Lung cancer prediction using CNN

Ms. Pooja Patle

*Department of M.Tech. AIML*

*Tulsiramji Gaikwad Patil College of Engineering & Technology,Nagpur,Maharashtra,India-441108*

poojapatle40@gmail.com

***Abstract: Lung cancer remains a leading cause of mortality worldwide.*** ***Early and accurate diagnosis is crucial for improving patient outcomes. This paper proposes a lung cancer prediction system utilizing optimized Convolution Neural Networks (CNNs).It is essential to detect malignant states of cancer at the earliest possible stage by developing an automated tool. The prediction with accuracy becomes a challenge for researchers despite the many algorithms proposed in the past by different researchers. This study proposes the system which leverages the power of CNN in extracting intricate features from lung CT scans for robust classification of malignant and begin nodules and to detect abnormal lung tissue growth. It presents a novel architecture incorporating residual connections and batch normalization for enhanced learning efficiency and reduced over fitting.*** ***The system is trained and validated on a publicly available lung cancer image dataset. The results demonstrate promising performance, achieving high accuracy in differentiating lung cancer from benign nodules. This study suggests the potential of optimized CNN-based systems to become valuable tools for computer-aided diagnosis (CAD) of lung cancer, potentially aiding healthcare professionals and improving patient survival rates.***

***Keywords: Lung Cancer, Convolution Neural Networks (CNNs), Deep Learning, Medical Image Analysis, Computer-Aided Diagnosis (CAD)***

1. INTRODUCTION

Unveiling the Threat: A Comprehensive Look at Lung Cancer

Lung cancer, a formidable foe in the fight for human health, arises from the uncontrolled growth of abnormal cells within the lungs. It holds the grim title of the world's leading cause of cancer-related deaths, surpassing breast, prostate, and colon cancer combined. This necessitates a deeper understanding of its causes, classifications, diagnosis, and the exciting advancements that offer a glimmer of hope for the future.

Unmasking the Culprits: Factors Contributing to Lung Cancer:

The Grip of Smoking: Cigarette smoking remains the primary villain in the lung cancer story. The multitude of harmful chemicals present in cigarettes wreaks havoc on lung cells, triggering mutations that propel them into uncontrolled growth.

Environmental Onslaught: Exposure to noxious elements in the environment significantly elevates lung cancer risk. This includes radioactive gas, air pollution, and secondhand smoke – all of which contribute to lung damage and potentially trigger cancerous mutations.

A Matter of Genetics: A person's genetic makeup can influence

their susceptibility to lung cancer. Inherited gene mutations can increase the likelihood of developing the disease.

Classifying the Enemy: Different Faces of Lung Cancer:

Primary vs. Secondary: It's crucial to distinguish between primary and secondary lung cancer. Primary lung cancer originates within the lungs themselves, while secondary lung cancer starts elsewhere in the body and metastasizes, or spreads, to the lungs.

Cancerous vs. Noncancerous: Lung tumors can be either malignant (cancerous) or benign (noncancerous). Malignant tumors are aggressive and have the potential to spread to other parts of the body. Benign tumors, on the other hand, are typically slow-growing and localized, posing a lesser threat.

Grading the Threat – A Matter of Severity: Doctors utilize a grading system to categorize lung cancer severity based on the rate and appearance of cell division. Lower grades (I & II) indicate slower growth and less abnormality in the cells. Conversely, higher grades (III & IV) signify faster growth and more significant cellular changes, suggesting a more aggressive form of cancer.

Unearthing the Enemy: Techniques for Diagnosis:

Visual Inspection: A crucial step in diagnosis involves microscopic examination of cells obtained through a biopsy. This allows pathologists to identify cancerous characteristics within the cells.

Imaging the Battlefield: Advanced technologies like CT scans, ultrasounds, and X-rays play a vital role in pinpointing and assessing tumors. CT scans, in particular, are invaluable due to their ability to generate high-resolution, three-dimensional images, providing a detailed picture of the lung tissue and the location and extent of any lesions.

The Rise of AI: A New Weapon in the Fight:

Researchers are at the forefront of exploring the potential of artificial intelligence (AI) to revolutionize lung cancer detection. Techniques like convolutional neural networks (CNNs) are being investigated to analyze CT scan images with the potential to improve detection accuracy. Early studies have shown promising results, with AI systems demonstrating encouraging levels of accuracy in differentiating normal and cancerous lung tissue based on CT scans. This technology holds the potential to become a valuable tool to assist medical professionals in early detection and potentially reduce the number of false positives. Ultimately, the aim is to empower clinicians with AI-driven insights for improved decision-making, leading to more effective lung cancer detection and treatment strategies.

