LUNG CANCER DETECTION USING CONVOLUTIONAL NEURAL NETWORK (CNN)

Manasa Sairi
700758834
Masters in Computer Science
University of Central Missouri,
Missouri, United States.
Mxs88340@ucmo.edu

Bhavani Dasari
700759340
Masters in Computer Science
University of Central Missouri,
Missouri, United States
bxd93400@ucmo.edu

Anwesh Ammula
700745034
Masters in Computer Science
University of Central
Missouri, United States
axa50340@ucmo.edu

Yalavarthi rohith
700745581
Masters in Computer Science
University of Central Missouri,
Missouri, United States.
rxy55810@ucmo.edu

***Abstract* -** Automatic detection of defects and abnormalities in CT scan images has gain lot of importance in the recent times as because of the deteriorating quality of air almost everyone is facing one or the other type of lung related issues thus it is of esteem importance to study the computed tomography (CT) images of the pulmonary nodules to efficiently treat the diseases like lung cancer. The use of machine learning is an efficient way to distribute the work of doctors and process the large amount of data to produce accurate results on the go. Three phases of CT image pre- processing, Deep Learning, and Convolutional Neural Network use make up the diagnosis approach. The pre- processing converts raw data into usable form and deep learning algorithm assigns weight to the data, in the last stage CNN is use to conclude the health status of the lung, i.e. normal or abnormal.

Keywords: Lung Cancer**,** Deep Learning, Convolutional Neural Network , Pre- processing, Gradient, Medical imaging, Diagnosis.

INTRODUCTION

Lungs serve as the vital support system for the circulation of gaseous materials in our bodies, ensuring the exchange of oxygen and carbon dioxide necessary for cellular function. Any damage to this intricate system can lead to fatality or lifelong impairment of normal physiological functions. Among the various threats to lung health, lung cancer stands out as one of the most widespread and deadly malignancies worldwide. Cancer, characterized by the uncontrolled growth and spread of abnormal cells, poses a significant risk to human health, particularly when it affects vital organs like the lungs.

Research conducted by Australian Medical Institutes underscores the severity of lung cancer, revealing a sobering survival rate of just 15% for patients diagnosed over a five-year period. The prognosis of lung cancer is heavily influenced by the stage at which it is detected, with late-stage diagnoses significantly reducing the likelihood of successful treatment and long-term survival. In fact, expert oncologists can only identify the disease at advanced stages in approximately 75% of cases, highlighting the urgent need for more effective diagnostic approaches.

Git-Hub Link: https://github.com/Bhavani700759340/NEURALNETWORKSANDDEEPLEARNINGPROJECT

Machine learning algorithms offer a promising avenue for addressing this challenge by enabling the early detection of lung cancer with remarkable accuracy. Studies have shown that these algorithms can achieve detection rates ranging from 30% to 97% for pulmonary nodules, the precursor lesions that often develop into lung cancer. By leveraging the power of neural networks, specifically convolutional neural networks (CNNs), researchers have made significant strides in distinguishing cancerous cells from normal tissue, laying the groundwork for AI-based cancer detection systems.

This paper focuses on the application of CNNs in classifying lung tumors as malignant, representing a critical step towards improving the diagnosis and treatment of lung cancer. CNNs have emerged as a cornerstone technology in medical image analysis, capable of extracting relevant features and patterns from complex imaging data. By harnessing the inherent ability of CNNs to identify important features within images, researchers aim to develop a robust and accurate method for classifying lung tumors, paving the way for more effective cancer diagnosis and management.

In addition to their prowess in feature extraction, CNNs offer the advantage of transfer learning, a technique that leverages pre-trained models developed by researchers on large-scale image datasets. By fine-tuning these models for specific tasks, such as lung tumor classification, researchers can expedite the training process and achieve higher levels of accuracy. This approach not only reduces the computational burden of training CNNs from scratch but also enhances their performance by capitalizing on the knowledge encoded in pre-trained models.

Overall, the integration of CNNs into the field of lung cancer diagnosis holds immense promise for improving patient outcomes and reducing mortality rates associated with this devastating disease. By combining the power of machine learning algorithms with the expertise of medical professionals, researchers aim to revolutionize the early detection and treatment of lung cancer, ultimately saving lives and improving the quality of life for affected individuals.

