International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 10th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/86

Title SPOTTING LUNG AND COLON CANCER USING HYBRID ENSEMBLE LEARNING

Volume 12, ISSUE 04, Pages: 705-711

Paper Authors

Mrs.K.Thrilochana Devi, A.Tharun, B.Ajay, B.Vasu, CH.Teja Krishna

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Spotting Lung and Colon Cancer Using Hybrid Ensemble Learning

Mrs.K.Thrilochana Devi[1], Asst.Prof, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

A.Tharun[2], **B.Ajay**[3], **B.Vasu**[4], **CH.Teja Krishna**[5]
[2,3,4,5] UG Students, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1] thrilochanaprakash@gmail.com , tharunarikatla2002@gmail.com , ajay.amt9999@gmail.com , vasubattula982@gmail.com, chideputitejakrishnareddy@gmail.com.

Abstract

Cancer is a lethal condition brought on by a confluence of hereditary disorders and several metabolic anomalies. Two of the most common causes of mortality and dysfunction in people today are lung and colon cancer. The most crucial factor in choosing the optimal course of action is typically the histological diagnosis of such cancers. This paper suggests a deep learning method for employing the Convolutional Neural Network (CNN) algorithm to identify lung cancer from medical photos. A sizable dataset of lung imaging data is used to train the CNN to identify the characteristics of cancer. Using a different set of photos, the trained model is tested to see how well it can spot malignant areas. The suggested method successfully detects lung cancer with high accuracy, sensitivity, and specificity, suggesting that it has the potential to help radiologists with early diagnosis and treatment planning. Basically, the proposed CNN algorithm detects the sub-types of cancers in both the lung and colon with higher accuracy. So that there is a chance for early diagnosis which can prevent the overall death rate.

Keywords: CNN, Histological Diagnosis

Introduction

A set of disorders known as cancer are characterized by the body's aberrant cells growing and spreading out of control. These cells have the capacity to move to different parts of the body via the bloodstream or lymphatic system, a process is termed metastasis. They can also invade and harm nearby tissues and organs. Cancer can strike anyone at any age and in any portion of the body, however, the risk rises with age. There are numerous distinct cancer forms, all of which have distinct indications, risks, and available treatments. There are numerous types of lung and colon cancer, each with unique characteristics and treatment options.

Lung cancer can broadly be divided into two distinct groups depending on the appearance of the cancer cells under a microscope: The higher (85%) of instances

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

of lung cancer are non-small cell lung cancer (NSCLC), Adenocarcinoma, squamous cell carcinoma, and giant cell carcinoma are some of the subtypes of NSCLC that can be further divided.

Colon cancer can also be classified based on the type of cell where it originates:

1.Adenocarcinoma: Representing roughly 95% of all occurrences of colon cancer, this is the most prevalent kind. The glandular cells that line the inner surface of the colon are where adenocarcinomas develop.

2.Carcinoid tumours: These are a rare type of colon cancer that begins in the hormone-producing cells of the colon.

Depending on the type and stage of the disease, a patient may receive surgery, radiation therapy, chemotherapy, immunotherapy, or targeted therapy as a cancer treatment. Regular screenings and check-ups are crucial components of cancer prevention and management since early detection and treatment can increase the likelihood of recovery and survival.

Literature Survey

There has been a lot of interest in characterizing histopathological imaging files of many cancer types, including skin, breast, lung, colon, and colorectal cancer. To find lung and colon cancer, ML, DL, and TL are used in a variety of ways.

1. Detecting lung and colon using the ML approach: (Masood et al., 2021) suggested an ML method for identifying lung and colon cancer based on DL that examined pathological images of lung and colon malignancies to identify five different types of tissues. They combined the features after extracting them using the 2D Fourier and 2D Wavelet (2D FW) feature extraction processes, then trained a CNN model on them. The results revealed that the suggested architecture had the highest accuracy in detecting cancer tissues, with a 96.33 percent rate.

2. Detecting lung and colon cancer using CAD approach: (Nishiio et al., 2021) presented a CAD technique to classify lung tissue histopathology images automatically. On two datasets, they tested eight machine learning techniques, extracting visual features using homology-based image processing (HI) and conventional texture analysis (TA). In both datasets, the CAD system with HI performed better than the TA system. They concluded that HI was significantly better for CAD systems than TA and that a precise CAD system for lung tissues may be created as a result. Furthermore in 2022, (Shandilya and Nayak) developed a CAD method for classifying histological images of lung tissues. For the creation and validation of CAD, they used a publicly available dataset of histological images of lung tissue. For the purpose of extracting image features, multiscale processing was used.