A Brighter Future: Advancements in Lung CancerManagement

The past five years have witnessed a surge in advancements in our understanding, diagnosis, and treatment of lung cancer. Continuous research efforts, coupled with cutting-edge technology and innovative therapeutic approaches, are continuously reshaping the landscape of lung cancer management. This offers a beacon of hope for improved patient

outcomes and a more promising future in the fight against this formidable disease. By harnessing these advancements, we can move closer to a world where lung cancer is not an inevitable death sentence, but a manageable condition.

Over the last five years, significant strides have been made in understanding, diagnosing, and treating lung cancer. Advances in research, technology, and therapeutic approaches have shaped the landscape of lung cancer management.

Fig: Pipeline of the CNN ensemble.

II. RELATED WORKS

In recent years, a plethora of research has been conducted to develop effective methods for the classification and detection of colon and lung cancers using deep learning techniques. Various convolution neural network (CNN) architectures, pre-trained models, and innovative approaches have been explored to achieve accurate and efficient cancer detection.

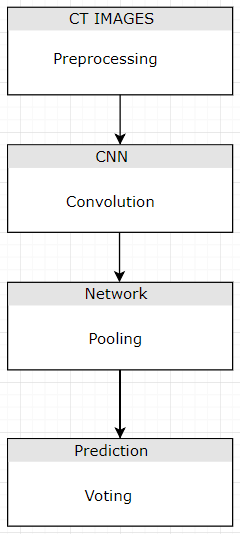
In [11], CNN methods were utilized to examine imaging data of colon cells, with average pooling, max pooling layers, and MobileNetV2 applied for colon cell image classification. Wahid et al. [12] introduced a computer-aided diagnosis (CAD) mechanism using CNNs to detect colon, lung, and cancer tissues. They employed pre-trained CNN models such as ResNet18, Shuffle Net V2, Google Net, and a customized CNN model for this purpose.

Kumar et al. [13] devised four CNN architectures, including 3-block, baseline, and 2-block CNNs with data augmentation, to classify colon tissue histopathological images (HPI). Garg and Garg [14] proposed the use of pre-trained CNN models like MobileNet, InceptionResNetV2, VGG16, and others for colon and lung cancer detection using HPI, along with improved augmentation approaches.

Adu et al. [15] introduced a novel DHS-CapsNet (dual horizontal squash capsule network) for categorizing colon and lung cancers on HPI. This method utilizes a horizontal squash (HSquash) function developed for encoder feature fusion (EFF).

In [16], gene elements with maximum correlation with tumors were chosen using Weighted Gene Co-expression Network Analysis (WGCNA), and a survival study was conducted using the least absolute shrinkage and selection operator (Lasso) technique. Sakr et al. [17] presented a lightweight DL method based on CNN for colon recognition, with the input HPI normalized before CNN processing.

In [18], a classification method for colon cancer was developed with two classes, polyps and adenocarcinomas, utilizing the CNN approach with the MobileNet structure. Mohalder et al. [19] presented a DL approach for forecasting CRC tumors in HPI, employing CNN to analyze tumor images and recognize abnormal or suspicious tumor designs.

Bhattacharya et al. [20] examined a framework utilizing DL or meta-heuristic techniques for predicting colon or lung cancer in HPIs with near-perfect precision. In [21], a colon cancer classification method was established to categorize colon adenocarcinomas and benign colonic tissues using CNN architectures such as VGG19, VGG16, ResNet152, and others.

Despite the advancements in ML and DL models, there is still a need to enhance lung and colon cancer classification performance. The increasing complexity of DL models can lead to overfitting, and hyperparameters such as epoch count, batch size, and learning rate selection play a crucial role in model efficiency. To address these challenges, metaheuristic algorithms like IAFO and TSA are employed for parameter selection in this work. These algorithms offer an efficient approach to tuning hyperparameters and optimizing DL models for improved cancer classification performance.

