Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 5-B, pp. 2011–2019
DOI : 10.47974/JIOS-2292

Speculation of lung cancer using deep learning sequential and CNN model

Ravi Raj Choudhary §
Himanshu Kumar Prashar †
Gaurav Meena *
Department of Computer Science
Central University of Rajasthan
Ajmer 305817
Rajasthan
India

Abstract

Lung cancer is the leading cause of cancer-related fatalities worldwide, owing mostly to its late detection and fast development. Early and precise detection is crucial to increasing survival rates. This research focuses on the conjecture and prediction of lung cancer using several deep learning models. Various DL architectures like CNN, GRU, LSTM and ANN are used for effective classification of lung cancer stages and types. These models were trained and evaluated using publicly available medical datasets, including radiological images and clinical data features. Feature extraction and selection techniques were applied to improve model performance and reduce overfitting. Experimental results demonstrate that deep learning models, particularly CNNs, outperform over sequential deep learning classifiers in terms of accuracy, sensitivity, and specificity. This research highlights potential of DL techniques for early speculation and diagnosis of lung cancer, paving the way for more efficient computer-aided diagnostic systems in the medical field.

Subject Classification: *68T07.*

Keywords: *Deep learning, GRU, Lung cancer, CNN.*

1. Introduction

One of the most aggressive and deadly types of cancer, lung cancer is responsible for a large portion of cancer-related fatalities worldwide.

§ *E-mail:* **raviraj@curaj.ac.in**
† *E-mail:* **2021imscs012@curaj.ac.in**
* *E-mail:* **gaurav.meena@curaj.ac.in (Corresponding Author)**

According to the World Health Organization (WHO), lung cancer contributes to nearly 1.8 million deaths annually, making it a critical public health concern. One of the main challenges in combating lung cancer is its late-stage diagnosis, which often limits treatment options and reduces patient survival rates. Early and accurate detection of lung cancer significantly increases the chances of effective treatment and patient recovery.

Recent developments in artificial intelligence (AI), especially in the fields of deep learning (DL) and machine learning (ML), have demonstrated enormous promise for improving medical diagnoses. Complex and high-dimensional medical data can be analyzed by these computer models to find hidden patterns and connections that conventional diagnostic techniques would find difficult to spot. The goal of this paper is to use deep-learning models to identify lung cancer early through testing models on DICOM images. By making lung nodule classification more accurate and reliable, this research hopes to help with faster and better diagnoses.

1.1 *Research Gaps*

Adenocarcinoma, squamous cell carcinoma, and large cell carcinoma are the three primary subtypes of lung cancer as well as normal lung pictures have been the focus of the majority of current research on lung cancer categorization using imaging data. However, they often overlook Small Cell Carcinoma, an aggressive and critical form of lung cancer. This oversight in model training restricts the breadth and practical application of the models developed. In contrast, the current study aims to bridge this gap by not only including the commonly studied subtypes and normal cases but also integrating Small Cell Carcinoma into the classification process. Additionally, this research seeks to improve the performance of the Convolutional Neural Network (CNN) model, enhancing its ability to distinguish among all four lung cancer subtypes and normal lung images, thus addressing a significant gap in current research.

Organization of the work: Section 2 shows various state of art models on lung cancer detection. Section 3 shows proposed architecture. Section 4 discusses the results obtained from the experiments, comparing them with recently published methods and analyzing how well the model performs in noisy environments. The job is comprehensively concluded in Section 5.

2. Literature Review

Employing a variety of CT and X-ray data, several research have employed deep learning techniques to identify lung cancer. Among the reviewed works, the models included CNNs, VGG16, ResNet, and UNet that were frequently combined with the methods of Grad-CAM, semantic multitask learning, or some ensemble models. LIDC-IDRI, MIMIC-CXR, and Kaggle CT scans were widely utilized as the datasets to classify the subtypes of the cancer such as Adenocarcinoma, Squamous Cell Carcinoma, and so on. The proven accuracies in these publications are between 84% and more than 99%, which proves the good prospects of AI in lung cancer classification and diagnosis. The published high levels of accuracy and reliability are on most reports. For examples, VGG-16, MobileNet, and custom CNN models can be used reasonably accurate with more than 90%, and some of the ensemble models can even be as high as 99%. In some research explainability methods such as Grad- CAM or SHAP have been incorporated in order to make more sense of model predictions. Author in paper [1] created an interpretable deep learning architecture, namely Hierarchical Semantic Convolutional Neural Network (HSCNN) to classify lung nodules based on their malignancy using CT images. The model was trained on 4,252 nodules in LIDC-IDRI, which included the following five semantically meaningful attributes (used clinically): margin, texture, subtlety, sphericity and calcification. HSCNN used a multitask learning framework 3D CNN architecture, which was able to predict malignancy together with the semantic features of the image.

