IES International Journal of Multidisciplinary Engineering Research

Lung Cancer Detection Using AI

Nowreen Mohamedmon [1], Sahal M A [2], Nibin P N [3], Sreehari E [4], Athira Bose [5]

[1,2,3,4] *Student, Department of Computer Science and Engineering, IES College of Engineering, Thrissur, Kerala, India*

[5] *Assistant Professor, Department of Computer Science and Engineering, IES College of Engineering, Thrissur, Kerala, India*

Email_id: nourinmohamedmon17@gmail.com, sahalma448@gmail.com, nibinpn5@gmail.com, sreehariplusone@gmail.com, athirabose@iesce.info

Abstract

The project's goal is to improve the detection and classification of lung cancer from medical imaging through advanced artificial intelligence techniques, particularly with Convolutional Neural Networks (CNNs). The goal is to support early diagnosis, enhance the accuracy and trustworthiness of clinical evaluations, save time, make diagnostic tools more accessible, and consequently, improve patient outcomes. The system architecture includes designing a machine learning model structured around CNNs; special attention is given to the steps of data gathering, data cleaning, feature selection, and model training. The main focus of the proposed approach is the application of advanced image analysis methods on top of the classifier based on CNNs which is used to classify the images. Under the project name “Disease Detection by Artificial Intelligence,” the model provides a diagnostic accuracy rate of 94.297%, demonstrating the potential of AI-based systems to autonomously formulate and execute disease diagnosis protocols in health care frameworks.

Keywords: CNN, Classification, Images, attention, analysis.

DOI: https://doi.org/10.5281/zenodo.15486578

1. Introduction

Detective AI has enlisted the power of modern advancements in AI and deep learning. The field of medicine can now rest easy knowing that AI will actively pursue unconventional and automated non-invasive methods to assist with triage checklists; the goal will always remain to improve the value of the initial attention. Chief suspects: millions of deaths and fatalistic nature of lung cancer pose extreme obstacle every year—the thin-eyed killer being explosive growth on lung epithelium tissues. Employees of public health are all too aware that such perpetrators often thrive in silence, and only inconvenient checks like CT scans, core tissue sampling via biopsy, and blood workish scrutinization reveal the infection. Progonosis usually invalidates hope as the stealth-like symptoms neatly swing under the radar, slicing vital nutritional arteries. They have, however, realized that using a deep convoluted neural network—alongside its rigid suboe and triological counterparts—provides new hope in the seemingly endless realm of biometric imaging for not only spotting, but fighting lung cancer. With this knowledge at hand, it is needless to say that this model and the capabilities imbued into it will give radiology technicians precise guidance to determine whether a patient’s CT scan depicts healthy anatomy or advanced cancer pathology.

2. Methodology

The system aims to automatically detect Lung cancer from CT Scan images by classifying the images into one of the following four categories: Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma and Normal. With the aid of deep learning algorithms, especially CNNs or Convolutional Neural Networks, the system intends to assist in timely as well as accurate diagnosis.

As the initial step of the model, a pre-labeled dataset of lung CT images containing all four categories is collected; every category is equally represented. In order to maximize versatility as well as uniformity across the model, certain uniformity enhancing preprocessing measures are taken. The images are first converted to a standard size, transformed into grey scale so that they have a single color channel instead of three which further decreases the processing power needed, scaled to a pre defined size, and cleared of unwanted pixel information to enhance the purity of the image; also known as noise reduction. In addition to this, data augmentation procedures like rotation, flipping, and scaling are also done in order to increase the dataset further.

After the data enhancement procedure is completed, the images are inserted into a CNN based classification model. The classification model contains a number of layers for extracting hierarchical features, including several convolutional layers through which the features will be identified, pooling layers with will reduce the scale of the data, dropout layers with the aim of offsetting overfitting, and fully connected layers.

2.1 Lung Cancer Detection System Architecture

The architecture of our system for cancer detection utilizes images from CT scans as biomedical inputs (features) that are examined using a CNN model capable of placing each case in one of four classes: Normal, Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma.

The procedure begins with medical image preprocessing and feature extraction aided by the convolutional layers of the network. The derived features undergo classification by several hidden layers of the network and is outputted through an output layer. After going through the training session, the model is sophisticated enough to make predictions by appropriately classifying all input images to their respective probable class labels with a computed likelihood of confidence ratio indicating certainty to the correct class per submission.

Also, the system is attuned with a straightforward web-based application enabling users to upload CT images and have them analyzed without interface interaction. The website's services are also automated, meaning it will provide answers with classification while keeping track of all performed actions within the framework including attributes as well as dates and times of each interaction which greatly boosts and supports the patient record archival system rich for longitudinal medical study and longitudinal medical research.