III.LITERATURE REVIEW

Lung cancer prediction systems utilizing Convolution Neural Networks (CNNs) have gained significant attention in recent years due to their potential to enhance early detection and treatment outcomes. Several studies have been conducted to develop and improve CNN-based models for lung cancer prediction. Here is a literature review summarizing the key findings and approaches from some of these studies:

Automated Lung Cancer Detection Using CNN with CT Images:

Researchers have explored the use of CNNs for automated lung cancer detection using CT images. Various CNN architectures, such as VGG16, ResNet, and Inception, have been employed to analyze high-resolution CT scans.

These studies typically involve training CNN models on large datasets of CT images, consisting of both normal and cancerous lung tissue.

Results from these studies have shown high accuracy in distinguishing between normal and cancerous lung tissue, with some models achieving over 90% accuracy.

Deep Learning for Lung Cancer Prediction from X-ray Images:

Other studies have focused on predicting lung cancer from X-ray images using deep learning techniques, particularly CNNs.

Researchers collected datasets of X-ray images from patients diagnosed with lung cancer and healthy individuals.

CNN models were trained to classify X-ray images as either cancerous or non-cancerous.

Promising results have been reported, indicating the potential of CNNs in accurately predicting lung cancer from X-ray images.

Multi-Modal Approach for Lung Cancer Prediction:

Some researchers have adopted a multi-modal approach, combining different imaging modalities such as CT scans, X-ray images, and clinical data for improved lung cancer prediction.

CNNs were employed to process and analyze the various types of data, extracting relevant features for cancer prediction.

Integrating multiple data sources enhanced the accuracy and reliability of the prediction models, leading to more robust lung cancer detection systems.

Transfer Learning for Lung Cancer Prediction:

Transfer learning techniques have also been explored in lung cancer prediction systems using CNNs.

Researchers fine-tuned pre-trained CNN models on lung cancer datasets to adapt them for specific prediction tasks.

Transfer learning facilitated training with limited data and accelerated the development of accurate prediction models.

Interpretability and Explainability of CNN Models:

Some studies have focused on enhancing the interpretability and explainability of CNN models for lung cancer prediction.

Techniques such as attention mechanisms and saliency mapping were employed to visualize regions of interest in CT or X-ray images, aiding clinicians in understanding the model's decision-making process.

In conclusion, CNN-based lung cancer prediction systems have shown great potential in improving early detection and prognosis. Continued advancements in deep learning techniques and imaging technology are expected to drive further improvements in lung cancer prediction systems.

Several studies have explored different approaches for detecting lung cancer, ranging from neural networks to Hidden Markov Models (HMM), decision trees, support vector machines (SVM), fuzzy clustering, and convolutional neural networks (CNNs).

In 1994, Ghosh and Reilly utilized a three-layer feedforward neural network for lung cancer detection. The network was trained on images containing signs of lung cancer to identify them accurately.

Abhinav and Amlan proposed a Hidden Markov Model (HMM) for lung cancer detection, which doesn't require past information to predict the further state of the disease.

Y. Sahin and E. Human introduced a method using decision trees and SVM for lung cancer detection. They compared the performance of various decision tree methods and SVM with different kernel functions.

Additionally, an approach combining fuzzy clustering and neural networks was suggested for lung cancer detection by Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang. Their method involved three phases: image collection, fuzzy clustering for normal usage behavior, and neural network-based verification for doubtful cases.

Nasser, Ibrahim M., et al. conducted extensive research on neural network-based models for classification, prediction, and diagnosis. They developed artificial neural network (ANN) models for predicting movie ratings, mobile phone prices, animal categories, tumor categories, and diagnosing autism.

Abu Naser et al. also developed numerous classification models based on artificial neural networks for various applications.

In summary, these studies showcase a range of techniques for lung cancer detection, including traditional machine learning methods like HMM, decision trees, SVM, as well as more advanced methods like CNNs. Each approach has its strengths and weaknesses, and ongoing research continues to refine and improve these methods for more accurate and reliable lung cancer detection.

1. PROPOSED SYSTEM

The proposed work outlines a comprehensive approach for the early detection of lung cancer using deep learning techniques, particularly Convolutional Neural Networks (CNNs). The project is planned to be carried out in the following steps:

Data Collection: Data will be collected from EBC, consisting of biopsy slide images of tumors. The EBC data contains patches cropped from the selected tumor region.

Classification of Whole Slide Images (WSI): WSI will be classified into positive (N+) or negative (N0) using the proposed DLCNB method.