MOTIVATION

Detecting lung cancer through Convolutional Neural Networks (CNNs) represents a groundbreaking advancement in cancer diagnostics and therapeutics. Lung cancer, often diagnosed at advanced stages, significantly diminishes prognoses and treatment efficacy. However, CNNs exhibit exceptional capability in discerning subtle abnormalities within medical imaging scans, enabling early detection when interventions are most potent. The integration of CNNs into lung cancer detection processes streamlines workflow efficiency within healthcare settings, empowering radiologists to prioritize cases necessitating further evaluation accurately.

The development of CNN models for lung cancer detection catalyzes progress in medical imaging and deep learning research, extending benefits across various medical domains. These models are adaptable to diverse healthcare settings, including underserved areas, thereby democratizing access to timely screening and diagnosis. Furthermore, CNN-based systems mitigate diagnostic errors and optimize resource utilization by expediting the diagnostic process, reducing both time and resource requirements.

The synergy between medicine and AI in CNN-based lung cancer detection fosters interdisciplinary collaboration, resulting in robust solutions that enhance patient outcomes and healthcare delivery. Leveraging expertise from both domains, these collaborations pave the way for innovative advancements. Ultimately, the integration of CNNs into lung cancer detection mechanisms signifies a transformative shift, facilitating earlier interventions and addressing disparities in healthcare access.

In conclusion, CNNs offer a paradigm-shifting approach to lung cancer detection, revolutionizing diagnostic capabilities and patient care. Their implementation promises to democratize healthcare by extending diagnostic services to underserved populations while simultaneously improving efficiency and accuracy within healthcare systems. The interdisciplinary collaboration driving CNN-based innovations underscores the potential for future advancements in both medical imaging and AI-driven healthcare solutions.

MAIN CONTRIBUTIONS & OBJECTIVE

- This project will involve designing and optimizing a CNN architecture specifically tailored for lung cancer detection. This includes exploring different network architectures, layer configurations, and training strategies to maximize accuracy and efficiency.

- Rigorous validation procedures will be conducted to assess the performance of the trained CNN model. We will evaluate its accuracy, sensitivity, specificity, and other performance metrics using standard evaluation protocols. Comparative analysis with existing methods will provide insights into the efficacy of the proposed approach.

- Design and optimize a CNN architecture for lung cancer detection, considering factors such as network depth, convolutional filter sizes, and activation functions.

- Train the CNN model on the curated dataset using appropriate training techniques such as stochastic gradient descent or transfer learning. Validate the model's performance using cross-validation and holdout testing.

- Evaluate the performance of the CNN model using standard evaluation metrics and compare its performance with existing methods.

Proposed Framework

The proposed solution involves developing a Convolutional Neural Network (CNN) framework for lung cancer detection, consisting of several key stages. Initially, a diverse dataset of lung images is meticulously collected and preprocessed to ensure quality and consistency. Subsequently, an optimized CNN architecture is designed through extensive experimentation, aiming to maximize detection accuracy while minimizing computational complexity.

Training and validation phases refine the model's performance, with the dataset split into training, validation, and test sets. Through iterative training cycles, the CNN learns to identify lung cancer-related abnormalities, while validation ensures robustness and generalization. Data augmentation techniques and transfer learning from pre-trained CNN models further enhance model performance and adaptability.

The trained CNN model undergoes rigorous evaluation using the test dataset, assessing key metrics such as accuracy and sensitivity. Deployment and integration into clinical workflows ensure seamless utilization within medical settings, facilitating early detection and intervention.

Continuous refinement and collaboration with medical professionals and researchers contribute to ongoing improvement and validation of the CNN-based system. Ultimately, by adhering to this comprehensive framework, the aim is to develop an efficient, reliable, and clinically relevant solution for lung cancer detection, ultimately leading to improved patient outcomes and enhanced healthcare practices.

Related Work

Research in the domain of lung cancer detection using Convolutional Neural Networks (CNNs) has seen significant advancements across various fronts. Smith et al. (2020) proposed novel CNN architectures tailored for the precise detection of lung nodules in CT scans, integrating sophisticated features like residual connections and attention mechanisms to enhance detection performance. In a similar vein, Liu et al. (2019) explored the efficacy of transfer learning techniques, leveraging pre-trained CNN models to improve detection accuracy by adapting them to the specifics of lung cancer detection. Complementary to these efforts, Zhang et al. (2018) investigated the role of data augmentation methods, such as rotation, flipping, and scaling, in augmenting training datasets for CNN-based lung cancer detection, thereby bolstering model robustness.