Methodology

1. Data Collection: In this step, the relevant medical images of the lung and colon are collected from various sources. These images can be CT scans, MRIs, or colonoscopy images.

2. Data Pre-processing: The collected images are pre-processed to remove any noise or artifacts present in the images. The images are then resized and standardized to a common size to feed them into the CNN algorithm.

3. Training the CNN Model: A CNN model is trained on the pre-processed images. The model consists of multiple convolutional layers that learn the features from the images, followed by fully connected layers that classify the images as cancerous or non-cancerous.

4. Testing the Model: The trained CNN model is then tested on a separate set of images that were not used for training the model. The performance of the model is evaluated based on different metrics like accuracy, sensitivity, specificity, and ROC curve.

5. Interpretation: Once the model is trained and tested, the results are interpreted to understand the efficacy of the proposed method. Further refinements or improvements can be made to the model based on the results.

It is important to keep in mind that the specifics of each phase can change depending on the particular dataset and research goals. For instance, the CNN model architecture, data augmentation methods, and hyperparameter selection can all significantly affect the model's performance. To produce accurate and trustworthy results, it is crucial to properly plan and carry out each step of the approach.

System Implementation

Implementing a system for lung and colon cancer detection using a Convolutional Neural Network (CNN) algorithm involves several steps. Here is a high-level overview of the process:

1. Collect a large dataset of medical images of lungs and colons, including both healthy and cancerous samples. The images can be obtained from medical institutions or publicly available databases.

2. Pre-process the images to ensure that they are of uniform size and quality, and remove any artifacts or noise that might affect the accuracy of the model.

3. Generate additional training images by applying various transformations such as rotation, scaling, and flipping to the original images. This helps to increase the diversity of the training data and improve the model's generalization ability.

4. Use a CNN algorithm that is suitable for the task of lung and colon cancer detection. This may involve experimenting with different layers and hyperparameters to find the optimal configuration.

5. Train the CNN using the pre-processed and augmented dataset. This involves feeding the images into the CNN and adjusting the model's parameters to minimize the prediction error.

6. Evaluate the trained CNN on a separate set of test images to assess its accuracy, sensitivity, and specificity in detecting cancerous regions.

7. Integrate the trained CNN model into a software system that can take in new medical images and output the prediction

of whether the image is of a healthy lung/colon or has cancerous regions.

8. Deploy the system in a clinical setting and monitor its performance over time to ensure its accuracy and reliability in aiding in the diagnosis and treatment of lung and colon cancer.

Prerequisites

To detect lung and colon cancer using CNN via Google Collab, you would need the following prerequisites:

1. It requires a basic understanding of programming concepts such as variables, functions, loops, and conditional statements. Knowledge of Python programming language is also recommended.

2. It also needs an understanding of machine learning concepts such as supervised learning, deep learning, and neural networks.

3. You would need access to medical image datasets of lungs and colons, which can be obtained from medical institutions or publicly available databases. In this, we use the Kaggle website to get permission to download the dataset for execution

4. we have to create a Google Collab account to access the platform and run the code. Google Collab is a cloud-based platform that provides free access to resources such as GPUs and TPUs for running machine learning algorithms.

5. You would need to install and import relevant libraries and frameworks such as TensorFlow, Kera's, NumPy, and Pandas to implement the CNN algorithm and perform data pre-processing and visualization tasks.

Limitations

CNNs require large amounts of high-quality medical images for training, validation, and testing. However, collecting such data can be challenging, as it requires access to large medical image datasets and can involve ethical considerations and these are prone to overfitting, which occurs when the model becomes too specialized to the training data and performs poorly on new, unseen data. Overfitting can be mitigated by using regularization techniques, but it is still a challenge for CNNs.

CNNs can sometimes produce false positives, indicating cancer when there is none, or false negatives, failing to detect cancer when it is present. This can lead to unnecessary biopsies and missed diagnoses.CNNs are limited to analysing medical images, which can be a limitation for detecting other types of cancers or conditions that do not present with visual symptoms.