Paper [2] determined the problem of explainable lung cancer classification through ensemble transfer learning solutions. The research was based on the use of the IQ-OTH/NCCD dataset that consists of 120 normal, 561 benign, 416 malignant CT images. Paper [3] study investigates the use of CNN for detecting lung cancer in CT scan images. The researchers utilize CNNs to categorize lung nodules as either cancerous or noncancerous. They implement various preprocessing methods to enhance the images and extract features, aiming to boost the model's performance. The CNN-based approach achieved an impressive accuracy rate of 95%, along with a precision of 89% and an F1-score of 90% for lung cancer detection. The training and evaluation dataset was obtained from Kaggle and mainly consists of annotated CT scan images that include both cancerous and healthy lung tissues. Paper [4] study explored different machine learning algorithms, including XGBoost, LightGBM, AdaBoost, logistic regression, and support vector machine (SVM), to detect lung cancer early among 5,000 patients in Dhaka, Bangladesh. An explainable method for lung cancer detection and localization from tissue images through convolutional neural networks was proposed in [5].

A deep learning approach has been developed for detecting lung cancer through CT images, which integrates 3D-VNet for segmenting lung nodules and 3D-ResNet for classification by in [6]. A model of explainable AI in lung cancer detection was developed [7] and was based on a modified Convolutional Neural Network (CNN) on CT data, which came from a Kaggle dataset. This combined model effectively minimizes false positives and enhances diagnostic accuracy. Paper [8] shows study on predicting lung cancer investigated a variety of machine learning models, including Extra Trees, Gradient Boosting, Random Forest, Light GBM, Decision Tree, and Logistic Regression. Feature selection was conducted using techniques such as Gain Ratio and importance derived from Random Forest. Paper [9] proposed CNN model was created to detect lung cancer in its early stages by analyzing CT scan images.

3. Proposed Work

Proposed architecture of the work showed in Figure 1. Model shows workflow of the lung cancer detection is described. The procedure begins at the stage of acquisition and data organization of lung imaging. Data Preprocessing, CNN modeling, Training and testing evaluations are performed at the level of classification.

3.1 *Data Collection & Organization:*

The Lung-PET-CT-Dx dataset (25%, 50000 images) from the Cancer Imaging Archive (TCIA), we have used the subset of the dataset which comprises the four main subtypes of lung cancer—adenocarcinoma, small cell carcinoma, large cell carcinoma, and squamous cell carcinoma—is the basis for this investigation (Include 10,000 each sub type). We also added normal lung images (10000 images).

3.2 *Data-preprocessing:*

Data preprocessing steps like data resizing and normalization are performed in this. In image resizing all input images are resized to a consistent dimension suitable for CNN input (e.g., 224x224 pixels). Pixel values are scaled (typically between 0 - 1) to reduce computational complexity and improve model convergence in normalization process.

3.3 *CNN Model Design:*

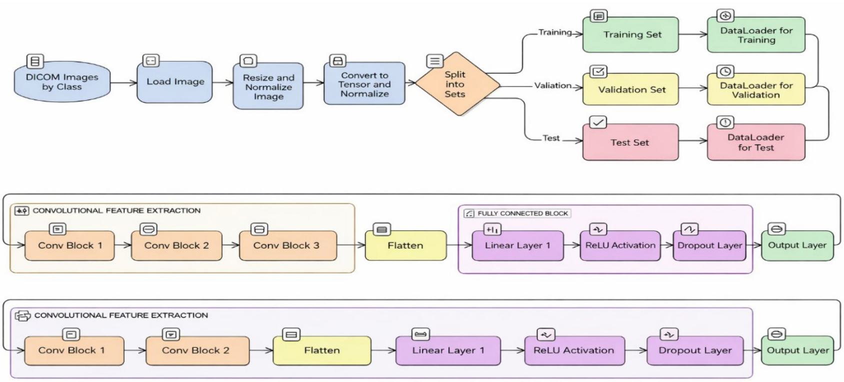

Figure 1

Proposed architecture of the work

Then deep learning model is constructed and trained to learn discriminative features. Performance evaluation follows the training part. Lastly, the optimized model is stored to make real-time prediction and give clinical application.

4. Result and Analysis

The bar chart in Figure 2 demonstrates the test accuracies of four different models: GRU, ANN, LSTM, and CNN. At 98.70% accuracy, CNN was found to capture spatial features in the data more successfully compared to the other models. LSTM achieved 90.78%, showing that recurrent structures are useful for recognizing sequences. ANN showed good performance, modeling nonlinear relationships with 84.89%, while GRU had the weakest performance at 74.49%. In general, the results highlight that CNNs are well-suited to this task, under- scoring the importance of selecting the appropriate model architecture for the problem.

