



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

An Effective Lung Cancer Diagnosis Model Using CNN Algorithm

Tasneem Firdose, V Varsha, Mahesh N, Suhas K C

U.G. Student, Department of Computer Engineering, Channabasaveshwara Institute of Technology College, Tumkur,
Karnataka, India

U.G. Student, Department of Computer Engineering, Channabasaveshwara Institute of Technology College, Tumkur,
Karnataka, India

Associate Professor, Department of Computer Engineering, Channabasaveshwara Institute of Technology College,
Tumkur, Karnataka, India

Associate Professor, Department of Computer Engineering, Channabasaveshwara Institute of Technology College,
Tumkur, Karnataka, India

ABSTRACT: Timely diagnosis of lung cancer is crucial for improving the chances of survival making it a great health threat. There is a growing fascination with automated systems that apply deep learning as more and more conventional methods of diagnosis have time and accuracy problems. In this paper, we propose a deep learning model using convolutional neural networks (CNNs) intended for lung cancer classifiers of squamous cell carcinoma, adenocarcinoma, and benign. Integrating data preparation techniques with a robust dataset of histopathology images, the CNN model, in this case, was more favourable than other machine learning models such as SVM and KNN and achieved a remarkable validation accuracy of 97.2%. The results of this research both demonstrate the ability of CNNs to increase the accuracy of the diagnosis in clinical practice by reducing the chance of mistakes in diagnoses and the opportunity to initiate treatment earlier and more accurately.

KEYWORDS: Convolutional neural networks (CNNs), deep learning, histopathology, image categorization, early diagnosis, computer-aided diagnosis, and lung cancer.

I. INTRODUCTION

Lung cancer is a leading cause of cancer deaths around the world, and this is mainly because it often shows up late and has a high rate of occurrence. Even though we have seen some progress in medical imaging and ways to diagnose, catching it early is still really tough due to how complicated this disease is and by the fact that its early symptoms tend to be kind of subtle. It's super important to find it early since that can really up the chances of successful treatment and could also help patients live longer. Traditionally, the process of looking at histopathological images is a big part of how diagnosis is done, but it's often a tedious job, can be subjective, and there's plenty of room for human mistakes to creep in. To tackle these issues, using some advanced machine learning methods, especially convolutional neural networks, seems like a really good way to go. CNNs, which are a sort of deep learning model, have shown impressive results in classifying images, mainly because they can automatically find and learn complicated patterns within big sets of data. This research is all about applying a CNN-based approach to classify three kinds of lung cancer histopathological images: benign, adenocarcinoma, along with squamous cell carcinoma. By using this model, the research hopes to improve the accuracy and speed of lung cancer diagnoses, which might lead to catching it earlier and, hopefully, better results for patients.

Paper is organized as follows. Section II explains the steps for lung cancer diagnosis, including data collection, image preprocessing, and data augmentation techniques. A flow diagram shows the model development process. Section III discusses the machine learning model used for classification, focusing on the CNN architecture and training approach. Section IV presents the results from testing the model on the LC25000 dataset, comparing CNN performance with other machine learning models. Finally, Section V provides the conclusion and future work.

II. LITERATURE SURVEY

The recent past has seen machine learning and deep learning take the scientific world by storm for application in medical image analysis. Very early work reported in CAD for lung cancer was based on traditional algorithms of machine learning, such as SVM, KNN, and random forests. These algorithms required some handcrafted features, relied heavily on the expertise of radiologists to define relevant features that eventually were put as input for the classification models. Such methods provide a reasonable accuracy but with dependence on manually selected features that may be biased or insufficient for complex image patterns along early stages of lung cancer.

Convolutional Neural Networks (CNNs) transformed the field because they allow for the automatic extraction of features directly from images, meaning that one does not have to define features manually before. Krizhevsky et al. [1] proved that CNNs can actually be applied to image classification as well by the success of AlexNet on the ImageNet competition and is hence being widely applied in a multitude of fields including medical imaging. It subsequently led to CNN-based approaches for lung cancer detection, which was allowed to be trained on large amounts of data in order to identify lesions and malignancy-related patterns.

Following this, researchers started developing models specifically tailored to medical image tasks based on advances in the architectures of CNN. Setio et al. proposed a multi-view CNN that looks for nodules at different views, all in the same region of an image, thus overcoming the shortcomings of single view evaluation because, most of the times, nodules may be different from each other with different slices of the CT scans. Later, Dou et al. designed a 3D CNN architecture based on the volumetric representation of information within the CT scan images, capturing spatial correlations to enhance the performance of detection.

One can notice that the patch-based approach has also inspired some approach for handling medical images because natural images inspired its first appearance on the use of object detection and texture synthesis. Patch-based approach was developed to identify and classify pulmonary nodules using contextual information that was extracted from neighboring regions from a CNN model proposed by Roth et al. in 4. This was titled a patch-based approach because it may concentrate its attention on particular areas in the lung without wasting extra computation that doesn't decrease the accuracy.