Through the application of artificial intelligence techniques, deep learning algorithms, advanced medical image processing, and user-friendly interfacing, the system significantly boosts the speed and accuracy of medical diagnosis.

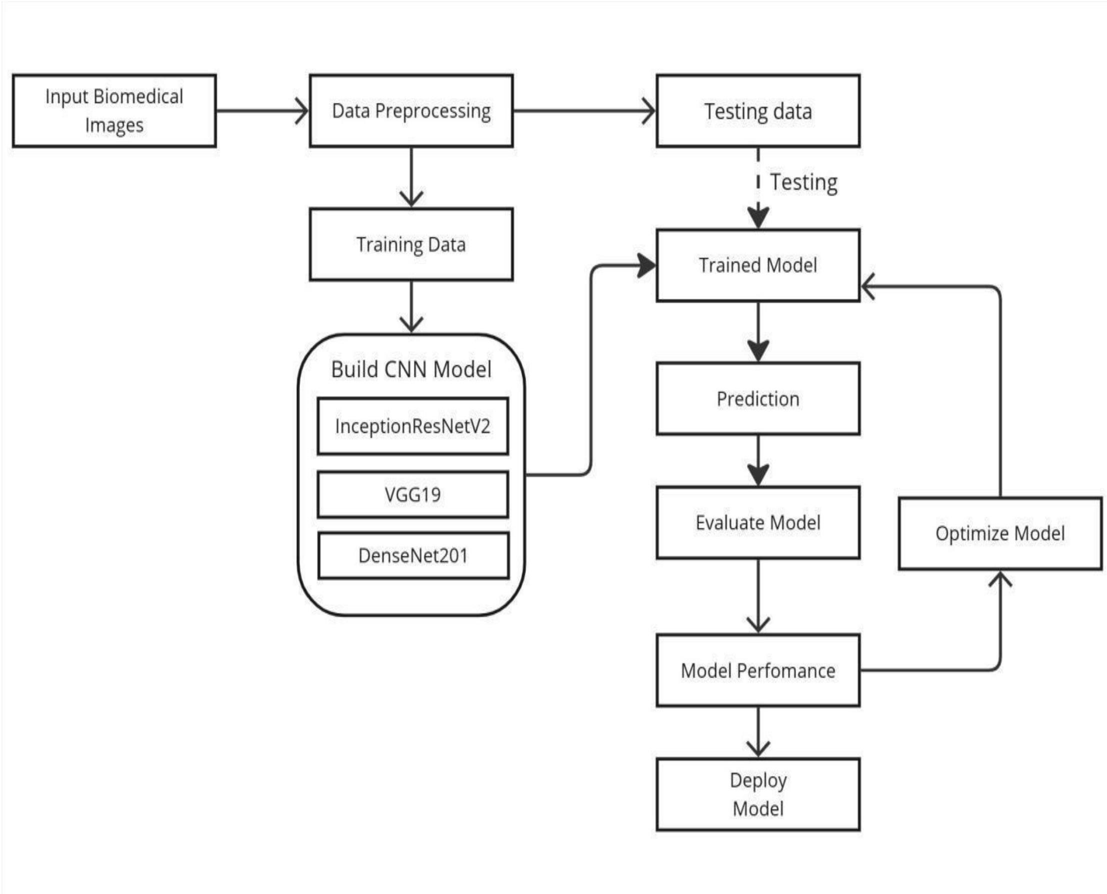

Figure 1: System Architecture for Lung Cancer Detection

2.2 Deployment and Frontend

The backend lung cancer detection system functions on a light and robust web application framework named Flask. Under this configuration, Flask acts as the middleman for communication between the users and the deep learning model. The web interface is made easy to use for users like radiologists, physicians, and researchers so that they can upload the images of the CT scan with ease and obtain real-time predictions.Once an image is uploaded through the interface, Flask processes the request, applies necessary preprocessing tasks, and makes the image run through the trained CNN model based on deep learning principles. The system then provides a classification output and a confidence score, indicating the probability of the uploaded image being in one of the four categories: Normal, Adenocarcinoma, Squamous Cell Carcinoma, or Large Cell Carcinoma.The solution is also deployable in a flexible manner, including local hosting on hospital networks as well as cloud-based deployments for remote safe access. The framework utilizes major Python packages such as TensorFlow, Keras, NumPy, Pandas, Seaborn, and Matplotlib, which are crucial for model running and result display. Everything from CT image processing to feature extraction and prediction is handled server-side through Flask, making it highly performance-oriented and secure.

This deployment model makes the system cost-efficient, simple to handle, scalable, and maintenance-free, yet still delivers a seamless user experience.

