Lecture Notes in Networks and Systems 1972

Abhishek Bhattacharya ·
Soumi Dutta · Debabrata Samanta ·
Paramartha Dutta ·
Adam Słowik *Editors*

Data Mining and Information Security

Proceedings of ICDMIS 2025, Volume 4

Springer

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Lung Cancer Detection Using CNN Model

Joktry Banerjee[1], Sinjini Rajguru[1], Nilufar Begum[1], Ankika Dey[2](✉), Anirban Ganguly[3], and Rajendrani Mukherjee[3]

[1] Department of CST, Institute of Engineering and Management, Kolkata University of Engineering and Management, Kolkata, India

[2] Department of CSE (AI and ML), Institute of Engineering and Management, Kolkata University of Engineering and Management, Kolkata, India
ankika067@gmail.com

[3] Department of CST and CSIT Institute of Engineering and Management, Kolkata University of Engineering and Management, Kolkata, India
anirban.ganguly@uem.edu.in

Abstract. The importance of timely and accurate detection of lung cancer cannot be overemphasized since it is a very common reason behind cancer-related deaths all over the world. We predict lung cancer from medical scans using Convolutional Neural Networks (CNN) that is a deep neural network architecture primarily seen in computer vision. It analyzes medical scans like CT scans or X-Rays to rapidly detect lung cancer. By analyzing them closely, it becomes possible to differentiate different classes of lung cancer, like adenocarcinoma, squamous cell carcinoma and healthy tissue. There are different stages in this process including breaking down the images into smaller units, pattern recognition and classification. We have shown that our approach is quick and deterministic providing a tool suitable for clinical diagnostics. The results show that it can support radiologists, by improving early detection rates and overall diagnostic accuracy. Ultimately, this study shows how advancement in imaging techniques can impact the healthcare and support in the fight against lung cancer.

Keywords: pattern recognition · lung cancer · X-Rays · Convolutional Neural Networks · deep neural network

1 Introduction

Lung cancer develops when lung cells multiply uncontrollably and form lumps, usually due to damage to the DNA from cigarette smoking or exposure to toxic chemicals. Dar upon cells that have started to become damaged and accumulate and essentially create an explosion of tumor growth. Without treatment the tumors disseminate throughout the lungs, compromising the lung function. Among malignancies, lung cancer is one of the most lethal and infectious, responsible for approximately 1.8 million yearly deaths worldwide. Its high death rate is mainly due to late diagnosis – early stages often show few or no symptoms. By the time when the symptoms are diagnosed, the disease has advanced to a stage where.

A. Bhattacharya et al. (Eds.): ICDMIS 2025, LNNS 1972, pp. 514–520, 2026.
https://doi.org/10.1007/978-3-032-25958-5_39

the treatment options are limited, which reduces survival chances. For this reason, early diagnosis is crucial to enhance treatment efficiency. Cigarette smoking, prolonged exposure to toxins is a major risk factor. CT scans and X-rays have improved lung cancer detection, but still radiologists rely on their experience to manually interpret these images. Early-stage tumors can be difficult to distinguish from noncancerous ones. However, the growing number of imaging studies are still the cause of extensive workload. Recent advancements in imaging technology have introduced new methods to resolve such situations. For example, CNNs are used to analyze complex CT scans and identify patterns in lung tissue. As it learns from a large, labelled scan, it can classify between the cancer affected and healthy tissues. This helps to detect lung cancer at early stages. This capability acts as a helping tool to radiologists. In this research, our aim is to build an automated, reliable and precise model for lung cancer detection by using CNNs. We will train this model on a set of lung CT scans to identify visual patterns for early lung cancer. The model will attempt to detect small, irregular nodules and other symptoms which are not usually visible, particularly when symptoms are minimal. The overall objective is to develop an AI based application that helps medical practitioners screen for lung cancer more efficiently and accurately. This system will help radiologists by flagging areas that need further review, reducing their workload. Detecting cancer in the earlier stages, improves treatment options and increases the chances of recovery. This paper is organized into sections that review related work, describe the functionality of the algorithm, and outline the overall workflow of our study.