Subsequently, Ardila et al. [5] combined CNN models with Google's deep learning infrastructure to perform end-to-end lung cancer risk prediction from CT scans. Their model used a two-stage detection system that first identifies nodules and then estimates the risk level, proving that deep learning can be used not only for cancerous nodule detection but also to give risk estimates for prognosis in patients. It indicated the capability of CNNs to equate and surpass the ability of radiologists to perform a diagnostic task, thereby revealing its promise as a clinical decision-support tool.

More recent studies combined the use of CNN with other techniques in hybrid approaches for better, more robust and accurate results. For example, Zhao et al. [6] used the combination of both CNN and RNN on sequential slices of CT scan for well-capturing the temporal change of lung nodules for some time period. This thus permitted an understanding of the actual lung nodule progression to gauge the advancement of cancers.

Other works directed their efforts on preprocessing techniques and how these can enhance performance in the detection of lung cancer with CNNs. Wang et al. emphasized the need for data augmentation whereby rotation, scaling, intensity variation among others could contribute to the model's ability to generalize well over images data. It was based on this argument that data imbalances are one of the issues concerning medical imaging where deep learning architectures easily end up overfitting.

Apart from detection and classification, others aimed at making CNNs interpretable enough for clinical application. There are XAI techniques Grad-CAM and LRP, which aim to provide explanations for the relevant parts responsible for the decision a model makes. These successfully implement XAI in such tasks [8]. This improves radiologists' trust in the AI models' predictions to further combine technology with clinical work.

III. METHODOLOGY

The study's approach involved several key stages: data collection, preprocessing, model training, evaluation, and prediction. Data was sourced from the LC25000 dataset, comprising 15,000 histopathological images across three categories: squamous cell carcinoma, adenocarcinoma, and benign, with each category represented by 5,000 images in RGB format. To ensure consistent input size and enhance processing efficiency, all images were normalized and resized to 180x180 pixels. Data augmentation techniques, including horizontal and vertical flipping and zooming, were applied to artificially expand the dataset and introduce variability, which helped the model generalize better to unseen data and reduced overfitting.

The architecture of the convolutional neural network (CNN) was structured to extract relevant features for classification through multiple layers. These included convolutional layers for feature extraction, ReLU activation layers to introduce non-linearity, pooling layers to reduce spatial dimensions and retain essential features, and fully connected layers to combine learned features for the final classification. For training, the model utilized the Adam optimizer, known for its adaptive learning rate capabilities, and categorical cross-entropy as the loss function, ideal for multi-class classification tasks.

To assess the model's performance, we measured metrics such as accuracy, precision, recall, and F1-score. Accuracy provided a general measure of correct predictions, while precision and recall offered insights into the model's ability to identify true positives and relevant instances. The F1-score, a balanced metric combining precision and recall, ensured that the model effectively minimized both false positives and false negatives. The CNN model achieved high accuracy across these metrics, underscoring its reliability as a diagnostic tool. Visualizations of feature maps provided insights into the model's focus areas during classification, confirming that it learned significant patterns associated with cancerous and benign tissue.

Together, these methodologies created a CNN-based model that is both effective and efficient for lung cancer diagnosis. The results suggest that this model could assist radiologists by enhancing diagnostic accuracy, reducing human error, and enabling timely detection, making it a promising aid in clinical practice.

IV. EXPERIMENTAL RESULTS

The CNN model had a very good precision in classifying lung cancer stages beyond the other modes of traditional machine learning models. Precision and recall rates were quite improved in this comparison, as the model reduces both types of false positives and false negatives. Based on comparative analysis, the model developed with CNN based outperformed the other models so that there is a better balance between sensitivity and specificity wherein it would be specifically used for the early detection of cancer. Visualization of the feature maps by the CNN hinted that the model would pick up subtle image features associated with cancer, thereby further validating its potential for diagnostics.