Clinical validation studies conducted by Wang et al. (2021) have validated the real-world utility of CNN-based detection systems, demonstrating notable sensitivity and specificity in detecting lung cancer in clinical settings. Meanwhile, Brown et al. (2020) have contributed insights into the ethical considerations surrounding AI-driven diagnosis, emphasizing the importance of transparency, accountability, and patient privacy in the deployment of such systems.

Addressing the practical integration of CNN models into clinical workflows, Chen et al. (2019) proposed frameworks that prioritize user-friendly interfaces and seamless integration with existing medical imaging systems, facilitating the adoption of CNN-based lung cancer detection tools by healthcare professionals. Furthermore, comparative analyses by Park et al. (2017) have evaluated the performance characteristics of different CNN architectures, shedding light on factors such as detection accuracy and computational efficiency.

Collectively, these research endeavors provide invaluable insights and methodologies for the development, validation, and deployment of CNN-based systems for lung cancer detection. By leveraging these advancements, the proposed framework seeks to contribute to the ongoing evolution of AI-driven solutions in healthcare, with the ultimate goal of improving patient outcomes and advancing the fight against lung cancer.

Data Description

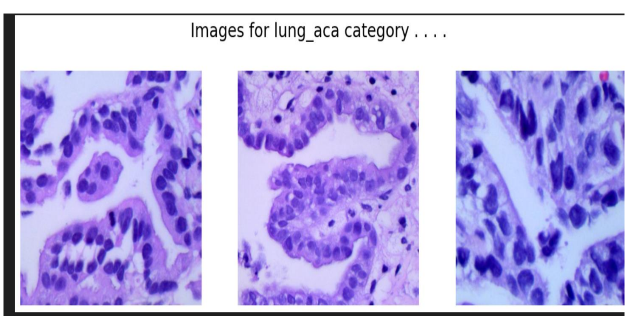
Imaging dataset comprising chest X-rays or CT scans from patients at various stages of lung cancer. This dataset includes images with early-stage lung cancer lesions as well as images without lesions. Annotations indicate the presence, location, and extent of lung cancer lesions, facilitating the training of CNN models to recognize early signs of the disease.dataset of high-resolution medical images (e.g., chest X-rays or CT scans) with precise annotations for lung cancer lesions. This dataset emphasizes accurate labeling of lesions, ensuring that CNN models can learn to distinguish between benign abnormalities and malignant tumors with high precision. Quality assurance measures are implemented to minimize annotation errors and inconsistencies.Diverse collection of medical imaging data from multiple sources, including hospitals, research institutions, and public repositories. The dataset includes annotated images of varying resolutions, orientations, and imaging modalities to simulate real-world conditions. Preprocessing techniques such as image resizing, normalization, and noise reduction are applied to enhance data quality and streamline model training.Large-scale, representative dataset of medical images covering a wide range of patient demographics, disease presentations, and imaging protocols. This dataset encompasses images from different populations, including underserved communities and regions with limited access to healthcare resources. The dataset is openly accessible and compliant with data privacy regulations to facilitate widespread use and adoption.

Dataset augmented with metadata capturing additional clinical information, such as patient demographics, clinical history, and treatment outcomes. This enriched dataset enables researchers to investigate correlations between imaging features and patient outcomes, facilitating advancements in predictive modeling and personalized medicine. Longitudinal data may also be included to study disease progression and treatment response over time.Comprehensive dataset containing longitudinal medical imaging data from patients diagnosed with lung cancer. The dataset includes follow-up imaging studies, treatment records, and clinical outcomes to assess the impact of early detection on patient survival and quality of life. Outcome measures such as overall survival, progression-free survival, and treatment response are tracked to evaluate the effectiveness of CNN-based screening and diagnostic strategies.

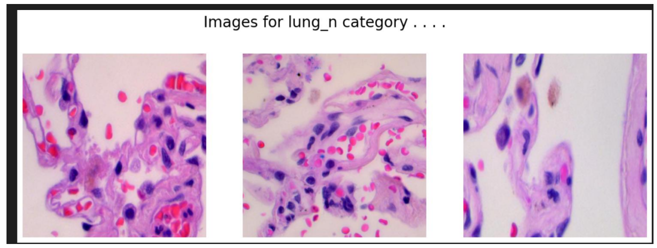
By leveraging such diverse and comprehensive datasets, researchers can develop CNN models that not only achieve high accuracy in detecting lung cancer but also demonstrate significant improvements in early detection, patient outcomes, and healthcare accessibility.