Feature Extraction: Features will be extracted from the images using CNNs. Additionally, clinical data will be preprocessed to extract features.

Max Pooling: Extracted features will be processed using max pooling.

Attention-Based Multiple-Instance Learning (MIL): The N feature vectors will pass through an attention-based MIL approach.

Data Splitting: The data will be split into training, testing, and validation sets.

Model Training: The model will be trained using the training dataset.

Testing: The saved weights from training will be tested on the testing dataset.

Evaluation: Various statistics of the model, such as accuracy, precision, and F1-score, will be calculated to analyze its performance.

Integration with Desktop/Web Apps: The developed model will be integrated into desktop or web applications for practical use.

Approach and steps for project building:

1. Problem Detection/Identification: Identifying the problem of early lung cancer detection using deep learning techniques.
2. Dataset Selection: Selecting datasets from open-source repositories containing biopsy slide images of lung tumors.
3. Selection of Algorithm, Framework, and Architecture: Choosing CNN as the algorithm, selecting suitable frameworks like TensorFlow or PyTorch, and designing the architecture.
4. Data Preprocessing: Preprocessing the collected data, including resizing, normalization, and cleaning.
5. Data Augmentation: Augmenting the dataset to increase its size and diversity, improving model generalization.
6. Training & Testing Data Preparation: Splitting the dataset into training and testing sets.
7. Model Training: Training the CNN model on the training dataset.
8. Validation: Validating the trained model on the validation set to fine-tune hyperparameters.
9. Testing: Evaluating the model's performance on the testing dataset.
10. Fine Tuning: Fine-tuning the model based on testing results to optimize performance.
11. Result Analysis: Analyzing the results using metrics like confusion matrix, accuracy, precision, recall, and F1 score.

These steps are essential for the early detection of lung cancer, which can significantly reduce risks and improve outcomes for patients. Preprocessing, scanning, training, testing, and confirmation are crucial stages, and the use of CNNs facilitates efficient image processing and classification. By integrating the developed model into desktop or web applications, the project aims to make lung cancer detection accessible and effective for a wide range of users.

These square measure steps projected within the system:- These steps can facilitate America police investigation the pre stages for early detection of carcinoma, this will facilitate America to urge the simplest accuracy through image process in these, {we will|we'll|we square measure going to} phase the image and supported that we have a tendency to are going to categories them to the process stage so comparison it with the past pictures of respiratory organ which is able to facilitate in police investigation cancer. This will facilitate individuals from reducing the risks of police investigation cancer within the last stages whereas doing a pre-test of cancer detection. These projects can facilitate individuals worldwide and from being late or delay in police investigation cancer. Preprocessing, Scanning, Training, Testing and confirmatory as a result of the info needed to coach a CNN is incredibly giant, it's usually fascinating to coach the model in batches. Loading all the coaching information into memory isn't invariably potential as a result of you would like enough memory to handle it and therefore the options too. I made a decision to load all pictures into a hdf’s dataset exploitation h5py library.

Adenocarcinomact scan article

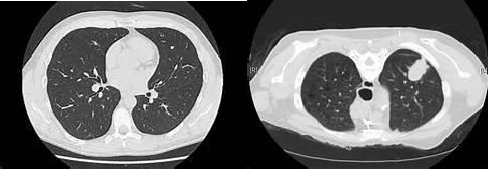


Adenocarcinoma is one of the most cancerous diseases which is creating a worldwide problem and increasing the death ratio.

So there are many upcoming projects that will help to reduce the chances and detect them in earlier stage.

So this image is the CT scan image of the lungs cancer, named Adenocarcinoma. We train all the images from the database.

Images of healthy lung to cancerous lung



Healthy lung Cancerous lung

V. REFERENCES

[1] M. S. N. Raju and B. S. Rao, ‘‘Lung and colon cancer classification using hybrid principle component analysis network-extreme learning machine,’’ Concurrency Comput., Pract. Exper., vol. 35, no. 1, Jan. 2023.

[2] M. A. Talukder, M. M. Islam, M. A. Uddin, A. Akhter, K. F. Hasan, and M. A. Moni, ‘‘Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning,’’ Expert Syst. Appl., vol. 205, Nov. 2022, Art. no. 117695.

