Deep Learning and Lung Disease

The recent developments of deep learning support the identification and classification of lung diseases in medical images.

1. Introduction

Lung diseases, also known as respiratory diseases, are diseases of the airways and the other structures of the lungs [1]. Examples of lung disease are pneumonia, tuberculosis and Coronavirus Disease 2019 (COVID-19). According to Forum of International Respiratory Societies [2], about 334 million people suffer from asthma, and, each year, tuberculosis kills 1.4 million people, 1.6 million people die from lung cancer, while pneumonia also kills millions of people. The COVID-19 pandemic impacted the whole world [3], infecting millions of people and burdening healthcare systems [4]. It is clear that lung diseases are one of the leading causes of death and disability in this world. Early detection plays a key role in increasing the chances of recovery and improve long-term survival rates [5][6]. Traditionally, lung disease can be detected via skin test, blood test, sputum sample test [7], chest X-ray examination and computed tomography (CT) scan examination [8]. Recently, deep learning has shown great potential when applied on medical images for disease detection, including lung disease.

Deep learning is a subfield of machine learning relating to algorithms inspired by the function and structure of the brain. Recent developments in machine learning, particularly deep learning, support the identification, quantification and classification of patterns in medical images [9]. These developments were made possible due to the ability of deep learning to learned features merely from data, instead of hand-designed features based on domain-specific knowledge. Deep learning is quickly becoming state of the art, leading to improved performance in numerous medical applications. Consequently, these advancements assist clinicians in detecting and classifying certain medical conditions efficiently [10].

Numerous works on the detection of lung disease using deep learning can be found in the literature. To the best of our knowledge, however, only one survey paper has been published in the last five years to analyse the state-of-the-art work on this topic [11]. In that paper, the history of deep learning and its applications in pulmonary imaging are presented. Major applications of deep learning techniques on several lung diseases, namely pulmonary nodule diseases, pulmonary embolism, pneumonia, and interstitial lung disease, are also described. In addition, the analysis of several common deep learning network structures used in medical image processing is presented. However, their survey is lacking in the presentation of taxonomy and analysis of the trend of recent work. A taxonomy shows relationships between previous work and categorises them based on the identified attributes that could improve reader understanding of the topic. Analysis of trend, on the other hand, provides an overview of the research direction of the topic of interest identified from the previous work.

2. The Basic Process to Apply Deep Learning for Lung Disease Detection

In this section, the process of how deep learning is applied to identify lung diseases from medical images is described. There are mainly three steps: image preprocessing, training and classification. Lung disease detection generally deals with classifying an image into healthy lungs or disease-infected lungs. The lung disease classifier, sometimes known as a model, is obtained via training. Training is the process in which a neural network learns to recognise a class of images. Using deep learning, it is possible to train a model that can classify images into their respective class labels. Therefore, to apply deep learning for lung disease detection, the first step is to gather images of lungs with the disease to be classified. The second step is to train the neural network until it is able to recognise the diseases. The final step is to classify new images. Here, new images unseen by the model before are shown to the model, and the model predicts the class of those images. The overview of the process is illustrated in Figure 1.

Figure 1. Overview of using deep learning for lung disease detection.

2.1. Image Acquisition Phase

The first step is to acquire images. To produce a classification model, the computer needs to learn by example. The computer needs to view many images to recognise an object. Other types of data, such as time series data and voice data, can also be used to train deep learning models. In the context of the work surveyed in this paper, the relevant data required to detect lung disease will be images. Images that could be used include chest X-ray, CT scan, sputum smear microscopy and histopathology image. The output of this step is images that will later be used to train the model.

2.2. Preprocessing Phase

The second step is preprocessing. Here, the image could be enhanced or modified to improve image quality. Contrast Limited Adaptive Histogram Equalisation (CLAHE) could be performed to increase the contrast of the images [12]. Image modification such as lung segmentation [13] and bone elimination [14] could be used to identify the region of interest (ROI), whereby the detection of the lung disease can then be performed on the ROI. Edge detection could also be used to provide an alternate data representation [15]. Data augmentation could be applied to the images to increase the amount of available data. Feature extraction could also be conducted so that the deep learning model could identify important features to identify a certain object or class. The output of this step is a set of images whereby the quality of the images is enhanced, or unwanted objects have been removed. The output of this step is images that were enhanced or modified that will later be used in training.

