In order to improve the accuracy of the results, practitioners may also utilize this model to aid in lung disease type diagnosis. This model distinguishes three types of lung cancer: Adeno, Squamous, and Carci. Moreover, this model can also classify four other types of lung cancer: Bacterial Pneumonia, Viral Pneumonia, Tuberculosis, and Corona Virus disease.

Both of these models are built on machine learning, which employs the scikit-learn and tensorflow libraries as its two major libraries. Both the picture and the CSV datasets are taken out of Kaggle for the model training [3, 4, 5]. Its goal is to assist program users in determining their risk of developing lung cancer so they may decide whether to get CT Scan. This objective is especially helpful for people who wish to avoid paying for needless CT scans. Additionally, if the patients decide to get a CT Scan, they can use this tool to classify the kind of lung cancer or potential lung illness using the findings of the analysis.

2. Propose Method

2.1 Prepare the kaggle's existing patient data in CSV format for model training using the scikit-learn (sklearn) library.

2.2 Create a model to predict the risk of lung cancer using the scikit-learn (sklearn) library's Decision Tree Classification.

2.3 Test the model to determine the accuracy of calculating the risk of lung cancer in patients.

2.4 Prepare the existing complete picture data of CT scan of lung disease from Kaggle for model training using the tensorflow library.

2.5 Prepare and preprocess the existing incomplete picture data of CT scan of lung cancer from Kaggle, add this folder into the lung disease dataset folder.

2.6 Utilize Convolutional Neural Network to build a model using keras from tensorflow.

2.7 Train the model using epochs=50 and step_per_epoch=100 along with testing the accuracy of lung disease CT scan identification.

3. Expected Result

We genuinely hope that the model we have developed will be able to determine a patient's risk of developing lung cancer with an accuracy of 85 percent or higher. In addition, we also anticipate that this model will be capable of accurately classifying lung disease categories from CT scans film with an accuracy of 80 percent or higher.

4. Project Planning

Description	Week							
	1	2	3	4	5	6	7	8
1. Review Paper								
2. Data preparation								
3.Implementation								
4. Evaluation								
5. Results discussion								
6. Conclusion and report								
7. Research presentation								

5. References

[1] Department of Medical Services. (2022). The Department of Medicine reveals that lung cancer is the second leading cause of death in Thai people. [online.]. Available: https://bit.ly/3JjLcHC. [Accessed 26 July 2022]

[2] KWANVAT AVUTTHONG. (2017). Why are people diagnosed with cancer at the last stage?. [online.]. Available: https://bit.ly/3OQcE0N [Accessed 26 July 2022]

[3] MYSAR AHMAD BHAT. (2021). Lung Cancer. [online.]. Available: https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer. [Accessed 1 July 2022]

[4] OMKAR MANOHAR DALVI. (2022). Lungs Disease Dataset (4 types). [online.]. Available: https://www.kaggle.com/datasets/omkarmanohardalvi/lungs-disease-dataset-4-types. [Accessed 1 July 2022]

[5] SHADY MOHAMED. (2022). LUNG CANCER (Adeno, Carci, Squarmo). [online.]. Available: https://www.kaggle.com/datasets/shadym0hamed/lung-cancer-adeno-carci-squarmo. [Accessed 1 July 2022]

[6] Siriraj Piyamaharajkarun Hospital. (2021). Lung cancer at any stage can be taken care of. [online.]. Available: https://www.siphhospital.com/th/news/article/share/621/Lungcancer. [Accessed 26 July 2022]

[7] Wikipedia. (2022). Template:COVID-19 pandemic data. [online.]. Available: https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data. [Accessed 26 July 2022]

[8] WISARUT JARUANATAPONG. (2020). Five serious diseases that cause the most deaths in Thai people.. [online.]. Available: https://www.tisco.co.th/th/advisory/top5-deadliest-group-diseases-thailand.html. [Accessed 26 July 2022]