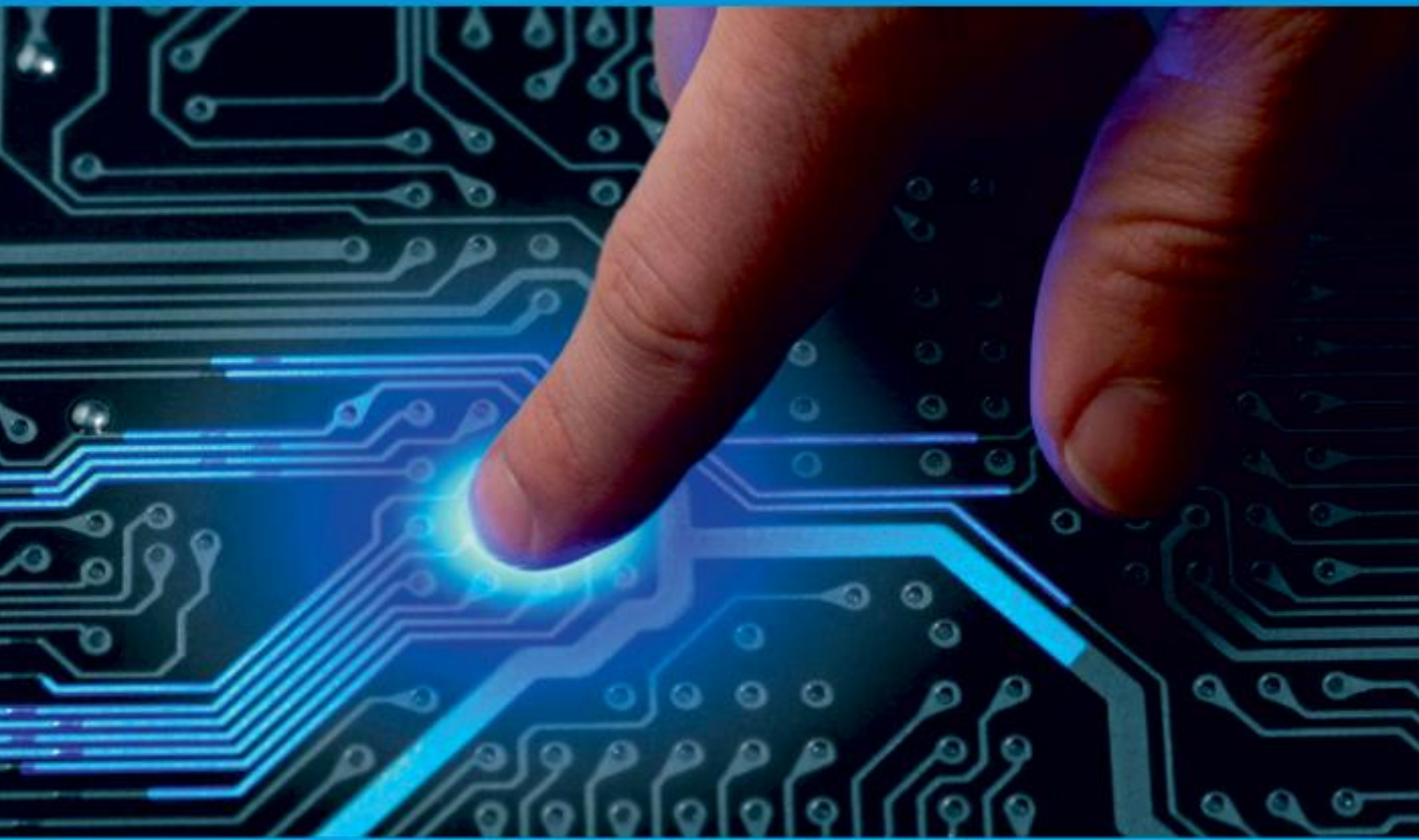
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204288 |