2.3. Training Phase

In the third step, namely training, three aspects could be considered. These aspects are the selection of deep learning algorithm, usage of transfer learning and usage of an ensemble. There are numerous deep learning algorithm, for example deep belief network (DBN), multilayer perceptron neural network (MPNN), recurrent neural network (RNN) and the aforementioned CNN. Different algorithms have different learning styles. Different types of data work better with certain algorithms. CNN works particularly well with images. Deep learning algorithm should be chosen based on the nature of the data at hand. Transfer learning refers to the transfer of knowledge from one model to another. Ensemble refers to the usage of more than one model during classification. Transfer learning and ensemble are techniques used to reduce training time, improve classification accuracy and reduce overfitting [16]. The output of this step is models generated from the data learned.

2.4. Classification Phase

In the fourth and final step, which is classification, the trained model will predict which class an image belongs to. For example, if a model was trained to differentiate X-ray images of healthy lungs and tuberculosis-infected lungs, it should be able to correctly classify new images (images that are never seen by the model before) into healthy lungs or tuberculosis-

infected lungs. The model will give a probability score for the image. The probability score represents how likely an image belongs to a certain class. At the end of this step, the image will be classified based on the probability score given to it by the model.

References

1. Bousquet, J. Global Surveillance, Prevention and Control of Chronic Respiratory Diseases; World Health Organization: Geneva, Switzerland, 2007; pp. 12–36.

2. Forum of International Respiratory Societies. The Global Impact of Respiratory Disease, 2nd ed.; European Respiratory Society: Sheffield, UK, 2017; pp. 5–42.

3. World Health Organization. Coronavirus Disease 2019 (COVID-19) Situation Report; Technical Report March; World Health Organization: Geneva, Switzerland, 2020.

4. Rahaman, M.M.; Li, C.; Yao, Y.; Kulwa, F.; Rahman, M.A.; Wang, Q.; Qi, S.; Kong, F.; Zhu, X.; Zhao, X. Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. J. X-Ray Sci. Technol. 2020, 28, 821–839.

5. Yahiaoui, A.; Er, O.; Yumusak, N. A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines. Biomed. Res. 2017, 28, 4208–4212.

6. Hu, Z.; Tang, J.; Wang, Z.; Zhang, K.; Zhang, L.; Sun, Q. Deep learning for image-based cancer detection and diagnosis-A survey. Pattern Recognit. 2018, 83, 134–149.

7. American Thoracic Society. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. Am. J. Respir. Crit. Care Med. 2000, 161, 1376–1395.

8. Setio, A.A.A.; Traverso, A.; de Bel, T.; Berens, M.S.; van den Bogaard, C.; Cerello, P.; Chen, H.; Dou, Q.; Fantacci, M.E.; Geurts, B.; et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Med. Image Anal. 2017, 42, 1–13.

9. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. Annu. Rev. Biomed. Eng. 2017, 19, 221–248.

10. Wu, C.; Luo, C.; Xiong, N.; Zhang, W.; Kim, T.H. A Greedy Deep Learning Method for Medical Disease Analysis. IEEE Access 2018, 6, 20021–20030.

11. Ma, J.; Song, Y.; Tian, X.; Hua, Y.; Zhang, R.; Wu, J. Survey on deep learning for pulmonary medical imaging. Front. Med. 2019, 14, 450–469.

12. Rajaraman, S.; Candemir, S.; Xue, Z.; Alderson, P.O.; Kohli, M.; Abuya, J.; Thoma, G.R.; Antani, S.; Member, S. A novel stacked generalization of models for improved TB detection in chest radiographs. In Proceedings of the 2018 40th Annual International Conference the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 718–721.

13. Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G.; et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat. Med. 2019, 25, 954–961.

14. Gordienko, Y.; Gang, P.; Hui, J.; Zeng, W.; Kochura, Y.; Alienin, O.; Rokovyi, O.; Stirenko, S. Deep Learning with Lung Segmentation and Bone Shadow Exclusion Techniques for Chest X-Ray Analysis of Lung Cancer. Adv. Intell. Syst. Comput. 2019, 638–647.

15. Kieu, S.T.H.; Hijazi, M.H.A.; Bade, A.; Yaakob, R.; Jeffree, S. Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images. IAES Int. J. Artif. Intell. 2019, 8, 429–435.

16. Dietterich, T.G. Ensemble Methods in Machine Learning. Int. Workshop Mult. Classif. Syst. 2000, 1–15.