[3] O. Attallah, M. F. Aslan, and K. Sabanci, ‘‘A framework for lung and colon cancer diagnosis via lightweight deep learning models and transformation methods,’’ Diagnostics, vol. 12, no. 12, p. 2926, Nov. 2022.

[4] M. A. Fahami, M. Roshanzamir, N. H. Izadi, V. Keyvani, and R. Alizadehsani, ‘‘Detection of effective genes in colon cancer: A machine learning approach,’’ Informat. Med. Unlocked, vol. 24, 2021, Art. no. 100605.

[5] H. C. Reis and V. Turk, ‘‘Transfer learning approach and nucleus segmentation with MedCLNet colon cancer database,’’ J. Digit. Imag., vol. 36, no. 1, pp. 306–325, Sep. 2022.

[6] N. Kumar, M. Sharma, V. P. Singh, C. Madan, and S. Mehandia, ‘‘An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images,’’ Biomed. Signal Process. Control, vol. 75, May 2022, Art. no. 103596.

[7] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, ‘‘A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework,’’ Sensors, vol. 21, no. 3, p. 748, Jan. 2021.

[8] A. H. Chehade, N. Abdallah, J.-M. Marion, M. Oueidat, and P. Chauvet, ‘‘Lung and colon cancer classification using medical imaging: A feature engineering approach,’’ Phys. Eng. Sci. Med., vol. 45, no. 3, pp. 729–746, Sep. 2022.

[9] M. Toğaçar, ‘‘Disease type detection in lung and colon cancer images using the complement approach of inefficient sets,’’ Comput. Biol. Med., vol. 137, Oct. 2021, Art. no. 104827.

[10] I. Pacal, D. Karaboga, A. Basturk, B. Akay, and U. Nalbantoglu, ‘‘A comprehensive review of deep learning in colon cancer,’’ Comput. Biol. Med., vol. 126, Nov. 2020, Art. no. 104003.

[11] Z. Tasnim, S. Chakraborty, F. M. J. M. Shamrat, A. N. Chowdhury, H. A. Nuha, A. Karim, S. B. Zahir, and M. M. Billah, ‘‘Deep learning predictive model for colon cancer patient using CNN-based classification,’’ Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 8, pp. 687–696, 2021.

[12] R. R. Wahid, C. Nisa’, R. P. Amaliyah, and E. Y. Puspaningrum, ‘‘Lung and colon cancer detection with convolutional neural networks on histopathological images,’’ AIP Conf. Proc., vol. 2654, Feb. 2023, Art. no. 020020.

[13] A. Kumar, A. Vishwakarma, V. Bajaj, A. Sharma, and C. Thakur, ‘‘Colon cancer classification of histopathological images using data augmentation,’’ in Proc. Int. Conf. Control, Autom., Power Signal Process. (CAPS), Dec. 2021, pp. 1–5.

[14] S. Garg and S. Garg, ‘‘Prediction of lung and colon cancer through analysis of histopathological images by utilizing pre-trained CNN models with visualization of class activation and saliency maps,’’ in Proc. 3rd Artif. Intell. Cloud Comput. Conf., Dec. 2020, pp. 38–45.

[15] K. Adu, Y. Yu, J. Cai, K. Owusu-Agyemang, B. A. Twumasi, and X. Wang, ‘‘DHS-CapsNet: Dual horizontal squash capsule networks for lung and colon cancer classification from whole slide histopathological images,’’ Int. J. Imag. Syst. Technol., vol. 31, no. 4, pp. 2075–2092, Dec. 2021.

[16] Y. Su, X. Tian, R. Gao, W. Guo, C. Chen, C. Chen, D. Jia, H. Li, and X. Lv, ‘‘Colon cancer diagnosis and staging classification based on machine learning and bioinformatics analysis,’’ Comput. Biol. Med., vol. 145, Jun. 2022, Art. no. 105409.

[17] A. S. Sakr, N. F. Soliman, M. S. Al-Gaashani, P. Pławiak, A. A. Ateya, and M. Hammad, ‘‘An efficient deep learning approach for colon cancer detection,’’ Appl. Sci., vol. 12, no. 17, p. 8450, Aug. 2022.

[18] Y. G. C. Kelana, S. Rizal, and S. Saidah, ‘‘Classification of histopathological images of colon cancer using convolutional neural network method,’’ in Proc. Int. Conf. Comput. Sci., Inf. Technol. Eng. (ICCoSITE), Feb. 2023, pp. 821–826.