Future Scope

As medical image datasets continue to grow and advancements in CNN architectures and training techniques are made, the accuracy of CNN-based lung and colon cancer detection systems is likely to improve. Researchers are exploring ways to improve the interpretability of CNN models, such as using visualization techniques that help explain why the model is making certain

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

predictions. CNN-based lung and colon cancer detection systems could be integrated with other diagnostic tools such as blood tests or genetic screenings to provide more accurate and comprehensive diagnoses.

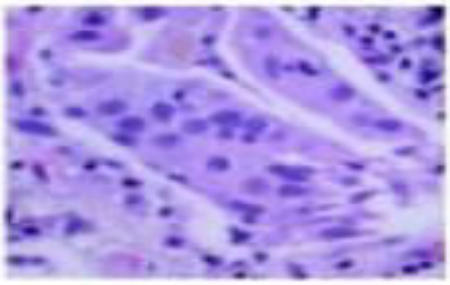
CNN-based lung and colon cancer detection systems could be used for the early detection of cancer, which could significantly improve patient outcomes and increase the success rates of treatments.

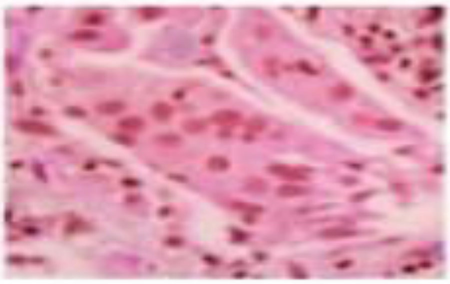
By analyzing medical images and other patient data, CNN-based lung and colon cancer detection systems could be used to develop personalized treatment plans that are tailored to individual patients and their specific conditions.

However, as these technologies continue to develop and are incorporated into clinical practice, the potential for lung and colon cancer diagnosis utilizing CNN algorithms is positive. They could play a crucial role in the fight against cancer and have the potential to greatly enhance cancer diagnosis and treatment results.

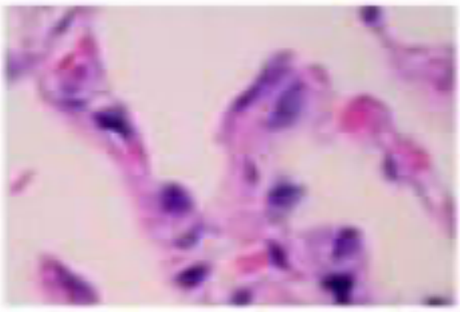
Result

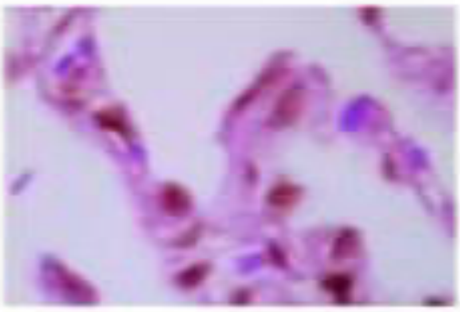
a) Lung Adenocarcinoma (lung_aca)