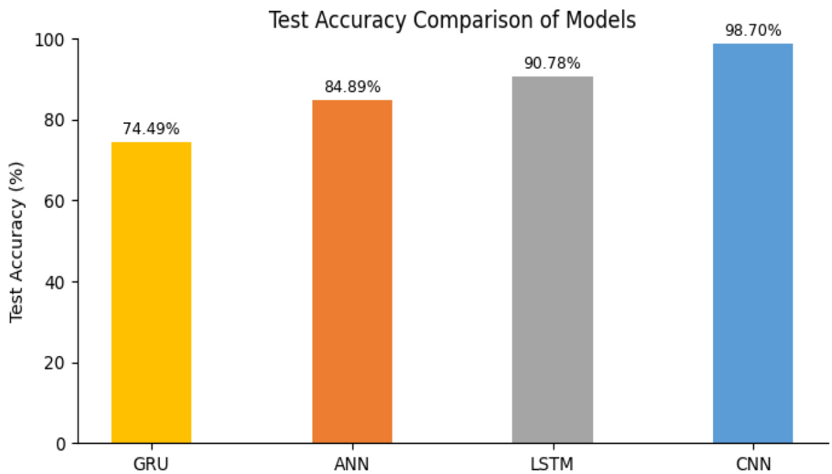

Figure 2

Test accuracy comparison of models

4.1 *Train vs Validation Accuracy*

The training and validation accuracy of our CNN across 20 epochs is displayed in Figure 3. Both training and validation accuracy increase and reach 99.5% and 98.7%, respectively. In other words, the model works well and is not overfitting.

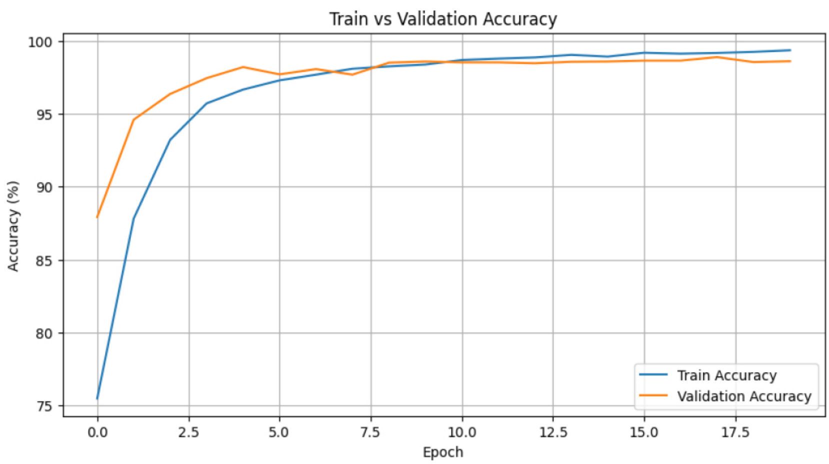

Figure 3

Train vs Validation Accuracy for CNN mode

a. Train vs Validation Loss

The CNN model's training and validation loss over 20 epochs is shown in Figure 4. Effective learning is indicated by a steady decrease in training loss. With just minimal overfitting, the validation loss first decreases and then varies somewhat, indicating that the model generalizes well. During training, the CNN shows consistent convergence overall. The Table 1 and 2 shows the results.

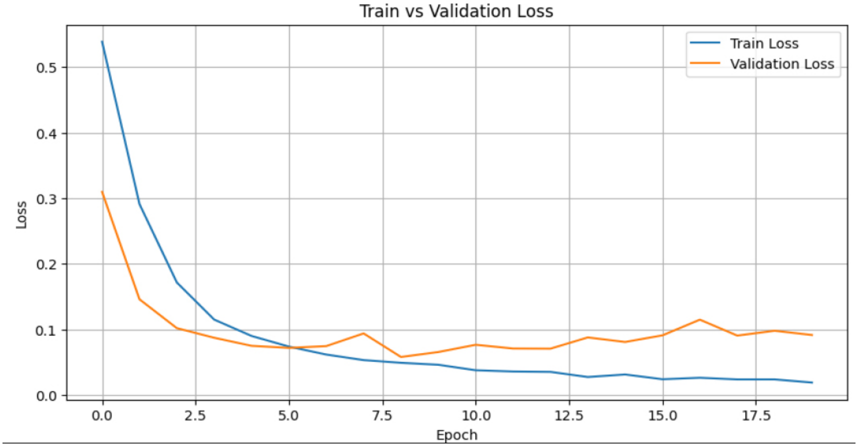

Figure 4

Train vs Validation Loss for CNN model

Table 1

Classification Report of CNN Model on Lung Cancer Dataset

Class	Precision	Recall	F1-Score	Support
Adenocarcinoma	0.98	0.97	0.98	1000
Large Cell Carcinoma	1.00	1.00	1.00	999
Small Cell Carcinoma	0.98	1.00	0.99	1000
Squamous Cell Carcinoma	0.98	0.97	0.97	1000
Normal Lungs	1.00	1.00	1.00	991
Accuracy	98.7	0.98	0.98	4990

Table 2

Shows comparative good results as compare to state of art methods.