A Deep Learning Model to Recognize Lung Cancer

Dr. T. Aravind, Sanjai S, Vasanthan G, Naveen S

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: The scarcity of labeled data might constrain deep learning model training and evaluation, reducing performance and generalizability. Furthermore, using varied annotation techniques and criteria across datasets can generate differences and biases in the training process. Deep learning models, particularly those based on convolutional neural networks (CNNs), can be computationally expensive, necessitating large amounts of computing power for training and inference. Image resizing, normalization, and augmentation, for example, must strike a balance between maintaining useful information and decreasing computational demands. Inadequate pre-processing procedures may result in either information loss or high computing requirements, affecting the deep learning pipeline's efficiency and scalability. Most deep-learning studies in lung cancer detection and diagnosis have concentrated on certain populations or datasets that may not effectively represent the diversity of patients and imaging practices. To ensure that deep learning models generalize successfully to varied populations and imaging circumstances, pre-processing techniques must address potential biases and constraints associated with individual datasets. To overcome these pre-processing restrictions, careful evaluation of dataset features, robust algorithmic approaches, and cooperation across various universities are required.

KEYWORDS: Lung Cancer; Deep Learning Techniques; Detection; Diagnosis; Classification; Segmentation

I. INTRODUCTION

Correctly acquiring medical images and interpreting them is crucial to correctly identifying and diagnosing malignant diseases. There are numerous high-resolution image capture devices available, including CT, MRI, and X-ray scans. After pre-processing, the illness identification system extracts pertinent information from these medical images and uses those features to train its models. A further application of the trained model is to identify the disease from corresponding unknown medical images.

The traditional machine learning method is unable to produce reliable findings since medical images of various individuals vary greatly. Deep learning techniques have been successfully applied in a variety of disciplines recently, particularly in the analysis of medical pictures. These techniques are useful and effective for evaluating medical images to find disorders, especially cancer.

Techniques for deep learning are a subset of machine learning techniques that enable estimating the outcome using the provided data set and training the model based on the outcome. Deep learning techniques use neural networks with numerous layers, including an input layer, several hidden layers, and an output layer. The deep learning model is taught more precisely because there are numerous layers present. Based on their learning methodologies, deep learning models can be divided into four groups: reinforced learning models, unsupervised learning models, semi-supervised learning models, and supervised learning models.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204288 |

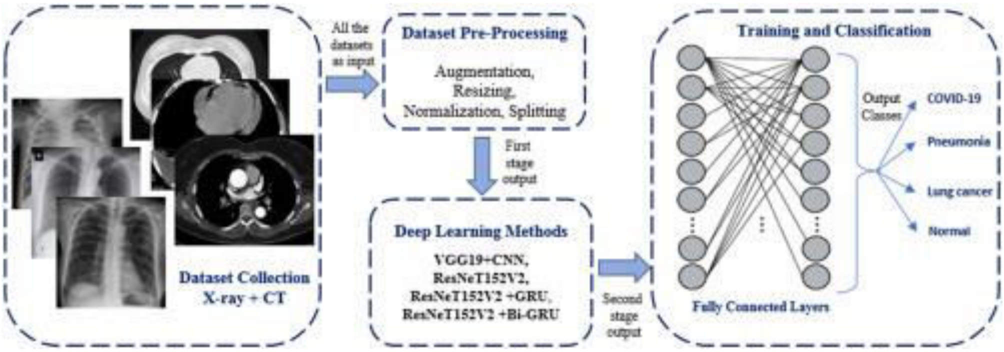


Fig 1: Deep-chest: Multi-classification

As CT scans offer highly detailed images of the lungs, they are frequently employed in the identification of lung cancer. By examining the size, shapes, textures, and intensities of the pictures, deep learning algorithms are able to recognize the affected nodules based on the CT images. Since they provide complete imaging of the lung capacity, 3D CT pictures offer a more comprehensive examination of the lungs compared to 2D CT images. In many previous works, deep learning methods have been used to analyze 3D CT images accurately to detect nodules and other abnormalities. To minimize the exposure to radiation during lung cancer screening, low-dose CT methods are the preferred modality, coupled with denoising and image augmentation techniques. Compared to CT scans, MR imaging has more capability for revealing information about tissue density and blood flow. Based on their distinctive features, deep learning systems can also analyze MR images to locate lung nodules along with additional abnormalities. To increase image quality and reduce noise, filtering techniques such as Gaussian and Median filtering are usually implemented to pre-process the incoming images. In lung cancer cases, candidate detection is the process of spotting areas of concern in the image to detect possible nodules or additional abnormalities. There are several methods in deep learning that can be used to identify prospective candidates, including region proposal networks and sliding window approaches.