3. Module Description

The system for detecting lung cancer is based on a number of interlinked modules, every one of which is responsible for executing particular activities connected with image processing, classification of the disease, and diagnostic support. The CT Image Input Module provides access for users to load lung CT scan images using a web-based interface. The Preprocessing Module cleans the images by resizing, denoising background noise, and transforming the images into the model analysis optimum format. The Disease Detection Module uses a Convolutional Neural Network (CNN) to analyze the CT images and classify them into one of four categories: Normal, Adenocarcinoma, Squamous Cell Carcinoma, or Large Cell Carcinoma. The Result Display Module returns the classification result and a confidence score through the web interface, thus giving important support to doctors at the time of diagnosis. Finally, the Data Logging and Storage Module stores the uploaded images, respective predictions, and timestamps to allow future retrieval, further analysis, and continuous improvements to the system.

4. Implementation

The system architecture of the lung cancer diagnosis system is made up of multiple functional components that collaborate to support efficient image processing, precise disease classification, and strong diagnostic assistance. The CT Image Input Module acquires lung CT scans from the database. The Preprocessing Module normalizes image dimensions, eliminates background noise, and reconverts the images into the ideal form for the model. The Disease Detection Module uses a Convolutional Neural Network (CNN) to identify the CT scan images and predict them as being in one of four categories: Normal, Adenocarcinoma, Squamous Cell Carcinoma, or Large Cell Carcinoma. The Result Display Module presents the prediction results via web interface, indicating both the category predicted and confidence score to support healthcare professionals to make a diagnosis. Lastly, the Data Logging and Storage Module logs input images, their respective predictions, and respective timestamps, allowing for effective retrieval for future use, for further analysis, and for ongoing system improvement.

4.1 Tools and Technologies Used

Category	**Tools & Technologies**
Programming Language	Python
Framework	flask
IDEs	PyCharm, Vs Code
Deep Learning Libraries	TensorFlow,Keras,NumPy,Pandas, Seaborn,Matplotlib
Frontend	HTML, CSS, JavaScript (optional enhancements with Bootstrap)

Image processing	OpenCV,PIL(Python Imaging Library)

Table 1: Tools and technologies used

4.2 Algorithm Details

The algorithm employs a Convolutional Neural Network (CNN) for the detection of lung cancer from CT scan images at an initial stage. The training dataset is pre-labeled into four classes: Normal, Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma. Every image is processed with a set of preprocessing operations such as resizing, grayscale transformation, filtering, and normalization to transform the image into a suitable format for analysis. The processed images are fed into a CNN capable of auto-extracting texture-pattern and structural-irregularity related features. The network's classification layer generates a predicted label along with a confidence metric reflecting how confident the network is in its prediction. The results are integrated into a web-based interface constructed with Flask, which displays uploaded images together with their classifications and, optionally, visual heatmaps of areas of interest. This automated methodology enhances classification quality and processing, decreases the burden of manual viewing of images, and enables accelerated clinical decision-making.

5. Results and Discussion

The model was tested on a labeled test set of CT scan images, covering four classes: Normal, Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma. To quantify the performance of the model, important metrics like accuracy, precision, and recall were used.

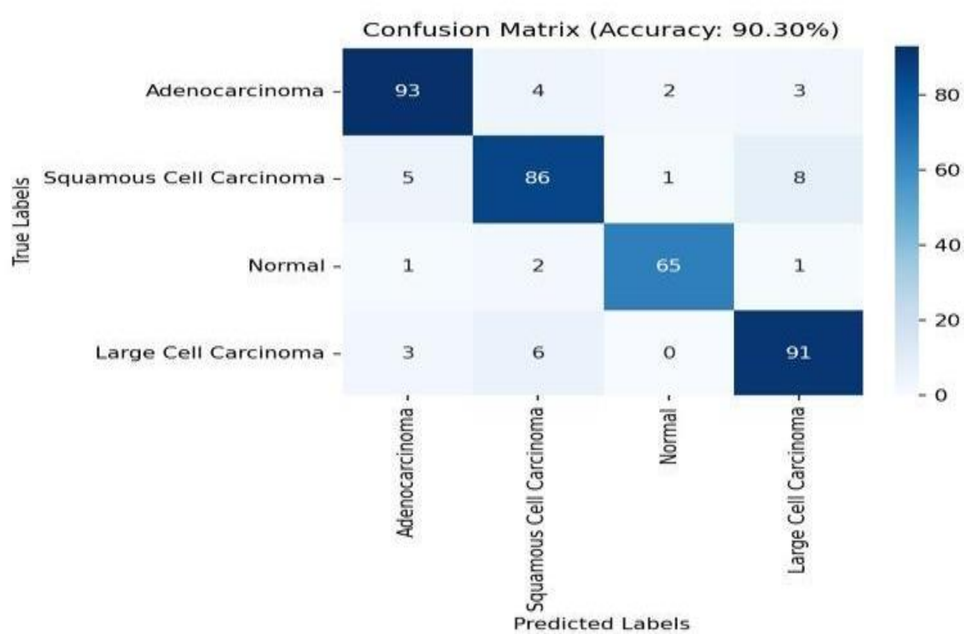

Figure 2: Confusion Matrix

The confusion matrix (as previously mentioned) is a clear summary of the classification performance of each category by the model. All cells in the matrix represent how many times there was a match or mismatch of the predicted and actual classes.

