PREPRINT

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Transfer Learning for Lung Nodules Classification with CNN and Random Forest

Abdulrazak Yahya Saleh[1]*, Chee Ka Chin[2] and Ros Ameera Rosdi[1]

[1]*Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak (UNIMAS), 94300 Kota Samarahan, Sarawak, Malaysia*

[2]*Department of Electrical and Electronic, Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), 94300 Kota Samarahan, Sarawak, Malaysia*

ABSTRACT

Machine learning and deep neural networks are improving various industries, including healthcare, which improves daily life. Deep neural networks, including Convolutional Neural Networks (CNNs), provide valuable insights and support in improving daily activities. In particular, CNNs enable the recognition and classification of images from CT and MRI scans and other tasks. However, training a CNN requires many datasets to attain optimal accuracy and performance, which is challenging in the medical field due to ethical worries, the lack of descriptive notes from experts and labeled data, and the overall scarcity of disease images. To overcome these challenges, this work proposes a hybrid CNN with transfer learning and a random forest algorithm for classifying lung cancer and non-cancer from CT scan images. This research aims include preprocessing lung nodular data, developing the proposed algorithm, and comparing its effectiveness with other methods. The findings indicate that the proposed hybrid CNN with transfer learning and random forest performs better than standard CNNs without transfer learning. This research demonstrates the potential of using machine learning algorithms in the healthcare industry, especially in disease detection and classification.

Keywords: Convolutional Neural Network, CT scan, lung nodules, random forest, transfer learning

ARTICLE INFO

Article history:
Received: 17 May 2023
Accepted: 15 August 2023
Published: 03 November 2023

DOI: https://doi.org/10.47836/pjst.32.1.25

E-mail addresses:
ysahabdulrazak@unimas.my (Abdulrazak Yahya Saleh)
cheekachin@yahoo.com.my (Chee Ka Chin)
rosameerar@gmail.com (Ros Ameera Rosdi)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Lung cancer is a common and deadly disease in modern times. Cancer cells initially develop in the lungs but can spread to other organs, including lymph nodes and the brain (Rajadurai et al., 2020). Lung cancer was the most common and deadly cancer worldwide in 2018, making up 11.6% of all cancer cases and deaths (Bray et al., 2018). It was also Malaysia's second leading cause of death from cancer that year, after breast cancer (Bray et al., 2018), according to the World Health Organization. Men comprise approximately 16.6% of the patients, while women comprise 5.4%. In Malaysia, lung cancer is the third most prevalent type of cancer, affecting men more than women, according to the Malaysia National Cancer Registry Report 2012-2016 (Azizah et al., 2019). Today, Deep Learning is often used in medical image analysis (Alom et al., 2019; Arabahmadi et al., 2022; Salahuddin et al., 2022; Zakaria et al., 2022). Deep Learning is becoming increasingly popular and necessary for reliable and accurate results (Anderson et al., 2018). Deep learning simulates how the human brain processes data and recognizes patterns to make decisions. As technology and algorithms improve, machines can offer more accurate and reliable medical analysis. Identifying cells that are cancerous or malignant is essential for lung cancer therapy.

Literature Review

Deep learning techniques can analyze CT scan images and identify cancer cells at an early stage to prevent them from becoming fatal (Primakov et al., 2022; Thai et al., 2021). Deep learning image analysis applications on Computed Tomography (CT) scan images to aid in the detection of malignant cells prior to their development and lethality (Primakov et al., 2022; Thai et al., 2021). Deep Learning is ideally suited for image processing tasks, especially object detection and localization (Singh & Gupta, 2019). Deep learning, especially CNN, can achieve high accuracy with abundant data (Zhao et al., 2018). Convolutional Neural Networks (CNNs) need large and precise labeled training data, such as ImageNet, to operate effectively. Unfortunately, large datasets are often unavailable for medical images because of the high cost of expert explanations, ethical concerns, and the lack of images of diseases (Zhao et al., 2018). In addition, models with a high parameter count tend to overfit and cannot learn patterns when working with lesser datasets (Li et al., 2020). Most traditional CNN architectures begin with a high parameter count, making their performance highly dependent on the data size. Therefore, datasets that consist of only hundreds or thousands of instances are incompatible with standard CNN models trained on large-scale datasets like ImageNet (Keshari et al., 2018). It is an issue that researchers need to address if they want to improve the performance of the model when handling a huge volume of annotated data. A standard CNN may prove insufficient in medical imaging with small datasets, where datasets usually comprise a few hundred to thousands of data. Numerous studies have investigated various types of CNN for lung nodule detection, false positive reduction, and classification to address this issue (Halder et al., 2020; Forte et al., 2022; Nakrani et al., 2021; Sharif et al., 2020). However, Table 1 in this document focuses specifically on CNNs employing innovative strategies instead of conventional techniques or conventional CNNs.