Athor's	Dataset	Accuracy (%)
Jacob and Menon [10]	Lung-PET-CT-Dx	95.5
Saad and Choi [11]	Lung1	0.78
Liu et al. [12]	Lung1	0.86
Marentakis et al. [13]	TCIA	0.74
Our approach	Lung-PET-CT-Dx **(Sub Set)**	98.7

5. Conclusion

This study presents an in-depth evaluation of various machine learning and deep learning models for lung cancer detection, focusing on image-based classification. Among the models tested—GRU, ANN, LSTM, and CNN—the Convolutional Neural Network (CNN) demonstrated superior performance, achieving a remarkable test accuracy of 98.70%. The CNN was the best model for this challenge because of its capacity to efficiently extract and learn spatial characteristics from medical images. High validation accuracy and low loss values across several epochs demonstrate that the CNN model generalizes well without experiencing considerable overfitting, as confirmed by the training and validation accuracy trends. The model's resilience is further supported by the classification report, which consistently displays strong precision, recall, and F1-scores across all five lung cancer classifications and normal lung cases. Future work may explore the integration of multi-modal data sources (e.g., combining imaging with clinical data), hyperparameter optimization, and deployment of the trained model into clinical decision support systems for real-time lung cancer screening and diagnosis

References

[1] D. Aberle, W. Hsu, A. A. T. Bui, S. Shen, and S. X. Han, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Syst. Appl.*, vol. 128, pp. 84–95 (2018).

[2] S. Alzahrani, "Explainable lung cancer classification with ensembled transfer learning of VGG16, ResNet50 and InceptionV3 using Grad-CAM," *BMC Med. Imaging*, vol. 24, art. no. 176 (2024).

[3] R. Maheshwari and V. Kapoor, "Predicting the NSE stock index trends considering global financial variables and ARIMA model," *J. Stat. Manage. Syst.*, vol. 25, no. 7, pp. 1513–1522 (2022).

[4] M. Bhuiyan, I. K. Chowdhury, M. Haider, and A. H. Jisan, "Advancements in early detection of lung cancer in public health: A comprehensive study utilizing machine learning algorithms and predictive models," *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 1, pp. 113–121 (2024).

[5] F. Mercaldo, M. G. Tibaldi, L. Lombardi, L. Brunese, A. Santone, and M. Cesarelli, "An explainable method for lung cancer detection and localization from tissue images through convolutional neural networks," *Electronics*, vol. 13, no. 7, p. 1393 (2024).

[6] L. J. Crasta, R. Neema, and A. R. Pais, "A novel deep learning architecture for lung cancer detection and diagnosis from computed tomography image analysis," *Healthcare Anal.*, vol. 5, p. 100316 (2024).

[7] A. El-Latif, P. Plawiak, M. A. ElAffendi, A. Ateya, M. Hammad, and G. Ali, "Explainable AI for lung cancer detection via a custom CNN on CT images," *Sci. Rep.*, vol. 15, art. no. 12707 (2025).

[8] A. Goyal, R. K. Shrivastava, M. Agarwal, and N. Joshi, "Enhanced prediction of lung cancer using machine learning," in *Proc. 7th Int. Conf. Contemp. Comput. Informatics (IC3I)*, vol. 7, pp. 1478–1483 (2024).

[9] N. K. Karthikeyan, S. S. Ali, and R. V. Sekhar, "Lung cancer classification using CT scan images through deep learning and CNN-based model," in *Proc. Int. Conf. Adv. Data Eng. Intell. Comput. Syst. (ADICS)*, pp. 1–5 (2024).

[10] C. Jacob and G. C. Menon, "Pathological categorization of lung carcinoma from multimodality images using convolutional neural networks," *Int. J. Imaging Syst. Technol.*, vol. 32, no. 5, pp. 1681–1695 (2022).

[11] M. Saad and T. S. Choi, "Deciphering unclassified tumors of non-small-cell lung cancer through radiomics," *Comput. Biol. Med.*, vol. 91, pp. 222–230 (2017).

[12] J. Liu, J. Cui, F. Liu, Y. Yuan, F. Guo, and G. Zhang, "Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model," *Med. Phys.*, vol. 46, no. 7, pp. 3091–3100 (2019).

[13] P. Marentakis, P. Karaiskos, V. Kouloulias, N. L. Kelekis, S. Argentos, N. Oikonomopoulos, and C. Loukas, "Lung cancer histology classification from CT images based on radiomics and deep learning models," *Med. Biol. Eng. Comput.*, vol. 59, no. 1, pp. 215–226 (2021), doi:10.1007/s11517-020-02302-w

Received April, 2025