In general, 2D deep learning algorithms such as recurrent neural networks (RNNs), CNNs, and hybrid models have been frequently employed for lung cancer diagnosis due to their ability to examine individual CT slices to detect nodules and other anomalies. Apart from 2D techniques, 3D deep learning architectures are also used to analyze the whole CT volumes to locate the nodules, especially on 3D CNNs, as well as additional models that are normally used for processing 3D data. In order to increase the precision of lung cancer detection, some methods applied a hybrid approach that incorporated various types of deep learning architectures for both 2D and 3D CNN to analyze both individual CT data slices and full volumes. To speed up the training process, many deep learning algorithms examine fresh data using transfer learning methods that pretrain the models using massive datasets of CT images before fine-tuning the models for any specific disease.

II. RELATED WORK

CT imaging produces detailed images of the human body, and it has been used as a non-invasive method for diagnosing and monitoring a number of diseases, including cancer, heart disease, and trauma. CT imaging has revolutionized the contemporary medicine field by providing specialists with accurate tools to provide internal organ images of the body, enabling them to make more informed diagnoses and treatment decisions. The attenuation of X-rays as they pass through various body tissues is the basic principle by which CT imaging maps the internal organs. A CT scan can be captured by requesting that the patient be positioned on a table that slides into a large, doughnut-shaped CT scanner. The scanner then emits a number of X-rays that go through the body before being detected by a number of detectors on the scanner’s opposite side.

The X-ray attenuation data is then processed by a computer to provide cross-sectional images of the body that can then be further transformed into 3D images. There are a number of advantages of CT imaging compared to other medical imaging modalities, whereby the former imaging technique is non-invasive since it does not involve any invasive treatments or incisions. A standard CT scan also requires a short amount of processing time, making it a rather speedy technique. Furthermore, a CT scan provides the imaging data inside the body in great detail, providing medical professionals with access to irregularities that might not be visible if traditional imaging techniques are used. CT imaging has, however, a number of disadvantages, including the use of ionizing radiation that could eventually increase

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 4, April 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1204288 |

the risk of getting cancer. Therefore, it should only be used when it is absolutely essential, and it must be administered with the appropriate radiation protection. Furthermore, CT imaging may detect abnormalities that are not actually there, a condition known as false-positive results. This could lead to unnecessary, expensive, and potentially harmful testing and treatment.

CT imaging-based lung cancer diagnosis has improved using DL methods. Deep learning uses multiple-layered networks to learn from enormous volumes of data. These networks are capable of identifying connections and patterns in data that are too complex for traditional machine-learning algorithms to grasp. Deep learning systems can be finetuned to take CT images as input to identify lesions, tumors, and nodules automatically. Additionally, they can be used to segment the images into several anatomical components, such as the liver, brain, and lungs. Medical professionals may be better able to diagnose anomalies in particular body parts thanks to this. The use of deep learning for CT image processing has a number of benefits. Deep learning algorithms can adapt to varied input data types, such as CT images with varying resolutions, noise levels, and contrast, and can learn from vast amounts of data. They can also perform real-time image analysis, which can boost the effectiveness of the diagnostic procedure. The accuracy and consistency of CT imaging analysis can also be improved by deep learning algorithms, lowering the possibility of false-positive and false-negative outcomes.