2 Related Work

Cancer diagnosis in the later stage remains as one of the crucial reasons of deaths globally in each year. The traditional deep neural network models, like CNN, have been extensively applied in timely detection of lung cancer in imaging, where the researchers anticipate improving the diagnostic accuracy and reducing false alarm cases. In 2019, Ardila et al. [1] studied the prediction of lung cancer risk from a 3D CNN model derived from chest radiographs. Their model was able to detect lung nodules with strong accuracy and estimate cancer risk. In 2016, Setio et al. [2] proposed a deep multi-view CNN architecture that can classify lung nodules more effectively. They employed a method able to address common problems—e.g., avoiding deceptive positive findings when conducting lung cancer screening—by fusing multiple medical imaging data views. With their approach, the sensitivity of the deep learning model was enhanced in distinguishing between the benign and malignant nodules. Litjens et al. [3] also defined the task of CNNs in radiology, its applications and challenges encountered while using deep learning in the clinic. Their study showed CNNs can assist radiologists by simplifying feature extraction and lesion classification which reduces the manual workload. In 2017, Kumar et al. [4] implemented CNN based model to automatically detect lung cancer. They optimized CNN architectures to prevent false positives. In 2020, Li et al. [5] proposed residual networks (ResNets) for enhanced feature extraction from CT scans. Hatuwal and Thapa [6] showed the application of CNN in detecting lung cancer from "histopathological images" with an improvement in classification accuracy. Asuntha and Srinivasan [7] developed a deep learning-based framework for the spotting and identification of class of lung cancer using multimedia dataset analysis. Siegel et al. [8] provided a statistical analysis of cancer incidence and death, which shows the need of early detection method in survival improvement. Samala et al. [9] applied deep convolutional features in lung nodule identification using computed tomography images. As CNNs continue to evolve, they are becoming a powerful tool for the analysis of biomedical images, e.g. X-rays and CT scans, to enhance lung cancer detection. These deep learning models are able to automatically identify patterns for malignancy and help to improve accurate and efficiency. CNN based techniques have that potential to change the healthcare system by providing early diagnosis which improves the patient outcomes. Further advancements in deep learning, such as inclusion of hybrid models and explainable AI, will be useful in overcoming various challenges and ensuring accurate clinical diagnosis.

3 Methodology

3.1 Data Preprocessing

The dataset, used here, is taken from [10]. The dataset (Fig. 1) consists of 5000 medical images of lung tissue samples classified into three different cases:

- Lung Adenocarcinoma [10] (lung_aca) – This is a subtype of lung cancer that emanates in outer regions of mucus secreting glandular cells.
- Normal Lung Tissue [10] (lung_n) – Healthy lung cells without any cancerous characteristics.
- Lung Squamous Cell Carcinoma [10] (lung_scc) – This particular type of lung cancer, forms in the cells residing in the squamous tissues of the lung.

The data set was preprocessed to maintain similarity in image quality as well as size. To improve model generalization pre-processing of the image information was done through the stages involving flipping, rotation and contrast adjustment. Such images per class have been constructed from 250 images by passing through Data Augmentation. Python libraries make it easier to handle the data as well as basic and advanced processes. Pandas (library used to load the data frame in 2D array), Numpy (for vector processing), Matplotlib (to plot the visualizations), Sklearn and OpenCV (an open-source package with primary focus on image processing and handling) have been employed in this work.

3.2 Model Description

Hidden Layer Information: - The given image displays a summary of a CNN architecture by using the Sequential API available in TensorFlow/ Keras.

The proposed CNN architecture initiates with a 2D convolutional layer that receives input images of size 256x256 with 32 filters followed by a MaxPooling layer. The first convolutional layer reduces the spatial dimensions to half (128x128). Another Conv2D layer follows, increasing the filter count to 64, again followed by MaxPooling2D, reducing the size to 64x64. This pattern repeats with a third Conv2D layer, increasing the filter count to 128, and another MaxPooling2D, bringing the size down to 32x32. After the convolutional and pooling stages, the aggregated feature vectors are flattened into a

131072-dimensional vector, followed by a Dense (fully connected) layer with 256 neurons, allowing the network to understand high-level representations. At the later stage batch Normalization layer is introduced to ensure stable and fast learning.