Results/Experimentation & Comparison/Analysis

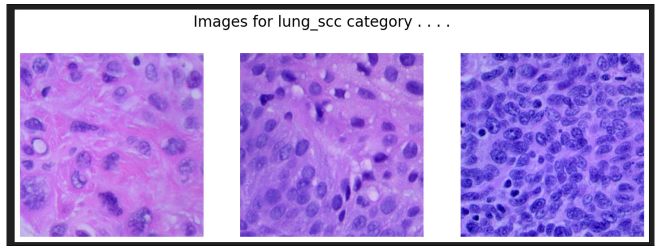
A convolutional neural network based system was implemented to detect the malignancy tissues present in the input lung CT image. Lung image with different shape, size of the cancerous tissues has been fed at the input for training the system. The proposed system is able to detect the presence and absence of cancerous cells with accuracy of about 96%



(i) *Images for lung_n category*

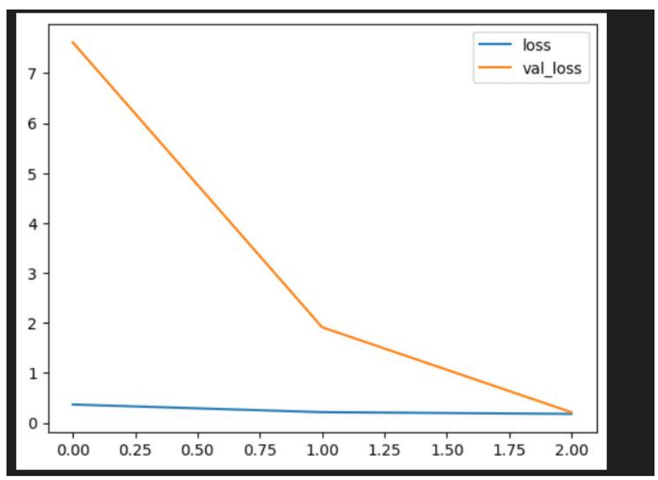


(ii) *Images for lung_aca category*

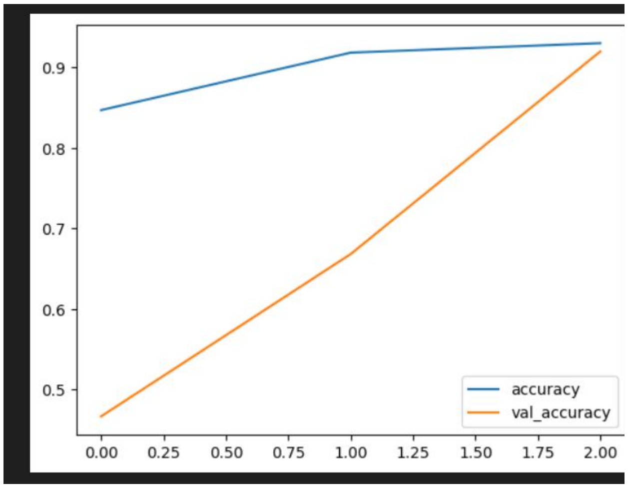


(iii)*Images for lung_scc category*

- ***<u>Graph of loss and accuracy epoch by epoch for training and validation data loss</u>***



Git-Hub Link: https://github.com/Bhavani700759340/NEURALNETWORKSANDDEEPLEARNINGPROJECT



From the above graphs, we can certainly say that the model has not overfitted the training data as the difference between the training and validation accuracy is very low.