III. METHODS

A particular kind of neural network that is excellent for image categorization is the convolutional neural network (CNN). The functioning of the human visual brain served as inspiration for this architecture. A neuron mechanism will be mimicked by a CNN filter with a set of receptive fields that only process a limited portion of the image. Each deeper layer of these neurons has a bigger receptive field and, as such, can learn and recognize more complex patterns, which are primarily arranged in hierarchical layers. CNNs can also be seen as numerous layers of sliding windows with tiny neural networks striding across the image.

One of the advantages of CNNs is their location invariance capability, which allows the filters to learn the patterns regardless of where they are located. This is because the use of sliding windows allows the filter to learn the patterns all over the images. Another advantage of CNNs is their hierarchical structure, which enables them to automatically pick up on increasingly abstract patterns. The first layers may pick up on boundaries and structures, while the intermediate layers may pick up on details like shapes, and even higher levels may pick up on general object shapes.

CNNs can be configured to process 3D images rather than the sliced 2D images of the CT scan. 3D CNNs can be implemented by using a sliding cube instead of a sliding pane that advances through three dimensions while extracting features at each step.

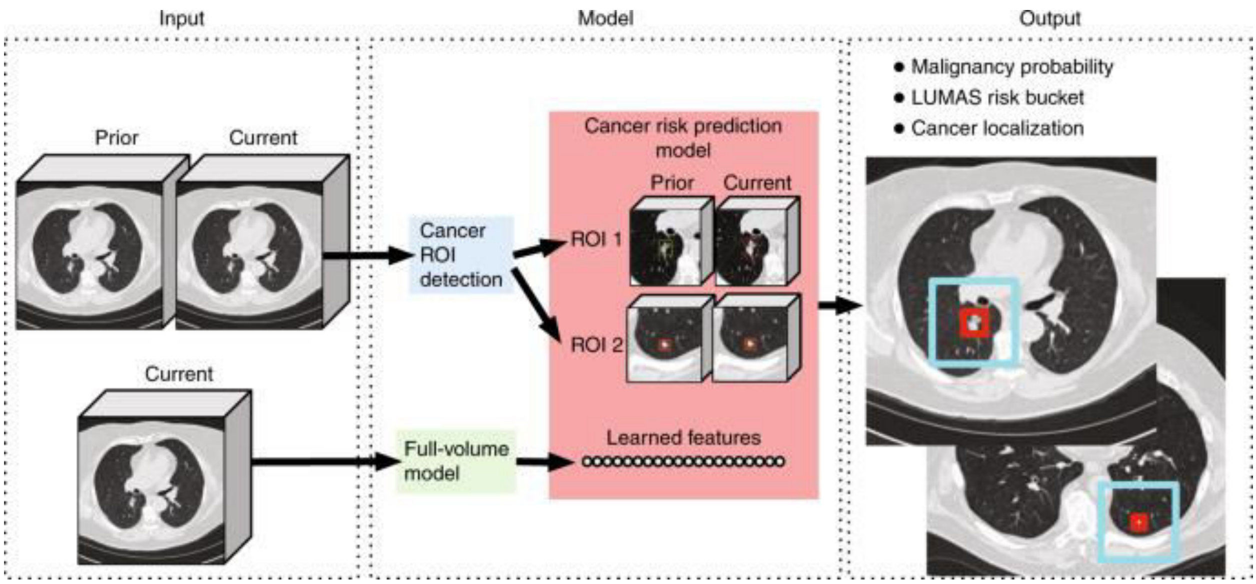


Fig 2: End-to-end lung cancer screening

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 4, April 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1204288 |

Researchers have been investigating the convolutional neural network's robustness for computer vision applications for the past ten years. For natural image processing and medical image analysis, various CNN-based methods have been proposed. The use of artificial intelligence and CT scans to identify and diagnose lung cancer has been suggested in a number of ways. To recognize and categorize lung nodules, for instance, the work in [26] created a three-dimensional convolutional neural network (CNN) with three modules. This technique beats manual assessment performance with a sensitivity rate of 84.4%.