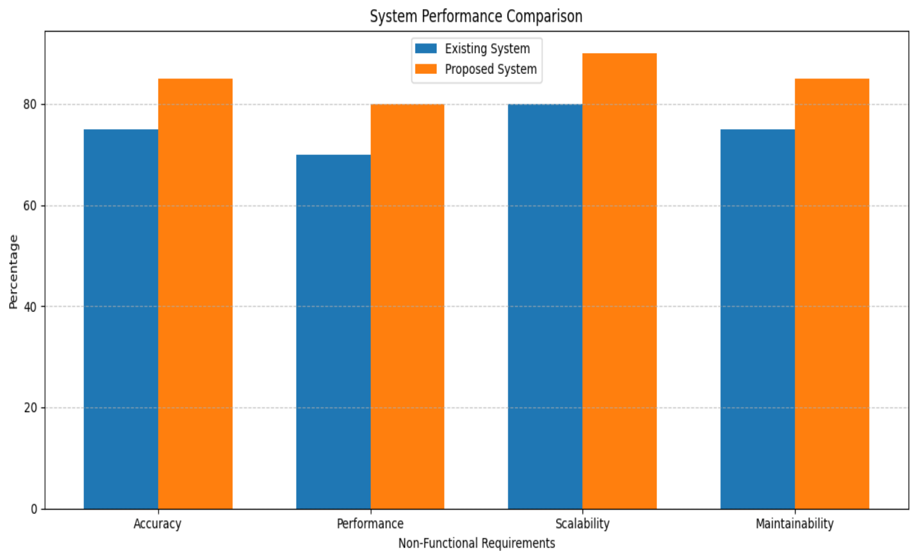

Figure 3: System performance comparison

6. Conclusions

This study introduces a lung cancer detection system that utilizes a Convolutional Neural Network (CNN) to support image classification in order to facilitate early and precise diagnosis. The model is trained to distinguish between four types: Normal, Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma. Utilizing deep learning algorithms, the system scans medical images to detect symptoms, if any, of lung cancer and provides real-time diagnostic results using a web application. This degree of automation improves diagnostic accuracy, aids radiologists in clinical decision-making, accelerates the initiation of treatment plans, and enhances patient outcomes through timely intervention.

7. References

[1]. Rajpurkar, P., Irvin, J., Zhu, K., et al. "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning." arXiv preprint arXiv:1711.05225, 2017.

[2]. Jiang, Z. "Chest X-ray pneumonia detection based on convolutional neural networks." In Proceedings of the 2020 International Conference on Big Data, Artificial Intelligence and IoT Engineering (ICBAIE), IEEE, pp. 341–344, 2020.

[3]. Chen, H., Dou, Q., Yu, L., Qin, J., and Heng, P.A. "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images." NeuroImage, vol. 170, pp. 446–455, 2018.

[4]. Simonyan, K., and Zisserman, A. "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv preprint arXiv:1409.1556, 2014.

[5]. Saba, S.M., Khan, M., and Zubair, M. "A CNN based automated lung cancer detection using high resolution CT scan images." In Proceedings of the 2020 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE, pp. 1–6, 2020.

[6]. Krizhevsky, A., Sutskever, I., and Hinton, G.E. "ImageNet classification with deep convolutional neural networks." Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[7]. Sharma, A., and Saba, M. "Lung cancer detection using artificial intelligence techniques: A comprehensive review." Computer Methods and Programs in Biomedicine, vol. 197, p. 105183, 2020.

[8]. Hossain, T., Shishir, F.S., Ashraf, M., Al Nasim, M.A., and Shah, F.M. "Brain tumor detection using convolutional neural network." In Proceedings of the 2019 International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, pp. 1–6, 2019.

[9]. Gandhi, C.S., Shinde, S.B., and Suryawanshi, D.M. "An efficient deep learning model for lung cancer classification using CT images." Procedia Computer Science, vol. 167, pp. 1810–1819, 2020.

[10]. Wang, J., Song, Y., Chen, L., Yang, L., and Jiang, M. "Lung cancer detection using deep learning methods: A systematic review." IEEE Access, vol. 9, pp. 112579–112591, 2021.