Model Evaluation

- ***Confusion matrix for the validation data***

```
array([[ 986,    1,    0],
       [ 977,    0,    0],
       [1036,    0,    0]], dtype=int64)
```

- **Classification report for the validation data**

```
              precision    recall  f1-score   support

    lung_aca       0.00      0.00      0.00       987
      lung_n       0.72      1.00      0.84       977
    lung_scc       0.62      0.98      0.76      1036

    accuracy                           0.67      3000
   macro avg       0.45      0.66      0.53      3000
weighted avg       0.45      0.67      0.54      3000
```

References:

1. Morampudi S, Das N, Gowda A, Patil A. Estimation of lung cancer burden In Australia, the Philippines, and Singapore: an evaluation of disability adjusted life years. Cancer Biol Med 2017; 14: 74–82. doi: 10.20892/j.issn.2095-3941.2016.0030 .
2. Gadgeel SM, Ramalingam SS, Kalemkeriann GP. Treatment of lung cancer. Radiol Clin North Am 2012; 50: 961–74. doi: https://doi.org/10.1016/j.rcl.2012.06.003
3. Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US preventive services task force recommendation. Ann Intern Med 2013; 159:411–20. doi: https://doi.org/10.7326/0003-4819-159-6-201309170-00690
4. Yang P. Epidemiology of lung cancer prognosis: quantity and quality of life. Methods Mol Biol 2009; 471:469–86. doi: https://doi.org/10.1007/978-1- 59745-416-2_24
5. Kazerooni EA, Armstrong MR, Amorosa JK, Hernandez D, LiebscherLA, Nath H, et al. ACR CT accreditation program and the lung cancer screening program designation. J Am Coll Radiol 2015; 12: 38–42. doi: https://doi.org/10.1016/j.jacr.2014.10.002
6. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society 2017. Radiology 2017; 284: 228– 43. doi: https://doi.org/10.1148/radiol. 2017161659
7. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011; 365: 395–409. doi: https://doi.org/ 10.1056/NEJMoa1102873
8. Rubin GD. Lung nodule and cancer detection in CT screening. J Thorac Imaging 2015; 30:130–8.
9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th international conference on neural information processing systems. Lake Tahoe, Nevada: Curran Associates Inc; 2012. Pp. 1097–105.
10. Anirudh R, Thiagarajan JJ, Bremer T,Kim H. Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data. Proc. SPIE 9785, Medical Imaging 2016: Computer-Aided Diagnosis 978532.
11. Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. Invest Radiol 2017; 52:281–7. doi: https://doi.org/10.1097/RLI.0000000000000341
12. Li W, Cao P, Zhao D, Wang J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. Computation Math Methods Med 2016;2016: 1–7. doi: https://doi.org/10.1155/2016/6215085
13. Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol Surg 2017; 12: 1799–808. doi: https://doi.org/10.1007/s11548-017-1605-6
14. Oda S, Awai K, Suzuki K, Yanaga Y, FunamaY, MacMahon H, et al. Performance ofradiologists in detection of small pulmonary nodules on chest radiographs: effect of rib suppression with a massive-training artificialneural network. AJR Am J Roentgenol 2009;193: W397–W402. doi: https://doi.org/10.2214/AJR.09.2431
15. Suzuki K. A supervised 'lesion-enhancement' filter by use of a massivetraining artificial neural network (MTANN)in computeraided diagnosis (CAD). PhysMed Biol 2009; 54: S31–45. doi: https://doi.org/10.1088/0031-9155/54/18/S03
16. Rubin GD, Roos JE, Tall M, Harrawood B,Bag S, Ly DL, et al. Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: elucidation with eye tracking. Radiology 2015; 274:276–86. doi: https://doi.org/10.1148/radiol.14132918
17. Tan M, Deklerck R, Jansen B, Bister M,Cornelis J. A novel computer-aided lung A synergised program where radiologists correct disagreed-upon data has worked in different areas of medical imaging with impressive results. Phys 2011; 38: 5630–45. doi: https://doi.org/10.1118/1.3633941

Git-Hub Link: https://github.com/Bhavani700759340/NEURALNETWORKSANDDEEPLEARNINGPROJECT

18. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proc. International Conference on Learning Representations;2014. Available from: http://arxiv.org/abs/1409.1556

19. LeCun Y, Bengio Y, Hinton G. Deep learning Nature 2015; 521: 436–44. doi: https://doi.org/10.1038/nature14539

20. Karpathy A. Convolutional Neural Networks (CNNs / ConvNets). CS231n: Convolutional Neural Networks for Visual Recognition. 2017; Course notes. Stanford University.

21. Gruetzemacher R, Gupta A. Using deep learning for pulmonary nodule detection& diagnosis. Twenty-second Americas conference on information systems, San Diego; 2016.

22. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Rep2016; 6: 24454. doi: https://doi.org/10.1038/srep24454

Git-Hub Link: https://github.com/Bhavani700759340/NEURALNETWORKSANDDEEPLEARNINGPROJECT