IV. RESULT ANALYSIS

One of the cancers that leads to mortality is lung cancer. Furthermore, it is very challenging to identify the cases because they normally emerge and manifest during the terminal stage. However, early disease detection and treatment tools can lower the mortality rate. Optimum imaging methods such as CT imaging may reveal all suspected and undetected lung cancer nodules, which makes it a trustworthy tool for diagnosing lung cancer. It may be challenging to identify the malignant cells, nevertheless, due to variations in CT scan intensity and anatomical structure misinterpretations by medical professionals and radiologists. Computer-aided diagnosis systems have recently become popular tools to help radiologists and clinicians effectively diagnose cancer. Numerous systems have been created, and research into the detection of lung cancer is still ongoing. However, some systems still need to be improved in order to obtain the best detection accuracy possible, which is going towards 100%.

Lung cancer can be effectively treated with the help of a thorough etiology, reliable early identification, and appropriate medications. Thus, early detection of lung cancer is essential, particularly in screening high-risk populations (such as smokers, those exposed to fumes, those working in oil fields or other toxic environments, etc.) with a pressing need to find new biomarkers. Furthermore, the best lung cancer treatment relies on the precision of the diagnostic tools. Therefore, the need to find sensitive and precise biomarkers for early detection is crucial, and current methods perform lung cancer screening procedures by using low-dose CT (LDCT). Additionally, compared to no screening cases, a study (NELSON) has demonstrated that this specific screening method has a selectivity of 85% and a specificity of 99%. Moreover, according to a recent study, the overall false-positive rate was actually less than 81%, which needs to be confirmed by further imaging or testing due to this extremely high number.

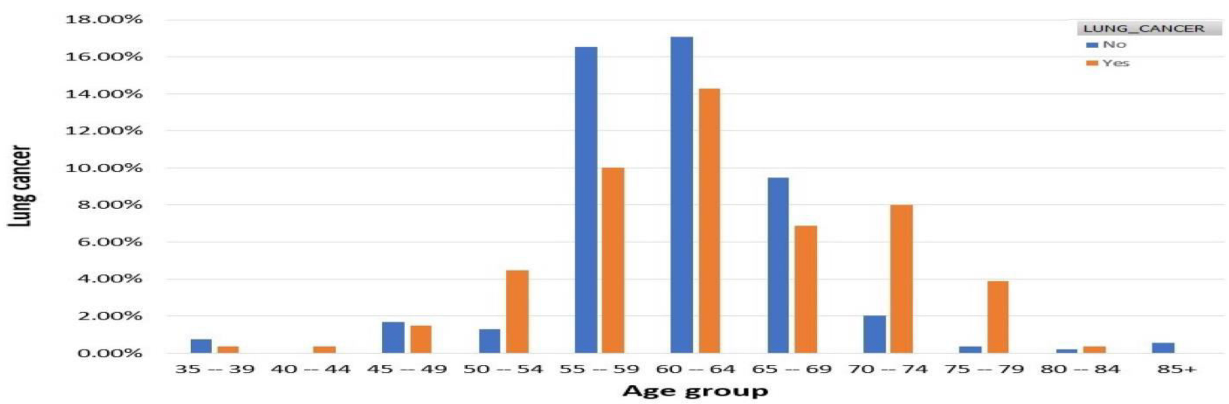


Fig 3: Lung Cancer Risk Prediction

A quick description of lung cancer staging is provided below to clarify the timing of the screening with respect to the development of lung cancer. Small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC) are the two main subtypes of lung cancer. SCLC is a central tumor that manifests as a perihilar mass in the airway submucosa. According to histological investigations, the neuroendocrine cells of the basal bronchial epithelium are the source of this particular type of cancer. Small, spindle-shaped, or rounded cells with little cytoplasm, granular chromatin, and necrosis are the most typical cell types found in this case. The subtypes of SCLC include pure and mixed NSCLC, which can be distinguished from the possibility of brain, liver, and bone metastases and is categorized as having limited or widespread stages. Either the ipsilateral mediastinum, mediastinal, or supraclavicular lymph nodes and one radiation site detected can be used to identify the restricted SCLC stage. As long as it is present on the same side of the cancerous chest, it is regarded as belonging to the category of supraclavicular lymph nodes. On the other hand, widespread SCLC spreads to the second lung lobe, lymph nodes, and additional organs like the bone marrow and is not constrained to a single radiation point in the lung.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204288 |