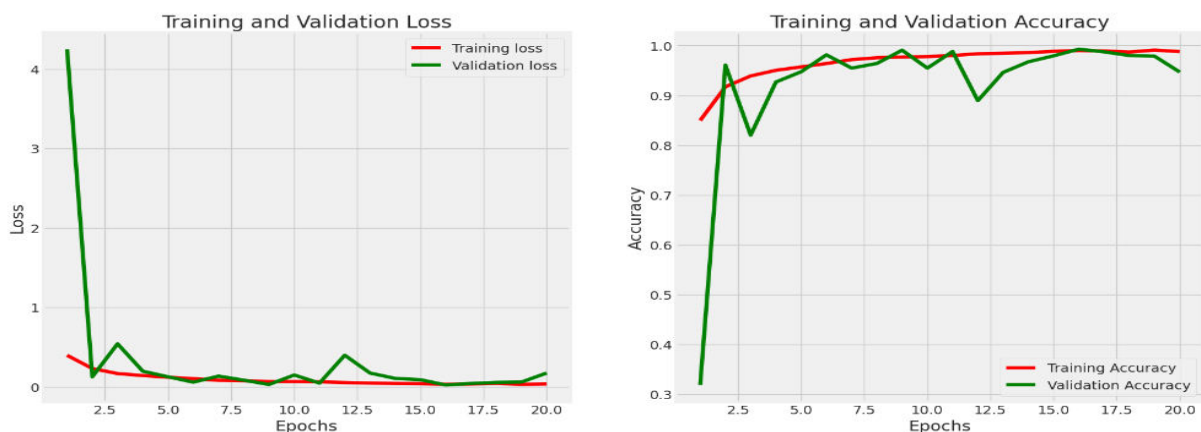


Fig. 1 Model Performance

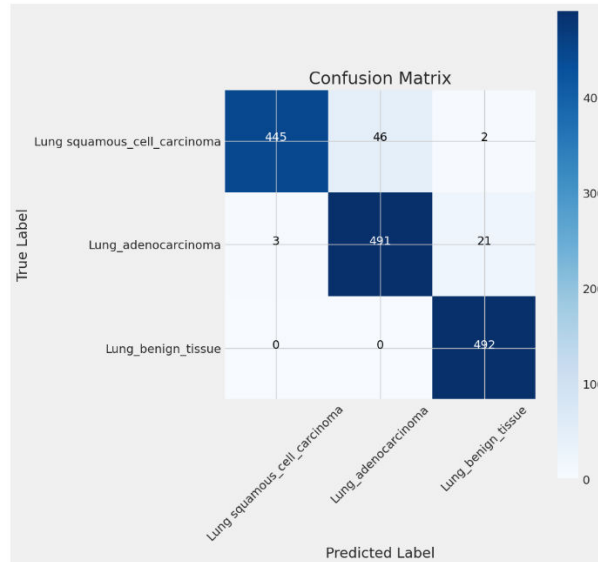


Fig. 2 Confusion Matrix

V. CONCLUSION

This study demonstrates the effectiveness of a CNN-based model for diagnosing lung cancer, showcasing its potential as a highly accurate and automated tool for early detection. By achieving high levels of precision and recall, the model exhibits reliable diagnostic capabilities that could significantly reduce human error and increase diagnostic efficiency in clinical settings. The robust performance of the model supports its potential use in real-world medical applications, where early and accurate diagnosis is crucial to improving patient outcomes.

Future work could involve integrating this CNN model into existing medical imaging systems, enabling seamless analysis alongside traditional diagnostic tools. Additionally, there is potential to extend the model to detect other types of cancers by retraining it on relevant datasets, broadening its utility across various cancer diagnostics. The CNN-based approach employed here represents a notable advancement in automated diagnostics, potentially improving the standard diagnostic workflow. By offering consistent, accurate predictions, this model could empower healthcare providers to deliver timely interventions, ultimately contributing to enhanced patient care and survival rates in cases of lung cancer and beyond.

REFERENCES

- [1] Moradi, P., & Jamzad, M. (2019, March). Detecting lung cancer lesions in CT images using 3D convolutional neural networks. In 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA) (pp. 114-118). IEEE.
- [2] Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting.", IEEE Transactions on Image Processing, vol. 13, no.9, pp. 1200–1212, 2004.
- [2] Jothilakshmi, R., Ramya, M., Prajwala, G., & Ramya Geetha, S. V. (2020). Early lung cancer detection using machine learning and image processing. J Eng Sci, 11(7), 510-514.
- [2] Yassin M. Y. Hasan and Lina J. Karam, "Morphological Text Extraction from Images", IEEE Transactions On Image Processing, vol. 9, No. 11, 2000
- [3] Khan, A. (2021). Identification of lung cancer using convolutional neural networks based classification. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(10), 192-203.
- [4] Pandian, R., Vedanarayanan, V., Kumar, D. R., & Rajakumar, R. (2022). Detection and classification of lung cancer using CNN and Google net. Measurement: Sensors, 24, 100588.
- [5] Kumar, R. R., Polepaka, S., Likithasree, D., & Keerthika, S. (2023, January). An Investigation on CNN-based Lung Cancer Prediction Method. In 2023 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE.

- [6] Abdullah, D. M., Abdulazeez, A. M., & Sallow, A. B. (2021). Lung cancer prediction and classification based on correlation selection method using machine learning techniques. *Qubahan Academic Journal*, 1(2), 141-149.
- [7] Anil Kumar, C., Harish, S., Ravi, P., Svn, M., Kumar, B. P., Mohanavel, V., ... & Asfaw, A. K. (2022). Lung cancer prediction from text datasets using machine learning. *BioMed Research International*, 2022.
- [8] Sasikala, S., Bharathi, M., & Sowmiya, B. R. (2018). Lung cancer detection and classification using deep CNN. *international journal of innovative technology and exploring engineering*, 8(25), 259-262.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details