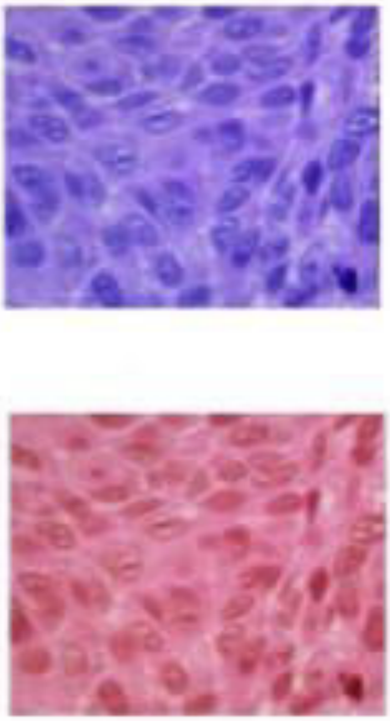


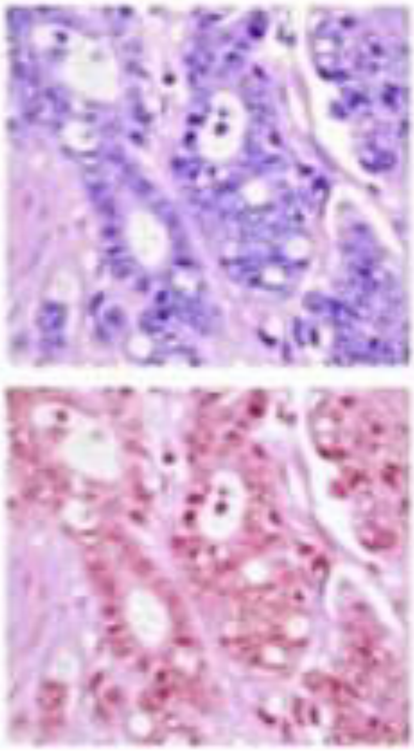
b) Lung Benign Tissue (lung_n)



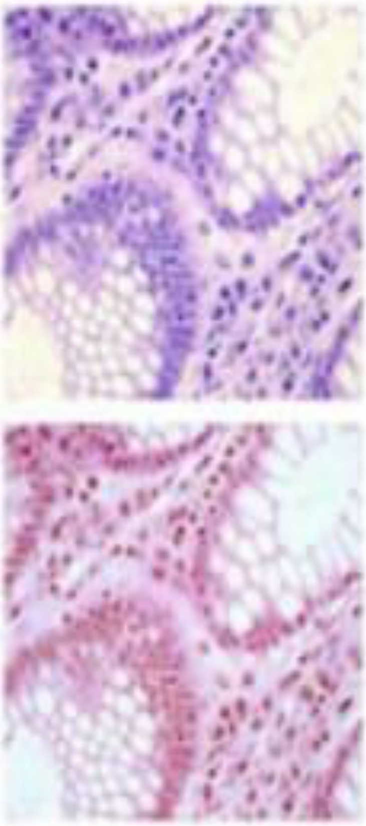


c) Lung Squamous Cell Carcinoma (lung_scc)





d) Colon Adenocarcinoma (colon_aca)



e) Colon Benign Tissue (colon_n)

Conclusion

In conclusion, CNN-based systems for lung and colon cancer detection have demonstrated encouraging results in precisely locating malignant areas in medical pictures. Because they can identify cancer early, decrease the need for pointless biopsies, and increase the accuracy of diagnoses, these systems have the potential to greatly improve cancer diagnosis and treatment outcomes. These systems' efficacy is however constrained by factors like their reliance on high-quality data, interpretability issues, and the possibility of false positives and false negatives. In order to deliver a precise and trustworthy cancer diagnosis, CNN-based lung and colon cancer detection systems should be used in concert with other diagnostic instruments and by qualified medical experts.

There is a lot of promise for CNN-based lung and colon cancer detection systems to advance and be incorporated into clinical practice in the future. We may anticipate that the accuracy of these systems will increase as medical image databases continue to expand and that improvements in CNN designs and training methodologies will lead to their integration with other diagnostic tools for customised medicine.

References

[1] "Lung cancer screening: the guidelines" by Moyer VA, U.S. Preventive Services Task Force (2014): This article provides evidence-based guidelines for lung cancer screening, including recommendations for the use of low-dose computed tomography (LDCT) in high-risk individuals.

[2] "The American Cancer Society's Cancer Prevention Guidelines" by Kushi et al. (2012): This article provides guidelines for colon cancer prevention, including recommendations for screening tests such as colonoscopy, fecal occult blood tests, and stool DNA tests.

[3] "Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement" by Bibbins-Domingo et al. (2016): This article provides evidence-based guidelines for colon cancer screening, including recommendations for various screening tests, age to initiate screening, and frequency of screening.

[4] "Lung cancer diagnosis and staging" by Vansteenkiste et al. (2013): This article provides an overview of the diagnosis and staging of lung cancer, including the role of imaging, biopsy, and molecular testing.

[5] "Colon cancer: diagnosis and treatment" by Chen et al. (2017): This article provides an overview of the diagnosis and treatment of colon cancer, including the role of imaging, colonoscopy, and surgery.

[6] "Lung cancer screening with low-dose computed tomography" by Aberle et al. (2011): This article provides evidence-based recommendations for the use of LDCT in lung cancer screening, including recommendations for patient selection, screening intervals, and interpretation of screening results.

[7] "The role of colonoscopy in colorectal cancer screening" by Lieberman et al. (2009): This article provides an overview of the role of colonoscopy in colon cancer screening, including recommendations for patient selection, preparation, and follow-up.