V. CONCLUSION

The body of knowledge on lung cancer diagnosis and presents several recommendations for further study in the area. The authors stress the requirement for additional lung image data from multiple imaging modalities, such as MRI and ultrasound, in addition to the significance of disclosing private datasets to enable comparison and research collaboration. The segmentation of big solid nodules is one area that has been specifically mentioned. This difficult task needs more investigation. The scientists also suggest creating a lung cancer detection model that can distinguish between early benign nodules and small malignant lesions, which would dramatically improve early identification and treatment. To increase the effectiveness of automated tumor identification, it is also recommended that deep features acquired from lung scan images be combined with additional patient data, including medical history and genetic reports. This all-encompassing method might offer a more precise diagnosis of the illness. The authors suggest using various pre-processing techniques and filters to improve image quality, such as edge-preserving methods and harmony search to improve grayscale image quality. These techniques can aid in improved picture analysis and more trustworthy diagnostic outcomes. A successful proposal for remote lung cancer detection serves as motivation for the authors' further suggestion to investigate the use of cloud computing technology for machine learning-based remote diagnosis of lung cancer. Large volumes of medical data can be processed and analyzed effectively by using cloud computing. The scientists suggest using a cat swarm-optimized deep belief network to extract features from lung medical images since it may perform better in feature extraction and classification tasks.

REFERENCES

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [**Google Scholar**] [**CrossRef**]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [**Google Scholar**] [**CrossRef**]
3. Shah, R.; Sabanathan, S.; Richardson, J.; Mearns, A.; Goulden, C. Results of surgical treatment of stage i and ii lung cancer. *J. Cardiovasc. Surg.* **1996**, *37*, 169–172. [**Google Scholar**]
4. Nesbitt, J.C.; Putnam, J.B., Jr.; Walsh, G.L.; Roth, J.A.; Mountain, C.F. Survival in early-stage non-small cell lung cancer. *Ann. Thorac. Surg.* **1995**, *60*, 466–472. [**Google Scholar**] [**CrossRef**] [**PubMed**]
5. Kuan, K.; Ravaut, M.; Manek, G.; Chen, H.; Lin, J.; Nazir, B.; Chen, C.; Howe, T.C.; Zeng, Z. Deep learning for lung cancer detection: Tackling the kaggle data science bowl 2017 challenge. *arXiv* **2017**, arXiv:1705.09435. [**Google Scholar**]
6. Ciompi, F.; Chung, K.; Gerke, P.K.; Jacobs, C.; Scholten, E.T.; SchaeferProkop, C.; Wille, M.M.W.; Marchianò, A.; Pastorino, U.; van Ginneken, B. Towards automatic pul-monary nodule management in lung cancer screening with deep learning. *Sci. Rep.* **2017**, *7*, 46479. [**Google Scholar**]
7. Sun, W.; Tseng, T.-L.B.; Qian, W.; Zhang, J.; Saltzstein, E.C.; Zheng, B.; Lure, F.; Yu, H.; Zhou, S. Using multiscale texture and density features for near-term breast cancer risk analysis. *Med. Phys.* **2015**, *42 Pt 1*, 2853–2862. [**Google Scholar**]
8. Hossain, M.S.; Muhammad, G. Cloud-Based Collaborative Media Service Framework for HealthCare. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, 858712. [**Google Scholar**]
9. Amin, S.U.; Alsulaiman, M.; Muhammad, G.; Mekhtiche, M.A.; Hossain, M.S. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Gener. Comput. Syst.* **2019**, *101*, 542–554. [**Google Scholar**]
10. Jiang, H.; Ma, H.; Qian, W.; Gao, M.; Li, Y. An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1227–1237. [**Google Scholar**]