The model has 33,684,291 total parameters, of which 33,683,523 are trainable and 768 are non-trainable (likely due to batch normalization). This model is likely de- signed for image classification or object detection tasks.

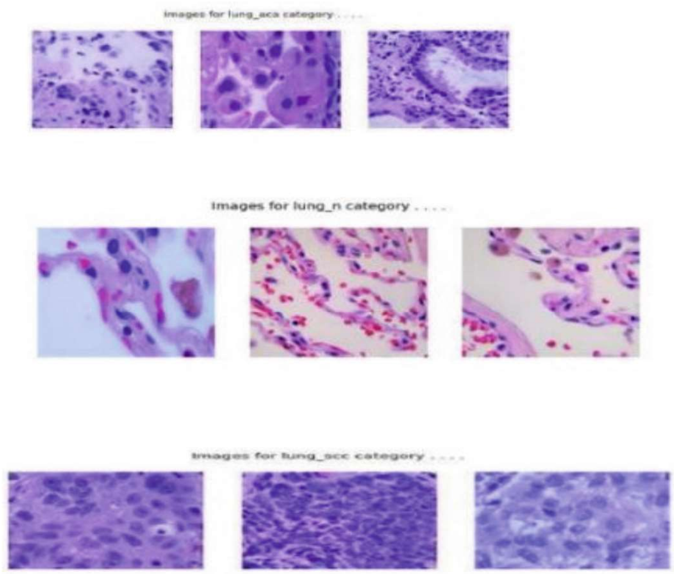

Fig. 1. Lung cancer data

3.3 Network Model

The 2D convolution is used to produce feature vectors from the full image. The reduced dimensionality after each convolution stage ensures specific feature map is obtained with filters applied at each convolution stage. The image is downscaled by the Pooling layer to decrease computation, afterwards the fully connected layer is used to get ultimate forecast. The network learns the best by incorporating backpropagation algorithm with gradient descent.

The combination of these layers enables CNNs to progressively analyze finer and more abstract details in an image. In each convolution layer significant number of filters are used to extract important feature maps. These feature map contains important information about vertical and horizontal edges, texture and shapes. This local connectivity and spatial sharing make CNNs ideal for analyzing images where spatial structure is crucial. Spatial dimension is being reduced by decimating the feature maps using several max pooling layers.

4 Result and Analysis

The model performance is good overall. Particularly for 'lung_n' and 'lung_scc' classes, it shows high true positive counts and low misclassifications.

However, the model struggles a bit with 'lung_aca' class, where a number of instances are misclassified as 'lung_scc'.

According to the Table 1, this model has a precision of 0.85 for 'lung_aca', which defines 85% of the instances predicted, were correct.

A recall of 0.60 for 'lung_aca' indicates that 60% of the actual 'lung_aca were correctly identified.

This model provides an F1-score of 0.70 for 'lung_aca' indicates a balance between precision and recall for that class.

Support number is 987, which is the number of samples of 'lung_aca' in the dataset.

The comprehensive accuracy of the model is 0.83, meaning it correctly classified 83% of the total instances across all classes.

The model is performing well overall, especially for the class lung_n, which has high precision and recall.

The lung_aca class, shows a significant gap between precision and recall, which indicates a scope for improvement, in identifying these instances without increasing more false positives.

Table 1. Performance parameters

	Precision	Recall	F1-score	Support
lung_aca	0.85	0.60	0.70	987
lung_n	1.00	0.93	0.96	977
lung_scc	0.72	0.97	0.82	1036
accuracy	_	_	0.83	3000
Macro avg	0.86	0.83	0.83	3000
Weighted avg	0.85	0.83	0.83	3000

In Fig. 2 the trend of diminishing loss is presented. The validation loss is indicated by val_loss line. In this specific graph, we observe that the validation loss increases significantly after the 4th epoch, and then decreases again. This behavior suggests that the neural network architecture is probably over fitting with the training image samples.

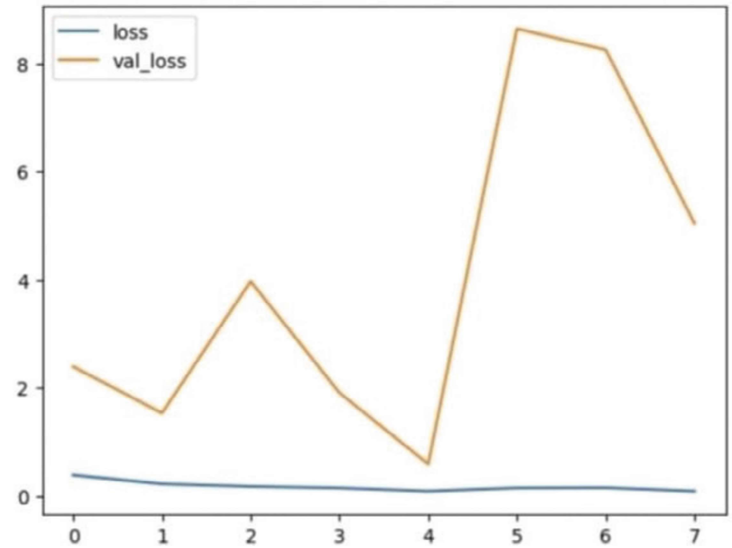

Fig. 2. Loss Indicators

5 Conclusion

Convolutional Neural Networks (CNNs) are transforming lung cancer detection by improving both the speed and accuracy of diagnosis. Conventional methods largely rely on radiologists to read scans manually, which can be laborious and fallible. For example, automated analysis of medical images like CT scans and histo-pathological slides is made possible through deep learning, particularly CNNs.

CNNs find trends in large datasets that can be difficult for humans to find. They have a high accuracy rate in detecting the type of lung nodule—whether it is a cancerous or non-cancerous growth and predicting how much at risk a patient is for cancer. The CNN based framework reduces the probability of any false alarm in the detection of cancerous tissues. Thus, it ensures any wrong diagnosis is avoided.

CNNs have the ability to detect cancer affected tissues at very early stage. When this artificial intelligence-based method assists radiologists, the diagnosis can become accurate and fast even for a large size of patient data volume. However, the need of labelling of large data volume creates challenge in many times. The CNN models are being continuously upgraded with advanced hybrid deep learning techniques. Nevertheless, the extensive usage of CNN based techniques are subject to clinical trials and necessary medical approvals.

In summary, CNNs offer an improved and efficient approach for the detection of lung cancer. They improve diagnostic accuracy, assist radiologists with their workflow and con- tribute to earlier diagnosis. With the constant refinement of CNN models and training mechanisms over time, AI facilitated technologies will become more reliable, cost- effective, and massively applicable in medical imaging and tumor detection.

References

1. Ardila, D., Kiraly, A.P., Bharadwaj, S., et al.: Lung cancer risk prediction using deep learning on chest radiographs: a retrospective cohort study. Nat. Med. **25**(6), 954–961 (2019)
2. Setio, A.A.A., Traverso, A., De Bel, T., et al.: Validation of a multi-view convolutional neural network for automated detection of pulmonary nodules in computed tomography scans. IEEE Trans. Med. Imaging **35**(5), 1160–1169 (2016)
3. Litjens, G., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
4. Kumar, D., Wong, A., Clausi, D.A.: Lung nodule classification using deep features in CT images. Proceedings of IEEE International Conference on Image Processing (ICIP), pp. 1192–1196 (2015)
5. Li, X., Zhang, S., Zhang, Y.: Deep residual learning for lung cancer detection in CT scans. Journal of Medical Imaging **7**(2), 1–12 (2020)
6. Hatuwal, B.K., Thapa, H.C.: Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images **68**(10), 21–24 (2020)
7. Asuntha, A., Srinivasan, A.: Deep learning for lung cancer detection and classification. Multimedia Tools and Applications, 1–32 (2020)
8. Rebecca, S., Deepa, N., Ahmedin, J.: Cancer statistics. A Cancer Journal for Clinicians **63**, 11–30 (2013)
9. Samala, R.K., et al.: Analysis of deep convolutional features for detection of lung nodules in computed tomography. Medical Imaging 2019: Computer-Aided Diagnosis **10950**, 109500Q (2019)
10. Kaggle: Lung and Colon Cancer Histopathological Images Dataset (n.d.). https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancerhistopathological-images